Paderborn University

PhD Thesis

Testing Coherence and Identifying Winners in Dueling Bandits: Theory and Algorithms

written by
Björn Haddenhorst

in partial fulfillment of the requirements for the academic degree of
Doctor rerum naturalium (Dr. rer. nat.)

submitted to
Paderborn University
Institute of Computer Science

supervised by
Prof. Dr. Eyke Hüllermeier

November, 2022

Danksagung

Vorab bedanke ich mich bei verschiedenen Personen, die mich auf dem Weg der Promotion begleitet und unterstützt haben.

Zuerst danke ich Prof. Dr. Eyke Hüllermeier. Er hat mir als Doktorvater nicht nur die Promotion an sich ermöglicht, sondern mich oft durch interessante Fragestellungen und fachlichen Input vorangebracht. Zudem hat er mich stets bei meinen Publikationen unterstützt und nicht zuletzt auch immer wieder motiviert.

Ein sehr großer Dank gebührt auch Dr. Viktor Bengs. Er hatte immer ein offenes Ohr und guten Rat für mich, hat mich bei den Veröffentlichungen tatkräftig unterstützt und sich immer wieder viel Zeit für regelmäßige fachliche und gewinnbringende Gespräche genommen. Ich danke Jasmin Brandt für ihre gute Kooperation an unterschiedlichen Projekten und Adil El-Mesaoudi Paul für seine Unterstützung bei der Einarbeitung in das Thema zu Beginn der Promotion.

Des Weiteren danke ich meinen Kolleg:innen und Freund:innen sowie meiner Familie und meiner Freundin Denise für ihre Unterstützung und die schöne Zeit, die ich während der Promotion mit ihnen verbringen durfte.

Zu guter Letzt danke ich der Deutschen Forschungsgemeinschaft (DFG), die meine Promotionsstelle finanziert hat (Projektnummer 317046553).

Danke!

Zusammenfassung

Zahlreiche Lernalgorithmen im stochastischen (Multi)-Dueling-Banditen-Szenario (engl.: (multi-)dueling bandits scenario; (M)DB) erfordern, dass die dem Feedback-Mechanismus zugrunde liegenden Gewinnwahrscheinlichkeiten gewisse Arten von Kohärenz erfüllen. In dieser Arbeit diskutieren wir das Testen derartiger Kohärenzannahmen und führen das Problem des *Testifizierens* des Condorcet Gewinners (engl.: Condorcet winner; CW) in DB ein, als das Problem, den CW zu identifizieren, falls er existiert, und andernfalls Nichtexistenz zu detektieren. Des Weiteren diskutieren wir die Identifikation des verallgemeinerten Condorcet Gewinners (engl.: generalized Condorcet winner; GCW) in MDB unter der Annahme, dass er existiert.

Wir zeigen unter anderem, dass die Kohärenz der Gewinnwahrscheinlichkeiten mit einem Plackett-Luce-Modell in MDB unter der sogenannten Low-Noise-Annahme nicht derart getestet werden kann, dass die erwartete Probenkomplexität (engl.: sample complexity) im schlechtesten Fall endlich ist, und gleiches gilt in DB für diverse Arten stochastischer Transitivität. Im Gegensatz dazu sind sowohl das Testen von schwacher stochastischer Transitivität (engl.: weak stochastic transitivity; WST) als auch das Testifizieren des CW in diesem Sinne möglich.

Für das Testen von WST, die Testifikation des CW als auch die Identifikation des GCW präsentieren wir algorithmische Lösungen im sogenannten *fixed-confidence* Setting und leiten instanzspezifische untere und obere Schranken an die zur Lösung der Probleme benötigten Probenkomplexität her, welche im schlechtesten Fall bis auf logarithmische Faktoren asymptotisch optimal sind. Zusätzlich untersuchen wir, in welchem Maße eine Plackett-Luce-Annahme an den stochastischen Feedback-Mechanismus das Lernproblem vereinfacht.

Abstract

Many learning algorithms in the stochastic (multi-)dueling bandits scenario assume the winning probabilities underlying the environment's feedback mechanism to be appropriately coherent. This thesis approaches the problem of checking the validity of several types of coherence in this regard. Moreover, we introduce the task of *CW identification* in dueling bandits, which consists of identifying the Condorcet winner (CW) if it exists, and detecting non-existence otherwise. Finally, we investigate the problem of identifying the generalized Condorcet winner (GCW) in multi-dueling bandits assuming its existence.

We show that, amongst others, coherence of the winning probabilities with a Plackett-Luce model cannot be tested under the low-noise assumption in multi-dueling bandits within a finite expected sample complexity in the worst case, and the same holds for several types of stochastic transitivity in dueling bandits. In contrast, testing of weak stochastic transitivity (WST) and even CW testification are solvable in this regard.

For all of WST testing, CW testification and GCW identification, we present multiple algorithmic solutions in the active fixed confidence setting and derive instance-dependent sample complexity upper and lower bounds that are in the worst case asymptotically tight up to logarithmic terms. In addition, we investigate to which extent a Plackett-Luce assumption on the probabilistic feedback mechanism simplifies the identification of the GCW.

Contents

List of Publications

The publications, which contributed to this PhD thesis, are:

- Björn Haddenhorst, Viktor Bengs, and Eyke Hüllermeier. On testing transitivity in online preference learning. *Machine Learning*, page 2063–2084, 2021b
- Björn Haddenhorst, Viktor Bengs, Jasmin Brandt, and Eyke Hüllermeier. Testification of Condorcet winners in dueling bandits. In *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)*, 2021a
- Björn Haddenhorst, Viktor Bengs, and Eyke Hüllermeier. Identification of the generalized Condorcet winner in multi-dueling bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 25904–25916, 2021c

List of Symbols

The following list contains those symbols, which are frequently used in this thesis.

Basics

$a \wedge b$	$\min\{a, b\}$
$a \vee b$	$\max\{a, b\}$
$\mathbf{1}_A$	indicator function on the set A
$[m]$	the set $\{1, 2, \ldots, m\}$
$[m]_k$	the set of all subsets of $[m]$ of size k
$(m)_2$	the set $\{(i, j) \in [m] \times [m] \mid i < j\}$
$\langle m \rangle_2$	the set $\{(i, j) \in [m] \times [m] \mid i \neq j\}$
$\mathcal{P}(A)$	the power set of the set A, i.e., $\mathcal{P}(A) = \{S : S \subseteq A\}$
$\mathbb{N}$	the set of natural numbers (without 0), i.e., $\mathbb{N} = \{1, 2, 3, \ldots\}$
$\mathbb{N}_0$	the set of natural numbers including zero, i.e., $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$
$\mathbb{N}_{\geq x}$	the set of all $n \in \mathbb{N}_0$ with $n \geq x$; $\mathbb{N}_{>x}$ analogously
$\mathbb{Z}$	the set of all integers
$\mathbb{R}$	the set of real numbers
$\mathbb{P}$	a probability measure
$\mathbb{E}$	an expectation
$\mathbb{S}_m$	the set of all permutations on $[m]$
α	desired bound on the type I error
β	desired bound on the type II error
γ	desired bound on both the type I and II error
$\vert\vert\cdot\vert\vert$	a norm
$\vert\vert\cdot\vert\vert_\infty$	the infinity norm, i.e., $\vert\vert(z_i)_{i\in I}\vert\vert_\infty = \sup_{i \in I} z_i$
A°	the interior of A
$\overline{A}$	the closure of A
∂A	the boundary of A

Basic Notation of (Multi-)Dueling Bandits

k	the query set size
m	the total number of arms
S	an element from $[m]_k$
Δ_S	set of all $\mathbf{w} = (w_i)_{i\in S} \in [0, 1]^S$ with $\sum_{i \in S} w_i = 1$, where S is a finite set
Δ_S^h	set of all $\mathbf{w} \in \Delta_S$, for which $i \in S$ exists with $\forall j \in S \setminus \{i\} : w_i \geq w_j + h$
Δ_k,Δ_k^h	$\Delta_{[k]}$ resp. $\Delta_{[k]}^h$
$\mathbf{p}$	an element from Δ_k or an element from Δ_S for some $S \in [m]_k$
mode($\mathbf{p}$)	$\mathrm{argmax}_{i\in[k]} p_i$ for $\mathbf{p} = (p_1, \ldots, p_k)$; the term mode($\mathbf{P}(\cdot\vert S)$) is defined accordingly
$h(\mathbf{p})$	$\max\{h \in [0, 1] \,\vert \mathbf{p} \in \Delta_k^h\}$ for $\mathbf{p} \in \Delta_k$
$h(\mathbf{P})$	$\max\{h \in [0, 1] \,\vert \mathbf{P} \in PM_k^m(\Delta^h)\}$
$\Delta_{(m)_2}$	the set of all $\mathbf{v} = (v_{i,j})_{(i,j)\in(m)_2}$ with $\min_{i<j} v_{i,j} \geq 0$ and $\sum_{i<j} v_{i,j} = 1$
$\boldsymbol{n}_t$	the family $((\boldsymbol{n}_t)_S)_{S\in[m]_k}$ that contains, for any $S \in [m]_k$, the number $(\boldsymbol{n}_t)_S$ of queries of S until time t

$\boldsymbol{w}_t$	the family $((\boldsymbol{w}_t)_{i\|S})_{S\in[m]_k, i\in S}$ that contains, for any $S \in [m]_k$ and $i \in S$, the number $(\boldsymbol{w}_t)_{i\|S}$ of times i has won in a multi-duel S until time t
$\boldsymbol{n}_t \geq t'$	short for "$\forall S \in [m]_k : (\boldsymbol{n}_t)_S \geq t'$"
S_t	the query (multi-duel) made at time t
X_{t,S_t}	the feedback observed at time t, i.e., the winner of the multi-duel S_t
$(\boldsymbol{n}_t)_{i,j}$	$(\boldsymbol{n}_t)_{\{i,j\}}$, i.e., the number of duels of arm i against arm j until time t
$(\boldsymbol{w}_t)_{i,j}$	$(\boldsymbol{w}_t)_{i,\{i,j\}}$, i.e., the number of wins of arm i against arm j until time t
$(\hat{\boldsymbol{q}}_t)_{i,j}$	the empirical preference probability that arm i wins against arm j after having seen the t first samples, i.e., $(\hat{\boldsymbol{q}}_t)_{i,j} = {(\boldsymbol{w}_t)_{i,j}}/{(\boldsymbol{n}_t)_{i,j}}$
$X_{i,j}^{[t]}$	$X_{t,\{i,j\}}$ if $S_t = \{i,j\}$, i.e., the outcome of the comparison between i and j at time t if $\{i,j\}$ is queried at time t
π	a sampling strategy
Π	set of all sampling strategies
Π_∞	set of $\pi \in \Pi$, which ensure $\lim_{t\to\infty}(\boldsymbol{n}_t)_S = \infty$ a.s. $\forall S \in [m]_k$
$\Pi_\infty^{\ln\ln}$	set of all $\pi \in \Pi_\infty$, which fulfill $\lim_{t\to\infty} {(\boldsymbol{n}_t)_S}/{\ln\ln t} = \infty$ a.s. $\forall S \in [m]_k$

Asymptotics

Ω	$f \in \Omega(g)$ as $x \to x_0$ if $\liminf_{x\to x_0} \frac{\|f(x)\|}{g(x)} > 0$
$\mathcal{O}$	$f \in \mathcal{O}(g)$ as $x \to x_0$ if $\limsup_{x\to x_0} \frac{\|f(x)\|}{g(x)} < \infty$
Θ	$f \in \Theta(g)$ as $x \to x_0$ if $f \in \Omega(g)$ and $f \in \mathcal{O}(g)$ as $x \to x_0$
o	$f \in o(g)$ as $x \to x_0$ if $\lim_{x\to x_0} \frac{\|f(x)\|}{g(x)} = 0$
$\widetilde{\Omega}$	Modification of Ω, which hides logarithmic factors; i.e., $f \in \widetilde{\Omega}(g)$ as $x \to x_0$ if $f \in \Omega(g')$ as $x \to x_0$ for $g'(x) = g(x)\ln^a(x)$ for some $a \in \mathbb{R}$
$\widetilde{\mathcal{O}}$	Modification of $\mathcal{O}$, which hides logarithmic factors; i.e., $f \in \widetilde{\mathcal{O}}(g)$ as $x \to x_0$ if $f \in \mathcal{O}(g')$ as $x \to x_0$ for $g'(x) = g(x)\ln^a(x)$ for some $a \in \mathbb{R}$
$\widetilde{\Theta}$	$f \in \widetilde{\Theta}(g)$ as $x \to x_0$ if $f \in \widetilde{\mathcal{O}}(g)$ and $f \in \widetilde{\Omega}(g)$ as $x \to x_0$
$\Omega_{\sup}$	$f \in \Omega_{\sup}(g)$ as $x \to x_0$ if $\limsup_{x\to x_0} \frac{\|f(x)\|}{g(x)} > 0$

Instance Classes and Reciprocal Relations

PM_k^m	set of all $\{\mathbf{P}(\cdot\|S)\}_{S\in[m]_k} \subseteq [0,1]^{\binom{m}{k}}$ with $\sum_{j\in S}\mathbf{P}(j\|S) = 1\ \forall S \in [m]_k$
$\mathbf{P}$	an element in PM_k^m
$\mathrm{GCW}(\mathbf{P})$	set of all GCWs of $\mathbf{P}$; if $\|\mathrm{GCW}(\mathbf{P})\| = 1$, it denotes the only element in $\mathrm{GCW}(\mathbf{P})$
$\mathbf{P}(\boldsymbol{\theta})$	that $\mathbf{P} \in PM_k^m(\mathrm{PL})$, which is coherent with the PL model with parameter $\boldsymbol{\theta}$, i.e., $\mathbf{P}(\boldsymbol{\theta})(i\|S) = \theta_i / \left(\sum_{a\in S}\theta_a\right)$ for all $S \in [m]_k, i \in S$
$PM_k^m(\mathrm{X})$	the set of all $\mathbf{P}$, which fulfill the condition(s) X
$\cdot\, PM_k^m(\Delta^h)$	the set $\{\mathbf{P} \in PM_k^m \mid \forall S \in [m]_k : \mathbf{P} \in \Delta_S^h\}$
$\cdot\, PM_k^m(\Delta^0)$	the set $\{\mathbf{P} \in PM_k^m \mid \forall S \in [m]_k : \mathbf{P} \in \Delta_S^0\}$
$\cdot\, PM_k^m(\mathrm{PL})$	the set $\{\mathbf{P} \in PM_k^m \mid \exists \boldsymbol{\theta} \in (0,\infty)^m : \mathbf{P} = \mathbf{P}(\boldsymbol{\theta})\}$
$\cdot\, PM_k^m(\mathrm{GCW})$	the set $\{\mathbf{P} \in PM_k^m \mid \|\mathrm{GCW}(\mathbf{P})\| \geq 1\}$
$\cdot\, PM_k^m(\mathrm{GCW}^*)$	the set $\{\mathbf{P} \in PM_k^m \mid \|\mathrm{GCW}(\mathbf{P})\| = 1\}$
$\cdot\, PM_k^m(h\mathrm{GCW})$	the set of all $\mathbf{P} \in PM_k^m$, for which some $i \in [m]$ exists s.t. for all $S \in [m]_k$ with $i \in [m]$ and all $j \in S$ the inequality $\|\mathbf{P}(i\|S) - \mathbf{P}(j\|S)\| \geq h$ holds
$\mathcal{U}_k^m$	the set $\{\mathbf{P} \in PM_k^m \mid \mathbf{P}(j\|S) > 0$ for all $S \in [m]_k$ and all $j \in S\}$
$\mathcal{Q}_m$	set of reciprocal relations on $[m]$; formally, $\mathcal{Q}_m = PM_2^m$
$\mathbf{Q}$	a relation $\mathbf{Q} = (q_{i,j})_{1\leq i,j\leq m} \in \mathcal{Q}_m$; $q_{i,j}$ is the (unknown) probability that i wins when compared to j
$\bar{q}_{i,j}$	the value $\|q_{i,j} - \frac{1}{2}\|$
$\mathcal{Q}_m^h$	set of all $\mathbf{Q} = (q_{i,j})_{1\leq i,j\leq m} \in \mathcal{Q}_m$ with $\|q_{i,j} - \frac{1}{2}\| > h$ for all $i \neq j \in [m]$

$\sigma_{\mathbf{Q}}$	a permutation σ on $[m]$ s.t. $q_{\sigma(i),\sigma(j)} > \frac{1}{2}$ whenever $i < j$; only defined for $\mathbf{Q} \in \mathcal{Q}_m^0$
$\mathcal{Q}_m(\text{X})$	set of all $\mathbf{Q}$, which fulfill condition(s) X
$\cdot\, \mathcal{Q}_m(\text{CW})$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which have a CW
$\cdot\, \mathcal{Q}_m(i)$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which have i as CW
$\cdot\, \mathcal{Q}_m(\text{XST})$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which fulfill the type of transitivity XST; here, XST $\in \{\text{WST}, \text{MST}, \nu\text{RST}, \text{SST}, \lambda\text{ST}\}$
$\cdot\, \mathcal{Q}_m(\text{STI})$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which fulfill the stochastic triangle inequality
$\cdot\, \mathcal{Q}_m(\text{GIA})$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which fulfill the general identifiability assumption
$\cdot\, \mathcal{Q}_m(\text{LNM})$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which have the property low noise model
$\cdot\, \mathcal{Q}_m(\text{Marg})$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which are marginal of a probability distribution on $\mathbb{S}_m$
$\cdot\, \mathcal{Q}_m(\text{Mal})$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which are marginal of a Mallows probability distribution on $\mathbb{S}_m$
$\cdot\, \mathcal{Q}_m(\text{BS})$	set of all $\mathbf{Q} \in \mathcal{Q}_m$, which are marginal of a Babington-Smith probability distribution on $\mathbb{S}_m$
$\mathcal{Q}_m(\neg X)$	$\mathcal{Q}_m \setminus \mathcal{Q}_m(X)$; here, X can be any property of a reciprocal relation
$\mathcal{Q}_m^h(X)$	$\mathcal{Q}_m^h \cap \mathcal{Q}_m(X)$; here, X can be any property of a reciprocal relation
$\mathcal{Q}_m^{\clubsuit}$	the set $\{\mathbf{Q} \in \mathcal{Q}_m \mid \forall (i,j) \in (m)_2 : q_{i,j} \geq 1/2\}$
$\mathcal{R}_m$	the set of deterministic reciprocal relations, i.e., $\mathcal{R}_m = \{\mathbf{Q} \in \mathcal{Q}_m : q_{i,j} \in \{0,1\}\, \forall (i,j) \in (m)_2\}$
$\mathbf{Q}^{\mathbb{P}}$	the reciprocal relation $\mathbf{Q}^{\mathbb{P}} = (q^{\mathbb{P}}_{i,j})_{1 \leq i,j \leq m} \in \mathcal{Q}_m$ with entries $q^{\mathbb{P}}_{i,j} = \sum_{\sigma \in \mathbb{S}_m \text{ with } \sigma(i) < \sigma(j)} \mathbb{P}(\sigma)$
$G(\mathbf{Q})$	the identifying tournament for a preference relation $\mathbf{Q} \in \mathcal{Q}_m^0$
$\rho(\mathbf{Q})$	the Slater index of $G(\mathbf{Q})$, cf. Def. 5.29 and the discussion thereafter

Graph Theory

G	a digraph
E_G	set of edges of the digraph G
$\mathcal{G}_m$	set of digraphs G on $[m]$ that have at most one edge between any two nodes $i,j \in [m]$
$\overline{\mathcal{G}}_m$	set of tournaments on $[m]$
$\overline{\mathcal{G}}_m(\text{X})$	set of tournaments on $[m]$ that have property X
$\cdot\, \overline{\mathcal{G}}_m(\emptyset)$	the set $\overline{\mathcal{G}}_m$
$\cdot\, \overline{\mathcal{G}}_m(\text{CW})$	set of tournaments on $[m]$, which have a Condorcet winner
$\cdot\, \overline{\mathcal{G}}_m(i)$	set of tournaments on $[m]$, which have i as Condorcet winner
$\cdot\, \overline{\mathcal{G}}_m(\text{acyclic})$	set of tournaments on $[m]$, which do not have a cycle
$\cdot\, \overline{\mathcal{G}}_m(\neg\text{X})$	the set $\overline{\mathcal{G}}_m \setminus \overline{\mathcal{G}}_m(\text{X})$
$\mathcal{G}_m(\text{X}\vert\text{Y})$	the set $\{G \in \mathcal{G}_m \mid$ every Y-extension G' of G fulfills $G' \in \overline{\mathcal{G}}_m(X)\}$
$\mathcal{G}_m(\text{X})$	the set $\mathcal{G}_m(\text{X}\vert\emptyset)$, i.e., the set of all $G \in \mathcal{G}_m$ for which any extension is in $\overline{\mathcal{G}}_m(\text{X})$
$\mathcal{G}_m(\Delta i)$	the set $\mathcal{G}_m(\neg\text{CW}) \cup \bigcup_{j \in [m] \setminus \{i\}} \mathcal{G}_m(j)$
$\mathcal{G}_m(\diamondsuit)$	the set $\bigcup_{i \in [m]} \mathcal{G}_m(i)$
$\text{CW}(G)$	the CW of a tournament $G \in \overline{\mathcal{G}}_m(\text{CW})$
$G_{i \leftrightarrow j}$	digraph defined via $E_{G_{i \leftrightarrow j}} = (E_G \setminus \{(i,j)\}) \cup \{(j,i)\}$
$\{i,j\}_G$	$\{i,j\}_G = (i,j)$ if $i \xrightarrow{G} j$ (which means $(i,j) \in E_G$), otherwise $\{i,j\}_G = (j,i)$; here, $G \in \overline{\mathcal{G}}_m$

Algorithms

$\mathcal{A}$	an algorithm
$\mathcal{A}(x_1, \ldots, x_l)$	an algorithm $\mathcal{A}$ called with the parameters $x_1, \ldots, x_l$
$\mathbf{D}(\mathcal{A})$	the return value/decision of algorithm $\mathcal{A}$

$T^{\mathcal{A}}$	the sample complexity of $\mathcal{A}$, i.e., the number of samples observed by $\mathcal{A}$ before termination
$T^{\mathcal{A}}_{\text{worst}}$	worst-case query complexity of $\mathcal{A}$, cf. Sec. 3.2; only defined if $\mathcal{A}$ is a DSTA
$T^{\mathcal{A}}_{\text{best}}$	best-case query complexity of $\mathcal{A}$, cf. Sec. 3.2; only defined if $\mathcal{A}$ is a DSTA
$\mathcal{A}_{\text{DSTA}}$	a deterministic sequential testing algorithm (DSTA)
$T^{\mathcal{A}}_G$	termination time of the DSTA $\mathcal{A}$ when started on G
$\mathfrak{A}_{\text{Coin}}$	the set of all algorithms for the coin tossing problem (2.7)
$\mathcal{A}_{\text{Coin}}$	an algorithm for the coin tossing problem (2.7)
$\cdot\, \mathcal{A}^{\text{Hoeffding}}_{\text{Coin}}(h,\gamma)$	the non-sequential solution to $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ from Lem. 2.10
$\cdot\, \mathcal{A}^{\text{SPRT}}_{\text{Coin}}(h,\gamma)$	the solution to $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ from Prop. 2.17
$\cdot\, \mathcal{A}^{\text{Farrell}}_{\text{Coin}}(\gamma)$	the solution to $\mathcal{P}^{\gamma}_{\text{Coin}}$ from Prop. 2.22
$\mathbf{D}(\mathcal{A}_{\text{Coin}}, t)$	$\mathbf{D}(\mathcal{A}_{\text{Coin}})$ if $\mathcal{A}_{\text{Coin}}$ has already terminated at after t samples, otherwise "N/A"
$i^{\mathcal{A}}_G(t), j^{\mathcal{A}}_G(t)$	distinct items compared by $\mathcal{A}$ at time t when started on G
$\mathfrak{G}^{\mathcal{A}}_G(t)$	the *picture* that $\mathcal{A}$ has of G at time t; formally defined via $E_{\mathfrak{G}^{\mathcal{A}}_G(t)} = \bigcup_{t' \leq t-1} \{\{i^{\mathcal{A}}_G(t'), j^{\mathcal{A}}_G(t')\}_G\}$
$\mathbf{D}^{\mathcal{A}}(G)$	output of the DSTA $\mathcal{A}$ when started on G

Probability Theory

$\text{Ber}(p)$	Bernoulli distribution with success probability $p \in [0,1]$
$\text{Bin}(n,p)$	Binomial distribution with success probability $p \in [0,1]$ and number of trials $n \in \mathbb{N}$
$\text{Cat}(\mathbf{p})$	Categorical distribution with parameter $\mathbf{p} = (p_1, \ldots, p_k) \in \Delta_{[k]}$
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathcal{U}(I)$	Uniform distribution on the set I; usually, I is a closed interval of $\mathbb{R}$
δ_x	Dirac measure on $\{x\}$, i.e., for any $A \subseteq \mathbb{R}$ we have $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise
$\chi^2_{(k)}$	χ^2-distribution with k degrees of freedom
$\text{KL}(\mathbf{p}, \mathbf{q})$	the Kullback-Leibler divergence of two independent random variables $X \sim \text{Cat}(\mathbf{p})$ and $Y \sim \text{Cat}(\mathbf{q})$ for $\mathbf{p}, \mathbf{q} \in \Delta_S$
$\text{kl}(p, q)$	the Kullback-Leibler divergence between two independent random variables $X \sim \text{Ber}(p)$ and $Y \sim \text{Ber}(q)$, i.e., $\text{kl}(p,q) = \text{KL}((p, 1-p), (q, 1-q))$

Problems

$\mathcal{P}^{\text{P}}_{\text{T}}(\text{A})$	problem to solve task T with parameters P for any instance fulfilling A
$\cdot\, \mathcal{P}^{\Theta_0,\Theta_1;\gamma}_{\text{Coin}}$	problem to test $\mathbf{H}_0 : p \in \Theta_0$ vs. $\mathbf{H}_1 : p \in \Theta_1$ with confidence $1-\gamma$ based on iid samples $\sim \text{Ber}(p)$
$\cdot\, \mathcal{P}^{p_0,p_1;\gamma}_{\text{Coin}}$	short for $\mathcal{P}^{\{p_0\},\{p_1\};\gamma}_{\text{Coin}}$
$\cdot\, \mathcal{P}^{p_0;\gamma}_{\text{Coin}}$	short for $\mathcal{P}^{p_0,1-p_0;\gamma}_{\text{Coin}}$
$\cdot\, \mathcal{P}^{h,\gamma}_{\text{Coin}}$	short for $\mathcal{P}^{[0,1/2-h),(1/2+h,1];\gamma}_{\text{Coin}}$
$\cdot\, \mathcal{P}^{\gamma}_{\text{Coin}}$	short for $\mathcal{P}^{0,\gamma}_{\text{Coin}} = \mathcal{P}^{[0,1/2),(1/2,1];\gamma}_{\text{Coin}}$
$\cdot\, \mathcal{P}^{k,h,\gamma}_{\text{Die}}$	problem to identify for any $\mathbf{p} \in \Delta^h_k$ with error probability $\leq \gamma$ correctly $\text{mode}(\mathbf{p})$ based on iid samples $\sim \text{Cat}(\mathbf{p})$; formally, $\mathcal{P}^{k,h,\gamma}_{\text{Die}} = \mathcal{P}^{k,k,\gamma}_{\text{GCWi}}(\Delta^h)$
$\cdot\, \mathcal{P}^{m,k,\alpha,\beta}_{\text{PL}}(\Delta^h)$	PL testing on $PM^m_k(\Delta^h)$ for α, β, cf. Sec. 2.5.1
$\cdot\, \mathcal{P}^{m,h,\alpha,\beta}_{\text{X}}$	X testing on $\mathcal{Q}^h_m$ for α, β; considered in Sec. 2.5.1 for X $\in$ {STI, GIA, LNM, Marg, Mal, BS}
$\cdot\, \mathcal{P}^{m,h,\alpha,\beta}_{\text{CWv}}$	CW verificiation on $\mathcal{Q}^h_m$ for α, β
$\cdot\, \mathcal{P}^{m,h,\alpha,\beta}_{\text{CWt}}$	CW testificiation on $\mathcal{Q}^h_m$ for α, β
$\cdot\, \mathcal{P}^{m,h,\alpha,\beta}_{\text{CWc}}$	CW checking on $\mathcal{Q}^h_m$ for α, β
$\cdot\, \mathcal{P}^{m,h,\alpha,\beta}_{\text{CWi}}(\text{CW})$	CW identification on $\mathcal{Q}^h_m(\text{CW})$ for α, β

$\cdot\, \mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWv}}(\mathrm{CW})$	CW verificiation on $\mathcal{Q}^h_m(\mathrm{CW})$ for α, β
$\cdot\, \mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{XST}}$	XST testing on $\mathcal{Q}^h_m$ for α, β; defined for XST $\in$ {WST, MST, SST, νRST, λST}
$\cdot\, \mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{X})$	GCW identification on $PM^m_k(\mathrm{X})$ for γ
$\cdot\, \mathcal{P}^{m,k,\gamma}_{\mathrm{GCWv}}(\mathrm{X})$	GCW identification on $PM^m_k(\mathrm{X})$ for γ
$\mathcal{D}^{\mathrm{P}}_{\mathrm{T}}(\mathrm{A})$	deterministic problem to solve task T with parameters P for any instance fulfilling A
$\cdot\, \mathcal{D}^m_{\mathrm{X_1,\dots,X_k}}[\mathcal{Z}](\mathrm{Y})$	problem to deterministically assign any tournament $G \in \overline{\mathcal{G}}_m(\mathrm{Y})$, given input $z \in \mathcal{Z}$, correctly one of the classes $\mathrm{X}_1(z), \dots, \mathrm{X}_k(z)$
$\cdot\, \mathcal{D}^m_{\mathrm{CWt}}$	CW testification for tournaments $G \in \overline{\mathcal{G}}_m$
$\cdot\, \mathcal{D}^m_{\mathrm{CWc}}$	CW checking for tournaments $G \in \overline{\mathcal{G}}_m$
$\cdot\, \mathcal{D}^m_{\mathrm{CWv}}$	CW verification for tournaments $G \in \overline{\mathcal{G}}_m$
$\cdot\, \mathcal{D}^m_{\mathrm{CWi}}(\mathrm{CW})$	CW identification for tournaments $G \in \overline{\mathcal{G}}_m(\mathrm{CW})$
$\cdot\, \mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$	CW verification for tournaments $G \in \overline{\mathcal{G}}_m(\mathrm{CW})$
$\cdot\, \mathcal{D}^m_{\mathrm{acyclic}}$	acyclicity testing for tournaments $G \in \overline{\mathcal{G}}_m$

Notation Related to the Analysis of Sticky Track-and-Stop

$\Delta^{\varepsilon}_{(m)_2}$	the set $\left\{(v_{i,j})_{1\le i<j\le m} \in \Delta_{(m)_2} \,:\, v_{i,j} \ge \varepsilon \text{ for all } (i,j) \in (m)_2\right\}$
$D(\mathbf{v}, \mathbf{Q}, \mathbf{Q}')$	$\sum_{(i,j)\in(m)_2} v_{i,j} \mathrm{kl}\left(q_{i,j}, q'_{i,j}\right)$
$D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}'_m)$	$\inf_{\mathbf{Q}'\in\mathcal{Q}'_m} D(\mathbf{v}, \mathbf{Q}, \mathbf{Q}')$
$D^{m,h}_{\mathrm{CWc}}(\mathbf{Q})$	$\sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}^h_m(\neg\mathrm{X}))$ if $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{X})$, $\mathrm{X} \in \{\mathrm{CW}, \neg\mathrm{CW}\}$
$D^{m,h}_{\mathrm{CWt}}(\mathbf{Q})$	$\sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}^h_m(\neg\mathrm{X}))$ if $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{X})$, $\mathrm{X} \in \{\neg\mathrm{CW}, 1, \dots, m\}$
$D^{m,h}_{\mathrm{WST}}(\mathbf{Q})$	$\sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}^h_m(\neg\mathrm{X}))$ if $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{X})$, $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$
$d_h(\mathbf{Q})$	$\max_{(i,j)\in(m)_2} \max\{\mathrm{kl}(q_{i,j}, 1/2+h), \mathrm{kl}(q_{i,j}, 1/2-h)\}$

Notation Related to the Likelihood-Ratio Test for WST

ϕ	the transformation $[0,1] \to [-\pi/2, \pi/2]$, $x \mapsto 2\arcsin(\sqrt{x}) - \pi/2$
$\boldsymbol{\theta}$	transformed ground-truth $\boldsymbol{\theta} = (\theta_{i,j})_{1\le i,j\le m} = (\phi(p_{i,j}))_{1\le i,j\le m}$
$\boldsymbol{z}_t$	transformed data vector with entries $(\boldsymbol{z}_t)_{i,j} = \phi((\boldsymbol{w}_t)_{i,j}/(\boldsymbol{n}_t)_{i,j})$
Θ_m	transformed parameter space $\Theta_m = \phi(\mathcal{Q}_m) = [-\pi/2, \pi/2]^{\binom{m}{2}}$
$\Theta_m(\mathrm{WST})$	the set $\overline{\phi(\mathcal{Q}_m(\mathrm{WST}))}$
$\Theta_m(\neg\mathrm{WST})$	the set $\overline{\phi(\mathcal{Q}_m(\neg\mathrm{WST}))}$
Θ^0_m	the set $\phi(\mathcal{Q}^0_m) = \{\boldsymbol{\theta} \in \Theta_m \,\vert\, \forall (i,j) \in (m)_2 : \theta_{i,j} \neq 0\}$
$\Theta^0_m(\mathrm{X})$	the set $\Theta^0_m \cap \Theta_m(\mathrm{X})$ for $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$
Θ^v_m	the set $\{\boldsymbol{\theta} \in \Theta_m \,\vert\, \forall (i,j) \in (m)_2 : \theta_{i,j} = 0 \text{ or } \vert\theta_{i,j}\vert > v\}$
$\Theta_m(\mathrm{X})^\circ$	the interior of $\Theta_m(\mathrm{X})$ w.r.t. the (from $\mathbb{R}^{m(m-1)/2}$) induced topology on Θ_m
$\Theta_m(\boldsymbol{\theta})$	the set $\{\boldsymbol{y} \in \Theta_m \,\vert\, \theta_{i,j} < 0 \;\Rightarrow\; y_{i,j} < 0 \text{ for every distinct } i,j \in [m]\}$
$\mathcal{L}$	likelihood function for $\mathbf{H}_0 : \mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST})$ vs. $\mathbf{H}_1 : \mathbf{Q} \in \mathcal{Q}_m(\mathrm{WST})$
$\tilde{\mathcal{L}}$	slightly modified version of $\mathcal{L}$
λ_t	LRT statistic for $\mathbf{H}_0 : \mathbf{Q} \in \mathcal{Q}_m(\mathrm{WST})$ vs. $\mathbf{H}_1 : \mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST})$
μ_t	LRT statistic for $\mathbf{H}_0 : \mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST})$ vs. $\mathbf{H}_1 : \mathbf{Q} \in \mathcal{Q}_m(\mathrm{WST})$
$\tilde{\lambda}_t$	LRT statistic for $\mathbf{H}_0 : \boldsymbol{\theta} \in \Theta_m(\mathrm{WST})$ vs. $\mathbf{H}_1 : \boldsymbol{\theta} \in \Theta_m \setminus \Theta_m(\mathrm{WST})$
$\tilde{\mu}_t$	LRT statistic for $\mathbf{H}_0 : \boldsymbol{\theta} \in \Theta_m(\neg\mathrm{WST})$ vs. $\mathbf{H}_1 : \boldsymbol{\theta} \in \Theta_m \setminus \Theta_m(\neg\mathrm{WST})$
$\mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}, \boldsymbol{\theta})$	weighted Euclidean distance from $\boldsymbol{z}$ to $\boldsymbol{\theta}$ with weights $\boldsymbol{n}_t = ((\boldsymbol{n}_t)_{i,j})_{1\le i,j\le m}$
$\mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}, \Theta_m(\mathrm{X}))$	$\inf_{\boldsymbol{\theta}\in\Theta_m(\mathrm{X})} \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}, \boldsymbol{\theta})$; here, $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$.
$l_{\mathrm{WST}}(t)$	decision boundary for the LRT for WST; cf. Thm. 5.13
$l_{\neg\mathrm{WST}}(t)$	decision boundary for the LRT for ¬WST; cf. Thm. 5.13
$\Theta_m(\mathbf{R})$	the set $\{\boldsymbol{\theta} \in \Theta_m \,\vert\, \forall \text{ distinct } i,j \in [m] : r_{i,j} = 1 \;\Rightarrow\; \theta_{i,j} \ge 0\}$
$\Theta_m(\boldsymbol{\theta})$	the set $\{\boldsymbol{y} \in \Theta_m \,\vert\, \forall \text{ distinct } i,j \in [m] : \theta_{i,j} < 0 \;\Rightarrow\; y_{i,j} < 0\}$

$\Theta_m(\boldsymbol{\theta} \wedge \mathbf{R})$	the set $\Theta_m(\boldsymbol{\theta}) \cap \Theta_m(\mathbf{R})$
$\mathcal{R}_m(\boldsymbol{\theta})$	the set $\{\mathbf{R} \in \mathcal{R}_m \,\vert\, \forall$ distinct $i, j \in [m] : \theta_{i,j} < 0 \Rightarrow r_{i,j} = 0\}$
$\Psi(\boldsymbol{\theta})$	the set $\Psi(\boldsymbol{\theta}) = \{(i, j) \in (m)_2 : \theta_{i,j} = 0\}$
$\psi(\boldsymbol{\theta})$	the size of $\Psi(\boldsymbol{\theta})$, i.e., $\psi(\boldsymbol{\theta}) = \vert\Psi(\boldsymbol{\theta})\vert$

List of Abbreviations

a.s.	almost surely
CW	Condorcet winner
DB	dueling bandits
DKW	Dvoretzky-Kiefer-Wolfowitz (inequality)
DKWT	Dvoretzky-Kiefer-Wolfowitz tournament
DSTA	deterministic sequential testing algorithm
iid	independent and identically distributed
λST	λ-stochastic transitive
GCW	generalized Condorcet winner
GIA	general identifiability assumption
GSPRT	generalized sequential probability ratio test
LIL	law of the iterated logarithm
LNM	low noise model
LRT	likelihood ratio test
MST	moderately stochastic transitive
MAB	multi-armed bandits
MDB	multi-dueling bandits
NTS	noisy tournament sampling
νRST	ν-relaxed stochastic transitive
PB-MAB	preference-based multi-armed bandit; also called *dueling bandit*
pdf	probability distribution function
PL	Plackett-Luce model/assumption
SPRT	sequential probability ratio test
SST	strongly stochastic transitive
STI	stochastic triangle inequality
w.h.p.	with high probability
WST	weakly stochastic transitive

List of Algorithms

1. Introduction

This thesis deals with two learning tasks in the realm of stochastic (multi-)dueling bandits, namely so-called best-arm identification and the problem to statistically test for various types of coherence in the underlying environment. While identifying the best arm is now a problem that is attracting a great deal of research interest in bandit learning scenarios, the aforementioned testing problems have not yet been considered.

We start with the introduction of the considerd (multi-)dueling bandits setting and then discuss different statistical assumptions common in this field. At this point, we are trying to provide already enough details necessary for a simplified overview of the theoretical results achieved in this thesis. Afterwards, we provide a short but far from complete literature overview of multi-armed bandits and its (multi-)dueling variants prior to end this introduction with some remarks on notational conventions made in this thesis.

1.1. The (Multi-)Dueling Bandits Setting

The stochastic *multi-armed bandits* (MAB) scenario is an often applicable learning scenario for sequential decision making problems, which has gained much attraction in recent years. It has its origins in [Thompson, 1933, Robbins, 1952].

In its standard form, it involves a set of m actions, that are indexed by $1, \ldots, m$ for convenience, a learner may choose from at each time step $t \in \mathbb{N} := \{1, 2, 3, \ldots\}$ upon which a stochastic real-valued feedback $X_{t,i(t)}$ for its action $i(t) \in [m] := \{1, \ldots, m\}$ is observed. The name of the setting comes from a prominent illustration thereof, which interprets each action as a possible pull of one of m many *(one-armed) bandits* – another term for a slot machine – and the feedback can be seen as a payoff, called *reward*, obtained by playing the chosen slot machine. For this reason, the different actions are also referred to as *arms* and one uses the phrase *to pull* an arm for saying that the corresponding action is chosen. In general, $X_{t,i}$ can be thought of as some information obtained for arm i at time t.

Yue and Joachims [2009] introduced *dueling bandits* (DB) as a preference-based variant of MABs, in which the learner chooses in each time step t not one but two distinct arms $i(t), j(t) \in [m]$ and then observes as feedback only one of the arms, which can be thought of as the *winner* of the *duel* $\{i(t), j(t)\}$ and in this sense as preferred over the other arm. Later on, this setting was generalized to the *multi-dueling bandits* (MDB) setting [Brost et al., 2016, Bengs et al., 2021] in the sense that the learner is supposed to choose at each time step t a non-empty subset S_t of all arms and observes a feedback X_{t,S_t}. The set S_t is referred to as a *query (set)* and, in analogy to DB, also as a *multi-duel* and supposed to be an element of the *set of allowed multi-duels* $\emptyset \neq \mathcal{S} \subseteq \mathcal{P}([m]) = \{S : S \subseteq [m]\}$.

The focus of this thesis is on the (multi-)dueling bandits scenario with *winner feedback without ties* meaning that, whenever a (multi-)duel is conducted, exactly one of the involved arms is observed as the winner, cf. Fig. 1 for an illustration. For simplicity, we

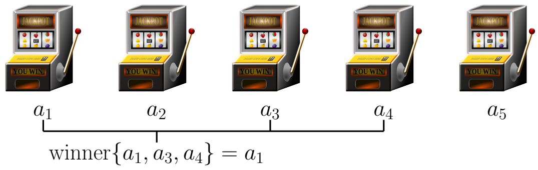


Figure 1.: Illustration of a MDB with winner feedback.

restrict ourselves to the homogeneous case

$$\mathcal{S} = [m]_k \coloneqq \{S \subseteq [m] \, : \, |S| = k\} \, ,$$

where the allowed query sets are the k-sized subsets of $[m]$ for some predefined $k \geq 2$. Throughout, we consider the *time-stationary stochastic* setting, in which the winner X_{t,S_t} of the t-th multi-duel S_t is supposed to be drawn from an unkown underlying categorical distribution $\mathrm{Cat}(\mathbf{P}(\cdot|S_t))$ with values in S_t. Formally, we suppose $\{X_{t,S}\}_{t\in\mathbb{N},S\in\mathcal{S}}$ to be an independent family of categorical random variables $X_{t,S} \sim \mathrm{Cat}(\mathbf{P}(\cdot|S))$ with parameter $\mathbf{P}(\cdot|S)$, of which the *learner* observes at time t only X_{t,S_t}, i.e., the feedback is supposed to be independent across time and query sets. This way, the family $\mathbf{P} = \{\mathbf{P}(i|S)\}_{S\in\mathcal{S},i\in S} \subseteq [0,1]$, which consists of the probabilities $\mathbf{P}(i|S)$ that i is the winner in the query set S, is an underlying parameter, that completely characterizes the stochastic feedback mechanism over time. We call $\mathbf{P}$ also a *probability model* and write PM_k^m for the set of all such probability models when $\mathcal{S} = [m]_k$, i.e., PM_k^m is given as

$$\left\{\mathbf{P} = \{\mathbf{P}(i|S)\}_{S\in[m]_k, i\in[m]} \,|\, \forall S \in [m]_k : \{\mathbf{P}(i|S)\}_{i\in S} \subseteq [0,1] \text{ and } \sum\nolimits_{i\in S} \mathbf{P}(i|S) = 1\right\}.$$

Any $\mathbf{P} \in PM_2^m$ can be written via $q_{i,j} \coloneqq \mathbf{P}(i|\{i,j\}) \in [0,1]$ as $\mathbf{Q} = (q_{i,j})_{1\leq i,j\leq m}$ and fulfills $q_{i,j} = 1 - q_{j,i}$ for all distinct $i,j \in [m]$ and w.l.o.g. $q_{i,i} = 1/2$ for all $i \in [m]$. This shows, that $\mathcal{Q}_m \coloneqq PM_2^m$ is the set of *reciprocal (preference) relations* $\mathbf{Q} = (q_{i,j})_{1\leq i,j\leq m}$ on $[m]$, and that any $\mathbf{Q} \in \mathcal{Q}_m$ is completely characterized by $(q_{i,j})_{1\leq i<j\leq m}$ and may thus conveniently be written as

$$\mathbf{Q} = \begin{pmatrix} - & q_{1,2} & \cdots & q_{1,m} \\ & \ddots & \ddots & \vdots \\ & & - & q_{m-1,m} \\ & & & - \end{pmatrix}.$$

The goals and assumptions made in the (multi-)dueling bandits setting are typically defined in terms of $\mathbf{P}$ resp. $\mathbf{Q}$, and we collect some of these in Sec. 1.2 below. Two fundamentally different ways how the queries S_t are chosen lead to the following two different learning scenarios:

- In the *active scenario*, the learner itself chooses the (multi-)duels at each time step. In particular, by trying to choose them in a favourable way, it may try to come as early as possible to a decision for its learning problem. Typical research questions in this scenario are “What is a good query strategy for choosing the (multi-)duels?” and “Which sample complexity is necessary/sufficient to solve the learning problem at hand?”.

- In the *passive scenario* instead, the learner can*not* choose the (multi-)duels itself, but instead the queries are supposed to be given by the environment or an external force. In this case, a goal of the learner could be to correctly solve the learning task for a huge class of environmental strategies, and its performance could be measured in dependence of a particular query strategy of the environment.

Let us write $(\boldsymbol{n}_t)_S$ for the number of times S has been queried until time t, which is formally given as $(\boldsymbol{n}_t)_S = \sum_{s=1}^{t} \mathbf{1}_{\{S_s=S\}}$. Here and throughout, $\mathbf{1}_{\{A\}}$ denotes the indicator function that is 1 if A is true and 0 otherwise. Moreover, we write $(\boldsymbol{w}_t)_{i|S}$ for the number of times arm $i \in S$ has been observed as a winner of S until time t, i.e., $(\boldsymbol{w}_t)_{i|S} = \sum_{s=1}^{t} \mathbf{1}_{\{S_s=S\}} \mathbf{1}_{\{X_{s,S_s}=i\}}$ and $(\boldsymbol{n}_t)_S = \sum_{i \in S} (\boldsymbol{w}_t)_{i|S}$. Then, we abbreviate $\boldsymbol{n}_t = ((\boldsymbol{n}_t)_S)_{S \in [m]_k}$ and $\boldsymbol{w}_t = ((\boldsymbol{w}_t)_{i|S})_{S \in [m]_k, i \in S}$.

To capture both the active and passive scenario in one framework, we define a *sampling strategy* to be a family of random mappings, which, depending on the time t and the observations $\boldsymbol{n}_0, \boldsymbol{w}_0, \ldots, \boldsymbol{n}_{t-1}, \boldsymbol{w}_{t-1}$ available before time t, determines a multi-duel $S_t \in [m]_k$ to be queried at time $t \in \mathbb{N}$. Let Π denote the set of all sampling strategies.

In the active scenario, the learner itself chooses the sampling strategy π. In the passive one, instead, π is supposed to be given by the environment and possibly unknown to the learner. In any case, the learner sequentially observes at each time the outcome of the duel queried by π and can decide after any observation (also called *sample*) either to stop the learning process and output a decision or to continue. Apart from providing a "good" decision for its learning task (which requires having seen enough samples), the learner typically tries to minimize the total number of observations before termination, the so-called *sample complexity*.

For the theoretical analysis, we will focus in the passive scenario at times on the family $\Pi_\infty \subsetneq \Pi$ of sampling strategies π, which sample every allowed query set $S \in [m]_k$ almost surely (a.s.) infinitely often, i.e., $(\boldsymbol{n}_t)_S \to \infty$ a.s. as $t \to \infty$. If $\pi \in \Pi \setminus \Pi_\infty$, then a sampling strategy $\hat{\pi} \in \Pi$ that chooses the same pair as π in each time step with probability $1 - \frac{1}{t}$, and otherwise (i.e., with probability $\frac{1}{t}$) picks a query set S_t uniformly at random from $[m]_k$ fulfills $\hat{\pi} \in \Pi_\infty$ and

$$\mathbb{P}\big(\pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0 \le t' \le t-1}) \neq \hat{\pi}(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0 \le t' \le t-1})\big) \le \frac{1}{t} \to 0 \text{ as } t \to \infty\,.$$

Thus, $\hat{\pi}$ and π behave similarly in the limit. This shows that the assumption $\pi \in \Pi_\infty$, which will be required for some of our theoretical results below, is rather mild.

1.2. Modeling Assumptions in (Multi-)Dueling Bandits

There are various learning tasks in (M)DB such as the identification of a best arm (also called *winner*), and the targets typically depend on the underlying unknown parameters $\mathbf{P}$ resp. $\mathbf{Q}$. To simplify a learning problem or even assure that a task can be solved at all, many works assume some type of coherence of these parameters. For example, winner identification is typically considered in scenarios, in which $\mathbf{P}$ resp. $\mathbf{Q}$ is coherent with the existence of a winner. In this section, we specify the theoretical learning targets and frequently assumed coherences in that setting, which we mainly focus on in this thesis. We start with a detailed description of those, which are relevant for the further course of this thesis, and mention in Sec. 1.4 further alternative concepts from related literature.

The Condorcet Winner and a Generalized Variant An often targeted learning task in the context of multi-armed bandits and its variants is the problem of identifying the best among all arms. While for standard MABs, the canonical definition of the "best arm" is that arm which provides the best feedback (e.g. the largest expected reward), the picture is less clear for its variants. In the realm of dueling bandits, a reasonable notion of "best arm" is the *Condorcet winner* (CW), which is an arm that is likely to win (i.e., with probability $\geq 1/2$) in each duel against another arm, i.e., formally $i \in [m]$ is the CW of $\mathbf{Q}$ (write $i = \mathrm{CW}(\mathbf{Q})$) if

$$q_{i,j} \geq 1/2 \quad \text{for every } j \in [m] \setminus \{i\}.$$

We write $\mathcal{Q}_m(\mathrm{CW}) := \{\mathbf{Q} \in \mathcal{Q}_m \,|\, \exists i \in [m] : i \text{ is a CW of } \mathbf{Q}\}$ resp. $\mathcal{Q}_m(\neg\mathrm{CW}) := \mathcal{Q}_m \setminus \mathcal{Q}_m(\mathrm{CW})$ for the set of reciprocal relations with resp. without a CW and further, for $i \in [m]$, $\mathcal{Q}_m(i)$ for the set of all $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{CW})$ with $\mathrm{CW}(\mathbf{Q}) = i$.

In practice, the Condorcet winner does not necessarily exist due to a possible presence of preferential cycles in the probabilistic model in the sense that i is likely to win against j, j against k, and k against i. In other words, $\mathbf{Q}$ might be incoherent with the existence of a CW. For the theoretical analysis of the best-arm identification problem, this issue is overcome in the literature either by simply assuming the existence of the CW or by the consideration of alternative optimality concepts such as *Borda winner* or *Copeland winner*, which will briefly be commented on in Sec. 1.4.

For the multi-dueling bandits setting, we consider the generalization of the CW as in [Agarwal et al., 2020], i.e., an arm $i \in [m]$ is called a *generalized Condorcet winner* (GCW) of $\mathbf{P} \in PM_k^m$ if it outperforms any other arm j in every query set S containing both i and j in the sense that

$$\forall S \in [m]_k \text{ with } i \in S, \forall j \in S \setminus \{i\} \,:\, \mathbf{P}(i \,|\, S) - \mathbf{P}(j \,|\, S) \geq 0.$$

We write $\mathrm{GCW}(\mathbf{P})$ for the set of all GCWs of $\mathbf{P}$ and define

$$PM_k^m(\mathrm{GCW}) := \{\mathbf{P} \in PM_k^m \,\big|\, \mathrm{GCW}(\mathbf{P}) \neq \emptyset\},$$
$$PM_k^m(\mathrm{GCW}^*) := \{\mathbf{P} \in PM_k^m \,\big|\, |\mathrm{GCW}(\mathbf{P})| = 1\}.$$

In addition, for $h \in (0, 1]$, we call $i \in [m]$ an hGCW of $\mathbf{P}$ if

$$\forall S \in [m]_k \text{ with } i \in S, \forall j \in S \setminus \{i\} \,:\, \mathbf{P}(i \,|\, S) - \mathbf{P}(j \,|\, S) \geq h,$$

and write $PM_k^m(h\mathrm{GCW})$ for the set of all $\mathbf{P} \in PM_k^m$ that have an hGCW.

With a look at the definitions, we directly observe that (a) any hGCW is a GCW, (b) if an hGCW exists, it is unique and (c) if $\mathbf{P} \in PM_k^m(h\mathrm{GCW})$ for some $h > 0$, the GCW of $\mathbf{P}$ is unique and coincides with the hGCW of $\mathbf{P}$. Clearly, it holds that $PM_k^m(\mathrm{GCW}^*) = \bigcup_{h>0} PM_k^m(h\mathrm{GCW})$ and every probability model $\mathbf{P} \in PM_k^m(\mathrm{GCW})$ has at least one GCW, while for any $\mathbf{P} \in PM_k^m(\mathrm{GCW}^*)$ the GCW is unique. Regarding the dueling bandits setting as the multi-dueling setting where the allowed multi-duels S are exactly those with $|S| = 2$, the GCW is indeed a generalization of the CW.

The Low-Noise Assumption In the dueling bandits setting, the term $\bar{q}_{i,j} := |q_{i,j} - 1/2|$ can be seen as a noise (or hardness) parameter for the outcomes of the duels between i

and j: If this value is large, the outcomes of the pairwise comparisons are relatively clear in the sense that one of i and j will clearly outperform the other one in a large portion of samples, and thus it is relatively "easy" to decide for the learner whether $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ holds. A small value of $\bar{q}_{i,j}$ instead implies that the winning probabilities of i and j are close to 1/2, and hence difficult to distinguish. Motivated by this, we define similar as [Braverman et al., 2016, Korba et al., 2017] that, for $h \in [0, 1/2)$, some $\mathbf{Q} \in \mathcal{Q}_m$ fulfills the *low-noise assumption with parameter h* if it is an element of

$$\mathcal{Q}_m^h := \left\{ \mathbf{Q} = (q_{i,j})_{1 \le i,j \le m} \in \mathcal{Q}_m \,|\, |q_{i,j} - 1/2| > h \text{ for all } (i,j) \in (m)_2 \right\}.$$

For the case of multi-dueling bandits, we introduce the following analogon of the low-noise assumption: We say that $\mathbf{P} = \{\mathbf{P}(i|S)\}_{S \in [m]_k, i \in S} \in PM_k^m$ fulfills the *low-noise assumption with parameter* $h \in (0, 1]$ if

$$\forall S \in [m]_k \, \exists i \in S \, \forall j \in S \setminus \{i\} : \mathbf{P}(i|S) - \mathbf{P}(j|S) \ge h,$$

and we write $PM_k^m(\Delta^h)$ for the set of all such $\mathbf{P}$. Moreover, $\mathbf{P}$ fulfills Δ^0 if

$$\forall S \in [m]_k \, \exists i \in S \, \forall j \in S \setminus \{i\} : \mathbf{P}(i|S) - \mathbf{P}(j|S) > 0$$

and we denote the set of all $\mathbf{P}$ fulfilling Δ^0 by $PM_k^m(\Delta^0)$. As for the case of reciprocal relations, the parameter h in this definition can be seen as a hardness parameter: If $\mathbf{P} \in PM_k^m(\Delta^h)$ for some large $h \in [0, 1]$, then, for any $S \in [m]_k$, the element $i_S = \mathrm{argmax}_{i \in S} \mathbf{P}(i|S)$ fulfills $\mathbf{P}(i_S|S) \ge \mathbf{P}(j|S) + h$ and is thus "easy" to identify from S based on independent and identically distributed (iid) samples with distribution $\mathrm{Cat}(\mathbf{P}(\cdot|S))$.

Parametric Assumptions Other prominent assumptions in the literature [Busa-Fekete et al., 2014a, Maystre and Grossglauser, 2017, Saha and Gopalan, 2019c] suppose that $\mathbf{P}$ is coherent with an underlying probability distribution $\mathbb{P}'$ on the set $\mathbb{S}_m$ of all rankings on $[m]$ in the sense that the distributions of the feedback $\mathbf{P}(i|S)$ are corresponding marginals of $\mathbb{P}'$. Formally, a ranking is a permutation $\sigma : [m] \to [m]$ and we say σ *ranks i better than j* if $\sigma(i) < \sigma(j)$. Given a probability measure $\mathbb{P}'$ on $\mathbb{S}_m$, the corresponding marginal (the probability that i is the winner of S) can be written as

$$\mathbf{P}(i|S) = \sum\nolimits_{\sigma \in \mathbb{S}_m \text{ with } \forall j \in S \setminus \{i\}: \sigma(i) < \sigma(j)} \mathbb{P}'(\sigma).$$

In case $\mathbb{P}'$ is instantiated with a *Plackett-Luce distribution* [Plackett, 1975, Luce, 1959] with unknown parameter $\boldsymbol{\theta} \in (0, \infty)^m$, we have

$$\mathbb{P}'(\sigma) = \prod\nolimits_{i=1}^{n} \frac{\theta_{\sigma(i)}}{\theta_{\sigma(i)} + \theta_{\sigma(i+1)} + \cdots + \theta_{\sigma(m)}}$$

and the corresponding marginals are (cf. e.g. [Cheng et al., 2010])

$$\mathbf{P}(i|S) = \frac{\theta_i}{\sum_{j \in S} \theta_j}. \tag{1.1}$$

Since $\theta_{i_1} \ge \cdots \ge \theta_{i_k}$ implies $\mathbf{P}(i_1|S) \ge \cdots \ge \mathbf{P}(i_k|S)$ for $S = \{i_1, \ldots, i_k\}$, θ_i can be seen as a skill parameter of arm i. We write $PM_k^m(\mathrm{PL})$ for the set of all $\mathbf{P} \in PM_k^m$, which

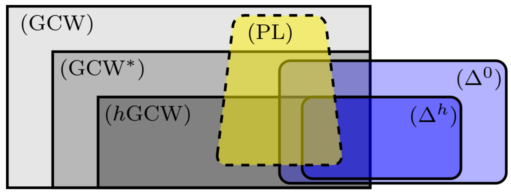


Figure 2.: Overview of coherences in MDB; here, we simply write (X) instead of PM_k^m(X) for X $\in \{\text{GCW}, \text{GCW}^*, h\text{GCW}, \Delta^0, \Delta^h, \text{PL}\}$.

fulfill (1.1) for some $\boldsymbol{\theta} \in (0,\infty)^m$. There are also other parametric distributions used as assumptions in this regard, and some of these will play a minor role in Sec. 2.5.1, but at this point we restrict ourselves to the Plackett-Luce distribution as it is the most relevant one for this thesis.

Fig. 2 illustrates the relationships between the different coherences in MDB in form of a Venn diagram, that will formally be justified in Lem. 6.1.

Transitivity Assumptions for Dueling Bandits Another commonly assumed coherence of the feedback mechanism in DB is some specific type of *transitivity*. An informal explanation of this term could be stated as "If i is better than j, and j is better than k, then i is also better than k". Such an assumption facilitates the identification of a "best arm", since observations on the duels $\{i,j\}$ and $\{j,k\}$ would also provide information on the winner of the duel $\{i,k\}$. For reciprocal (preference) relations, it is not uniquely determined what "i is better than j" means, since the preference of i over j is not deterministic but instead i is preferred over j with some probability $q_{i,j} \in [0,1]$. In dueling bandits, appropriate modified versions of transitivity, so-called *stochastic transitivities* [Fishburn, 1973, Haddenhorst et al., 2020], play an important role: First, they may assure that the learning task itself is actually well defined, for example that a naturally "best" arm actually exists. Second, they are on the basis of the design of efficient learning algorithms, which exploit generalized transitivity to reduce sample complexity [Yue and Joachims, 2011, Mohajer et al., 2017, Falahatgar et al., 2018]. This is comparable to how standard sorting algorithms avoid the comparison of all pairs of items and achieve an $\mathcal{O}(n \log n)$ (instead of an $\mathcal{O}(n^2)$) complexity.

Two prominent types of stochastic transitivity are weak and strong stochastic transitivity. Formally, $\mathbf{Q} = (q_{i,j})_{1 \le i,j \le m} \in \mathcal{Q}_m$ satisfies *weak stochastic transitivity* (WST) if

$$\forall \text{ distinct } i,j,k \in [m] : \big(\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge 1/2 \big) \ ,$$

and *strong stochastic transitivity* (SST) if

$$\forall \text{ distinct } i,j,k \in [m] : \big(\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge \max(q_{i,j}, q_{j,k}) \big) \, .$$

We denote the set consisting of all stochastically transitive reciprocal relations of a certain type XST as $\mathcal{Q}_m(\text{XST}) := \{\mathbf{Q} \in \mathcal{Q}_m : \mathbf{Q} \text{ is XST}\}$ and write $\mathcal{Q}_m(\neg\text{XST}) := \mathcal{Q}_m \setminus \mathcal{Q}_m(\text{XST})$. Then, $\mathcal{Q}_m(\text{SST}) \subsetneq \mathcal{Q}_m(\text{WST})$ holds [Haddenhorst et al., 2020]. Further types of stochastic transitivity will be introduced and discussed in Ch. 5.

Deterministic Winner Feedback in Dueling Bandits We will also take a look at deterministic analoga of several dueling bandits problems, i.e., those where the outcomes of a duel between two arms is supposed to be deterministic instead of random. Formally, this can be realized by assuming the reciprocal relation $\mathbf{Q}$ to be an element of

$$\mathcal{R}_m \coloneqq \{\mathbf{Q} = (q_{i,j})_{1 \leq i,j \leq m} \in \mathcal{Q}_m \,:\, q_{i,j} \in \{0,1\} \text{ for all distinct } i,j \in [m]\},$$

which can be identified with the set of tournaments on $[m]$. Graph-theoretical considerations will result in deterministic solutions to our problems of interest and also prove fruitful for their corresponding probabilistic counterparts.

Mixed Assumptions From a theoretical point of view, it is often interesting to combine certain assumptions, since this allows to analyze to what extent a learning problem under a particular assumption is influenced by the addition of another assumption. Therefore, for any assumptions X and Y from above such as GCW, Δ^h, PL for $k \geq 2$ and WST, CW, SST in case $k = 2$, we implicitly define $\neg$X, X $\wedge$ Y and X $\vee$ Y via $PM_k^m(\neg \mathrm{X}) = PM_k^m \backslash PM_k^m(\mathrm{X})$, $PM_k^m(\mathrm{X} \wedge \mathrm{Y}) = PM_k^m(\mathrm{X}) \cap PM_k^m(\mathrm{Y})$ and $PM_k^m(\mathrm{X} \vee \mathrm{Y}) = PM_k^m(\mathrm{X}) \cup PM_k^m(\mathrm{Y})$, respectively. Furthermore, we abbreviate $\mathcal{R}_m(\mathrm{X}) \coloneqq \mathcal{R}_m \cap \mathcal{Q}_m(\mathrm{X})$ and $\mathcal{Q}_m^h(\mathrm{X}) \coloneqq \mathcal{Q}_m^h \cap \mathcal{Q}_m(\mathrm{X})$.

1.3. Outline and Contribution of this Thesis

This thesis is split into three parts, for each of which we provide a brief overview below.

Part I

The first part serves as preparation for the analysis of the learning problems in (M)DB considered in Parts II and III. In it, we restate useful concentration inequalities and other helpful lemmata. We discuss the problem of identifying the mode of categorical random variable with a desired confidence and with as few samples as possible, starting with the special case of two categories and generalizing it afterwards to $k > 2$ categories. The obtained upper and lower bounds on the sample complexity of solutions to this problem seem to partly improve upon the state-of-the-art and lay a basis for the theoretical results obtained in Parts II and III. We conclude Ch. 2 with a change-of-measure inequality from the field of multi-armed bandits, which is a main ingredient for our lower bound results in Part III. We demonstrate its usefulness by inferring general impossibility results for testing for several coherences in (M)DB under the low-noise assumption, which imply, e.g. to some extent the impossibility of testing for the stochastic triangle inequality assumption in DB and for the Plackett-Luce assumption in MDB.

Ch. 3 further prepares Part II with some graph-theoretical observations. More precisely, we take a look at testing properties such as CW for deterministic reciprocal relations $\mathbf{R} \in \mathcal{R}_m$, which can be identified with tournaments on $[m]$. For this, we focus on *deterministic sequential testing algorithms (DSTAs)*, which may query at each time step exactly one of the edges of the tournament. We establish necessary and sufficient conditions for the termination time of solutions to such problems. Aside of preparations for Part II, we also address the problem of checking acyclicity of tournaments in a sequential manner. For the case where m is even, we provide a non-trivial lower bound for the worst-case query complexity necessary to solve this task as well as a non-trivial sequential procedure, which is able to test acyclicity without having to query all edges of the tournament.

Part II

The second part is dedicated to discuss the following statistical learning problems in the dueling bandits scenario.

CW Testification & Related Problems Ch. 4 analyzes problems related to the CW in dueling bandits. Here, our main focus is on combined test ing and ident ification of the Condorcet winner, which we have termed *CW testification*. More precisely, we say that a (possibly probabilistic sequential) algorithm $\mathcal{A}$ *solves CW testification on* $\mathcal{Q}_m^h$ *for* γ if both of the following are true:

- whenever $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$, $\mathcal{A}$ identifies the correct CW of $\mathbf{Q}$ with error probability at most γ,
- whenever $\mathbf{Q} \in \mathcal{Q}_m^h(\neg\mathrm{CW})$, $\mathcal{A}$ outputs with error probability at most γ that $\mathbf{Q}$ has no CW.

We provide a general framework called *Noisy Tournament Sampling* (NTS), which is is able to passively solve the CW testification problem under mild assumptions on π. Instantiated with an appropriate sampling strategy π, which mimics the behaviour of an appropriately chosen DSTA for CW testification of tournaments, NTS results in a solution to the problem, which has w.r.t. $\mathcal{Q}_m^h$ a worst-case expected sample complexity of order $\tilde{\mathcal{O}}(\frac{m}{h^2}\ln\frac{1}{\gamma})$; here and throughout, we write $\tilde{\mathcal{O}}$, $\widetilde{\Omega}$ and $\widetilde{\Theta}$ for those versions of the Bachmann-Landau notations, which hide logarithmic factors. Empirically, we show that this solution outperforms a naive two-stage procedure consisting of a state-of-the-art CW identification algorithm and a verification procedure. The lower bounds from Part I allow us to deduce an instance-wise lower bound for solutions to the problem, which is formulated in terms of the parameters m, γ as well as the *gaps* $\bar{q}_{i,j} := |1/2 - q_{i,j}|$ of the underlying instance $\mathbf{Q} = (q_{i,j})_{1 \le i,j \le m} \in \mathcal{Q}_m$.

In addition to CW testification, we also tackle the following learning problems:

- *CW checking on* $\mathcal{Q}_m^h$ *for* γ*:* Whenever $\mathbf{Q} \in \mathcal{Q}_m^h$, check with error probability $\le \gamma$ whether $\mathbf{Q}$ has a CW or not.
- *CW verification on* $\mathcal{Q}_m^h$ *for* γ*:* Whenever $\mathbf{Q} \in \mathcal{Q}_m^h$ and given some input $i \in [m]$, decide with error probability $\le \gamma$ whether i is the CW of $\mathbf{Q}$ or not.
- *CW identification on* $\mathcal{Q}_m^h(\mathrm{CW})$ *for* γ*:* Whenever $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$, identify the CW of $\mathbf{Q}$ with error probability $\le \gamma$.
- *CW verification on* $\mathcal{Q}_m^h(\mathrm{CW})$ *for* γ*:* Whenever $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$ and given some input $i \in [m]$, decide with error probability $\le \gamma$ whether i is the CW of $\mathbf{Q}$ or not.

We provide worst-case sample complexity bounds for these problems, that asymptotically match the corresponding lower bounds in the worst-case sense up to logarithmic factors. We show that any of these problems requires in the worst case $\tilde{\Theta}(\frac{m}{h^2}\ln\frac{1}{\gamma})$ samples to be solved, and provide more sophisticated instance-wise sample complexity upper and lower bounds for solutions to these, again in terms of the gaps $\bar{q}_{i,j}$. As far as we know, most of our upper and lower bounds were novel when published in [Haddenhorst et al., 2021a], the exception being those for CW identification on $\mathcal{Q}_m^h(\mathrm{CW})$ for γ. Via a reduction to *pure-exploration multi-armed bandits* (PE-MABs), results from Degenne and Koolen [2019] are applicable for any of these problems and yield sharp bounds on certain asymptotics of the expected sample complexity of solutions. These bounds are different, but consistent with the aforementioned.

Testing for Stochastic Transitivity In Ch. 5, we discuss the problem of testing stochastic transitivity in dueling bandits. Even though this problem is, due to frequently made stochastic transitivity assumptions, of importance for the DB literature, it has apparently not been tackled before [Haddenhorst et al., 2021b]. Formally, for a fixed transitivity type $\mathrm{XST} \in \{\mathrm{WST}, \mathrm{SST}, \dots\}$ a (possibly probabilisitic, sequential) algorithm $\mathcal{A}$ is said to solve XST *testing on* $\mathcal{Q}_m^h$ *for* γ if the following holds: Whenever $\mathbf{Q} \in \mathcal{Q}_m^h$, $\mathcal{A}$ correctly decides with error probability $\leq \gamma$ whether $\mathbf{Q}$ is XST or not.

So far, we have only introduced WST and SST, but we will consider in Ch. 5 several alternative types XST of stochastic transitivity, which are of interest for dueling bandits. With the help of a general impossibility result established in Part I, we are able to show that any $\mathrm{XST} \neq \mathrm{WST}$ cannot adequately be tested in the sense that the worst-case expected sample complexity of any solution to XST testing on $\mathcal{Q}_m^h$ for γ is infinite. Hence, we focus on WST in the further course of the work. For this type of transitivity, we provide similarly as for the CW-related problems above instance-wise lower and upper bounds on the sample complexity, which are in the worst-case asymptotically optimal up to logarithmic factors and of order $\widetilde{\Theta}(\frac{m^2}{h^2} \ln \frac{1}{\gamma})$. We see that the optimal sample complexity may already be achieved with a rather naive solution. However, incorporating graph-theoretical considerations from Ch. 3 results in an improved algorithm, which significantly outperforms the other one in our experiments. Both the naive and improved algorithm are supplemented with appropriate passive versions. The rough idea behind our improved algorithm is based on conducting multiple binomial tests (one for each query set), whilst terminating as soon as an estimated digraph is *acyclic in extension* (meaning that any of its supergraphs is acyclic) and restaining from querying those pairs that correspond to *negligible edges* (those, which cannot be contained in a cycle of a supergraph) of the digraph.

Additionally, we address the WST testing problem in a different way, namely based on the likelihood-rato test statistics λ_t resp. μ_t of

$$\begin{aligned} &\mathbf{H}_0 : \mathbf{Q} \in \mathcal{Q}_m(\mathrm{WST}) \quad \text{vs.} \quad \mathbf{H}_1 : \mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST}) \\ \text{resp.} \quad &\mathbf{H}_0' : \mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST}) \quad \text{vs.} \quad \mathbf{H}_1' : \mathbf{Q} \in \mathcal{Q}_m(\mathrm{WST}). \end{aligned}$$

With slightly modified versions $\tilde{\lambda}_t$ and $\tilde{\mu}_t$ of these, we provide another passive solution to WST testing on $\mathcal{Q}_m^h$ for γ, which requires mild assumptions on the sampling strategy. This solution is computationally more expensive and w.r.t. the expected sample complexity asymptotically worse than our other solution, but nevertheless, we state it for the sake of completeness in full detail.

We establish bounds on $\sup_{\mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST})} \limsup_{t\to\infty} \mathbb{P}_{\mathbf{Q}}(\mu_t > l)$ and the analogon for λ_t, which allow for asymptotic level-α tests for WST and $\neg$WST. These estimates extend upon results from Iverson and Falmagne [1985], who restricted themselves to an offline setting, where every pair is assumed to be queried exactly the same number of times. In comparison, our bounds are weaker but valid for the broader scenario of DB.

Similarly as for the CW-related problems, a reduction to PE-MABs allows to state further slightly different upper and lower bounds for WST testing, which appear overall consistent with the other results.

In fact, we also analyzed appropriate asymmetric versions of the above mentioned learning problems in DB with different type-I and type-II confidences $1-\alpha$ and $1-\beta$, but for the sake of simplicity we restricted ourselves to the symmetric case $\alpha = \beta =: \gamma$ in this overview.

Part III

In the last part, we switch from the dueling to the more general multi-dueling bandits setting and restrict ourselves to the active scenario. Ch. 6 discusses the problem to identify the GCW in multi-dueling bandits under different assumptions X, which are at least as strong as the existence of a GCW, i.e., $PM_k^m(\mathrm{X}) \subseteq PM_k^m(\mathrm{GCW})$. Formally, we say that a (possibly probabilistic, sequential) algorithm $\mathcal{A}$ solves GCW *identification on* $PM_k^m(\mathrm{X})$ *for* γ (short: $\mathcal{A}$ solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{X})$) if it identifies for any $\mathbf{P} \in PM_k^m(\mathrm{X})$ with error probability $\leq \gamma$ a GCW of $\mathbf{P}$.

We provide instance-wise sample complexity lower and upper bounds for different choices of X. Moreover, we prove all of the worst-case sample complexity bounds shown in Table 1.1, where the worst-case is meant w.r.t. instances in $PM_k^m(\mathrm{X} \wedge \mathrm{Y})$ for some further assumption Y. The bounds for those problems including the assumption PL are inferred from corresponding instance-wise versions, which have been proven by Saha and Gopalan [2020b]. The remaining lower bound is based on a measure-changing argument from Kaufmann et al. [2016], that we already prepare in Sec. 2.5. The upper bounds from Thm. 6.12 and Thm. 6.13 are achieved via different versions of our algorithm Dvoretzky-Kiefer-Wolfowitz Tournament (DKWT). This is basically a knockout-procedure and consists of different executions of an algorithm, which determines the mode of a categorical distribution based on the *Dvoretzky-Kiefer-Wolfowitz inequality* and is already presented in Sec. 2.3.

Table 1.1.: Sample complexity bounds of solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{X})$.

(X)	(Y)	Type	Asymptotic bounds	References
(PL)	(hGCW)	in exp.	$\Omega(\frac{m}{h^2k}(\frac{1}{k}+h)\ln\frac{1}{\gamma})$	Thm. 6.3
($\Delta^h \wedge$ GCW)	(Δ^h)	in exp.	$\Omega(\frac{m}{h^2k}\ln\frac{1}{\gamma})$	Thm. 6.4
(PL $\wedge$ GCW*)	(hGCW)	w.h.p.	$\mathcal{O}(\frac{m}{h^2k}(\frac{1}{k}+h)\ln(\frac{k}{\gamma}\ln\frac{1}{h}))$	Thm. 6.11
(GCW $\wedge \Delta^0$)	(Δ^h)	w.h.p.	$\mathcal{O}(\frac{m}{h^2k}\ln(\frac{m}{k})(\ln\ln\frac{1}{h}+\ln\frac{1}{\gamma}))$	Thm. 6.12
(hGCW $\wedge \Delta^0$)	(hGCW)	a.s.	$\mathcal{O}(\frac{m}{h^2k}\ln\frac{m}{k\gamma})$	Thm. 6.13

Thm. 6.4 implies that any solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW})$ has w.r.t. $PM_k^m(h\mathrm{GCW})$ a worst-case sample complexity of $\Omega(\frac{m}{kh^2}\ln\frac{1}{\gamma})$. As Thm. 6.11 and Thm. 6.12 indicate that the bounds in Thm. 6.3 and Thm. 6.4 are asymptotically sharp up to logarithmic factors, the GCW identification problem seems to be easier under the PL assumption by a factor $\frac{1}{k}+h$.

Finally, we conclude this thesis in Ch. 7 with a brief discussion on the limitations of this work, remaining open questions and potentially interesting further research directions.

The results presented in this thesis have mostly been published in [Haddenhorst et al., 2021a], [Haddenhorst et al., 2021b] and [Haddenhorst et al., 2021c], but there are some improvements (e.g., the sample complexity lower and upper bounds for solutions to $\mathcal{P}_{\mathrm{CWt}}^{m,0,\alpha,\beta}$ in Sec. 4) and even further results which have not been published such as the upper and lower bounds for acyclicity testing of tournaments, the impossibility results in Sec. 2.5.1 and all of Sec. 5.4. More details on such differences and unpublished results are briefly mentioned at the end of each corresponding chapter.

1.4. Related Work

In general, multi-armed bandits (MABs) describe a sequential decision making problem, in which a learner subsequently chooses at each time step a possible action and observes a feedback for its choice from an environment. Possible differences in the available actions of the learner (e.g. pull one or multiple "arms", cf. below), the type of feedback (e.g. a numerical or a binary value), assumptions on the environment (e.g. stochastic or non-stochastic) as well as distinct learning objectives (e.g. to identify the "best arm" or a "ranking over all arms") result in a variety of possible learning scenarios and thus lead to different variants of MABs.

In this thesis, we restrict ourselves to the particular variants of *dueling bandits* (DB) and *multi-dueling bandits* (MDB) as introduced above. To give an intuition how these fit into the bigger picture and literature on MABs, we give a small overview of the (standard) MAB scenario and its subfields DB and MDB in general. As the field of MABs is increasingly growing and already contains a large body of literature, giving a complete survey would be far out of scope of this section. Instead, we can only provide a limited overview of MAB and (M)DB at this point. A more extensive overview of and introduction to MABs in general can be found in [Bubeck and Cesa-Bianchi, 2012, Lattimore and Szepesvári, 2020, Slivkins, 2022]. For a survey on further real-world applications of MABs confer [Bouneffouf and Rish, 2019], and more information on the particular subfields of DB and MDB can be found in the surveys [Sui et al., 2018] and [Bengs et al., 2021].

(Standard) Multi-Armed Bandits The *multi-armed bandit* (MAB) scenario has been introduced by Thompson [1933] and describes a simple sequential decision making problem, in which a learner can choose at each time step from m many options, upon which it observes a corresponding feedback value. Here, the options may be regarded as different slot machines, and the choice of an option as a "pull" of the "arm" of the corresponding slot machine; this explains the notion "multi-armed bandit". The learner's behaviour which arm to choose at time t based on the *history* at time t (i.e., all previously pulled arms and observed feedback values until time t) is modelled by means of a *policy* [Lattimore and Szepesvári, 2020] or a *sampling strategy* [Kalyanakrishnan et al., 2012, Kaufmann and Kalyanakrishnan, 2013], and our definition of a sampling strategy is an appropriate modification for (M)DB of the latter term.

In (standard) MABs, the observed feedback upon pulling an arm is oftentimes a numerical value $X_{t,i(t)}$ and understood as *reward* of pulling that particular arm $i(t)$. A typical learning objective is to achieve a large cumulative reward $R_T = \sum_{t=1}^{T} X_{t,i(t)}$ until some predefined time $T \in \mathbb{N} \cup \{\infty\}$, or, alternatively, minimize the (cumulative) *regret*, which is the corresponding gap between the reward of the chosen arm and the largest mean of the reward distributions of all arms. Whilst solving this problem, a learner naturally faces the so-called *exploration-exploitation dilemma* [Lai and Robbins, 1985, Auer et al., 2002a, Cesa-Bianchi and Lugosi, 2006]: It wants to find a suitable tradeoff between gathering information about all arms to identify "good" ones (known as "exploration") and exploiting this knowledge by playing supposedly good arms. Theoretical questions in this scenario include, e.g. lower and upper bounds on the cumulative reward R_T, and prominent solutions for such kind of learning scenarios are, e.g. *upper-confidence bound* (UCB) algorithms [Auer et al., 2002b] and *Thompson sampling* [Thompson, 1933].

An alternative learning scenario is that of *pure-exploration multi-armed bandits* (PE-MABs), where the learner restricts itself on the identification of a good arm without further exploiting it. More precisely, a typical theoretical objective therein consists of identifying with confidence at least $1-\gamma$ an ε-approximately correct (i.e., best) arm, for an appropriate notion of "approximately", and this is known as the (ε, γ)-PAC (*probably approximately correct*) paradigm[1]. A typical research question in this regard is "How many arm pulls are necessary for finding with confidence at least $1-\gamma$ an ε-best arm?" [Even-Dar et al., 2002, Mannor and Tsitsiklis, 2004, Even-Dar et al., 2006]. In case $\varepsilon = 0$, where one is in fact interested in finding a best (not only approx. best) target, this scenario is also known as γ-PAC [Kaufmann et al., 2016].

Apart from the just mentioned, there are further modifications of MABs and its variants. For example, the works [Berry et al., 1997, Wang et al., 2009] and [Carpentier and Valko, 2015] consider MABs with infinitely many arms, whereas [Eick, 1988, Vernade et al., 2020] and [Gael et al., 2020] consider MABs with feedback, which is not immediately observable for the learner but instead comes with some delay. In the *contextualized bandits* scenario [Auer et al., 2002b, Tewari and Murphy, 2017] the learner observes prior to its action an additional side information, which is of relevance for observable feedbacks of the arms; in the field of recommandation systems, this could, e.g. be some knowledge about the customer who provides the next feedback. Recently, Badanidiyuru et al. [2013] introduced *bandits with knapsacks* (cf. e.g. [Slivkins, 2022, p.122]) as a constrained MAB version, where the learner is equipped with a certain amount of different resources, each arm pull results (in addition to a reward value) also in a consumption of these resources, and if one of the resources is exceeded, the learning process is terminated.

Dueling Bandits In many practically relevant applications, instead of a numerical feedback signal, only a preference over $k \geq 2$ alternatives may be observable. For example, this is the case when ranking XBox gamers according to duel outcomes [Guo et al., 2012] or when rating different objects according to pairwise user preferences, which can nowadays conveniently be collected via crowdsourcing services such as Amazon Mechanical Turk [Chen et al., 2013, Yan et al., 2022]. This led to an increased research in the field of preference learning [Fürnkranz and Hüllermeier, 2011] and on the multi-dueling variant of MAB. The particular case $k = 2$ is known as *dueling bandits* (DB).

Despite its obviously different type of feedback, the standard MAB setting is still of

[1]Usually, this setting is referred to (ε, δ)-PAC, but we use γ instead of δ here to be more consistent with the overall notation of this thesis.

interest for its dueling variant, because certain DB problems may be reduced to MAB problems [Owen, 1982, Ailon et al., 2014b] and in fact, for proving a lower bound result, a reduction to MABs will prove helpful for us in Sec. 4.1. Moreover, several algorithmic ideas may be transferred from MAB to DB for solving DB problems. Worth mentioning are, e.g. the two adaptations of Thompson sampling known as *Double Thompson Sampling* (DTS) [Wu and Liu, 2016] and MergeDTS [Li et al., 2020], and also several variants of UCB such as RUCB [Zoghi et al., 2014b], MergeRUCB [Zoghi et al., 2015b] and UCB-TS [Ramamohan et al., 2016].

To formulate the objective of regret minimization, the notion of regret from MAB has been adapted to DB in several ways: If i^* is the target arm (i.e., a notion of best arm such as the CW) and the learner chooses the duel $\{i(t), j(t)\}$ at time t, the suffered *strong, weak* and *average regret* are defined as $r_{t,\max} := \max\{q_{i^*,i(t)}, q_{i^*,j(t)}\} - 1/2$, $r_{t,\min} := \min\{q_{i^*,i(t)}, q_{i^*,j(t)}\} - 1/2$ and $r_{t,\mathrm{avg}} := \frac{r_{t,\max}+r_{t,\min}}{2}$, respectively. The learning objective consists of minimizing the corresponding cumulative regret and is investigated, e.g. by Yue and Joachims [2009], Urvoy et al. [2013], Zoghi et al. [2014a] for the average regret, by Peköz et al. [2020] for the weak regret and by Chen and Frazier [2017] for both strong and weak regret. Chen and Frazier [2016] consider as further variant a weak utility-based regret.

As for MABs, an alternative goal is pure exploration of a “best arm”, either in an (ε, γ)-PAC fashion [Falahatgar et al., 2018, Lin and Lu, 2018] or in a γ-PAC sense [Karnin, 2016, Mohajer et al., 2017, Ren et al., 2020]. In contrast to MABs, the notion of “best arm” is much less clear for DBs, and in addition to the CW introduced above there exist further prominent choices in the literature. Following Copeland’s method [Copeland, 1951], one may define for a reciprocal relation $\mathbf{Q} \in \mathcal{Q}_m$ and any $i \in [m]$ the number $c_i(\mathbf{Q})$ of arms j, for which i is likely to win in the sense that $q_{i,j} > 1/2$, and define the *Copeland winner* as that arm with largest Copeland score. A related but different notion is the *Borda winner* of $\mathbf{Q}$, which is that arm i with largest *Borda score* $b_i(\mathbf{Q}) = \frac{1}{m-1}\sum_{j\neq i} q_{i,j}$. Note that $b_i(\mathbf{Q})$ is the expected probability of i winning its duel in a randomly chosen query set. In contrast to the CW both a Copeland winner and a Borda winner always exist, but may be not unique. Further alternative best arm concepts of interest in the literature include the *von Neumann winner* [Balsubramani et al., 2016] and the *random walk winner*, see [Bengs et al., 2021] for an overview. Apart from that, further alternative goals in this regard are the identification [Braverman et al., 2016, Mohajer et al., 2017, Ren et al., 2020] or ranking [Mohajer et al., 2017] of the k best arms, and Ramamohan et al. [2016] consider as a learning objective the identification of the *Banks set*, *top cycle*, and *uncovered set*.

In the literature, the identification of the best arm w.r.t. some of the previously mentioned notions is oftentimes done whilst assuming $\mathbf{Q}$ to have a certain type of coherence. Some of these are formally required to assure that the learning problem is well-defined, e.g. [Falahatgar et al., 2018] assume the existence of an ε-best arm for the sake of identifying it, but some assumptions go beyond this. For example, making the low-noise assumption [Braverman and Mossel, 2008] simplifies the learning problem. Another assumption we have already sketched above states that $\mathbf{Q}$ is coherent with a probability model on rankings such as the Plackett-Luce (PL) model [Szörényi et al., 2015]. Apart from PL also the *Mallows model* [Busa-Fekete et al., 2014a] as well as the more general *random utility model* (RUM) [Saha and Gopalan, 2020a], which we briefly state in the paragraph on MDB below, are of interest in this regard. Moreover, also different types of *stochastic transitivity* are assumptions frequently made in the realm of dueling bandits. Falahatgar et al. [2017b,a, 2018] analyze the best-arm identification problem in an (ε, γ)-PAC manner under weak, moderate and strong stochastic transitivity (and partly further assumptions) and show that the stronger the assumption, the lower the required sample complexity, whereas strong resp. relaxed stochastic transitivity is considered by Yue et al. [2012] resp. Yue and Joachims [2011] for the sake of regret minimization. A more thorough discussion of the CW identification problem is provided in Ch. 4, and Ch. 5 contains more information on the different types of stochastic transitivity. For more details on further theoretical coherences like the *stochastic triangle inequality* we refer to Sec. 2.5.1, where we present negative results for statistically testing for such coherences under the low-noise assumption.

According to our definition of the DB setting, ties are not allowed as outcomes for the duels, i.e., each duel has its winner. When allowing ties, there are at least two ways how to deal with them in the DB literature: One possibility is *tie breaking*, i.e., declare one of the involved arms as the winner, and this may either be done randomly or in a favourable way, cf. the differences between DTS and DTS^+ in [Wu and Liu, 2016]. Another possibility is to count a tie between i and j as half a win for both i and j [Busa-Fekete et al., 2013, 2014b].

There exist several further variants of DB, of which we only mention a small portion at this point. Similar as for MAB, the case of infinitely many arms has been dealt with for DB in [Yue and Joachims, 2009, Ailon et al., 2014b], and Dudík et al. [2015] introduced a contextualized variant of DB that has recently been considered, e.g. in [Bengs et al., 2022, Saha, 2021]. Gupta and Saha [2021] and Kolpaczki et al. [2022] consider regret minimization in the case of non-stationarity, i.e., when $\mathbf{Q}$ may vary over time, and the works [Ailon et al., 2014a, Dudík et al., 2015, Gajane and Urvoy, 2015, Zimmert and Seldin, 2019] and [Ailon et al., 2014b] tackle the case where feedback is non-stochastic or even adversarially generated.

Multi-Dueling Bandits Another modification of MAB and generalization of DB is the *multi-dueling bandits* (MDB) scenario [Brost et al., 2016, Sui et al., 2017, Saha and Gopalan, 2018], in which a learner chooses $k \geq 2$ arms at once at each time step and then observes one of them as a winner. MDB is also known as *battling* [Saha and Gopalan, 2018], *preselection* [Bengs and Hüllermeier, 2020] and *choice bandits* [Agarwal et al., 2020] in the literature, and real-world applications include, e.g. algorithm configuration [El Mesaoudi-Paul et al., 2020] and online retrieval evaluation [Schuth et al., 2016].

As for MAB and DB, both regret minimization [Saha and Gopalan, 2019b, Bengs and Hüllermeier, 2020, Agarwal et al., 2020] and best-arm identification in an (ε, γ)-PAC setting [Saha and Gopalan, 2019c, 2020a,b] have been considered as learning objectives in the literature, where the “best arm” can again be defined in multiple ways: Similarly as we generalized the notion of the CW to the GCW above, one can also generalize the Copeland and Borda winner [Bengs et al., 2021]. But also alternative learning objectives such as identification of the top-k arms [Chen et al., 2018] or an underlying ranking (assuming its existence) in an (ε, γ)-PAC manner [Saha and Gopalan, 2019a,c, 2020a] have been considered so far.

Similarly as in DB, many works make assumptions on the underlying parameter $\mathbf{P}$

to assure, e.g. the well-definedness of the learning task itself or to simplify it. For example, Bengs and Hüllermeier [2020] assume $\mathbf{P}$ to fulfill the Plackett-Luce assumption whereas [Saha and Gopalan, 2020a] suppose $\mathbf{P}$ to be coherent with a general *random utility model* (RUM). This latter assumption means that there exists a parameter vector $\mathbf{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m$, giving each arm $i \in [m]$ a latent score v_i, as well as a probability distribution function (pdf) f on $\mathbb{R}$ such that, for any $i \in [m]$ and $S \subseteq [m]$ with $i \in S$, $\mathbf{P}(i|S) = \mathbb{P}(v_i + \xi_i = \max_{j \in S} v_j + \xi_j)$ for iid samples $\xi_j \sim f$, and in the special case where the ξ_j's are standard Gumbel distributed, it is equivalent to the PL assumption. Of a different type is the assumption made by Ren et al. [2021] for the identification of the top-k arms and the ranking of all arms. They assume coherence of $\mathbf{P}$ with an underlying ranking over all arms and also that for some unknown parameter $\vartheta \in (1/2, 1)$ any multi-duel S_t reveals with probability ϑ the "correct winner" i (the mode of $\mathbf{P}(\cdot|S_t)$) whereas the corresponding probabilities for observing any other arm of S_t can be arbitrary.

A scenario related to MDB is that of *combinatorial bandits* [Cesa-Bianchi and Lugosi, 2012, Kveton et al., 2015]. It resembles MDB in the sense that the learner chooses a set of arms at each time, but differs in the sense that the learner obtains as feedback not the mere winner information but instead rather quantitative feedback such as corresponding reward values for each involved arm or the total sum over such rewards. Apart from that, there are further variations of this setting, which are of interest in the literature: Saha and Gopalan [2018] discuss regret minimization in an MDB scenario with infinitely many arms, Brandt et al. [2022] investigate GCW identification in a non-stochastic setting and El Mesaoudi-Paul et al. [2020] apply a contextualized variant of MDB to the problem of algorithm configuration.

The Learning Scenario in This Thesis To put it in the broader context of MAB and its variants, one could describe the particular setting of (M)DB, which we focus on in this thesis, as *(multi-)dueling bandits with winner feedback in the stochastic setting*: A learner can pull at each time step $k \geq 2$ many arms $i_1, \ldots, i_k$, which compete against each other in a multi-duel, and exactly one of these is observed as a winner, namely i_l is supposed to win with some (unknown, underlying) probability $\mathbf{P}(i_l \,|\, \{i_1, \ldots, i_k\})$. Moreover, we suppose that this latter probability does not depend on the time the (multi-)duel is conducted and that the overall number of arms is finite. Throughout, our learning objective is to solve with a predefined confidence $1 - \gamma$ a decision problem for the underlying parameter $\mathbf{Q}$ in DB resp. $\mathbf{P}$ in MDB such as "Is $\mathbf{Q}$ weakly stochastic transitive?" or "What is the GCW of $\mathbf{P}$?". Thus, we are interested in *pure exploration* learning tasks in a γ-PAC manner.

At this point, we end our literature overview for the moment. Any of the subsequent chapters contains in its last section more thorough remarks on the corresponding related literature as well as a discussion on the obtained results and points out possible directions for future research.

1.5. Notation and Conventions

In this thesis, we discuss a variety of learning problems. In an attempt to standardize the notation, we write $\mathcal{P}_{\mathrm{T}}^{\mathrm{P}}(\mathrm{A})$ for a learning problem with task T, that has parameters P and shall be solved under assumptions A; e.g., $\mathcal{P}_{\mathrm{CWi}}^{m,h,\gamma}(\mathrm{CW})$ is the problem to identify the CW of any $\mathbf{Q} \in \mathcal{Q}_m^h$ with confidence $1 - \gamma$ under the assumption that it exists (i.e., that $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$). If no ssumptions are made, we simply write $\mathcal{P}_{\mathrm{T}}^{\mathrm{P}}$, and when talking about the problem in general, we will leave out parts (or all) of P. Formally, any algorithm which tackles $\mathcal{P}_{\mathrm{T}}^{\mathrm{P}}(\mathrm{A})$ is supposed to have knowledge of all of T, P and A; e.g., if $\mathcal{A}$ tackles $\mathcal{P}_{\mathrm{GCWi}}^{m,k,\gamma}(\mathrm{GCW} \wedge \Delta^h)$, it knows that it should identify the GCW of some $\mathbf{P} \in PM_k^m(\mathrm{GCW} \wedge \Delta^h)$ with error probability at most γ and it a priori knows the values of m, k, γ and h. We distinguish the deterministic sequential learning problems from Ch. 3 from the probabilistic ones by writing $\mathcal{D}_{\mathrm{T}}^{\mathrm{P}}(\mathrm{A})$ instead of $\mathcal{P}_{\mathrm{T}}^{\mathrm{P}}(\mathrm{A})$ for these, e.g., $\mathcal{D}_{\mathrm{acyclic}}^m$ denotes the problem to deterministically test whether a tournament on $[m]$ is acyclic or not.

To formally address such problems, we oftentimes have to restrict ourselves to those instances with certain properties, e.g., the property to fulfill a particular assumption. Thus, let us write, for any set A and property X, $A(\mathrm{X})$ for the set of all elements in A, which fulfill X. If $\mathrm{X}, \mathrm{X}_1, \dots, \mathrm{X}_n$ are properties, we write as usual $\neg\mathrm{X}$ for "not X", $\mathrm{X}_1 \wedge \dots \wedge \mathrm{X}_n$ for "$\forall i \in [n] : \mathrm{X}_i$" and $\mathrm{X}_1 \vee \dots \vee \mathrm{X}_n$ for "$\exists i \in [n] : \mathrm{X}_i$". Hence, we have $A(\neg\mathrm{X}) = A \setminus A(\mathrm{X})$, $A(\mathrm{X}_1 \wedge \dots \wedge \mathrm{X}_n) = \bigcap_{i\in[n]} A(\mathrm{X}_i)$ and $A(\mathrm{X}_1 \vee \dots \vee \mathrm{X}_n) = \bigcup_{i\in[n]} A(\mathrm{X}_i)$. Note that this notation is consistent with the definitions of $PM_k^m(\mathrm{X}_1 \wedge \mathrm{X}_2)$ etc. from above.

For a convenient statement of the asymptotic behaviour of proven sample complexity upper and lower bounds, we use the standard Bachmann-Landau notations $\mathcal{O}(\cdot)$, $\Omega(\cdot)$, $\Theta(\cdot)$ and $o(\cdot)$ and, if not explicitly stated otherwise, they are to be understood as any of the corresponding involved parameters m, k, $\frac{1}{h}$, $\frac{1}{\alpha}$, $\frac{1}{\beta}$ and $\frac{1}{\gamma}$ tend to infinity. We write $\widetilde{\mathcal{O}}(\cdot)$ and $\widetilde{\Omega}$ for those modifications of $\mathcal{O}$ and Ω, which hide logarithmic factors, and $f(x) \in \Omega_{\sup}(g(x))$ as $x \to x_0$ for $\limsup_{x\to x_0} \frac{|f(x)|}{g(x)} > 0$. Moreover, we may abbreviate $f(k,h) \in \Theta(\ln k) \cap \Theta(\frac{\ln h}{\ln\ln h})$ for $f(k,h) \in \Theta(\ln k)$ and $f(k,h) \in \Theta(\frac{\ln h}{\ln\ln h})$.

Apart from that, in order to provide detailed versions of all proven results, we introduce lots of further convenient notation throughout this thesis. Every such notation is formally introduced along the lines where it first comes up. As additional guidance for the reader, we collected the most frequently used notations in a list of symbols at the beginning of this document, and we also added a list of algorithms and a list of abbreviations there.

Since the main focus of this thesis is on the theoretical guarantees, we present most of the proofs in the main text. For the sake of readability, some technical results and proofs are not instantly presented but outsourced to corresponding individual sections or subsections, and some proofs of the fundamental Chapters 2 and 3 are deferred to the appendix.

Before starting with Ch. 2, let us briefly mention two minor inconsistencies in this thesis. Firstly, the low-noise assumptions for $k = 2$ and $k \geq 3$ are not fully consistent, namely formally $PM_2^m(\Delta^h) = \{\mathbf{Q} \in \mathcal{Q}_m \,|\, \forall (i,j) \in (m)_2 : |q_{i,j} - 1/2| \geq h/2\} \neq \mathcal{Q}_m^h$ holds. For this reason, one may say that $\mathcal{P}_{\mathrm{CWi}}^{m,h/2,\gamma}(\mathrm{CW}) \approx \mathcal{P}_{\mathrm{GCWi}}^{m,2,\gamma}(h\mathrm{GCW}) = \mathcal{P}_{\mathrm{GCWi}}^{m,2,\gamma}(\mathrm{GCW} \wedge \Delta^h)$. Despite this slight inconsistency, we follow this notation throughout this thesis for the sake of convenience due to technical reasons, which we do not want to specify at this point. As we are mainly interested in the asymptotic behavior of sample complexity bounds in terms of h (and further parameters) for solving problems under the low-noise assumption in (M)DB, $\mathcal{Q}_m^h = \bigcup_{h'>h/2} PM_k^m(\Delta^{h'})$ and $PM_k^m(\Delta^h) = \bigcap_{h'<2h} \mathcal{Q}_m^{h'}$ assure that this small difference in the definitions is negligible for such asymptotic results.

Secondly, as already seen in Table 1.1, the sample complexity upper and lower bounds for discussed learning problems may differ in their corresponding type of guarantee. Some of them are to be understood with respect to the expected sample complexity, whereas others bound this term with high probability or even almost surely. Even though formally and theoretically incorrect, it is common practice in the (M)DB literature [Bengs et al., 2021] to compare such results of different type with each other, and throughout we proceed in a similar manner. For example, Thm. 6.12 merely shows that the sample complexity of an appropriate solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$ is w.r.t. the worst-case on $PM^m_k(\mathrm{GCW} \wedge \Delta^h)$-instances w.h.p. at most $\mathcal{O}(\frac{m}{h^2 k}(\ln \frac{m}{k})(\ln \ln \frac{1}{h} + \ln \frac{1}{\gamma}))$, and this does not imply a corresponding upper bound for the expected termination time of that solution[2] but we compare it anyway with the lower bound from Thm. 6.4 on the expected termination time of solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$ and call it then asymptotically optimal up to logarithmic factors.

[2] Note that, for any fixed $\gamma \in (0,1)$, a bound of the form $\mathbb{P}_{\mathbf{Q}}(T^{\mathcal{A}} \leq C) \geq 1-\gamma$ on the sample complexity $T^{\mathcal{A}}$ of a learner $\mathcal{A}$ does not even imply $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] < \infty$.

Part I.

Fundamentals

2. Probabilistic Prerequisites

This chapter prepares our analysis of the (multi-)dueling bandit problems in the Parts II and III of this thesis with some statistical and probabilistic discussions. It is split into six sections. In the first of these, we provide a collection of rather basic results and concentration inequalities, which will be of aid later on. In the second section, we tackle the problem of identifying the mode of a biased coin and provide sample complexity upper and lower bounds for solutions to this problem. Sec. 2.3 treats in a similar fashion the more general problem of identifying the mode of a generalized die, i.e., an arbitrary categorical random variable. After a brief empirical evaluation of the stated solutions in Sec. 2.4, we focus in Sec. 2.5 on a change-of-measure argument, which will be a key ingredient in some sample complexity lower bounds in the further course of this thesis. As an application, we illustrate how this argument implies negative results on the testibility of several statistical assumptions in (multi-)dueling bandits. We conclude this chapter in Sec. 2.6 with further remarks on our obtained results and related literature.

For the sake of readability, the proofs of some of the results in this chapter have been moved to the appendix.

2.1. Concentration Inequalities and Convergence Results

In this section, we collect some concentration inequalities, which will be of interest later on. We start with two famous inequalities named after Chernoff [1952] and Hoeffding [1963]. For convenience, we will simply refer to these as *Chernoff bound* resp. *Hoeffding's inequality* instead of Lem. 2.1 resp. Lem. 2.2.

Lemma 2.1 (Chernoff bound)**.** *If $X_1, \ldots, X_n$ are independent random variables with values in $\{0,1\}$ and $\gamma \in (0,1)$, then $S_n := \sum_{i=1}^n X_i$ fulfills*

$$\mathbb{P}\left(S_n \leq (1-\gamma)\mathbb{E}[S_n]\right) \leq e^{-\frac{\mathbb{E}[S_n]\gamma^2}{2}}.$$

Proof. This is (4.5) in [Mitzenmacher and Upfal, 2017]. □

Lemma 2.2 (Hoeffding's inequality)**.** *If $X_1, \ldots, X_n$ are independent random variables with $\mathbb{P}(\forall i \in [n] : X_i \in [a_i, b_i]) = 1$ for $\{a_i\}_{i\in[n]}, \{b_i\}_{i\in[n]} \subseteq \mathbb{R}$ with $a_i < b_i$ for all $i \in [n]$, then*

$$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

holds for any $\varepsilon > 0$

Proof. This is Prop. 2.7 in [Massart, 2007]. □

Throughout the rest of Sec. 2.1, we suppose $\gamma \in (0, 1/2)$ to be arbitrary but fixed and write, for any $p \in [0,1]$, $\{X_k^{(p)}\}_{k\in\mathbb{N}}$ for a sequence of random variables, which are independent and identically distributed (iid) as $X_k^{(p)} \sim \mathrm{Ber}(p)$, a Bernoulli distribution with parameter p. Under such assumptions, both previous lemmata are applicable. As immediate consequences of Hoeffding's inequality, we obtain the following two results. They will be of help for the proofs of Thm. 5.15 and Thm. 5.13.

Lemma 2.3. *For every $h \in (0, \frac{1}{2})$ we have*

$$\sup_{p:|p-1/2|>h} \sum_{n\in\mathbb{N}} \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} = \frac{1}{2}\right) \leq \frac{1}{h^2}. \tag{2.1}$$

Lemma 2.4. *For $\kappa > 1$ and $U_\kappa(n,\gamma) \coloneqq \sqrt{\frac{\ln\left(n^\kappa/\gamma\right)}{2n}}$ we have*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \geq U_\kappa(n,\gamma)\right) \leq \frac{\gamma}{n^\kappa},$$
$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \leq -U_\kappa(n,\gamma)\right) \leq \frac{\gamma}{n^\kappa}.$$

In particular, if $n' \in \mathbb{N}$ is such that $\sum_{n\geq n'} \frac{1}{n^\kappa} \leq 1$, then

$$\mathbb{P}\left(\exists n \geq n' : \frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \geq U_\kappa(n,\gamma)\right) \leq \gamma,$$
$$\mathbb{P}\left(\exists n \geq n' : \frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \leq -U_\kappa(n,\gamma)\right) \leq \gamma.$$

The following lemma is a rather direct consequence of the Chernoff bound and an auxiliary result that will be used in Sec. 5.4.

Lemma 2.5. *Let $\phi : [0,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ be given as $\phi(x) \coloneqq 2\arcsin(\sqrt{x}) - \pi/2$ and $Z_n^{(p)} \coloneqq \phi(\frac{1}{n}\sum_{i=1}^{n} X_i^{(p)})$ for any $n \in \mathbb{N}, p \in [0,1]$. Then, for fixed $c \in (0, \frac{1}{2})$ and every $p \in [0,c) \cup (1-c, 1]$ we obtain*

$$\mathbb{P}\left(Z_n^{(p)}\phi(p) < 0\right) \leq \exp\left(-\frac{(1-2c)^2 n}{4(2-2c)^2}\right). \tag{2.2}$$

For $\gamma \in (0,1)$, $q \coloneqq \exp\left(-\frac{(1-2c)^2}{4(2-2c)^2}\right)$ and $\tilde{n} \coloneqq \lceil \log_q\left((1-q)\gamma\right)\rceil$ we thus have

$$\sup_{p\in[0,c)\cup(1-c,1]} \mathbb{P}\left(\exists n \geq \tilde{n} \,:\, Z_n^{(p)}\phi(p) < 0\right) \leq \gamma. \tag{2.3}$$

Next, we provide an anytime confidence bound, which is due to Jamieson et al. [2013] and e.g. of interest for Sec. 2.2.

Lemma 2.6. *Let $\varepsilon = \varepsilon(\gamma) \in (0,1)$ and $\delta = \delta(\gamma) \in \left(0, \frac{\ln(1+\varepsilon)}{e}\right)$ be such that $\gamma = \frac{2+\varepsilon}{\varepsilon}\left(\frac{\delta}{\ln(1+\varepsilon)}\right)^{1+\varepsilon}$ holds. Define $U_\gamma(n) \coloneqq U_{\varepsilon(\gamma),\delta(\gamma)}$ as*

$$U_\gamma(n) \coloneqq \left(1+\sqrt{\varepsilon}\right)\sqrt{\frac{1}{2}(1+\varepsilon)n\ln\left(\frac{\ln((1+\varepsilon)n)}{\delta}\right)}.$$

Then, the centered sequence $Y_k^{(p)} := X_k^{(p)} - p$, $k \in \mathbb{N}$, *fulfills*

$$\mathbb{P}\left(\forall n \in \mathbb{N} : \sum\nolimits_{k=1}^{n} Y_k^{(p)} \leq U_\gamma(n)\right) \geq 1 - \gamma \tag{2.4}$$

as well as

$$\mathbb{P}\left(\forall n \in \mathbb{N} : \left|\sum\nolimits_{k=1}^{n} Y_k^{(p)}\right| \leq U_\gamma(n)\right) \geq 1 - 2\gamma. \tag{2.5}$$

The next lemma is a preparation for a result in Sec. 5.4.

Lemma 2.7. *Let* $c \in (0, 1/2)$, $\gamma \in (0,1)$, $\gamma' := \gamma/4$ *and* $\varepsilon' \in (0,1)$, $\delta' \in \left(0, \frac{\ln(1+\varepsilon')}{e}\right)$ *be such that* $\gamma' = \frac{2+\varepsilon'}{\varepsilon'}\left(\frac{\delta'}{\ln(1+\varepsilon')}\right)^{1+\varepsilon'}$. *As in Lem. 2.5 let* $\phi : [0,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\phi(x) := 2\arcsin(\sqrt{x}) - \pi/2$ *and* $Z_n^{(p)} := \phi\left(\frac{1}{n}\sum_{i=1}^{n} X_i^{(p)}\right)$ *for all* $n \in \mathbb{N}$ *and* $p \in [0,1]$ *and let*

$$L := L(c) := \sup_{x\in[c/2,1-c/2]} |\phi'(x)|.$$

Define for $n \in \mathbb{N}$

$$l(n) := \frac{1}{2}L^2\left(1 + \sqrt{\varepsilon'}\right)^2 (1+\varepsilon')\ln\left(\frac{\ln((1+\varepsilon')n)}{\delta'}\right)$$

and further

$$\tilde{n} := \frac{d'}{c^2}\ln\left(\frac{2}{\delta'}\ln\frac{(1+\varepsilon')d'}{c^2\delta'}\right) + 1$$

with $d' := 2(1+\sqrt{\varepsilon'})^2(1+\varepsilon')$. *Then, we have*

$$\sup_{p\in[c,1-c]} \mathbb{P}\left(\exists\, n \geq \tilde{n} \,:\, n(Z_n^{(p)} - \phi(p))^2 > l(n)\right) \leq \gamma.$$

The next lemma will show us that an assumption, which we make on the sample strategy for a theoretical result below, is rather a mild one.

Lemma 2.8. *Let* $a \geq 1$, $m \in \mathbb{N}$ *and suppose* $f : \mathbb{N} \to [1, \infty)$ *is monotonically increasing with* $f(t) \in o(\frac{t}{\ln^a(t)})$ *as* $t \to \infty$. *Let* $\{Z_t\}_{t\in\mathbb{N}}$ *be a family of independent random variables with* $Z_t \sim \mathrm{Ber}(\frac{1}{f(t)m})$, $t \in \mathbb{N}$. *Then,*

$$\frac{1}{\ln^a(t)}\sum\nolimits_{t'=1}^{t} Z_{t'} \to \infty \quad a.s. \text{ as } t \to \infty. \tag{2.6}$$

For the proof of Thm. 5.17 and Cor. 5.19, the following lemma will be helpful. It is a generalization of Lem. 3 in [Iverson and Falmagne, 1985]. Here and throughout, $\chi^2_{(k)}$ denotes the χ^2-distribution with k degrees of freedom, and for $a, b \in \mathbb{R}$ we also abbreviate $a \wedge b := \min\{a,b\}$ as well as $a \vee b := \max\{a,b\}$.

Lemma 2.9. *For fixed* $c, l > 0$ *the sequence* $\{a_t\}_{t\in\mathbb{N}}$ *given as*

$$a_t = \frac{1}{2^t}\sum\nolimits_{r=0}^{t}\binom{t}{r}\mathbb{P}\left(\chi^2_{(r\wedge c)} > l\right)$$

is monotonically increasing.

2.2. Sequentially Testing for the Mode of a Biased Coin

In the dueling bandits scenario, when $\{i, j\}$ is pulled at time step t, the observed feedback is supposed to be a sample $\sim \mathrm{Ber}(q_{i,j})$. Most of the properties of the underlying parameter $\mathbf{Q} = (q_{i,j})_{1 \leq i,j \leq m}$ introduced in Sec. 1.2 such as CW, WST and SST highly depend on whether $q_{i,j}$ is larger or lower than some particular threshold (that may depend on the other entries of $\mathbf{Q}$). Consequently, for being able to test for those properties, it is of benefit to be able to test whether the bias p of a coin is larger than some threshold p_0 or not. This section is dedicated to discuss the sample complexities necessary and sufficient for solving this problem and also to provide appropriate solutions. We start with a brief formalization of the problem.

Problem Description Suppose $p \in [0, 1]$ to be fixed but unknown to us and $\Theta_0, \Theta_1 \subseteq [0, 1]$ be two non-empty and disjoint parameter spaces for p that we know a priori. In fact, we will restrict ourselves to some specific choices of Θ_0 and Θ_1 below. Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of iid samples $X_n \sim \mathrm{Ber}(p)$, $n \in \mathbb{N}$, which are w.l.o.g. defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here, we write $\mathbb{P}_p(A)$ for the probability of an event $A \in \sigma(X_n, n \in \mathbb{N}) \subseteq \mathcal{F}$ when $X_n \sim \mathrm{Ber}(p)$, and $\mathbb{E}_p$ for the expectation w.r.t. $\mathbb{P}_p$. We are interested in deciding

$$\mathbf{H}_0 : p \in \Theta_0 \quad \text{versus} \quad \mathbf{H}_1 : p \in \Theta_1 \tag{2.7}$$

with predefined confidence $1 - \gamma \in (0, 1)$, based on as few of the samples X_n as possible, which are assumed to arrive in a sequential manner in the order $X_1, X_2, X_3, \ldots$. If a testing algorithm $\mathcal{A}$ is given, it may decide at any time to stop the learning process (i.e., the process of observing X_i's) and terminate. We write

$$T^{\mathcal{A}} := \sup\{n \in \mathbb{N} \,|\, \mathcal{A} \text{ observes } X_n\}$$

for the *termination time/sample complexity* of $\mathcal{A}$. Here, $T^{\mathcal{A}}$ is a random variable, which depends on the randomness in the observed samples $X_1, X_2, \ldots$ as well as the innate randomness of $\mathcal{A}$ itself, and takes values in $[0, \infty]$. We write $\mathbf{D}(\mathcal{A}) = 0$ or $\mathbf{D}(\mathcal{A}) = 1$ if $\mathcal{A}$ decides for $\mathbf{H}_0$ or $\mathbf{H}_1$, respectively. Moreover, we denote by $\mathfrak{A}_{\mathrm{Coin}}$ the set of all testing algorithms for (2.7).

We say that an algorithm $\mathcal{A}$ *solves* $\mathcal{P}^{\Theta_0,\Theta_1;\gamma}_{\mathrm{Coin}}$ if it terminates a.s. for any $p \in \Theta_0 \cup \Theta_1$ and tests (2.7) with type I and II errors at most γ, respectively, i.e., if[1]

$$\begin{aligned}
\forall p \in \Theta_0 \cup \Theta_1 &: \mathbb{P}_p\left(T^{\mathcal{A}} < \infty\right) = 1, \\
\forall p \in \Theta_0 &: \mathbb{P}_p\left(\mathbf{D}(\mathcal{A}) = 0\right) \geq 1 - \gamma, \\
\forall p \in \Theta_1 &: \mathbb{P}_p\left(\mathbf{D}(\mathcal{A}) = 1\right) \geq 1 - \gamma.
\end{aligned}$$

In the further course of this text, we restrict ourselves to the following choices of Θ_0 and Θ_1, and define the related problems $\mathcal{P}^{p_0,p_1;\gamma}_{\mathrm{Coin}}$, $\mathcal{P}^{p_0;\gamma}_{\mathrm{Coin}}$, $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ and $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ as follows:

- If $\Theta_0 = \{p_0\}$ and $\Theta_1 = \{p_1\}$ for some distinct $p_0, p_1 \in [0, 1]$, we write $\mathcal{P}^{p_0,p_1;\gamma}_{\mathrm{Coin}}$ for $\mathcal{P}^{\Theta_0,\Theta_1;\gamma}_{\mathrm{Coin}}$. For any $p_0 \neq 1/2$, we further abbreviate $\mathcal{P}^{p_0,1-p_0;\gamma}_{\mathrm{Coin}}$ as $\mathcal{P}^{p_0;\gamma}_{\mathrm{Coin}}$.

[1] The probability $\mathbb{P}_p(T^{\mathcal{A}} < \infty$ and $\mathbf{D}(\mathcal{A}) = b)$ is taken w.r.t. both the randomness of the $\{X_n\}_{n \in \mathbb{N}}$ (with $X_n \sim \mathrm{Ber}(p)$) as well as the randomness involved in the (possibly) probabilistic behaviour of $\mathcal{A}$.

- If $\Theta_0 = [0, 1/2 - h)$ and $\Theta_1 = (1/2 + h, 1]$ for some known $h \in [0, 1/2)$, we also write $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ instead of $\mathcal{P}^{\Theta_0,\Theta_1;\gamma}_{\text{Coin}}$. For convenience, we abbreviate $\mathcal{P}^{\gamma}_{\text{Coin}} := \mathcal{P}^{0,\gamma}_{\text{Coin}}$.

If $h < h'$, any solution to $\mathcal{P}^{h',\gamma}_{\text{Coin}}$ also solves $\mathcal{P}^{h,\gamma}_{\text{Coin}}$. In this regard, h can be seen as a hardness parameter, which allows us to model the low-noise assumption in dueling bandits, cf. Sec. 1.1. Moreover, we obtain as a direct consequence of the definitions the equivalences

$$\mathcal{A} \text{ solves } \mathcal{P}^{\gamma}_{\text{Coin}} \quad \Leftrightarrow \quad \forall h \in (0, 1/2) : \mathcal{A} \text{ solves } \mathcal{P}^{h,\gamma}_{\text{Coin}} \quad \Leftrightarrow \quad \forall p_0 > 1/2 : \mathcal{A} \text{ solves } \mathcal{P}^{p_0;\gamma}_{\text{Coin}},$$

as well as, for any $h \in (0, 1/2)$,

$$\mathcal{A} \text{ solves } \mathcal{P}^{h,\gamma}_{\text{Coin}} \quad \Leftrightarrow \quad \forall p_0 \in (1/2 + h, 1] : \mathcal{A} \text{ solves } \mathcal{P}^{p_0;\gamma}_{\text{Coin}}.$$

which are valid for any testing algorithm $\mathcal{A}$ and any fixed γ.

Before treating sequential testing algorithms, let us discuss the limitations of non-sequential testing alorithms to the problems at hand using the examples of $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ and $\mathcal{P}^{\gamma}_{\text{Coin}}$. In Sec. 2.2.1 we present sample complexity lower bounds of solutions to $\mathcal{P}^{h,\gamma}_{\text{Coin}}$, $\mathcal{P}^{\gamma}_{\text{Coin}}$ and also $\mathcal{P}^{p_0,p_1;\gamma}_{\text{Coin}}$, which will be of use for proving sample complexity lower bounds in the dueling bandits scenario in Sec. 4 and 5. Then, we continue in Sec. 2.2.2 with the discussion of solutions to $\mathcal{P}^{p_0;\gamma}_{\text{Coin}}$, $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ and $\mathcal{P}^{\gamma}_{\text{Coin}}$ provide corresponding sample complexity bounds.

Non-sequential Testing Algorithms

A first, naive idea for tackling $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ would be to restrict ourselves to *non-sequential* testing algorithms. Such a non-sequential algorithm $\mathcal{A}$ may choose a priori both $T \in \mathbb{N}$ and a function $f : \{0,1\}^T \to \{0,1\}$, observe $X_1, \dots, X_T$ and output $f(X_1, \dots, X_T)$ as decision. Obviously, the termination time $T^{\mathcal{A}}$ of such an algorithm is T. An easy application of Hoeffding's inequality shows: If we choose $T \in \mathcal{O}(\frac{1}{h^2} \ln \frac{1}{\gamma})$ large enough and let $f(X_1, \dots, X_T) = 0$ if $\sum_{t \le T} X_t \ge T/2$ and 1 otherwise, we obtain a non-sequential testing algorithm $\mathcal{A}$, which solves $\mathcal{P}^{h,\gamma}_{\text{Coin}}$. According to Anthony and Bartlett [1999], this solution is already asymptotically optimal among all possible non-sequential solutions. Its details and optimality are provided in the following lemma.

Lemma 2.10. *Let $h \in (0, 1/2)$ be fixed.*

(i) Let $\gamma \in (0,1)$ be arbitrary. Choose $T := \left\lceil \frac{1}{2h^2} \ln \frac{1}{\gamma} \right\rceil$ and define $f : \{0,1\}^T \to \{0,1\}$ via

$$f(x_1, \dots, x_T) := \begin{cases} 0, & \text{if } \frac{1}{T} \sum_{i=1}^T x_i < \frac{1}{2}, \\ 1, & \text{if } \frac{1}{T} \sum_{i=1}^T x_i \ge \frac{1}{2}. \end{cases}$$

The corresponding non-sequential testing algorithm $\mathcal{A}$ solves $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ and $\mathcal{P}^{1/2+h;\gamma}_{\text{Coin}}$.

(ii) Let $\gamma \in (0, 1/4)$ and suppose $\mathcal{A}$ to be a non-sequential testing algorithm, which solves $\mathcal{P}^{h,\gamma}_{\text{Coin}}$. Then, we have a.s.

$$T^{\mathcal{A}} \ge \frac{1}{2} \left\lfloor \frac{1 - 4h^2}{h^2} \ln \left(\frac{1}{8\gamma(1 - 2\gamma)} \right) \right\rfloor.$$

In particular, $T^{\mathcal{A}} \in \Omega\left(\frac{1}{h^2} \ln \frac{1}{\gamma}\right)$ as $\min\{h, \gamma\} \to 0$.

A restriction to non-sequential testing algorithms would come with several drawbacks:

- As we will see later in Prop. 2.17, the solution to $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$ from Lem. 2.10 is not optimal with respect to the more general class of sequential testing algorithms.
- As a solution to $\mathcal{P}_{\text{Coin}}^{\gamma}$ clearly solves $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ for every $h \in (0, 1/2)$, the lower bound from Lem. 2.10 shows that there does not exist a non-sequential solution to $\mathcal{P}_{\text{Coin}}^{\gamma}$.
- Of course, the lower bound from Lem. 2.10 only holds for non-sequential testing algorithms and is not applicable to sequential ones.

To overcome these issues, let us from now on consider the broader class of *sequential* testing algorithms. Since the feedback in the dueling bandits is observed in an online manner, a sequential testing algorithm is applicable there. Every such algorithm $\mathcal{A}$ executes in each time step $t \in \mathbb{N}$ both of the following steps:

(i) Observe X_t.

(ii) Either skip this step or terminate. In the latter case, return either 0 or 1 as decision.

The decision whether $\mathcal{A}$ decides at time t for executing (i) or (ii) may depend on the observations $X_1, \ldots, X_t$ it has made until time t. Thus the number $T^{\mathcal{A}}$ of samples, which $\mathcal{A}$ observes before termination, is a random variable with values in $\mathbb{N} \cup \{\infty\}$, where $T^{\mathcal{A}} = \infty$ means that $\mathcal{A}$ does not terminate.

In the following and throughout the entire thesis, the expectation $\mathbb{E}_p[T^{\mathcal{A}}]$ of such a probabilistic algorithm $\mathcal{A}$ is also taken with regard to the possibly probabilistic behavior of $\mathcal{A}$, i.e., formally we have $\mathbb{E}_{p,\mathcal{A}}[T^{\mathcal{A}}]$. Nevertheless, we may simply write $\mathbb{E}_p[T^{\mathcal{A}}]$ for convenience.

2.2.1. Lower Bounds

We start with a lower bound for the sample complexity of solutions to $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$, which is based on the optimality of the later on discussed *sequential probability ratio test*, cf. Prop. 2.17 below.

Proposition 2.11. *Suppose $0 < \gamma < \gamma_0 < 1/2$ and $0 < h < h_0 < 1/2$ to be fixed. If $\mathcal{A}$ is a solution to $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$, then*

$$\mathbb{E}_{1/2 \pm h}[T^{\mathcal{A}}] \geq \frac{1-2\gamma}{2h} \left\lceil \frac{\ln \frac{1-\gamma}{\gamma}}{\ln \frac{1/2+h}{1/2-h}} \right\rceil \geq \frac{c(h_0,\gamma_0)}{h^2} \ln \frac{1}{\gamma}$$

for some appropriate constant $c(h_0, \gamma_0)$, which does not depend on γ or h. In particular, any solution $\mathcal{A}$ to $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ fulfills

$$\sup_{p:|p-1/2|>h} \mathbb{E}_p\left[T^{\mathcal{A}}\right] \geq \frac{1-2\gamma}{2h} \left\lceil \frac{\ln \frac{1-\gamma}{\gamma}}{\ln \frac{1/2+h}{1/2-h}} \right\rceil \geq \frac{c(h_0,\gamma_0)}{h^2} \ln \frac{1}{\gamma}.$$

Proof. The first statement follows from the optimality of the sequential probability ratio test for $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$ and is stated and proven in detail in Prop. 2.17 below. To prove the second statement, note that any solution $\mathcal{A}$ to $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ also solves $\mathcal{P}_{\text{Coin}}^{1/2+h+\varepsilon;\gamma}$ for any $\varepsilon \in (0, 1-1/2-h]$, therefore, it can be inferred from the first one by taking the limit $\varepsilon \searrow 0$. $\square$

Alternatively, the worst-case asymptotic lower bound from Prop. 2.11 can be deduced in a slightly weaker version (that holds only for small values of γ) from results in [Mannor and Tsitsiklis, 2004]. For the sake of completeness, we state this result in the following.

Proposition 2.12. *Let $\gamma \in \left(0, \frac{1}{40e^8}\right)$ and $h \in \left(0, \frac{1}{4}\right)$ be fixed. If $\mathcal{A}$ solves $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$ resp. $\mathcal{P}_{\text{Coin}}^{h,\gamma}$, then*

$$\max_{p\in\{1/2\pm h\}} \mathbb{E}_p\left[T^{\mathcal{A}}\right] \geq \frac{c}{h^2}\ln\frac{1}{\gamma} \quad \textit{resp.} \quad \sup_{p:|p-1/2|>h} \mathbb{E}_p\left[T^{\mathcal{A}}\right] \geq \frac{c}{h^2}\ln\frac{1}{\gamma},$$

where $c > 0$ is a universal constant, which does not depend on h or γ.

If $\mathcal{A}$ solves $\mathcal{P}_{\text{Coin}}^{\gamma}$, then it is also a solution to $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ for every $h \in (0, 1/2)$. In this case, Prop. 2.11 shows that $\sup_{p\neq 1/2} \mathbb{E}_p[T^{\mathcal{A}}] = \infty$ holds. In contrast to solutions to $\mathcal{P}_{\text{Coin}}^{h,\gamma}$, it is not possible that $\mathcal{A}$ terminates for any $p \neq 1/2$ almost surely before some time $N \in \mathbb{N}$.

Proposition 2.13. *Let $\gamma \in (0, 1/2)$ be fixed. If $\mathcal{A}$ solves $\mathcal{P}_{\text{Coin}}^{\gamma}$, then*

$$\limsup_{h\to 0} \frac{\mathbb{E}_{1/2\pm h}\left[T^{\mathcal{A}}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \geq \frac{1}{2}\mathbb{P}_{1/2}(T^{\mathcal{A}} = \infty) \geq \frac{1-2\gamma}{2} > 0.$$

Proof. This is stated in Thm. 1 in [Farrell, 1964]. To verify this, note that $\frac{1}{|\ln|\ln|h|||} = \frac{1}{\ln\ln\frac{1}{h}}$ holds for $h < \frac{1}{e}$ and also confer the remark directly after Thm 1 therein. $\square$

Chen and Li [2015] demonstrate that the limes superior in Prop. 2.13 may not be avoided. Moreover, they provide a slightly stronger result than Prop. 2.13 (cf. Thm. D.1 there), but we do not discuss it here as it is not required in the further course of this thesis.

As a preparation for the proof of Prop. 5.3, we want to show lower bounds on the expected sample complexity of solutions to $\mathcal{P}_{\text{Coin}}^{p_0;p_1;\gamma}$ for arbitrary distinct $p_0, p_1 \in [0,1]$. Note that Prop. 2.17 provides a sample complexity lower bound for the particular case $p_0 = 1/2 - h$, $p_1 = 1/2 + h$. Reducing $\mathcal{P}_{\text{Coin}}^{1/2+h;1/2-h}$ to $\mathcal{P}_{\text{Coin}}^{p_0,p_1;\gamma}$ will thus allow us to infer a sample complexity lower bound for solutions to $\mathcal{P}_{\text{Coin}}^{p_0,p_1;\gamma}$. The following lemma plays a crucial role in this reduction.

Lemma 2.14. *Suppose $\{X^{(p)}\}_{p\in[0,1]}$ and $\{U^{(r)}\}_{r\in[0,1]}$ to be families of random variables $X^{(p)} \sim \text{Ber}(p)$ and $U^{(r)} \sim \text{Ber}(r)$ such that, for any $p, r \in [0,1]$, $X^{(p)}$ and $U^{(r)}$ are independent.*

(i) $Y^{(p,r)} := X^{(p)} + \mathbf{1}_{\{X^{(p)}=0\}}U^{(r)}$ fulfills $Y^{(p,r)} \sim \text{Ber}(p + (1-p)r)$.

(ii) If $p_0, p_1 \in [0,1]$ with $p_1 > p_0$ and $p_0 + p_1 \geq 1$ are fixed, we obtain with the choices

$$h := \frac{p_1 - p_0}{2(2 - p_1 - p_0)} \quad \textit{and} \quad r' := \frac{p_0 - (1/2 - h)}{1/2 + h}$$

that $Y^{(1/2-h,r')} \sim \text{Ber}(p_0)$ and $Y^{(1/2+h,r')} \sim \text{Ber}(p_1)$.

With this preparation, we can prove the following sample complexity lower bound for solutions to $\mathcal{P}_{\text{Coin}}^{p_0,p_1;\gamma}$.

Corollary 2.15. *For any fixed $\gamma_0 \in (0,1/2)$ and $\varepsilon_0 \in (0,1)$ there exists a constant $c = c(\varepsilon_0,\gamma_0) > 0$ with the following property: Let $\gamma \in (0,\gamma_0)$, $p_0,p_1 \in (0,1)$ with $p_0 < p_1 < p_0 + \varepsilon_0(1-p_0)$ and $p_0 + p_1 \geq 1$, and suppose $\mathcal{A}$ is any sequential testing algorithm, which solves $\mathcal{P}_{\text{Coin}}^{p_0,p_1;\gamma}$. Then,*

$$\min_{p\in\{p_0,p_1\}} \mathbb{E}_p[T^{\mathcal{A}}] \geq c \left(\frac{2(2-p_1-p_0)}{p_1-p_0}\right)^2 \ln\frac{1}{\gamma}.$$

Proof of Cor. 2.15. Suppose $\gamma_0 \in (0,1/2)$ and $\varepsilon_0 \in (0,1)$ to be fixed. Due to $\frac{4x}{1+2x} \to 1$ as $x \nearrow \frac{1}{2}$, we can fix $h_0 := h_0(\varepsilon_0) \in (0,1/2)$ with $\varepsilon_0 < \frac{4h_0}{1+2h_0}$. By assumption, we have

$$p_1 < p_0 + \frac{4h_0}{1+2h_0}(1-p_0) = \frac{4h_0 + p_0(1-2h_0)}{1+2h_0}$$

and thus

$$h := \frac{p_1 - p_0}{2(2-p_1-p_0)} \in (0,h_0).$$

Choose $c := c(h_0,\gamma_0)$ to be the corresponding constant from Prop. 2.17. Moreover, define $r := \frac{p_0-(1/2-h)}{1/2+h}$ and let $U_t^{(r)} \sim \text{Ber}(r)$ be such that the families $\{U_t^{(r)}\}_{t\in\mathbb{N}}$ and $\{X_t^{(p)}\}_{t\in\mathbb{N}}$ are independent. Let $\mathcal{A}'$ be the sequential algorithm, which simulates $\mathcal{A}$ in the following way: At each time step $t \in \mathbb{N}$, if $\mathcal{A}$ has not yet terminated, throw the coin C and observe feedback $X_t^{(p)}$ and pass the feedback $Y_t^{(p,r)} := X_t^{(p)} + \mathbf{1}_{\{X_t^{(p)}=0\}} U_t^{(r)}$ to $\mathcal{A}$. If $\mathcal{A}$ terminates with decision $w \in \{0,1\}$, $\mathcal{A}'$ terminates as well and outputs w.

Note that $\mathcal{A}'$ knows both the value of h as well as of p_0, which are necessary to compute r. Moreover, by construction $T^{\mathcal{A}} = T^{\mathcal{A}'}$ holds. From Lem. 2.14 we infer $Y_t^{(1/2-h,r)} \sim \text{Ber}(p_0)$ and $Y_t^{(1/2+h,r)} \sim \text{Ber}(p_1)$. As $\mathbf{D}(\mathcal{A}') = w$ iff $\mathbf{D}(\mathcal{A}) = w$, it follows that $\mathbb{P}_{1/2-h}(\mathbf{D}(\mathcal{A}') = 0) = \mathbb{P}_{p_0}(\mathbf{D}(\mathcal{A}) = 0) \geq 1-\gamma$ and $\mathbb{P}_{1/2+h}(\mathbf{D}(\mathcal{A}') = 1) = \mathbb{P}_{p_1}(\mathbf{D}(\mathcal{A}) = 1) \geq 1-\gamma$. Consequently, $\mathcal{A}'$ solves $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$. Thus, recalling the choice of c, Prop. 2.17 lets us infer that

$$\min_{p\in\{p_0,p_1\}} \mathbb{E}_p[T^{\mathcal{A}}] = \min_{p\in\{1/2\pm h\}} \mathbb{E}_p[T^{\mathcal{A}'}] \geq \frac{c}{h^2}\ln\frac{1}{\gamma},$$

which completes the proof due to the choice of h. □

Table 2.1 summarizes the previously presented sample complexity lower bounds for solutions to the different variants of the coin-tossing problem in a partly simplified manner.

2.2.2. Upper Bounds

Next, we discuss different approaches for sequential solutions to $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ and $\mathcal{P}_{\text{Coin}}^{\gamma}$. As already mentioned, $\mathcal{P}_{\text{Coin}}^{\gamma}$ cannot be solved by a non-sequential testing algorithm but requires non-sequential solutions. Apart from that, considering sequential tests for $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ seems reasonable as well, since these possibly allow for earlier termination than the non-sequential solution from above.

Table 2.1.: Sample complexity lower bounds for sequential solutions to the coin-tossing problems.

Problem	Theoretical statement	Bound
$\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$	$\mathbb{E}_{1/2\pm h}[T^{\mathcal{A}}] \in \Omega(\frac{1}{h^2}\ln\frac{1}{\gamma})$	Prop. 2.11
$\mathcal{P}_{\text{Coin}}^{h,\gamma}$	$\sup_{p:\lvert p-1/2\rvert>h} \mathbb{E}_p[T^{\mathcal{A}}] \in \Omega(\frac{1}{h^2}\ln\frac{1}{\gamma})$	Prop. 2.11
$\mathcal{P}_{\text{Coin}}^{\gamma}$	$\limsup_{h\to 0} \mathbb{E}_{1/2+h}\left[T^{\mathcal{A}}\right]/(\frac{1}{h^2}\ln\ln\frac{1}{h}) \geq 1/2-\gamma$	Prop. 2.13
$\mathcal{P}_{\text{Coin}}^{p_0,p_1;\gamma}$	$\min_{p\in\{p_0,p_1\}} \mathbb{E}_p[T^{\mathcal{A}}] \geq c\left(\frac{2(2-p_1-p_0)}{p_1-p_0}\right)^2 \ln\frac{1}{\gamma}$	Cor. 2.15

The Sequential Probability Ratio Test

A first approach for a sequential solution to coin-tossing is the *sequential probability ratio test* (SPRT), which has its origins in [Wald, 1945]. Let us illustrate the idea of the SPRT on a small example taken from [Siegmund, 1985, p. 10] and suppose we are interested in solving $\mathcal{P}_{\text{Coin}}^{p_0,p_1;\gamma}$, i.e., we want to test $\mathbf{H}_0 : p = p_0$ versus $\mathbf{H}_1 : p = p_1$ for some known distinct $p_0, p_1 \in [0,1]$. Performing a SPRT $\mathcal{A}$ here means to choose boundary values $-\infty < A' < B' < \infty$, sample until the sequential probability ratio $\lambda_n := \mathbb{P}_{p_0}(\forall i \leq n : X_i = x_i)/\mathbb{P}_{p_1}(\forall i \leq n : X_i = x_i)$ is not in (A', B') and then return 0, if $\lambda_n \geq B'$, and 1 otherwise. With $q_b := 1 - p_b$ for $b \in \{0,1\}$ and $S_n := \sum_{i=1}^n (2X_i - 1) = 2\sum_{i=1}^n X_i - n$ we get

$$\lambda_n = (p_1 p_0^{-1})^{(n+S_n)/2}(q_1 q_0^{-1})^{(n-S_n)/2} = (p_1 q_0 p_0^{-1} q_1^{-1})^{S_n/2}(p_1 p_0^{-1} q_1 q_0^{-1})^{n/2}.$$

This shows that there exist $A < B$ with $\lambda_n \leq A'$ iff $S_n \leq A$ and $\lambda_n \geq B'$ iff $S_n \geq B$, i.e., $\mathcal{A}$ may decide based on S_n instead of λ_n. This justifies the following definition of the SPRT that will be used throughout this section.

Definition 2.16. *For $-\infty < A < B < \infty$, the* sequential probability ratio test (SPRT) *with barriers A and B is the testing algorithm, which stops at the first time n where $S_n := \sum_{i=1}^n (2X_i - 1)$ is not in (A, B) and then outputs 0 if $S_n \geq B$ and 1 if $S_n \leq A$. For $B > 0$, the* symmetric SPRT with barrier B *is the SPRT with barriers $-B$ and B.*

Writing δ_x for the Dirac measure on $\{x\}$, which assigns any set $W \subseteq \mathbb{R}$ the measure 1 if $x \in W$ and 0 otherwise, $X_n \sim \text{Ber}(p)$ implies $2X_n - 1 \sim p\delta_1 + (1-p)\delta_{-1}$ and thus S_n is a random walk on $\mathbb{Z}$ with drift p. In Chapters 4 and 5, we will rather use the test statistic $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ instead of S_n, but since $S_n \geq B$ iff $\overline{X}_n \geq 1/2 + B/n$ and $S_n \leq A$ iff $\overline{X}_n \leq 1/2 - B/n$, this is unproblematic.

With the particular choices $-A = B = \left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+h)/(1/2-h))} \right\rceil$, the resulting SPRT can be shown to be an optimal solution to $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$ and also to $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ with sample complexity $\mathcal{O}(\frac{1}{h^2}\ln\frac{1}{\gamma})$. The details are presented in the upcoming proposition, which also contains and justifies the first statement of Prop. 2.11 from above.

Proposition 2.17. *Suppose $0 < \gamma < \gamma_0 < 1/2$ and $0 < h < h_0 < 1/2$ to be fixed.*

(i) The symmetric SPRT $\mathcal{A}$ with barrier $B := \left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+h)/(1/2-h))} \right\rceil$ solves $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$ and $\mathcal{P}_{\text{Coin}}^{h,\gamma}$, i.e.,

$$\forall p \geq 1/2+h : \mathbb{P}_p(\mathbf{D}(\mathcal{A}) = 0) \geq 1-\gamma \quad \textit{and} \quad \forall p \leq 1/2-h : \mathbb{P}_p(\mathbf{D}(\mathcal{A}) = 1) \geq 1-\gamma.$$

Moreover, the termination time $T^{\mathcal{A}}$ of $\mathcal{A}$ fulfills

$$\sup_{p\in[0,1/2-h]\cup[1/2+h,1]} \mathbb{E}_p[T^{\mathcal{A}}] = \mathbb{E}_{1/2\pm h}[T^{\mathcal{A}}] = \frac{1-2\gamma}{2h}\left\lceil \frac{\ln\frac{1-\gamma}{\gamma}}{\ln\frac{1/2+h}{1/2-h}} \right\rceil, \tag{2.8}$$

which is in $\mathcal{O}(\frac{1}{h^2}\ln\frac{1}{\gamma})$ as $\max\{\frac{1}{h},\frac{1}{\gamma}\}\to\infty$.

(ii) The testing algorithm $\mathcal{A}$ from (i) is w.r.t. $\mathbb{E}_{1/2+h}[T^{\mathcal{A}}]$ and $\mathbb{E}_{1/2-h}[T^{\mathcal{A}}]$ optimal among all solutions to $\mathcal{P}^{1/2+h;\gamma}_{\mathrm{Coin}}$. In other words: If $\mathcal{A}'$ is an algorithm, which fulfills

$$\mathbb{P}_{1/2+h}(\mathbf{D}(\mathcal{A}')=0)\geq 1-\gamma \quad \textit{and} \quad \mathbb{P}_{1/2-h}(\mathbf{D}(\mathcal{A}')=1)\geq 1-\gamma,$$

then it fulfills

$$\mathbb{E}_{1/2\pm h}[T^{\mathcal{A}'}] \geq \mathbb{E}_{1/2\pm h}[T^{\mathcal{A}}] = \frac{1-2\gamma}{2h}\left\lceil \frac{\ln\frac{1-\gamma}{\gamma}}{\ln\frac{1/2+h}{1/2-h}} \right\rceil \geq \frac{c(h_0,\gamma_0)}{h^2}\ln\frac{1}{\gamma}$$

for some appropriate constant $c(h_0,\gamma_0)$, which does not depend on γ or h.

Prop. 2.11 and Prop. 2.17 show that the SPRT from Prop. 2.17 is, w.r.t. the worst-case expected sample complexity, an asymptotically optimal solution to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$. For $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$, we have the following negative result.

Lemma 2.18. *There does not exist a SPRT that solves $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$.*

The Generalized Sequential Probability Ratio Test

For solving $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ we generalize the notion of SPRTs in a similar way as Farrell [1964].

Definition 2.19. *For $A,B:\mathbb{N}\to\mathbb{R}\cup\{\pm\infty\}$ with $A(n)\leq B(n)$ for all $n\in\mathbb{N}$, the* generalized sequential probability ratio test (GSPRT) *$\mathcal{A}$ with barriers A and B is that testing algorithm, which samples until the first time n where $S_n = 2\sum_{i=1}^n X_i - n \notin [A(n),B(n)]$ and outputs 0 if $S_n > B(n)$ and 1 if $S_n < A(n)$. If $B=-A$, we simply call $\mathcal{A}$ the* symmetric GSPRT *with barrier B.*

Example 2.20. *(i) Every SPRT is a GSPRT with constant barriers. In particular, the symmetric SPRT from Prop. 2.17 with barrier $B=\left\lceil\frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+h)/(1/2-h))}\right\rceil$ is a GSPRT that solves $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$.*

(ii) If T is odd, the test from Lem. 2.10 can be regarded as a symmetric GSPRT with a barrier B given by $B(n)=\infty$ for $n\leq T$ and $B(n)=0$ for $n>T$, which solves $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$. If T is even, the GSPRT with barrier B may terminate at time $T+1$ and does not coincide with the test from Lem. 2.10.

For the tests from Lem. 2.10 and Prop. 2.17, if γ is fixed, the expected termination time on the instance $\frac{1}{2}+h>\frac{1}{2}$ is monotonically increasing in h. The next lemma shows that this property holds in fact for any symmetric GSPRT. Its proof is based on a coupling argument and given in the appendix.

Lemma 2.21. *If $\mathcal{A}$ is a symmetric GSPRT, then*

$$\forall 1/2 \le p_1 \le p_2, \forall n \in \mathbb{N} : \mathbb{P}_{p_1}\left(T^{\mathcal{A}} \le n\right) \le \mathbb{P}_{p_2}\left(T^{\mathcal{A}} \le n\right).$$

In particular, the function $[\frac{1}{2}, 1] \to \mathbb{R} \cup \{\infty\}$, $h \mapsto \mathbb{E}_{\frac{1}{2}+h}[T^{\mathcal{A}}]$ is monotonically decreasing.

The following proposition provides us a solution $\mathcal{A}$ to $\mathcal{P}^{\gamma}_{\text{Coin}}$, which is, according to Prop. 2.13, for fixed γ w.r.t. its asymptotic behavior of $h \mapsto \mathbb{E}_{1/2\pm h}[T^{\mathcal{A}}]$ optimal. Therein, we abbreviate for the sake of convenience $\ln_2(\cdot) := \ln\ln(\cdot)$ and $\ln_3(\cdot) := \ln\ln\ln(\cdot)$. According to Farrell [1964], it is based on a version of the law of the iterated logarithm due to Cantelli [1933].

Proposition 2.22. *Let $\gamma \in (0, 1/2)$ be fixed. Let S'_n be a symmetric random walk on $\mathbb{Z}$, i.e. $S'_n = \sum_{i=1}^n X'_i$ where $X'_i \sim \frac{1}{2}(\delta_1 + \delta_{-1})$ are iid. For arbitrary $c > 3$ the number*

$$n_0 := \min\left\{ n \in \mathbb{N} \,\middle|\, \mathbb{P}\left(\exists \tilde{n} \ge n+1 \,:\, \left|S'_{\tilde{n}}\right| \ge \frac{\tilde{n}}{\sqrt{2}} \ln_2(\tilde{n}+e) + c\ln_3(\tilde{n}+e^e) \right) \le \gamma \right\}$$

is finite. The corresponding symmetric GSPRT $\mathcal{A}$ with the barrier $B^{\text{Farrell}}_{\gamma} : \mathbb{N} \to [0, \infty]$ given by

$$B^{\text{Farrell}}_{\gamma}(n) := \begin{cases} \sqrt{n\ln_2(n+e) + c\ln_3(n+e^e)}/\sqrt{2}, & \text{if } n \ge n_0 + 1, \\ n, & \text{otherwise} \end{cases}$$

solves $\mathcal{P}^{\gamma}_{\text{Coin}}$ and fulfills

$$\lim_{h\to 0} \frac{\mathbb{E}_{1/2\pm h}\left[T^{\mathcal{A}}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} = \frac{1}{2}\mathbb{P}_{1/2}(T^{\mathcal{A}} = \infty) > 0. \tag{2.9}$$

Note that $|S_n| = |\sum_{i=1}^n (2X_i - 1)| \le n = B^{\text{Farrell}}_{\gamma}(n)$ for $n \le n_0$ implies that the GSPRT from Prop. 2.22 does not terminate before n_0; in fact, we could have equivalently defined $B^{\text{Farrell}}_{\gamma}(n) = \infty$ for $n \le n_0$. As it seems hard to determine n_0 in Prop. 2.22 – and this value is presumably very large – this algorithm does not appear to be very practicable to us. For this reason, we provide further below two other solutions to $\mathcal{P}^{\gamma}_{\text{Coin}}$, which are obtained by means of a reduction to the multi-armed bandits scenario.

Before doing so, let us state yet another GSPRT, which solves $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ with different theoretical guarantees as the previous solutions. It is mainly based Lem. 2.6 from above, and for the sake of generality we state and analyze a general parameterized version of it.

Proposition 2.23. *(i) For $h \in (0, 1/2)$, $\gamma \in (0, 1)$ let $\varepsilon = \varepsilon(\gamma) \in (0, 1)$ and $\delta = \delta(\gamma) \in \left(0, \frac{\ln(1+\varepsilon)}{e}\right)$ be such that $\gamma = \frac{2+\varepsilon}{\varepsilon}\left(\frac{\delta}{\ln(1+\varepsilon)}\right)^{1+\varepsilon}$. Then, the symmetric GSPRT with barrier $B^{\text{LiL}}_{h,\gamma}$ defined via*

$$B^{\text{LiL}}_{h,\gamma}(n) := \max\left\{ 0, (1+\sqrt{\varepsilon})\sqrt{\frac{1}{2n}(1+\varepsilon)\ln\left(\frac{\ln((1+\varepsilon)n)}{\delta}\right)} - h \right\}. \tag{2.10}$$

solves $\mathcal{P}^{h,\gamma}_{\text{Coin}}$. Moreover, $n \mapsto B^{\text{LiL}}_{h,\gamma}(n)$ is monotonically decreasing.

(ii) Let $\gamma_0, \varepsilon_0 \in (0,1)$ be such that $(\frac{\varepsilon_0}{2+\varepsilon_0})^{\frac{1}{1+\varepsilon_0}} < \frac{1}{e}$ holds[2]*. With $\varepsilon(\gamma) := \varepsilon_0$ and $\delta(\gamma) := (\frac{\gamma\varepsilon_0}{2+\varepsilon_0})^{\frac{1}{1+\varepsilon_0}} \ln(1+\varepsilon_0) \in \left(0, \frac{\ln(1+\varepsilon_0)}{e}\right)$ we have $\gamma = \frac{2+\varepsilon(\gamma)}{\varepsilon(\gamma)} \left(\frac{\delta(\gamma)}{\ln(1+\varepsilon(\gamma))}\right)^{1+\varepsilon(\gamma)}$ and the symmetric GSPRT $\mathcal{A}$ from (i) started with h and γ fulfills*

$$\inf_{p\in[0,1/2-h)\cup(1/2+h,1]} \mathbb{P}_p\left(T^{\mathcal{A}} \leq N_0(h,\gamma)\right) = 1,$$

where

$$N_0(h,\gamma) := \frac{d_0}{h^2} \ln\left(\frac{1}{\delta(\gamma)} \ln \frac{(1+\varepsilon_0)d_0}{h^2\delta(\gamma)}\right) + 2 \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$$

with $d_0 := \frac{1}{2}(1+\sqrt{\varepsilon_0})^2(1+\varepsilon_0)$.

In Prop. 2.23 we have to choose $\varepsilon_0 \in (0,1)$ such that $f(\varepsilon_0) < 1/e$ for $f(x) := (\frac{x}{2+x})^{\frac{1}{1+x}}$. As $\ln\frac{x}{x+2} < 0$ for $x \in (0,1)$, we see that

$$f'(x) = \left(\frac{x}{x+2}\right)^{1/(x+1)-1} (x+1)^{-2}(x+2)^{-2}\left(2(x+1) - x(x+2)\ln\left(\frac{x}{x+2}\right)\right)$$

is positive on $(0,1)$. Together with $f(0.54) < 1/e$ this shows that in particular any $\varepsilon_0 \leq 0.54$ would be a valid choice. If $\varepsilon(\gamma) = \varepsilon_0 = 1/2$ and $\delta = (\frac{\gamma\varepsilon_0}{2+\varepsilon_0})^{\frac{1}{1+\varepsilon_0}} \ln(1+\varepsilon_0) = (\gamma/5)^{2/3}\ln(3/2)$, we obtain

$$B^{\mathrm{LiL}}_{h,\gamma}(n) := \max\left\{0, \frac{2+\sqrt{2}}{2}\sqrt{\frac{3}{4n}\ln\left(5^{2/3}\gamma^{-2/3}\ln(3n/2)/\ln(3/2)\right)} - h\right\}$$

and

$$N_0(h,\gamma) = \frac{3(2\sqrt{2}+3)}{8h^2}\ln\left(\frac{5^{2/3}}{\gamma^{2/3}\ln(3/2)}\ln\left(\frac{9(2\sqrt{2}+3)5^{2/3}}{16\ln(3/2)h^2\gamma^{2/3}}\right)\right) + 2.$$

The proof of Prop. 2.23 makes use of the following auxiliary lemma.

Lemma 2.24. *For any $N \geq 1$, $\varepsilon \in (0,1)$, $c > 0$ and $\delta \in (0,1]$ we have the implication*

$$N > \frac{1}{c}\ln\left(\frac{2}{\delta}\ln\frac{1+\varepsilon}{c\delta}\right) \quad\Rightarrow\quad \frac{1}{N}\ln\left(\frac{\ln((1+\varepsilon)N)}{\delta}\right) < c.$$

Proof. This is the contraposition of (1) on p. 6 in [Jamieson et al., 2013]. □

Note that the solution from Prop. 2.23 is guaranteed to terminate before some known time of order $\mathcal{O}(\frac{1}{h^2}(\ln\ln\frac{1}{h})\ln\frac{1}{\gamma})$ whilst still having the chance to terminate early in case $|p-1/2|$ is large. Such guarantees could not possibly be given by any SPRT: If $\mathcal{A}$ was a SPRT with barriers A and B that solves $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ for appropriately small fixed h and γ and terminates a.s. before some time $N \in \mathbb{N}$, then we would necessarily have $B, -A > 1$. But then, $\mathbb{P}_p(T^{\mathcal{A}} > 0) \geq \mathbb{P}_p(\forall n' \leq n : |S_n| \leq 1) > 0$ would hold for any $p \in (0,1)$ and every $n \in \mathbb{N}$, a contradiction to the existence of N.

For solving $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ whilst terminating a.s. before some time $\mathcal{O}(\frac{1}{h^2}\ln\frac{1}{\gamma})$, one may also combine the test from Prop. 2.23 with the non-sequential solution to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ from Lem. 2.10.

[2] E.g., $\varepsilon_0 = 1/2$ works here.

Corollary 2.25. *For fixed $\gamma \in (0,1)$, $h \in (0,1/2)$ let ε and δ be as in Prop. 2.23 and define $B : \mathbb{N} \to [0,\infty]$ via*

$$B(n) := \begin{cases} B^{\mathrm{LIL}}_{h,\gamma}(n), & \text{if } n \le \frac{1}{2h^2}\ln\frac{1}{\gamma} \\ 0, & \text{otherwise.} \end{cases}$$

with $B^{LIL}_{h,\gamma}$ as in Prop. 2.23. Then, the symmetric GSPRT $\mathcal{A}$ with barrier B solves $\mathcal{P}^{h,3\gamma}_{\mathrm{Coin}}$ and fulfills $P_p\left(T^{\mathcal{A}} \le \frac{1}{2h^2}\ln\frac{1}{\gamma} + 2\right) = 1$ for any $p \in [0,1]$.

Substituting $\gamma' = \frac{\gamma}{3}$, Cor. 2.25 yields a GSPRT $\mathcal{A}$, which solves $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ with the guarantee $\mathbb{P}_p\left(T^{\mathcal{A}} \le \frac{1}{2h^2}\ln\frac{3}{\gamma} + 2\right) = 1$ whenever $|p - 1/2| > h$. In particular, it fulfills $\sup_{p:|p-1/2|>h} \mathbb{E}_p[T^{\mathcal{A}}] \in \mathcal{O}(\frac{1}{h^2}\ln\frac{1}{\gamma})$ and is thus asymptotically optimal with regard to Prop. 2.11. In addition to Prop. 2.17 and Lem. 2.10, this can be seen as a third proof that the asymptotic lower bound from Prop. 2.13 is sharp. In comparison to $\mathcal{A}$, the test $\mathcal{A}'$ from Prop. 2.23 is only up to a factor of $\ln\ln\frac{1}{h}$ asymptotically optimal in this sense. However, until time $\left\lfloor \frac{1}{2h^2}\ln\frac{3}{\gamma} \right\rfloor$, the barriers of $\mathcal{A}$ are larger than those of $\mathcal{A}'$. Thus, if $\mathcal{A}$ and $\mathcal{A}'$ terminate early, $\mathcal{A}'$ is likely to terminate earlier than $\mathcal{A}$. In particular, one cannot say that $\mathcal{A}$ is per se better than $\mathcal{A}'$.

A Reduction to Multi-Armed Bandits

We continue the discussion on the problem $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ by presenting yet another approach to solve it, namely reducing it to the *best-arm identification problem* in the multi-armed bandits setting. Suppose we are given m arms $a^{[1]}, \dots, a^{[m]}$ and are allowed to pull at each time step one of them. When pulling arm $a^{[m]}$ for the n-th time, we obtain a (noisy) reward $Z^{[i]}_n$. We assume that $\{Z^{[i]}_n\}_{i\in[m],n\in\mathbb{N}}$ are independent and that there exists $p^{[1]}, \dots, p^{[m]}$ with $Z^{[i]}_n \sim \mathrm{Ber}(p^{[i]})$ for every $i \in [m]$. The arm $a^{[i^*]}$ is considered to be the best arm if it yields on average the highest reward among all the available arms, i.e., $p^{[i^*]} = \max\{p^{[i]} : i \in [m]\}$.

Alg. 1 is a general reduction of $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ to the best-arm identification problem with two arms. In order todecide whether the bias p of the observed iid samples $X_1, X_2, \dots$ is larger or smaller than $\frac{1}{2}$, it assigns the X_i's as reward values to the first arm and assigns iid samples from $\mathrm{Ber}(\frac{1}{2})$ to the second arm. Then, it executes a MAB algorithm $\mathcal{A}_{\mathrm{MAB}}$ for best-arm identification and outputs 0 iff the first arm is detected as the best arm and 1 otherwise. Since the first arm is the best one iff $p > \frac{1}{2}$, the decision will be correct provided the MAB algorithm decided correctly.

For $t \in \mathbb{N}$ we write $n^{[i]}(t)$ for the number of times arm $a^{[i]}$ has been pulled up to time t and note that $t = n^{[1]}(t) + \dots + n^{[m]}(t)$ holds. For the sake of convenience, we restrict ourselves to two choices of $\mathcal{A}_{\mathrm{MAB}}$ in the following. The first of these is lil'UCB from Jamieson et al. [2013] and provided as Alg. 2, and the second is Exponential-Gap Estimation from Karnin et al. [2013] and stated (for the sake of convenience only for 2 arms) as Alg. 3. The two propositions below state the guarantees we obtain when properly initializing Alg. 1 with these choices of $\mathcal{A}_{\mathrm{MAB}}$.

Algorithm 1 Reduction of $\mathcal{P}^{\gamma}_{\text{Coin}}$ to best-arm identification

Input: access to an iid sequence $\{X_n\}_{n\in\mathbb{N}}$ of Bernoulli random variables, a MAB algorithm $\mathcal{A}_{\text{MAB}}$ for best-arm identification

Initialization: Let $\{Y_n\}_{n\in\mathbb{N}}$ be iid with $Y_n \sim \text{Ber}(1/2)$ and independent of $\{X_n\}_{n\in\mathbb{N}}$
Create arms $a^{[1]}, a^{[2]}$ with rewards $Z^{[1]}_n = X_n$ and $Z^{[2]}_n = Y_n$, for all $n \in \mathbb{N}$, respectively

1: Simulate $\mathcal{A}_{\text{MAB}}$ with the two-armed bandit with arms $a^{[1]}$ and $a^{[2]}$
2: **if** $\mathcal{A}_{\text{MAB}}$ predicts $a^{[1]}$ to be the best arm **then return** 0
3: **else return** 1

Algorithm 2 LIL'UCB

Input: algorithm parameters $\delta, \varepsilon, \lambda, \beta > 0$, and a MAB instance with arms $a^{[1]}, \dots, a^{[m]}$ with associated rewards $\{Z^{[1]}_n\}_{n\in\mathbb{N}}, \dots, \{Z^{[m]}_n\}_{n\in\mathbb{N}}$

Initialization: Sample each of the m arms once (i.e., $n^{[i]}(t) = 1$ for all i and $t = n$)

1: **while** $\forall i \in [m] : n^{[i]}(t) < 1 + \lambda \sum_{j\neq i} n^{[j]}(t)$ **do**
2: Pull arm $a^{[I_t]}$, where I_t is

$$\text{argmax}_{i\in[m]} \left\{ \frac{1}{n^{[i]}(t)} \sum_{n'=1}^{n^{[i]}(t)} Z^{[i]}_{n'} + (1+\beta)(1+\sqrt{\varepsilon}) \sqrt{\frac{(1+\varepsilon)\ln\left(\ln\left((1+\varepsilon)n^{[i]}(t)\right)/\delta\right)}{2n^{[i]}(t)}} \right\}$$

3: Update $n^{[k]}(t+1) \leftarrow n^{[k]}(t) + \mathbf{1}_{\{k=I_t\}}$ for all $k \in [m]$, then let $t \leftarrow t+1$
4: **return** $\text{argmax}_{i\in[m]} n^{[i]}(t)$.

Proposition 2.26. *Let $\varepsilon \in (0,1)$ be fixed, define $c_\varepsilon := \frac{2+\varepsilon}{\varepsilon}\left(\frac{1}{\ln(1+\varepsilon)}\right)^{1+\varepsilon}$. Let $\gamma \in (0,1)$ be such that $\delta := \frac{\gamma}{8c_\varepsilon} \in \left(0, \frac{\ln(1+\varepsilon)}{ec_\varepsilon}\right)$ holds and choose $\beta \in (0,3]$ arbitrarily. Then, there exists a constant $\lambda > 0$ with the following property: Denote by $\mathcal{A}$ Alg. 1 with sample access to $\{X_n\}_{n\in\mathbb{N}}$, where the black-box component $\mathcal{A}_{\text{MAB}}$ is* LIL'UCB *instantiated with $m = 2$, $\delta, \varepsilon, \lambda$ and β. Then, $\mathcal{A}$ solves $\mathcal{P}^{\gamma}_{\text{Coin}}$ and fulfills*

$$\sup_{p:|p-1/2|>h} \mathbb{P}_p\left(T^{\mathcal{A}} \leq C(h,\gamma)\right) \geq 1-\gamma \tag{2.11}$$

where $C(h,\gamma) \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ as $\min\{h,\gamma\} \to 0$. Moreover, for any $t' \in \mathbb{N}$ and $p \in [0,1]$ we have $\mathbb{P}_p(T^{\mathcal{A}} > t') > 0$.

Jamieson et al. [2013, p. 5] suggest to use $\varepsilon = 0.01, \beta = 1$ and $\lambda = \left(\frac{2+\beta}{\beta}\right)^2$ as parameters for LIL'UCB, for which their theoretical guarantees are proven to hold. However, they point out that the choice $\varepsilon = 0$, $\beta = 1/2$, $\lambda = 1 + 1/m$ and $\gamma \in (0,1)$ works well in their experiments even if they are not formally allowed.

Proposition 2.27. *Let $\gamma \in (0,1)$ be arbitrary. Write $\mathcal{A}$ for Alg. 1 with sample access to $\{X_n\}_{n\in\mathbb{N}}$ with $\mathcal{A}_{\text{MAB}}$ chosen to be Alg. 3 called with γ. Then, $\mathcal{A}$ solves $\mathcal{P}^{\gamma}_{\text{Coin}}$ and fulfills*

$$\sup_{p:|p-1/2|>h} \mathbb{P}_p\left(T^{\mathcal{A}} \leq C(h,\gamma)\right) \geq 1-\gamma, \tag{2.12}$$

where $C(h,\gamma) \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ as $\min\{h,\gamma\} \to 0$. Moreover, for any $t' \in \mathbb{N}$ and any $p \in [0,1]$ we have $\mathbb{P}_p(T^{\mathcal{A}} > t') > 0$.

Algorithm 3 EXPONENTIAL-GAP ESTIMATION for 2 arms

Input: $\gamma > 0$

Initialization: $r \leftarrow 1$

Notation: Write $i^\perp := 3 - i$, i.e., $1^\perp = 2$ and $2^\perp = 1$.

```
while True do
    S ← {1,2}, ε_r ← 1/2^{r+2}, γ_r ← γ/(50 r^3)
    Initialize ε̃_1 ← ε_r/8, γ̃_1 ← γ_r
    while |S| > 1 do
        ∀k ∈ {1,2}: Pull a^[k] for ⌈(4/ε̃_l^2) ln(3/γ̃_l)⌉ times, let p̃_l^[i] be its average reward
        if ∃ k ∈ {1,2} with p̃_l^[k⊥] < p̃_l^[k] then S ← {k}
        Update ε̃_{l+1} ← (3/4) ε̃_l, γ̃_{l+1} ← γ̃_l/2 and then l ← l + 1
    i ← the unique element in S
    ∀k ∈ {1,2}: Continue pulling a^[k] s.t. it has been pulled ≥ ⌈(2/ε_r^2) ln(2/γ_r)⌉ times at this
        iteration of the outer while-loop; let p̂_r^[k] be its average reward
    if p̂_r^[i⊥] < p̂_r^[i] − ε_r then return i
    Update r ← r + 1
```

We remark that the solutions from Prop. 2.26 and Prop. 2.27 to $\mathcal{P}^{\gamma}_{\text{Coin}}$ are *not* GSPRTs in the sense of Def. 2.19. To see this indirectly for the solution $\mathcal{A}$ from Prop. 2.26, suppose $\mathcal{A}$ was a GSPRT with barriers $A, B : \mathbb{N} \to [0, \infty]$ with $A(n) \le B(n)$ for all $n \in \mathbb{N}$. According to its definition, $\mathbb{P}_p(T^{\mathcal{A}} = 0) = 0$ holds, and Prop. 2.26 yields $\mathbb{P}_p(T^{\mathcal{A}} > t') > 0$ for every $t' \in \mathbb{N}$. Hence, we have $\mathbb{P}_p(T^{\mathcal{A}} \in \{0,1\}) \in (0,1)$. Now, a closer look at $\mathcal{A}$ (the algorithm from Prop. 2.26) reveals that

$$\forall p \in [0,1]\ \forall n \in \mathbb{N} : \mathbb{P}_p(T^{\mathcal{A}} = n \text{ and } \mathbf{D}(\mathcal{A}) = 0) = \mathbb{P}_{1-p}(T^{\mathcal{A}} = n \text{ and } \mathbf{D}(\mathcal{A}) = 1).$$

Due to this, we may suppose w.l.o.g. that $B(n) = -A(n)$ holds for every $n \in \mathbb{N}$, i.e., $\mathcal{A}$ is a symmetric GSPRT. As $S_1 = (2X_1 - 1) \sim p\delta_1 + (1-p)\delta_{-1}$, $\mathbb{P}_p(T^{\mathcal{A}} > 1) > 0$ is only possible if $B(1) \ge 1$, and $B(1) \ge 1$ implies $T^{\mathcal{A}} > 1$ a.s. Consequently, $\mathbb{P}_p(T^{\mathcal{A}} \le 1) \in \{0,1\}$ has to hold, a contradiction. Exactly the same argumentation shows that the solution from Prop. 2.27 to $\mathcal{P}^{\gamma}_{\text{Coin}}$ is nto a GSPRT.

There exist further best-arm identification algorithms $\mathcal{A}_{\text{MAB}}$ for MABs, which could be of interest for this latter approach such as [Kalyanakrishnan et al., 2012, Shah et al., 2020] and [Kaufmann and Kalyanakrishnan, 2013] just to mention a few. Since our main focus is on asymptotic optimality (up to logarithmic factors), which is already provided by the solutions from Prop. 2.17 and Prop. 2.22, we do not further elaborate on this.

A Bayesian Approach

Suppose $p \neq 1/2$ fixed but unknown and write $X_t = X_t^{(p)}$. An alternative approach for solving $\mathcal{P}^{\gamma}_{\text{Coin}}$ could be to assume a prior distribution μ_0 for the underlying value of p and update the corresponding posterior μ_t when observing sample X_t motivated by Bayes' rule as

$$\mu_t(\hat{p}) \coloneqq \frac{\mu_{t-1}(\hat{p})\hat{p}^{\mathbf{1}\{X_t=1\}}(1-\hat{p})^{\mathbf{1}\{X_t=0\}}}{\int_{\tilde{p}=0}^{1}\mu_{t-1}(\tilde{p})\tilde{p}^{\mathbf{1}\{X_t=1\}}(1-\tilde{p})^{\mathbf{1}\{X_t=0\}}}.$$

When choosing as μ_0 the beta distribution (e.g., the uniform distribution), which is a well-known *conjugate prior* in Bayesian statistics, the resulting posterior μ_t is assured to be a beta distribution as well. Recently, Waudby-Smith and Ramdas [2020] showed that the *prior-posterior ratio* (PPR) $R_t(\hat{p}) \coloneqq \frac{\mu_0(\hat{p})}{\mu_t(\hat{p})}$ is for the true value $\hat{p} = p$ a martingale, and an appropriate version of Ville's inequality for nonnegative supermartingales allowed them to show that the sets $C_t \coloneqq \{\hat{p} : R_t(\hat{p}) < 1/\gamma\}$ form a $(1-\gamma)$-*confidence sequence* in the sense that $\mathbb{P}(\exists t \geq 1 : p \notin C_t) \leq \gamma$. As pointed out by Jain et al. [2021], when choosing $\mu_0 \equiv 1$ (the uniform distribution on $[0,1]$), the corresponding belief distribution μ_t has the convenient form $\mu_t(\hat{p}) = f_{\mathrm{Beta}}(\hat{p}; W_t^1 + 1, W_t^0 + 1)$, where $W_t^1 = \sum_{t'=1}^{t} X_t$ resp. $W_t^0 \coloneqq \sum_{t'=1}^{t}(1 - X_t)$ is the number of observed 1's resp. 0's until time t and $f_{\mathrm{Beta}}(\hat{p}; a, b)$ denotes the pdf of a Beta-distribution with parameters a and b evaluated at $\hat{p}$. By terminating with the most frequent observation as soon as the $(1-\gamma)$-confidence sequence does not contain 1/2, they obtained a solution PPR-Bernoulli to $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$. We state it in our notations in Alg. 4, where we conveniently write $(W_t^{(0)}, W_t^{(1)})$ for the order statistics of $\left(W_t^0, W_t^1\right)$, i.e., $W_t^{(0)} = \max\{W_t^0, W_t^1\}$ and $W_t^{(1)} = \min\{W_t^0, W_t^1\}$.

Algorithm 4 PPR-Bernoulli

Input: $\gamma > 0$, sample access to $\mathrm{Ber}(p)$
Initialization: $t \leftarrow 1$

while True **do**
  Observe X_t and update W_t^0, W_t^1
  if $f_{\mathrm{Beta}}(\frac{1}{2}; W_t^{(0)} + 1, W_t^{(1)} + 1) \leq \gamma$ **then**
    return $\mathrm{argmax}_{b \in \{0,1\}} W_t^b$
  $t \leftarrow t + 1$

Proposition 2.28. *For any $\gamma \in (0,1)$, $\mathcal{A} \coloneqq$Alg. 4 solves $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ and for all $p > 1/2$*

$$\mathbb{P}_p\left(T^{\mathcal{A}} \leq T_\gamma(p)\right) \geq 1 - \gamma, \tag{2.13}$$

where $T_\gamma(p) \coloneqq \frac{20.775p}{(p-1/2)^2} \ln\left(\frac{2.49}{(p-1/2)^2\gamma}\right)$ fulfills $\sup_{p:p>1/2+h} T_\gamma(p) \in \Theta\left(\frac{1}{h^2}\ln\frac{1}{h^2\gamma}\right)$.

Proof. The proof of Thm. 7 in [Jain et al., 2021] shows that $\mathbb{P}_p\left(\frac{1}{2} \notin C_{T_\gamma(p)}\right) \geq 1 - \gamma$ is fulfilled for any $p > 1/2$, i.e., (2.13) holds. Moreover, monotonicity of $\gamma \mapsto T_\gamma(p)$ assures

$$\mathbb{P}_p(T^{\mathcal{A}} < \infty) \geq \mathbb{P}_p\left(\bigcup\nolimits_{\gamma \in (0,1)} \{T^{\mathcal{A}} < T_\gamma(p)\}\right) = \lim_{\gamma \to 0} \mathbb{P}_p\left(T^{\mathcal{A}} < T_\gamma(p)\right) \geq \lim_{\gamma \to 0} 1 - \gamma = 1$$

for all $p \neq 1/2$. □

In comparison to the solutions from Prop. 2.22, Prop. 2.26 and Prop. 2.27, the theoretical guarantee of PPR-Bernoulli is thus in a worst-case sense asymptotically suboptimal by a factor $\frac{\ln\frac{1}{h}}{\ln\ln\frac{1}{h}}$. However, to the best of our knowledge, it is w.r.t. the empirical

Table 2.2.: Sample complexity upper bounds for $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$, $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ and $\mathcal{P}_{\text{Coin}}^{\gamma}$.

Alg.	**Idea**	**Problem**	**Theoretical Guarantees**
Prop. 2.17	SPRT	$\mathcal{P}_{\text{Coin}}^{h,\gamma}$, $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$	w.r.t. $\mathbb{E}_{1/2\pm h}[T^{\mathcal{A}}]$ best solution to $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$ $\sup_{p:\lvert p-1/2\rvert>h} \mathbb{E}_p[T^{\mathcal{A}}] \in \mathcal{O}\left(\frac{1}{h^2}\ln\frac{1}{\gamma}\right)$
Prop. 2.22	GSPRT & LIL (from Cantelli)	$\mathcal{P}_{\text{Coin}}^{\gamma}$	$\forall\gamma < 1/2$: $\sup_{p:\lvert p-1/2\rvert>h} \mathbb{E}_p[T^{\mathcal{A}}] \in \mathcal{O}\left(\frac{1}{h^2}\ln\ln\frac{1}{h}\right)$ $T^{\mathcal{A}} < \infty$ a.s. $\forall p \neq 1/2$
Prop. 2.23	GSPRT & LIL	$\mathcal{P}_{\text{Coin}}^{h,\gamma}$	$T^{\mathcal{A}} \le f(h,\gamma)$ a.s., $f(h,\gamma) \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ $T^{\mathcal{A}} < \infty$ a.s. $\forall p \neq 1/2$
Cor. 2.25	GSPRT, Prop. 2.23 & Lem. 2.10	$\mathcal{P}_{\text{Coin}}^{h,\gamma}$	$T^{\mathcal{A}} \le f(h,\gamma)$ a.s., $f(h,\gamma) \in \mathcal{O}\left(\frac{1}{h^2}\ln\frac{1}{\gamma}\right)$ $T^{\mathcal{A}} < \infty$ a.s. $\forall p \neq 1/2$
Prop. 2.26	Reduction to lil'UCB	$\mathcal{P}_{\text{Coin}}^{\gamma}$	$T^{\mathcal{A}} \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ with prob. $\ge 1-\gamma$ $T^{\mathcal{A}} < \infty$ a.s. $\forall p \neq 1/2$
Prop. 2.27	Reduction to Exp.-gap est.	$\mathcal{P}_{\text{Coin}}^{\gamma}$	$T^{\mathcal{A}} \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ with prob. $\ge 1-\gamma$ $T^{\mathcal{A}} < \infty$ a.s. $\forall p \neq 1/2$
Prop. 2.28	Conf. sequ. for PPR	$\mathcal{P}_{\text{Coin}}^{\gamma}$	$T^{\mathcal{A}} \in \mathcal{O}\left(\frac{1}{h^2}\ln\frac{1}{h\gamma}\right)$ with prob. $\ge 1-\gamma$ $T^{\mathcal{A}} < \infty$ a.s. $\forall p \neq 1/2$
Prop. 2.40	DKWT for $k=2$	$\mathcal{P}_{\text{Coin}}^{\gamma}$	$T^{\mathcal{A}} \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h} + \ln\frac{1}{\gamma}\right)\right)$ with prob. $\ge 1-\gamma$ $T^{\mathcal{A}} < \infty$ a.s. $\forall p \neq 1/2$

performance currently the best solution in the literature, and we will see in Sec. 2.4 that it in fact performs very well in practice and outperforms the other solutions.

Table 2.2 summarizes the previously stated upper bounds in partly simplified form. For the sake of completeness, we also added the solution from Prop. 2.40 below for the case $k = 2$, in which $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$ conincides with $\mathcal{P}_{\text{Coin}}^{\gamma}$. Apart from the fact that Prop. 2.26 and Prop. 2.27 have the same worst-case asymptotic guarantees and imply those from Prop. 2.28 and Prop. 2.40 for $k = 2$, none of the presented theoretical guarantees seems to formally imply any of the others: E.g., the bound on $\mathbb{E}_p[T^{\mathcal{A}}]$ from Prop. 2.22 is neither necessary nor sufficient for the almost sure bound from Prop. 2.23. Hence, all the presented bounds appear to be of interest for themselves.

2.2.3. Lower Bounds for Multiple Coin Problems

In the folllowing, we prove lower bounds for testing problems involving multiple coins, which will be of use later on in Part II. Let J be some finite index set and suppose we are given independent coins $C_j, j \in J$, with unknown head probabilities $p_j, j \in J$, respectively. For fixed (unknown) $\mathbf{p} = (p_j)_{j\in J}$, throwing coin C_j at time t results in the feedback $Y_{t,j} \sim \text{Ber}(p_j)$, and we suppose the feedback is independent over time and coins, i.e., $\{Y_{t,j}\}_{j\in J, t\in\mathbb{N}}$ is independent. Let us define the hypothesis

$$\mathbf{H}_{0;J} : \forall j \in J : p_j > \frac{1}{2} \qquad \text{and} \qquad \mathbf{H}_{1;J} : \exists j \in J : p_j \le \frac{1}{2}. \tag{2.14}$$

If $\mathcal{A}$ is a (sequential probabilistic) testing algorithm for $\mathbf{H}_{0;J}$ versus $\mathbf{H}_{1;J}$, we may write

$\mathbf{D}(\mathcal{A}) = 0$ if $\mathcal{A}$ decides for $\mathbf{H}_{0;J}$ and $\mathbf{D}(\mathcal{A}) = 1$ if it decides for $\mathbf{H}_{1;J}$. Similarly as before, we write $T^{\mathcal{A}}$ for the stopping time of the algorithm, that is, the number of samples queried until termination, i.e., the total number of coin tosses until termination. We denote by $\mathbb{P}_{\mathbf{p}}$ the probability distribution on the different possible states of the algorithm, if the true parameter is $\mathbf{p}$, and write $\mathbb{E}_{\mathbf{p}}$ for the expectation with respect to $\mathbb{P}_{\mathbf{p}}$.

For the proof of Thm. 5.4 we will make use of the following lower bound on the expected termination time $\mathbb{E}_{\boldsymbol{p}}[T^{\mathcal{A}}]$ for algorithms $\mathcal{A}$ which test the above mentioned hypothesis, if it is known in advance that $|p_j - \frac{1}{2}| = h_j$ holds for every $j \in J$. For its proof, we will make use of the optimality of the SPRT as stated in Prop. 2.17. The following theorem is a slightly more general version of the result, which we presented as Lem. G.1 in [Haddenhorst et al., 2021a] and as Lem. A.7 in [Haddenhorst et al., 2021b]; it is given in this form merely for the sake of generality, we do not explicitly require this stronger version in the further course of this thesis, but instead its weaker counterpart would suffice.

Theorem 2.29. *Let $h_0, \gamma_0 \in (0, 1/2)$ be fixed, $\gamma \in (0, \gamma_0)$ and J be some arbitrary finite index set. Suppose $\{h_j\}_{j\in J} \subseteq (0, h_0)$, define $\mathbf{p}' := (\frac{1}{2} + h_j)_{j\in J} \in (0, 1)^J$ and for each $j \in J$ also $\mathbf{p}^{(j)} \in (0, 1)^J$ via $p^{(j)}_j := 1/2 - h_j$ and $p^{(j)}_{j'} := \frac{1}{2} + h_{j'}$ for all $j' \neq j$. Suppose $\mathcal{A}$ to be a (probabilistic) testing algorithm, which, provided the fact*

$$\mathbf{p} \in \mathfrak{P} := \{\mathbf{p}'\} \cup \bigcup\nolimits_{j\in J} \left\{\mathbf{p}^{(j)}\right\}$$

is known whereas the concrete value of $\mathbf{p}$ is unknown beforehand, is able to test $\mathbf{H}_{0;J}$ versus $\mathbf{H}_{1;J}$ with error probability at most γ. In other words, $\mathcal{A}$ fulfills

$$\mathbb{P}_{\mathbf{p}'}(\mathbf{D}(\mathcal{A}) = 0) \geq 1 - \gamma \quad \textit{and} \quad \forall \mathbf{p} \in \mathfrak{P} \setminus \{\mathbf{p}'\} : \mathbb{P}_{\mathbf{p}}(\mathbf{D}(\mathcal{A}) = 1) \geq 1 - \gamma.$$

Then, there exist some constant $c = c(h_0, \gamma_0) > 0$, which does not depend on h,γ or m, such that the corresponding stopping time $T^{\mathcal{A}}$ of $\mathcal{A}$ fulfills

$$\mathbb{E}_{\mathbf{p}'}[T^{\mathcal{A}}] \geq \sum\nolimits_{j\in J} \frac{1-2\gamma}{2h_j} \left\lceil \frac{\ln \frac{1-\gamma}{\gamma}}{\ln \frac{1/2+h_j}{1/2-h_j}} \right\rceil \geq c \sum\nolimits_{j\in J} \frac{1}{h_j^2} \ln \frac{1}{\gamma}.$$

and for any $j \in J$ also

$$\mathbb{E}_{\mathbf{p}^{(j)}}[T^{\mathcal{A}}] \geq \frac{1-2\gamma}{2h_j} \left\lceil \frac{\ln \frac{1-\gamma}{\gamma}}{\ln \frac{1/2+h_j}{1/2-h_j}} \right\rceil \geq \frac{c}{h_j^2} \ln \frac{1}{\gamma},$$

which naturally coincides with the lower bound for $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ from Prop. 2.17.

Proof of Thm. 2.29. At first, note that the case $|J| = 1$ of Thm. 2.29 corresponds to the testing problem considered in Prop. 2.17. For the sake of convenience, suppose without loss of generality that $J = [N]$. For an algorithm $\mathcal{A}'$ with sample access to $(p_1, \ldots, p_N)$, write $T^{\mathcal{A}'}_j$ for the number of times $\mathcal{A}'$ queries the coin C_j (with bias p_j) until termination. Moreover, for $j \in [N]$ and $p \in [0, 1]$ define $\mathbf{p}^{[j]}(p) = (p^{[j]}_{j'}(p))_{j'\in[N]}$ where $p^{[j]}_j(p) = p$ and $p^{[j]}_{j'}(p) = \frac{1}{2} + h_{j'}$ for $j' \neq j$. Note that $\mathbf{p}^{[j]}(\frac{1}{2} + h_j) \in \mathfrak{P}$ and $\mathbf{p}^{[j]}(\frac{1}{2} - h_j) \in \mathfrak{P}$ hold.
Now, suppose $\mathcal{A}$, $\mathbf{p}'$ and $\mathbf{p}^{(j)}$ to be as in the statement of this lemma. Let $\mathcal{A}(j)$ be the

algorithm, which is given sample access to p_j as input, simulates $\mathcal{A}$ with sample access to $\mathbf{p}^{[j]}(p_j)$ as input, terminates as soon as $\mathcal{A}$ terminates and outputs 0 if $\mathcal{A}$ outputs 0 and outputs 1 if $\mathcal{A}$ outputs 1. As $\mathcal{A}$ is able to decide $\mathbf{H}_{0;[N]} : \forall j \in [N] : p_j > \frac{1}{2}$ versus $\mathbf{H}_{1;[N]} : \exists j \in [N] : p_j < \frac{1}{2}$ with error probability at most γ for every $\mathbf{p} \in \mathfrak{P}$, $\mathcal{A}(j)$ is able to decide whether $p_j > \frac{1}{2}$ or $p_j < \frac{1}{2}$ with error probability at most γ in both cases $p_j = \frac{1}{2} + h_j$ and $p_j = \frac{1}{2} - h_j$. Prop. 2.17 assures that $\mathcal{A}(j)$ fulfills

$$\mathbb{E}_{\frac{1}{2}\pm h_j}[T^{\mathcal{A}(j)}] \geq \frac{1-2\gamma}{2h_j}\left\lceil \frac{\ln\frac{1-\gamma}{\gamma}}{\ln\frac{1/2+h_j}{1/2-h_j}} \right\rceil \geq \frac{c}{h_j^2}\ln\frac{1}{\gamma}$$

with $c = c(h_0, \gamma_0) > 0$ as in Prop. 2.17, where $T^{\mathcal{A}(j)}$ denotes the number of times algorithm $\mathcal{A}(j)$ with sample access to $\mathbf{p}^{[j]}(p_j)$ queries any of the coins $C_1, \ldots, C_N$ before termination. As deciding whether $p_j > \frac{1}{2}$ or $p_j < \frac{1}{2}$ does not require knowledge about any of the coins $C_{j'}$, $j' \neq j$, which are independent of C_j, we may assume without loss of generality that $\mathcal{A}(j)$ throws *only* coin C_j for this purpose.[3] Regarding that $\mathbf{p}^{[j]}(\frac{1}{2} + h_j) = \mathbf{p}'$ and $\mathbf{p}^{[j]}(\frac{1}{2} - h_j) = \mathbf{p}^{(j)}$ hold, we obtain

$$\mathbb{E}_{\mathbf{p}'}[T_j^{\mathcal{A}}] = \mathbb{E}_{\frac{1}{2}+h_j}[T^{\mathcal{A}(j)}] \geq \frac{1-2\gamma}{2h_j}\left\lceil \frac{\ln\frac{1-\gamma}{\gamma}}{\ln\frac{1/2+h_j}{1/2-h_j}} \right\rceil$$

and

$$\mathbb{E}_{\mathbf{p}^{(j)}}[T^{\mathcal{A}}] \geq \mathbb{E}_{\mathbf{p}^{(j)}}[T_j^{\mathcal{A}}] = \mathbb{E}_{\frac{1}{2}-h_j}[T^{\mathcal{A}(j)}] \geq \frac{1-2\gamma}{2h_j}\left\lceil \frac{\ln\frac{1-\gamma}{\gamma}}{\ln\frac{1/2+h_j}{1/2-h_j}} \right\rceil \geq \frac{c}{h_j^2}\ln\frac{1}{\gamma}.$$

As this holds for each $j \in [N]$ we get

$$\begin{aligned}\mathbb{E}_{\mathbf{p}'}[T^{\mathcal{A}}] = \sum\nolimits_{j\in[N]} \mathbb{E}_{\mathbf{p}'}[T_j^{\mathcal{A}}] &\geq \sum\nolimits_{j\in[N]} \frac{1-2\gamma}{2h_j}\left\lceil \frac{\ln\frac{1-\gamma}{\gamma}}{\ln\frac{1/2+h_j}{1/2-h_j}} \right\rceil \\ &\geq c\sum\nolimits_{j\in[N]} \frac{1}{h_j^2}\ln\frac{1}{\gamma},\end{aligned}$$

which completes the proof. □

2.3. Sequentially Testing for the Mode of a Biased Die

Similarly as Sec. 2.2 prepares our results on dueling bandits in Part II, this section provides useful tools for those results in the multi-dueling bandits scenario in Part III. Recall that in the multi-dueling bandits setting with winner feedback, which we focus on in this

[3] To see this formally, suppose on the contrary that $T_j^{\mathcal{A}(j)} < \frac{c}{h^2}\ln\frac{1}{\gamma}$. Let $\widetilde{\mathcal{A}}(j)$ be given sample access to $\mathbf{p}^{[j]}(p_j)$ and behave as $\mathcal{A}(j)$ with the only difference that samples from any coin $C_{j'} \neq C_j$ are replaced by an artificial sample $\mathrm{Ber}(p_{j'})$, which is independent of all the coins. Then, none of the coins $C_{j'} \neq C_j$ are thrown, we have $T^{\widetilde{\mathcal{A}}(j)} = T_j^{\widetilde{\mathcal{A}}(j)}$ and thus $\mathbb{E}_{1/2+h_j}[T^{\widetilde{\mathcal{A}}(j)}] = \mathbb{E}_{1/2+h_j}[T_j^{\widetilde{\mathcal{A}}(j)}] = \mathbb{E}_{1/2+h_j}[T_j^{\mathcal{A}(j)}] < \frac{c}{h^2}\ln\frac{1}{\gamma}$, a contradiction to Prop. 2.17.

thesis, the learner chooses at each time a set (called *multi-duel*) $S_t \in [m]_k$ and observes as feedback exactly one element of S_t as the winner of this multi-duel. As described in Sec. 1.1, when $S_t = S$, the winner feedback is distributed as a categorical random variable with parameter $(\mathbf{P}(i|S)_{i\in S})_{i\in S}$ and values in S. In this section, we focus on the special case $m = k$. Due to $[k]_k = \{[k]\}$ the set $[k]$ is the only query set and the feedback observed reduces to samples of a categorical random variable (i.e., a generalized die) with parameter $\mathbf{p} \coloneqq \mathbf{P}(\cdot|[k])$, which is an element in the simplex $\Delta_k \coloneqq \{(p_j)_{j\in[k]} \in [0,1]^k \mid \sum_{j\in[k]} p_j = 1\}$. We write $\mathrm{mode}(\mathbf{p}) \coloneqq \mathrm{argmax}_{j\in[k]} p_j$ for the modes of $\mathbf{p} = (p_j)_{j\in[k]} \in \Delta_k$ resp. its only element if $|\mathrm{mode}(\mathbf{p})| = 1$ and note that this notion coincides with $\mathrm{GCW}(\mathbf{P})$ if $\mathbf{P} \in PM_k^k$ and $\mathbf{p} = \mathbf{P}(\cdot|[k])$. Regarding that

$$PM_k^k(\Delta^h) = \left\{\mathbf{p} = (p_j)_{j\in[k]} \in \Delta_k \,|\, \forall j \neq \mathrm{mode}(\mathbf{p}) : p_{\mathrm{mode}(\mathbf{p})} \geq p_j + h\right\} = PM_k^k(h\mathrm{GCW})$$

holds for any $h \in (0,1]$, we do not have to distinguish between the sets $PM_k^m(\Delta^h)$ and $PM_k^m(h\mathrm{GCW})$ in case $m = k$, but may only consider $\Delta_k^h \coloneqq PM_k^k(\Delta^h)$. Similarly, the set $\Delta_k^0 \coloneqq \bigcup_{h\in(0,1]} \Delta_k^h = \{\mathbf{p} \in \Delta_k \,|\, \forall j \neq \mathrm{mode}(\mathbf{p}) : p_{\mathrm{mode}(\mathbf{p})} > p_j\}$ coincides with both $PM_k^m(\mathrm{GCW}^*)$ and $PM_k^m(\Delta^0)$ if $m = k$. Hence, we may conveniently reformulate $\mathcal{P}_{\mathrm{Die}}^{k,h,\gamma} \coloneqq \mathcal{P}_{\mathrm{GCWi}}^{k,k,\gamma}(\Delta^h)$ introduced in Ch. 1 with less terminology as follows:

For $h \in [0,1]$, $\gamma \in (0,1)$, a multi-dueling bandits algorithm $\mathcal{A}$ *solves* $\mathcal{P}_{\mathrm{Die}}^{k,h,\gamma}$ if it terminates a.s for any $\mathbf{p} \in \Delta_k^h$ and outputs for any $\mathbf{p} \in \Delta_k^h$ with error probability at most γ correctly the mode of $\mathbf{p}$, i.e., if

$$\forall \mathbf{p} \in \Delta_k^h : \mathbb{P}_{\mathbf{p}}\left(T^{\mathcal{A}} < \infty\right) = 1,$$
$$\forall \mathbf{p} \in \Delta_k^h : \mathbb{P}_{\mathbf{p}}\left(\mathbf{D}(\mathcal{A}) = \mathrm{mode}(\mathbf{p})\right) \geq 1 - \gamma.$$

Below, we provide several sample complexity lower and upper bounds for solutions to $\mathcal{P}_{\mathrm{Die}}^{k,h,\gamma}$. Some of them are stated in an instance-wise manner in terms of the quantity[4]

$$h(\mathbf{p}) \coloneqq \max\{h \in [0,1] \,|\, \mathbf{p} \in \Delta^h\}.$$

It can be seen as a hardness parameter for the problem at hand: The larger $h(\mathbf{p})$, the easier the mode of $\mathbf{p}$ could be identified based on iid samples from $\mathrm{Cat}(\mathbf{p})$.

2.3.1. Lower Bounds

We start with a sample complexity lower bound of solutions to $\mathcal{P}_{\mathrm{Die}}^{k,h,\gamma}$, that is based on the optimality of the SPRT as well as on Wald's identity.

Proposition 2.30. *Let $0 < \gamma < \gamma_0 < 1/2$ and $0 < h < h_0 < 1$ be fixed. Suppose $\mathcal{A}$ solves $\mathcal{P}_{\mathrm{Die}}^{k,h,\gamma}$, let $\mathbf{p} \in \Delta_k^h$ be arbitrary and write $i \coloneqq \mathrm{mode}(\mathbf{p})$. Then,*

$$\mathbb{E}_{\mathbf{p}}\left[T^{\mathcal{A}}\right] \geq \frac{f\left(\frac{p_i - p_j}{2(p_i+p_j)}, \gamma\right)}{p_i + p_j}$$

[4]Let us remark here that the term $h(\mathbf{p})$ coincides with that of $h(\mathbf{P})$ from Ch. 6 below if $\mathbf{P} \in PM_k^k$, and in case $k = 2$ and $\mathbf{p} = (p, 1-p)$ we have $h(\mathbf{p}) = 2\overline{p}$ with $\overline{p} = |1/2 - p|$ as e.g. used in Thm. 4.1 later on.

holds for all $j \in [k] \setminus \{i\}$ *with* $f(z,\gamma) := \frac{1-2\gamma}{2z} \left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+z)/(1/2-z))} \right\rceil$, *which fulfills* $\forall z \in (0, h_0/2) : f(z,\gamma) \geq \frac{c(h_0,\gamma_0)}{z^2} \ln \frac{1}{\gamma}$ *for some appropriate constant* $c(h_0,\gamma_0) > 0$ *that does not depend on* γ *or* h. *In particular, we obtain the worst-case bound*

$$\sup_{\mathbf{p}\in\Delta_k^h} \mathbb{E}_{\mathbf{p}}[T^{\mathcal{A}}] \geq \frac{4c(h_0,\gamma_0)}{h^2} \ln \frac{1}{\gamma} \tag{2.15}$$

and the instance-wise bound

$$\forall \mathbf{p} \in \Delta_k^h : \mathbb{E}_{\mathbf{p}}[T^{\mathcal{A}}] \geq \frac{2c(h_0,\gamma_0)}{(h(\mathbf{p}))^2} \left(\frac{1}{k} + h\right) \ln \frac{1}{\gamma}. \tag{2.16}$$

This result is similar to one that has recently been given by Shah et al. [2020, Thm. 3]. In comparison to ours, their bound is based on Lem. 2.42 and does not provide the asymptotic behavior w.r.t. k in a worst-case sense.

In the proof of Prop. 2.30 we will exploit sample complexity lower bounds of solutions to $\mathcal{P}_{\text{Die}}^{2,h,\gamma}$. For the sake of convenience, we write p for $(p, 1-p) \in \Delta_2$. Note that solving $\mathcal{P}_{\text{Die}}^{2,h,\gamma}$ resp. $\mathcal{P}_{\text{Die}}^{2,0,\gamma}$ reduces to deciding with error probability at most γ

$$\mathbf{H}_0 : p > 1/2 \quad \text{vs.} \quad \mathbf{H}_1 : p < 1/2 \tag{2.17}$$

based on iid samples $X_1, X_2, \dots \sim \text{Ber}(p)$ for any $p \in [0,1]$ with $|p - 1/2| \geq h$ resp. $|p - 1/2| > 0$. Regarding that $\mathcal{Q}_m^{h/2} \approx PM_2^m(\Delta^h)$, we have $\mathcal{P}_{\text{Die}}^{2,h,\gamma} \approx \mathcal{P}_{\text{Coin}}^{\gamma,h/2}$, for which we already stated an appropriate lower bound in Prop. 2.17.

Before we give the proof of Prop. 2.30, we state two further auxiliary lemmata. The first one is a simplified version of *Walds identity* [Bauer and Burckel, 1996, Thm. 17.7], which we shortly prove for the sake of convenience in the appendix, and the second one is only required for the instance-wise bound in Prop. 2.30. Here and throughout, we denote by $\mathcal{E} \perp\!\!\!\perp \mathcal{E}'$ independence of families $\mathcal{E}, \mathcal{E}'$ of events, and if X is a random variable, we simply write $X \perp\!\!\!\perp \mathcal{E}'$ for independence of X (i.e., of its generated sigma algebra $\sigma(X)$) and $\mathcal{E}'$.

Lemma 2.31. *Let* $k \in \mathbb{N}$ *and* $(p_1, \dots, p_k) \in \Delta_k$ *be fixed. Suppose* $\{X_t\}_{t\in\mathbb{N}}$ *to be an iid family of random variables* $X_t \sim \text{Cat}(p_1, \dots, p_k)$ *on some joint probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $\{\mathcal{F}_t\}_{t\in\mathbb{N}} \subseteq \mathcal{F}$ *to be a filtration, such that* $\{X_t\}_t$ *is* $\{\mathcal{F}_t\}_t$*-adapted and* $\forall t : X_t \perp\!\!\!\perp \mathcal{F}_{t-1}$, *e.g.* $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$. *If* τ *is an* $\{\mathcal{F}_t\}_t$*-stopping time, then the random variables*

$$T_i(\tau) := \sum\nolimits_{t\leq\tau} \mathbf{1}_{\{X_t = i\}}, \quad i \in [k],$$

fulfill $\mathbb{E}[T_i(\tau)] = p_i \mathbb{E}[\tau]$ *for each* $i \in [k]$. *In particular, we obtain*

$$\mathbb{E}[\tau] = \frac{\sum_{i\in I} \mathbb{E}[T_i(\tau)]}{\sum_{i\in I} p_i}$$

for any $I \subseteq [k]$ *with* $\sum_{i\in I} p_i > 0$.

Lemma 2.32. *Suppose* $\mathbf{p} \in \Delta_k^h \setminus \Delta_k^{\tilde{h}}$ *for some* $0 < h < \tilde{h} < 1$ *and let* $i := \text{mode}(\mathbf{p})$ *and* $j \in \text{argmax}_{l\in[k]\setminus\{i\}} p_l$. *Then, we have* $p_i + p_j \geq \frac{2+(k-2)h}{k}$ *and* $p_i - p_j < \tilde{h}$.

We refer for the proof of Lem. 2.32 to the appendix and proceed here with the proof of Prop. 2.30.

Proof of Prop. 2.30. We may suppose w.l.o.g. $i = 1$ and fix $j = 2$. Let us define $a \coloneqq \frac{p_1}{p_1+p_2}$ and suppose we have a coin $C \sim \mathrm{Ber}(p)$ with $p \in \{a, 1-a\}$. By simulating $\mathcal{A}$, we will construct an algorithm $\mathcal{A}'$ for testing

$$\mathbf{H}_0' : p = a \qquad \mathbf{H}_1' : p = 1 - a$$

in the following way: Whenever $\mathcal{A}$ makes a query at time t, we generate an independent sample $U_t \sim \mathcal{U}([0,1])$. Then, we return the feedback $X_t = i' \in \{3, \ldots, k\}$ iff $U_t \in (\sum_{j' \leq i'-1} p_{j'}, \sum_{j' \leq i'} p_{j'}]$ and in case $U_t \in [0, p_1 + p_2]$ we generate an independent sample $C_t \sim \mathrm{Ber}(p)$ from our coin C and return

$$X_t = \begin{cases} 1, & \text{if } C_t = 1, \\ 2, & \text{if } C_t = 0. \end{cases}$$

As soon as $\mathcal{A}$ terminates, we terminate and return $\mathbf{D}(\mathcal{A}') = 0$ if $\mathbf{D}(\mathcal{A}) = 1$ and $\mathbf{D}(\mathcal{A}') = 1$ otherwise. By our construction, we have $\mathbb{P}_p(X_t = i) = p_i$ for each $i \in \{3, \ldots, k\}$

$$\mathbb{P}_a(X_t = 1) = (p_1 + p_2)\mathbb{P}(C_t = 1) = p_1, \quad \mathbb{P}_a(X_t = 2) = (p_1 + p_2)\mathbb{P}(C_t = 0) = p_2$$

and similarly $\mathbb{P}_{1-a}(X_t = 1) = p_2$ and $\mathbb{P}_{1-a}(X_t = 2) = p_1$. Thus if $p = a$, $\mathcal{A}$ behaves as started on $\mathbf{p}$ and if $p = 1 - a$, $\mathcal{A}$ behaves as started on $\mathbf{p}' \coloneqq (p_2, p_1, p_3, \ldots, p_k) \in \Delta_k^h$. Since $\mathcal{A}$ solves $\mathcal{P}_{\mathrm{Die}}^{k,h,\gamma}$, we obtain

$$\mathbb{P}_a(\mathbf{D}(\mathcal{A}') = 0) = \mathbb{P}_{\mathbf{p}}\left(\mathbf{D}(\mathcal{A}) = 1\right) \geq 1 - \gamma$$

and (due to $2 = \mathrm{argmax}_{j' \in [k]} p_j'$)

$$\mathbb{P}_{1-a}(\mathbf{D}(\mathcal{A}') = 1) = \mathbb{P}_{\mathbf{p}'}(\mathbf{D}(\mathcal{A}) \neq 1) \geq \mathbb{P}_{\mathbf{p}'}(\mathbf{D}(\mathcal{A}) = 2) \geq 1 - \gamma,$$

i.e., $\mathcal{A}'$ is able to decide $\mathbf{H}_0'$ versus $\mathbf{H}_1'$ with error probability at most γ. From Prop. 2.17 we infer that it has to throw the coin C (in both cases $p \in \{a, 1-a\}$) in expectation at least $f(a - 1/2, \gamma)$ times for this. Regarding that C is thrown in our construction iff we return as feedback an element from $\{1, 2\}$, we get that

$$\mathbb{E}_{\mathbf{p}}[T_1(T^{\mathcal{A}}) + T_2(T^{\mathcal{A}})] \geq f(a - 1/2, \gamma) \quad \text{where } T_i(T^{\mathcal{A}}) \coloneqq \textstyle\sum_{t \leq T^{\mathcal{A}}} \mathbf{1}_{\{X_t = i\}}.$$

An application of Lem. 2.31 yields

$$\mathbb{E}_{\mathbf{p}}[T^{\mathcal{A}}] \geq \frac{f(a - 1/2, \gamma)}{p_1 + p_2} = \frac{f\left(\frac{p_1 - p_2}{2(p_1+p_2)}, \gamma\right)}{p_1 + p_2},$$

which completes the proof of the first statement.

The worst-case bound (2.15) then follows from the just proven bound via

$$\sup_{\mathbf{p} \in \Delta_k^h} \mathbb{E}_{\mathbf{p}}[T^{\mathcal{A}}] \geq \mathbb{E}_{\left(\frac{1+h}{2}, \frac{1-h}{2}, 0, \ldots, 0\right)}[T^{\mathcal{A}}] \geq f\left(\frac{h}{2}, \gamma\right) \geq \frac{4c(h_0, \gamma_0)}{h^2} \ln \frac{1}{\gamma}$$

for some $c(h_0, \gamma_0) > 0$, that is assured to exist by Prop. 2.17. To prove (2.16) suppose at first $\tilde{h} \in (h, 1)$ and $\mathbf{p} \in \Delta_k^h \setminus \Delta_k^{\tilde{h}}$ to be fixed and write $i \coloneqq \mathrm{mode}(\mathbf{p})$. Lem. 2.32 reveals that there exists some $j \in [k] \setminus \{i\}$ with $p_i + p_j \geq \frac{2+(k-2)h}{k}$ and $p_i - p_j < \tilde{h}$. Consequently, the above proven bound and the estimate $f(h,\gamma) \geq \frac{c(h_0,\gamma_0)}{h^2} \ln \frac{1}{\gamma}$ yield

$$\begin{aligned}\mathbb{E}_{\mathbf{p}}[T^{\mathcal{A}}] \geq \frac{f\left(\frac{p_i-p_j}{2(p_i+p_j)}, \gamma\right)}{p_i+p_j} &\geq 4c(h_0,\gamma_0)\frac{p_i+p_j}{(p_i-p_j)^2}\ln\frac{1}{\gamma} \\ &\geq \frac{4c(h_0,\gamma_0)}{\tilde{h}^2}\cdot\frac{2+(k-2)h}{k}\ln\frac{1}{\gamma} \\ &\geq \frac{2c(h_0,\gamma_0)}{\tilde{h}^2}\left(\frac{1}{k}+h\right)\ln\frac{1}{\gamma}.\end{aligned}$$

Since $\mathbf{p} \in \Delta_k^{h(\mathbf{p})} \setminus \left(\bigcup_{\tilde{h}>h(\mathbf{p})} \Delta_k^{\tilde{h}}\right) = \bigcap_{\tilde{h}>h(\mathbf{p})} (\Delta_k^{h(\mathbf{p})} \setminus \Delta_k^{\tilde{h}})$ for any $\mathbf{p} \in \Delta_k^h$, (2.16) can be inferred from this by taking the limit $\tilde{h} \searrow h(\mathbf{p})$. □

Based on Prop. 2.13 instead of Prop. 2.17, we also obtain the following bound. In comparison to that from Prop. 2.30, it does not involve a dependence on k and its dependence of h is of the order $\frac{1}{h^2}\ln\ln\frac{1}{h}$ instead of $\frac{1}{h^2}$.

Proposition 2.33. *Let $\gamma \in (0,1/2)$ be fixed and suppose $\mathcal{A}$ solves $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$. Let $\mathbf{p} \in \Delta_k^0$ be arbitrary, $i := \text{mode}(\mathbf{p})$ and $j := \text{argmax}_{j\in[k]\setminus\{i\}} p_j$. Then, the family $\{\mathbf{p}(h)\}_{h\in(0,p_i-p_j)} \subseteq \Delta_k^0$ defined via $(\mathbf{p}(h))_i := \frac{(p_i+p_j+h)}{2}$, $(\mathbf{p}(h))_j := \frac{(p_i+p_j-h)}{2}$ and $(\mathbf{p}(h))_l := p_l$ for $l \in [k] \setminus \{i,j\}$ fulfills $\mathbf{p}(h) \in \Delta_k^h$ as well as*

$$\limsup_{h\to 0} \frac{\mathbb{E}_{\mathbf{p}(h)}[T^{\mathcal{A}}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \geq (1-2\gamma)(p_i+p_j) > 0.$$

Proof of Prop. 2.33. We suppose w.l.o.g. $(i,j) = (1,2)$ throughout the proof. For $h \in (0, p_1 - p_2)$ we have $(\mathbf{p}(h))_1 > (\mathbf{p}(h))_2 > (\mathbf{p}(h))_l$ for every $l \in \{3,\ldots,k\}$ and together with $|(\mathbf{p}(h))_1 - (\mathbf{p}(h))_2| = h$ this shows $\mathbf{p}(h) \in \Delta_k^h$. Suppose we have a coin $C \sim \text{Ber}(p)$ for $p \neq 1/2$. By simulating $\mathcal{A}$ as in the proof of Prop. 2.30 we obtain an algorithm $\mathcal{A}'$ for testing $\mathbf{H}_0 : p > 1/2$ versus $\mathbf{H}_1 : p < 1/2$, which has (due to the theoretical guarantees of $\mathcal{A}$) an error probability $\leq \gamma$ for every $p \neq 1/2$. Consequently, Prop. 2.13 guarantees the existence of a sequence $\{h'_l\}_{l\in\mathbb{N}} \subseteq (0, \frac{1}{e^4})$ with

$$\forall l \in \mathbb{N} : \frac{\mathbb{E}_{1/2\pm h'_l}[T^{\mathcal{A}'}]}{\frac{1}{h_l'^2}\ln\ln\frac{1}{h'_l}} \geq \frac{1-2\gamma}{2} - \varepsilon > 0$$

for some arbitrarily small but fixed $\varepsilon \in (0, \frac{1-2\gamma}{2})$. If we choose $h_l := 2(p_1+p_2)h'_l$, then the corresponding bias of the coin C in the reduction (cf. the proof of Prop. 2.30) is exactly

$$\frac{(\mathbf{p}(h_l))_1}{(\mathbf{p}(h_l))_1 + (\mathbf{p}(h_l))_2} = \frac{\frac{p_1+p_2}{2} + \frac{h_l}{2}}{p_1+p_2} = \frac{1}{2} + \frac{h_l}{2(p_1+p_2)} = \frac{1}{2} + h'_l$$

Hence, if $\mathcal{A}'$ is started on $1/2+h'_l$, its internal method $\mathcal{A}$ works as if started on $\mathbf{p}(h_l)$. From $h_l \leq e^{-4}$ we obtain $4 = (\frac{1}{2})^{-2} \leq \ln\frac{1}{h_l}$ and thus $-2\ln\frac{1}{2} \leq \ln\ln\frac{1}{h_l}$, i.e., $\ln\frac{1}{2} \geq -\frac{1}{2}\ln\ln\frac{1}{h_l} \geq$

$-\frac{1}{2}\ln\frac{1}{h_l}$. Consequently,

$$\ln\ln\frac{1}{h_l'} = \ln\ln\left(\frac{1}{2h_l(p_1+p_2)}\right) \geq \ln\left(\ln\frac{1}{2}+\ln\frac{1}{h_l}\right) \geq \ln\left(\frac{1}{2}\ln\frac{1}{h_l}\right)$$
$$= \ln\frac{1}{2}+\ln\ln\frac{1}{h_l} \geq \frac{1}{2}\ln\ln\frac{1}{h_l}$$

holds, and we obtain similarly as in the proof of Prop. 2.30

$$\mathbb{E}_{\mathbf{p}(h_l)}[T_1(T^{\mathcal{A}})+T_2(T^{\mathcal{A}})] \geq \mathbb{E}_{\frac{1}{2}+h_l'}[T^{\mathcal{A}'}] \geq \left(\frac{1}{2}(1-2\gamma)-\varepsilon\right)\cdot\frac{1}{{h_l'}^2}\ln\ln\frac{1}{h_l'}$$
$$\geq 2(p_1+p_2)^2\left(\frac{1}{2}(1-2\gamma)-\varepsilon\right)\cdot\frac{1}{h_l^2}\ln\ln\frac{1}{h_l}.$$

Regarding that this holds for arbitrarily small $\varepsilon > 0$, Lem. 2.31 shows[5] that

$$\frac{\mathbb{E}_{\mathbf{p}(h_l)}[T^{\mathcal{A}}]}{\frac{1}{h_l^2}\ln\ln\frac{1}{h_l}} \geq (1-2\gamma)(p_1+p_2)$$

holds for every $l \in \mathbb{N}$, which completes the proof. □

2.3.2. Upper Bounds

Next, we construct a solution to $\mathcal{P}_{\text{Die}}^{k,h,\gamma}$. An algorithm $\mathcal{A}$, which tackles $\mathcal{P}_{\text{Die}}^{k,h,\gamma}$, has to decide in a sequential manner at each time t, whether it wants to make a further query $S_t \in [k]_k$ resulting in a sample X_t or to output an answer $D(\mathcal{A}) \in [k]$. As $[k]_k = \{[k]\}$, we can only choose $S_t = [k]$ in each time step t, upon which we observe as feedback $X_t \sim \text{Cat}(\mathbf{p})$, i.e., $\mathbb{P}_{\mathbf{p}}(X_t = i) = p_i$ for any $i \in [k]$. Having observed $X_1, \dots, X_t$, which are iid by assumption, a straightforward idea for a prediction $\mathbf{D}(\mathcal{A})$ would be to use the mode of the empirical distribution $\hat{\mathbf{p}}^t \coloneqq (\hat{p}_1^t, \dots, \hat{p}_k^t)$ given by $\hat{p}_i^t \coloneqq \frac{1}{t}\sum_{t' \leq t} \mathbf{1}_{\{X_{t'}=i\}}$.

In fact, the prominent *Dvoretzky-Kiefer-Wolfowitz inequality* [Dvoretzky et al., 1956, Massart, 1990], which we state for simplicity only for categorical random variables in the following, assures that $\hat{\mathbf{p}}^t$ is w.r.t. the infinity norm $||\cdot||_\infty$ close to $\mathbf{p}$ with high confidence for large values of t.

Lemma 2.34 (Dvoretzky-Kiefer-Wolfowitz inequality for categorical random variables)**.** *Suppose $X_1, X_2, \dots$ to be iid random variables $X_n \sim \text{Cat}(\mathbf{p})$ for some $\mathbf{p} \in \Delta_k$. For $t \in \mathbb{N}$ let $\hat{\mathbf{p}}^t$ be the corresponding empirical distribution after the t observations $X_1, \dots, X_t$, i.e., $\hat{p}_i^t = \frac{1}{t}\sum_{s=1}^t \mathbf{1}_{\{X_s=i\}}$ for all $i \in [k]$. Then, we have for any $\varepsilon > 0$ and $t \in \mathbb{N}$ the estimate*

$$\mathbb{P}\left(\left|\left|\hat{\mathbf{p}}^t - \mathbf{p}\right|\right|_\infty > \varepsilon\right) \leq 4e^{-t\varepsilon^2/2}.$$

Lem. 2.34 shows that, for large values of t, predicting the mode of $\hat{\mathbf{p}}^t$ would be the correct prediction for mode($\mathbf{p}$) with high probability. To solve $\mathcal{P}_{\text{Die}}^{k,h,\gamma}$, we have to choose t so large that mode($\hat{\mathbf{p}}^t$) = mode($\mathbf{p}$) with confidence at least $1-\gamma$. For this, the following lemma is of use.

[5] Note here that $(\mathbf{p}(h))_1 + (\mathbf{p}(h))_2 = p_1 + p_2$.

Lemma 2.35. *For $h \in [0,1], \varepsilon \in (-h,1], \mathbf{p} \in \Delta_k^h$ and $\tilde{\mathbf{p}} \in \Delta_k$ we have*

$$(\exists i : \tilde{p}_i - \max_{j \neq i} \tilde{p}_j \geq \varepsilon \text{ and } p_i \neq \max_j p_j) \quad \Rightarrow \quad ||\tilde{\mathbf{p}} - \mathbf{p}||_\infty \geq \frac{h+\varepsilon}{2}.$$

The bounds from Lem. 2.35 are apparently sharp, as $\mathbf{p} \in \Delta_k^h$ and $\tilde{\mathbf{p}} \in \Delta_k$ defined via

$$p_i = \begin{cases} \frac{1}{2} - \frac{h}{2}, & \text{if } i = 1, \\ \frac{1}{2} + \frac{h}{2}, & \text{if } i = 2, \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \tilde{p}_i = \begin{cases} \frac{1}{2} + \frac{\varepsilon}{2}, & \text{if } i = 1, \\ \frac{1}{2} - \frac{\varepsilon}{2}, & \text{if } i = 2, \\ 0, & \text{otherwise.} \end{cases}$$

fulfill $\tilde{p}_1 - \max_{j \neq 1} \tilde{p}_j = \varepsilon$ and $p_1 \neq \max_{j \in [k]} p_j$ and at the same time $||\mathbf{p} - \tilde{\mathbf{p}}||_\infty = \frac{h+\varepsilon}{2}$.

With the choices $\varepsilon = 0$ and $\tilde{\mathbf{p}} = \hat{\mathbf{p}}^t$, Lem. 2.35 shows us that $\left|\left|\hat{\mathbf{p}}^t - \mathbf{p}\right|\right|_\infty > \frac{h}{2}$ is necessary for $\mathrm{mode}(\hat{\mathbf{p}}^t) \neq \mathrm{mode}(\mathbf{p})$. Combining this with Lem. 2.34, we see that simply querying $S_t = [k]$ for $T = \left\lceil \frac{8}{h^2} \ln \frac{4}{\gamma} \right\rceil$ many times and returning the mode of $\hat{\mathbf{p}}^T$ as the decision results in a solution to $\mathcal{P}_{\text{Die}}^{k,h,\gamma}$. For the sake of convenience, we provide the corresponding pseudocode as Alg. 5 and state its guarantees as the next proposition.

Algorithm 5 DKW mode identification – (non-sequential) solution to $\mathcal{P}_{\text{Die}}^{k,h,\gamma}$

Input: $\gamma \in (0,1)$, $h \in (0,1)$, $k \in \mathbb{N}$, access to iid samples $X_t \sim \mathrm{Cat}(\mathbf{p})$

1: Let $T \leftarrow \left\lceil \frac{8}{h^2} \ln \frac{4}{\gamma} \right\rceil$
2: Observe $X_1, \ldots, X_T \sim \mathrm{Cat}(\mathbf{p})$
3: **return** $\mathrm{mode}(\hat{\mathbf{p}}^T) = \mathrm{argmax}_{i \in [k]} \sum_{t=1}^T \mathbf{1}_{\{X_t = i\}}$

Proposition 2.36. *For any $k \in \mathbb{N}$, $h \in (0,1)$ and $\gamma \in (0,1)$, Alg. 5 called with parameters γ, h, k solves $\mathcal{P}_{\text{Die}}^{k,h,\gamma}$ and terminates after exactly $\left\lceil \frac{8}{h^2} \ln \frac{4}{\gamma} \right\rceil$ time steps.*

Proof. This is a direct consequence of Lem. 2.35 and Lem. 2.34. □

According to Prop. 2.30, Alg. 5 is in a worst-case sense asymptotically optimal as a solution to $\mathcal{P}_{\text{Die}}^{k,h,\gamma}$.

Now, we intend to solve the more challenging problem $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$. Note that any solution to $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$ is also a solution to $\mathcal{P}_{\text{Die}}^{k,h,\gamma}$ for any $h > 0$, hence Prop. 2.30 shows that $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$ cannot be solved by any non-sequential algorithm, i.e., one which decides a priori the number of samples it observes. To construct a solution, we make use of Alg. 6, which also tackles the problem of finding the mode of $\mathbf{p}$ in a non-sequential manner but is allowed to return UNSURE as an indicator that it is not confident enough for its prediction. In other words, the algorithm is allowed to abstain from making a decision. We obtain the following guarantees of Alg. 6.

Lemma 2.37. *$\mathcal{A} :=$ Alg. 6 init. with parameters $\gamma, h \in (0,1)$ fulfills $T^{\mathcal{A}} = \left\lceil \frac{8}{h^2} \ln \frac{4}{\gamma} \right\rceil$,*

$$\forall \mathbf{p} \in \Delta_k \quad : \mathbb{P}_{\mathbf{p}}(\mathbf{D}(\mathcal{A}) \in [k] \text{ and } p_{\mathbf{D}(\mathcal{A})} < \max_{j \in [k]} p_j) \leq \gamma, \tag{2.18}$$

$$\forall \mathbf{p} \in \Delta_k^0 \quad : \mathbb{P}_{\mathbf{p}}(\mathbf{D}(\mathcal{A}) \in \{\mathrm{mode}(\mathbf{p}), \text{UNSURE}\}) \geq 1 - \gamma, \tag{2.19}$$

$$\forall \mathbf{p} \in \Delta_k^{3h} : \mathbb{P}_{\mathbf{p}}(\mathbf{D}(\mathcal{A}) = \mathrm{mode}(\mathbf{p})) \geq 1 - \gamma. \tag{2.20}$$

Lem. 2.37 reveals that Alg. 6 has a low failure rate (2.18) by appropriate choice of γ, while in turn by an appropriate choice of h, namely $h \leq \frac{1}{3}h(\mathbf{p})$, the correct decision will be returned (2.20) with high probability.

Algorithm 6 DKW mode-identification with abstention

Input: $\gamma \in (0,1)$, $h \in (0,1)$, access to iid samples $X_t \sim \mathrm{Cat}(\mathbf{p})$

1: $T \leftarrow \left\lceil \frac{8}{h^2} \ln \frac{4}{\gamma} \right\rceil$
2: Observe samples $X_1, \ldots, X_T$
3: Calculate $\hat{\mathbf{p}}^T = (\hat{p}_1^T, \ldots, \hat{p}_k^T)$ as $\hat{p}_i^T := \frac{1}{T}\sum_{t=1}^T \mathbf{1}_{\{X_t = i\}}$, $i \in [k]$
4: Choose $i^* \in \mathrm{mode}(\hat{\mathbf{p}}^T)$
5: **if** $\hat{p}_{i^*}^T > \max_{j \neq i^*} \hat{p}_j^T + h$ **then return** i^*
6: **else return** UNSURE

We prepare the proof of Lem. 2.37 with the following result. More precisely, it will assure us that Alg. 6 returns with probability $\geq 1-\gamma$ the correct mode in case $\mathbf{p} \in \Delta_k^{3h}$.

Lemma 2.38. *Let $h > 0$, $\mathbf{p} \in \Delta_k^{3h}$ and $\tilde{\mathbf{p}} \in \Delta_k$ be fixed. Then,*

$$\forall i : \tilde{p}_i \leq \max\nolimits_{j \neq i} \tilde{p}_j + h \quad \Rightarrow \quad ||\mathbf{p} - \tilde{\mathbf{p}}||_\infty \geq h.$$

Before giving a detailed proof of Lem. 2.37, let us note the constraint $\mathbf{p} \in \Delta_k^{3h}$ in the statement above is sharp in the sense that we have for any $h \in (0, 1/8)$

$$\inf\left\{ s > 0 \,\middle|\, \forall \mathbf{p} \in \Delta_k^{sh}\, \forall \tilde{\mathbf{p}} \in \Delta_k : (\forall i : \tilde{p}_i \leq \max\nolimits_{j \neq i} \tilde{p}_j + h \Rightarrow ||\mathbf{p} - \tilde{\mathbf{p}}||_\infty \geq h) \right\} = 3.$$

This is justified in the next lemma.

Lemma 2.39. *For any $h \in (0, \frac{1}{8})$, $\varepsilon \in (0, \frac{1}{3})$ and $k \in \mathbb{N}_{\geq 3}$ there exist $\mathbf{p} \in \Delta_k^{(3-\varepsilon)h}$ and $\tilde{\mathbf{p}} \in \Delta_k$ such that*

$$\forall i \in [k] : \tilde{p}_i \leq \max\nolimits_{j \neq i} \tilde{p}_j + h \quad \text{and} \quad ||\mathbf{p} - \tilde{\mathbf{p}}||_\infty < h.$$

We proceed with the proof of Lem. 2.37.

Proof of Lem. 2.37. Let $\mathbf{p} \in \Delta_k$ be fixed, and note that Alg. 6 terminates after exactly $\left\lceil \frac{8}{h^2} \ln \frac{4}{\gamma} \right\rceil$ time steps. Lem. 2.35 and Lem. 2.34 let us directly infer

$$\begin{aligned}
&\mathbb{P}_{\mathbf{p}}\left(\mathbf{D}(\mathcal{A}) \in [k] \text{ and } p_{\mathbf{D}(\mathcal{A})} < \max\nolimits_{j \in [k]} p_j\right) \\
&\quad = \mathbb{P}\left(\exists i \in [k] : \hat{p}_i^t - \max\nolimits_{j \neq i} \hat{p}_j^t > h \text{ and } p_i \neq \max\nolimits_{j \in [k]} p_j\right) \\
&\quad \leq \mathbb{P}\left(||\hat{\mathbf{p}}^t - \mathbf{p}||_\infty > \frac{h}{2}\right) \leq \gamma. \qquad (2.21)
\end{aligned}$$

Next, suppose $\mathbf{p} \in \Delta_k^0$ and let $i' := \mathrm{mode}(\mathbf{p}) \in [k]$. Again, Lem. 2.35 yields

$$\begin{aligned}
\left\{\mathbf{D}(\mathcal{A}) \in [k] \setminus \{i'\}\right\} &= \left\{\exists i \neq i' : \hat{p}_i^t - \max\nolimits_{j \neq i} \hat{p}_j^t > h \text{ and } p_{i'} > \max\nolimits_{j \neq i'} p_j\right\} \\
&\subseteq \left\{||\hat{\mathbf{p}}^t - \mathbf{p}||_\infty > \frac{h}{2}\right\}, \qquad (2.22)
\end{aligned}$$

and thus

$$\mathbb{P}_{\mathbf{p}}(\mathbf{D}(\mathcal{A}) \notin \{i', \text{UNSURE}\}) \leq \mathbb{P}_{\mathbf{p}}\left(\left|\left|\hat{\mathbf{p}}^t - \mathbf{p}\right|\right|_\infty > \frac{h}{2}\right) \leq \gamma$$

follows from Lem. 2.34 and the choice of t. Now, let us suppose $\mathbf{p} \in \Delta_k^{3h}$. A look at Lem. 2.38 reveals

$$\{\mathbf{D}(\mathcal{A}) = \text{UNSURE}\} = \left\{\forall i \in [k] : \hat{p}_i^t \leq \max_{j \neq i} \hat{p}_j^t + h\right\} \subseteq \left\{\left|\left|\hat{\mathbf{p}}^t - \mathbf{p}\right|\right|_\infty > h\right\},$$

and combining this with (2.22) yields

$$\begin{aligned}\mathbb{P}_{\mathbf{p}}(\mathbf{D}(\mathcal{A}) \neq \text{mode}(\mathbf{p})) &= \mathbb{P}_{\mathbf{p}}\left(\mathbf{D}(\mathcal{A}) \in [k] \setminus \{i'\} \text{ or } \mathbf{D}(\mathcal{A}) = \text{UNSURE}\right) \\ &\leq \mathbb{P}_{\mathbf{p}}\left(\left|\left|\hat{\mathbf{p}}^t - \mathbf{p}\right|\right|_\infty > \frac{h}{2}\right) \leq \gamma,\end{aligned}$$

where the last estimate is again due to Lem. 2.34. □

There are two drawbacks of Alg. 6: First of all, it can also abstain from making a decision (see (2.19)), and more importantly, the value of $h(\mathbf{p})$ is unknown when solving $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$. As a remedy, we could run Alg. 6 successively with appropriately decreasing choices for γ and h until a (real) decision is returned. This approach is followed by Alg. 7 and the following proposition shows that it is indeed a solution to $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$; its proof is an adaptation of Lem. 11 in [Ren et al., 2019a].

Algorithm 7 DKW mode-identification – Solution to $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$

Input: $\gamma \in (0,1)$, sample access to $\text{Cat}(\mathbf{p})$
Initialization: $\widetilde{\mathcal{A}} := \text{Alg. 6}$, $s \leftarrow 1$, $\forall r \in \mathbb{N} : \gamma_r := \frac{6\gamma}{\pi^2 r^2}, h_r := 2^{-r-1}$

```
feedback ← UNSURE
while feedback is UNSURE do
    feedback ← Ã(γ_s, h_s, sample access to Cat(p))
    s ← s + 1
return feedback
```

Proposition 2.40. *$\mathcal{A} :=$ Alg. 7 initialized with the parameter $\gamma \in (0,1)$ solves $\mathcal{P}_{\text{Die}}^{k,0,\gamma}$ s.t.*

$$\mathbb{P}_{\mathbf{p}}\left(\mathbf{D}(\mathcal{A}) = \text{mode}(\mathbf{p}) \text{ and } T^{\mathcal{A}} \leq t_0(\gamma, h(\mathbf{p}))\right) \geq 1 - \gamma$$

for any $\mathbf{p} \in \Delta_k^0$, where $t_0(\gamma,h)$ is mononotically decreasing w.r.t. h with $t_0(\gamma,h) \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h} + \ln\frac{1}{\gamma}\right)\right)$.

Proof of Prop. 2.40. Let $\mathbf{p} \in \Delta_k^0$ be fixed and abbreviate $h := h(\mathbf{p})$. Moreover, denote by $\mathbf{D}(\mathcal{A}_s)$ the output of the instance of Alg. 6 with parameters γ_s, h_s that is called in iteration s of the while loop of $\mathcal{A}$ (Alg. 7). Let us define for each $s \in \mathbb{N}$ the set

$$\begin{aligned}\mathcal{E}_1^s &:= \left\{h_s > \frac{h}{3} \text{ and } \mathbf{D}(\mathcal{A}_s) \in \{\text{UNSURE}, \text{mode}(\mathbf{p})\}\right\}, \\ \mathcal{E}_2^s &:= \left\{h_s \leq \frac{h}{3} \text{ and } \mathbf{D}(\mathcal{A}_s) = \text{mode}(\mathbf{p})\right\}\end{aligned}$$

and

$$\mathcal{E} := \bigcup_{s\in\mathbb{N}} (\mathcal{E}_1^s \cup \mathcal{E}_2^s)^c .$$

From the equivalence $h' \leq \frac{1}{3}h(\mathbf{p}) \Leftrightarrow \mathbf{p} \in \Delta_k^{3h'}$ and Lem. 2.37 we infer

$$\mathbb{P}_{\mathbf{p}}\left((\mathcal{E}_1^s \cup \mathcal{E}_2^s)^c\right) = \begin{Bmatrix} \mathbb{P}_{\mathbf{p}}\left((\mathcal{E}_1^s)^c\right), & \text{if } h_s > h/3 \\ \mathbb{P}_{\mathbf{p}}\left((\mathcal{E}_2^s)^c\right), & \text{if } h_s \leq h/3 \end{Bmatrix} \leq \gamma_s$$

and therefore

$$\mathbb{P}_{\mathbf{p}}(\mathcal{E}) \leq \sum_{s\in\mathbb{N}} \gamma_s = \sum_{s\in\mathbb{N}} \frac{6\gamma}{\pi^2 s^2} = \gamma. \tag{2.23}$$

Now, let $s_0 := s_0(h) \in \mathbb{N}$ be such that $h_{s_0} \leq \frac{h}{3} < h_{s_0-1}$ and note that

$$\begin{aligned} &\mathcal{E}^c \subseteq \mathcal{E}_2^{s_0} \subseteq \{\mathbf{D}(\mathcal{A}_{s_0}) \neq \text{UNSURE}\} \\ &\quad \subseteq \{\mathcal{A} \text{ terminates at latest after the } s_0\text{-th iteration of the while loop}\}. \end{aligned} \tag{2.24}$$

In particular, $\mathcal{A}$ terminates almost surely on $\mathcal{E}^c$. Regarding the construction[6] of $\mathcal{A}$ we also have

$$\begin{aligned} \mathcal{E}^c = \bigcap_{s\in\mathbb{N}} (\mathcal{E}_1^s \cup \mathcal{E}_2^s) &\subseteq \bigcap_{s\in\mathbb{N}} \{\mathbf{D}(\mathcal{A}_s) \in \{\text{UNSURE}, \text{mode}(\mathbf{p})\}\} \\ &\subseteq \{\mathbf{D}(\mathcal{A}) = \text{mode}(\mathbf{p})\}. \end{aligned} \tag{2.25}$$

Since $\mathcal{A}$ makes in its s-th iteration of the while loop (according to Alg. 6) exactly $\left\lceil \frac{8}{h_s^2} \ln \frac{4}{\gamma_s} \right\rceil$ queries, combining (2.23), (2.24) and (2.25) yields

$$\mathbb{P}_{\mathbf{p}}\left(\mathbf{D}(\mathcal{A}) = \text{mode}(\mathbf{p}) \text{ and } T^{\mathcal{A}} \leq t_0(h,\gamma)\right) \geq \mathbb{P}_{\mathbf{p}}\left(\mathcal{E}^c\right) \geq 1-\gamma,$$

with $t_0(h,\gamma) := \sum_{s\leq s_0(h)} \left\lceil \frac{8}{h_s^2} \ln \frac{4}{\gamma_s} \right\rceil$. As the choice of $s_0 = s_0(h)$ guarantees $\frac{h}{3} < h_{s_0-1} = 2^{-s_0}$ and thus $s_0 < \log_2 \frac{3}{h}$, we obtain with regard to the choices of $h_s = 2^{-s-1}$ and $\gamma_s = \frac{6\gamma}{\pi^2 s^2}$ that

$$\begin{aligned} t_0(h,\gamma) \leq 2^7 \sum_{s=1}^{s_0(h)} 2^{2s-1} \ln\left(\frac{2\pi^2 s^2}{3\gamma}\right) &\in \mathcal{O}\left(\sum_{s=1}^{s_0(h)} 2^{2s-1} \ln\left(\frac{s_0(h)}{\gamma}\right)\right) \\ &\subseteq \mathcal{O}\left(4^{s_0(h)} \ln\left(\frac{s_0(h)}{\gamma}\right)\right) \\ &\subseteq \mathcal{O}\left(4^{\log_2 \frac{3}{h}} \ln\left(\frac{\log_2 \frac{3}{h}}{\gamma}\right)\right) \\ &\subseteq \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h} + \ln\frac{1}{\gamma}\right)\right) \end{aligned}$$

as $\min\{h,\gamma\} \to 0$. It remains to show that $T^{\mathcal{A}}$ is almost surely finite w.r.t. $\mathbb{P}_{\mathbf{p}}$. For an arbitrary integer $s \geq \log_2 \frac{3}{h}$ we have $h_s \leq \frac{h}{3}$ and thus

$$\begin{aligned} \mathbb{P}_{\mathbf{p}}\left(T^{\mathcal{A}} = \infty\right) &\leq \mathbb{P}_{\mathbf{p}}\left(\forall s' \in \mathbb{N} \text{ with } h_{s'} \leq \frac{h}{3} : \mathbf{D}(\mathcal{A}_{s'}) = \text{UNSURE}\right) \\ &\leq \mathbb{P}_{\mathbf{p}}\left(\mathbf{D}(\mathcal{A}_s) = \text{UNSURE}\right) \leq \mathbb{P}_{\mathbf{p}}((\mathcal{E}_2^s)^c) \leq \gamma_s, \end{aligned}$$

which directly implies $\mathbb{P}_{\mathbf{p}}(T^{\mathcal{A}} = \infty) \leq \lim_{s\to\infty} \gamma_s = 0$. □

[6]Note here that $\mathbf{D}(\mathcal{A}) \in [k]$ holds, i.e., $\mathcal{A}$ cannot terminate with UNSURE as output.

Last but not least, we discuss yet another solution, which has recently been developed and performs very well in practice. [Jain et al., 2021] adapted the *one-vs-one* and *one-vs-rest* paradigms, which are known in the machine learning community for multi-class classification, for generalizing solutions to $\mathcal{P}^{\gamma}_{\text{Coin}} = \mathcal{P}^{2,0,\gamma}_{\text{Die}}$ to $\mathcal{P}^{k,0,\gamma}_{\text{Die}}$ for general $k \geq 2$. As solutions to $\mathcal{P}^{\gamma}_{\text{Coin}}$ the authors consider in addition to their method PPR-Bernoulli (Alg. 4) also $\mathcal{A}_1$ from [Shah et al., 2020] and KL-SN from [Garivier, 2013]. They observe that the one-vs-one variant of PPR-Bernoulli clearly outperforms any of the others, hence we restrict ourselves at this point to giving the details of this variant.

For this, we write $(W_t)_i = \sum_{t'=1}^{t} \mathbf{1}_{\{X_{t'}=i\}}$ for the number of times i has been observed as feedback until time t, and further $(W_t)_{(1)}, \ldots, (W_t)_{(k)}$ for the order statistic of $((W_t)_i)_{i\in[k]}$, i.e., $\{(W_t)_i \,|\, i \in [k]\} = \{(W_t)_{(i)} \,|\, i \in [k]\}$ and $(W_t)_{(1)} \geq \cdots \geq (W_t)_{(k)}$. The resulting algorithm is called PPR1v1 and stated as Alg. 8.

Algorithm 8 PPR1v1 – Solution to $\mathcal{P}^{k,0,\gamma}_{\text{Die}}$

Input: $\gamma > 0$, sample access to $\text{Cat}(\mathbf{p})$
Initialization: $t \leftarrow 1$

```
1: while True do
2:     Observe X_t and update ((W_t)_i)_{i∈[k]}
3:     if f_Beta(1/2; (W_t)_(1) + 1, (W_t)_(2) + 1) ≤ γ/(k − 1) then
4:         return argmax_{i∈[k]} (W_t)_i
5:     t ← t + 1
```

To state the theoretical guarantee for PPR1v1, we write $(p_{(1)}, \ldots, p_{(k)})$ for the order statistic of $\mathbf{p} = (p_i)_{i\in[k]}$, i.e, $\{p_i \,|\, i \in [k]\} = \{p_{(i)} \,|\, i \in [k]\}$ and $p_{\text{mode}(\mathbf{p})} = p_{(1)} \geq \cdots \geq p_{(k)}$.

Proposition 2.41. *For any* $\gamma \in (0, 1)$, $\mathcal{A} := \text{PPR1v1}$ *solves* $\mathcal{P}^{k,0,\gamma}_{\text{Die}}$ *s.t. for any* $\mathbf{p} \in \Delta_k$

$$\mathbb{P}_{\mathbf{p}}\left(T^{\mathcal{A}} \leq T_{\gamma}(\mathbf{p})\right) \geq 1 - \gamma$$

for $T_{\gamma}(\mathbf{p}) = \frac{194.07 p_{(1)}}{(p_{(1)}-p_{(2)})^2} \ln\left(\sqrt{\frac{79.68(k-1)}{\gamma}} \frac{p_{(1)}}{p_{(1)}-p_{(2)}}\right)$ *with* $\sup_{\mathbf{p}\in\Delta^h_k} T_{\gamma}(\mathbf{p}) \in \Theta\left(\frac{1}{h^2} \ln\left(\frac{k}{h^2\gamma}\right)\right)$.

Proof. The first part is Thm. 9 in [Jain et al., 2021], where $\inf_{\mathbf{p}\in\Delta^0_k} \mathbb{P}_{\mathbf{P}}(T^{\mathcal{A}} < \infty) = 1$ can be seen similarly as the corresponding result in the proof of Prop. 2.28. The worst-case result on $T_{\gamma}(\mathbf{p})$ follows from the fact that $h(\mathbf{p}) = p_{(1)} - p_{(2)}$ holds for any $\mathbf{p} \in \Delta_k$. □

In comparison to our bound from Prop. 2.40, the worst-case upper bound w.r.t Δ^h_k of PPR1v1 is thus suboptimal by a factor $\frac{\ln \frac{k}{h^2\gamma}}{\ln\ln\frac{1}{h} + \ln\frac{1}{\gamma}} \in \Theta(\ln k) \cap \Theta(\frac{\ln h}{\ln\ln h})$.

2.4. Empirical Evaluation

Before continuing, we empirically compare the mode identification algorithms which we previously presented. For simplicity, we restrict ourselves to the case $\gamma = 0.05$ and report only the accuracy together with the sample mean and standard error of the termination time of the algorithms when executed on particular instances $p \in [0, 1]$ resp. $\mathbf{p} \in \Delta_k$.

2.4.1. Mode Identification of a Coin

We start with the empirical comparison of the solutions to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$. For this, we write $\mathcal{A}_{\mathrm{Hoeffding}}$ for the non-sequential solution from Lem. 2.10, $\mathcal{A}_{\mathrm{SPRT}}$ for that from Prop. 2.17 and $\mathcal{A}_{\mathrm{LIL}}$ for the one from Prop. 2.23 with the particular choices $\varepsilon = \varepsilon_0 = \frac{1}{2}, \delta = (\frac{\gamma}{5})^{2/3} \ln \frac{3}{2}$. Table 2.3 shows the performance of these procedures when started with parameters $h = 0.1$ and $\gamma = 0.05$ on different instances $p \in [0, 1]$, the values are averaged over 1000 repetitions each.

In case $p \in \{0.51, 0.55\}$ the low-noise assumption $|p - 1/2| > h = 0.1$ is clearly violated, hence it is not surprising that the algorithms partly have an error > 0.05. However, in all other cases, all algorithms achieve an accuracy of at least 0.95 as desired. We see that $\mathcal{A}_{\mathrm{SPRT}}$ outperforms both $\mathcal{A}_{\mathrm{Hoeffding}}$ and $\mathcal{A}_{\mathrm{LIL}}$ in any case, and regarding the strong optimality result for the SPRT from Prop. 2.17 this is not at all surprising. $\mathcal{A}_{\mathrm{LIL}}$ appears suboptimal for practical purposes, and for small values of $|p - 1/2|$ it performs even worse than the naive non-sequential solution $\mathcal{A}_{\mathrm{Hoeffding}}$.

Table 2.3.: Comparison of solutions to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$.

	$T^{\mathcal{A}}$			Accuracy		
p	$\mathcal{A}_{\mathrm{Hoeffding}}$	$\mathcal{A}_{\mathrm{SPRT}}$	$\mathcal{A}_{\mathrm{LIL}}$	$\mathcal{A}_{\mathrm{Hoeffding}}$	$\mathcal{A}_{\mathrm{SPRT}}$	$\mathcal{A}_{\mathrm{LIL}}$
0.51	150.0 (0)	**79.1** (2.1)	942.6 (6.8)	0.61	0.58	0.76
0.55	150.0 (0)	**65.9** (1.6)	560.0 (5.1)	0.91	0.85	1.00
0.60	150.0 (0)	**42.0** (0.9)	310.9 (2.8)	1.00	0.97	1.00
0.65	150.0 (0)	**29.2** (0.5)	197.4 (1.7)	1.00	0.99	1.00
0.70	150.0 (0)	**22.5** (0.4)	137.8 (1.2)	1.00	1.00	1.00
0.80	150.0 (0)	**15.1** (0.2)	74.8 (0.6)	1.00	1.00	1.00
0.90	150.0 (0)	**11.4** (0.1)	47.8 (0.3)	1.00	1.00	1.00

Next, we compare different solutions to $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ with each other. We write $\mathcal{A}_{\mathrm{lilUCB}}$ for the solution from Prop. 2.26 with the choices $\varepsilon = 0.01, \beta = 1$ and $\lambda = \left(\frac{2+\beta}{\beta}\right)^2 = 9$ since these assure its theoretical guarantees to hold, cf. our discussion after Prop. 2.26. Moreover, we write $\mathcal{A}_{\mathrm{DKW}}$ for Alg. 7 (with $k = 2$), $\mathcal{A}_{\mathrm{EGE}}$ for the solution from Prop. 2.27 and $\mathcal{A}_{\mathrm{PPR-Ber}}$ for Alg. 4. We ran each of these algorithms with parameter $\gamma = 0.05$ for 100 repetitions for different values of p, the achieved results are shown in Table 2.4. The observed accuracy was throughout 1.00 with the only exception that $\mathcal{A}_{\mathrm{PPR-Ber}}$ achieved 0.99 for $p = 0.55$.

We see that $\mathcal{A}_{\mathrm{PPR-Ber}}$ outperforms all other solutions on any of the considered instances. The algorithms $\mathcal{A}_{\mathrm{EGE}}$ and $\mathcal{A}_{\mathrm{lilUCB}}$ perform poorly. The round-based structure of $\mathcal{A}_{\mathrm{EGE}}$ and $\mathcal{A}_{\mathrm{DKW}}$ imply that the sample complexity is not strictly decreasing in $|p - 1/2|$, but instead there seems to be a minimum sample complexity required for any $p \in [0, 1]$, namely 625.0 for $\mathcal{A}_{\mathrm{DKW}}$ and 123150.0 for $\mathcal{A}_{\mathrm{EGE}}$ for the considered choice of $\gamma = 0.05$.

We repeated this experiment for smaller values of $|p - 1/2|$ and restricted ourselves to $\mathcal{A}_{\mathrm{DKW}}$ and $\mathcal{A}_{\mathrm{PPR-Ber}}$ at this point, the achieved results are shown in Table 2.4. The observed accuracy was 1.00 in all cases, with the only exceptions that $\mathcal{A}_{\mathrm{PPR-Ber}}$ achieved accuracy 0.99 for $p = 0.51$ and for $p = 0.54$. Again, $\mathcal{A}_{\mathrm{PPR-Ber}}$ apparently performs better

Table 2.4.: Comparison of solutions to $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$.

	$T^{\mathcal{A}}$			
p	$\mathcal{A}_{\mathrm{lilUCB}}$	$\mathcal{A}_{\mathrm{DKW}}$	$\mathcal{A}_{\mathrm{EGE}}$	$\mathcal{A}_{\mathrm{PPR-Ber}}$
0.55	67467.9 (1589.3)	16733 (453.5)	3783214.0 (0.0)	**1129.2** (79.2)
0.60	16618.9 (420.1)	3256.4 (123.3)	797091.0 (0.0)	**253.62** (15.7)
0.65	7144.0 (169.2)	785.5 (70.0)	132150.0 (0.0)	**77.8** (5.6)
0.70	4048.7 (100.1)	625.0 (0.0)	132150.0 (0.0)	**54.0** (3.9)
0.80	1780.3 (40.6)	625.0 (0.0)	132150.0 (0.0)	**25.2** (1.41)
0.90	968.8 (23.9)	625.0 (0.0)	132150.0 (0.0)	**12.6** (0.51)

than $\mathcal{A}_{\mathrm{DKW}}$ on any of the considered instances.

Table 2.5.: Comparison of $\mathcal{A}_{\mathrm{DKW}}$ and $\mathcal{A}_{\mathrm{PPR-Ber}}$ for small values of p.

	$T^{\mathcal{A}}$	
p	$\mathcal{A}_{\mathrm{DKW}}$	$\mathcal{A}_{\mathrm{PPR-Ber}}$
0.51	357484.9 (11037.3)	**33323.5** (1970.6)
0.52	82416.0 (2794.4)	**8484.7** (462.3)
0.53	57821.4 (3026.6)	**3264.4** (220.5)
0.54	18328.0 (0.0)	**1826.6** (126.7)

Mode Identification of a Die

Similarly as above, we want to compare several procedures for identifying the mode of a categorical random variable. For the sake of convenience, we restrict ourselves to the instances

- $\mathbf{p}_1 \coloneqq (0.5, 0.25, 0.25) \in \Delta_3^0$ with $h(\mathbf{p}_1) = 0.25$,
- $\mathbf{p}_2 \coloneqq (0.4, 0.2, 0.2, 0.2) \in \Delta_4^0$ with $h(\mathbf{p}_2) = 0.2$,
- $\mathbf{p}_3 \coloneqq (0.2, 0.1, 0.1, \ldots, 0.1) \in \Delta_9^0$ with $h(\mathbf{p}_3) = 0.1$,
- $\mathbf{p}_4 \coloneqq (0.1, 0.05, 0.05, \ldots, 0.05) \in \Delta_{19}^0$ with $h(\mathbf{p}_4) = 0.05$,
- $\mathbf{p}_5 \coloneqq (0.35, 0.33, 0.12, 0.1, 0.1) \in \Delta_5^0$ with $h(\mathbf{p}_5) = 0.02$,
- $\mathbf{p}_6 \coloneqq (0.35, 0.33, 0.04, 0.04, \ldots, 0.04) \in \Delta_{10}^0$ with $h(\mathbf{p}_6) = 0.02$,

which have also been considered by Jain et al. [2021] in their experimental evaluation. We write $\mathcal{A}_{\mathrm{DKW}}$ for Alg. 7 and $\mathcal{A}_{\mathrm{PPR1v1}}$ for Alg. 8. Table 2.6 shows sample mean (with standard error in brackets) for both procedures started with parameter $\gamma = 0.05$ on the instances $\mathbf{p}_1, \ldots, \mathbf{p}_6$, averaged over 100 repetitions each. For the sake of comparison, we have also included Alg. 5, denoted by $\mathcal{A}_{\mathrm{DKW}h}$, as a non-sequential solution to $\mathcal{P}^{k,h,\gamma}_{\mathrm{Die}}$, which obtains in addition to γ also the underlying (usually unknown) value $h(\mathbf{p}_i)$ when executed on $\mathbf{p}_i$.

We observed an accuracy of 1.00 for any procedure on any of $\mathbf{p}_1, \ldots, \mathbf{p}_6$ with the only exception that $\mathcal{A}_{\mathrm{PPR1v1}}$ achieved 0.99 for $\mathbf{p}_5$. With regard to the required sample complexity, $\mathcal{A}_{\text{PPR1v1}}$ clearly outperforms $\mathcal{A}_{\text{DKW}}$ and even $\mathcal{A}_{\text{DKW}h}$ on any of the six considered instances. The results for $\mathcal{A}_{\text{PPR1v1}}$ appear by and large consistent with those from Jain et al. [2021, Table 1], who evaluated $\mathcal{A}_{\text{PPR1v1}}$ on the same instances; the differences in termination time and standard error from our results are presumably due to partly different and small numbers of repetitions.

Table 2.6.: Comparison of $\mathcal{A}_{\text{DKW}}$, $\mathcal{A}_{\text{PPR1v1}}$ and $\mathcal{A}_{\text{DKW}h}$.

		$T^{\mathcal{A}}$		
$\mathbf{p}$	k	$\mathcal{A}_{\text{DKW}}$	$\mathcal{A}_{\text{PPR1v1}}$	$\mathcal{A}_{\text{DKW}h}$
$\mathbf{p}_1$	3	2614.6 (155.8)	**172.8** (8.2)	561.0 (0.0)
$\mathbf{p}_2$	4	3769.8 (44.9)	**241.0** (11.6)	877.0 (0.0)
$\mathbf{p}_3$	9	18328.0 (0.0)	**684.8** (24.5)	3506.0 (0.0)
$\mathbf{p}_4$	19	81016.0 (0.0)	**1661.8** (49.5)	14023.0 (0.0)
$\mathbf{p}_5$	5	357484.9 (11037.3)	**29234.2** (1562.2)	87641.0 (0.0)
$\mathbf{p}_6$	10	346392.0 (0.0)	**33390.8** (1492.4)	87641.0 (0.0)

2.5. A Change-of-Measure Argument

In this section, we will see a useful tool for proving lower bounds in the (multi-)dueling bandit scenario, which will be of use in Part III. As an application of it, we provide already in this section impossibility results for testing for some coherence in (multi-)dueling bandits under the low-noise assumption. Before stating it, we require some preparation.

For $S \in [m]_k$ and $\mathbf{p}, \mathbf{q} \in \Delta_S := \{(p'_j)_{j\in S} \in [0,1]^S \mid \sum_{j\in S} p'_j = 1\}$ let us write $\mathrm{KL}(\mathbf{p}, \mathbf{q})$ for the *Kullback-Leibler divergence* of random variables $X \sim \mathrm{Cat}(\mathbf{p})$ and $Y \sim \mathrm{Cat}(\mathbf{q})$, i.e.,

$$\mathrm{KL}(\mathbf{p}, \mathbf{q}) = \begin{cases} \sum_{x\in S: p_x>0} p_x \ln\left(\frac{p_x}{q_x}\right), & \text{if } \forall y \in S : q_y = 0 \Rightarrow p_y = 0, \\ \infty, & \text{otherwise.} \end{cases}$$

For the sake of convenience, we write in the binary case $k = 2$ simply $\mathrm{kl}(x, y) := \mathrm{KL}((x, 1-x), (y, 1-y))$ for any $x, y \in [0, 1]$.

Given a sequential testing algorithm $\mathcal{A}$ for the multi-dueling bandits scenario, let us write $S_t^{\mathcal{A}}$ for the query (element of $[m]_k$) made at time step t. Moreover, define $T_S^{\mathcal{A}}$ to be the number of times $\mathcal{A}$ makes the query $S \in [m]_k$ before termination, i.e., $T_S^{\mathcal{A}} = \sum_{t=1}^{T^{\mathcal{A}}} \mathbf{1}_{\{S_t^{\mathcal{A}}=S\}}$ and $T^{\mathcal{A}} = \sum_{S\in[m]_k} T_S^{\mathcal{A}}$ are fulfilled. Let $i_t^{\mathcal{A}} \in S_t^{\mathcal{A}}$ be the feedback observed by $\mathcal{A}$ at time step t, after having queried $S_t^{\mathcal{A}}$, and write $\mathcal{F}_t^{\mathcal{A}} := \sigma(S_1^{\mathcal{A}}, i_t^{\mathcal{A}}, \dots, S_t^{\mathcal{A}}, i_t^{\mathcal{A}})$ for the sigma algebra generated by the behaviour and observed feedback of $\mathcal{A}$ until time t, and as usual $\mathcal{F}_{T^{\mathcal{A}}} := \{E \in \mathcal{F} : E \cap \{T^{\mathcal{A}} \le t\} \in \mathcal{F}_t \forall t \in \mathbb{N}\}$ with $\mathcal{F} := \sigma\left(\bigcup_{t\in\mathbb{N}} \mathcal{F}_t\right)$.

Since $\mathcal{A}$ may be thought of as a multi-armed bandit with $\binom{m}{k}$ arms (one for each $S \in [m]_k$) and "rewards" $i_t^{\mathcal{A}} \in S_t^{\mathcal{A}}$, we may translate Lem. 1 from [Kaufmann et al., 2016] to our setting as follows.

Lemma 2.42. *Let $\mathbf{P}, \mathbf{P}' \in PM_k^m$ be such that $\mathbb{P}_{\mathbf{P}}$ and $\mathbb{P}_{\mathbf{P}'}$ are mutually absolutely continuous. If $\mathcal{A}$ is a sequential testing algorithm such that $T^{\mathcal{A}}$ is a.s. finite[7] w.r.t. $\mathbb{P}_{\mathbf{P}}$,*

[7]Since $\mathbb{P}_{\mathbf{P}}$ and $\mathbb{P}_{\mathbf{P}'}$ are mutually absolutely continuous, this is the case iff $T^{\mathcal{A}}$ is a.s. finite w.r.t. $\mathbb{P}_{\mathbf{P}}$.

then

$$\sum_{S\in[m]_k} \mathbb{E}_{\mathbf{P}}\left[T_S^{\mathcal{A}}\right] \mathrm{KL}\left(\mathbf{P}(\cdot|S), \mathbf{P}'(\cdot|S)\right) \geq \sup_{\mathcal{E}\in\mathcal{F}_{T^{\mathcal{A}}}} \mathrm{kl}\left(\mathbb{P}_{\mathbf{P}}(\mathcal{E}), \mathbb{P}_{\mathbf{P}'}(\mathcal{E})\right).$$

Proof. Cf. Lem. 1 in [Kaufmann et al., 2016]. □

Lemma 2.43. *(i) For any $S \in [m]_k$ and $\mathbf{p}, \mathbf{q} \in \Delta_S$ we have*

$$\mathrm{KL}\left(\mathbf{p}, \mathbf{q}\right) \leq \sum_{x\in S} \frac{(p_x - q_x)^2}{q_x}.$$

In particular, $\mathrm{kl}\left(p, 1-p\right) \leq \frac{(1-2p)^2}{p(1-p)}$ *and* $\mathrm{kl}\left(1/2 \pm h, 1/2 \mp h\right) \leq \frac{4h^2}{1/4-h^2}$ *hold for any* $p \in (0,1)$ *and* $h \in [0, 1/2)$*.*

(ii) The inequality $\mathrm{kl}(\gamma, 1-\gamma) \geq \ln \frac{1}{2.4\gamma}$ *holds for any* $\gamma \in (0,1)$*.*

Proof. The first statement from (i) is Lem. 3 in [Chen and Wang, 2018] and implies

$$\mathrm{kl}\left(p, 1-p\right) = \mathrm{KL}\left((p, 1-p), (1-p, p)\right) \leq (1-2p)^2 \left(\frac{1}{p} + \frac{1}{p-1}\right) = \frac{(1-2p)^2}{p(p-1)}$$

and thus also $\mathrm{kl}\left(1/2 \pm h, 1/2 \mp h\right) \leq \frac{4h^2}{1/4-h^2}$.
For (ii) confer Equation (3) in [Kaufmann et al., 2016]. □

2.5.1. Application: Impossibility Results

Note that we may regard any $\mathbf{P} = (\mathbf{P}(j|S))_{S\in[m]_k, j\in S} \in PM_k^m$ in a natural way as an element in $\mathbb{R}^{k\cdot\binom{m}{k}}$. This allows us to restrict the standard Euclidean norm $||\cdot||$ on $\mathbb{R}^{k\cdot\binom{m}{k}}$ to PM_k^m such that

$$\left|\left|\mathbf{P} - \mathbf{P}'\right|\right|^2 = \sum_{S\in[m]_k} \sum_{j\in S} (\mathbf{P}(j|S) - \mathbf{P}(j'|S))^2$$

for any $\mathbf{P}, \mathbf{P}' \in PM_k^m$. For non-empty subsets $\mathfrak{P}, \mathfrak{P}' \subseteq PM_k^m$ this gives us the distance

$$d(\mathfrak{P}, \mathfrak{P}') \coloneqq \inf_{\mathbf{P}\in\mathfrak{P}} \inf_{\mathbf{P}'\in\mathfrak{P}'} \left|\left|\mathbf{P} - \mathbf{P}'\right|\right|.$$

For the sake of convenience, let us write

$$\mathcal{U}_k^m \coloneqq \{\mathbf{P} \in PM_k^m \,|\, \mathbf{P}(j|S) > 0 \text{ for all } S \in [m]_k \text{ and all } j \in S\}.$$

Theorem 2.44. *Suppose $\gamma \in (0, \frac{1}{2.4})$ and let $\mathfrak{P}, \mathfrak{P}' \subseteq PM_k^m$ be disjoint with $\mathfrak{P} \cap \mathcal{U}_k^m \neq \emptyset \neq \mathfrak{P}' \cap \mathcal{U}_k^m$ and $d(\mathfrak{P} \cap \mathcal{U}_k^m, \mathfrak{P}' \cap \mathcal{U}_k^m) = 0$. If $\mathcal{A}$ is a sequential testing algorithm, which tests*

$$\mathbf{H}_0 : \mathbf{P} \in \mathfrak{P} \qquad \textit{versus} \qquad \mathbf{H}_1 : \mathbf{P} \in \mathfrak{P}' \tag{2.26}$$

with error probability at most γ correctly for any $\mathbf{P} \in \mathfrak{P} \cup \mathfrak{P}'$ in the sense that

$$\forall \mathbf{P} \in \mathfrak{P} : \mathbb{P}_{\mathbf{P}}(\mathbf{D}(\mathcal{A}) = 0) \geq 1-\gamma \quad \textit{and} \quad \forall \mathbf{P} \in \mathfrak{P}' : \mathbb{P}_{\mathbf{P}}(\mathbf{D}(\mathcal{A}) = 1) \geq 1-\gamma,$$
$$\textit{and } \forall \mathbf{P} \in \mathfrak{P} \cup \mathfrak{P}' \,:\, \mathbb{P}_{\mathbf{P}}(T^{\mathcal{A}} < \infty) = 1,$$

then we have for any $\mathbf{P}' \in \mathfrak{P}' \cap \mathcal{U}_k^m$ with $d(\mathbf{P}', \mathfrak{P} \cap \mathcal{U}_k^m) = 0$ that

$$\mathbb{E}_{\mathbf{P}'}[T^{\mathcal{A}}] = \sup_{\mathbf{P}\in\mathfrak{P}} \mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] = \infty.$$

Proof of Thm. 2.44. Suppose $\mathbf{P}' \in \mathfrak{P}' \cap \mathcal{U}_k^m$ to be fixed with $d(\mathbf{P}', \mathfrak{P} \cap \mathcal{U}_k^m) = 0$. Then, $c := \min_{S\in[m]_k, j\in S} \mathbf{P}'(j|S) > 0$ and due to $d(\mathbf{P}', \mathfrak{P} \cap \mathcal{U}_k^m) = 0$ there exists a sequence $\mathbf{P}_n = ((\mathbf{P}_n(j|S))_{S\in[m]_k, j\in S})_{n\in\mathbb{N}} \subseteq \mathfrak{P} \cap \mathcal{U}_k^m$ with $||\mathbf{P}_n - \mathbf{P}'|| \to 0$ as $n \to \infty$. Lem. 2.43 guarantees

$$\mathrm{KL}\left(\mathbf{P}_n(\cdot|S), \mathbf{P}'(\cdot|S)\right) \leq \sum\nolimits_{j\in S} \frac{(\mathbf{P}_n(j|S) - \mathbf{P}'(j|S))^2}{\mathbf{P}'(j|S)} \leq \frac{||\mathbf{P}_n - \mathbf{P}'||}{c}$$

for any $S \in [m]_k$. As $\mathcal{A}$ is able to test (2.26) with confidence $1-\gamma$, $\mathcal{E}_n := \{\mathbf{D}(\mathcal{A}) = 0\} \in \mathcal{F}_{T^{\mathcal{A}}}$ fulfills $\mathbf{P}_n(\mathcal{E}_n) \geq 1-\gamma$ and $\mathbf{P}'(\mathcal{E}_n) \leq \gamma$ for any $n \in \mathbb{N}$. Consequently, Lem. 2.42 and Lem. 2.43 let us infer

$$\begin{aligned}
\ln \frac{1}{2.4\gamma} &\leq \mathrm{kl}\left(\mathbf{P}_n(\mathcal{E}_n), \mathbf{P}'(\mathcal{E}_n)\right) \\
&\leq \sup\nolimits_{\mathcal{E}\in\mathcal{F}_{T^{\mathcal{A}}}} \mathrm{kl}\left(\mathbf{P}_n(\mathcal{E}), \mathbf{P}'(\mathcal{E})\right) \\
&\leq \sum\nolimits_{S\in[m]_k} \mathbb{E}_{\mathbf{P}_n}\left[T_S^{\mathcal{A}}\right] \mathrm{KL}\left(\mathbf{P}_n(\cdot|S), \mathbf{P}'(\cdot|S)\right) \\
&\leq \frac{||\mathbf{P}_n - \mathbf{P}'||}{c} \sum\nolimits_{S\in[m]_k} \mathbb{E}_{\mathbf{P}_n}[T_S^{\mathcal{A}}],
\end{aligned}$$

i.e., $||\mathbf{P}_n - \mathbf{P}'|| \to 0$ assures

$$\sup\nolimits_{\mathbf{P}\in\mathfrak{P}} \mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] \geq \sup\nolimits_{n\in\mathbb{N}} \mathbb{E}_{\mathbf{P}_n}[T^{\mathcal{A}}] \geq \sup\nolimits_{n\in\mathbb{N}} \frac{c}{||\mathbf{P}_n - \mathbf{P}'||} \ln \frac{1}{2.4\gamma} = \infty.$$

It remains to prove $\mathbb{E}_{\mathbf{P}'}[T^{\mathcal{A}}] = \infty$. Due to $||\mathbf{P}_n - \mathbf{P}'|| \to 0$ as $n \to \infty$, we can assume with regard to the choice of c w.l.o.g. $\min_{n\in\mathbb{N}} \min_{S\in[m]_k, j\in S} \mathbf{P}_n(j|S) > \frac{c}{2}$. Then, Lem. 2.43 yields

$$\mathrm{KL}\left(\mathbf{P}'(\cdot|S), \mathbf{P}_n(\cdot|S)\right) \leq \frac{2\,||\mathbf{P}_n - \mathbf{P}'||}{c}.$$

Similarly as above, we obtain

$$\ln \frac{1}{2.4\gamma} \leq \mathrm{kl}\left(\mathbf{P}'(\mathcal{E}_n), \mathbf{P}_n(\mathcal{E}_n)\right) \leq \frac{2\,||\mathbf{P}_n - \mathbf{P}'||}{c} \sum\nolimits_{S\in[m]_k} \mathbb{E}_{\mathbf{P}'}[T_S^{\mathcal{A}}] = \frac{2\,||\mathbf{P}_n - \mathbf{P}'||\, \mathbb{E}_{\mathbf{P}'}[T^{\mathcal{A}}]}{c}$$

and thus $\mathbb{E}_{\mathbf{P}'}[T^{\mathcal{A}}] = \infty$ follows by taking the limit $n \to \infty$. □

As a consequence, we will see in the following that testing several properties under the low-noise assumption is impossible in the beforementioned sense.

Testing for Plackett-Luce Marginals in Multi-Dueling Bandits We start with the *Plackett-Luce property* (PL), which e.g. serves as modeling assumption for dueling bandits in [Szörényi et al., 2015] and for multi-dueling bandits in [Saha and Gopalan, 2020b]. For $\boldsymbol{\theta} \in (0,\infty)^m$ we denote by $\mathbf{P}(\boldsymbol{\theta}) \in PM_k^m(\mathrm{PL})$ the corresponding PM, that is coherent with the Plackett-Luce model with parameter $\boldsymbol{\theta}$ on $\mathbb{S}_m :=$ {permutations on $[m]$}, i.e., $\mathbf{P}(\boldsymbol{\theta}) = \{\mathbf{P}(\boldsymbol{\theta})(\cdot|S)\}_{S\in[m]_k}$ is defined via

$$\mathbf{P}(\boldsymbol{\theta})(i|S) := \frac{\theta_i}{\sum_{a\in S} \theta_a} \quad \text{for any } S \in [m]_k \text{ and } i \in S.$$

As a direct consequence, we see that for any $i \in \{2, \ldots, m\}$ the parameter θ_i can be reconstructed from θ_1 and $\mathbf{P}(\boldsymbol{\theta})$ via $\theta_i = \frac{\mathbf{P}(\boldsymbol{\theta})(i|S)}{\mathbf{P}(\boldsymbol{\theta})(1|S)}\theta_1$, which holds for any $S \in [m]_k$ containing both i and 1. In other words, the underlying PL-parameter $\boldsymbol{\theta}$ is already determined up to multiplicity by parts of the marginals, namely by the values $\{\mathbf{P}(\boldsymbol{\theta})(i|S)\}_{S\in[m]_k:1\in S, i\in S}$. Since $\mathbf{P}(x\boldsymbol{\theta}) = \mathbf{P}(\boldsymbol{\theta})$ holds trivially for any $x > 0$, the knowledge of $\boldsymbol{\theta}$ up to multiplicities is sufficient for knowing $\mathbf{P}(\boldsymbol{\theta})$. This shows that any $\mathbf{P} \in PM_k^m(\mathrm{PL})$ is already fully determined by $\{\mathbf{P}(\boldsymbol{\theta})(i|S)\}_{S\in[m]_k:1\in S,i\in S}$.

Combining these observations with Thm. 2.44 allows us to deduce an impossibility result for the following problem, which is defined for parameters $m \in \mathbb{N}$, $k \in \mathbb{N}$ with $2 \le k < m$, $h \in [0, 1)$ and $\alpha, \beta \in (0, 1)$: We say that a multi-dueling bandits algorithm $\mathcal{A}$ solves PL *testing on* $PM_k^m(\Delta^h)$ *for* α *and* β (short: $\mathcal{P}_{\mathrm{PL}}^{m,k,\alpha,\beta}(\Delta^h)$) if it is able to test $\mathbf{H}_0 : \mathbf{P} \in PM_k^m(\mathrm{PL})$ versus $\mathbf{H}_1 : \mathbf{P} \in PM_k^m(\neg\mathrm{PL})$ with type I/II error at most α/β for any $\mathbf{P} \in PM_k^m(\Delta^h)$ and a.s. terminates for any of these, i.e., if it fulfills

$$\begin{aligned} \forall \mathbf{P} \in PM_k^m(\Delta^h) &: \mathbb{P}_{\mathbf{P}}(T^{\mathcal{A}} < \infty) = 1, \\ \forall \mathbf{P} \in PM_k^m(\Delta^h \wedge \mathrm{PL}) &: \mathbb{P}_{\mathbf{P}}(\mathbf{D}(\mathcal{A}) = \mathrm{PL}) \ge 1 - \alpha, \\ \forall \mathbf{P} \in PM_k^m(\Delta^h \wedge \neg\mathrm{PL}) &: \mathbb{P}_{\mathbf{P}}(\mathbf{D}(\mathcal{A}) = \neg\mathrm{PL}) \ge 1 - \beta. \end{aligned}$$

Before stating the negative result for PL testing, we prepare its proof with two observations on $PM_k^m(h\mathrm{GCW} \wedge \mathrm{PL})$ and $PM_k^m(\Delta^h \wedge \mathrm{PL})$, which will again be of use in Sec. 6.1.

Lemma 2.45. *For* $\boldsymbol{\theta} \in (0, \infty)^m$ *with* $\theta_1 \ge \cdots \ge \theta_m$ *we have*

$$\begin{aligned} \mathbf{P}(\boldsymbol{\theta}) \in PM_k^m(h\mathrm{GCW}) &\Leftrightarrow \forall j \in \{2, \ldots, k\} : h(\theta_1 + \cdots + \theta_k) + \theta_j - \theta_1 \le 0 \\ &\Leftrightarrow h(\theta_1 + \cdots + \theta_k) + \theta_2 - \theta_1 \le 0 \end{aligned}$$

and

$$\mathbf{P}(\boldsymbol{\theta}) \in PM_k^m(\Delta^h) \Leftrightarrow \forall i \in [m-k] : h(\theta_i + \cdots + \theta_{i+k-1}) + \theta_{i+1} - \theta_i \le 0.$$

Proof. This follows directly from the definitions. □

From this, we obtain the following result, whose proof is given in the appendix.

Lemma 2.46. *For any* $h \in (0, 1)$ *and* $m, k \in \mathbb{N}$ *with* $k \le m$ *we have* $PM_k^m(\mathrm{PL} \wedge h\mathrm{GCW}) \supseteq PM_k^m(\mathrm{PL} \wedge \Delta^h) \neq \emptyset$.

Corollary 2.47. *If* $\mathcal{A}$ *solves* $\mathcal{P}_{\mathrm{PL}}^{m,k,\alpha,\beta}(\Delta^h)$ *for* $\alpha, \beta \in (0, \frac{1}{2.4})$*, then* $\mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] = \infty$ *for any* $\mathbf{P} \in PM_k^m(\Delta^h \wedge \mathrm{PL}) \cap \mathcal{U}_k^m$ *with* $d(\mathbf{P}, PM_k^m(\Delta^h \wedge \neg\mathrm{PL}) \cap \mathcal{U}_k^m) = 0$ *and also any* $\mathbf{P} \in PM_k^m(\Delta^h \wedge \neg\mathrm{PL}) \cap \mathcal{U}_k^m$ *with* $d(\mathbf{P}, PM_k^m(\Delta^h \wedge \mathrm{PL}) \cap \mathcal{U}_k^m) = 0$*. In particular,*

$$\sup\nolimits_{\mathbf{P}\in PM_k^m(\Delta^h\wedge\mathrm{PL})} \mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] = \infty = \sup\nolimits_{\mathbf{P}\in PM_k^m(\Delta^h\wedge\neg\mathrm{PL})} \mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}].$$

Proof. At first, note that $\mathbf{P} \in \mathcal{U}_k^m$ holds trivially for any $\mathbf{P} \in PM_k^m(\mathrm{PL})$. Let us define $\mathfrak{P} := PM_k^m(\Delta^h \wedge \mathrm{PL})$ and $\mathfrak{P}' := PM_k^m(\Delta^h \wedge \neg\mathrm{PL})$. From Lem. 2.46 we know $\mathfrak{P} \neq \emptyset$. Thus, we may fix $\mathbf{P} \in \mathfrak{P}$ and $S' \in [m]_k$ with $1 \notin S \supseteq \{2, 3\}$. There exists $\varepsilon \in (-1, 1)$ such that defining $\mathbf{P}'(2|S') := \mathbf{P}(2|S') + \varepsilon$, $\mathbf{P}'(3|S) := \mathbf{P}'(3|S') - \varepsilon$ and $\mathbf{P}' := \mathbf{P}(i|S)$ for $S \in [m]_k$, $i \in S$, for which $S \neq S'$ or $i \notin \{2, 3\}$, results in a probability model $\mathbf{P}' = \{\mathbf{P}'(i|S)\}_{S\in[m]_k, i\in S} \in PM_k^m(\Delta^h) \cap \mathcal{U}_k^m$. Assuming $\mathbf{P}' \in PM_k^m(\mathrm{PL})$ for the moment, the fact that $\mathbf{P}(\cdot|\cdot)$ and $\mathbf{P}'(\cdot|\cdot)$ coincide on the determining set $\{(i,S)\,|\,S \in [m]_k \text{ with } 1 \in S, i \in S\}$ would imply $\mathbf{P} = \mathbf{P}'$, a contradiction. Hence, $\mathbf{P}' \in PM_k^m(\Delta^h \wedge \neg\text{PL}) \cap \mathcal{U}_k^m = \mathfrak{P}' \cap \mathcal{U}_k^m$ follows.

In the construction above, ε may be chosen such that $|\varepsilon|$ is arbitrarily small. Thus, we obtain $d(\mathfrak{P} \cap \mathcal{U}_k^m, \mathfrak{P}' \cap \mathcal{U}_k^m) = 0$, and the rest follows from Thm. 2.44. □

The class $PM_k^m(\text{PL})$ coincides with that of the *multi-nomial logit model*, which in turn is a special case of a *random utility model* (RUM); note that the RUM has briefly been introduced in Sec. 1.4. Recently, Saha and Gopalan [2020a] have investigated multi-dueling bandit problems under such generalized assumptions, and it may be of interest for future work to check whether analoga of the above result are valid for RUMs as well. Nevertheless, we stop at this point and continue with further types of coherence of $\mathbf{Q}$ that are prominent assumptions in the dueling bandits scenario.

Testing for Further Coherences in Dueling Bandits In the dueling bandits scenario, there are further coherences of $\mathbf{Q}$ assumed in the literature. Via Thm. 2.44, we can see that many of these are impossible to test in the above mentioned sense.

Formally, given a property X of reciprocal relations and parameters $m \in \mathbb{N}$, $h \in (0, 1/2)$ and $\alpha, \beta \in (0,1)$, let us say that a dueling bandit algorithm $\mathcal{A}$ solves X *testing on* $\mathcal{Q}_m^h$ (short: $\mathcal{P}_{\text{X}}^{m,h,\alpha,\beta}$) if it is able to decide with type I/II errors at most α/β for any $\mathbf{Q} \in \mathcal{Q}_m^h$ whether $\mathbf{H}_0 : \mathbf{Q} \in \mathcal{Q}_m(\text{X})$ or $\mathbf{H}_1 : \mathbf{Q} \in \mathcal{Q}_m(\neg\text{X})$ is true and terminates a.s. for any $\mathbf{Q} \in \mathcal{Q}_m^h$, i.e., if

$$\forall \mathbf{Q} \in \mathcal{Q}_m^h : \mathbb{P}_{\mathbf{Q}}(T^{\mathcal{A}} < \infty) = 1,$$
$$\forall \mathbf{Q} \in \mathcal{Q}_m^h(\text{X}) : \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \text{X}) \geq 1 - \alpha,$$
$$\forall \mathbf{Q} \in \mathcal{Q}_m^h(\neg\text{X}) : \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \neg\text{X}) \geq 1 - \beta.$$

For proving worst-case sample complexity bounds of solutions to $\mathcal{P}_{\text{X}}^{m,h,\alpha,\beta}$, the following simplified version of Thm. 2.44 will be sufficient. For the sake of convenience, we abbreviate $\mathcal{Q}_m^{\clubsuit} := \{\mathbf{Q} \in \mathcal{Q}_m \,|\, \forall (i,j) \in (m)_2 : q_{i,j} \geq 1/2\}$.

Corollary 2.48. *Let* X *be a property for reciprocal relations with* $\mathcal{Q}_m^h(\text{X}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m \neq \emptyset \neq \mathcal{Q}_m^h(\neg\text{X}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$ *and* $\alpha, \beta \in (0, \frac{1}{2.4})$*. Then, any solution* $\mathcal{A}$ *to* $\mathcal{P}_{\text{X}}^{m,h,\alpha,\beta}$ *fulfills*

$$\sup\nolimits_{\mathbf{Q} \in \mathcal{Q}_m^h(\text{X})} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \sup\nolimits_{\mathbf{Q} \in \mathcal{Q}_m^h(\neg\text{X})} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty.$$

Proof. Choose $\mathfrak{P} = \mathcal{Q}_m^h(\text{X}) \cap \mathcal{Q}_m^{\clubsuit}$ and $\mathfrak{P}' = \mathcal{Q}_m^h(\neg\text{X}) \cap \mathcal{Q}_m^{\clubsuit}$. Since $(\mathfrak{P} \cap \mathcal{U}_2^m) \cup (\mathfrak{P}' \cap \mathcal{U}_2^m) = \mathcal{Q}_m^h \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$ is connected, the statement follows from Thm. 2.44. □

Let us give a brief overview of those properties X, for which we will infer negative results by means of Cor. 2.48 below. For more information on the relations among themselves (including further properties such as stochastic transitivities) confer [Bengs et al., 2021]. A reciprocal relation $\mathbf{Q} \in \mathcal{Q}_m$ is said to satisfy the *stochastic triangle inequality* (STI) if it fulfills

$$q_{i,j}, q_{j,k} \geq \frac{1}{2} \;\Rightarrow\; q_{i,k} \leq q_{i,j} + q_{j,k} - \frac{1}{2}$$

for any distinct $i,j,k \in [m]$. As a direct consequence of this definition, we see that $\mathcal{Q}_m^h \subseteq \mathcal{Q}_m(\text{STI})$ for any $h \geq \frac{1}{4}$. For this reason, we will restrict ourselves to the case $h < \frac{1}{4}$

for STI testing. The assumption STI has e.g. been made in [Yue et al., 2012, Falahatgar et al., 2017b, Ren et al., 2020].

Another property, which is e.g. of interest in [Zimmert and Seldin, 2018] and [Bengs et al., 2021], is the *general identifiablity assumption* (GIA). Formally, $\mathbf{Q} \in \mathcal{Q}_m$ fulfills this property if there exists $i \in [m]$ s.t.

$$\forall j \in [m] \setminus \{i\} : \forall l \in [m] \setminus \{i\} : \, q_{i,l} > q_{j,l}.$$

Moreover, Bengs et al. [2021] considered the property *low noise model* (LNM), which is defined to hold for $\mathbf{Q} \in \mathcal{Q}_m$ if $q_{i,j} \neq 1/2$ and

$$q_{i,j} > 1/2 \quad \Rightarrow \quad \sum\nolimits_{l \in [m]} q_{i,l} > \sum\nolimits_{l \in [m]} q_{j,l}$$

is fulfilled for any distinct i, j in $[m]$.

As mentioned in Sec. 1, one frequently assumed coherence is that $\mathbf{Q}$ is the marginal of some certain type of reciprocal probability distribution on $\mathbb{S}_m$. Recall here that, if $\mathbb{P}$ is such a distribution, its corresponding marginals $\mathbf{Q}^{\mathbb{P}} = (q^{\mathbb{P}}_{i,j})_{1 \leq i,j \leq m} \in \mathcal{Q}_m$ are given as

$$q^{\mathbb{P}}_{i,j} = \sum\nolimits_{\sigma \in \mathbb{S}_m \text{ with } \sigma(i) < \sigma(j)} \mathbb{P}(\sigma).$$

We say that $\mathbf{Q} \in \mathcal{Q}_m$ fulfills

- Marg if $\mathbf{Q} = \mathbf{Q}^{\mathbb{P}}$ for an arbitrary probability distribution $\mathbb{P}$ on $\mathbb{S}_m$,
- Mal if $\mathbf{Q} = \mathbf{Q}^{\mathbb{P}}$ for a *Mallows distribution* [Mallows, 1957] $\mathbb{P}$ on $\mathbb{S}_m$,
- BS if $\mathbf{Q} = \mathbf{Q}^{\mathbb{P}}$ for a *Babington Smith distribution* [Smith, 1950] $\mathbb{P}$ on $\mathbb{S}_m$,

where the notions “Mallows distribution” and “Babington Smith distribution” are clarified below.

The property Marg is clearly the most general one. Actually, it has not gained much attention in the dueling bandit literature so far, but nevertheless, we added it for the sake of completeness.

The assumption $\mathbf{Q} \in \mathcal{Q}_m(\text{Mal})$ has e.g. been made in [Busa-Fekete et al., 2014a] for identifying the most-preferred arm according to the underlying Mallows probability distribution. There, it has also been shown that $\mathbf{Q} \in \mathcal{Q}_m$ fulfills Mal iff there exist $\nu \in \mathbb{S}_m$ and $\phi \in (0, 1]$ s.t.

$$q_{i,j} = \omega_\phi(\nu(j) - \nu(i) + 1) - \omega_\phi(\nu(j) - \nu(i)) \tag{2.27}$$

holds for any $1 \leq i < j \leq m$ with $\omega_\phi(k) := \frac{k}{1-\phi^k}$.

The condition BS is not yet discussed in the realm of dueling bandits. However, it is a generalization of Mal and may be of interest for further research. Even though we do not require knowledge of the concrete marginal formulas for our purposes, we provide the definition of the model for the sake of completeness. Formally, given an appropriate parameter $\mathbf{W} = (w_{i,j})_{1 \leq i,j \leq m} \in \mathcal{Q}_m$, the corresponding Babington Smith distribution $\mathbb{P}^{\text{BS}(\mathbf{W})}$ is defined by $\mathbb{P}^{\text{BS}(\mathbf{W})}(\sigma) := \hat{\mathbb{P}}^{\text{BS}(\mathbf{W})}(\sigma) / C(\mathbf{W})$ for any $\sigma \in \mathbb{S}_m$, where

$$\hat{\mathbb{P}}^{\text{BS}(\mathbf{W})}(\sigma) := \prod\nolimits_{1 \leq i < j \leq m} w_{i,j}^{I_{i,j}(\sigma)} (1 - w_{i,j})^{1 - I_{i,j}(\sigma)} \quad \text{and} \quad C(\mathbf{W}) := \sum\nolimits_{\sigma \in \mathbb{S}_m} \hat{\mathbb{P}}^{\text{BS}(\mathbf{W})}(\sigma)$$

with $I_{i,j}(\sigma) := \mathbf{1}_{\{\sigma(i)<\sigma(j)\}}$. Here, $\mathbf{W}$ is meant to be appropriate if the above is well-defined, i.e., if $C(\mathbf{W}) \neq 0$.

The impossibility result on Marg testing is prepared with the following lemma.

Lemma 2.49. *If* $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{Marg})$*, then*

$$q_{i,j} + q_{j,k} - 1 \leq q_{i,k} \leq q_{i,j} + q_{j,k} \quad \textit{for any distinct } i,j,k \in [m]. \tag{2.28}$$

Proof of Lem. 2.49. This is Thm. 3.2.1 in [Fligner and Verducci, 1993]. □

This result has a close connection to the concept of fuzzy transitivity as e.g. discussed in [Świtalski, 2003, Haddenhorst et al., 2020]: The property (2.28) is fulfilled iff $\mathbf{Q}$ is T_L-transitive, where $T_L(x,y) = \max(x+y-1,0)$ is the *Łukasiewicz T-norm.* As noted by Fishburn and Falmagne [1989, p. 479], (2.28) is also sufficient for $\mathbf{Q} \in \mathbf{Q}_m(\mathrm{Marg})$ iff $m \leq 5$.

Corollary 2.50. *Suppose* $\mathrm{X} \in \{\mathrm{GIA}, \mathrm{Marg}, \mathrm{Mal}, \mathrm{BS}\}$ *and* $h \in (0,1/2)$*, or* $\mathrm{X} = \mathrm{STI}$ *and* $h < \frac{1}{4}$*, or* $\mathrm{X} = \mathrm{LNM}$ *and* $h < \frac{m-2}{2m}$*. If* $\alpha, \beta \in (0, \frac{1}{2.4})$ *and* $\mathcal{A}$ *solves* $\mathcal{P}_{\mathrm{X}}^{m,h,\alpha,\beta}$*, then*

$$\sup\nolimits_{\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{X})} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \sup\nolimits_{\mathbf{Q} \in \mathcal{Q}_m^h(\neg \mathrm{X})} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty.$$

Proof. To apply Cor. 2.48 it is sufficient to show, for any choice of X, existence of two reciprocal relations $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{X}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$ and $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\mathrm{X}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$. With some abuse of notation, after replacing for convenience h by $h+\delta$ for some small $\delta > 0$, we may suppose w.l.o.g. $\mathcal{Q}_m^h = \{\mathbf{Q} \in \mathcal{Q}_m \,|\, \forall (i,j) \in (m)_2 : |q_{i,j} - \frac{1}{2}| \geq h\}$. We treat each case separately.

(i) **Case** $\mathrm{X} = \mathrm{GIA}$**.** Define $\mathbf{Q}$ and $\mathbf{Q}'$ via

$$q_{i,j} := \begin{cases} 1 - \frac{h}{2}, & \text{if } i = 1, \\ \frac{1}{2} + h, & \text{otherwise,} \end{cases} \quad \text{and} \quad q'_{i,j} := \begin{cases} 1 - \frac{h}{2}, & \text{if } j - i = 1, \\ \frac{1}{2} + h, & \text{otherwise} \end{cases}$$

for any $1 \leq i < j \leq m$ and note that $\mathbf{Q}, \mathbf{Q}' \in \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$. Since $q_{1,l} - q_{j,l} \geq \frac{1}{2} - \frac{3h}{2} > 0$ for all $\forall j,l \in [m] \setminus \{1\}$, we have $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{GIA})$. For $i \in [m-2]$ we have $q'_{i,i+2} - q'_{i+1,i+2} = \frac{h}{2} - \frac{1}{2} < 0$, and similarly for $i \in \{m-1, m\}$ we have $q'_{i,i-1} - q'_{i-2,i-1} = -\frac{1}{2} - \frac{h}{2} < 0$, hence $\mathbf{Q}' \in \mathcal{Q}_m(\neg\mathrm{GIA})$ holds.

(ii) **Case** $\mathrm{X} = \mathrm{Marg}$**.** There exists $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{PL}) \cap \mathcal{Q}_m^{\clubsuit}$ according to Lem. 2.46, which is trivially also an element in $\mathcal{U}_2^m$. Due to $\mathcal{Q}_m^h(\mathrm{PL}) \subseteq \mathcal{Q}_m^h(\mathrm{Marg})$ we obtain $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{Marg}) \cap \mathcal{Q}_m^{\clubsuit}$. Moreover, for sufficiently small $\varepsilon > 0$ the relation $\mathbf{Q}'$, defined by $q'_{1,m} = 1 - 3\varepsilon$ and $q'_{i,j} = 1 - \varepsilon$ for any $1 \leq i < j \leq m$ with $(i,j) \neq (1,m)$, is an element in $\mathcal{Q}_m^h \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$ and fulfills $q'_{1,m-1} + q'_{m-1,m} - 1 = 1 - 2\varepsilon > q'_{1,m}$. Hence, Lem. 2.49 assures $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\mathrm{Marg}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$.

(iii) **Case** $\mathrm{X} = \mathrm{Mal}$**.** With $\nu = \mathrm{id}_{[m]} \in \mathbb{S}_m$ the expression in (2.27) becomes $\omega_\phi(j - i + 1) - \omega_\phi(j-i)$, which is in $(0,1)$ and tends to 1 as $\phi \to 0$ whenever $i < j$. Thus, a sufficiently small choice of ϕ yields a reciprocal relation $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{Mal}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$. By (ii), there exists $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\mathrm{Marg}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m \subseteq \mathcal{Q}_m^h(\neg\mathrm{Mal}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$.

(iv) **Case** X = BS. As shown in [Mallows, 1957], any Mallows distribution is in particular also a Babington Smith distribution. Consequently, the element $\mathbf{Q}$ from (iii) fulfills $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{Mal}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m \subseteq \mathcal{Q}_m^h(\mathrm{BS}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$. Moreover, we have for $\mathbf{Q}'$ from (ii) that $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\mathrm{Marg}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m \subseteq \mathcal{Q}_m^h(\neg\mathrm{BS}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$.

(v) **Case** X = STI. Choose $\mathbf{Q}$ via $q_{i,j} \coloneqq 1 - \frac{h}{2}$ for any $1 \le i < j \le m$. Moreover, let $\mathbf{Q}'$ be defined via $q'_{1,2} = q'_{2,3} = \frac{1}{2} + h$, $q'_{1,3} = 1$ and $q'_{i,j} = 1 - \frac{h}{2}$ for any other $1 \le i < j \le m$.

(vi) **Case** X = LNM. In case $h < \frac{m-2}{2m}$, we can fix an arbitrary $\varepsilon \in \left(0, \frac{1}{2} - \frac{mh}{m-2}\right)$. The reciprocal relations $\mathbf{Q}$ and $\mathbf{Q}'$, defined via

$$q_{i,j} \coloneqq 1/2 + h \quad \text{and} \quad q'_{i,j} \coloneqq \begin{cases} \frac{1}{2} + h, & \text{if } i = 1, \\ 1 - \varepsilon, & \text{otherwise} \end{cases}$$

for any $1 \le i < j \le m$, are elements of $\mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$. If $q_{i,j} > \frac{1}{2}$, then $i < j$ and $\sum_{l \in [m]} q_{i,l} - q_{j,l} = 2h(j-i) > 0$ holds, hence $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{LNM})$. Moreover, we have $q'_{1,2} > 0$, and $\varepsilon < \frac{1}{2} - \frac{mh}{m-2}$ lets us infer

$$\begin{aligned}\sum\nolimits_{l \in [m]} q'_{1,l} - q'_{2,l} &= (q'_{1,1} - q'_{2,1}) + (q'_{1,2} - q'_{2,2}) + \sum\nolimits_{3 \le l \le m} q'_{1,l} - q'_{2,l} \\ &= h + h + (m-2)\left(-\frac{1}{2} + h + \varepsilon\right) = mh + (m-2)\left(\varepsilon - \frac{1}{2}\right) > 0.\end{aligned}$$

This shows $\mathbf{Q}' \in \mathcal{Q}_m(\neg\mathrm{LNM})$.

□

2.6. Discussion and Related Work

The collection of concentration inequalities presented in Sec. 2.1 is far from complete and restriced to those, which we actually make use of in the course of this thesis, a more thorough overview can e.g. be found in [Massart, 2007].

Based on Hoeffding's inequality, we obtained in Lem. 2.10 already a non-sequential solution to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ that is asymptotically optimal. However, the SPRT plays a major role, as its optimality result allowed us to infer a lower bound for sequential testing algorithms.

Prop. 2.13 is interesting since it provides a sharper lower bound for solutions to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ in case $h = 0$, namely the worst-case sample complexity of solutions to $\mathcal{P}^{0,\gamma}_{\mathrm{Coin}}$ w.r.t. instances in $\mathcal{Q}_m^h$ is by a factor $\ln\ln\frac{1}{h}$ larger than corresponding worst-case sample complexity of solutions to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$. The original bound by Farrell [1964] underlying Prop. 2.13 holds for the more general case where the X_t's belong to the exponential family, but we have restricted ourselves to the case of Bernoulli random variables for simplicity. Jamieson et al. [2013] provide a weaker version of this result, which comes with a simpler proof than the original one but additionally assumes $\mathcal{A}$ to be a GSPRT as defined above and the X_t to be normally distributed. Chen and Li [2015] improved upon Farrell's original bound in the sense that they weakened the exponential family assumption and also provide guarantees that are slightly stronger than merely bounding the limes superior of the expected termination time. We restricted ourselves to Farrell's lower bound result as it already suffices our purposes.

As the theoretically (in an appropriate sense) optimal solution to $\mathcal{P}^{\gamma}_{\text{Coin}}$ from Prop. 2.22 seems infeasible for practical purposes, we also included two solutions based on reductions to multi-armed bandits as well as PPR-Bernoulli from [Jain et al., 2021]. The latter one is to the best of our knowledge w.r.t. empirical performance currently the best solution to $\mathcal{P}^{\gamma}_{\text{Coin}}$ in the literature. As already mentioned, there exist even more solutions to $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ and $\mathcal{P}^{\gamma}_{\text{Coin}}$. For example, Zhao et al. [2016] and Balsubramani [2014] presented further anytime LIL confidence bounds that are similar to Lem. 2.6 and could potentially be used to construct in a similar fashion as Prop. 2.23 corresponding GSPRTs that solve $\mathcal{P}^{h,\gamma}_{\text{Coin}}$. The works [Karp and Kleinberg, 2007] and [Ren et al., 2020] contain alternative solutions to $\mathcal{P}^{\gamma}_{\text{Coin}}$, and apart from that, numerous further solutions to the best-arm identification problem in MABs such as those in [Kalyanakrishnan et al., 2012, Shah et al., 2020, Kaufmann and Kalyanakrishnan, 2013] can be used in Alg. 1 for creating solutions to $\mathcal{P}^{\gamma}_{\text{Coin}}$; cf. [Chen et al., 2017] for a more extensive literature overview and a more thorough analysis of best-arm identification in MABs. Last but not least, $\mathcal{P}^{2,h,\gamma}_{\text{Die}} \approx \mathcal{P}^{h/2,\gamma}_{\text{Coin}}$ and $\mathcal{P}^{2,0,\gamma}_{\text{Die}} = \mathcal{P}^{\gamma}_{\text{Coin}}$ assure that any solution to $\mathcal{P}^{k,h,\gamma}_{\text{Die}}$ can be seen with $k = 2$ as a solution to the corresponding mode identification problem for a coin. Comparing all solutions prevalent in the literature would certainly exceed the scope of this chapter, hence we restricted ourselves to those ones stated in Sec. 2.2.2, because they apparently suffice to achieve (almost) asymptotically optimal sample complexity results for our problems of interest in Part II.

Lem. 17 in [Ren et al., 2019a] resp. Lem. 15 in the corrected preprint version [Ren et al., 2019b] provides a sample complexity lower bound for a problem related to that discussed in Sec. 2.2.3. Informally, they consider the problem of testing for any of m biased coins whether the bias is head or tails, whereas we are interested in deciding whether all of them are biased towards head or not. Apparently, their problem is more difficult than ours, hence proving sample complexity lower bounds of corresponding solutions is easier in their setting than in ours. They establish a bound of order $\Omega_{\sup}(\frac{m}{h^2} \ln\ln \frac{1}{h})$ by using the above mentioned improved version of Farrell's lower bound underlying Prop. 2.13 from Chen and Li [2015]. Unfortunately, this argument does not seem applicable in our setting and we merely obtained a bound of order $\Omega_{\sup}(\frac{1}{h^2} \ln\ln \frac{1}{h})$.

The Dvoretzky-Kiefer-Wolfowitz (DKW) inequality stated in Lem. 2.34 actually holds for quite general distributions of X_t [Kosorok, 2008]. For the particular case of categorical random variables, Devroye [1983] and Berend and Kontorovich [2012] provide improved concentration inequalities for $||\mathbf{p} - \hat{\mathbf{p}}_t||_1 = \sum_{i \in [k]} |p_i - (\hat{\mathbf{p}}_t)_i|$ instead of $||\mathbf{p} - \hat{\mathbf{p}}_t||_\infty$. For our purpose of mode identification, these did not result in improved solutions to $\mathcal{P}^{k,h,\gamma}_{\text{Die}}$.

Alg. 7 will be a main ingredient for our solution to $\mathcal{P}^{m,k,\gamma}_{\text{GCWi}}(\text{GCW} \wedge \Delta^0)$ in Ch. 6. Its idea is similar to that underlying the solution SEEBS from [Ren et al., 2020] to $\mathcal{P}^{\gamma}_{\text{Coin}}$, a major difference thereof is that its guarantees are based on the DKW inequality instead of Hoeffding's inequality and the Chernoff bound. Alg. 6 is an important building block of Alg. 7, but it will also be of importance for our solution to $\mathcal{P}^{m,k,\gamma}_{\text{GCWi}}(h\text{GCW} \wedge \Delta^0)$ in Ch. 6.

We have seen in Prop. 2.23 how an appropriate anytime confidence bound (Lem. 2.6) can be used to construct a solution to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$, which has a worst-case sample complexity of order $\mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ whilst having the chance to terminate early. This raises the question whether an appropriate anytime version of the DKW inequality may lead to similar solutions to $\mathcal{P}^{k,h,\gamma}_{\mathrm{Die}}$, i.e., whether an analogon of Lem. 2.6 holds with essentially the same $U_\gamma(n)$ but $\sum_{k=1}^n Y_k^{(p)}$ replaced by $||\mathbf{p}-\hat{\mathbf{p}}_n||_\infty$. The proof technique used in Lem. 3 in [Jamieson et al., 2013], which is underlying Lem. 2.6, does not appear transferrable for this purpose. Recently, Howard and Ramdas [2021, Thm. 2] gave an anytime version of the DKW inequality, which could be of use for the construction of such a solution.

Shah et al. [2020] also analyzed the mode identification problem for categorical random variables. Based on Lem. 2.42, they proved an instance-wise lower bound (cf. Thm. 3 in their paper), which is asymptotically comparable to Prop. 2.30. In contrast to theirs, our bound also captures the asymptotical behavior as $k\to\infty$. Moreover, the authors provide a solution for $\mathcal{P}^{k,0,\gamma}_{\mathrm{Die}}$ (cf. Thm. 2 in their paper). For this, they make use of confidence intervals that are given via improved empirical Bernstein bounds. Our solution from Prop. 2.40 theoretically outperforms theirs with respect to two essential aspects: First, its sample complexity bound is constant instead of increasing in k and second, their dependence on the hardness parameter $h(\mathbf{p})$ is $\frac{1}{h^2(\mathbf{p})}\ln\frac{1}{h(\mathbf{p})}$ instead of $\frac{1}{h^2(\mathbf{p})}\ln\ln\frac{1}{h(\mathbf{p})}$ as in Prop. 2.30. Their solution empirically performs worse than PPR1v1 [Jain et al., 2021] and was thus not added to our empirical comparison.

Jain et al. [2021] showed that their solution $\mathcal{A}=$ PPR1v1 has in addition to its good empirical performance also the appealing property that $\lim_{\gamma\searrow 0}\frac{\mathbb{E}_{\mathbf{p}}[T^{\mathcal{A}}]}{\mathrm{LB}(\mathbf{p},\gamma)}=1$ for any $\mathbf{p}\in\Delta_k^0$, where $\mathrm{LB}(\mathbf{p},\gamma)$ is the above mentioned sample complexity lower bound for an arbitrary solution to $\mathcal{P}^{k,0,\gamma}_{\mathrm{Die}}$ on instance $\mathbf{p}$ from [Shah et al., 2020]. This raises the question whether the other solutions to $\mathcal{P}^{k,h,\gamma}_{\mathrm{Die}}$ or $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ come with analogous guarantees.

We think that the idea of constructing solutions to $\mathcal{P}^{k,h,\gamma}_{\mathrm{Die}}$ from solutions to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ via a *one-vs-one* paradigm could possibly be fruitful for the case $h>0$ as well. More precisely, one may ask whether the SPRT as optimal solution to $\mathcal{P}^{h',\gamma}_{\mathrm{Coin}}$ can be used to construct a good solution to $\mathcal{P}^{k,h,\gamma}_{\mathrm{Die}}$, and presumably the difficulty lies in deciding which value $h'=h'(h)$ to choose here.

The measure-changing argument stated as Lem. 2.42 is a frequently used approach for proving sample complexity lower bounds in (multi-)dueling bandits [Saha and Gopalan, 2020b, Agarwal et al., 2020, Bengs and Hüllermeier, 2020] and it could also be used as an argument for proving lower bounds in Chapters 4 and 5 below. However, we will exploit the optimality of the SPRT for this purpose, because the SPRT has apparently the best guarantees for solutions to $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$, cf. Prop. 2.17.

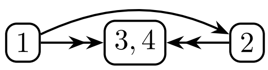
In Sec. 2.5, we already discussed the problem to test several properties of probability models and reciprocal relations, which are of relevance in (multi-)dueling bandits, including the stochastic triangle inequality (STI) [Yue et al., 2012, Falahatgar et al., 2017b, Ren et al., 2020] as well as the properties of being marginals of a Plackett-Luce [Szörényi et al., 2015, Saha and Gopalan, 2020b], a Mallows [Busa-Fekete et al., 2014a] or any probability distribution on rankings on $[m]$. However, our impossibility results from Sec. 2.5 must be used with caution: They do not at all imply that testing any of the aforementioned properties is per se impossible, but only show that testing these with finite worst-case termination time under the low-noise assumption is not possible. When restricting e.g. to instances in $\mathcal{Q}_m^{\mathrm{STI}>h} := \{\mathbf{Q} \in \mathcal{Q}_m \,|\, d(\mathbf{Q}, \partial\mathcal{Q}_m(\mathrm{STI})) > h\}$, one could presumably construct a solution $\mathcal{A}$ to STI testing on $\mathcal{Q}_m^{\mathrm{STI}>h}$ with finite worst-case termination time, and similarly for other properties $\mathrm{X} \neq \mathrm{STI}$. But in contrast to the low-noise assumption, assumptions like $\mathbf{Q} \in \mathcal{Q}_m^{\mathrm{STI}>h}$ seem rather artificial and unreasonable to us. Also, one could still formulate instance-wise bounds for solutions to the STI testing on $\mathcal{Q}_m^h$ for α and β. Even though our results from Sec. 2.5 may appear rather trivial, we think they could be of interest as they indicate to some extend the hardness of checking such frequently made statistical assumptions in (multi-)dueling bandits. There exists some literature on testing hypothesis in the context of such assumptions, e.g., Busa-Fekete et al. [2021] tested in an offline manner for the parameter of a Mallows model by means of ranking feedback and Rastogi et al. [2020] analyzed the impact of WST, SST, MST and the Plackett-Luce assumption as parameter restrictions in a two-sample hypothesis testing scenario. However, such results appear rather loosely related to the learning tasks in this thesis, thus we did not consider them in more detail.

3. Graph-Theoretical Considerations

This chapter is dedicated to the graph-theoretical prerequisites required in Chapters 4 and 5 and also analyzes deterministic variants of the problems CW identification, CW testification, CW verification and WST testing. Throughout this chapter, if not explicitly stated otherwise, we assume $m \geq 3$.

3.1. Basic Terminology

Let us write $\mathcal{G}_m$ for the set of all simple *directed graphs* (short: *digraphs*) on $[m]$ without loops and with at most one edge between each two nodes. In other words, $\mathcal{G}_m$ contains all digraphs $G = ([m], E_G)$ with $E_G \subseteq \langle m \rangle_2 := \{(i,j) \in [m] \times [m] \mid i \neq j\}$ such that $(i,j) \notin E_G$ or $(j,i) \notin E_G$ holds for every distinct $i,j \in [m]$. Let $\overline{\mathcal{G}}_m$ be the set of *tournaments* on $[m]$, i.e., $\overline{\mathcal{G}}_m \subseteq \mathcal{G}_m$ contains all digraphs $G = ([m], E_G)$, where for every $(i,j) \in \langle m \rangle_2$ either $(i,j) \in E_G$ (we write $i \to j$ in G, or $i \xrightarrow{G} j$) or $(j,i) \in E_G$. For $G \in \mathcal{G}_m$ and disjoint $V_1, V_2 \subseteq [m]$ we use in illustrations a double arrow $V_1 \twoheadrightarrow V_2$ to indicate that G contains all the edges $i_1 \to i_2$ with $i_1 \in V_1, i_2 \in V_2$. For example, the graph $G = ([m], E_G) \in \mathcal{G}_m$ with the set of edges $E_G = \{(1,2),(1,3),(1,4),(2,3),(2,4)\}$ may be illustrated as follows.



Given a tournament $G \in \overline{\mathcal{G}}_m$, a permutation $\sigma \in \mathbb{S}_m$ is called a *topological sorting* if $\sigma(i) < \sigma(j)$ iff $i \to j$ in G. Moreover, we call $(i_1, \ldots, i_k) \in [m]^k$ a k-cycle (or simply a cycle) in G if $i_1, \ldots, i_k$ are distinct, $i_1 \to \ldots \to i_{k-1} \to i_k \to i_1$ holds and $k \geq 3$ is fulfilled. We say that $G \in \mathcal{G}_m$ is *acyclic* if it does not contain any cycle. We say that a *property* X, that is applicable to tournaments on $[m]$, is *possible* if the restricted set

$$\overline{\mathcal{G}}_m(\mathrm{X}) := \left\{G \in \overline{\mathcal{G}}_m \mid G \text{ fulfills X}\right\}$$

is non-empty. Given possible properties $\mathrm{X}, \mathrm{X}_1, \ldots, \mathrm{X}_n$, we implicitly define the properties $\neg\mathrm{X}$, $\mathrm{X}_1 \wedge \cdots \wedge \mathrm{X}_n$ and $\mathrm{X}_1 \vee \cdots \vee \mathrm{X}_n$ via $\overline{\mathcal{G}}_m(\neg\mathrm{X}) := \overline{\mathcal{G}}_m \setminus \overline{\mathcal{G}}_m(\mathrm{X})$, $\overline{\mathcal{G}}_m(\mathrm{X}_1 \wedge \cdots \wedge \mathrm{X}_n) := \bigcap_{1 \leq l \leq n} \overline{\mathcal{G}}_m(\mathrm{X}_l)$ and $\overline{\mathcal{G}}_m(\mathrm{X}_1 \vee \cdots \vee \mathrm{X}_n) := \bigcup_{1 \leq l \leq n} \overline{\mathcal{G}}_m(\mathrm{X}_l)$, respectively. We focus on combinations of the possible properties $\emptyset$, i, CW and acyclic given by

$$\begin{aligned}
\overline{\mathcal{G}}_m(\emptyset) &:= \overline{\mathcal{G}}_m, \\
\overline{\mathcal{G}}_m(i) &:= \left\{G \in \overline{\mathcal{G}}_m \mid \forall j \in [m] \setminus \{i\} \,:\, i \to j\right\}, \\
\overline{\mathcal{G}}_m(\mathrm{CW}) &:= \bigcup\nolimits_{i \in [m]} \overline{\mathcal{G}}_m(i), \\
\overline{\mathcal{G}}_m(\text{acyclic}) &:= \left\{G \in \overline{\mathcal{G}}_m \mid \exists \text{ topological sorting of } G\right\}.
\end{aligned}$$

Clearly, $G \in \overline{\mathcal{G}}_m$ is acyclic iff G has a topological sorting, that is, we have $\{G \in \overline{\mathcal{G}}_m \mid G \text{ is acyclic}\} = \overline{\mathcal{G}}_m(\text{acyclic})$.

For $\mathbf{Q} \in \mathcal{Q}_m^0$, the *associated tournament* $G(\mathbf{Q}) \in \overline{\mathcal{G}}_m$ *of* $\mathbf{Q}$ is defined via

$$i \to j \text{ in } G(\mathbf{Q}) \quad :\Leftrightarrow \quad q_{i,j} > \frac{1}{2}.$$

The map $\Phi_m : \mathcal{R}_m \to \overline{\mathcal{G}}_m, \mathbf{Q} \mapsto G(\mathbf{Q})$ provides a one-to-one connection between the tournaments on $[m]$ and the deterministic reciprocal relations on $[m]$. Note that $\mathbf{Q} \in \mathcal{Q}_m^0 \supsetneq \mathcal{R}_m$ has i as Condorcet winner iff $i \to j$ holds in $G(\mathbf{Q})$ for any $j \in [m] \setminus \{i\}$, i.e., we have $\Phi_m(\mathcal{R}_m(i)) = \overline{\mathcal{G}}_m(i)$ and $\Phi_m(\mathcal{R}_m(\mathrm{CW})) = \overline{\mathcal{G}}_m(\mathrm{CW})$. Hence, we may say that i is the *Condorcet winner* of $G \in \overline{\mathcal{G}}_m$ if $\Phi_m^{-1}(G) \in \mathcal{R}_m$ and then write $\mathrm{CW}(G) = i$. We obtain the following connection between stochastically transitive reciprocal relations and acyclic tournaments.

Proposition 3.1. *If* $\mathbf{Q} \in \mathcal{Q}_m^0$, *then*

$$\mathbf{Q} \in \mathcal{Q}_m^0(\mathrm{WST}) \quad \Leftrightarrow \quad G(\mathbf{Q}) \in \overline{\mathcal{G}}_m(\text{acyclic}).$$

Moreover, for $\mathrm{XST} \in \{\mathrm{MST}, \mathrm{SST}, \nu\mathrm{RST}\}$ *we have* $\Phi_m(\mathcal{R}_m(\mathrm{XST})) = \overline{\mathcal{G}}_m(\text{acyclic})$.

We prepare its proof with the following lemma, which is common knowledge [Moon, 2015]. For the sake of completeness, we restate it here and provide a proof in the appendix.

Lemma 3.2. *Some* $G \in \overline{\mathcal{G}}_m$ *is acyclic iff it does not contain a* 3*-cycle.*

Proof of Prop. 3.1. Note that $\mathbf{Q} = (q_{i,j})_{1 \le i,j \le m} \in \mathcal{Q}_m^0$ is not WST if and only if there exist distinct $i, j, k \in [m]$ such that $q_{i,j}, q_{j,k} \ge \frac{1}{2}$ and $q_{i,k} < \frac{1}{2}$. Regarding the definition of $\mathcal{Q}_m^0$, this is equivalent to $q_{i,j}, q_{j,k}, q_{k,i} > \frac{1}{2}$, which is fulfilled if and only if $G(\mathbf{Q})$ contains a 3-cycle. The first statement then follows from Lem. 3.2, which shows that the existence of a 3-cycle in $G(\mathbf{Q})$ is equivalent to the existence of a k-cycle for $k \ge 3$ in $\Phi_m(\mathbf{Q})$.
The second statement follows from the first one by using that any $\mathbf{Q} \in \mathcal{R}_m \subsetneq \mathcal{Q}_m^0$ fulfills

$$\begin{aligned} \mathbf{Q} \in \mathcal{R}_m(\mathrm{XST}) \quad &\Leftrightarrow \quad \text{for all distinct } i, j, k \in [m] : (q_{i,j} = q_{j,k} = 1 \;\Rightarrow\; q_{i,k} = 1) \\ &\Leftrightarrow \quad \mathbf{Q} \in \mathcal{R}_m(\mathrm{WST}). \end{aligned}$$

□

By exploiting the connection of reciprocal relations to tournaments, one obtains the following result. This has already been stated in [Haddenhorst et al., 2020] and for the sake of completeness we provide a proof in the appendix.

Lemma 3.3. *For any* $\mathbf{Q} \in \mathcal{Q}_m^0$ *there exists a permutation* σ *on* $[m]$ *s.t.* $q_{\sigma(i),\sigma(i+1)} > \frac{1}{2}$ *for every* $i \in [m-1]$.

The remainder of this chapter is dedicated to the analysis of the following deterministic variants of the CW-related and transitivity testing problems introduced in Sec. 1.3:

- *CW testification on* $\overline{\mathcal{G}}_m$ (short: $\mathcal{D}^m_{\mathrm{CWt}}$): For any $G \in \overline{\mathcal{G}}_m(\mathrm{CW})$, if G has a CW, return it, and otherwise return $\neg$CW.

- *CW checking on* $\overline{\mathcal{G}}_m$ (short: $\mathcal{D}^m_{\mathrm{CWc}}$): For any $G \in \overline{\mathcal{G}}_m(\mathrm{CW})$, decide whether G has a CW or not.

- *CW verification on* $\overline{\mathcal{G}}_m$ (short: $\mathcal{D}^m_{\text{CWv}}$): For any $G \in \overline{\mathcal{G}}_m(\text{CW})$ and any input $z \in [m]$, decide whether $\text{CW}(G) = z$ or not.
- *CW identification on* $\overline{\mathcal{G}}_m(\text{CW})$ (short: $\mathcal{D}^m_{\text{CWi}}(\text{CW})$): For any $G \in \overline{\mathcal{G}}_m(\text{CW})$, identify the CW of G.
- *CW verification on* $\overline{\mathcal{G}}_m(\text{CW})$ (short: $\mathcal{D}^m_{\text{CWv}}(\text{CW})$): For any $G \in \overline{\mathcal{G}}_m(\text{CW})$ and any input $z \in [m]$, decide whether $\text{CW}(G) = z$ or not.
- *acyclicity testing on* $\overline{\mathcal{G}}_m$ (short: $\mathcal{D}^m_{\text{acyclic}}$): For any $G \in \overline{\mathcal{G}}_m$, decide whether G is acyclic or not.

Formally, any of these problems can be regarded as a deterministic variant of its corresponding probabilistic counterpart with error probability $\gamma = 0$ and the additional assumption $\mathbf{Q} \in \mathcal{R}_m$. Unlike in the dueling bandits scenario, we mainly restrict ourselves throughout this chapter to deterministic solutions of these problems, i.e., to such, which are not allowed to choose their queries in a probabilistic manner. For this purpose, we start with some observations for such algorithms and more general decision problems.

3.2. Deterministic Sequential Testing Algorithms

Let Y be an arbitrary property of tournaments on $[m]$. We say a *deterministic sequential testing algorithm (DSTA) for* $\overline{\mathcal{G}}_m(\text{Y})$ is an algorithm $\mathcal{A}$, which, when started on any $G \in \overline{\mathcal{G}}_m$, chooses until its termination at each time step t a *query* $\{i^{\mathcal{A}}_G(t), j^{\mathcal{A}}_G(t)\} \in [m]_2$ and then observes whether $i^{\mathcal{A}}_G(t) \to j^{\mathcal{A}}_G(t)$ or $j^{\mathcal{A}}_G(t) \to i^{\mathcal{A}}_G(t)$ holds in G. Here, we suppose that $\mathcal{A}$ receives only m as parameter and has – apart from the fact $G \in \overline{\mathcal{G}}_m(\text{Y})$ – no a priori knowledge of G. For any G, let $\mathbf{D}^{\mathcal{A}}(G)$ be the return value of $\mathcal{A}$ started on G and write $T^{\mathcal{A}}_G$ for the termination time of $\mathcal{A}$ started on G, i.e., the number of queries made by $\mathcal{A}$ before termination. We measure the performance of $\mathcal{A}$ by means of its *worst-case termination time*

$$T^{\mathcal{A}} := T^{\mathcal{A}}_{\text{worst}} := \max_{G \in \overline{\mathcal{G}}_m(\text{Y})} T^{\mathcal{A}}_G,$$

and at times, we will also consider its *best-case termination time*

$$T^{\mathcal{A}}_{\text{best}} := \min_{G \in \overline{\mathcal{G}}_m(\text{Y})} T^{\mathcal{A}}_G.$$

As querying any already queried $\{i, j\} \in [m]_2$ is of no benefit, we assume w.l.o.g. that all queries made by $\mathcal{A}$ are disjoint, hence we have $T^{\mathcal{A}}_G \le |E_G| \le \binom{m}{2}$ for any $G \in \overline{\mathcal{G}}_m$, and $T^{\mathcal{A}} \le \binom{m}{2}$. Moreover, we define the *picture of* $\mathcal{A}$ *started on* G *at time* t as that digraph $\mathfrak{G}^{\mathcal{A}}_G(t) := ([m], E') \in \mathcal{G}_m$ with

$$E' := \left\{(i, j) \,|\, \exists t' \le t : \left\{i^{\mathcal{A}}_G(t'), j^{\mathcal{A}}_G(t')\right\}_G = (i, j)\right\},$$

where we have used the notation

$$\{i', j'\}_G := \begin{cases} (i', j'), & \text{if } i' \to j' \text{ in G}, \\ (j', i'), & \text{if } j' \to i' \text{ in G} \end{cases}$$

for distinct $i', j' \in [m]$. Note that $\mathfrak{G}^{\mathcal{A}}_G(t)$ contains all information gathered by $\mathcal{A}$ started on G until time t.

If $\mathrm{X}_1, \dots, \mathrm{X}_n$ are possible properties of tournaments on $[m]$ s.t. $\overline{\mathcal{G}}_m(\mathrm{Y}) = \bigcup_{k\in[n]} \overline{\mathcal{G}}_m(\mathrm{X}_k)$ is a distinct union, we say that a DSTA $\mathcal{A}$ for $\overline{\mathcal{G}}_m(\mathrm{Y})$ *solves* $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}(\mathrm{Y})$ if

$$\forall k \in [n]\, \forall G \in \overline{\mathcal{G}}_m(\mathrm{X}_k) \,:\, \mathbf{D}^{\mathcal{A}}(G) = \mathrm{X}_k.$$

We call a solution $\mathcal{A}$ to $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}(\mathrm{Y})$ *optimal* if every solution $\mathcal{A}'$ to $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}(\mathrm{Y})$ fulfills $T^{\mathcal{A}'} \geq T^{\mathcal{A}}$. In case $\mathrm{Y} = \emptyset$, we simply write $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}$ for $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}(\mathrm{Y})$.

Note that, with this general framework, we are already able to model four of the six deterministic decision problems from above as follows:

$$\mathcal{D}^m_{\mathrm{CWt}} = \mathcal{D}^m_{\mathrm{1,\dots,m,\neg CW}}, \quad \mathcal{D}^m_{\mathrm{CWc}} = \mathcal{D}^m_{\mathrm{CW,\neg CW}},$$
$$\mathcal{D}^m_{\mathrm{CWi}}(\mathrm{CW}) = \mathcal{D}^m_{\mathrm{1,\dots,m}}(\mathrm{CW}) \quad \text{and} \quad \mathcal{D}^m_{\mathrm{acyclic}} = \mathcal{D}^m_{\mathrm{acyclic,\neg acyclic}}.$$

For being able to incorporate CW verification, we require yet a further generalization. Given a non-empty set $\mathcal{Z}$, let us say that $\mathcal{A}$ is a *DSTA for* $\overline{\mathcal{G}}_m(\mathrm{Y})$ *with input space* $\mathcal{Z}$ if, for any $z \in \mathcal{Z}$, $\mathcal{A}$ started with z (denoted by $\mathcal{A}(z)$) is a DSTA for $\overline{\mathcal{G}}_m(\mathrm{Y})$. If $\mathcal{A}$ is started with z on G, we write $\{i^{\mathcal{A}(z)}_G, j^{\mathcal{A}(z)}_G\}$ for the query at time t, $\mathfrak{G}^{\mathcal{A}(z)}_G(t)$ for its picture at time t of G and $D^{\mathcal{A}(z)}(G)$ for the corresponding output, and the worst- resp. best-case termination times of $\mathcal{A}$ are simply defined as

$$T^{\mathcal{A}} = T^{\mathcal{A}}_{\mathrm{worst}} \coloneqq \max_{z\in\mathcal{Z}} T^{\mathcal{A}(z)}_{\mathrm{worst}} \quad \text{and} \quad T^{\mathcal{A}}_{\mathrm{best}} \coloneqq \min_{z\in\mathcal{Z}} T^{\mathcal{A}(z)}_{\mathrm{best}}.$$

If $\mathrm{X}_1, \dots, \mathrm{X}_n \,:\, \mathcal{Z} \mapsto \{\text{possible properties of tournaments on } [m]\}$ are s.t. $\overline{\mathcal{G}}_m(\mathrm{Y}) = \bigcup_{k\in[n]} \overline{\mathcal{G}}_m(\mathrm{X}_k(z))$ is a disjoint union for any $z \in \mathcal{Z}$, we say that a DSTA $\mathcal{A}$ for $\overline{\mathcal{G}}_m(\mathrm{Y})$ with input space $\mathcal{Z}$ *solves* $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}[\mathcal{Z}](\mathrm{Y})$ if

$$\forall z \in \mathcal{Z} \,:\, \mathcal{A}(z) \text{ solves } \mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}[z](\mathrm{Y}) \coloneqq \mathcal{D}^m_{\mathrm{X_1(z),\dots,X_n(z)}}(\mathrm{Y}).$$

Furthermore, let us call a solution $\mathcal{A}$ to $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}[\mathcal{Z}](\mathrm{Y})$ *optimal* if every solution $\mathcal{A}'$ to $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}[\mathcal{Z}](\mathrm{Y})$ fulfills $T^{\mathcal{A}'} \geq T^{\mathcal{A}}$. In case $\mathrm{Y} = \emptyset$, we simply write $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}[\mathcal{Z}]$ for $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}[\mathcal{Z}](\mathrm{Y})$.

Clearly, this notion of DSTAs with inputs generalizes that of DSTAs, and in the particular case $|\mathcal{Z}| = 1$ these concepts are equivalent. With the particular choices $\mathcal{Z} \coloneqq [m]$ and $\mathrm{X}^*_1(z) \coloneqq z$, $\mathrm{X}^*_2(z) \coloneqq \neg z$ for each $z \in \mathcal{Z}$, we obtain the characterizations

$$\mathcal{D}^m_{\mathrm{CWv}} = \mathcal{D}^m_{\mathrm{X^*_1,X^*_2}}[\mathcal{Z}] \quad \text{and} \quad \mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW}) = \mathcal{D}^m_{\mathrm{X^*_1,X^*_2}}[\mathcal{Z}](\mathrm{CW}).$$

When m, properties $\mathrm{Y}, \mathrm{X}_1, \dots, \mathrm{X}_n$ and an input space $\mathcal{Z}$ are fixed, the number of different[1] DSTAs is finite and thus conducting an extensive brute-force search would trivially allow to find an optimal solution to $\mathcal{D}^m_{\mathrm{X_1,\dots,X_n}}[\mathcal{Z}](\mathrm{Y})$. However, this approach seems practically infeasible, because the set of possible DSTAs is extremely large. To get an intuition how large this value could be, note that a DSTA $\mathcal{A}$ for $\overline{\mathcal{G}}_m(\mathrm{Y})$ with input space $\mathcal{Z}$, which fulfills $|\mathcal{Z}| = 1$ for simplicity, is specified by its initital query $\{i_1, j_1\}$ as well as the actions it takes when observing a feedback. Here, the set of possible actions consists of querying an unqueried edge or returning a decision. $\mathcal{A}$ may be depicted in form of a tree as in Fig. 3, where each node represents an action. Upon querying $\{i^x_t, j^x_t\}$ (the x-th query in

[1]Here, we regard two DSTAs $\mathcal{A}_1$ and $\mathcal{A}_2$ as equal iff both $\mathbf{D}^{\mathcal{A}_1}(G) = \mathbf{D}^{\mathcal{A}_2}(G)$ and $T^{\mathcal{A}_1}_G = T^{\mathcal{A}_2}_G$ hold for any $G \in \overline{\mathcal{G}}_m$.

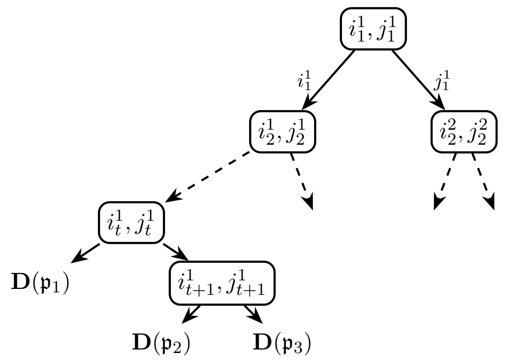


Figure 3.: Representation of a DSTA as a tree.

the t-th row of the tree), $\mathcal{A}$ observes as feedback either i_t^x or j_t^x and then decides for the action in the left resp. child if it observed i_t^x resp. j_t^x. It continues this way until it finally reaches a leave node, where it outputs a decision $\mathbf{D}(\mathfrak{p}) \in \{\mathrm{X}_1, \ldots, \mathrm{X}_n\}$ depending on the path $\mathfrak{p}$ from this leave node to the root i_1, j_1. Here, $\mathfrak{p}$ encodes all the queries chosen and feedback observed by $\mathcal{A}$. If N is the number of leave nodes, there are n^N different possible assignments of $\mathrm{X}_1, \ldots, \mathrm{X}_n$ to the leave nodes and any such assignment leads to a different DSTA. If $\mathcal{A}$ queries for every $G \in \overline{\mathcal{G}}_m$ all $\binom{m}{2}$ edges of G, the number of leave nodes is $N = 2^{\binom{m}{2}}$. Hence, there are at least $n^{(2^{m(m-1)/2})}$ different DSTAs. Regarding that we have ignored all cases where this representation of $\mathcal{A}$ is not a complete tree, the total number of possible DSTAs is in fact by far larger than this lower bound, and a brute-force search for an optimal solution to $\mathcal{D}^m_{\mathrm{X_1,\ldots,X_n}}[\mathcal{Z}](\mathrm{Y})$ appears infeasible.

3.3. Properties in Extension

To construct a solution $\mathcal{A}$ to $\mathcal{D}^m_{\mathrm{X_1,\ldots,X_n}}[\mathcal{Z}](\mathrm{Y})$ with small worst-case termination time, we would like to decide (for each $i \in [n]$) as early as possible and only based on the input z and $\mathfrak{G}_G^{\mathcal{A}(z)}(t)$ whether $G \in \overline{\mathcal{G}}_m(\mathrm{Y})$ fulfills $G \in \overline{\mathcal{G}}_m(\mathrm{X}_i(z))$ or not. For being at time t absolutely sure that $G \in \overline{\mathcal{G}}_m(\mathrm{X}_i(z))$ holds, each supergraph $G' \in \mathcal{G}_m(\mathrm{Y})$ of $\mathfrak{G}_G^{\mathcal{A}(z)}(t)$ should fulfill $\mathrm{X}_i(z)$. This motivates the following definitions, where Y is an arbitrary possible property of tournaments on $[m]$:

- For $G = ([m], E_G) \in \mathcal{G}_m$, a Y*-extension of* G is a tournament $G' = ([m], E_{G'}) \in \overline{\mathcal{G}}_m(\mathrm{Y})$ with $E_G \subseteq E_{G'}$. If $\mathrm{Y} = \emptyset$, such a G' is simply called an *extension of* G.
- Given a possible property X of tournaments on $[m]$ with $\overline{\mathcal{G}}_m(\mathrm{X}) \subseteq \overline{\mathcal{G}}_m(\mathrm{Y})$, a digraph $G \in \mathcal{G}_m$ *is* X *in* Y*-extension* if the set of all Y-extensions of G is a non-empty subset of $\overline{\mathcal{G}}_m(\mathrm{X})$. We write $\mathcal{G}_m(\mathrm{X} \mid \mathrm{Y})$ for the set of all $G \in \mathcal{G}_m$ that are X in Y-extension. In case $\mathrm{Y} = \emptyset$, we abbreviate $\mathcal{G}_m(\mathrm{X}) := \mathcal{G}_m(\mathrm{X} \mid \mathrm{Y})$ and simply say that each $G \in \mathcal{G}_m(\mathrm{X})$ fulfills X in extension. For the sake of convenience, we write $\mathcal{G}_m(\mathrm{X} \mid \mathrm{Y}) := \mathcal{G}_m(\mathrm{X} \wedge \mathrm{Y} \mid \mathrm{Y})$ in case $\overline{\mathcal{G}}_m(\mathrm{X}) \not\subseteq \overline{\mathcal{G}}_m(\mathrm{Y})$.

In our notation, $\mathcal{G}_m(\text{acyclic})$ is *not* the set of all acyclic digraphs in $\mathcal{G}_m$, instead it is the set of all $G \in \mathcal{G}_m$ that are *acyclic in extension.* In case $m \geq 3$ we have

$$\{G \in \mathcal{G}_m \mid G \text{ is acyclic}\} \supsetneq \mathcal{G}_m(\text{acyclic}),$$

e.g., $([m], \emptyset)$ is an acyclic digraph that is not acyclic in extension.

To get used to the newly introduced notation, let us consider a small example. The only extensions of $G \coloneqq ([3], \{(1,2),(2,3)\}) \in \mathcal{G}_3$ are the tournaments

$$G_1 \coloneqq ([3], \{(1,2),(2,3),(1,3)\}) \quad \text{and} \quad G_2 \coloneqq ([3], \{(1,2),(2,3),(3,1)\}).$$

In contrast to G_2, G_1 is also an acyclic-extension of G; the topological sorting of G_2 is $\mathrm{id}_{[3]} : [3] \to [3], x \mapsto x$. From $G_2 \notin \overline{\mathcal{G}}_3(\text{acyclic})$ we infer $G \notin \mathcal{G}_3(\text{acyclic})$, and $G_1 \in \overline{\mathcal{G}}_3(\text{acyclic})$ implies $G \notin G_3(\neg\text{acyclic})$. Since G_1 is the only acyclic-extension of G and it is an element of $\mathcal{G}_3(1 \,|\, \text{acyclic})$ but not of $\mathcal{G}_3(2 \,|\, \text{acyclic})$, we have $G \in \mathcal{G}_3(1 \,|\, \text{acyclic})$ and $G \notin \mathcal{G}_3(2 \,|\, \text{acyclic})$. Moreover, $1 \overset{G}{\to} 2$ implies that G has no 3-extension, and thus $G \notin \mathcal{G}_3(\mathrm{X} \,|\, 3)$ for any property X.

The following Prop. 3.4 provides a link between X testing under Y-assumption and the notion of X in Y-extension. Even though it may appear a bit technical, its basic idea is rather naive and not new to the literature [Bollobás, 1978, p. 429ff.]. In simple words, it states that a correct DSTA can only terminate with decision X as soon as it is absolutely sure that G has property X.

Proposition 3.4. *If a DSTA $\mathcal{A}$ for $\overline{\mathcal{G}}_m(\mathrm{Y})$ with input space $\mathcal{Z}$ solves $\mathcal{D}^m_{\mathrm{X}_1,\ldots,\mathrm{X}_n}[\mathcal{Z}](\mathrm{Y})$, then*

$$\forall z \in \mathcal{Z}\, \forall k \in [n]\, \forall G \in \overline{\mathcal{G}}_m(\mathrm{X}_k(z)) \,:\, \mathfrak{G}_G^{\mathcal{A}(z)}\left(T_G^{\mathcal{A}(z)}\right) \in \mathcal{G}_m(\mathrm{X}_k(z) \,|\, \mathrm{Y}).$$

Proof. Let $z \in \mathcal{Z}$ be fixed. Recall $\overline{\mathcal{G}}_m(\mathrm{Y}) = \bigcup_{k \in [n]} \overline{\mathcal{G}}_m(\mathrm{X}_k(z))$. To show the contraposition, suppose there is some $k \in [n]$ and $G \in \overline{\mathcal{G}}_m(\mathrm{X}_k(z))$ with $\mathfrak{G}_G^{\mathcal{A}(z)}\left(T_G^{\mathcal{A}(z)}\right) \notin \mathcal{G}_m(\mathrm{X}_k(z) \,|\, \mathrm{Y})$. Then, there exists some extension G' of $\mathfrak{G}_G^{\mathcal{A}(z)}\left(T_G^{\mathcal{A}(z)}\right)$ with $G' \in \overline{\mathcal{G}}_m(\mathrm{Y}) \setminus \overline{\mathcal{G}}_m(\mathrm{X}_k[z])$, i.e., $G' \in \overline{\mathcal{G}}_m(\mathrm{X}_{k'}[z])$ for some $k' \neq k$. Since $\mathcal{A}(z)$ started on G' observes until termination exactly the same feedback as if started on G, we have $\mathbf{D}^{\mathcal{A}(z)}(G') = \mathbf{D}^{\mathcal{A}(z)}(G) = \mathrm{X}_k(z) \neq \mathrm{X}_{k'}(z)$. Thus, $\mathcal{A}$ does not solve $\mathcal{D}^m_{\mathrm{X}_1,\ldots,\mathrm{X}_n}[\mathcal{Z}](\mathrm{Y})$. □

Since $\mathcal{A}(z)$ started on G is assumed to make every query $\{i,j\}$ at most once, $\mathfrak{G}_G^{\mathcal{A}(z)}\left(T_G^{\mathcal{A}(z)}\right)$ contains exactly $T_G^{\mathcal{A}(z)}$ edges, and Prop. 3.4 directly lets us infer the following result.

Corollary 3.5. *If a DSTA $\mathcal{A}$ for $\overline{\mathcal{G}}_m(\mathrm{Y})$ with input space $\mathcal{Z}$ solves $\mathcal{D}^m_{\mathrm{X}_1,\ldots,\mathrm{X}_n}[\mathcal{Z}](\mathrm{Y})$, then*

$$\forall z \in \mathcal{Z}\, \forall k \in [n]\, \forall G \in \overline{\mathcal{G}}_m(\mathrm{X}_k(z)) \,:\, T_G^{\mathcal{A}(z)} \geq \min\left\{|E| : ([m], E) \in \mathcal{G}_m(\mathrm{X}_k(z) \,|\, \mathrm{Y})\right\}.$$

In particular,

$$T^{\mathcal{A}}_{\text{best}} \geq \min\left\{|E| : ([m], E) \in \mathcal{G}_m(\mathrm{X}_k(z) \,|\, \mathrm{Y}) \text{ for some } z \in \mathcal{Z}, k \in [n]\right\}.$$

Proof. This is a direct consequence of Prop. 3.4. □

Before continuing with more specific choices of X and Y, we state in the next lemma some rather straight-forward observations concerning the newly introduced notions. For the sake of completeness, we provide its proof in the appendix.

Lemma 3.6. *Let* $\mathrm{X}, \mathrm{X}_1, \mathrm{X}_2$ *and* Y *be possible properties of tournaments on* $[m]$*. Then, we have:*

(i) If $G \in \overline{\mathcal{G}}_m$ *fulfills* $G \in \overline{\mathcal{G}}_m(\mathrm{Y})$ *and has a subgraph* $\tilde{G} \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$*, then* $G \in \overline{\mathcal{G}}_m(\mathrm{X})$*.*

(ii) $\mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y}) \cap \overline{\mathcal{G}}_m = \overline{\mathcal{G}}_m(\mathrm{X} \wedge \mathrm{Y})$*.*

(iii) $\mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \subseteq \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$ *iff* $\overline{\mathcal{G}}_m(X_1 \wedge \mathrm{Y}) \subseteq \overline{\mathcal{G}}_m(\mathrm{X}_2 \wedge \mathrm{Y})$*.*

(iv) If $G \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$ *and* $G' \in \mathcal{G}_m(\mathrm{Y})$ *with* $E_G \subseteq E_{G'}$*, then* $G' \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$*.*

(v) $\mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y}) \subseteq \mathcal{G}_m \setminus \mathcal{G}_m(\neg \mathrm{Y})$ *with equality iff* $\overline{\mathcal{G}}_m(\mathrm{X}) = \overline{\mathcal{G}}_m(\mathrm{Y})$*.*

(vi) $\mathcal{G}_m(\mathrm{X}_1 \wedge \mathrm{X}_2 \,|\, \mathrm{Y}) = \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cap \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$*.*

(vii) $\mathcal{G}_m(\mathrm{X}_1 \vee \mathrm{X}_2 \,|\, \mathrm{Y}) \supseteq \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cup \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$*.*

(viii) If $\overline{\mathcal{G}}_m(\mathrm{X}_1) \cap \overline{\mathcal{G}}_m(\mathrm{X}_2) = \emptyset$*, then* $\mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cap \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y}) = \emptyset$*.*

(ix) If $\overline{\mathcal{G}}_m(\mathrm{Y}) \subseteq \overline{\mathcal{G}}_m(\mathrm{X})$*, then* $\mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y}) = \{G \in \mathcal{G}_m \,|\, \exists\, \mathrm{Y}\text{-extension of } G\}$*.*

Under the assumptions of Lem. 3.6, we do not necessarily have $\mathcal{G}_m(\mathrm{X}_1 \vee \mathrm{X}_2 \,|\, \mathrm{Y}) \subseteq \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cup \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$. A look at the definitions reveals that any $G \in \mathcal{G}_m$ fulfills $G \in \mathcal{G}_m(\mathrm{X}_1 \vee \mathrm{X}_2 \,|\, \mathrm{Y}) \setminus (\mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cup \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y}))$ iff

$$\begin{aligned} \forall\, \mathrm{Y}\text{-extensions } G' \text{ of } G &: G' \in \overline{\mathcal{G}}_m(\mathrm{X}_1 \vee \mathrm{X}_2), \\ \exists\, \mathrm{Y}\text{-extension } G_1 \text{ of } G &: G_1 \notin \overline{\mathcal{G}}_m(\mathrm{X}_1), \\ \exists\, \mathrm{Y}\text{-extension } G_2 \text{ of } G &: G_2 \notin \overline{\mathcal{G}}_m(\mathrm{X}_2). \end{aligned}$$

To illustrate this, suppose $m = 3$, $\mathrm{Y} = \emptyset$, $\mathrm{X}_1 = 1$, $\mathrm{X}_2 = 2$ and $G := ([3], \{(1,3),(2,3)\})$. The set of $\emptyset$-extensions of G consists of the following two graphs G_1 and G_2:

Due to $G_1 \in \overline{\mathcal{G}}_3(1) \setminus \overline{\mathcal{G}}_3(2)$ and $G_2 \in \overline{\mathcal{G}}_3(2) \setminus \overline{\mathcal{G}}_3(1)$ the above conditions are fulfilled and we have $G \in \mathcal{G}_3(1 \vee 2) \setminus (\mathcal{G}_3(1) \cup \mathcal{G}_3(2))$. By Prop. 3.7 stated below, we similarly see that for arbitrary $m \geq 3$

$$\mathcal{G}_m(\mathrm{CW}) = \mathcal{G}_m(1 \vee \cdots \vee m) \subsetneq \bigcup\nolimits_{i \in [m]} \mathcal{G}_m(i).$$

3.4. CW in Extension

In this section, we characterize the sets $\mathcal{G}_m(\mathrm{CW})$, $\mathcal{G}_m(\neg\mathrm{CW})$, $\mathcal{G}_m(i)$, $\mathcal{G}_m(\neg i)$, $\mathcal{G}_m(i \,|\, \mathrm{CW})$ and $\mathcal{G}_m(\neg i \,|\, \mathrm{CW})$ as well as particular combinations of these. The insights obtained will be of value for solving the CW-related problems in the dueling bandits scenario. Moreover, they will be of use in Sec. 3.5, where we discuss solutions to $\mathcal{D}^m_{\mathrm{CWt}}$, $\mathcal{D}^m_{\mathrm{CWc}}$, $\mathcal{D}^m_{\mathrm{CWv}}$, $\mathcal{D}^m_{\mathrm{CWi}}(\mathrm{CW})$ as well as $\mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$. We start with a characterization of $\mathcal{G}_m(\mathrm{CW})$.

Proposition 3.7. *For* $G \in \mathcal{G}_m$ *we have* $G \in \mathcal{G}_m(\mathrm{CW})$ *iff there exists some* $i_0, i_1 \in [m]$ *such that* G *contains at least one of the subgraphs*

In other words, $G \in \mathcal{G}_m(\mathrm{CW})$ *is fulfilled iff at least one of the following holds:*

(a) There exists $i_0 \in [m]$ *such that* $i_0 \xrightarrow{G} j$ *holds for every* $j \in [m] \setminus \{i_0\}$*.*

(b) There exist distinct $i_0, i_1 \in [m]$ *with* $i_0 \xrightarrow{G} j$ *and* $i_1 \xrightarrow{G} j$ *for every* $j \in [m] \setminus \{i_0, i_1\}$*.*

In particular, $|E_G| \geq m - 1$ *holds for every* $G \in \mathcal{G}_m(\mathrm{CW})$*.*

In order to prove Prop. 3.7, we at first need some prerequisites. Given some $G \in \overline{\mathcal{G}}_m$ and distinct $i, j \in [m]$ we define $G_{i\leftrightarrow j} \in \overline{\mathcal{G}}_m$ to be the tournament in which the edge between i and j is reversed (in comparison to G) and all the other edges are the same, i.e., if $(i, j) \in E_G$ then

$$E_{G_{i\leftrightarrow j}} = (E_G \setminus \{(i, j)\}) \cup \{(j, i)\}.$$

Note in particular that $G_{i\leftrightarrow j}$ is *not* the graph G with nodes i and j interchanged. For example, G_1 and G_2 from p. 69 fulfill $G_1 = (G_2)_{1\leftrightarrow 2}$.

Lemma 3.8. *If* $G \in \overline{\mathcal{G}}_m(\mathrm{CW})$*,* $i_0 := \mathrm{CW}(G)$ *and* $i_1 \in [m] \setminus \{i_0\}$ *are such that* $G' := G_{i_0\leftrightarrow i_1} \in \overline{\mathcal{G}}_m(\mathrm{CW})$*, then* $i_1 = \mathrm{CW}(G')$ *holds.*

Proof of Lem. 3.8. For every $j \in [m]\setminus\{i_0, i_1\}$ we can infer from $i_0 = \mathrm{CW}(G)$ that $i_0 \xrightarrow{G} j$ and thus also $i_0 \xrightarrow{G'} j$ hold. Together with $i_1 \xrightarrow{G'} i_0$ this shows $\mathrm{CW}(G') \notin [m] \setminus \{i_1\}$, and thus further $\mathrm{CW}(G') = i_1$. □

With this, we are able to prove Prop. 3.7.

Proof of Prop. 3.7. To show "⇒" indirectly, assume that there was some $G \in \mathcal{G}_m(\mathrm{CW})$ such that neither (a) nor (b) holds. Choose an arbitrary extension $G_0 \in \overline{\mathcal{G}}_m$ of G and note that $G \in \mathcal{G}_m(\mathrm{CW})$ implies that $i_0 := \mathrm{CW}(G_0)$ is well-defined. As (a) does not hold, there exists some $i_1 \in [m] \setminus \{i_0\}$ with $\neg(i_0 \xrightarrow{G} i_1)$. Moreover, the definition of i_0 ensures[2] $\neg(i_1 \xrightarrow{G} i_0)$. Consequently, $G_1 := (G_0)_{i_0\leftrightarrow i_1} \in \overline{\mathcal{G}}_m$ is also an extension of G, and by assumption on G we have $G_1 \in \overline{\mathcal{G}}_m(\mathrm{CW})$. Thus, we can infer $i_1 = \mathrm{CW}(G_1)$ from Lem. 3.8. Since (b) does not hold, there exist $b \in \{0, 1\}$ and $k \in [m] \setminus \{i_0, i_1\}$ such that $\neg(i_b \xrightarrow{G} k)$ holds. From $i_b \xrightarrow{G_b} k$ we can infer $\neg(k \xrightarrow{G} i_b)$. As we have seen that G does neither contain any edge between i_b and k nor between i_{1-b} and i_b and G_{1-b} is an extension of G, also the graph

$$G' := \begin{cases} (G_{1-b})_{i_{1-b}\leftrightarrow i_b} = G_b, & \text{if } k \xrightarrow{G_{1-b}} i_b, \\ \left((G_{1-b})_{i_{1-b}\leftrightarrow i_b}\right)_{i_b\leftrightarrow k}, & \text{if } i_b \xrightarrow{G_{1-b}} k, \end{cases}$$

is an extension of G. Due to $G \in \mathcal{G}_m(\mathrm{CW})$ we obtain $G' \in \overline{\mathcal{G}}_m(\mathrm{CW})$, hence Lem. 3.8 guarantees $\mathrm{CW}(G') \in \{i_b, i_{1-b}, k\}$. This is a contradiction, since G' contains by its definition the edges $i_{1-b} \to k$, $k \to i_b$ and $i_b \to i_{1-b}$.

It remains to show "⇐". For this, suppose $G \in \mathcal{G}_m$ to be such that (a) or (b) holds and let $G' \in \overline{\mathcal{G}}_m$ be an arbitrary extension of G. In case (a) holds, we obtain for each

[2] In fact, assuming $i_1 \to i_0$ in G would also imply $i_1 \to i_0$ in G_0, which is according to $i_0 = \mathrm{CW}(G_0)$ not possible.

$j \in [m] \setminus \{i_0\}$ due to $i_0 \xrightarrow{G} j$ that $i_0 \xrightarrow{G'} j$ is fulfilled. Thus, we can infer $\mathrm{CW}(G') = i_0$ and in particular $G' \in \overline{\mathcal{G}}_m(\mathrm{CW})$. In case (b) holds, G' contains all the edges $i_0 \to j$, $i_1 \to j$, $j \in [m] \setminus \{i_0, i_1\}$. Moreover, for one $b \in \{0,1\}$ we have $i_b \xrightarrow{G'} i_{1-b}$ and we obtain $i_b = \mathrm{CW}(G')$, i.e., $G' \in \overline{\mathcal{G}}_m(\mathrm{CW})$. ☐

Proposition 3.9. *For $G \in \mathcal{G}_m$ we have the equivalence*

$$G \in \mathcal{G}_m(\neg\mathrm{CW}) \quad \Leftrightarrow \quad \forall\, i \in [m]\, \exists\, j \in [m] \setminus \{i\} : j \xrightarrow{G} i.$$

In particular, $|E_G| \geq m$ holds for every $G \in \mathcal{G}_m(\neg\mathrm{CW})$.

Proof of Prop. 3.9. Let $G \in \mathcal{G}_m$ be fixed. To see "$\Leftarrow$" suppose that there is for all $i \in [m]$ some $j = j(i) \in [m] \setminus \{i\}$ with $j \xrightarrow{G} i$ and let G' be an arbitrary extension of G. Then, for any $i \in [m]$, $j(i) \xrightarrow{G'} i$ shows that i cannot be the Condorcet winner of G'. We infer $G' \in \overline{\mathcal{G}}_m(\neg\mathrm{CW})$, and arbitrariness of G' lets us conclude $G \in \mathcal{G}_m(\neg\mathrm{CW})$.

To show "$\Rightarrow$" we prove its contraposition. Thus, let us suppose there exists some $i \in [m]$ such that $\neg(j \xrightarrow{G} i)$ holds for every $j \in [m] \setminus \{i\}$. Then, we can choose an extension G' of G with $i \xrightarrow{G'} j$ for every $j \in [m] \setminus \{i\}$. Thus, $G' \in \overline{\mathcal{G}}_m(\mathrm{CW})$ holds with $i = \mathrm{CW}(G')$, which implies $G \notin \mathcal{G}_m(\neg\mathrm{CW})$. ☐

Proposition 3.10. *If $G \in \mathcal{G}_m$ and $i^* \in [m]$, then*

$$G \in \mathcal{G}_m(i^*) \Leftrightarrow \forall j \in [m] \setminus \{i^*\} : i^* \xrightarrow{G} j \quad \text{and} \quad G \in \mathcal{G}_m(\neg i^*) \Leftrightarrow \exists j \in [m] : j \xrightarrow{G} i^*.$$

In particular, $|E_G| \geq m-1$ holds for every $G \in \mathcal{G}_m(i^*)$, and $|E_G| \geq 1$ holds for every $G \in \mathcal{G}_m(\neg i^*)$.

Proof. Let $G \in \mathcal{G}_m$ be fixed. For showing "$\Rightarrow$" indirectly suppose there was some $j \in [m] \setminus \{i^*\}$ with $\neg(i^* \xrightarrow{G} j)$. Then, there exists an extension $G' \in \overline{\mathcal{G}}_m$ of G with $j \xrightarrow{G'} i^*$, which is trivially not in $\overline{\mathcal{G}}_m(i^*)$. Thus, we would obtain that $G \notin \mathcal{G}_m(i^*)$, which is a contradiction.

In order to see "$\Leftarrow$" suppose on the contrary $G \notin \mathcal{G}_m(i^*)$. Then, there exists some extension $G' \in \overline{\mathcal{G}}_m$ of G with $G' \notin \overline{\mathcal{G}}_m(i^*)$. Now, $i^* \neq \mathrm{CW}(G')$ implies the existence of some $j \in [m]$ with $j \overset{G'}{\to} i^*$, and as G' is an extension of G this shows $\neg(i^* \xrightarrow{G} j)$. ☐

For identifying the CW of a tournament $G \in \overline{\mathcal{G}}_m$ under the assumption that it exists, the sets $\mathcal{G}_m(i \,|\, \mathrm{CW})$ and $\mathcal{G}_m(\neg i \,|\, \mathrm{CW})$ are of interest. These are characterized in the following.

Proposition 3.11. *We have*

$$\begin{aligned} \mathcal{G}_m(i \,|\, \mathrm{CW}) &= \left\{ G \in \mathcal{G}_m \,|\, \forall l \neq i \,:\, \neg(l \overset{G}{\to} i) \text{ and } \forall j \neq i\, \exists k \neq j : k \overset{G}{\to} j \right\} \\ &= \bigcap\nolimits_{j \neq i} \mathcal{G}_m(\neg j) \setminus \mathcal{G}_m(\neg\mathrm{CW}) \end{aligned}$$

and

$$\mathcal{G}_m(\neg i \,|\, \mathrm{CW}) = \mathcal{G}_m(\neg i) \setminus \mathcal{G}_m(\neg\mathrm{CW}).$$

Proof. At first, suppose $G \in \mathcal{G}_m(i \,|\, \mathrm{CW})$. Then, there is a CW-extension G' of G with $G' \in \mathcal{G}_m(i)$, i.e., $i \overset{G'}{\to} l$ for every $l \neq i$ and in particular $G \notin \mathcal{G}_m(\neg\mathrm{CW})$. Since G is a subgraph of G', we thus have $\forall l \neq i : \neg(l \overset{G}{\to} i)$. Moreover, assuming $\exists j \neq i \forall k \neq j : \neg(k \overset{G}{\to} j)$ would imply the existence of some extension G'' of G with $G'' \in \overline{\mathcal{G}}_m(j) \subseteq \mathcal{G}_m(j \,|\, \mathrm{CW})$, which is due to $\overline{\mathcal{G}}_m(i) \cap \overline{\mathcal{G}}_m(j) = \emptyset$ not in $\mathcal{G}_m(i \,|\, \mathrm{CW})$. Thus, $\forall j \neq i \exists k \neq j : k \overset{G}{\to} j$ holds, which directly implies $G \in \bigcap_{j \neq i} \mathcal{G}_m(\neg j)$. This proves "$\subseteq$" in both equalities of the first statement.
To see "$\supseteq$", fix $G \in \bigcap_{j \neq i} \mathcal{G}_m(\neg j) \setminus \mathcal{G}_m(\neg\mathrm{CW})$. For each $j \neq i$, we infer from $G \in \mathcal{G}_m(\neg j)$ that some $k \neq j$ exists with $k \overset{G}{\to} j$. Furthermore, $G \notin \mathcal{G}_m(\neg\mathrm{CW})$ implies indirectly $\forall l \neq i : \neg(l \overset{G}{\to} i)$. In particular, there exists some extension G' of G with $G' \in \overline{\mathcal{G}}_m(i)$. Due to $\forall j \neq i \exists k \neq j : k \overset{G}{\to} j$, each CW-extension of G is in $\overline{\mathcal{G}}_m \setminus (\bigcup_{j \neq i} \overline{\mathcal{G}}_m(j)) = \overline{\mathcal{G}}_m(i)$. Thus, $G \in \mathcal{G}_m(i \,|\, \mathrm{CW})$ holds, which concludes the proof of the first statement.
The second statement is a consequence of the equivalences

$$\begin{aligned} G \in \mathcal{G}_m(\neg i \,|\, \mathrm{CW}) &\Leftrightarrow \exists \text{ CW-extension of } G \text{ and all CW-extensions of } G \text{ are in } \overline{\mathcal{G}}_m(\neg i) \\ &\Leftrightarrow \exists \text{ CW-extension of } G \text{ and } \exists j \neq i : j \overset{G}{\to} i \\ &\Leftrightarrow G \in \mathcal{G}_m(\neg i) \text{ and } G \notin \mathcal{G}_m(\neg\mathrm{CW}) \\ &\Leftrightarrow G \in \mathcal{G}_m(\neg i) \setminus \mathcal{G}_m(\neg\mathrm{CW}), \end{aligned}$$

which hold for any $G \in \mathcal{G}_m$. □

We finish this section with the following lemma, which will be crucial for the proofs of Theorems 4.6 and 4.12 below. It allows us to project graphs in $\mathcal{G}_m(\mathrm{CW})$, $\mathcal{G}_m(\neg\mathrm{CW})$, $\mathcal{G}_m(i^*)$, $\mathcal{G}_m(\neg i^*)$ as well as in

$$\mathcal{G}_m(\Delta i^*) \coloneqq \mathcal{G}_m(\neg\mathrm{CW}) \cup \bigcup_{j \in [m] : j \neq i^*} \mathcal{G}_m(j) \quad \text{and} \quad \mathcal{G}_m(\diamondsuit) \coloneqq \bigcup_{i \in [m]} \mathcal{G}_m(i)$$

to characteristic subgraphs, respectively. Note here that $\mathcal{G}_m(\Delta i^*) \neq \mathcal{G}_m(\neg i^*)$, as for instance the graph $([m], \{(2,1)\})$ is contained in $\mathcal{G}_m(\neg 1)$ but not in $\mathcal{G}_m(\Delta 1)$. According to Prop. 3.7, $\mathcal{G}_m(\mathrm{CW}) \supsetneq \mathcal{G}_m(\diamondsuit)$, hence these notions are not redundant.

Lemma 3.12. *Let $i^* \in [m]$. There exist mappings $\mathfrak{l}_{\mathrm{CW}}, \mathfrak{l}_{\neg\mathrm{CW}}, \mathfrak{l}_{i^*}, \mathfrak{l}_{\neg i^*}, \mathfrak{l}_{\Delta i^*}, \mathfrak{l}_{\diamondsuit} : \mathcal{G}_m \to \mathcal{G}_m$ with the following properties:*

(a) $E_{\mathfrak{l}_{\mathrm{CW}}(G)}, E_{\mathfrak{l}_{\neg\mathrm{CW}}(G)}, E_{\mathfrak{l}_{i^*}(G)}, E_{\mathfrak{l}_{\neg i^*}(G)}, E_{\mathfrak{l}_{\Delta i^*}(G)}, E_{\mathfrak{l}_{\diamondsuit}(G)} \subseteq E_G$ *for every* $G \in \mathcal{G}_m$,

(b) for every $G \in \mathcal{G}_m$ *we have* $|E_{\mathfrak{l}_{\mathrm{CW}}(G)}| \in \{0, m-1, 2m-4\}$, $|E_{\mathfrak{l}_{\neg\mathrm{CW}}(G)}| \in \{0, m\}$, $|E_{\mathfrak{l}_{i^*}(G)}| \in \{0, m-1\}$, $|E_{\mathfrak{l}_{\neg i^*}(G)}| \in \{0, 1\}$, $|E_{\mathfrak{l}_{\Delta i^*}(G)}| \in \{0, m-1, m\}$ *as well as* $|E_{\mathfrak{l}_{\diamondsuit}(G)}| \in \{0, m-1\}$,

(c) for every $G \in \mathcal{G}_m$ *we have the equivalences*

$$\begin{gathered} G \in \mathcal{G}_m(\mathrm{CW}) \Leftrightarrow \mathfrak{l}_{\mathrm{CW}}(G) \in \mathcal{G}_m(\mathrm{CW}), \quad G \in \mathcal{G}_m(\neg\mathrm{CW}) \Leftrightarrow \mathfrak{l}_{\neg\mathrm{CW}}(G) \in \mathcal{G}_m(\neg\mathrm{CW}), \\ G \in \mathcal{G}_m(i^*) \Leftrightarrow \mathfrak{l}_{i^*}(G) \in \mathcal{G}_m(i^*), \quad G \in \mathcal{G}_m(\neg i^*) \Leftrightarrow \mathfrak{l}_{\neg i^*}(G) \in \mathcal{G}_m(\neg i^*), \\ G \in \mathcal{G}_m(\Delta i^*) \Leftrightarrow \mathfrak{l}_{\Delta i^*}(G) \in \mathcal{G}_m(\Delta i^*), \quad G \in \mathcal{G}_m(\diamondsuit) \Leftrightarrow \mathfrak{l}_{\diamondsuit}(G) \in \mathcal{G}_m(\diamondsuit). \end{gathered}$$

Proof. To define $\mathfrak{l}_{\mathrm{CW}}$ suppose $G \in \mathcal{G}_m$ to be fixed for the moment. In case $G \notin \mathcal{G}_m(\mathrm{CW})$ we may simply define $\mathfrak{l}_{\mathrm{CW}}(G) := ([m], \emptyset)$, and in case $G \in \mathcal{G}_m(\mathrm{CW})$ there exist according to Prop. 3.7 two distinct $i_0, i_1 \in [m]$ such that at least one of

$$E[i_0] := \{(i_0, j) \,:\, j \in [m] \setminus \{i_0\}\}$$

and

$$E[i_0; i_1] := \{(i_0, j) \,:\, j \in [m] \setminus \{i_0, i_1\}\} \cup \{(i_1, j) \,:\, j \in [m] \setminus \{i_0, i_1\}\}$$

is a subset of E_G, i.e., we may define

$$\mathfrak{l}_{\mathrm{CW}}(G) := \begin{cases} ([m], E[i_0]), & \text{if } E[i_0] \subset E_G, \\ ([m], E[i_0; i_1]), & \text{otherwise.} \end{cases}$$

It is straightforward to check that $\mathfrak{l}_{\mathrm{CW}}$ fulfills all the desired properties.

The existence of $\mathfrak{l}_{\neg\mathrm{CW}}$, $\mathfrak{l}_{i^*}$ and $\mathfrak{l}_{\neg i^*}$ follow from Prop. 3.9 and Prop. 3.10.

For defining $\mathfrak{l}_{\Delta i^*}$ let $G \in \mathcal{G}_m$ be given. In case $G \notin \mathcal{G}_m(\Delta i^*)$ we define $\mathfrak{l}_{\Delta i^*}(G) := ([m], \emptyset)$. In the remaining case $G \in \mathcal{G}_m(\Delta i^*)$ we choose

$$\mathfrak{l}_{\Delta i^*}(G) := \begin{cases} \mathfrak{l}_j(G), & \text{if } \exists j \in [m] \setminus \{i^*\} \text{ with } G \in \mathcal{G}_m(j), \\ \mathfrak{l}_{\neg\mathrm{CW}}(G), & \text{otherwise.} \end{cases}$$

Note that this is due to $\mathcal{G}_m(j) \cap \mathcal{G}_m(j') = \emptyset$ for $j \neq j'$ well-defined. Then, $\mathfrak{l}_{\Delta i^*}$ has all the properties stated above.

Finally, we define $\mathfrak{l}_{\diamondsuit}$ via

$$\mathfrak{l}_{\diamondsuit}(G) := \begin{cases} \mathfrak{l}_j(G), & \text{if } \exists j \in [m] \text{ with } G \in \mathcal{G}_m(j), \\ ([m], \emptyset), & \text{otherwise.} \end{cases}$$

□

3.5. DSTAs for CW-related Problems

Next, let us discuss DSTAs for $\mathcal{D}^m_{\mathrm{CWt}}$, $\mathcal{D}^m_{\mathrm{CWc}}$, $\mathcal{D}^m_{\mathrm{CWv}}$, $\mathcal{D}^m_{\mathrm{CWi}}(\mathrm{CW})$ and $\mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$. Recall that, for any of these problems, Prop. 3.4 provides us necessary conditions on solutions to these problems, e.g. if a DSTA $\mathcal{A}$ solves $\mathcal{D}^m_{\mathrm{CWc}} = \mathcal{D}^m_{\mathrm{CW},\neg\mathrm{CW}}$, then

$$\mathfrak{G}^{\mathcal{A}}_G\left(T^{\mathcal{A}}_G\right) \in \mathcal{G}_m(\mathrm{CW}) \quad \text{for all } G \in \overline{\mathcal{G}}(\mathrm{CW})$$

and

$$\mathfrak{G}^{\mathcal{A}}_G\left(T^{\mathcal{A}}_G\right) \in \mathcal{G}_m(\neg\mathrm{CW}) \quad \text{for all } G \in \overline{\mathcal{G}}_m(\neg\mathrm{CW}).$$

Together with the characterizations of the sets $\mathcal{G}_m(\mathrm{CW})$, $\mathcal{G}_m(\neg\mathrm{CW})$, $\mathcal{G}_m(i)$, $\mathcal{G}_m(\neg i)$, $\mathcal{G}_m(i \,|\, \mathrm{CW})$ and $\mathcal{G}_m(\neg i \,|\, \mathrm{CW})$ from Sec. 3.4 this will allow us to provide lower query complexity bounds for solutions to these problems.

Proposition 3.13. *(i) Alg. 9 is an optimal solution to $\mathcal{D}^m_{\mathrm{CWv}}$ with worst- and best-case termination time $m-1$.*

Algorithm 9 An optimal DSTA for $\mathcal{D}^m_{\mathrm{CWv}}$ and $\mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$

Input: $i \in [m]$
Initialization: $W \leftarrow [m] \setminus \{i\}$
$\triangleright$ W = set of nodes $j \in [m]$, which have not yet been compared to i

while $|W| \geq 1$ **do**
  Choose an arbitrary $j \in W$
  $(i', j') \leftarrow \{i, j\}_G$ $\triangleright$ Query $\{i, j\}$
  if $(i, j') = (j, i)$ **then return** $\neg i$ $\triangleright$ i cannot be the CW
  else $W \leftarrow W \setminus \{j\}$ $\triangleright$ j has been compared to i
return i

(ii) Alg. 9 is an optimal solution to $\mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$ *with worst- and best-case termination time* $m-1$.

Proof. It is easy to check that $\mathcal{A}$:=Alg. 9 solves $\mathcal{D}^m_{\mathrm{CWv}}$ and makes exactly $m-1$ queries when started with any input $i \in [m]$ on any $G \in \overline{\mathcal{G}}_m$, hence it has worst-case termination time $m-1$. In particular, it is also a solution to the less difficult task $\mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$ with worst-case termination time $m-1$ and has – for both $\mathcal{D}^m_{\mathrm{CWv}}$ and $\mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$ – a best-case termination time of $m-1$.

To prove the lower bounds, let $\mathcal{A}$ be any solution to $\mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$ and fix arbitrary $i \in [m]$ and $G \in \overline{\mathcal{G}}_m(i)$. By Prop. 3.4 we have $\mathfrak{G}^{\mathcal{A}(i)}_G\left(T^{\mathcal{A}(i)}_G\right) \in \mathcal{G}_m(i \,|\, \mathrm{CW})$, hence Prop. 3.11 assures that for each $j \neq i$ there exists $k \neq j$ such that $k \to j$ in $\mathfrak{G}^{\mathcal{A}(i)}_G\left(T^{\mathcal{A}(i)}_G\right)$, i.e., $\mathfrak{G}^{\mathcal{A}(i)}_G\left(T^{\mathcal{A}(i)}_G\right)$ has $\geq m-1$ edges. Thus, $T^{\mathcal{A}} \geq T^{\mathcal{A}(i)}_G \geq m-1$ holds. Since any solution to $\mathcal{D}^m_{\mathrm{CWv}}$ also solves $\mathcal{D}^m_{\mathrm{CWv}}(\mathrm{CW})$, an optimal solution to $\mathcal{D}^m_{\mathrm{CWv}}$ has worst-case termination time $\geq m-1$. Consequently, $\mathcal{A}$ is optimal for both problems. $\square$

Algorithm 10 An optimal DSTA for $\mathcal{D}^m_{\mathrm{CWt}}$ [Procaccia, 2008]

Initialization: Construct an *almost complete* binary tree T of height $D := \lceil \log m \rceil$ with m leaves, which are labeled by $1, \ldots, m$. Here, *almost complete* means that there are exactly 2^d nodes on each level $d \leq D-1$.

while height$(T) > 0$ **do**
  Pick two sibling leave nodes $i, j \in [m]$ of T and compare them
  if $\{i, j\}_G = (i, j)$ **then** $\triangleright$ j cannot be the CW
    Label the unique parent of i and j with i, then remove its children from T
  else $\triangleright$ i cannot be the CW
    Label the unique parent of i and j with j, then remove its children from T
Let i^* be the label of the only node in T
Compare i^* with all other alternatives, with which it has *not* been compared yet
if i^* has won all of its duels **then return** i^*
else return $\neg\mathrm{CW}$

Proposition 3.14. *Alg. 10 is an optimal solution to* $\mathcal{D}^m_{\mathrm{CWt}}$ *with worst- and best-case termination time* $2m - \lfloor \log_2 m \rfloor - 2$.

Proof. Cf. Thm. 2.1 in [Procaccia, 2008] and Lem. 3.2 in [Balasubramanian et al., 1997]. ☐

Algorithm 11 An optimal DSTA for $\mathcal{D}^m_{\text{CWi}}(\text{CW})$

Initialization: $S \leftarrow [m]$, $i \leftarrow 1$ ▷ S = set of candidates for CW
▷ i = the current candidate

while $|S| > 1$ **do**
  Choose an arbitrary $j \in S \setminus \{i\}$
  $(i', j') \leftarrow \{i, j\}_G$ ▷ Query $\{i, j\}$
  if $(i, j') = (i, j)$ **then** $S \leftarrow S \setminus \{j\}$ ▷ j cannot be the CW
  else $S \leftarrow S \setminus \{i\}$, $i \leftarrow j$ ▷ i is not the CW, j is the new candidate
return i ▷ $S = \{i\}$

Proposition 3.15. *Alg. 11 is an optimal solution to* $\mathcal{D}^m_{\text{CWi}}(\text{CW})$ *and has worst- and best-case termination time* $m - 1$.

Proof. Write $\mathcal{A}$:=Alg. 11 for the moment. In Alg. 11, $|S|$ decreases by 1 in each iteration of the while loop, hence $\mathcal{A}$ terminates after exactly $m - 1$ time steps when started on any $G \in \overline{\mathcal{G}}_m(\text{CW})$. Hence, $T^{\mathcal{A}}_{\text{worst}} = T^{\mathcal{A}}_{\text{best}} = m - 1$ holds. For fixed $G \in \overline{\mathcal{G}}_m(\text{CW})$, $\mathcal{A}$ returns at termination the only remaining element i of S. The construction of S assures $\text{CW}(G) \neq j$ for each $j \in [m] \setminus S$, i.e., $\mathbf{D}^{\mathcal{A}}(G) = \text{CW}(G)$ has to be fulfilled. Hence, $\mathcal{A}$ solves $\mathcal{D}^m_{\text{CWi}}(\text{CW})$.
To see the lower bound, suppose $\mathcal{A}$ to be any solution to $\mathcal{D}^m_{\text{CWi}}(\text{CW})$ and fix $G \in \overline{\mathcal{G}}_m(\text{CW})$ and $i := \text{CW}(G)$. According to Prop. 3.4 we have $\mathfrak{G}^{\mathcal{A}(i)}_G\left(T^{\mathcal{A}(i)}_G\right) \in \mathcal{G}_m(i \,|\, \text{CW})$. Therefore, Prop. 3.11 guarantees that for each $j \neq i$ some $k \neq j$ exists with $k \to j$ in $\mathfrak{G}^{\mathcal{A}(i)}_G\left(T^{\mathcal{A}(i)}_G\right)$, i.e., $\mathfrak{G}^{\mathcal{A}(i)}_G\left(T^{\mathcal{A}(i)}_G\right)$ has $\geq m - 1$ edges. Thus, $T^{\mathcal{A}} \geq T^{\mathcal{A}(i)}_G \geq m - 1$ holds. Consequently, the worst-case termination time of $\mathcal{A}$ is at least $m - 1$. ☐

Algorithm 12 An optimal DSTA for $\mathcal{D}^m_{\text{CWc}}$

Simulate Alg. 10 until it terminates, let d be its return value
if $d = \neg\text{CW}$ **then return** $\neg CW$
else return CW

Proposition 3.16. *Alg. 12 is an optimal solution to* $\mathcal{D}^m_{\text{CWc}}$ *and has worst- and best-case termination time* $2m - \lfloor \log_2 m \rfloor - 2$.

Proof. Since Alg. 10 solves $\mathcal{D}^m_{\text{CWt}}$, Alg. 12 solves $\mathcal{D}^m_{\text{CWc}}$. Moreover, they have the same worst-and best-case termination time, which is $2m - \lfloor \log_2 m \rfloor - 2$. For the lower bound confer e.g. Lem. 3.2 in [Balasubramanian et al., 1997]. ☐

3.6. Acyclicity in Extension

This section is dedicated to discuss the property *acyclicity in extension*, i.e., the set $\mathcal{G}_m(\text{acyclic})$. Even though this notion is only defined by means of acyclicity of tournament graphs, acyclicity itself is applicable to *any* digraph $G \in \mathcal{G}_m$. In fact, we can show the following result.

Lemma 3.17. *Any $G \in \mathcal{G}_m$ is acyclic in extension iff every supergraph $\tilde{G} \in \mathcal{G}_m$ of it is a acyclic.*

Proof. Obviously, if any supergrah $\tilde{G} \in \mathcal{G}_m$ of G is acyclic, G is acyclic in extension. On the other side, if any of the supergraphs $\tilde{G} \in \mathcal{G}_m$ of G was not acyclic, any arbitrary supergraph $\tilde{G}' \in \overline{\mathcal{G}}_m$ of $\tilde{G}$ would be a non-acyclic supergraph of G and hence G would not be acyclic in extension. □

Let us fix some further notation. For $G \in \mathcal{G}_m$, we denote for convenience, with a slight abuse of notation, by E_G also the set of all $\{i,j\} \in [m]_2$ with $(i,j) \in E_G$ or $(j,i) \in E_G$. This way, "$\{i,j\} \in E_G$" means "$(i,j) \in E_G$ or $(j,i) \in E_G$", "$\{i,j\} \notin E_G$" means "$(i,j) \notin E_G$ and $(j,i) \notin E_G$", and thus the set $[m]_2 \setminus E_G$ is given as $\{\{i,j\} \in [m]_2 \,|\, (i,j) \notin E_G \text{ and } (j,i) \notin E_G\}$. Before characterizing the set $\mathcal{G}_m(\text{acyclic})$, we prove a necessary condition for $G \in \mathcal{G}_m(\text{acyclic})$.

Proposition 3.18. *(i) For every $G \in \mathcal{G}_m(\text{acyclic})$ we have $|E_G| \geq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$.*

(ii) For any $m \geq 3$ there exists $G \in \mathcal{G}_m(\text{acyclic})$ with $|E_G| = \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$

Proof. (i) Let $G \in \mathcal{G}_m(\text{acyclic})$ with $|E_G| \leq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor - 1$. Then

$$K \coloneqq \{\{i,j\} \in [m]_2 \,|\, (i,j),(j,i) \notin E_G\}$$

fulfills $|K| = \binom{m}{2} - |E_G| \geq \lfloor \frac{m}{2} \rfloor + 1$. Thus, there exist distinct $i,j,k \in [m]$ with $\{i,j\},\{j,k\} \in K$. In the case $(i,k) \in E_G$ we may define $\tilde{G} \in \overline{\mathcal{G}}_m$ with $E_G \subseteq E_{\tilde{G}}$ and $(j,i),(k,j) \in E_G$, hence $\tilde{G}$ would contain the cycle $i \to k \to j \to i$, and thus $G \notin \mathcal{G}_m(\text{acyclic})$ follows. In the remaining case $(i,k) \notin E_G$ we can similarly construct $\tilde{G} \in \overline{\mathcal{G}}_m$ fulfilling $E_G \subseteq E_{\tilde{G}}$ as well as $(k,i),(i,j),(j,k) \in E_G$, which also implies $G \notin \mathcal{G}_m(\text{acyclic})$.

(ii) If $m = 3$, choose $G \coloneqq ([m],(1,2),(2,3))$, and for $m = 5$ and $m = 6$ appropriate choices of G are depicted below this proof. We proceed with a formal proof for the case $m \geq 4$.

First, suppose $m \geq 4$ to be even. Define $G = ([m], E_G) \in \mathcal{G}_m$ via

$$(i,j) \in E_G \quad \text{iff} \quad i < j \text{ and } (i,j) \notin K$$

with $K \coloneqq \{(1,2),(3,4),\ldots,(m-1,m)\}$. Due to $|K| = \lfloor \frac{m}{2} \rfloor$, G fulfills $|E_G| = \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$, it remains to show $G \in \mathcal{G}_m(\text{acyclic})$. For this, let $G' \in \overline{\mathcal{G}}_m$ be an arbitrary extension of G. For proving acyclicity of G' it is according to Lem. 3.2 sufficient to show that for any distinct $i,j,k \in [m]$ with $(i,j),(j,k) \in E_{G'}$ also $(i,k) \in E_{G'}$ holds. Let such $i,j,k \in [m]$ be fixed in the following. By its definition, K contains at most one of the tuples $(i,j),(j,i),(j,k),(k,j),(i,k),(k,i)$.

Case 1: $(i,j),(j,i),(i,k),(k,i) \notin K$ and $((j,k) \in K$ or $(k,j) \in K)$.
Then, $(i,j) \in E_{G'} \cap E_G$ implies $i < j$. Since $k \in \{j-1, j+1\}$ has to be fulfilled, we obtain $i < k$ and thus $(i,k) \in E_{G'} \cap E_G$.

Case 2: $(j,k),(k,j),(i,k),(k,i) \notin K$ and $((i,j) \in K$ or $(j,i) \in K)$.
Similarly to Case 1 we get $j < k$ from $(j,k) \in E_{G'} \cap E_G$ and thus $i \in \{j-1, j+1,\}$, which implies $i < k$ and consequently $(i,k) \in E_{G'} \cap E_G$.

Case 3: $(i,j),(j,i),(j,k),(k,j) \notin K$.
We infer $\{(i,j),(j,k)\} \subseteq E_{G'} \cap E_G$, which implies $i < j < k$. In particular, $(i,k) \notin K$ and thus $(i,k) \in E_G \subseteq E_{G'}$ holds.

Thus, we obtain $(i,k) \in E_G$ in every case, which shows $G' \in \overline{\mathcal{G}}_m(\text{acyclic})$.

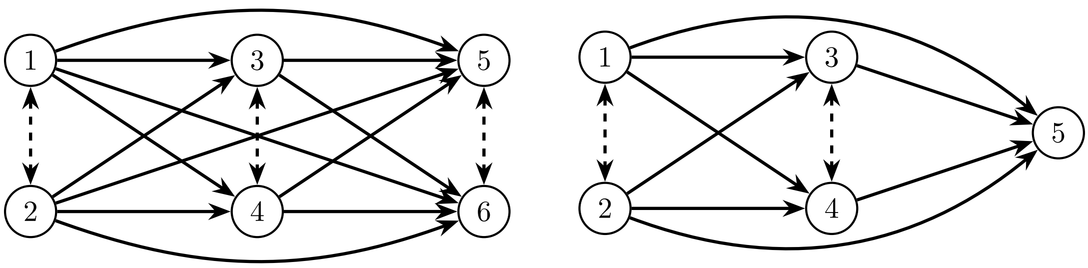
Finally, suppose $m \geq 4$ to be odd. We have just proven that there exists a graph $\tilde{G} \in \mathcal{G}_{m-1}(\neg\text{acyclic})$ with $|E_{\tilde{G}}| = \binom{m-1}{2} - \lfloor \frac{m-1}{2} \rfloor$ edges. Let $G = ([m], E_G) \in \mathcal{G}_m$ be the graph with

$$(i,j) \in E_G \quad \text{iff} \quad j = m \text{ or } (i,j) \in E_{\tilde{G}},$$

i.e., G contains the same edges as $\tilde{G}$ and in addition also all the edges (i,m), $i \in [m-1]$. Then, G is acyclic in extension and has exactly $\binom{m}{2} - \lfloor \frac{m-1}{2} \rfloor = \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$ edges.

□

For $m = 6$ and $m = 5$, the graphs $G \in \mathcal{G}_m(\text{acyclic})$ with $|E_G| = \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$ edges, which we constructed in the proof of Prop. 3.18, can be illustrated as follows. In the picture, we denote the missing edges by dashed lines.



For $G \in \mathcal{G}_m$, we call a pair $\{i,j\} \in [m]_2$ *negligible for* G if for every $k \in [m] \setminus \{i,j\}$ either $(i,k),(j,k) \in E_G$ or $(k,i),(k,j) \in E_G$ holds. The following result provides a link between acyclicity in extension and the notion of negligibility.

Proposition 3.19. *The set* $\mathcal{G}_m(\text{acyclic})$ *is*

$$\{G \in \mathcal{G}_m \,|\, G \text{ is acyclic and } \forall \{i,j\} \in [m]_2 \setminus E_G : \{i,j\} \text{ is negligible for } G\}.$$

We prepare the proof of Prop. 3.19 with two rather straight-forward observations on negligibility.

Lemma 3.20. *Let* $G_1 = ([m], E_{G_1}) \in \mathcal{G}_m$ *and* $G_2 \in \mathcal{G}_m$ *be a subgraph of* G_1, *i.e.,* $G_2 = ([m], E_{G_2})$ *and* $E_{G_2} \subseteq E_{G_1}$. *If some* $\{i,j\} \in [m]_2$ *is negligible for* G_2, *then it is also negligible for* G_1.

Proof. This is a direct consequence of the definition of negligibility together with $E_{G_2} \subseteq E_{G_1}$. $\square$

Proposition 3.21. *Let $G \in \mathcal{G}_m$. If $\{i,j\} \in [m]_2$ is negligible for G, then i and j are not contained in a 3-cycle in G.*

Proof. Let $G \in \mathcal{G}_m$ and $\{i,j\} \in [m]_2$ be fixed. If i and j are contained in a 3-cycle, there exists some $k \in [m] \setminus \{i,j\}$ with either $k \to i \to j \to k$ or $k \to j \to i \to k$, that is either $(j,k),(k,i) \in E_G$ or $(k,j),(i,k) \in E_G$. Thus, $\{i,j\}$ is not negligible for G. $\square$

Now, we are ready to prove Prop. 3.19.

Proof of Prop. 3.19. At first, suppose $G \in \mathcal{G}_m$ to be acyclic and such that every $\{i,j\} \in [m]_2 \setminus E_G$ is negligible. Let $\tilde{G} \in \overline{\mathcal{G}}_m$ with $E_G \subseteq E_{\tilde{G}}$ and assume there was a cycle in $\tilde{G}$. According to Lem. 3.2 there was also a 3-cycle in $\tilde{G}$. Since G does not contain a 3-cycle, this 3-cycle in $\tilde{G}$ has to contain at least one edge $(i^*,j^*) \in E_{\tilde{G}} \setminus E_G$. Then, $\{i^*,j^*\} \notin E_G$ holds and $\{i^*,j^*\}$ is thus negligible for G by assumption. Lem. 3.20 ensures that $\{i^*,j^*\}$ is also negligible for $\tilde{G}$, and thus Prop. 3.21 implies that (i^*,j^*) cannot be contained in any 3-cycle in $\tilde{G}$. This is a contradiction to the choice of (i^*,j^*). ↯ We conclude that $\tilde{G}$ is acyclic. Since $\tilde{G}$ was arbitrary, $G \in \mathcal{G}_m(\text{acyclic})$ follows. This completes the first part of the proof.

It remains to show that whenever $G \in \mathcal{G}_m(\text{acyclic})$ and $\{i,j\} \in [m]_2 \setminus E_G$, then $\{i,j\}$ is negligible for G. To prove the contraposition of this, suppose $G \in \mathcal{G}_m(\text{acyclic})$ and $\{i,j\} \in [m]_2 \setminus E_G$ are such that $\{i,j\}$ is *not* negligible for G. By the definition of negligibility, there exists some $k \in [m] \setminus \{i,j\}$ with

$$((i,k) \notin E_G \text{ or } (j,k) \notin E_G) \quad \text{and} \quad ((k,i) \notin E_G \text{ or } (k,j) \notin E_G).$$

Consequently, one of the following holds:

$(a.1)$ $\{i,k\} \notin E_G$ and $j \to k$ in G, $(a.2)$ $\{i,k\} \notin E_G$ and $k \to j$ in G,
$(b.1)$ $\{j,k\} \notin E_G$ and $i \to k$ in G, $(b.2)$ $\{j,k\} \notin E_G$ and $k \to i$ in G,
$(c.1)$ $i \to k \to j$ in G, $(c.2)$ $j \to k \to i$ in G,
(d) $\{i,j\},\{j,k\} \notin E_G$.

Then, we may straigthforwardly construct an extension G' of G, which contains the cycle

$(a.1)$ $j \to k \to i \to j$, $(a.2)$ $k \to j \to i \to k$,
$(b.1)$ $i \to k \to j \to i$, $(b.2)$ $k \to i \to j \to k$,
$(c.1)$ $i \to k \to j \to i$, $(c.2)$ $j \to k \to i \to j$,
(d) $i \to j \to k \to i$.

In particular, $G \notin \mathcal{G}_m(\text{acyclic})$ holds, which completes the proof. $\square$

The remainder of this section is dedicated to the following result, which will serve in the proof of Thm. 5.8 as justification for the scaling factor $\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ of the desired type II error of our improved WST testing algorithm.

Proposition 3.22. *For any $G \in \mathcal{G}_m(\text{acyclic})$ with $|E_G| > \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ there exists some $\tilde{G} \in \mathcal{G}_m(\text{acyclic})$ with $E_{\tilde{G}} \subsetneq E_G$ and $|E_{\tilde{G}}| = \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$.*

As preparation for the proof of Prop. 3.22 we formulate two auxiliary results.

Proposition 3.23. *Let $G \in \overline{\mathcal{G}}_m(\text{acyclic})$ and σ be its topological sorting. Then, $\{i,j\} \in [m]_2$ is negligible for G if and only if $|\sigma(i) - \sigma(j)| = 1$.*

Proof. To show sufficiency, we show the contraposition and may assume w.l.o.g. $\sigma(i) < \sigma(j)$. If $|\sigma(i) - \sigma(j)| > 1$, then there exists some $k \in [m]$ with $\sigma(i) < \sigma(k) < \sigma(j)$. This implies $(i,k),(k,j) \in E_G$ and thus $\{i,j\}$ is not negligible for G.

In order to show the necessity, let $k \in [m]\backslash\{i,j\}$. Then, either $\sigma(k) < \sigma(i), \sigma(j)$ or $\sigma(i), \sigma(j) < \sigma(k)$ hold, so that negligibility of $\{i,j\}$ follows directly. □

The proof of the following lemma is straight-forward and thus deferred to the appendix.

Lemma 3.24. *Let $m \geq 3$ and $S \subseteq [m]$.*

(a) If $|S| > m - \lfloor \frac{m}{3} \rfloor$, there exists $i \in [m-2]$ with $i, i+1, i+2 \in S$.

(b) If $|S| > m - \lfloor \frac{m+2}{3} \rfloor$, then $\{1,2\} \subseteq S$ or $\{m-1, m\} \subseteq S$ or there exists $i \in [m-2]$ with $i, i+1, i+2 \in S$.

Equipped with the graph theoretical results just shown, we are now in the position to verify Prop. 3.22.

Proof of Prop. 3.22. Let $G \in \mathcal{G}_m(\text{acyclic})$ with $|E_G| > \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ be fixed. Let $G' \in \overline{\mathcal{G}}_m$ be an arbitrary extension of G. As G is acyclic in extension, G' is acyclic and has thus a topological order σ such that $i \to j$ in G' iff $\sigma(i) < \sigma(j)$. Moreover, let $i_1, \dots, i_m \in [m]$ be such that $\sigma(i_1) < \sigma(i_2) < \dots < \sigma(i_m)$, i.e., $i_1 \to i_2 \to \dots \to i_m$ in G'. Define $S := \{l \in [m-1] \,|\, i_l \to i_{l+1} \text{ in } G\}$. By assumption on G, at most $\lfloor \frac{m+1}{3} \rfloor$ edges $\{i,j\} \in [m]_2$ are *not* contained in E_G, hence we have $|S| > (m-1) - \lfloor \frac{(m-1)+2}{3} \rfloor$. Lem. 3.24 assures that (a) $\exists l \in [m-3]$ with $l, l+1, l+2 \in S$ or (b) $1, 2 \in S$ or (c) $m-2, m-1 \in S$, i.e., we have (a) $i_l \to i_{l+1} \to i_{l+2} \to i_{l+3}$ or (b) $i_1 \to i_2 \to i_3$ or (c) $i_{m-2} \to i_{m-1} \to i_m$ in G (and in G').

Let us first suppose that (a) holds. Define $\tilde{G} \in \mathcal{G}_m$ via $E_{\tilde{G}} := E_G \setminus \{(i_{l+1}, i_{l+2})\}$.

Claim 1: Any $\{i,j\} \in [m]_2 \setminus (E_{\tilde{G}} \cup \{\{i_{l+1}, i_{l+2}\}\})$ is negligible for $\tilde{G}$.
Proof: Let $\{i,j\} \in [m]_2 \setminus (E_{\tilde{G}} \cup \{\{i_{l+1}, i_{l+2}\}\}) = [m]_2 \setminus E_G$ be fixed. As $G \in \mathcal{G}_m(\text{acyclic})$, Prop. 3.19 lets us infer that $\{i,j\}$ is negligible for G, and Lem. 3.20 implies that $\{i,j\}$ is negligible for G'. Consequently, Prop. 3.23 yields $\{i,j\} = \{i_{l'}, i_{l'+1}\}$ for some $l' \in [m-1]$. Recalling that $i_l \to i_{l+1} \to i_{l+2} \to i_{l+3}$ holds in G (and thus in G'), we have $\{i_{l'}, i_{l'+1}\} \cap \{i_{l+1}, i_{l+2}\} = \emptyset$. With a look at the definition of $\tilde{G}$ we may thus infer from the negligibility of $\{i_{l'}, i_{l'+1}\}$ for G that $\{i_{l'}, i_{l'+1}\}$ is also negligible for $\tilde{G}$. ■
Claim 2: $\{i_{l+1}, i_{l+2}\}$ is negligible for $\tilde{G}$.
Proof: Let $k \in [m] \setminus \{l+1, l+2\}$ be arbitrary. As $G' \in \overline{\mathcal{G}}_m$ is acyclic with $i_1 \to \dots \to i_m$ in G', we have $(i_k, i_{l+1}), (i_k, i_{l+2}) \in E_{G'}$ if $k \leq l$ and $(i_{l+1}, i_k), (i_{l+2}, i_k) \in E_{G'}$ if $k \geq l+3$. Suppose for the moment $k \leq l$. From $G' \in \mathcal{G}_m(\text{acyclic})$ and $|(l+2) - k| \geq 2$ we infer via Prop. 3.23 that $\{i_k, i_{l+2}\}$ is not negligible for G'. Now, Lem. 3.20 guarantees that $\{i_k, i_{l+2}\}$ is not negligible for G, and due to $G \in \mathcal{G}_m(\text{acyclic})$ Prop. 3.19 yields $\{i_k, i_{l+2}\} \in E_G$ and

thus also $\{i_k, i_{l+2}\} \in E_{\tilde{G}}$. As $\tilde{G}$ is a subgraph of G' we obtain $i_k \to i_{l+2}$ in $\tilde{G}$. If $k = l$, we have by choice of l that $i_l \to i_{l+1}$ in G and thus, by choice of $\tilde{G}$, also $i_k = i_l \to i_{l+1}$ in $\tilde{G}$. If not, then $l + 1 - k > 2$ and a similar argumentation as before (for proving $i_k \to i_{l+2}$ in $\tilde{G}$) yields that $i_k \to i_{l+1}$ in $\tilde{G}$.
We have shown $i_k \to i_{l+1}$ and $i_k \to i_{l+2}$ in $\tilde{G}$ for any $k \le l$. An analogue argumentation yields $i_{l+1} \to i_k$ and $i_{l+2} \to i_k$ in $\tilde{G}$ for any $k \ge l + 3$. Hence, $\{i_{l+1}, i_{l+2}\}$ is negligible for $\tilde{G}$. ■
Due to Claims 1 and 2, $\tilde{G} \in \mathcal{G}_m(\text{acyclic})$ follows from Prop. 3.19. This finishes our discussion of (a).

Next, let us consider the case (b), i.e., suppose $i_1 \to i_2 \to i_3$ in G. Define $\tilde{G} \in \mathcal{G}_m$ via $E_{\tilde{G}} \coloneqq E_G \setminus \{(i_1, i_2)\}$. Similarly as in (a), we see that each $\{i, j\} \in [m]_2 \setminus E_{\tilde{G}}$ is negligible for $\tilde{G}$: Any $\{i, j\} \in [m]_2 \setminus (E_G \cup \{\{i_1, i_2\}\})$ is according to Prop. 3.19 negligible for G, can be shown to be of the form $\{i_{l'}, i_{l'+1}\}$ and fulfills $\{i, j\} \cap \{i_1, i_2\} = \emptyset$, and is thus also negligible for $\tilde{G}$. Moreover, negligibility of $\{i_1, i_2\}$ follows from $i_2 \to i_3$ and $i_k \to i_{k+2}$ in $\tilde{G}$ for any $k \in [m - 2]$. Thus, $\tilde{G}$ is acyclic in extension according to Prop. 3.19. The case (c) can be treated exactly similar to (b).

As we have proven in each of the possible cases (a), (b) and (c) the existence of some $\tilde{G} \in \mathcal{G}_m(\text{acyclic})$ with $|E_{\tilde{G}}| = |E_G| - 1$, the statement follows via induction. To be more precise, as long as the resulting $\tilde{G}$ fulfills $|E_{\tilde{G}}| > \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$, we can iteratively repeat the argumentation above by substituting G with $\tilde{G}$ and in this way obtain $\tilde{G}_1, ..., \tilde{G}_r \in \mathcal{G}_m(\text{acyclic})$ such that $|E_{\tilde{G}_{i+1}}| = |E_{\tilde{G}_i}| - 1$, $E_{\tilde{G}_i} \subsetneq E_G$ for all $i \in [r]$ and $|E_{\tilde{G}_r}| = \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$.

□

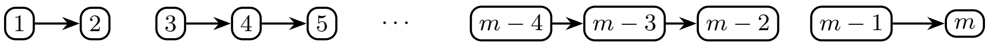
We conclude this section with the following result, which shows optimality of the term $\lfloor \frac{m+1}{3} \rfloor$ in Prop. 3.22.

Proposition 3.25. *For any $m \in \mathbb{N}_{\ge 3}$ there exists some $G \in \mathcal{G}_m(\text{acyclic})$ with $|E_G| = \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ such that any proper subgraph $\tilde{G}$ of G (i.e., $E_{\tilde{G}} \subsetneq E_G$) is* not *in $\mathcal{G}_m(\text{acyclic})$.*

Proof. At first, consider the case $m \equiv 1 \bmod 3$, i.e., $\lfloor \frac{m+1}{3} \rfloor = \frac{m-1}{3}$. Let $G' \in \overline{\mathcal{G}}_m(\text{acyclic})$ be the acyclic tournament with $1 \to 2 \to \ldots \to m$ in G'. Now, define $G \in \mathcal{G}_m$ via

$$E_G \coloneqq E_{G'} \setminus \left(\bigcup_{l=1}^{(m-1)/3} \{(3l-1, 3l)\} \right) = E_{G'} \setminus \{(2,3), (5,6), \ldots, (m-2, m-1)\},$$

where $\frac{m-1}{3} \in \mathbb{N}$ by assumption on m. That is, G contains exactly all those edges $i \to j$ with $j > i + 1$ and the following edges:



Note that each $\{3l - 1, 3l\}$, $1 \le l \le \frac{m-1}{3}$, is negligible for G. Thus, $G \in \mathcal{G}_m(\text{acyclic})$ follows from Prop. 3.19. Next, we show

$$\forall \{i, j\} \in E_G \;:\; \{i, j\} \text{ is } \textit{not} \text{ negligible for } G. \tag{3.1}$$

To see this indirectly, suppose there was some $\{i, j\} \in E_G$, $i < j$, which was negligible for G. Let $G'' \in \overline{\mathcal{G}}_m$ be an arbitrary extension of G. Recalling $G \in \mathcal{G}_m(\text{acyclic})$ and the construction of G, G'' has a topological sorting of the form

$$1 \to i_1 \to j_1 \to 4 \to i_2 \to j_2 \to 7 \to \cdots \to m - 3 \to i_{\frac{m-1}{3}} \to j_{\frac{m-1}{3}} \to m$$

for some distinct $i_1, j_1, \ldots, i_{\frac{m-1}{3}}, j_{\frac{m-1}{3}}$ with $\{i_l, j_l\} = \{3l-1, 3l\}$ for $1 \le l \le \frac{m-1}{3}$. According to Prop. 3.23 and Lem. 3.20, i and j have to be consecutive elements in the topological sorting of G''. Due to $i < j$ and the fact that E_G does not contain any $\{i_l, j_l\}$, we have either $i = 3l-2, j = 3l-1$ or $i = 3l, j = 3l+1$ for some $1 \le l \le \frac{m-1}{3}$. In the first case, we could find an extension of G that contains the cycle $3l \to 3l-1 \to 3l-2 \to 3l$, and in the latter case one with the cycle $3l+1 \to 3l \to 3l-1 \to 3l+1$, which would both contradict the fact that Lem. 3.20 guarantees neglibility of $\{i,j\}$ for any extension of G. This proves (3.1).

To finish the proof of the proposition, suppose $\tilde{G}$ to be any proper subgraph of G and let $\{i,j\} \in E_G \setminus E_{\tilde{G}}$. If $\tilde{G}$ was acyclic in extension, Prop. 3.19 would imply that $\{i,j\}$ was negligible for $\tilde{G}$, and thus also negligible for G by Lem. 3.20. This would clearly contradict Claim (3.1), hence we conclude $\tilde{G} \notin \mathcal{G}_m(\text{acyclic})$.

In the remaining cases $m \equiv 0 \bmod 3$ resp. $m \equiv 2 \bmod 3$ let again $G' \in \overline{\mathcal{G}}_m(\text{acyclic})$ with $1 \to 2 \to \ldots \to m$ in G', define

$$E_G \coloneqq E_{G'} \setminus \left(\bigcup_{l=1}^{m/3} \{(3l-2, 3l-1)\} \right) \text{ resp. } E_G \coloneqq E_{G'} \setminus \left(\bigcup_{l=1}^{(m+1)/3} \{(3l-2, 3l-1)\} \right)$$

and argue similarly as above. □

3.7. DSTAs for Testing Acyclicity of Tournaments

In this section, we present a solution to $\mathcal{D}^m_{\text{acyclic}}$ for even $m \ge 4$ and then discuss lower bounds for this problem.

Proposition 3.26. *If $m \ge 4$ is even, Alg. 13 solves $\mathcal{D}^m_{\text{acyclic}}$ with worst-case termination time $\binom{m}{2} - 1$.*

Proof. Let $m \ge 4$ be even and fixed, and write $\mathcal{A}$ for Alg. 13. Suppose $\mathcal{A}$ is started on an arbitrary $G \in \overline{\mathcal{G}}_m$. If $\mathcal{A}$ terminates in Step 2, $\mathbf{D}(\mathcal{A}) = \text{acyclic}$ is correct. Thus, suppose that the queries $\{i,j\}$, $1 \le i < j \le m-1$ made in Step 1 do not reveal a cycle of G. Then, $G|_{[m-1]} = ([m-1], \{(i,j) \in E_G \mid 1 \le i, j \le m-1\})$ is acyclic and there exists a topological sorting $\tilde{\sigma} : [m-1] \to [m-1]$ of $G_{[m-1]}$. In particular, the mapping σ in Step 3 exists and is given by $\sigma(i) = \tilde{\sigma}^{-1}(i)$ for $i \in [m-1]$ and $\sigma(m) = m$.

In Step 4, $\mathcal{A}$ queries all the edges $\{\sigma(l), m\}$ for $l \in L = \{2, 4, \ldots, m-2\}$ and these queries have not been made before. We proceed with a case distinction for the observed feedback for these queries.

Case 1: $\forall l \in L : m \to \sigma(l)$.
According to Step 5, $\mathcal{A}$ chooses $e \coloneqq \{\sigma(1), m\}$, and then queries in Step 9 all edges from $[m]_2 \setminus \{e\}$, which have not been queried yet. Thus, it has made $T = \binom{m}{2} - 1$ queries in total when it terminates in Step 10 or 11. In fact, the only edge that is not queried by $\mathcal{A}$ is $e = \{\sigma(1), m\}$. If there exists $\tilde{l} \in L$ with $\sigma(\tilde{l}+1) \to m$, then $\sigma(\tilde{l}+1) \to m \to \sigma(\tilde{l}) \to \sigma(\tilde{l}+1)$ is a cycle in $\mathfrak{G}^{\mathcal{A}}_G(T)$ (because of the choice of σ and the fact that this cycle does not contain e), and thus $\mathcal{A}$ correctly outputs $\mathbf{D}(\mathcal{A}) = \neg\text{acyclic}$ in Step 11. Otherwise, we have $m \to \sigma(j)$ for all $j \in [m-1] \setminus \{1\}$ and by definition of σ also $\sigma(1) \to \sigma(j)$ in $\mathfrak{G}^{\mathcal{A}}_G(T)$ for all $j \in [m-1]$, hence $e = \{\sigma(1), m\}$ is negligible for $\mathfrak{G}^{\mathcal{A}}_G(T)$. Consequently, Prop. 3.19 assures that the output of $\mathcal{A}$ will be correct.

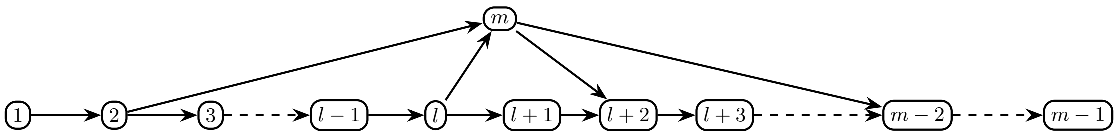


Figure 4.: Illustration of Case 2 in the proof of Prop. 3.26, where $\sigma = \mathrm{id}$.

Case 2: $\exists l \in L : \forall l' \in L \cap [0, l] : \sigma(l') \to m$ and $\forall l'' \in L \cap (l, \infty) : m \to \sigma(l'')$.
In this case, $\mathcal{A}$ queries all edges in $[m]_2$ except $e = \{\sigma(l+1), m\}$, i.e., we have again $T = \binom{m}{2} - 1$ in Step 9. In case there exists a $j \in \{1, 3, \ldots, l-1\}$ with $m \to \sigma(j)$, $j+1 \in L \cap [0, l]$ implies that $m \to \sigma(j) \to \sigma(j+1) \to m$ is a cycle in $\mathfrak{G}_G^{\mathcal{A}}(T)$. Moreover, if there is a $j \in \{l+3, l+5, \ldots, m-1\}$ with $\sigma(j) \to m$, then $j-1 \in L \cap (l, \infty)$ assures that $\sigma(j) \to m \to \sigma(j-1) \to \sigma(j)$ is a cycle in $\mathfrak{G}_G^{\mathcal{A}}(T)$. Consequently, there is either a cycle in $\mathfrak{G}_G^{\mathcal{A}}(T)$, which leads to the correct decision of $\mathcal{A}$, or we have for any $j \in [m-1] \setminus \{l+1\}$ the equivalence $\sigma(j) \to m$ iff $\sigma(j) \to \sigma(l+1)$. In the latter case, $e = \{\sigma(l+1), m\}$ is negligible, and thus $\mathcal{A}$ terminates according to Prop. 3.19 with the correct decision.

Case 3: None of the Cases 1 and 2 are true.
Then, there exist $l', l'' \in L$ with $l' < l''$ and $\sigma(l'') \to m \to \sigma(l')$. Consequently, we have the cycle $\sigma(l') \to \sigma(l'') \to m \to \sigma(l')$ in G and $\mathcal{A}$ correctly terminates with $\mathbf{D}(\mathcal{A}) = \neg\mathrm{acyclic}$ in Step 8. □

Algorithm 13 Solution to $\mathcal{D}^m_{\mathrm{acyclic}}$ if $m \geq 4$ is even.

Input: Query access to a tournament G on $[m]$

1: Query all edges $\{i, j\}$ with $1 \leq i < j \leq m-1$.
2: **if** the observations in Step 1 reveal a cycle in G **then return** ¬acyclic
3: Fix the permutation $\sigma : [m] \to [m]$ with

$$\sigma(m) = m \text{ and } \sigma(i) \to \sigma(j) \text{ for all } 1 \leq i < j \leq m-1.$$

4: Query all the edges $\{\sigma(l), m\}$ for $l \in L \coloneqq \{2, 4, \ldots, m-2\}$.
5: **if** $\forall l \in L : m \to \sigma(l)$ **then** let $e \coloneqq \{\sigma(1), m\}$
6: **else if** $\exists l \in L : \forall l' \in L \cap [0, l] : \sigma(l') \to m$ and $\forall l'' \in L \cap (l, \infty) : m \to \sigma(l'')$ **then**
7: let $e \coloneqq \{\sigma(l+1), m\}$
8: **else return** ¬acyclic
9: Query edges from $[m]_2 \setminus \{e\}$, which have not been queried yet. Let T be the total number of queried edges.
10: **if** $\mathfrak{G}_G^{\mathcal{A}}(T)$ is acyclic in extension **then return** acyclic
11: **else return** ¬acyclic

Corollary 3.27. *If $\mathcal{A}$ solves $\mathcal{D}^m_{\mathrm{acyclic}}$, then $T_G^{\mathcal{A}} \geq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$ is fulfilled for any $G \in \overline{\mathcal{G}}_m(\mathrm{acyclic})$.*

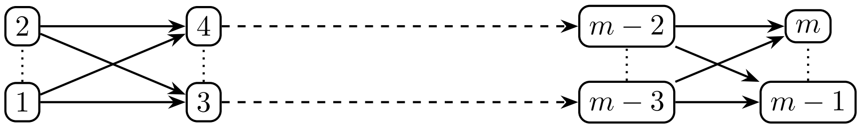
Proof. This follows directly from Prop. 3.4 and Prop. 3.18. □

Theorem 3.28. *Suppose $m \geq 4$ is even and let $\mathcal{A}$ be a solution to $\mathcal{D}^m_{\text{acyclic}}$. Then, either $T^{\mathcal{A}}_G \geq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor + 1$ for every $G \in \overline{\mathcal{G}}_m(\text{acyclic})$ or there exists some $\tilde{G} \in \overline{\mathcal{G}}_m(\text{acyclic})$ with $T^{\mathcal{A}}_{\tilde{G}} \geq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor + 2$. In particular, $T^{\mathcal{A}} \geq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor + 1$ is fulfilled.*

Proof of Thm. 3.28. Let $m \geq 4$ be even and $\mathcal{A}$ be a fixed solution to $\mathcal{D}^m_{\text{acyclic}}$. In case $T^{\mathcal{A}}_G \geq \binom{m}{2} - \frac{m}{2} + 1$ for any $G \in \overline{\mathcal{G}}_m$, there is nothing to show, hence let us suppose w.l.o.g. $T^{\mathcal{A}}_G \leq \binom{m}{2} - \frac{m}{2}$ for some fixed $G \in \overline{\mathcal{G}}_m(\text{acyclic})$. Then, G is topologically sortable, and after possibly relabeling the items in $[m]$ we may suppose w.l.o.g. that G has the topological ordering $1 \to 2 \to \cdots \to m$. As $\mathcal{A}$ is correct and $G \in \overline{\mathcal{G}}_m(\text{acyclic})$, Prop. 3.4 assures $\mathfrak{G}^{\mathcal{A}}_G\left(T^{\mathcal{A}}_G\right) \in \mathcal{G}_m(\text{acyclic})$ and thus $T^{\mathcal{A}}_G \geq |E_{\mathfrak{G}^{\mathcal{A}}_G(T^{\mathcal{A}}_G)}| \geq \binom{m}{2} - \frac{m}{2}$ by Prop. 3.18. Thus, we have $T^{\mathcal{A}}_G = \binom{m}{2} - \frac{m}{2}$ and from Prop. 3.19 we infer[3]

$$E_{\mathfrak{G}^{\mathcal{A}}_G(T^{\mathcal{A}}_G)} = E_G \setminus \{(i, i+1) \mid i \in [m] \text{ is odd}\} \tag{3.2}$$

That is, $\mathfrak{G}^{\mathcal{A}}_G\left(T^{\mathcal{A}}_G\right)$ can be depicted as follows.



Therein, we have left out those edges $(i, j) \in E_{\mathfrak{G}^{\mathcal{A}}_G(T^{\mathcal{A}}_G)}$ with $j - i \geq 4$, and have depicted those edges which are *not* contained in $\mathfrak{G}^{\mathcal{A}}_G\left(T^{\mathcal{A}}_G\right)$ as dotted lines. Now, let

$$E' := \left\{(i, j) \in E_{\mathfrak{G}^{\mathcal{A}}_G(T^{\mathcal{A}}_G)} \mid (i \text{ odd and } j - i \leq 3) \text{ or } (i \text{ even and } j - i \leq 2)\right\}$$

be the set of edges of $\mathfrak{G}^{\mathcal{A}}_G\left(T^{\mathcal{A}}_G\right)$ which are depicted as solid lines in the picture, and define the time

$$T' := \max\left\{t \in \mathbb{N} \mid \{i^{\mathcal{A}}_G(t), j^{\mathcal{A}}_G(t)\}_G \in E'\right\}$$

at which $\mathcal{A}$ (if started on G) queries the last edge contained in E'. As $E' \subseteq E_{\mathfrak{G}^{\mathcal{A}}_G(T^{\mathcal{A}}_G)}$ holds, we trivially have $T' \leq \binom{m}{2} - \frac{m}{2}$. Denote the corresponding last query as $\{i', j'\} = \{i^{\mathcal{A}}_G(T'), j^{\mathcal{A}}_G(T')\}$ with $i' < j'$. Due to $(i', j') \in E' \subseteq E_{\mathfrak{G}^{\mathcal{A}}_G(T^{\mathcal{A}}_G)}$, there exists an odd $z \in [m-3]$ with $\{i', j'\} \subsetneq \{z, z+1, z+2, z+3\}$, and a look at (3.2) reveals that $i' \in \{z, z+1\}$ and $j' \in \{z+2, z+3\}$ has to hold. Define $i'', j'' \in [m]$ such that $\{i', i''\} = \{z, z+1\}$ and $\{j', j''\} = \{z+2, z+3\}$.

Now, let $\tilde{G} \in \overline{\mathcal{G}}_m$ be that tournament with topological ordering

$$1 \to 2 \to \cdots \to i'' \to j' \to i' \to j'' \to \cdots \to m-1 \to m.$$

Comparing this with the topological ordering of G and regarding the choice of z above, we see $z = \sigma_{\tilde{G}}(i'')$, $z + 1 = \sigma_{\tilde{G}}(j')$, $z + 2 = \sigma_{\tilde{G}}(i')$ and $z + 3 = \sigma_{\tilde{G}}(j'')$.

Claim 1: We have

$$\left\{\{i, j\} \in [m]_2 \mid \{i, j\}_G \neq \{i, j\}_{\tilde{G}}\right\} \subseteq \{\{i', i''\}, \{j', j''\}, \{i', j'\}\}.$$

[3]Note here: According to Prop. 3.19, every edge $(i, j) \in E_G \setminus \mathfrak{G}^{\mathcal{A}}_G\left(T^{\mathcal{A}}_G\right) =: \tilde{E}$ has to be of the form $(i, i+1)$ and there cannot be two consecutive ones, i.e., if $(i, i+1) \in \tilde{E}$ then neither $(i-1, i) \in \tilde{E}$ nor $(i+1, i+2) \in \tilde{E}$. Thus, $|\tilde{E}| = \frac{m}{2}$ implies $\tilde{E} = \{(1, 3), (3, 5), \ldots, (m-1, m)\}$.

Proof: Comparing the topological orderings of G and $\tilde{G}$, we see on the one hand

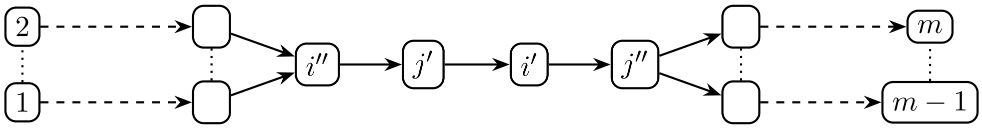
$$\forall \text{ distinct } i, j \in [m] \setminus \{i', j', i'', j''\} : (i, j) \in E_G \Leftrightarrow (i, j) \in E_{\tilde{G}}$$

and on the other hand that for every $i \in [m]$ and $j \in \{i', i'', j', j''\}$ the two implications

$$i < j \Rightarrow (i, j) \in E_{\tilde{G}} \cap E_G \quad \text{and} \quad i > j \Rightarrow (j, i) \in E_{\tilde{G}} \cap E_G$$

hold, i.e., $\{i, j\}_G = \{i, j\}_{\tilde{G}}$ is fulfilled in every such cases. It remains to consider the three possible choices of $\{i, j\}$ where $i, j \in \{i', j', i'', j''\}$ are distinct and $\{i, j\} \notin \{\{i', i''\}, \{j', j''\}, \{i', j'\}\}$, i.e., the choices $\{i'', j'\}, \{i', j''\}$ and $\{i'', j''\}$. From $z = \sigma_{\tilde{G}}(i'')$, $z+1 = \sigma_{\tilde{G}}(j')$ and $i'' \in \{z, z+1\}, j' \in \{z+2, z+3\}$ we infer $\{i'', j'\}_{\tilde{G}} = (i'', j') = \{i'', j'\}_G$, and similarly we see $\{i', j''\}_{\tilde{G}} = (i', j'') = \{i', j''\}_G$ as well as $\{i'', j''\}_{\tilde{G}} = (i'', j'') = \{i'', j''\}_G$. ■

Equation (3.2) assures that $\mathcal{A}$ started on G queries neither $\{i', i''\} = \{z, z+1\}$ nor $\{j', j''\} = \{z+2, z+3\}$ before termination, and according to the definition of T' it does not query $\{i', j'\}$ before time T'. Consequently, Claim 1 implies that $\mathcal{A}$ started on G makes until time T' exactly the same queries as started on $\tilde{G}$, and $\mathfrak{G}^{\mathcal{A}}_{\tilde{G}}(T')$ looks as follows.



In particular, the set

$$\begin{aligned} X := \{\{i, j\} \in [m]_2 : \{i, j\} = \{\sigma_{\tilde{G}}(l), \sigma_{\tilde{G}}(l+1)\} \text{ for some } l \in [m-1] \\ \text{and } \mathcal{A} \text{ started on } \tilde{G} \text{ queries } \{i, j\} \text{ before termination}\} \\ \supseteq \{\{l, l+1\} \mid l \in [m-1] \text{ even}\} \cup \{\{z, z+1\}, \{z+2, z+3\}\}. \end{aligned}$$

contains at least $\lfloor \frac{m-1}{2} \rfloor + 2 = \frac{m}{2} + 1$ elements. From Propositions 3.4 and 3.19 we know that $\mathcal{A}$ started on $\tilde{G}$ may only leave out those queries $\{i, j\}$, which are neglibible for $\mathfrak{G}^{\mathcal{A}}_{\tilde{G}}\left(T^{\mathcal{A}}_{\tilde{G}}\right)$, and according to Prop. 3.23 these can only be of the form $\{i, j\} = \{\sigma_{\tilde{G}}(l), \sigma_{\tilde{G}}(l+1)\}$ for some $l \in [m-1]$. Thus, $|X| \geq \frac{m}{2} + 1$ assures us $|E_{\mathfrak{G}^{\mathcal{A}}_{\tilde{G}}(T^{\mathcal{A}}_{\tilde{G}})}| \geq \binom{m}{2} - (m-1) + \frac{m}{2} + 1 = \binom{m}{2} - \frac{m}{2} + 2$, and we conclude $T^{\mathcal{A}} \geq T^{\mathcal{A}}_{\tilde{G}} \geq \binom{m}{2} - \frac{m}{2} + 2$. □

Corollary 3.29. *Alg. 13 is an optimal solution to* $\mathcal{D}^4_{\text{acyclic}}$.

Proof. This is a direct consequence of Prop. 3.26 and Thm. 3.28. □

Consequences for Randomized Testing Algorithms Above, we have seen that any *deterministic* sequential testing algorithm, which correctly classifies any $G \in \overline{\mathcal{G}}_m$ as acyclic or ¬acyclic, has to make in the worst case at least $\binom{m}{2} - \lfloor \frac{m}{2} \rfloor + 1$ queries before termination. But from Thm. 3.28 we can also obtain a result for possibly randomized solutions of this classification problem, as we will discuss in the following.

For this, suppose $\mathcal{A}$ to be any sequential testing algorithm for this problem, which may incorporate in each time step some randomness in the choice of its next query. Let us call $\mathcal{A}$ *correct* if it outputs with probability 1 for any $G \in \overline{\mathcal{G}}_m$ the correct decision, i.e., writing $\mathbb{P}_G$ for the probability measure on the possible states of $\mathcal{A}$ started on G, and as usual $\mathbf{D}(\mathcal{A})$ for the decision of $\mathcal{A}$ started on G, $\mathcal{A}$ is correct if $\mathbb{P}_G(\mathbf{D}(\mathcal{A}) = \text{acyclic}) = 1$ for all $G \in \overline{\mathcal{G}}_m(\text{acyclic})$ and $\mathbb{P}_G(\mathbf{D}(\mathcal{A}) = \neg\text{acyclic}) = 1$ for all $G \in \overline{\mathcal{G}}_m(\neg\text{acyclic})$. Write $\mathfrak{A}$ for the set of all correct DSTAs that solve $\mathcal{D}^m_{\text{acyclic}}$ and note that $|\mathfrak{A}| < \infty$. If $\mathcal{A}$ is correct, there exists $(p_a)_{a\in\mathfrak{A}} \in [0,1]^{|\mathfrak{A}|}$ with $\sum_{a\in\mathfrak{A}} p_a = 1$ s.t. $\mathcal{A}$ behaves for any $G \in \overline{\mathcal{G}}_m$ with probability p_a exactly as $a \in \mathfrak{A}$. By *Yao's minimax principle* [Yao, 1977], we obtain for any subclass $\overline{\mathcal{G}}'_m \subset \overline{\mathcal{G}}_m$ of instances and any probability distribution μ' on $\overline{\mathcal{G}}'_m$ for $\mathcal{A}$'s sample complexity $T^{\mathcal{A}}$ the worst-case lower bound

$$\max_{G\in\overline{\mathcal{G}}'_m} \mathbb{E}_G[T^{\mathcal{A}}] \geq \min_{a\in\mathfrak{A}} \mathbb{E}_{G\sim\mu'}[T^a_G].$$

For arbitrary $G \in \mathcal{G}_m(\text{acyclic})$ we know from Cor. 3.27 that $T^a_G \geq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$ for any $a \in \mathfrak{A}$ and thus applying Yao's principle with $\overline{\mathcal{G}}'_m = \{G\}$ and $\mu'(G) = 1$ trivially yields

$$\mathbb{E}_G[T^{\mathcal{A}}] \geq \binom{m}{2} - \left\lfloor \frac{m}{2} \right\rfloor.$$

Based on the proof of Thm. 3.28, we are able to obtain a slightly stronger result.

Corollary 3.30. *If $\mathcal{A}$ is a (possibly random) sequential algorithm, which correctly tests any $G \in \overline{\mathcal{G}}_m$ for acyclicity, then*

$$\max_{G\in\overline{\mathcal{G}}'_m} \mathbb{E}_G[T^{\mathcal{A}}] \geq \binom{m}{2} - \left\lfloor \frac{m}{2} \right\rfloor + \frac{2}{3(m-3)+1}$$

for a subclass $\overline{\mathcal{G}}'_m \subsetneq \overline{\mathcal{G}}_m(\text{acyclic})$ of size $|\overline{\mathcal{G}}'_m| \leq 3(m-3)+1$.

Proof. As in the proof of Thm. 3.28 let $G \in \overline{\mathcal{G}}_m(\text{acyclic})$ be the tournament with topological sorting $1 \to \ldots \to n$. Moreover, suppose $a \in \mathfrak{A}$ to be a DSTA, which solves $\mathcal{D}^m_{\text{acyclic}}$. In the proof, we have constructed an element $\tilde{G} \in \overline{\mathcal{G}}_m(\text{acyclic})$, which was dependent on the algorithmic behaviour of a and is thus denoted by $\tilde{G}_a$ in the following, with the property

$$\forall a \in \mathfrak{A} : T^a_G = \binom{m}{2} - \lfloor m/2 \rfloor \quad \Rightarrow \quad T^a_{\tilde{G}_a} \geq \binom{m}{2} - \left\lfloor \frac{m}{2} \right\rfloor + 2.$$

More precisely, we have seen that $\tilde{G}_a$ is a tournament with topological sorting

$$1 \to 2 \to \cdots \to i'' \to j' \to i' \to j'' \to \cdots \to m-1 \to m$$

for some appropriate values i'', j', i', j'' with $\{i', i''\} = \{z, z+1\}$ and $\{j', j''\} = \{z+2, z+3\}$ for an appropriate choice of $z \in [m-3]$, which depend on a itself. Let $\overline{\mathcal{G}}'_m$ be the set of all tournaments on $[m]$, which fulfill these constraints. Note that $G \in \overline{\mathcal{G}}'_m$. Moreover, there are $m-3$ choices for z, and for each of these there are four choices of i', j', i'', j'' such that the constraints are met, and one of these four yields exactly G as a tournament. Consequently, there are at most $3(m-3)+1$ elements in $\overline{\mathcal{G}}'_m$. Now, let μ' be defined via $\mu' : \overline{\mathcal{G}}'_m \to [0,1], \mu'(G') \coloneqq 1/|\overline{\mathcal{G}}'_m|$ for all $G' \in \overline{\mathcal{G}}'_m$. Abbreviate $x \coloneqq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$. By the proof of Thm. 3.28 we have, for any $a \in \mathfrak{A}$, that (a) $\min_{G'\in\overline{\mathcal{G}}'_m} T^a_{G'} \geq x+1$ or (b) $\exists G' \in \overline{\mathcal{G}}'_m : T^a_{G'} \geq x+2$ and $\forall G' \in \overline{\mathcal{G}}'_m \setminus \{G\} : T^a_{G'} \geq x$. Hence,

$$\mathbb{E}_{G'\sim\mu'}[T^a_{G'}] \geq \min\left\{ x+1, \frac{(|\overline{\mathcal{G}}'_m|-1)x + (x+2)}{|\overline{\mathcal{G}}'_m|} \right\} = x + \frac{2}{|\overline{\mathcal{G}}'_m|}$$

for any $a \in \mathfrak{A}$, and we conclude via Yao's minimax principle [Yao, 1977]

$$\max_{G' \in \overline{\mathcal{G}}'_m} \mathbb{E}_{G'}[T^{\mathcal{A}}] \geq \min_{a \in \mathfrak{A}} \mathbb{E}_{G' \sim \mu'}[T^a_{G'}] \geq x + \frac{2}{|\overline{\mathcal{G}}'_m|} \geq \binom{m}{2} - \left\lfloor \frac{m}{2} \right\rfloor + \frac{2}{3(m-3)+1}.$$

□

3.8. Discussion and Related Work

With regard to the further course of this thesis, the main contributions of this chapter are the optimal solution to $\mathcal{D}^m_{\mathrm{CWc}}$ with its guarantees resp. the theory on negligible edges and acyclicity in extension, as these will be of particular interest in Chapters 4 resp. 5. But along the road, we achieved further insights in testing properties of tournaments in general, with a special emphasis on (non-)acyclicity, (non-)existence of a CW and related properties. Many results were already published in [Haddenhorst et al., 2021a,b], but in comparison to these, we defined for general graph properties X and Y the term X *in* Y*-extension* and generalized the notion of a DSTA accordingly, corrected some minor mistakes, and also discussed acyclicity testing on $\overline{\mathcal{G}}_m$.

The notion of a tournament is quite basic in graph theory, and it may not be surprising that learning problems for tournaments have already attained some attention in the literature. Moon [2015] provides an extensive overview of observations on tournaments in graph theory, and the author discusses problems such as "Provided $G \in \overline{\mathcal{G}}_m(\text{acyclic})$, how many queries of G are in the worst-case necessary and sufficient to find the topological sorting of G?" (Sec. 16) or also "Given a property X, what is the minimal/maximal size of m such that a tournament $G \in \mathcal{G}_m$ with property X exists?" (Sec. 11) for a particular choice X. However, such questions are rather loosely related to our work, because they apparently do not help with our particular problems at hand.

Bollobás [1978] considered the problem of sequentially testing properties of undirected graphs on $[m]$ in the form of a two-player game, in which a player A tries to detect (as early as possible) whether a graph G, that is iteratively constructed by an adversary player B, has some property X or not. Here, at each time step, A may either terminate and decide for X or $\neg$X or alternatively choose a query $\{i, j\} \in [m]_2$, whereupon B decides whether $\{i, j\}$ is an edge in G or not. A tries to correctly terminate as early as possible, whereas B tries to work against this. The number $c(\mathrm{X})$ of queries A makes before termination, when both A and B are optimal players, is considered as complexity of X and it can be seen as an analogon of the worst-case termination time of DSTAs for tournaments as introduced above. The author is particularly interested whether X is *elusive* meaning that $c(\mathrm{X}) = \binom{m}{2}$, and shows that a huge class of properties X is non-elusive. Thm. VIII.1.2 in [Bollobás, 1978] states that acyclicity (of undirected graphs) is elusive in case $m \geq 3$. In contrast, transferring this language to our setting, Thm. 3.26 shows that acyclicity of tournaments (i.e., particular digraphs) is non-elusive if $m \geq 4$ is even. As far as we know, this is a novel result, and we are not aware whether an analogon is true for odd $m \geq 4$.

It has already been known that testing acyclicity (amongst several other properties) on $\overline{\mathcal{G}}_m$ requires $\Omega(m^2)$ queries in a worst-case sense [Rivest and Vuillemin, 1975, Holt and Reingold, 1972]; more precisely, they show that at least $\frac{m^2-1}{4}$ queries are necessary for acyclicity testing. We have shown that actually at most $\mathcal{O}(m)$ of the $\binom{m}{2}$ possible queries $\{i, j\} \in [m]_2$ may be left out by a solution whenever $G \in \overline{\mathcal{G}}_m(\text{acyclic})$. More precisely, our observations on acyclicity in extension rather straight-forwardly implied that any deterministic solution $\mathcal{A}$ to $\mathcal{D}^m_{\text{acyclic}}$ fulfills $\min_{G\in\overline{\mathcal{G}}_m(\text{acyclic})} T^{\mathcal{A}}_G \geq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$ (Cor. 3.27). From this, an application of Yao's minimax principle [Yao, 1977] can already assure $\max_{G\in\overline{\mathcal{G}}_m(\text{acyclic})} \mathbb{E}_G[T^{\mathcal{A}}] \geq \binom{m}{2} - \lfloor \frac{m}{2} \rfloor$ for any probabilistic solution $\mathcal{A}$ to $\mathcal{D}^m_{\text{acyclic}}$. For even $m \geq 4$, we improved these bounds in Thm. 3.28 and Cor. 3.30 for deterministic and probabilistic algorithms by additional summands 1 and $\frac{2}{3(m-3)+1}$, respectively. We have to admit that this is only a small improvement, but the corresponding proofs are non-trivial and thus added for the sake of completeness. To the best of our knowledge, these results are novel, and we believe that similar lower bounds can also be obtained for odd $m \geq 4$.

Both the upper and lower bounds on solutions to $\mathcal{D}^m_{\text{acyclic}}$ are not going to be exploited in the further course of this paper. Instead, they can rather be seen as interesting by-products of our graph-theoretical observations and are certainly of interest on their own.

The problem $\mathcal{D}^m_{\text{CWt}}$ has apparently been solved independently of each other by Bollobás and Eldridge [1978], Balasubramanian et al. [1997] and Procaccia [2008], and we stated its provably optimal solution in Prop. 3.14. For the sake of completeness, we also gave solutions and lower bound results for the related problems $\mathcal{D}^m_{\text{CWc}}$, $\mathcal{D}^m_{\text{CWv}}$, $\mathcal{D}^m_{\text{CWi}}(\text{CW})$, $\mathcal{D}^m_{\text{CWv}}(\text{CW})$ even though they may appear rather trivial.

Throughout, we have mainly focused on worst-case query complexity of DSTAs, as the theoretical analysis of solutions to the dueling bandits in Chapters 4 and 5 mainly require such type of guarantees. From a theoretical perspective, also best- or average-case or even weighted analoga appear interesting, and (where directly obtainable) we already stated some results regarding the best-case query complexity. Apart from that, it might be interesting to further discuss the sample complexity of randomized algorithms for the testing problems from above: We have merely given an expected sample complexity bound of randomized solutions to $\mathcal{D}^m_{\text{acyclic}}$ that are with probability 1 correct for any $G \in \overline{\mathcal{G}}_m$, but one might e.g. also ask for corresponding bounds of algorithms that are only correct with confidence $1-\gamma$ for all $G \in \overline{\mathcal{G}}_m$.

Part II.

Learning Problems in Dueling Bandits

4. Testification for the Existence of a Condorcet Winner

In Part II, we discuss several decision problems in the realm of dueling bandits (DB), starting with the CW-related decision problems introduced in Sec. 1.1 in this chapter. Recall from Sec. 1.2 that in the field of DB any arm that is likely to win (i.e., with probability $\geq 1/2$) in each duel against another arm is called the *Condorcet winner* (CW). Even though the existence of a CW in the dueling bandits problem is required in a variety of papers – either explicitly or implicitly (cf. e.g. [Bengs et al., 2021] or Sec. 4.7 for a small overview) –, Zoghi et al. [2015a] noted that its existence can neither be guaranteed in theory nor in real-life scenarios. However, the arguments put forward by the authors are purely empirical, and the conclusion that a CW does not exist in the considered applications are only derived in hindsight, after having seen all the data. In particular, the authors do not provide a statistical framework to verify or reject the CW assumption, neither in an offline nor in an online manner. For this reason, we suggested the problem *CW testification* as combined <u>test</u>ing and ver<u>ification</u> of the CW: If a CW exists, return it, otherwise return ¬CW. Here, testing means finding the CW if it exists, which is usually referred to as "CW identification" throughout this thesis for clearer separation from CW testification, whereas verification refers to CW verification as introduced already in Ch. 1. A naive approach for tackling CW testification may be to treat both CW identification and CW verification therein separately in two consecutive phases, and in fact this will result in a solution to CW testification. However, a natural question is whether a more sophisticated approach, which interleaves both phases, can do better. In a similar problem, which we briefly elaborate on directly before Sec. 4.1, the answer is "yes", but for the more difficult problem of CW testification we can merely show that our particular improved, interleaved solution outperforms a particular reasonable naive one both empirically and theoretically.

Overview of Decision Problems In Ch. 1, we have already introduced several CW-related decision problems in dueling bandits. Informally, CW testification is the problem

"Is $\mathbf{Q}$ in $\mathcal{Q}_m(\mathrm{CW})$? If so, determine the CW and return it, otherwise return ¬CW.",

and similarly, CW checking and CW verification are binary classification problems with classes CW and ¬CW resp. i and $\neg i$ (given as input $i \in [m]$) of the form

"Is $\mathbf{Q}$ in $\mathcal{Q}_m(\mathrm{CW})$? If so, return CW, otherwise return ¬CW.",

"Is i the CW of $\mathbf{Q}$? If so, return i, otherwise return $\neg i$.".

For the latter, the type I/II error of the testing algorithm corresponds to a false positive/negative classification.

Now, let us formally introduce appropriate asymmetric versions for any of these problems.

As usually, we write $\mathbf{D}(\mathcal{A})$ for the decision of an algorithm $\mathcal{A}$ and $\mathcal{A}(z)$ for $\mathcal{A}$ started with input z. Throughout, we focus on algorithms $\mathcal{A}$, which might be probabilistic and interact with the underlying dueling bandits environment, as stipulated by the definition of a sampling strategy π (cf. Sec. 1.1). Moreover, we suppose $m \in \mathbb{N}$, a hardness parameter $h \in [0, 1/2)$ and desired error probabilities $\alpha, \beta \in (0, 1)$ to be fixed for the moment. Given m, h, α, β as parameters, we say that $\mathcal{A}$ solves

- *CW testification on $\mathcal{Q}_m^h$ for α, β* (short: $\mathcal{P}_{\mathrm{CWt}}^{m,h,\alpha,\beta}$) if

$$\begin{aligned} \inf_{i^* \in [m]} \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(i^*)} \mathbb{P}_{\mathbf{Q}}\big(\mathbf{D}(\mathcal{A}) = i^*\big) &\geq 1 - \alpha, \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\neg \mathrm{CW})} \mathbb{P}_{\mathbf{Q}}\big(\mathbf{D}(\mathcal{A}) = \neg \, \mathrm{CW}\big) &\geq 1 - \beta, \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h} \mathbb{P}_{\mathbf{Q}}\left(T^{\mathcal{A}} < \infty\right) &= 1, \end{aligned} \tag{4.1}$$

- *CW checking on $\mathcal{Q}_m^h$ for α, β* (short: $\mathcal{P}_{\mathrm{CWc}}^{m,h,\alpha,\beta}$) if, whenever given any $i \in [m]$ as input,

$$\begin{aligned} \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \mathrm{CW}) &\geq 1 - \alpha, \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\neg \mathrm{CW})} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \neg \, \mathrm{CW}) &\geq 1 - \beta, \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h} \mathbb{P}_{\mathbf{Q}}\left(T^{\mathcal{A}} < \infty\right) &= 1, \end{aligned}$$

- *CW verification on $\mathcal{Q}_m^h$ for α, β* (short: $\mathcal{P}_{\mathrm{CWv}}^{m,h,\alpha,\beta}$) if, whenever given any $i \in [m]$ as input, and

$$\begin{aligned} \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW}) : \mathrm{CW}(\mathbf{Q}) = i} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}(i)) = i) &\geq 1 - \alpha \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h : \mathbf{Q} \in \mathcal{Q}_m(\neg \mathrm{CW}) \text{ or } \mathrm{CW}(\mathbf{Q}) \neq i} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}(i)) = \neg i) &\geq 1 - \beta, \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h} \inf_{i \in [m]} \mathbb{P}_{\mathbf{Q}}\left(T^{\mathcal{A}(i)} < \infty\right) &= 1, \end{aligned} \tag{4.2}$$

- *CW identification on $\mathcal{Q}_m^h(\mathrm{CW})$ for α* (short: $\mathcal{P}_{\mathrm{CWi}}^{m,h,\alpha}(\mathrm{CW})$) if

$$\begin{aligned} \inf_{i \in [m]} \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW}) : \mathrm{CW}(\mathbf{Q}) = i} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = i) &\geq 1 - \alpha, \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})} \mathbb{P}_{\mathbf{Q}}\left(T^{\mathcal{A}} < \infty\right) &= 1, \end{aligned}$$

- *CW verification on $\mathcal{Q}_m^h(\mathrm{CW})$ for α, β* (short: $\mathcal{P}_{\mathrm{CWv}}^{m,h,\alpha,\beta}(\mathrm{CW})$) if, whenever given any $i \in [m]$, the following holds:

$$\begin{aligned} \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW}) : \mathrm{CW}(\mathbf{Q}) = i} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}(i)) = i) &\geq 1 - \alpha, \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW}) : \mathrm{CW}(\mathbf{Q}) \neq i} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}(i)) = \neg i) &\geq 1 - \beta, \\ \inf_{\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})} \inf_{i \in [m]} \mathbb{P}_{\mathbf{Q}}\left(T^{\mathcal{A}(i)} < \infty\right) &= 1. \end{aligned}$$

In the symmetric case $\alpha = \beta = \gamma$, we abbreviate $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma} := \mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma,\gamma}$, $\mathcal{P}_{\mathrm{CWc}}^{m,h,\gamma} := \mathcal{P}_{\mathrm{CWc}}^{m,h,\gamma,\gamma}$ and so forth. Our primary interest lies in the discussion of the sample complexity necessary and sufficient to solve the previously introduced problems. Not surprisingly, the corresponding sample complexities depend on the predefined error bounds α, β, the number of available arms m, as well as on the parameter h of the class of preference relations $\mathcal{Q}_m^h$ satisfying the low noise assumption. As we will see, any of these problems requires in the symmetric case $\alpha = \beta = \gamma$ an expected worst-case sample complexity of order $\tilde{\Theta}(\frac{m}{h^2} \ln \frac{1}{\gamma})$.

The problem $\mathcal{P}_{\text{CWi}}^{m,h,\alpha}(\text{CW})$ is also known as *best-arm identification* in the dueling bandits literature and (matching) sample complexity lower and upper bounds for it have already been established by Braverman et al. [2016].

Degenne and Koolen [2019] consider the pure exploration multi-armed bandit problem with multiple correct answers in a quite general setting, which also allows for CW testification. From their results one can obtain instance-wise optimal lower and upper bounds on the asymptotics of CW testification algorithms, which we elaborate on in Sec. 4.5. Unfortunately, these bounds do not provide any information of the sample complexity for solving the CW testification task with a predefined level of confidence, which is probably the most common use case in reality. Apart from this, the CW testification problem has merely been addressed in the deterministic scenario, in which the outcome of a duel between two arms, if queried repeatedly, is always the same. This problem, termed $\mathcal{D}_{\text{CWt}}^m$ by us in Sec. 3.5, has already been investigated by Bollobás and Eldridge [1978], Balasubramanian et al. [1997] and Procaccia [2008], and we have restated in Prop. 3.14 an optimal deterministic solution for it.

An important use case of a solution to $\mathcal{P}_{\text{CWv}}^{m,k,\alpha,\beta}$ may be to verify the validity of a ranking over the arms, say $\sigma \in \mathbb{S}_m$, which shall be coherent with $\mathbf{Q}$ in the sense that $q_{\sigma(i),\sigma(j)} > 1/2$ iff $i < j$. Such a ranking could be the output of some ranking learning algorithm in the realm of the dueling bandits setting, for instance. By iteratively verifying the arm with rank $i \in [m]$ to be the CW among the arms with lower ranks, one needs at most $m - 1$ executions of $\mathcal{P}_{\text{CWv}}^{m,k,\alpha,\beta}$ to decide, up to some adjustable confidence, whether σ is correct.

Outline of This Chapter We start our analysis in Sec. 4.1 with worst-case sample complexity lower bounds for solutions to $\mathcal{P}_{\text{CWt}}^{m,h,\alpha,\beta}$, $\mathcal{P}_{\text{CWc}}^{m,h,\alpha,\beta}$, $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}$ and $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}(\text{CW})$. Asymptotically, these turn out to be of the form $\Omega(\frac{m}{h^2} \ln \frac{1}{\alpha \vee \beta})$ and thus basically resemble in the case $\alpha = \beta$ the bound for $\mathcal{P}_{\text{CWi}}^{m,h,\alpha}(\text{CW})$ given by Braverman et al. [2016]. Afterwards, we provide in Sec. 4.2 solutions to these problems, which are proven to be (up to logarithmic factors and in case $\alpha = \beta = \gamma$) asymptotically optimal w.r.t. the quantities γ, m and h. For the case of CW testification, this almost optimal sample complexity is already achieved by means of a simple two-stage approach $\mathcal{A}$-then-Verify, in which at first a solution $\mathcal{A}$ to the *best-arm identification problem* finds a candidate for the CW and then verifies this in a subsequent phase. However, we provide a more sophisticated algorithmic framework called *Noisy Tournament Sampling* (NTS), which exploits a connection of the testing problems to the graph-theoretical considerations from Ch. 3 and has several advantages over $\mathcal{A}$-then-Verify: On the one hand, it outperforms the latter one both theoretically and empirically, and on the other hand it also allows for CW testification in a passive scenario. For the sake of readability, we deferred the technical proof of the theoretical guarantees of NTS to Ch. 4.3.

In Sec. 4.4 we briefly discuss the problems $\mathcal{P}_{\text{CWi}}^{m,h,\gamma}(\text{CW})$ and $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}(\text{CW})$, which assume the existence of a CW. The former one is also known as *best-arm identification* in the literature. Afterwards, we empirically investigate the performance of our CW testification algorithm.

In Sec. 4.5, we discuss for the sake of comparison and completeness to which extent results from pure exploration bandits with multiple correct answers as defined in [Degenne and Koolen, 2019] are suited to solve the problems at hand. For this, we focus on the two problems CW testification and CW checking and reduce these to the setting of Degenne and Koolen [2019]. More precisely, we will see that any solution $\mathcal{A}(\gamma)$ to $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$ necessarily fulfills

$$\liminf_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln \frac{1}{\gamma}} \geq \frac{1}{D_{\mathrm{CWt}}^{m,h}(\mathbf{Q})} \tag{4.3}$$

for some known constant $D_{\mathrm{CWt}}^{m,h}(\mathbf{Q}) > 0$, and that there exists a solution $\mathcal{A}(\gamma)$ to $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$ with

$$\lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} \leq \frac{1}{D_{\mathrm{CWt}}^{m,h}(\mathbf{Q})}. \tag{4.4}$$

We will prove that

$$\frac{(m-1)(1/4-h^2)}{4h^2} \leq \sup_{\mathbf{Q} \in \mathcal{Q}_m^h} \frac{1}{D_{\mathrm{CWt}}^{m,h}(\mathbf{Q})} \leq \frac{m}{8h^2} \tag{4.5}$$

holds, hence any *optimal* solution $\mathcal{A}(\gamma)$ to $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$ fulfills

$$\sup_{\mathbf{Q} \in \mathcal{Q}_m^h} \lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} \in \Theta\left(\frac{m}{h^2}\right). \tag{4.6}$$

Similar results are provided for $\mathcal{P}_{\mathrm{CWc}}^{m,h,\gamma}$. Unfortunately and in contrast to our results in Sec. 4.1 and 4.2, these results do not yield any information for cases where γ is fixed. Moreover, the algorithmic solution $\mathcal{A}(\gamma)$ presented by Degenne and Koolen [2019] is very inefficient if not infeasible in practice, which is due to a hard min-max problem that has to be solved at each time step.

Before continuing with the discussion on the CW-related problems, we give, as promised, a related problem where a combined testification approach can be provably better than a two-stage procedure. For this purpose, suppose we want to testify for the total order $\sigma_{\mathbf{R}} : [m] \to [m]$ of a relation $\mathbf{R} = (r_{i,j})_{1 \leq i,j \leq m} \in \mathcal{R}_m$ (in the sense that $\sigma_{\mathbf{R}}(i) < \sigma_{\mathbf{R}}(j)$ iff $r_{i,j} = 1$), where each $r_{i,j}$ is a random sample from $\mathrm{Ber}(q_{i,j})$ for some *a priori known* $q_{i,j} \in [0,1]$. Let $\mathcal{A}$ be any two-stage solution to the problem, which at first finds a candidate $\hat{\sigma}$ (e.g. via conducting the quicksort algorithm on the observed feedback) and then checks whether $\hat{\sigma}$ is in fact the true underlying total order. In case $\mathbf{Q} \approx (1/2)_{1 \leq i,j \leq m}$, the well-known average-case lower bound for sorting lets us infer that $\mathcal{A}$ has an expected sample complexity $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \Omega(m \ln m)$, cf. e.g. [Knuth, 1973].

On the other side, consider a testification algorithm $\mathcal{A}'$ that chooses distinct (i_1, j_1), $(j_1, k_1), (k_1, i_1), \ldots, (i_m, j_m), (j_m, k_m), (k_m, i_m) \in \langle m \rangle_2$ (which is possible if m is large enough) and checks a priori whether there is any 3-cycle of the form $i_l \to j_l \to k_l \to i_l$ or $j_l \to i_l \to k_l \to j_l$ in $\mathbf{R}$, when interpreting $i \to j$ as $r_{i,j} = 1$. As soon as it detects such a cycle, it terminates, otherwise it continues sampling until it has seen each entry of $\mathbf{R}$ once and then outputs the observed total order. In case $\mathbf{Q} \approx (1/2)_{1 \leq i,j \leq m}$, we have for all $l \in [m]$

$$\mathbb{P}_{\mathbf{Q}}\left(i_l, j_l, k_l \text{ form a cycle}\right) \approx \frac{2}{2^3} = \frac{1}{4},$$

$$\mathbb{P}_{\mathbf{Q}}\left(i_l, j_l, k_l \text{ form a cycle and } \forall l' \leq l-1 : i_{l'}, j_{l'}, k_{l'} \text{ do not form a cycle}\right) \approx \frac{3^{l-1}}{4^l},$$

$$\mathbb{P}_{\mathbf{Q}}\left(\forall l' \leq m : i_{l'}, j_{l'}, k_{l'} \text{ do not form a cycle}\right) \approx \frac{3^m}{4^m}.$$

Consequently, the average sample complexity of $\mathcal{A}'$ is then

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}'}\right] \lesssim \sum\nolimits_{l=1}^{m} 3l \cdot \frac{3^{l-1}}{4^{l-1}} \cdot \frac{1}{4} + \frac{m^2 3^m}{4^m} = 3\left(4 - \frac{(m+4)3^m}{4^m}\right) + \frac{m^2 3^m}{4^m} \in o(m \ln m).$$

Hence, $\mathcal{A}'$ provably outperforms $\mathcal{A}$ on average on these instances for large values of m.

4.1. Lower Bounds

In this section, we provide sample complexity lower bounds for solutions to the previously defined problems. They are mainly based on Thm. 2.29 and Prop. 2.13. Thm. 4.2 from below reveals that the impact of the key quantities (i.e., α, β, m, h) on the order of the worst-case sample complexity for the testification problem is similar as for the worst-case sample complexity for the sole CW identification task under the stricter assumption of existence of a total order [Braverman et al., 2016]. Moreover, the dependency on the number of arms in (4.7) below coincides with (4.5). The following lower bound is instance-wise and formulated in terms of the gaps $\bar{q}_{i,j} = |q_{i,j} - 1/2|$ of the underlying instance $\mathbf{Q}$.

Theorem 4.1. *For any fixed $h_0, \gamma_0 \in (0, 1/2)$ there exists a constant $c = c(h_0, \gamma_0) > 0$ such that the following holds:*

(i) Let $h \in (0, h_0)$, $\alpha, \beta \in (0, \gamma_0)$ and suppose that $\mathcal{A}$ is some (probabilistic) algorithm, which solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWv}}$. Then, for all $i \in [m]$ and every $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})$ with $\mathrm{CW}(\mathbf{Q}) = i$, we have

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(i)}\right] \geq c \sum\nolimits_{j \in [m] \setminus \{i\}} \frac{1}{\bar{q}^2_{i,j}} \ln \frac{1}{\alpha \vee \beta}.$$

In particular, $\mathcal{A}$ fulfills for each $i \in [m]$ the estimate

$$\sup\nolimits_{\mathbf{Q} \in \mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(i)}\right] \geq \frac{c(m-1)}{h^2} \ln \frac{1}{\alpha \vee \beta}.$$

(ii) Let $h \in (0, h_0)$, $\alpha, \beta \in (0, \gamma_0)$ and suppose that $\mathcal{A}$ is some (probabilistic) algorithm, which solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWc}}$. Moreover, let $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})$ be fixed and suppose σ to be a permutation[1] on $[m]$ such that $q_{\sigma(i),\sigma(j)} > 1/2$ iff $i < j$. Then,

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] \geq c \sum\nolimits_{j=3}^{m} \frac{1}{\bar{q}^2_{\sigma(1),\sigma(j)}} \ln \frac{1}{\alpha \vee \beta}$$

$$\geq c \min\nolimits_{j \in [m] \setminus \{\mathrm{CW}(\mathbf{Q})\}} \sum\nolimits_{j' \in [m] \setminus \{\mathrm{CW}(\mathbf{Q}), j\}} \frac{1}{\bar{q}^2_{\mathrm{CW}(\mathbf{Q}),j'}} \ln \frac{1}{\alpha \vee \beta}.$$

[1] For the existence of such a permutation confer Lem. 3.3.

In particular, $\mathcal{A}$ fulfills

$$\sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \geq \frac{c(m-2)}{h^2} \ln \frac{1}{\alpha \vee \beta}.$$

Proof of Thm. 4.1. (i) As any (probabilistic) algorithm $\mathcal{A}$ that solves $\mathcal{P}_{\mathrm{CWv}}^{m,h,\alpha,\beta}$, trivially also fulfills the guarantees

$$\inf_{\mathbf{Q}\in\mathcal{Q}_m^h(\mathrm{CW}):\mathrm{CW}(\mathbf{Q})=i} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}(i)) = i) \geq 1 - \max\{\alpha,\beta\} \quad \text{and}$$
$$\inf_{\mathbf{Q}\in\mathcal{Q}_m^h:\mathbf{Q}\in\mathcal{Q}_m(\neg\mathrm{CW}) \text{ or } \mathrm{CW}(\mathbf{Q})\neq i} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}(i)) = \neg i) \geq 1 - \max\{\alpha,\beta\}$$

for every $i \in [m]$, which are weaker than (4.2), we may suppose w.l.o.g. $\alpha = \beta =: \gamma$ from now on.

Let $i \in [m]$ be fixed. Choose $J := \{i\} \times ([m] \setminus \{i\})$. As $\mathcal{A}$ solves $\mathcal{P}_{\mathrm{CWv}}^{m,h,\gamma}$, $\mathcal{A}(i)$ is able to decide[2]

$$\mathbf{H}_{0;J} : \forall j \in [m] \setminus \{i\} : p_{i,j} \geq 1/2 \qquad \mathbf{H}_{1;J} : \exists j \in [m] \setminus \{i\} : p_{i,j} < 1/2$$

for each $\boldsymbol{p} = (p_{i,j})_{j\in[m]\setminus\{i\}} \in \prod_{j\in[m]\setminus\{i\}}\{1/2 \pm \bar{q}_{i,j}\}$ with error probability $\leq \gamma$. If $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$ with $\mathrm{CW}(\mathbf{Q}) = i$, then $q_{i,j} = 1/2 + \bar{q}_{i,j} > 1/2$ for every $(i,j) \in J$ and thus Thm. 2.29 implies

$$\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(i)}] \geq c \sum_{j\in[m]\setminus\{i\}} \frac{1}{\bar{q}_{i,j}^2} \ln \frac{1}{\gamma}$$

with $c = c(h_0, \gamma_0)$ as in Thm. 2.29. By choosing $\mathbf{Q}(\varepsilon) \in \mathcal{Q}_m^h(\mathrm{CW})$ with $\mathrm{CW}(\mathbf{Q}(\varepsilon)) = i$ such that $|q_{i',j'} - 1/2| = h + \varepsilon$ for all $(i',j') \in (m)_2$ and arbitrarily small $\varepsilon > 0$ we obtain

$$\sup_{Q\in\mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(i)}] \geq \mathbb{E}_{\mathbf{Q}(\varepsilon)}[T^{\mathcal{A}(i)}] \geq \frac{c(m-1)}{(h+\varepsilon)^2} \ln \frac{1}{\gamma},$$

hence taking the limit $\varepsilon \searrow 0$ completes the proof of (i).

(ii) As in part (i), we may suppose w.l.o.g. $\alpha = \beta =: \gamma$ from now on. As $\mathbf{Q} = (q_{i,j})_{1\leq i,j\leq m}$ has a CW iff $(q_{\sigma(i),\sigma(j)})_{1\leq i,j\leq m} \in \mathcal{Q}_m$ has a CW, we may suppose w.l.o.g. $\sigma = \mathrm{id}$ in the following, i.e., $\mathrm{CW}(\mathbf{Q}) = 1$ and $q_{i,i+1} > 1/2$ for every $i \in [m-1]$. For $\mathbf{p} = (p_{1,3}, \ldots, p_{1,m}) \in [0,1]^{m-2}$ define $\hat{\mathbf{Q}}(\mathbf{p}) \in \mathcal{Q}_m$ via

$$\hat{\mathbf{Q}}(\mathbf{p})_{i,j} = \begin{cases} p_{i,j}, & \text{if } i = 1 \text{ and } j \in \{3,\ldots,m\}, \\ q_{i,j}, & \text{otherwise}, \end{cases}$$

for any $1 \leq i < j \leq m$. As $\min_{i\in[m-1]} q_{i,i+1} > 1/2$ by assumption on σ, for any $\mathbf{p} \in [0,1]^{m-2}$ either $\hat{\mathbf{Q}}(\mathbf{p}) \in \mathcal{Q}_m(\neg\mathrm{CW})$ or $\mathrm{CW}(\hat{\mathbf{Q}}(\mathbf{p})) = 1$ is fulfilled. Provided $\mathbf{p} \in ([0,1/2) \cup (1/2,1])^{m-2}$, we thus have the equivalence

$$\hat{\mathbf{Q}}(\mathbf{p}) \in \mathcal{Q}_m(\mathrm{CW}) \;\Leftrightarrow\; \forall j \in \{3,\ldots,m\} : p_{1,j} > 1/2.$$

[2]Note here that for every $i \in [m]$ and $(p_{i,j})_{j\in[m]\setminus\{i\}} \in \prod_{j\in[m]\setminus\{i\}}\{1/2 \pm h_{i,j}\}$ with $\min_{i\neq j} h_{i,j} > h$, there is a $\mathbf{Q}' = (q'_{i',j'})_{1\leq i',j'\leq m} \in \mathcal{Q}_m^h(i)$ with $q'_{i,j} = p_{i,j}$ for every $j \in [m] \setminus \{i\}$, and $\mathrm{CW}(\mathbf{Q}') = i$ holds iff $p_{i,j} \geq 1/2$ for any $j \in [m] \setminus \{i\}$.

Suppose $\mathcal{A}'$ to be the algorithm, which gets as input sample access to $\mathbf{p} \in [0,1]^{m-2}$, simulates $\mathcal{A}$ on $\hat{\mathbf{Q}}(\mathbf{p})$ and then outputs 0 if $\mathbf{D}(\mathcal{A}) = \mathrm{CW}$, and 1 if $\mathbf{D}(\mathcal{A}) = \neg\mathrm{CW}$. As $\mathcal{A}$ solves $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWc}}$, the algorithm $\mathcal{A}'$ is able to decide

$$\mathbf{H}_0 : \forall j \in \{3,\dots,m\} : p_{1,j} \geq 1/2 \quad \text{versus} \quad \mathbf{H}_1 : \exists j \in \{3,\dots,m\} : p_{1,j} < 1/2$$

with error probability at most γ for every $\mathbf{p} \in ([0,1/2-h) \cup (1/2+h,1])^{m-2}$. Regarding that $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})$ with $\mathrm{CW}(\mathbf{Q}) = 1$ implies $\mathbf{p}' := (q_{1,3},\dots,q_{1,m}) = (1/2+\bar{q}_{1,3},\dots,1/2+\bar{q}_{1,m}) \in (1/2+h,1]^{m-2}$, Thm. 2.29 yields

$$\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \mathbb{E}_{\mathbf{p}'}\left[T^{\mathcal{A}'}\right] \geq c\sum\nolimits_{j=3}^{m} \frac{1}{\bar{q}^2_{1,j}} \ln\frac{1}{\gamma} \geq c \min_{j\in[m]\setminus\{1\}} \sum\nolimits_{j'\in[m]\setminus\{1,j\}} \frac{1}{\bar{q}^2_{1,j'}} \ln\frac{1}{\gamma}.$$

The rest follows as in Part (i), i.e., by considering relations $\mathbf{Q}(\varepsilon)$ with entries in $\{1/2 \pm (h+\varepsilon)\}$ and taking the limit $\varepsilon \searrow 0$.

□

The following theorem provides an instance-wise lower bound on the sample complexity of any algorithm solving $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$. With Thm. 4.1, the proof of it becomes trivial.

Theorem 4.2. *For $h_0, \gamma_0 \in (0,1/2)$ there exists a constant $c = c(h_0,\gamma_0) > 0$ with the following property: Let $h \in (0,h_0)$, $\alpha,\beta \in (0,\gamma_0)$ and $\mathcal{A}$ be any solution to $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$. Then, for any $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})$, we have*

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] \geq c\sum\nolimits_{j\neq \mathrm{CW}(\mathbf{Q})} \frac{1}{\bar{q}^2_{\mathrm{CW}(\mathbf{Q}),j}} \ln\frac{1}{\alpha\vee\beta}.$$

In particular, $\mathcal{A}$ fulfills

$$\sup\nolimits_{\mathbf{Q}\in\mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \geq \frac{c(m-1)}{h^2} \ln\frac{1}{\alpha\vee\beta}. \tag{4.7}$$

Proof of Thm. 4.2. Suppose $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})$ to be fixed and let $i^* := \mathrm{CW}(\mathbf{Q})$. If $\mathcal{A}$ solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$, then the algorithm $\widehat{\mathcal{A}}$ which takes any $i \in [m]$ as input, simulates $\mathcal{A}$ until it terminates and then outputs

$$\mathbf{D}\left(\widehat{\mathcal{A}}[i]\right) := \begin{cases} i, & \text{if } \mathbf{D}(\mathcal{A}) = \mathrm{CW}, \\ \neg i, & \text{if } \mathbf{D}(\mathcal{A}) = \neg\mathrm{CW} \text{ or } \mathbf{D}(\mathcal{A}) \in [m]\setminus\{i\}, \end{cases}$$

solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWv}}$. Therefore, the statement follows directly from Part (i) of Thm. 4.1 with the choice $i = i^*$.

□

Since any solution to $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{CWt}}$ also solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$, the above theorem also provides us a sample complexity lower bound for solutions to $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{CWt}}$. In case m, α and β are fixed, we can improve upon this lower bound w.r.t. its asymptotics in terms of h as follows.

At first sight, the following result may appear to contradict (4.4) and (4.5), which does not involve an additional $\ln\ln\frac{1}{h}$-factor. However, note that (4.4) only yields an upper bound on the *asymptotics* of $\frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln\frac{1}{\gamma}}$ as $\gamma \searrow 0$, whereas the lower bound from Thm. 4.3 holds for any fixed γ.[3] Thus, there is actually no contradiction.

[3]To illustrate this difference, note that $f : (0,1)^2 \to \mathbb{R}$ defined via $f(\gamma,h) := \frac{1}{h^2}\ln\ln\frac{1}{h}$ if $h \leq \gamma$ and $f(\gamma,h) := \frac{1}{h^2}$ if $h > \gamma$ fulfills $\lim_{\gamma\to 0} f(\gamma,h) = \frac{1}{h^2}$ for all fixed $h \in (0,1)$, but at the same time we have $\lim_{h\to 0} \frac{f(\gamma,h)}{\frac{1}{h^2}\ln\ln\frac{1}{h}} = 1$.

Theorem 4.3. *If $\alpha, \beta \in (0, \frac{1}{2})$ and $\mathcal{A}$ is a (probabilistic) algorithm, which solves $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{CWt}}$, then*

$$\limsup_{h \searrow 0} \sup_{\mathbf{Q} \in \mathcal{Q}^h_m} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \geq \frac{1-2(\alpha \vee \beta)}{2} > 0.$$

Proof. Since $\mathcal{A}$ solves $\mathcal{P}^{m,0,\gamma}_{\mathrm{CWt}}$ with $\gamma := \max\{\alpha, \beta\}$, we may suppose w.l.o.g. $\alpha = \beta = \gamma$. To exploit Prop. 2.13, we will show that $\mathcal{A}$ can be used to construct a solution to $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$. For this, define

$$\mathbf{Q}(p) := \begin{pmatrix} - & 1 & \dots & 1 & p \\ & - & 1 & \dots & 1 \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & 1 \\ & & & & - \end{pmatrix} \in \mathcal{Q}_m$$

for every $p \in [0,1]$ and note

$$\mathbf{Q}(p) \in \mathcal{Q}_m(\mathrm{CW}) \quad \Leftrightarrow \quad p > \frac{1}{2}$$

and also that, for any $h \in [0, 1/2)$,

$$\mathbf{Q}(p) \in \mathcal{Q}^h_m \quad \Leftrightarrow \quad \left|p - \frac{1}{2}\right| > h.$$

Let $\mathcal{A}'$ be that algorithm, which is given sample access to a coin $C \sim \mathrm{Ber}(p)$, and then simulates $\mathcal{A}$ on $\mathbf{Q}(p)$ and returns $\mathbf{D}(\mathcal{A}') = 0$ if $D(\mathcal{A}) \in [m]$ and otherwise it returns $\mathbf{D}(\mathcal{A}') = 1$. This algorithm has the guarantees

$$\mathbb{P}_p\left(\mathbf{D}(\mathcal{A}') = 0\right) = \mathbb{P}_{\mathbf{Q}(p)}\left(\mathbf{D}(\mathcal{A}) \in [m]\right) \geq 1 - \gamma$$

and

$$\mathbb{P}_p\left(\mathbf{D}(\mathcal{A}') = 1\right) = \mathbb{P}_{\mathbf{Q}(p)}\left(\mathbf{D}(\mathcal{A}) = \neg\mathrm{CW}\right) \geq 1 - \gamma,$$

and since $\mathcal{A}$ terminates a.s. for any $\mathbf{Q} \in \mathcal{Q}^0_m$, $\mathcal{A}'$ terminates a.s. for any $p \neq \frac{1}{2}$. In other words, $\mathcal{A}'$ solves $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$. By Prop. 2.13, $\mathcal{A}'$ has to throw the coin C sufficiently often for this, and as each throw of C corresponds to a query of $\{1, m\}$ from $\mathcal{A}$, we obtain

$$\begin{aligned}\limsup_{h \searrow 0} \sup_{\mathbf{Q} \in \mathcal{Q}^h_m} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} &\geq \limsup_{h \searrow 0} \sup_{p \in [0,1/2-h) \cup (1/2-h]} \frac{\mathbb{E}_{\mathbf{Q}(p)}\left[T^{\mathcal{A}}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \\ &\geq \limsup_{h \searrow 0} \sup_{p \in [0,1/2-h) \cup (1/2-h]} \frac{\mathbb{E}_{p}\left[T^{\mathcal{A}'}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \\ &\geq \frac{1-2\gamma}{2} > 0.\end{aligned}$$

□

Thm. 4.1, Thm. 4.3 and Thm. 4.2 do not provide a lower bound for solutions to the problem $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWv}}(\mathrm{CW})$. This is due to the fact, that the proof technique used in the proofs of these theorems does not seem to work for $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWv}}(\mathrm{CW})$. However, we can obtain an appropriate lower bound by means of the change-of-measure argument from Sec. 2.5. In fact, it even allows for a more general result for multi-dueling bandits, which we present in full detail as Thm. 6.9 in Sec. 6.1 below. For the sake of convenience, we merely deduce the following lower bound from this more general version at this point and leave the correctness of it for Sec. 6.1.

Proposition 4.4. *Suppose $\mathcal{A}$ solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWv}}(\mathrm{CW})$. Then, for any $i \in [m]$ and any $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})$ with $\mathrm{CW}(\mathbf{Q}) = i$, we have*

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(i)}\right] \geq \sum\nolimits_{j\in[m]\setminus\{i\}} \frac{\ln \frac{1}{2.4(\alpha\vee\beta)}}{\mathrm{kl}\,(q_{i,j}, 1-q_{j,i})} \geq \frac{\ln \frac{1}{2.4(\alpha\vee\beta)}}{4} \sum\nolimits_{j\neq i} \frac{1-\bar{q}^2_{i,j}}{\bar{q}^2_{i,j}}.$$

In particular, $\mathcal{A}$ fulfills for each $i \in [m]$ the estimate

$$\sup\nolimits_{\mathbf{Q}\in\mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(i)}] \geq \frac{(m-1)(1-h^2)}{4h^2} \ln \frac{1}{\alpha\vee\beta}.$$

Proof. Let $\gamma \coloneqq \alpha \vee \beta$. According to Thm. 6.9 we have

$$\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(i)}] \geq \sum\nolimits_{j\in[m]\setminus\{i\}} \frac{\ln \frac{1}{2.4\gamma}}{\mathrm{kl}\,(q_{i,j}, 1-q_{j,i})},$$

and since Lem. 2.43 assures

$$\mathrm{kl}\,(q_{i,j}, 1-q_{i,j}) = \mathrm{kl}\,(1/2+\bar{q}_{i,j}, 1/2-\bar{q}_{i,j}) \leq \frac{4\bar{q}^2_{i,j}}{1/4-\bar{q}^2_{i,j}},$$

the statement follows. □

4.2. Solutions for CW Testification

In this section, we systematically provide practically feasible algorithmic solutions to $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$, starting with the straightforward approach that performs identification and verification of the CW separately one after the other. For the construction of our dueling bandits algorithms, both the testing algorithms $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$ presented in Sec. 2 as well as the insights achieved in Sec. 3 will play an important role. Recalling the definitions of $\boldsymbol{n}_t$ and $\boldsymbol{w}_t$ from Ch. 1, we will conveniently write $(\boldsymbol{n}_t)_{i,j} \coloneqq (\boldsymbol{n}_t)_{i|\{i,j\}}$ and $(\boldsymbol{w}_t)_{i,j} \coloneqq (\boldsymbol{w}_t)_{i|\{i,j\}}$ for the rest of Part II.

4.2.1. Naive Approaches

Next, we give a first naive attempt to interleave identification and testing in an algorithmic solution whose obvious flaws together with graph-theoretical considerations from Sec. 3 will help us to design a more sophisticated algorithmic solution. For the sake of convenience, let us consider first the symmetric case $\alpha = \beta =: \gamma$. To construct the first solution, suppose $\mathcal{A}$ to be an algorithm with parameters m, h, γ, which is able to find the CW whenever[4]

[4]In particular, $\mathcal{A}$ is not required to have any guarantees for $\mathbf{Q} \in \mathcal{Q}^h_m(\neg\mathrm{CW})$ here.

$\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$ with an error probability at most γ, i.e., $\mathcal{A}$ is a solution to $\mathcal{P}_{\mathrm{CWi}}^{m,h,\gamma}(\mathrm{CW})$. We define $\mathcal{A}$-THEN-VERIFY via Alg. 14. It executes $\mathcal{A}(m, h, \gamma/2)$, observes its output i (the alleged CW) and afterwards verifies with error probability at most $\gamma/2$ whether i is indeed the CW by querying each of the pairs $\{i, j\}, j \neq i$, sufficiently often. Actually, the verification phase (Lines 2–4) of Alg. 14 could be replaced by any solution to $\mathcal{P}_{\mathrm{CWv}}^{m,h,\gamma/2}$, but for simplicity we restrict ourselves to the stated one.

At first sight, our framework may resemble the EXPLORE-VERIFY FRAMEWORK from Karnin [2016], but in contrast to ours the latter can only be used to identify the CW provided it exists and does *not* to solve the CW testification task. Prop. 4.5 assures that $\mathcal{A}$-THEN-VERIFY indeed solves $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$.

Algorithm 14 $\mathcal{A}$-THEN-VERIFY

Input: m, h, γ, a testing algorithm $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$
Initialization: for any $(i, j) \in (m)_2$ let $\mathcal{A}_{\mathrm{Coin}}^{i,j}$ be an instance of $\mathcal{A}_{\mathrm{Coin}}$.

1: $i \leftarrow \mathcal{A}(m, h, \gamma/2)$
2: **for** $j \in [m] \setminus \{i\}$ **do**
3: Via executing $\mathcal{A}_{\mathrm{Coin}}^{i,j}$ until its termination test whether $\mathbf{H}_0 : q_{i,j} > 1/2$ or $\mathbf{H}_1 : q_{i,j} < 1/2$ with an error $\leq \frac{\gamma}{2(m-1)}$.
4: **if** $\mathbf{D}(\mathcal{A}_{\mathrm{Coin}}^{i,j}) = 1$ for any $j \in [m] \setminus \{i\}$ **then return** $\neg\mathrm{CW}$
5: **return** i

Proposition 4.5. *Let $\mathcal{A}_{\mathrm{Coin}}$ be a any solution to $\mathcal{P}_{\mathrm{Coin}}^{h,\gamma/(2m-2)}$. If $\mathcal{A}$ solves $\mathcal{P}_{\mathrm{CW}}^{m,h,\gamma/2}(\mathrm{CW})$, then $\mathcal{A}$-THEN-VERIFY (Alg. 14) started with m, h, γ and $\mathcal{A}_{\mathrm{Coin}}$ solves $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma,\gamma/2}$ and in particular $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$.*

Proof. Suppose $\mathcal{A}$ to be a fixed solution to $\mathcal{P}_{\mathrm{CWi}}^{m,h,\gamma/2}(\mathrm{CW})$ and write for convenience $\mathcal{A}$-THEN-VERIFY for the corresponding version of Alg. 14 with parameters m, h, γ and $\mathcal{A}_{\mathrm{Coin}}$. Due to a union bound, for any $\mathbf{Q} \in \mathcal{Q}_m^h$, the probability that the sign of any $q_{i,j} - 1/2$, $j \neq i$, is estimated incorrectly in lines 2–4 of Alg. 14 is at most $\gamma/2$.

Suppose at first $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$. If $\mathcal{A}$-THEN-VERIFY makes an error, then the output of $\mathcal{A}$ in line 1 is incorrect or a mistake is made in lines 2–4. Since both of these happen happen with error probability $\leq \gamma/2$, the overall error of $\mathcal{A}$-THEN-VERIFY is at most γ. Now, suppose $\mathbf{Q} \in \mathcal{Q}_m^h(\neg\mathrm{CW})$. If an error occurs, the candidate i of $\mathcal{A}$ from Step 1 is falsely verified to fulfill $\min_{j \neq i} q_{i,j} > 1/2$ in lines 2–4, hence

$$
\begin{aligned}
&\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}\text{-THEN-VERIFY}) \in [m]) \\
&= \sum\nolimits_{i \in [m]} \mathbb{P}_{\mathbf{Q}} \left(\text{ an error is made in steps 2–4} \,\middle|\, \mathbf{D}(\mathcal{A}) = i \right) \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = i) \\
&\leq \gamma/2.
\end{aligned}
$$

□

Recall that the problem of identifying the CW (i.e., $\mathcal{P}_{\mathrm{CWi}}^{m,h,\gamma}(\mathrm{CW})$) is also referred to as the *best-arm identification* problem in the dueling bandits literature [Karnin, 2016, Bengs et al., 2021]. As many solutions to this problem have stronger requirements than the mere existence of the CW, they cannot be used without further adaptations as a candidate for $\mathcal{A}$ in Prop. 4.5. For example, SEEBS from [Ren et al., 2020] formally requires strong stochastic transitivity (SST) as well as the stochastic triangle inequality (STI) to hold, which we introduced above in Sec. 1.2 and Sec. 2.5.1; thus, SEEBS is proven to correctly identify the CW with error probability $\leq \gamma$ only for any $\mathbf{Q}$ in a proper subset $\mathcal{Q}_m^h(\mathrm{SST} \wedge \mathrm{STI}) \subsetneq \mathcal{Q}_m^h(\mathrm{CW})$. As a consequence, the error probability of SEEBS-THEN-VERIFY could only be guaranteed to be $\leq \gamma$ whenever $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{SST} \wedge \mathrm{STI}) \cup \mathcal{Q}_m^h(\neg\mathrm{CW}) \subsetneq \mathcal{Q}_m^h$. In other words, we can*not* infer that SEEBS-THEN-VERIFY solves $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma,\gamma}$.

The algorithm SELECT from Mohajer et al. [2017] is a state-of-the-art solution to the best-arm identification problem. Their authors suppose for its theoretical analysis weak stochastic transitivity (WST) to hold. Fortunately, it can be shown that this assumption is not necessary and instead the mere existence of a CW suffices, hence SELECT-THEN-VERIFY is a solution to $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$. SELECT conducts a knockout-tournament between all m arms, in which two competing arms i, j are dueled for a fixed number of times N and i wins this comparison if it wins at least $N/2$ of the duels. Choosing

$$N := \frac{(1+\varepsilon)\ln(2)\log_2(\log_2 m)}{2h^2} \quad \text{with} \quad \varepsilon := -\frac{\ln(\gamma/2)}{\ln(\log_2 m)},$$

it can be shown that SELECT is a solution to $\mathcal{P}_{\mathrm{CWi}}^{m,h,\gamma/2}(\mathrm{CW})$ with constant sample complexity $\lceil (m-1)N \rceil$. With the help of Prop. 4.5 and Prop. 2.17 we can infer that SELECT-THEN-VERIFY solves $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$ with a worst-case expected sample complexity on $\mathcal{Q}_m^h$ of order $\tilde{O}(\frac{m}{h^2}\ln\frac{1}{\gamma})$, which is with regard to Thm. 4.2 (up to logarithmic terms) asymptotically optimal.

Despite this (almost) satisfactory theoretical guarantee of the latter approach, it seems unfavorable to separate identification and testing in the learning process, as an unnecessary verification might be conducted at the end of the learning process. Quite naturally, the question arises how testing and identification could be interleaved in a suitable way, which formally boils down to construct an appropriate decision criterion.

As a gentle start for the development of our decision criterion, consider the following naive criterion: For each pair $(i,j) \in (m)_2$, sample repeatedly noisy pairwise comparisons of the corresponding pairwise probability $q_{i,j}$ until being confident enough whether $q_{i,j}$ is above or below $1/2$ with confidence $\geq 1-\gamma'$ for some $\gamma' \in (0,1)$, based on the pairwise probability estimates $(\hat{\boldsymbol{q}}_t)_{i,j} := \frac{(\boldsymbol{w}_t)_{i,j}}{(\boldsymbol{n}_t)_{i,j}}$. Then, decide for $\neg$CW in case $\hat{\boldsymbol{q}}_t := ((\hat{\boldsymbol{q}}_t)_{i,j})_{1\leq i,j\leq m}$ is in $\mathcal{Q}_m(\neg\mathrm{CW})$, and otherwise decide for $i^* = \mathrm{argmax}_{i\in[m]} \sum_{j\neq i} \mathbf{1}_{\{(\hat{\boldsymbol{q}}_t)_{i,j}>1/2\}}$, which necessarily exists in this case. Let us denote the resulting algorithm by $\mathcal{A}^{\mathrm{naive}}$ and let $\gamma' = \gamma/\binom{m}{2}$. Then, by virtue of independence of the individual stopping decisions in the pairwise samplings and Bernoulli's inequality,

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}^{\mathrm{naive}}) = \neg\mathrm{CW}) = \mathbb{P}_{\mathbf{Q}}(\hat{\boldsymbol{q}}_t \in \mathcal{Q}_m(\neg\mathrm{CW})) \geq (1-\gamma')^{\binom{m}{2}} \geq 1-\gamma$$

holds for any $\mathbf{Q} \in \mathcal{Q}_m^h(\neg\mathrm{CW})$. Similarly, the first inequality in (4.1) holds, i.e., $\mathcal{A}^{\mathrm{naive}}$ is a solution to $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$. Evidently, however, this algorithm has an expected sample complexity that depends quadratically on the number of arms m. In addition, it is not clearly specified what it means that $\mathcal{A}^{\mathrm{naive}}$ is "confident enough" about the sign of $q_{i,j} - 1/2$.

In order to overcome the obvious flaws of $\mathcal{A}^{\mathrm{naive}}$, we formulate the following questions, the answers of which will lead us to a more sophisticated algorithm for the testification problem:

(i) How can we decide, as early as possible and with confidence $\geq 1 - \gamma'$, based on $(\hat{\boldsymbol{q}}_t)_{i,j}$, whether $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ holds?
(ii) Do we need to be sure about the sign of $q_{i,j} - 1/2$ for *all* pairs $(i,j) \in (m)_2$?
(iii) Is the choice $\gamma' = \gamma/\binom{m}{2}$ necessary, or is $\gamma' = \gamma/m'$ with some $m' < \binom{m}{2}$ sufficient?

For answering the first question, the preparations made in Sec. 2 come into play: Instead of deciding non-sequentially after a fixed number of time steps, we may apply a more sophisticated, sequential testing algorithm $\mathcal{A}_{\text{Coin}} \in \mathfrak{A}_{\text{Coin}}$.

For questions (ii) and (iii), it will be fruitful to exploit a connection of the testification problem to graph-theoretical concepts of tournaments. In particular, we will see that question (ii) can be answered negatively, while the answer to question (iii) is $\gamma' = \gamma/m$ in the symmetric case and $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$ in the asymmetric case.

4.2.2. Noisy Tournament Sampling

We incorporate the graph-theoretical observations from Sec. 3 into a more sophisticated testification algorithm, which we call the *Noisy Tournament Sampling* (NTS) and denote by $\mathcal{A}^{\text{NTS}}$. The name stems from the resemblance of its underlying sampling idea to noisy sorting algorithms [Braverman and Mossel, 2008], which will be described more thoroughly in the following. The procedure is shown in Alg. 15, and therein, the updates of $\boldsymbol{n}_t$ and $\boldsymbol{w}_t$ after having seen the sample $X_{i,j}^{[t]}$ are formally given as

$$\begin{cases} (\boldsymbol{n}_t)_{i,j} := (\boldsymbol{n}_{t-1})_{i,j} + \mathbf{1}_{\{\{i(t),j(t)\}=\{i,j\}\}}, \\ (\boldsymbol{w}_t)_{i,j} := (\boldsymbol{w}_{t-1})_{i,j} + \mathbf{1}_{\{\{i(t),j(t)\}=\{i,j\} \text{ and } X_{i,j}^{[t]}=1\}}. \end{cases} \tag{4.8}$$

Recall from Sec. 1.1 that a sampling strategy in dueling bandits is a family of random mappings, which, depending on the time t and the observations $\boldsymbol{n}_0, \boldsymbol{w}_0, \ldots, \boldsymbol{n}_{t-1}, \boldsymbol{w}_{t-1}$ available before time t, determines two arms i_t, j_t to be dueled at time $t \in \mathbb{N}$.

The algorithm $\mathcal{A}^{\text{NTS}}$ maintains a graph $\hat{G}_t := ([m], \hat{E}_t)$ and successively adds edges (corresponding to pairs $(i,j) \in \langle m \rangle_2$) to $\hat{G}_t$, for which at time t the algorithm $\mathcal{A}^{\text{NTS}}$ is confident with level $1 - \gamma'$ that $q_{i,j} > \frac{1}{2}$ holds (lines 7–10). $\mathcal{A}^{\text{NTS}}$ stops only in two cases: One in which the graph $\hat{G}_t$ is in $\mathcal{G}_m(\neg\text{CW})$, i.e., none of its tournament extensions can bring forth a CW (line 13), the other in which the graph $\hat{G}_t$ is in $\mathcal{G}_m(i^*)$ for some $i^* \in [m]$, i.e., all tournament extensions are preference relations with i^* as CW. According to which event caused the termination, either the supposed CW (i.e., i^*) or $\neg$CW is returned (lines 12–13). Formally, we have $\mathbf{D}(\mathcal{A}^{\text{NTS}}) = i^*$ if $\hat{G}_t \in \mathcal{G}_m(i^*)$ and $\mathbf{D}(\mathcal{A}^{\text{NTS}}) = \neg\text{CW}$ if $\hat{G}_t \in \mathcal{G}_m(\neg\text{CW})$. Regarding the definition of $\mathcal{G}_m(i^*)$ and $\mathcal{G}_m(\neg\text{CW})$ (as well as Prop. 3.4), termination is only reasonable if $\hat{G}_t$ is in $\bigcup_{i^* \in [m]} \mathcal{G}_m(i^*) \cup \mathcal{G}_m(\neg\text{CW})$.

4.2.3. The Passive Scenario

In this section we discuss the passive testification scenario, where the sampling strategy π might *not* be specifically designed in order to ensure a quick termination of the testing algorithm. In other words, π might be any sampling strategy as defined in Sec. 1.1. Recall that Π_∞ is the set of all sampling strategies $\pi \in \Pi$, which sample every pair $\{i,j\} \in [m]_2$ a.s. infinitely often, and that the assumption $\pi \in \Pi_\infty$ is rather mild (cf. Lem. 2.8).

Algorithm 15 $\mathcal{A}^{\mathrm{NTS}}$: Noisy tournament sampling

Input: m, a sampling strategy π, a testing algorithm $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$

Initialization: for any $(i,j) \in (m)_2$ let $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ be an instance of $\mathcal{A}_{\mathrm{Coin}}$, $\boldsymbol{n}_0 \leftarrow \boldsymbol{w}_0 \leftarrow (0)_{1\le i,j\le m}$, $\hat{E}_0 \leftarrow \emptyset$,

1: **for** $t \in \mathbb{N}$ **do**
2: $\quad \{i,j\} \sim \pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0\le t'\le t-1})$, w.l.o.g. $i < j$
3: $\quad \hat{E}_t \leftarrow \hat{E}_{t-1}$
4: $\quad$ Observe $X^{[t]}_{i,j} \sim \mathrm{Ber}(q_{i,j})$
5: $\quad$ Update $\boldsymbol{n}_t$ and $\boldsymbol{w}_t$ according to (4.8)
6: $\quad$ Reveal $X^{[t]}_{i,j}$ to $\mathcal{A}^{i,j}_{\mathrm{Coin}}$
7: $\quad$ **if** $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ terminated with $\mathbf{D}(\mathcal{A}^{i,j}_{\mathrm{Coin}}) = 0$ **then**
8: $\quad\quad \hat{E}_t \leftarrow \hat{E}_t \cup \{(i,j)\}$
9: $\quad$ **if** $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ terminated with $\mathbf{D}(\mathcal{A}^{i,j}_{\mathrm{Coin}}) = 1$ **then**
10: $\quad\quad \hat{E}_t \leftarrow \hat{E}_t \cup \{(j,i)\}$
11: $\quad \hat{G}_t \leftarrow ([m], \hat{E}_t)$
12: $\quad$ **if** $\exists i^* \in [m] : \hat{G}_t \in \mathcal{G}_m(i^*)$ **then return** i^*
13: $\quad$ **if** $\hat{G}_t \in \mathcal{G}_m(\neg\mathrm{CW})$ **then return** $\neg$CW

Theorem 4.6. *Let $\pi \in \Pi_\infty$, $h \in [0, 1/2)$ and $\alpha, \beta \in (0,1)$ be fixed, write $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$ and let $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$ be a solution to $\mathcal{P}^{h,\gamma'}_{\mathrm{Coin}}$. Let $\mathcal{A}$ be Alg. 15, called with the parameters π, m and $\mathcal{A}_{\mathrm{Coin}}$. Then, $\mathcal{A}$ solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$.*

Prop. 3.9 and Prop. 3.10 indicate that the used correction term in the choice of γ' in Thm. 4.6 is optimal. Note here that, in contrast to Thm. 5.2 in [Haddenhorst et al., 2021a], Thm. 4.6 also provides a solution to $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{CWt}}$.

For the sake of readability, the proof of Thm. 4.6 is deferred to Sec. 4.3. By (passively) monitoring the statistical validity of the CW assumption, the algorithmic framework presented in Thm. 4.6 can be utilized in order to justify the usage of dueling bandits algorithms focusing on alternative best arm concepts for the goal of regret minimization, if the test component detects a violation of the CW assumption. Finally, it is worth noting that $\mathcal{A}$-THEN-VERIFY cannot be used in a sensible way for this passive scenario due to the strictly separated identification and testing phases.

4.2.4. The Active Scenario

The key question is how to construct a sampling strategy π such that $\mathcal{A}^{\mathrm{NTS}}$ terminates as soon as possible. Apparently, one needs to construct the internal tournament $\hat{G}_t$ in Alg. 15 such that it quickly becomes clear whether each extension admits a CW or not (cf. lines 12 and 13). Thus, a natural approach would be to build this tournament according to a deterministic sequential testing algorithm (DSTA) for testification of the CW in a tournament, as those are commonly designed specifically for that purpose. However, as the outcome of a duel in the underlying problem is in general not deterministic, one has to conduct the duels several times until having enough confidence on the actual pairwise probability.

Based on these considerations, we define an epoch-based sampling strategy (implicitly defined by lines 1, 9 and 13 in Alg. 16) using such a DSTA, say $\mathcal{A}_{\text{DSTA}}$, to determine which pair shall be sampled repeatedly during an epoch. To be more precise, at the beginning of each epoch, the Noisy Tournament Sampling strategy queries $\mathcal{A}_{\text{DSTA}}$ to provide a pair (i,j) for a duel. This duel is repeated until the sign of $q_{i,j} - 1/2$ is determined with a specific confidence (based on α, β and h) leading to both, the end of the current epoch, and no consideration of the pair in any upcoming epoch (lines 3–14). If the sign is assumed to be positive resp. negative, $\mathcal{A}_{\text{DSTA}}$ is provided with the feedback $i \to j$ resp. $j \to i$, as if no randomness was involved, leading $\mathcal{A}_{\text{DSTA}}$ either to suggest the next pair to be queried (lines 9 and 13) or to terminate. If $\mathcal{A}_{\text{DSTA}}$ terminates before $\mathcal{A}^{\text{NTS}}$ came to a decision, we suppose $\mathcal{A}^{\text{NTS}}$ to continue until its termination by choosing the duels uniformly at random from $\langle m \rangle_2$ (lines 18–20). As a result, we obtain Alg. 16, which is essentially a modification of Alg. 15, where line 2 is replaced by the just described sampling mechanism based on interaction with $\mathcal{A}_{\text{DSTA}}$. Its theoretic guarantees are provided in the theorem below. As $\mathcal{A}_{\text{Coin}}$ we use the SPRT that solves $\mathcal{P}_{\text{Coin}}^{h,\gamma}$ from Prop. 2.17 due to its optimality guarantees shown in that lemma and denote it by $\mathcal{A}_{\text{Coin}}^{\text{SPRT}}(h,\gamma)$ in the following.

Theorem 4.7. *Let $\mathcal{A}_{\text{DSTA}}$ be any DSTA for $\overline{\mathcal{G}}_m$ and $\gamma_0 \in (0,1/2)$ be fixed. Then, for any $\alpha, \beta \in (0,\gamma_0)$ and $h \in (0,1/2)$, the noisy sorting algorithm $\mathcal{A}^{\text{NTS}}$ (Alg. 16) called with the parameters $\mathcal{A}_{\text{DSTA}}$ as its black-box DSTA and $\mathcal{A}_{\text{Coin}}^{\text{SPRT}}(h,\gamma')$ with $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$ as $\mathcal{A}_{\text{Coin}}$, solves $\mathcal{P}_{\text{CWt}}^{m,h,\alpha,\beta}$. Moreover, if $\mathcal{A}_{\text{DSTA}}$ solves $\mathcal{D}_{\text{CWt}}^{m}$, then*

$$\sup\nolimits_{\mathbf{Q} \in \mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}} \left[T^{\mathcal{A}^{\text{NTS}}} \right] \in \mathcal{O}\left(\frac{T^{\mathcal{A}_{\text{DSTA}}}}{h^2} \ln \frac{1}{\gamma'} \right)$$

as $\max\{\frac{1}{h}, \frac{1}{\gamma'}\} \to \infty$.

Proof of Thm. 4.7. Write $\mathcal{A} := \mathcal{A}^{\text{NTS}}$ for the algorithm as considered in the statement of this theorem. Note that the choices of the queries in Alg. 16 can indeed be described by an appropriate sampling strategy π, i.e., $\mathcal{A}$ is of the form as stipulated by Alg. 15 with $\alpha = \beta = \gamma$. Lines 18–20 in Alg. 16 and the same argumentation as in the proof of Thm. 4.6 assure that $\mathcal{A}$ – and hence also π – fulfills $(\boldsymbol{n}_t)_{i,j} \to \infty$ almost surely for every $(i,j) \in (m)_2$. Thus, according to Thm. 4.6, $\mathcal{A}$ terminates a.s. for any $\mathbf{Q} \in \mathcal{Q}_m^h$. Moreover, Prop. 2.17 assures that each duel proposed by $\mathcal{A}_{\text{DSTA}}$ is conducted in expectation at most $\mathcal{O}(\frac{1}{h^2} \ln \frac{1}{\gamma'})$ times.

Now, suppose $\mathcal{A}_{\text{DSTA}}$ to be a solution to $\mathcal{D}_{\text{CWt}}^{m}$. If $\mathcal{A}_{\text{DSTA}}$ terminates, then we have according to Prop. 3.4 that $\hat{G}_t \in \mathcal{G}_m(\neg\text{CW}) \cup \bigcup_{i \in [m]} \mathcal{G}_m(i)$. Consequently, $\mathcal{A}$ terminates before reaching Line 18; at termination it has queried only those edges, which have been proposed by $\mathcal{A}_{\text{DSTA}}$, i.e., at most $T^{\mathcal{A}_{\text{DSTA}}}$ many. From this, we can directly infer that $\sup_{\mathbf{Q}} \mathbb{E}_{\mathbf{Q} \in \mathcal{Q}_m^h}[T^{\mathcal{A}}] \in \mathcal{O}(\frac{T^{\mathcal{A}_{\text{DSTA}}}}{h^2} \ln \frac{1}{\gamma'})$ as $\max\{\frac{1}{h}, \frac{1}{\gamma'}\} \to \infty$. □

When replacing $\mathcal{A}_{\text{Coin}}^{\text{SPRT}}(h,\gamma')$ by the solution to $\mathcal{P}_{\text{Coin}}^{\gamma'}$ from Prop. 2.22, which we denote by $\mathcal{A}_{\text{Coin}}^{\text{Farrell}}(\gamma')$ from now on, we obtain a solution to $\mathcal{P}_{\text{CWt}}^{m,0,\alpha,\beta}$ with the following guarantees.

Theorem 4.8. *Let $\mathcal{A}_{\text{DSTA}}$ be any DSTA for $\overline{\mathcal{G}}_m$ and $\gamma_0 \in (0,1/2)$ be fixed. Then, for any $\alpha, \beta \in (0,\gamma_0)$, the noisy sorting algorithm $\mathcal{A} = \mathcal{A}^{\text{NTS}}$ (Alg. 16) called with the parameters $\mathcal{A}_{\text{DSTA}}$ as its black-box DSTA and $\mathcal{A}_{\text{Coin}}^{\text{Farrell}}(\gamma')$ with $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$ as $\mathcal{A}_{\text{Coin}}$, solves $\mathcal{P}_{\text{CWt}}^{m,0,\alpha,\beta}$.*

Moreover, if $\mathcal{A}_{\mathrm{DSTA}}$ *solves* $\mathcal{D}^m_{\mathrm{CWt}}$*, then there exists* $h_0 \in (0,1/2)$ *s.t. for all* $\mathbf{Q} \in \mathcal{Q}^0_m$ *with* $0 < \tilde{h} \leq \bar{q}_{i,j} \leq h_0$ *for all distinct* $i,j \in [m]$

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] \leq \frac{T^{\mathcal{A}_{\mathrm{DSTA}}}}{2\tilde{h}^2} \ln\ln \frac{1}{\tilde{h}}.$$

Proof. Recall that Prop. 2.22 assures that $\mathcal{A}_{\mathrm{Coin}} = \mathcal{A}^{\mathrm{Farrell}}_{\mathrm{Coin}}(\gamma')$ solves $\mathcal{P}^{\gamma'}_{\mathrm{Coin}}$ and thus $\mathcal{A}$ solves $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{CW}}$ according to Thm. 4.6.

Now, suppose $\mathcal{A}_{\mathrm{DSTA}}$ to be a solution to $\mathcal{D}^m_{\mathrm{CWt}}$. Prop. 2.22 lets us choose $h_0 \in (0,1/2)$ such that

$$\frac{\mathbb{E}_{1/2\pm h}\left[T^{\mathcal{A}_{\mathrm{Coin}}}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \leq \frac{1}{2}$$

holds for any $h \leq h_0$. As each $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ is an instance of $\mathcal{A}_{\mathrm{Coin}}$, we have for any $\mathbf{Q} \in \mathcal{Q}^0_m$ with $\tilde{h} \leq \bar{q}_{i,j} \leq h_0$ for all $(i,j) \in (m)_2$ the estimate

$$\mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}^{i,j}_{\mathrm{Coin}}}\right] \leq \frac{1}{2\bar{q}^2_{i,j}} \ln\ln\frac{1}{\bar{q}_{i,j}} \leq \frac{1}{2\tilde{h}^2}\ln\ln\frac{1}{\tilde{h}} \tag{4.9}$$

As in the proof of Thm. 4.7, a look at Alg. 16 shows that $\mathcal{A}$ queries at most $T^{\mathcal{A}_{\mathrm{DSTA}}}$ distinct $\{i,j\} \in [m]_2$, and it does not query $\{i,j\}$ after $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ has terminated. Hence, $\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] \leq \sum_{(i,j)\in(m)_2} \mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}^{i,j}_{\mathrm{Coin}}}\right]$ and combining this with (4.9) completes the proof. □

Without much effort, we obtain the following instance-wise bound for the expected sample complexity of Alg. 16 when $\mathcal{A}_{\mathrm{Coin}}$ is chosen as the SPRT from Prop. 2.17.

Theorem 4.9. *Let* $m \in \mathbb{N}$, $\alpha,\beta \in (0,1/2)$, $h \in (0,1/2)$ *and* $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$. *Suppose* $\mathcal{A}_{\mathrm{DSTA}}$ *to be a solution to* $\mathcal{D}^m_{\mathrm{CWt}}$ *and let* $\mathcal{A} = \mathcal{A}^{\mathrm{NTS}}$ *be Alg. 16 called with* $\mathcal{A}_{\mathrm{DSTA}}$ *as its black-box DSTA and* $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{Coin}}(h,\gamma')$ *as* $\mathcal{A}_{\mathrm{Coin}}$. *Then,* $\mathcal{A}$ *solves* $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$. *Let* $\mathbf{Q} \in \mathcal{Q}^h_m$ *and suppose* $(i_1,j_1),\ldots,(i_{\binom{m}{2}},j_{\binom{m}{2}}) \in (m)_2$ *to be distinct and such that* $\bar{q}_{i_1,j_1} \leq \cdots \leq \bar{q}_{i_{\binom{m}{2}},j_{\binom{m}{2}}}$ *holds. Then, we have with* $c(h,\gamma') \coloneqq \left\lceil \frac{\ln((1-\gamma')/\gamma')}{\ln((1/2+h)/(1/2-h))} \right\rceil$ *that*

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] \leq \sum_{k=1}^{T^{\mathcal{A}_{\mathrm{DSTA}}}} \frac{c(h,\gamma')}{2\bar{q}_{i_k,j_k}} \left| 1 - 2\left(1 + (1/2+\bar{q}_{i_k,j_k})^{c(h,\gamma')}(1/2-\bar{q}_{i_k,j_k})^{-c(h,\gamma')}\right)^{-1}\right|.$$

Proof. Similarly as in the proof of Thm. 4.7 we see that $\mathcal{A}$ solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$ and also that it only queries those edges, which have been proposed by $\mathcal{A}_{\mathrm{DSTA}}$. According to the choice of $\mathcal{A}_{\mathrm{Coin}}$ and the identity (A.6) stated in the proof of Prop. 2.17, any such edge (i',j') proposed by $\mathcal{A}_{\mathrm{DSTA}}$ is queried in expectation at most

$$\frac{c(h,\gamma')}{2\bar{q}_{i',j'}} \left| 1 - 2\left(1 + (1/2+\bar{q}_{i',j'})^{c(h,\gamma')}(1/2-\bar{q}_{i',j'})^{-c(h,\gamma')}\right)^{-1}\right|$$

times by $\mathcal{A}$. This immediately concludes the proof. □

We have seen in Sec. 3.5 a solution $\mathcal{A}_{\mathrm{DSTA}}$ to $\mathcal{D}^m_{\mathrm{CWt}}$, which achieves the optimal worst-case sample complexity $T^{\mathcal{A}_{\mathrm{DSTA}}} = 2m - \lfloor \log_2 m \rfloor - 2$ (Alg. 10, Prop. 3.14). As a direct consequence of Thm. 4.7, we thus obtain the following result.

Algorithm 16 $\mathcal{A}^{\mathrm{NTS}}$ using $\mathcal{A}_{\mathrm{DSTA}}$ as sampling strategy

Input: m, $\mathcal{A}_{\mathrm{DSTA}}$,a testing algorithm $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$.
Initialization: For any $(i,j) \in (m)_2$ let $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ be an instance of $\mathcal{A}_{\mathrm{Coin}}$.
Let $\hat{E}_0 \leftarrow \emptyset$, $\boldsymbol{n}_0 \leftarrow (0)_{1\le i,j\le m}$, $\boldsymbol{w}_0 \leftarrow (0)_{1\le i,j\le m}$, $t' \leftarrow 1$, $t \leftarrow 1$
▷ $\hat{E}_t$ = set of edges (i,j), for which $q_{i,j} > \frac{1}{2}$ w.h.p.
▷ t = number of observations for π
▷ t' = number of observations for $\mathcal{A}_{\mathrm{DSTA}}$

1: $(i,j) \leftarrow (i^{\mathcal{A}_{\mathrm{DSTA}}}(1), j^{\mathcal{A}_{\mathrm{DSTA}}}(1))$ ▷ Get first query of $\mathcal{A}_{\mathrm{DSTA}}$
2: **while** $\mathcal{A}_{\mathrm{DSTA}}$ did not terminate yet **do**
3: Observe $X^{[t]}_{i,j} \sim \mathrm{Ber}(q_{i,j})$
4: $\hat{E}_t \leftarrow \hat{E}_{t-1}$
5: Reveal $\mathrm{X}^{[t]}_{i,j}$ to $\mathcal{A}^{i,j}_{\mathrm{Coin}}$
6: **if** $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ terminated with $\mathbf{D}(\mathcal{A}^{i,j}_{\mathrm{Coin}}) = 0$ **then** ▷ $i \to j$ in $G(\mathbf{Q})$ w.h.p.
7: $\hat{E}_t \leftarrow \hat{E}_t \cup \{(i,j)\}$
8: Forward 1 to $\mathcal{A}_{\mathrm{DSTA}}$ and set $t' \leftarrow t'+1$ ▷ $\mathcal{A}_{\mathrm{DSTA}}$ observes $i^{\mathcal{A}_{\mathrm{DSTA}}}(t') \to j^{\mathcal{A}_{\mathrm{DSTA}}}(t')$
9: Let $(i,j) \leftarrow (i^{\mathcal{A}_{\mathrm{DSTA}}}(t'), j^{\mathcal{A}_{\mathrm{DSTA}}}(t'))$ ▷ Choose next query from $\mathcal{A}_{\mathrm{DSTA}}$
10: **else if** $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ terminated with $\mathbf{D}(\mathcal{A}^{i,j}_{\mathrm{Coin}}) = 1$ **then** ▷ $j \to i$ in $G(\mathbf{Q})$ w.h.p.
11: $\hat{E}_t \leftarrow \hat{E}_t \cup \{(j,i)\}$
12: Forward 0 to $\mathcal{A}_{\mathrm{DSTA}}$ and set $t' \leftarrow t'+1$ ▷ $\mathcal{A}_{\mathrm{DSTA}}$ observes $j^{\mathcal{A}_{\mathrm{DSTA}}}(t') \to i^{\mathcal{A}_{\mathrm{DSTA}}}(t')$
13: Let $(i,j) \leftarrow (i^{\mathcal{A}_{\mathrm{DSTA}}}(t'), j^{\mathcal{A}_{\mathrm{DSTA}}}(t'))$ ▷ Choose next query from $\mathcal{A}_{\mathrm{DSTA}}$
14: $t \leftarrow t+1$
15: **if** $\exists i^* \in [m] : \hat{G}_t \in \mathcal{G}_m(i^*)$ **then**
16: **return** i^*
17: **if** $\hat{G}_t \in \mathcal{G}_m(\neg\mathrm{CW})$ **then return** $\neg$CW
18: **while** True **do** ▷ No interaction with $\mathcal{A}_{\mathrm{DSTA}}$ anymore
19: Sample a pair (i,j) uniformly at random from $\langle m\rangle_2$.
20: Do Steps 3–7, 10, 11 and 14–17

Corollary 4.10. *Let $\gamma_0 \in (0,1/2)$ be fixed, $\alpha,\beta \in (0,\gamma_0)$ and $h \in (0,1/2)$. The algorithm $\mathcal{A}^{\mathrm{NTS}}$ called with $\mathcal{A}_{\mathrm{DSTA}}$ as defined in Alg. 10 as its black-box DSTA as well as $\mathcal{A}_{\mathrm{Coin}} = \mathcal{A}^{\mathrm{SPRT}}_{\mathrm{Coin}}(h,\gamma')$ with $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$ as in Thm. 4.9, solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{CWt}}$ such that*

$$\sup_{\mathbf{Q}\in\mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}^{\mathrm{NTS}}}\right] \in \mathcal{O}\left(\frac{m\ln m}{h^2}\ln\frac{1}{\gamma'}\right).$$

According to Thm. 4.2, the algorithm $\mathcal{A}^{\mathrm{NTS}}$ from Cor. 4.10 is optimal w.r.t. the worst-case expected sample complexity up to a factor of $\ln m$ for the problem $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWt}}$.

In contrast to (4.6), Thm. 4.2 and Cor. 4.10 specifically yield asymptotic lower and upper bounds on the sample complexity for solving $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWt}}$ if γ is fixed. The next lemma compares the algorithmic solution $\mathcal{A}^{\mathrm{NTS}}$ to $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWt}}$ from Cor. 4.10 with the two-stage approach $\mathcal{A}$-THEN-VERIFY from Sec. 4.2.1. More precisely, we show that the solution from Cor. 4.10 theoretically outperforms SELECT-THEN-VERIFY when the corresponding testing algorithm $\mathcal{A}_{\text{Coin}}$ is chosen to be the SPRT from Prop. 2.17. Again, we decided for this choice of $\mathcal{A}_{\text{Coin}}$ as it is known to be a solution to $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ that is optimal w.r.t. its worst-case expected sample complexity. Of course, there are other candidates for $\mathcal{A}$ in $\mathcal{A}$-THEN-VERIFY, e.g., as already indicated in Sec. 4.2.1, an appropriate solution $\mathcal{A}$ to $\mathcal{P}^{m,h,\gamma}_{\text{CW}_\text{V}}$ may be inferred from [Karnin, 2016]. For the sake of convenience and simplicity, we have restricted ourselves to SELECT at this point, because the algorithm itself and its theoretical guarantees fit into our setting – e.g., it is implicitly assumed that $\mathbf{Q} \in \mathcal{Q}^h_m$ – and thus makes the discussed optimality (up to logarithmic terms) of SELECT-THEN-VERIFY rather easy to see. Even though a theoretical comparison of our solution to SELECT-THEN-VERIFY on arbitrary instances appears infeasible, we are able to show the following result.

Lemma 4.11. *Let $m \in \mathbb{N}$ and $\gamma \in (0, 1/2)$ be arbitrary. For sufficiently small $h > 0$ and sufficiently large $\tilde{h} \in (h, 1/2)$ we have: Whenever $\mathcal{A}^{\text{NTS}}$ from Cor. 4.10 and $\mathcal{A}'$:=SELECT-THEN-VERIFY are started with parameters m, h, γ and the SPRT from Prop. 2.17 as $\mathcal{A}_{\text{Coin}}$, then*

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}^{\text{NTS}}}\right] \le \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}'}\right]}{2}$$

holds for any $\mathbf{Q} \in \mathcal{Q}^{\tilde{h}}_m \subsetneq \mathcal{Q}^h_m$.

Proof. At first, let us recall the corresponding lower and upper bounds for $\mathcal{A}^{\text{NTS}}$ and $\mathcal{A}'$, which we will use. For any $m \in \mathbb{N}, \gamma \in (0, 1), h \in (0, 1/2), \tilde{h} \in (h, 1/2)$ and $\mathbf{Q} \in \mathcal{Q}_m \setminus Q^{\tilde{h}}_m$ the instance-wise upper bound from Thm. 4.9 (with $\mathcal{A}_{\text{DSTA}}$ chosen as in Cor. 4.10) yields

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}^{\text{NTS}}}\right] \le \frac{c(h,\gamma')(2m - \lfloor \log_2 m \rfloor - 2)}{2\tilde{h}} \left| 1 - 2\left(1 + \left((1/2+\tilde{h})/(1/2-\tilde{h}) \right)^{c(h,\gamma')} \right)^{-1} \right| =: g(\tilde{h})$$

for any $\mathbf{Q} \in \mathcal{Q}^{\tilde{h}}_m$, where $\gamma' := \gamma/m$ and

$$c(h,\gamma') := \left\lceil \frac{\ln((1-\gamma')/\gamma')}{\ln(1/2+h)/(1/2-h))} \right\rceil .$$

Moreover, the "SELECT-part" of SELECT-THEN-VERIFY (started with m, γ, h) requires for any $\mathbf{Q} \in \mathcal{Q}_m$ exactly

$$\frac{(m-1)(1+\epsilon(m,\gamma))\ln(2)\log_2(\log_2 m)}{2h^2} \quad \text{with } \epsilon(m,\gamma) := -\frac{\ln(\gamma/2)}{\ln(\log_2 m)}$$

samples. Note that this is trivially a lower bound for $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}'}]$.

Now, let $m \in \mathbb{N}$ and $\gamma \in (0, 1/2)$ be fixed. Define $\gamma' := \gamma/m$ and choose $h \in (0, 1/2)$ with

$$h < \frac{(m-1)(1+\epsilon(m,\gamma))\ln(2)\log_2(\log_2 m)}{2\ln((1-\gamma')/\gamma')(2m - \lfloor \log_2 m \rfloor - 2)}.$$

Then, $c(h,\gamma')$ is fixed. As $\ln((1/2+\hat{h})/(1/2-\hat{h})) > 4\hat{h}$ holds for any $\hat{h} \in (0, 1/2)$, we have $c(h,\gamma') < \frac{\ln((1-\gamma')/\gamma')}{4h}$. Regarding that $\frac{1}{2\hat{h}} \to 1$ and $\left(\frac{1/2+\hat{h}}{1/2-\hat{h}}\right)^{c(h,\gamma')} \to \infty$ as $\hat{h} \to 1/2$, we obtain

$$g(\hat{h}) \longrightarrow c(h,\gamma')(2m - \lfloor \log_2(m) \rfloor - 2) \le \frac{\ln((1-\gamma')/\gamma')(2m - \lceil \log_2 m \rceil - 2)}{4h}$$

Consequently, we have for sufficiently large $\tilde{h} \in (h, 1/2)$ and any $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}$

$$\begin{aligned}\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}^{\mathrm{NTS}}}\right] \leq g(\tilde{h}) &\leq \frac{\ln((1-\gamma')/\gamma')(2m - \lceil \log_2 m \rceil - 2)}{2h} \\ &\leq \frac{(m-1)(1+\epsilon(m,\gamma))\ln(2)\log_2(\log_2 m)}{4h^2} \leq \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}'}\right]}{2},\end{aligned}$$

where the third inequality holds due to the choice of h. $\square$

4.3. Proof of Theorem 4.6

We continue with the proof of Thm. 4.6, which states that Alg. 15 indeed solves $\mathcal{P}_{\mathrm{CWt}}^{m,h,\alpha,\beta}$, provided π is a sampling strategy in Π_∞. Note that the correction terms $\frac{1}{m}$ and $\frac{1}{m-1}$ associated for the type I/II error probabilities α/β are chosen in an optimal way with regard to a Bonferroni correction due to Prop. 3.9 and Prop. 3.10, respectively. For convenience, if $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$, we may write in the following

$$\mathbf{D}(\mathcal{A}_{\mathrm{Coin}}, t) := \begin{cases} \text{“N/A”}, & \text{if } t < T^{\mathcal{A}_{\mathrm{Coin}}}, \text{ i.e. } \mathcal{A}_{\mathrm{Coin}} \text{ has not terminated yet,} \\ \mathbf{D}(\mathcal{A}_{\mathrm{Coin}}), & \text{if } t \geq T^{\mathcal{A}_{\mathrm{Coin}}} \end{cases} \tag{4.10}$$

for the outcome of $\mathcal{A}_{\mathrm{Coin}}$ after having observed t samples.

Proof of Thm. 4.6. For convenience we abbreviate $T := T^{\mathcal{A}}$ and write $(i(t), j(t))$ for the query sampled from π at time t. With regard to the definition of $\mathcal{A}$, the testing algorithm $\mathcal{A}_{\mathrm{Coin}}^{i,j}$ observes the sample $X_{i,j}^{[t]} \sim \mathrm{Ber}(q_{i,j})$ iff $(i(t), j(t)) = (i,j)$, and after time t it has observed exactly $(\boldsymbol{n}_t)_{i,j}$ of these samples. Let $\mathbf{Q} \in \mathcal{Q}_m^h$ be fixed for the moment. Recall that $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$. We split the remaining proof into four parts.

Part 1: Almost sure finiteness of T
Let $\mathcal{A}'$ be that modification of $\mathcal{A}$, which simulates $\mathcal{A}$ until it terminates, memorizes $\mathbf{D}(\mathcal{A})$ but then continues until all $\mathcal{A}_{\mathrm{Coin}}^{i,j}$'s are terminated and afterwards terminates with the decision $\mathbf{D}(\mathcal{A})$. Since a.s. termination of $\mathcal{A}'$ would directly imply a.s. termination of $\mathcal{A}$, we may suppose w.l.o.g. $\mathcal{A}$ to be replaced by $\mathcal{A}'$ throughout our proof of Part 1. By the assumption $\pi \in \Pi_\infty$, we have $(\boldsymbol{n}_t)_{i,j} \to \infty$ almost surely as $t \to \infty$ for each $(i,j) \in (m)_2$. For any $(i,j) \in (m)_2$, $\mathcal{A}_{\mathrm{Coin}}^{i,j}$ solves by assumption $\mathcal{P}_{\mathrm{Coin}}^{h,\gamma'}$ and hence the stopping time

$$T_{i,j} = \min\left\{t \in \mathbb{N} \,:\, \mathbf{D}\left(\mathcal{A}_{\mathrm{Coin}}^{i,j}, (\boldsymbol{n}_t)_{i,j}\right) \neq \text{“N/A”}\right\}$$

is a.s. finite. In particular,

$$T' := \max_{(i,j)\in(m)_2} T_{i,j}$$

is a.s. finite, since each $T_{i,j}$ is a.s. finite by assumption. Regarding the definitions of $\mathcal{A}$ and T' we see that $\hat{G}_{T'}$ is almost surely an element of $\overline{\mathcal{G}}_m = \bigcup_{i^* \in [m]} \overline{\mathcal{G}}_m(i^*) \cup \overline{\mathcal{G}}_m(\neg\mathrm{CW})$, i.e., $\hat{G}_{T'}$ is with probability 1 either in $\overline{\mathcal{G}}_m(\neg\mathrm{CW}) \subsetneq \mathcal{G}_m(\neg\mathrm{CW})$ or in $\overline{\mathcal{G}}_m(i^*) \subsetneq \mathcal{G}_m(i^*)$ for some $i^* \in [m]$. Consequently, we obtain

$$T = \min\left\{t \in \mathbb{N} \,:\, \hat{G}_t \in \mathcal{G}_m(\neg\mathrm{CW}) \text{ or } \hat{G}_t \in \mathcal{G}_m(i^*) \text{ for some } i^* \in [m]\right\} \leq T' < \infty \quad \text{a.s.,}$$

which completes the proof of Part 1. ■

Before showing the guarantees on the type I and II errors we fix some further notation: For $\mathbf{Q} \in \mathcal{Q}_m$, $E \subseteq [m] \times [m]$ and $\{i,j\} \in [m]_2$ we say that $\{i,j\}$ *is assigned incorrectly (ass. inc.) in* E *w.r.t.* $\mathbf{Q}$ if

$$(i,j) \in E \text{ and } q_{i,j} < 1/2 \quad \text{or} \quad (j,i) \in E \text{ and } q_{i,j} > 1/2$$

holds, where we may omit the term "w.r.t. $\mathbf{Q}$" in case $\mathbf{Q}$ is clear from the context.

Part 2: Showing $\mathbb{P}_{\mathbf{Q}}(\{i,j\}$ is ass. inc. in $\hat{E}_T) \leq \gamma'$ in case $|q_{i,j} - 1/2| > h$
A look at lines 8 and 10 of Alg. 15 reveals $\hat{E}_{t-1} \subseteq \hat{E}_t$ for all $t \leq T$ and moreover

$$(i,j) \in \hat{E}_T \quad \Leftrightarrow \quad T_{i,j} \leq T \text{ and } \mathbf{D}\left(\mathcal{A}^{i,j}_{\text{Coin}}\right) = \mathbf{D}\left(\mathcal{A}^{i,j}_{\text{Coin}}, T\right) = 0.$$

As $\mathcal{A}^{i,j}_{\text{Coin}}$ solves $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$, we infer in case $q_{i,j} < 1/2 - h$ that

$$\begin{aligned} \mathbb{P}_{\mathbf{Q}}\left(\{i,j\} \text{ is assigned incorrectly in } \hat{E}_T\right) &= \mathbb{P}_{\mathbf{Q}}((i,j) \in \hat{E}_T) \\ &\leq \mathbb{P}_{\mathbf{Q}}\left(\mathbf{D}\left(\mathcal{A}^{i,j}_{\text{Coin}}\right) = 0\right) \leq \gamma', \end{aligned}$$

and in case $q_{j,i} > 1/2 + h$ we similarly obtain

$$\mathbb{P}_{\mathbf{Q}}(\{i,j\} \text{ is assigned incorrectly in } \hat{E}_T) = \mathbb{P}_{\mathbf{Q}}((j,i) \in \hat{E}_T) \leq \gamma'.$$

This shows the assertion of Part 2. ■

Part 3: Bounding the type I error
In the following let $\iota_{\Delta i^*}$ for $i^* \in [m]$ and $\iota_{\diamondsuit}$ be defined as in Lem. 3.12. Suppose $i^* \in [m]$ and $\mathbf{Q} \in \mathcal{Q}^h_m(i^*)$. Part (c) of Lem. 3.12 yields the identity

$$\begin{aligned} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq i^*) &= \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \neg\text{CW or } \mathbf{D}(\mathcal{A}) = j \text{ for some } j \in [m] \setminus \{i^*\}) \\ &= \mathbb{P}_{\mathbf{Q}}(\hat{G}_T \in \mathcal{G}_m(\Delta i^*)) = \mathbb{P}_{\mathbf{Q}}(\iota_{\Delta i^*}(\hat{G}_T) \in \mathcal{G}_m(\Delta i^*)). \end{aligned} \tag{4.11}$$

For the sake of convenience, we write $\overline{G}$ for the set $\{\{i,j\} \,|\, (i,j) \in E_G \text{ or } (j,i) \in E_G\}$ for $G \in \mathcal{G}_m$. By Part 2 we have

$$\mathbb{P}_{\mathbf{Q}}\left(\{i,j\} \text{ is ass. inc. in } E_{\iota_{\Delta i^*}(\hat{G}_T)} \,\middle|\, \overline{\iota_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right) \leq \begin{cases} 0, & \text{if } \{i,j\} \notin \overline{G}, \\ \gamma', & \text{if } \{i,j\} \in \overline{G}, \end{cases}$$

for any $G \in \mathcal{G}_m$ with $\mathbb{P}_{\mathbf{Q}}\left(\overline{\iota_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right) > 0$. If no $\{i,j\} \in \overline{\iota_{\Delta i^*}(\hat{G}_T)}$ was assigned incorrectly in $E_{\iota_{\Delta i^*}(\hat{G}_T)}$, then $\iota_{\Delta i^*}(\hat{G}_T) \in \mathcal{G}_m(\Delta i^*)$ (i.e., in particular $i^* \neq \text{CW}(\iota_{\Delta i^*}(\hat{G}_T)))$ would imply $\text{CW}(\mathbf{Q}) \neq i^*$. Consequently, $\mathbf{Q} \in \mathcal{Q}^h_m(i^*)$ lets us infer that $\iota_{\Delta i^*}(\hat{G}_T) \in \mathcal{G}_m(\Delta i^*)$ is only possible if there exists some $\{i,j\} \in \overline{\iota_{\Delta i^*}(\hat{G}_T)}$, which is assigned incorrectly in $E_{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)}$. Regarding that $|\overline{G}| = |E_G|$, we thus get

$$\begin{aligned}
&\mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\Delta i^*}(\hat{G}_T) \in \mathcal{G}_m(\Delta i^*) \text{ and } \overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \mathbb{P}_{\mathbf{Q}}\left(\exists \{i,j\} \in \overline{G}, \text{ which is ass. inc. in } E_{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} \text{ and } \overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \sum\nolimits_{\{i,j\}\in\overline{G}} \mathbb{P}_{\mathbf{Q}}\left(\{i,j\} \text{ is ass. inc. in } E_{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} \text{ and } \overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \gamma' |E_G| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right)
\end{aligned}$$

for every $G \in \mathcal{G}_m$. Together with (4.11) and $\mathbb{P}_{\mathbf{Q}}\left(\left|E_{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)}\right| \leq m\right) = 1$, which holds according to (b) of Lem. 3.12, we infer

$$\begin{aligned}
&\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq i^*) = \mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\Delta i^*}(\hat{G}_T) \in \mathcal{G}_m(\Delta i^*)\right)\\
&= \sum\nolimits_{\overline{G}:G\in\mathcal{G}_m} \mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\Delta i^*}(\hat{G}_T) \in \mathcal{G}_m(\Delta i^*) \text{ and } \overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \sum\nolimits_{\overline{G}:G\in\mathcal{G}_m} \mathbb{P}_{\mathbf{Q}}\left(\exists \{i,j\} \in \overline{G}, \text{ which is ass. inc. in } E_{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} \text{ and } \overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \gamma' \sum\nolimits_{\overline{G}:G\in\mathcal{G}_m} |\overline{G}| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right) \leq \gamma' m \leq \alpha,
\end{aligned}$$

where we have used that $\sum_{\overline{G}:G\in\mathcal{G}_m} \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right) = 1$ holds trivially. ■

Part 4: Bounding the type II error
Now, we consider the case $\mathbf{Q} \in \mathcal{Q}_m^h(\neg\mathrm{CW})$. Similarly as above in Part 3, Lem. 3.12 yields

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq \neg\mathrm{CW}) = \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \in [m]) = \mathbb{P}_{\mathbf{Q}}(\hat{G}_T \in \mathcal{G}_m(\Diamond)). \tag{4.12}$$

Next, using $\mathfrak{l}_{\Diamond}$ as defined in Lem. 3.12, an analogue argumentation as above shows that $\mathfrak{l}_{\Diamond}(\hat{G}_T) \in \mathcal{G}_m(\Diamond)$ is only possible if there exists some $\{i,j\} \in \overline{\mathfrak{l}_{\Diamond}(\hat{G}_T)}$, which is assigned incorrectly in $E_{\mathfrak{l}_{\Diamond}(\hat{G}_T)}$. From this and Part 2 we can infer that

$$\mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\Diamond}(\hat{G}_T) \in \mathcal{G}_m(\Diamond) \text{ and } \overline{\mathfrak{l}_{\Diamond}(\hat{G}_T)} = \overline{G}\right) \leq \gamma' |E_G| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right) \tag{4.13}$$

holds for every $G \in \mathcal{G}_m$. According to Lem. 3.12(b) we have $\mathbb{P}_{\mathbf{Q}}\left(\left|E_{\mathfrak{l}_{\Diamond}(\hat{G}_T)}\right| \leq m-1\right) = 1$, hence combining (4.12) with (4.13) yields

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq \neg\mathrm{CW}) \leq \gamma' \sum\nolimits_{\overline{G}:G\in\mathcal{G}_m} |\overline{G}| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\Delta i^*}(\hat{G}_T)} = \overline{G}\right) \leq \gamma'(m-1) \leq \beta.$$

This completes the proof of Part 4 and also the proof of the theorem. □

4.4. Solutions to Other CW-Related Problems

In this section we discuss solutions to $\mathcal{P}_{\text{CWc}}^{m,h,\alpha,\beta}$ and $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}$. Suppose $h \in [0,1/2)$, $\gamma_0 \in (0,1)$ and $\alpha, \beta \in (0,\gamma_0)$ to be fixed. Denote by $\mathcal{A}^{\text{NTS}}$ the corresponding algorithm from Cor. 4.10 called with parameters h, α and β. Now, let $\widehat{\mathcal{A}}_1$ be the algorithm, which simulates $\mathcal{A}^{\text{NTS}}$, terminates as soon as $\mathcal{A}^{\text{NTS}}$ terminates, and outputs

$$\mathbf{D}(\widehat{\mathcal{A}}_1) = \begin{cases} \neg\text{CW}, & \text{if } \mathbf{D}(\mathcal{A}^{\text{NTS}}) = \neg\text{CW}, \\ \text{CW}, & \text{if } \mathbf{D}(\mathcal{A}^{\text{NTS}}) = i \text{ for some } i \in [m]. \end{cases}$$

Similarly, define $\widehat{\mathcal{A}}_2$ to be the algorithm, which, given any $i \in [m]$, simulates $\mathcal{A}^{\text{NTS}}$, terminates as soon as $\mathcal{A}^{\text{NTS}}$ terminates and then returns

$$\mathbf{D}(\widehat{\mathcal{A}}_2[i]) = \begin{cases} i, & \text{if } \mathbf{D}(\mathcal{A}^{\text{NTS}}) = i, \\ \neg i, & \text{otherwise.} \end{cases}$$

Since $\mathcal{A}^{\text{NTS}}$ solves $\mathcal{P}_{\text{CWt}}^{m,h,\alpha,\beta}$, it follows that $\widehat{\mathcal{A}}_1$ resp. $\widehat{\mathcal{A}}_2$ solves $\mathcal{P}_{\text{CWc}}^{m,h,\alpha,\beta}$ resp. $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}$. Moreover, both $\widehat{\mathcal{A}}_1$ and $\widehat{\mathcal{A}}_2$ have exactly the same runtime as $\mathcal{A}^{\text{NTS}}$. Consequently, we have with regard to Cor. 4.10 in case $h > 0$ with $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$

$$\sup_{\mathbf{Q} \in \mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}} \left[T^{\widehat{\mathcal{A}}_1} \right] \in \mathcal{O}\left(\frac{m \ln m}{h^2} \ln \frac{1}{\gamma'} \right)$$

and also

$$\sup_{\mathbf{Q} \in \mathcal{Q}_m^h} \max_{i \in [m]} \mathbb{E}_{\mathbf{Q}} \left[T^{\widehat{\mathcal{A}}_2[i]} \right] \in \mathcal{O}\left(\frac{m \ln m}{h^2} \ln \frac{1}{\gamma'} \right)$$

as $\max\{m, \frac{1}{h}, \frac{1}{\gamma'}\} \to \infty$, respectively. We conclude this section with an enhanced solution to $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}$.

Algorithm 17 $\mathcal{A}_{\text{CWv}}^{\text{NTS}}$: Noisy tournament sampling for $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}$

Input: $\alpha, \beta, h, \pi, i^*$, a testing algorithm $\mathcal{A}_{\text{Coin}} \in \mathfrak{A}_{\text{Coin}}$
Initialization: $\boldsymbol{n}_0, \boldsymbol{w}_0 \leftarrow (0)_{1 \le i,j \le m}$, $\hat{E}_0 \leftarrow \emptyset$, $\gamma' \leftarrow \min\{\alpha, \frac{\beta}{m-1}\}$

1: **for** $t \in \mathbb{N}$ **do**
2: Do steps 2–11 of Alg. 15
3: **if** $\hat{G}_t \in \mathcal{G}_m(i^*)$ **then return** i^*
4: **if** $\hat{G}_t \in \mathcal{G}_m(\neg i^*)$ **then return** $\neg i^*$

Theorem 4.12. *Let $\pi \in \Pi_\infty$, $h \in [0,1/2)$ and $\alpha, \beta \in (0,1)$ be fixed. Then, $\mathcal{A}_{\text{CWv}}^{\text{NTS}}$ (Alg. 17) called with parameters h, α, β, and π as the sampling strategy, solves $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}$.*

Proof. Let us abbreviate $\mathcal{A} := \mathcal{A}_{\text{CWv}}^{\text{NTS}}$, write $T := T^{\mathcal{A}}$ and suppose $\mathcal{A}$ is given a fixed $i^* \in [m]$ as input. For the sake of convenience we simply write $\mathcal{A}$ for $\mathcal{A}(i^*)$ in the following. Let $\mathbf{Q} \in \mathcal{Q}_m^h$ be arbitrary but fixed. Due to $\overline{\mathcal{G}}_m(i^*) \cup \overline{\mathcal{G}}_m(\neg i^*)$, we infer similarly as in the proof of Thm. 4.6 that $\mathcal{A}$ terminates almost surely. Moreover, using the notation from the proof of Thm. 4.6 we obtain also

$$\mathbb{P}_{\mathbf{Q}}\left(\{i,j\} \text{ is assigned incorrectly in } \hat{E}_T \right) \le \gamma'$$

for every $\{i,j\} \in [m]_2$ with $|q_{i,j} - 1/2| > h$.

To bound the type I error, suppose that $\mathbf{Q} \in \mathcal{Q}_m^h(i^*)$. Then, Lem. 3.12 ensures

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq i^*) = \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \neg i^*) = \mathbb{P}_{\mathbf{Q}}(\mathfrak{l}_{\neg i^*}(\hat{G}_T) \in \mathcal{G}_m(\neg i^*)). \tag{4.14}$$

The same argumentation as in the proof of Thm. 4.6 yields

$$\mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\neg i^*}(\hat{G}_T) \in \mathcal{G}_m(\neg i^*) \text{ and } \overline{\mathfrak{l}_{\neg i^*}(\hat{G}_T)} = \overline{G}\right) \leq \gamma' |E_G| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\neg i^*}(\hat{G}_T)} = \overline{G}\right)$$

for every $G \in \mathcal{G}_m$. Combining this with (4.14), using $\sum_{\overline{G}:G \in \mathcal{G}_m} \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\neg i^*}(\hat{G}_T)} = \overline{G}\right) \leq 1$ and the fact that $|E_{\mathfrak{l}_{\neg i^*}(\hat{G}_T)}| = 1$ holds a.s. (see (b) of Lem. 3.12) shows that

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq i^*) = \sum_{\overline{G}:G \in \mathcal{G}_m} \mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\neg i^*}(\hat{G}_T) \in \mathcal{G}_m(\neg i^*) \text{ and } \overline{\mathfrak{l}_{\neg i^*}(\hat{G}_T)} = \overline{G}\right) \leq \gamma'.$$

For showing the guarantee on the type II error, let $\mathbf{Q} \in \mathcal{Q}_m^h \setminus \mathcal{Q}_m^h(i^*)$ be fixed. Again, a similar argumentation as in the proof of Thm. 4.6 yields

$$\begin{aligned}\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq \neg i^*) &= \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = i^*) = \mathbb{P}_{\mathbf{Q}}(\hat{G}_T \in \mathcal{G}_m(i^*)) = \mathbb{P}_{\mathbf{Q}}(\mathfrak{l}_{i^*}(\hat{G}_T) \in \mathcal{G}_m(i^*)) \\ &\leq \gamma' \sum_{\overline{G}:G \in \mathcal{G}_m} |\overline{G}| \mathbb{P}_{\mathbf{Q}}(\mathfrak{l}_{i^*}(\hat{G}_T) = G) \leq (m-1)\gamma' \leq \beta,\end{aligned}$$

where we have used that $|E_{\mathfrak{l}_{i^*}(\hat{G}_T)}| \leq m-1$ holds a.s. according to (b) of Lem. 3.12. □

Similar to Thm. 4.9, one may obtain an instance-wise expected sample complexity bound of Alg. 17, when initialized with the SPRT from Prop. 2.17 as $\mathcal{A}_{\text{Coin}}$. The more restrictive problem $\mathcal{P}_{\text{CWv}}^{m,h,\gamma}(\text{CW})$ has also been considered by Urvoy et al. [2013] and Karnin [2016] for constructing an algorithmic solution to identify the CW. Note that an upper bound for $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}(\text{CW})$ is a direct consequence of the upper bound for $\mathcal{P}_{\text{CWv}}^{m,h,\alpha,\beta}$.

4.5. A Reduction to Pure Exploration Multi-Armed Bandits

Degenne and Koolen [2019] have proposed the Sticky Track-and-Stop algorithm for the setting of pure exploration bandits with multiple correct answers, which also covers our problems of interest. In this section we explicitly state Sticky Track-and-Stop for $\mathcal{P}_{\text{CWc}}^{m,h,\alpha,\beta}$ and $\mathcal{P}_{\text{CWt}}^{m,h,\alpha,\beta}$ and state and discuss its guarantees in these scenarios. For the sake of simplicity, we restrict ourselves to the symmetric case $\alpha = \beta =: \gamma$. Moreover, we omit the problems $\mathcal{P}_{\text{CWv}}^{m,h,\gamma}$, $\mathcal{P}_{\text{CWv}}^{m,h,\gamma}(\text{CW})$ and $\mathcal{P}_{\text{CWi}}^{m,h,\gamma}(\text{CW})$ here, because minor changes of the version for $\mathcal{P}_{\text{CWt}}^{m,h,\gamma}$ would result in corresponding solutions with similar guarantees. For the sake of convenience, we start with the easier problem $\mathcal{P}_{\text{CWc}}^{m,h,\gamma}$. Let us define

$$\Delta_{(m)_2} := \left\{(v_{i,j})_{1 \leq i < j \leq m} \in \mathbb{R}^{\binom{m}{2}} \,:\, \sum_{(i,j) \in (m)_2} v_{i,j} = 1 \text{ and } v_{i,j} \geq 0 \text{ for all } (i,j) \in (m)_2\right\}$$

and for any $\varepsilon > 0$ also

$$\Delta_{(m)_2}^{\varepsilon} := \left\{(v_{i,j})_{1 \leq i < j \leq m} \in \Delta_{(m)_2} \,:\, v_{i,j} \geq \varepsilon \text{ for all } (i,j) \in (m)_2\right\}.$$

Recall $\mathrm{kl}\,(p,q) = p\ln\frac{p}{q} + (1-p)\ln\frac{1-p}{1-q}$ from Sec. 2.5 and define for $\mathbf{v} = (v_{i,j})_{1\leq i<j\leq m} \in \Delta_{(m)_2}$ and $\mathbf{Q}, \mathbf{Q}' \in \mathcal{Q}_m$ the value

$$D(\mathbf{v}, \mathbf{Q}, \mathbf{Q}') := \sum\nolimits_{(i,j)\in(m)_2} v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right),$$

and for $\mathcal{Q}'_m \subseteq \mathcal{Q}_m$ let further

$$D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}'_m) := \inf\nolimits_{\mathbf{Q}'\in\mathcal{Q}'_m} D(\mathbf{v}, \mathbf{Q}, \mathbf{Q}').$$

4.5.1. Sticky Track-and-Stop for CW Checking

In the setting of $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWc}}$, the Sticky Track-and-Stop algorithm from Degenne and Koolen [2019] can be stated as Alg. 18. Note that Steps 2 and 11 of Alg. 18 are already computationally expensive, but the calculation of $D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}^h_m(\neg\mathrm{X}))$ in step 5 is even more involved, especially in the case $\mathrm{X} = \neg\mathrm{CW}$, because $\mathcal{Q}^h_m(\mathrm{CW})$ is non-convex[5]. Hence, the algorithm appears to be infeasible for practical applications to us.

For fixed $m \in \mathbb{N}$ and $h \in [0, 1/2)$, we define for any $\mathbf{Q} \in \mathcal{Q}_m$ the value

Algorithm 18 : Sticky Track-and-Stop for CW checking

Input: $\gamma \in (0,1)$, $h \in [0, 1/2)$, a sequence $(\varepsilon_t)_{t\in\mathbb{N}}$, functions $t \mapsto f(t)$ and $(t,\gamma) \mapsto \beta(t,\gamma)$
Initialization: $t \leftarrow 1$, $\hat{\boldsymbol{Q}}_0 \leftarrow (0)_{1\leq i<j\leq m}$, $\boldsymbol{n}_0 \leftarrow (0)_{1\leq i,j\leq m}$.

1: **while** True **do**
2: Let $\mathcal{C}_t \leftarrow \{\mathbf{Q}' \in \mathcal{Q}^h_m : D(\boldsymbol{n}_{t-1}/(t-1), \hat{\mathbf{Q}}_{t-1}, \mathbf{Q}') \leq \ln(f(t-1))\}$
3: Compute $I_t = \{\mathrm{X} \in \{\mathrm{CW}, \neg\mathrm{CW}\} \,|\, \exists \mathbf{Q}' \in \mathcal{Q}^h_m(\mathrm{X}) \cap \mathcal{C}_t\}$
4: Choose an element X from I_t, prefer CW over $\neg$CW
5: Compute that weight $\mathbf{v}_t \in \Delta_{(m)_2}$, which maximizes $D(\mathbf{v}_t, \hat{\mathbf{Q}}_{t-1}, \mathcal{Q}^h_m(\neg\mathrm{X}))$
6: Compute the projection $\mathbf{v}^{\varepsilon_t}_t$ of $\mathbf{v}_t$ onto $\Delta^{\varepsilon_t}_{(m)_2}$
7: Pull $(i,j) = \mathrm{argmin}_{(i',j')\in(m)_2} (\boldsymbol{n}_t)_{i',j'} - \sum_{s=1}^t (\mathbf{v}^{\varepsilon_s}_s)_{i',j'}$, observe $X_{i,j} \sim \mathrm{Ber}(q_{i,j})$
8: Update $\boldsymbol{w}_t$ via $(\boldsymbol{w}_t)_{k,l} \leftarrow (\boldsymbol{w}_{t-1})_{k,l} + \mathbf{1}_{\{\{k,l\}=\{i,j\} \text{ and } X_{k,l}=1\}}$ $\quad\forall 1 \leq k,l \leq m$
9: Update $\boldsymbol{n}_t$ via $(\boldsymbol{n}_t)_{k,l} \leftarrow (\boldsymbol{n}_{t-1})_{k,l} + \mathbf{1}_{\{\{k,l\}=\{i,j\}\}}$ $\quad\forall 1 \leq k,l \leq m$
10: Update $\hat{\mathbf{Q}}_t \leftarrow \frac{\boldsymbol{w}_t}{\boldsymbol{n}_t}$.
11: Let $\mathcal{D}_t \leftarrow \{\mathbf{Q}' \in \mathcal{Q}^h_m : D(\boldsymbol{n}_t/t, \hat{\mathbf{Q}}_t, \mathbf{Q}') \leq \beta(t,\gamma)\}$
12: **if** $\exists \mathrm{X} \in \{\mathrm{CW}, \neg\mathrm{CW}\}$ with $\mathcal{D}_t \cap \mathcal{Q}^h_m(\neg\mathrm{X}) = \emptyset$ **then**
13: **return** X
14: Update $t \leftarrow t+1$

$$D^{m,h}_{\mathrm{CWc}}(\mathbf{Q}) := \begin{cases} \sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}^h_m(\neg\mathrm{CW})), & \text{if } \mathbf{Q} \in \mathcal{Q}_m(\mathrm{CW}), \\ \sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}^h_m(\mathrm{CW})), & \text{if } \mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{CW}). \end{cases}$$

This characteristic plays a crucial role in the theoretical results proven by Degenne and Koolen [2019], which we will state and comment on below in Prop. 4.16. As a first step, we prove upper and lower bounds for $D^{m,h}_{\mathrm{CWc}}(\mathbf{Q})$. For this purpose, we will make use of the following lemma. It is taken from [Bubeck and Cesa-Bianchi, 2012] and is a mere consequence of Pinsker's theorem as well as the inequality $\ln x \leq x - 1$, which holds for all $x > 0$.

[5] It is the union of disjoint convex sets.

Lemma 4.13. *For $p, q \in [0,1]$ we have*

$$2(p-q)^2 \leq \mathrm{kl}\,(p,q) \leq \frac{(p-q)^2}{q(1-q)}.$$

Moreover, we will make use of the following result, which follows immediately by the definition of the CW.

Lemma 4.14. *Suppose $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{CW})$ with $i = \mathrm{CW}(\mathbf{Q})$, $j \in [m] \setminus \{i\}$ and let $\mathbf{Q}' = (q'_{i,j})_{1 \leq i,j \leq m}$ be defined via $q'_{i,j} = 1 - q_{i,j}$ and $q'_{i',j'} = q_{i',j'}$ for every $(i',j') \in \langle m \rangle_2 \setminus \{(i,j),(j,i)\}$. Then, either $j = \mathrm{CW}(\mathbf{Q}')$ or $\mathbf{Q}' \in \mathcal{Q}_m(\neg\mathrm{CW})$.*

We obtain the following lower and upper bounds for $D^{m,h}_{\mathrm{CWc}}(\mathbf{Q})$. In it, we restrict ourselves to the interesting case $m \geq 3$. Note that the factor $m-2$ in (i) in the following lemma is in accordance with the factor $m-2$ in our lower bound for solutions to $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWc}}$ from Thm. 4.1.

Lemma 4.15. *Let $h \in [0, 1/2)$ and $m \geq 3$ be fixed.*

(i) For any $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})$ we have

$$D^{m,h}_{\mathrm{CWc}}(\mathbf{Q}) \leq \frac{d_h(\mathbf{Q})}{m-2}$$

with $d_h(\mathbf{Q}) := \max_{(i,j)\in(m)_2} \max\{\mathrm{kl}\,(q_{i,j}, 1/2+h), \mathrm{kl}\,(q_{i,j}, 1/2-h)\}$.

(ii) For any $\tilde{h} \in [h, 1/2) \setminus \{0\}$ and any $\mathbf{Q} \in \mathcal{Q}^{\tilde{h}}_m$ we have

$$D^{m,h}_{\mathrm{CWc}}(\mathbf{Q}) \geq \frac{\mathrm{kl}\left(1/2+h, 1/2-\tilde{h}\right)}{m-1} \geq \frac{2(h+\tilde{h})^2}{m-1}. \tag{4.15}$$

Proof. (i) Let $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})$ be fixed, and suppose $\mathbf{v} \in \Delta_{(m)_2}$ to be fixed for the moment. Let $i := \mathrm{CW}(\mathbf{Q})$. By assumption on $\mathbf{v}$ there exists[6] some distinct $j', j'' \in [m] \setminus \{i\}$ with $\max\{v_{i,j'}, v_{i,j''}\} \leq \frac{1}{m-2}$. According to Lem. 4.14 we can choose $j \in \{j', j''\}$ such that $q_{j,k} < 1/2$ for at least one $k \in [m] \setminus \{i\}$. Thus, for arbitrarily small $\delta \in (0, 1/2-h)$, $\mathbf{Q}' \in \mathcal{Q}^h_m$ defined via

$$q'_{r,s} := \begin{cases} 1/2 - (h+\delta), & \text{if } (r,s) = (i,j), \\ 1/2 + h + \delta, & \text{if } (r,s) = (j,i), \\ q_{r,s}, & \text{otherwise,} \end{cases}$$

for each $(r,s) \in (m)_2$, fulfills $\mathbf{Q}' \in \mathcal{Q}^h_m(\neg\mathrm{CW})$. As $q_{i,j} > 1/2+h$ holds by assumption on $\mathbf{Q}$ and i, the definition of $\mathbf{Q}'$ assures

$$\sum\nolimits_{(r,s)\in(m)_2} v_{r,s}\mathrm{kl}\left(q_{r,s}, q'_{r,s}\right) \leq v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right) \leq \frac{\mathrm{kl}\,(q_{i,j}, 1/2-(h+\delta))}{m-2}.$$

[6] Indeed, as $\sum_{j' \neq i} v_{i,j'} \leq 1$ one can choose $j' \neq i$ with $v_{i,j'} \leq \frac{1}{m-1}$. Now, $\sum_{j'' \notin \{i,j'\}} v_{i,j''} \leq 1$ allows us to choose $j'' \in [m] \setminus \{i, j'\}$ with $v_{i,j''} \leq \frac{1}{m-2}$. Then, $\max\{v_{i,j'}, v_{i,j''}\} \leq \frac{1}{m-2}$ holds.

Regarding $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\text{CW})$ and that this estimate is obtained for any $\mathbf{v} \in \Delta_{(m)_2}$, we can conclude that

$$D_{\text{CWc}}^{m,h}(\mathbf{Q}) = \sup_{\mathbf{v}\in\Delta_{(m)_2}} \inf_{\mathbf{Q}'\in\mathcal{Q}_m^h(\neg\text{CW})} \sum\nolimits_{(r,s)\in(m)_2} v_{r,s}\text{kl}\left(q_{r,s}, q'_{r,s}\right) \leq \frac{\text{kl}\left(q_{i,j}, 1/2-(h+\delta)\right)}{m-2}.$$

Taking the limit $\delta \searrow 0$ yields

$$D_{\text{CWc}}^{m,h}(\mathbf{Q}) \leq \frac{\text{kl}\left(q_{i,j}, 1/2-h\right)}{m-2} \leq \frac{d_h(\mathbf{Q})}{m-2}.$$

(ii) Let $\tilde{h} \in [h, 1/2) \setminus \{0\}$ and $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}$ be fixed. We distinguish two cases.
Case 1: $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}(\text{CW})$. Let $i := \text{CW}(\mathbf{Q})$ and define $\mathbf{v} = (v_{r,s})_{(r,s)\in(m)_2} \in \Delta_{(m)_2}$ via $v_{r,s} = \mathbf{1}_{\{i\in\{r,s\}\}}(m-1)^{-1}$ for each $(r,s) \in (m)_2$. For any $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\text{CW})$ there exists some $j \in [m] \setminus \{i\}$ with $q'_{i,j} < 1/2 - h$, as otherwise $\text{CW}(\mathbf{Q}') = i$ would hold. But by assumption on i, $q_{i,j} > 1/2 + \tilde{h}$ holds, hence we can estimate with Lem. 4.13 that

$$\sum\nolimits_{(r,s)\in(m)_2} v_{r,s}\text{kl}\left(q_{r,s}, q'_{r,s}\right) \geq v_{i,j}\text{kl}\left(q_{i,j}, q'_{i,j}\right) \geq \frac{\text{kl}\left(1/2+\tilde{h}, 1/2-h\right)}{m-1} \geq \frac{2(h+\tilde{h})^2}{m-1}.$$

As this holds for arbitrary $\mathbf{Q}'$, (4.15) follows.
Case 2: $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}(\neg\text{CW})$. For every $i \in [m]$, $\text{CW}(\mathbf{Q}) \neq i$ implies the existence of some $j(i) \in [m] \setminus \{i\}$ with $q_{i,j(i)} < 1/2 - \tilde{h}$. Now, choose $\mathbf{v} = (v_{r,s})_{(r,s)\in(m)_2} \in \Delta_{(m)_2}$ such that

$$v_{r,s} = \begin{cases} \frac{1}{m}, & \text{if } (r,s) \in \{(i,j(i)),(j(i),i)\} \text{ for some } i \in [m], \\ 0, & \text{otherwise,} \end{cases}$$

for any $(r,s) \in (m)_2$. Let $\mathbf{Q}' \in \mathcal{Q}_m^h(\text{CW})$ be arbitrary and write $i' = \text{CW}(\mathbf{Q}')$. Then, $q'_{i',j(i')} > 1/2 + h$ but at the same time $q_{i',j(i')} < 1/2 - \tilde{h}$ holds. Therefore, assuming for convenience w.l.o.g.[7] $i' < j(i')$, we obtain again with the help of Lem. 4.13 the estimate

$$\sum\nolimits_{(r,s)\in(m)_2} v_{r,s}\text{kl}\left(q_{r,s}, q'_{r,s}\right) \geq v_{i',j(i')}\text{kl}\left(q_{i',j(i')}, q'_{i',j(i')}\right) \geq \frac{\text{kl}\left(1/2+h, 1/2-\tilde{h}\right)}{m} \geq \frac{2(h+\tilde{h})^2}{m}.$$

As $\mathbf{Q}'$ was arbitrary, we obtain (4.15).

□

Equipped with these results, we show the following:

[7] In case $i' > j(i')$ estimate the following sum by $v_{j(i'),i'}\text{kl}\left(q_{j(i'),i'}, q'_{j(i'),i'}\right)$ and argue analogously.

Proposition 4.16. *(i) Let $h \in [0, 1/2)$ and $m \in \mathbb{N}_{\geq 3}$ be fixed. If $\mathcal{A}(\gamma)$ solves $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWc}}$ for any $\gamma > 0$, then*

$$\liminf_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} \geq \frac{1}{D^{m,h}_{\mathrm{CWc}}(\mathbf{Q})}$$

for any $\mathbf{Q} \in \mathcal{Q}^h_m$, and consequently in case $h > 0$

$$\sup_{\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{CW})} \liminf_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln \frac{1}{\gamma}} \geq \frac{(m-2)(1/4 - h^2)}{4h^2},$$

and in case $h = 0$ for every $\tilde{h} > 0$

$$\sup_{\mathbf{Q} \in \mathcal{Q}^{\tilde{h}}_m(\mathrm{CW})} \liminf_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln \frac{1}{\gamma}} \geq \frac{m-2}{4\tilde{h}^2}.$$

(ii) Let $m \in \mathbb{N}$ and $C > 0$ s.t. $C \geq e \sum_{t=1}^{\infty} t^{-2} \left(e/\binom{m}{2}\right)^{\binom{m}{2}} \left(\ln^2(Ct^2)\ln(t)\right)^{\binom{m}{2}}$, suppose $h \in [0, 1/2)$ and

$$\varepsilon_t := \frac{1}{2}\left(\binom{m}{2}^2 + t\right)^{-\frac{1}{2}}, \quad f(t) := Ct^{10} \quad \text{and} \quad \beta(t, \gamma) := \ln \frac{Ct^2}{\gamma}.$$

Write $\mathcal{A}(\gamma)$ for Alg. 18 called with parameters $\gamma, h, (\varepsilon_t)_t, f$ and β. Then, $\mathcal{A}(\gamma)$ solves $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWc}}$ and fulfills

$$\lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} = \frac{1}{D^{m,h}_{\mathrm{CWc}}(\mathbf{Q})}$$

for any $\mathbf{Q} \in \mathcal{Q}^h_m$. In particular, we have in case $h > 0$

$$\sup_{\mathbf{Q} \in \mathcal{Q}^h_m} \lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln \frac{1}{\gamma}} \leq \frac{m}{8h^2}$$

and in case $h = 0$ for any $\tilde{h} > 0$

$$\sup_{\mathbf{Q} \in \mathcal{Q}^{\tilde{h}}_m} \lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln \frac{1}{\gamma}} \leq \frac{m}{2\tilde{h}^2}.$$

Proof. (i) The first statement corresponds to Thm. 1 in [Degenne and Koolen, 2019]. Next, suppose $h > 0$ and let $\mathbf{Q} = \mathbf{Q}(\delta) \in \mathcal{Q}^h_m(\mathrm{CW})$ be such that $q_{i,j} \in \{1/2 \pm (h+\delta)\}$ for all $(i,j) \in (m)_2$ and some arbitrarily small $\delta > 0$. It holds that

$$\begin{aligned} d_h(\mathbf{Q}) &= \max_{(i,j)\in(m)_2} \max\{\mathrm{kl}\,(q_{i,j}, 1/2 + h), \mathrm{kl}\,(q_{i,j}, 1/2 - h)\} \\ &= \mathrm{kl}\,(1/2 + (h+\delta), 1/2 - h) \leq \frac{4(h+\delta/2)^2}{1/4 - h^2}, \end{aligned}$$

where we have used Lem. 4.13 in the last step. Thus, the second statement follows from part (i) of Lem. 4.15 by taking the limit $\delta \searrow 0$.

Now, consider the case $h = 0$. Let $\tilde{h} \in (0, 1/2)$ and $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}$ with entries $q_{i,j} \in \{1/2 \pm (\tilde{h} + \delta)\}$ for all $(i, j) \in (m)_2$ and some $\delta \in (0, 1/2 - h)$. As

$$d_0(\mathbf{Q}) = \max_{(i,j)\in(m)_2} \mathrm{kl}\,(q_{i,j}, 1/2) = \mathrm{kl}\left(1/2 + (\tilde{h} + \delta), 1/2\right) \leq 4(\tilde{h} + \delta)^2$$

is assured by Lem. 4.13, the statement thus follows again from part (i) of Lem. 4.15 via taking the limit $\delta \searrow 0$.

(ii) Thm. 11 in [Degenne and Koolen, 2019] implies the first statement. For the choice of ε_t confer p. 7 in [Garivier and Kaufmann, 2016], for $f(t)$ see Lem. 14 on p. 9 in [Degenne and Koolen, 2019] and for $\beta(t, \gamma)$ see Thm. 10 on p. 6 in [Degenne and Koolen, 2019]. The second statement follows directly from the bound on $D_{\mathrm{CWc}}^{m,h}(\mathbf{Q})$ stated in Lem. 4.15.

□

4.5.2. Sticky Track-and-Stop for CW Testification

Define $\mathcal{I} \coloneqq \{\neg\mathrm{CW}, 1, \ldots, m\}$ and recall that, for any $h \in [0, 1/2)$, $\mathcal{Q}_m^h = \bigcup_{\mathrm{X}\in\mathcal{I}} \mathcal{Q}_m^h(\mathrm{X})$ is a disjoint union. We endow $\mathcal{I}$ with the ordering $\succ_{\mathcal{I}}$ defined[8] via $1 \succ_{\mathcal{I}} 2 \succ_{\mathcal{I}} \cdots \succ_{\mathcal{I}} m \succ_{\mathcal{I}} \neg\mathrm{CW}$; This way, choosing, e.g., an element from $\{2, 3, \neg\mathrm{CW}\} \subset \mathcal{I}$ according to $\succ_{\mathcal{I}}$ means to choose 2. Let $\Delta_{(m)_2}$ and $\Delta_{(m)_2}^{\varepsilon}$ be defined as above. For $\mathbf{v} = (v_{i,j})_{1\leq i<j\leq m} \in \Delta_{(m)_2}$, $\mathbf{Q}, \mathbf{Q'} \in \mathcal{Q}_m$ and $\mathcal{Q}'_m \subseteq \mathcal{Q}_m$ recall the definitions of $D(\mathbf{v}, \mathbf{Q}, \mathbf{Q'})$ and $D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}'_m)$ from Sec. 4.5.1. For $\mathbf{Q} \in \mathcal{Q}_m$ define

$$i_F(\mathbf{Q}) \coloneqq \mathrm{argmax}_{\mathrm{X'}\in\mathcal{I}} \left(\mathrm{X'} \mapsto \sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}_m^h(\neg\mathrm{X'}))\right).$$

and note that $(\neg\mathrm{CW}) \notin i_F(\mathbf{Q})$ whenever $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$ holds[9].

In the setting of $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$, the Sticky Track-and-Stop algorithm from Degenne and Koolen [2019] can be stated as Alg. 19. Steps 2 and 11 of Alg. 19 are the same as in Alg. 18 and thus similarly computationally very expensive, and analogously step 5 is expensive, in particular if $\mathrm{X} = \neg\mathrm{CW}$. Moreover, as there are $m + 1$ possible answers (namely $1, \ldots, m, \neg\mathrm{CW}$) for $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$ whereas $\mathcal{P}_{\mathrm{CWc}}^{m,h,\gamma}$ is a problem with a binary outcome (CW or $\neg$CW), Step 3 in Alg. 19 is far more complex than the corresponding step in Alg. 18. This step requires to calculate for each $\mathbf{Q} \in \mathcal{C}_t$ the set $i_F(\mathbf{Q}) \subseteq \mathcal{I}$, which is the set of maximizers of $\mathrm{X'} \mapsto \sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}_m^h(\neg\mathrm{X'}))$. Finding $i_F(\mathbf{Q})$ for one fixed $\mathbf{Q}$ already requires the solution of a difficult min-max problem; doing this for any $\mathbf{Q} \in \mathcal{C}_t$ is seemingly infeasible. This indicates that Alg. 19 is computationally even far more complex than Alg. 18.

To analyze its theoretical performance, define for $m \in \mathbb{N}, h \in [0, 1/2)$, $\mathrm{X} \in \mathcal{I}$ and any $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{X})$ the value

$$\begin{aligned} D_{\mathrm{CWt}}^{m,h}(\mathbf{Q}) &\coloneqq \sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}_m^h(\neg\mathrm{X})) \\ &= \sup_{\mathbf{v}\in\Delta_{(m)_2}} \inf_{\mathbf{Q'}\in\mathcal{Q}_m^h(\neg\mathrm{X})} \sum_{(r,s)\in(m)_2} v_{r,s} \mathrm{kl}\left(q_{r,s}, q'_{r,s}\right). \end{aligned}$$

[8] Here, we merely have to choose *any* fixed ordering on $\mathcal{I}$, which one is not of importance.

[9] In fact, if $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$ with $\mathrm{CW}(\mathbf{Q}) = k$, then $D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}_m^h(\neg k)) > 0 = D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}_m^h(\mathrm{CW}))$.

Algorithm 19 : STICKY TRACK-AND-STOP for CW testification

Input: $\gamma \in (0,1)$, $h \in [0,1/2)$, a sequence $(\varepsilon_t)_{t\in\mathbb{N}}$, functions $t \mapsto f(t)$ and $(t,\gamma) \mapsto \beta(t,\gamma)$

Initialization: $t \leftarrow 1$, $\hat{\boldsymbol{Q}}_0 \leftarrow (0)_{1\le i<j\le m}$, $\boldsymbol{n}_0 \leftarrow (0)_{1\le i,j\le m}$.

while True **do**
  Let $\mathcal{C}_t \leftarrow \{\mathbf{Q}' \in \mathcal{Q}_m^h : D(\boldsymbol{n}_{t-1}/(t-1), \hat{\mathbf{Q}}_{t-1}, \mathbf{Q}') \le \ln(f(t-1))\}$
  Let $I_t = \bigcup_{\mathbf{Q}'\in\mathcal{C}_t} i_F(\mathbf{Q}')$
  Choose an element X from I_t according to $\succ_{\mathcal{I}}$
  Compute that weight $\mathbf{v}_t \in \Delta_{(m)_2}$, which maximizes $D(\mathbf{v}_t, \hat{\mathbf{Q}}_{t-1}, \mathcal{Q}_m^h(\neg \mathrm{X}))$
  Compute the projection $\mathbf{v}_t^{\varepsilon_t}$ of $\mathbf{v}_t$ onto $\Delta_{(m)_2}^{\varepsilon_t}$
  Pull $(i,j) = \operatorname{argmin}_{(i',j')\in(m)_2} (\boldsymbol{n}_t)_{i',j'} - \sum_{s=1}^t (\mathbf{v}_s^{\varepsilon_s})_{i',j'}$, observe $X_{i,j} \sim \mathrm{Ber}(q_{i,j})$
  Update $\boldsymbol{w}_t$ via $(\boldsymbol{w}_t)_{k,l} \leftarrow (\boldsymbol{w}_{t-1})_{k,l} + \mathbf{1}_{\{\{k,l\}=\{i,j\} \text{ and } X_{k,l}=1\}}$ $\forall 1 \le k,l \le m$
  Update $\boldsymbol{n}_t$ via $(\boldsymbol{n}_t)_{k,l} \leftarrow (\boldsymbol{n}_{t-1})_{k,l} + \mathbf{1}_{\{\{k,l\}=\{i,j\}\}}$ $\forall 1 \le k,l \le m$
  Update $\hat{\mathbf{Q}}_t \leftarrow \frac{\boldsymbol{w}_t}{\boldsymbol{n}_t}$.
  Let $\mathcal{D}_t \leftarrow \{\mathbf{Q}' \in \mathcal{Q}_m^h : D(\boldsymbol{n}_t/t, \hat{\mathbf{Q}}_t, \mathbf{Q}') \le \beta(t,\gamma)\}$
  if $\exists \mathrm{X} \in \mathcal{I}$ with $\mathcal{D}_t \cap \mathcal{Q}_m^h(\neg\mathrm{X}) = \emptyset$ **then**
    return X
  Update $t \leftarrow t+1$

As $\mathcal{Q}_m^h = \bigcup_{\mathrm{X}\in\mathcal{I}} \mathcal{Q}_m^h(\mathrm{X})$ is a disjoint union, $D_{\mathrm{CWt}}^{m,h}(\mathbf{Q})$ is well-defined for any $\mathbf{Q} \in \mathcal{Q}_m^h$. Similarly as in Lem. 4.15 we obtain the following result. Therein, the term $m-1$ is in accordance to our lower bound from Thm. 4.2.

Lemma 4.17. *Let $h \in [0,1/2)$ be fixed.*

(i) For any $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$ we have

$$D_{\mathrm{CWt}}^{m,h}(\mathbf{Q}) \le \frac{d_h(\mathbf{Q})}{m-1}$$

with $d_h(\mathbf{Q}) := \max_{(i,j)\in(m)_2} \max\{\mathrm{kl}\,(q_{i,j}, 1/2+h), \mathrm{kl}\,(q_{i,j}, 1/2-h)\}$.

(ii) For any $\tilde{h} \in [h,1/2)\setminus\{0\}$ and any $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}$ we have

$$D_{\mathrm{CWt}}^{m,h}(\mathbf{Q}) \ge \frac{2(h+\tilde{h})^2}{m}$$

and in case $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}(\mathrm{CW})$ we even obtain $D_{\mathrm{CWt}}^{m,h}(\mathbf{Q}) \ge \frac{2(h+\tilde{h})^2}{m-1}$.

Proof. (i) Let $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{CW})$ and $\mathbf{v} \in \Delta_{(m)_2}$ be fixed for the moment. Let $i := \mathrm{CW}(\mathbf{Q})$. By assumption on $\mathbf{v}$ there exists some $j \in [m]\setminus\{i\}$ with $v_{i,j} \le 1/(m-1)$. For arbitrary small but fixed $\delta \in (0, 1/2-h)$ define $\mathbf{Q}' \in \mathcal{Q}_m^h$ via

$$q'_{r,s} := \begin{cases} 1/2-(h+\delta), & \text{if } (r,s)=(i,j), \\ 1/2+h+\delta, & \text{if } (r,s)=(j,i), \\ q_{r,s}, & \text{otherwise,} \end{cases}$$

for each $(r,s) \in (m)_2$. Then, $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg i)$ due to $q'_{i,j} < 1/2-h$. As $q_{i,j} > 1/2+h$ holds by assumption on $\mathbf{Q}$ and i, the definition of $\mathbf{Q}'$ assures

$$\sum\nolimits_{(r,s)\in(m)_2} v_{r,s}\mathrm{kl}\left(q_{r,s}, q'_{r,s}\right) \le v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right) \le \frac{\mathrm{kl}\,(q_{i,j}, 1/2-(h+\delta))}{m-1}.$$

Regarding $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg i)$ and that this estimate is obtained for any $\mathbf{v} \in \Delta_{(m)_2}$, we can conclude (taking into account that $\mathbf{Q} \in \mathcal{Q}_m(i)$) that

$$\begin{aligned} D^{m,h}_{\mathrm{CWt}}(\mathbf{Q}) &= \sup\nolimits_{\mathbf{v}\in\Delta_{(m)_2}} \inf\nolimits_{\mathbf{Q}\in\mathcal{Q}_m^h(\neg i)} \sum\nolimits_{(r,s)\in(m)_2} v_{r,s}\mathrm{kl}\left(q_{r,s}, q'_{r,s}\right) \\ &\leq v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right) \leq \frac{\mathrm{kl}\left(q_{i,j}, 1/2-(h+\delta)\right)}{m-1} \end{aligned}$$

and taking the limit $\delta \searrow 0$ yields

$$D^{m,h}_{\mathrm{CWt}}(\mathbf{Q}) \leq \frac{\mathrm{kl}\left(q_{i,j}, 1/2-h\right)}{m-1} \leq \frac{d_h(\mathbf{Q})}{m-1}.$$

(ii) Let $\tilde{h} \in [h, 1/2) \setminus \{0\}$, $\mathrm{X} \in \mathcal{I}$ and $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}(\mathrm{X})$ be arbitrary but fixed. In case $\mathrm{X} = \neg\mathrm{CW}$ we have due to $\mathcal{Q}_m^h(\neg(\neg\mathrm{CW})) = \mathcal{Q}_m^h(\mathrm{CW})$ the equality $D^{m,h}_{\mathrm{CWc}}(\mathbf{Q}) = D^{m,h}_{\mathrm{CWt}}(\mathbf{Q})$, hence $D^{m,h}_{\mathrm{CWt}}(\mathbf{Q}) \geq \frac{2(h+\tilde{h})^2}{m}$ follows from Lem. 4.15.
Now, consider the case $\mathrm{X} = i \in \{1,\ldots,m\}$. Define $\mathbf{v} = (v_{r,s})_{(r,s)\in(m)_2} \in \Delta_{(m)_2}$ via $v_{r,s} = \mathbf{1}_{\{i\in\{r,s\}\}}(m-1)^{-1}$ for each $(r,s) \in (m)_2$. For arbitrary $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg i)$ there exists some $j \in [m]\setminus\{i\}$ with $q'_{i,j} < 1/2 - h$, as otherwise $\mathrm{CW}(\mathbf{Q}') = i$ would hold. But by assumption on i, $q_{i,j} > 1/2 + \tilde{h}$ holds, hence we can estimate with Lem. 4.13 that

$$\begin{aligned} \sum\nolimits_{(r,s)\in(m)_2} v_{r,s}\mathrm{kl}\left(q_{r,s}, q'_{r,s}\right) &\geq v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right) \\ &\geq \frac{\mathrm{kl}\left(1/2+h, 1/2-\tilde{h}\right)}{m-1} \geq \frac{2(h+\tilde{h})^2}{m-1}. \end{aligned}$$

As $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg i)$ was arbitrary, we get $D^{m,h}_{\mathrm{CWt}}(\mathbf{Q}) \geq \frac{2(h+\tilde{h})^2}{m-1} > \frac{2(h+\tilde{h})^2}{m}$.

□

With this, we obtain the following result, which is an analogon to Prop. 4.16.

Proposition 4.18. *(i) Let $h \in [0,1/2)$ and $m \in \mathbb{N}$ be fixed. If $\mathcal{A}(\gamma)$ solves $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWt}}$ for any $\gamma > 0$, then*

$$\liminf_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln\frac{1}{\gamma}} \geq \frac{1}{D^{m,h}_{\mathrm{CWt}}(\mathbf{Q})}$$

for any $\mathbf{Q} \in \mathcal{Q}_m^h$. In particular, we have in case $h > 0$

$$\sup\nolimits_{\mathbf{Q}\in\mathcal{Q}_m^h(\mathrm{CW})} \liminf_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln\frac{1}{\gamma}} \geq \frac{(m-1)(1/4-h^2)}{4h^2}$$

and in case $h = 0$ for any $\tilde{h} > 0$

$$\sup\nolimits_{\mathbf{Q}\in\mathcal{Q}_m^{\tilde{h}}(\mathrm{CW})} \liminf_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln\frac{1}{\gamma}} \geq \frac{m-1}{4\tilde{h}^2}.$$

(ii) Let $h \in [0, 1/2)$ and $m \in \mathbb{N}$ be fixed. Choose $C > 0$, $(\varepsilon_t)_{t\in\mathbb{N}}$, $t \mapsto f(t)$ and $(t, \gamma) \mapsto \beta(t, \gamma)$ as in Prop. 4.16. Write $\mathcal{A}(\gamma)$ for Alg. 19 called with parameters $\gamma, h, (\varepsilon_t)_t, f$ and β. Then, $\mathcal{A}(\gamma)$ solves $\mathcal{P}^{m,h,\gamma}_{\mathrm{CWt}}$ and fulfills

$$\lim_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} = \frac{1}{D^{m,h}_{\mathrm{CWt}}(\mathbf{Q})}$$

for any $\mathbf{Q} \in \mathcal{Q}^h_m$. In particular, we have in case $h > 0$

$$\sup_{\mathbf{Q}\in\mathcal{Q}^h_m} \lim_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} \leq \frac{m}{8h^2}$$

and in case $h = 0$ for any $\tilde{h} > 0$

$$\sup_{\mathbf{Q}\in\mathcal{Q}^{\tilde{h}}_m} \lim_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} \leq \frac{m}{2\tilde{h}^2}.$$

Proof. (i) The proof is essentially the same as that of Prop. 4.16. In fact, the statements can be seen via following the lines of the latter one and using part (i) of Lem. 4.17 instead of part (i) of Lem. 4.15.

(ii) Thm. 11 in [Degenne and Koolen, 2019] implies the first statement. For the choice of ε_t confer p. 7 in [Garivier and Kaufmann, 2016], for $f(t)$ see Lem. 14 on p.9 in [Degenne and Koolen, 2019] and for $\beta(t, \gamma)$ see Thm. 10 on p.8 in [Degenne and Koolen, 2019]. The second statement follows directly from the bound on $D^{m,h}_{\mathrm{CWt}}(\mathbf{Q})$ stated in part (ii) of Lem. 4.17.

□

4.6. Empirical Evaluation

In this section, we present an experimental study to illustrate the performance of our algorithm $\mathcal{A}^{\mathrm{NTS}}$ for CW testification. Throughout these experiments, we denote by $\mathcal{A}^{\mathrm{NTS}}$ (or simply NTS) Alg. 16 initiated as in Cor. 4.10 with Alg. 10 as $\mathcal{A}_{\mathrm{DSTA}}$ and with the SPRT $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{Coin}}(h, \gamma')$ from Prop. 2.17 as $\mathcal{A}_{\mathrm{Coin}}$, and parameters m, h and $\alpha = \beta = \gamma$. Prop. 3.14 and Prop. 2.17 indicate optimality of this choices of $\mathcal{A}_{\mathrm{DSTA}}$ and $\mathcal{A}_{\mathrm{Coin}}$. Moreover, we write $\mathcal{A}^{\mathrm{NTS}}_{\mathrm{PPR}}$ for Alg. 16 initiated with Alg. 10 as $\mathcal{A}_{\mathrm{DSTA}}$ and with with PPR-Bernoulli (Alg. 4) as $\mathcal{A}_{\mathrm{Coin}}$, and parameters m, h and $\alpha = \beta = \gamma$. Since PPR-Bernoulli solves $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$, Thm. 4.6 assures us that $\mathcal{A}^{\mathrm{NTS}}_{\mathrm{PPR}}$ solves $\mathcal{P}^{m,0,\gamma}_{\mathrm{CWt}}$. Throughout this section, all experiments, which involved a variation of γ, were conducted with the values 0.001, 0.005, 0.01, 0.015, 0.02, 0.03, 0.05, 0.075, 0.1, 0.125, 0.15, 0.2, 0.25, 0.35, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.95 and 0.99 for γ.

4.6.1. The Active Scenario

We start with a comparison between $\mathcal{A}^{\mathrm{NTS}}$ and SELECT-then-verify (StV). For this, we sample uniformly at random relations $\mathbf{Q}$ from $\mathcal{Q}^{0.05}_m$, execute $\mathcal{A}^{\mathrm{NTS}}$ and StV on this instance with the same parameters h and γ and report the average termination times and observed accuracies. To guarantee stability of the results, we average over 25000 runs for $m = 5$ and over 100000 runs for $m \in \{8, 10\}$. Since $\mathbf{Q}$ is usually unknown in practice, we try out and compare different values of h and γ as parameters for $\mathcal{A}^{\mathrm{NTS}}$ and StV, which causes a variation in the number of iterations before the algorithms terminate. These free parameters cause a variation in the used confidence bounds and thus in the number of iterations before the algorithms terminate.

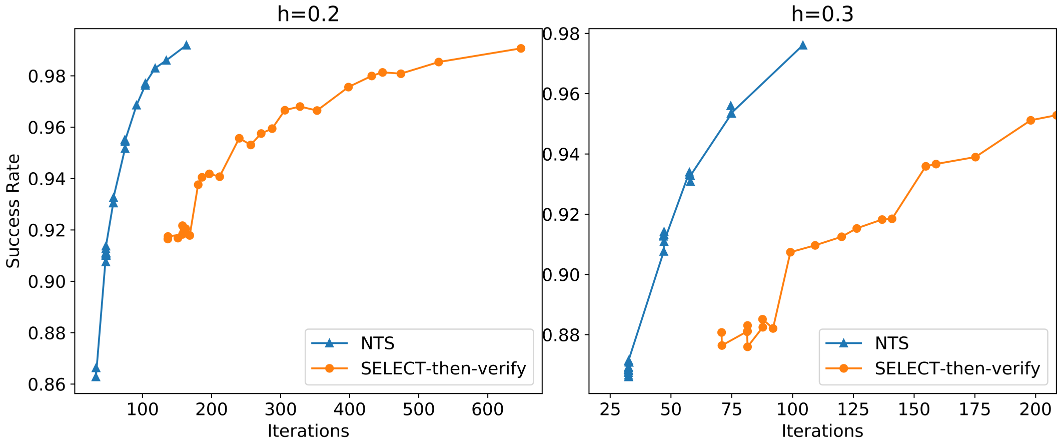


Figure 5.: Success rate and total number of comparisons until termination for the proposed $\mathcal{A}^{\mathrm{NTS}}$ and StV for different values of the gap h to 1/2.

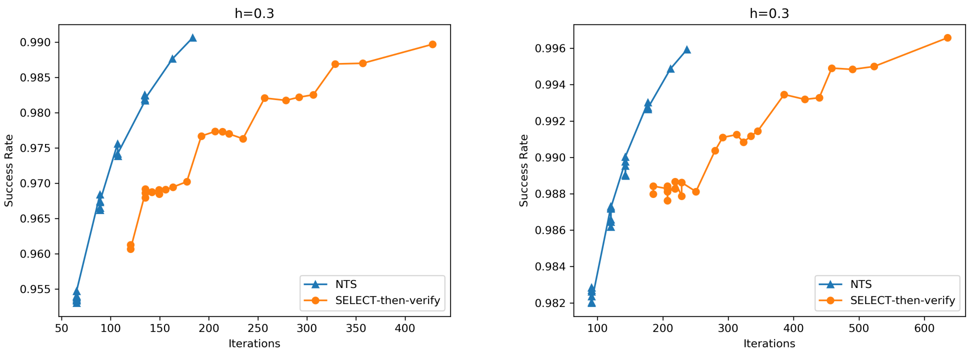


Figure 6.: Accuracy and termination time of $\mathcal{A}^{\mathrm{NTS}}$ and StV for 8 arms (on the left) and 10 arms (on the right).

The curves in Fig. 5 and Fig. 6 have been produced through variation of the parameter γ and illustrate the compromise between the success rate and the number of iterations of the algorithms (decreasing γ increases the success rate but also the sample complexity). As can be seen, the curves of $\mathcal{A}^{\mathrm{NTS}}$ dominate the curves of StV for any considered choices of m and h. Indeed, with the same number of comparisons, $\mathcal{A}^{\mathrm{NTS}}$ achieves a higher success rate than StV, regardless of m and h, and in fact $\mathcal{A}^{\mathrm{NTS}}$ seems to be quite robust towards incorrect choices of this parameter.

Fig. 6 illustrates the effect of increasing the number of arms on the success rate and termination time of both algorithms. It is clearly visible that the larger the value of m, the larger both the success rate and the termination times of both algorithms.

Table 4.1.: Experimental results for $m = 20, \gamma = 0.05, h = 0.05$, $N = 100$ and varying h'.

	$T^{\mathcal{A}}$		Accuracy	
h'	$\mathcal{A}^{\mathrm{NTS}}$	StV	$\mathcal{A}^{\mathrm{NTS}}$	StV
0.45	**887** (4.1)	21751 (42.4)	1.00	1.00
0.40	**980** (5.4)	21730 (43.7)	1.00	1.00
0.35	**1085** (6.5)	21707 (48.1)	1.00	1.00
0.30	**1262** (8.5)	21773 (47.0)	1.00	1.00
0.25	**1447** (10.8)	21730 (52.3)	1.00	1.00
0.20	**1798** (15.6)	21832 (47.7)	1.00	1.00
0.15	**2331** (21.8)	21870 (51.0)	1.00	1.00
0.10	**3383** (29.3)	22096 (50.8)	1.00	1.00
0.05	**6607** (88.6)	22544 (92.3)	1.00	1.00
0.02	**14155** (234.6)	23567 (167.9)	1.00	1.00

To further compare $\mathcal{A}^{\mathrm{NTS}}$ with StV, we conduct the following experiment: We fix $m \in \mathbb{N}$, $\gamma \in (0, 1/2)$ and $h \in (0, 1/2)$ in advance, sample relations $\mathbf{Q}_1, \dots, \mathbf{Q}_N$ uniformly at random from $\widehat{\mathcal{Q}}_m^{h'} := \{\mathbf{Q} \in \mathcal{Q}_m \,|\, q_{i,j} \in \{1/2 \pm h'\} \forall (i,j) \in (m)_2\}$ and execute $\mathcal{A}^{\mathrm{NTS}}$ and StV with parameters m, h, γ on every instance $\mathbf{Q}_i$, $i \in \{1, \dots, N\}$. Table 4.1 shows the observed mean sample complexities (with standard errors in brackets) as well as the accuracies of both algorithms for $N = 100$, $m = 20$, $\gamma = 0.05$ and $h = 0.05$ for different values of h'. Both algorithms achieve an accuracy of 100% for any $h' \geq h = 0.05$ and even for $h' = 0.02$. Moreover, $\mathcal{A}^{\mathrm{NTS}}$ clearly outperforms StV for any $h' > h$, and the magnitude to which extend this happens (i.e., the sample complexity gap) appears to be increasing in h'.

Existence or Non-Existence of the CW

Next, we repeat our experiment from above with the only difference that we sample $\mathbf{Q}$ uniformly at random from $\mathcal{Q}_5^{0.05}(\mathrm{CW})$ or from $\mathcal{Q}_5^{0.05}(\neg\mathrm{CW})$, respectively. Again, the plots are generated by averaging over 25000 repetitions each. The results are shown in Fig. 7 and 8 and demonstrate that $\mathcal{A}^{\mathrm{NTS}}$ outperforms StV in both cases.

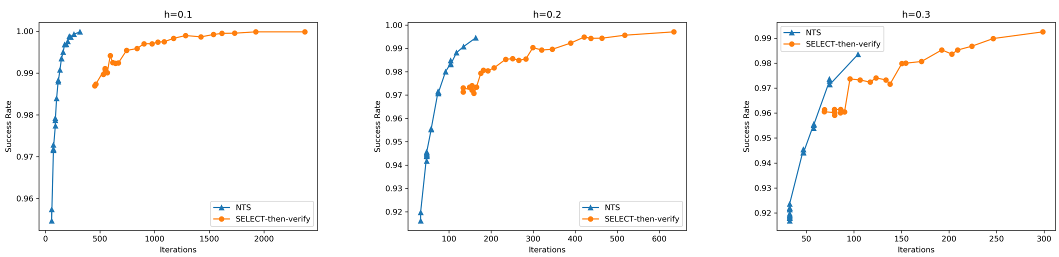


Figure 7.: Accuracy and termination time of $\mathcal{A}^{\mathrm{NTS}}$ and StV for 5 arms provided a CW exists.

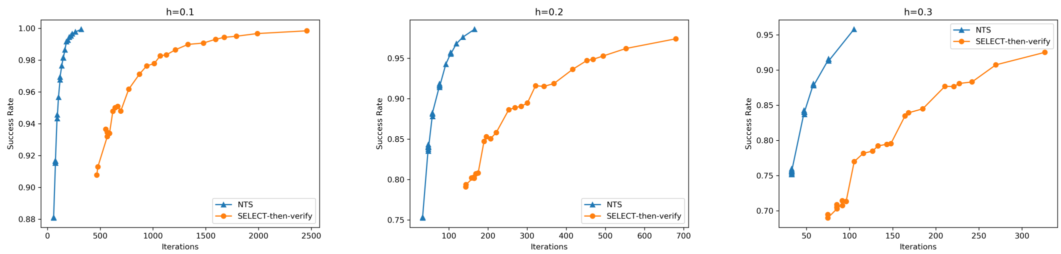


Figure 8.: Accuracy and termination time of $\mathcal{A}^{\text{NTS}}$ and StV for 5 arms provided a CW does not exist.

Comparison of $\mathcal{A}^{\text{NTS}}$ and $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$

Next, we investigate the influence that a priori knowledge of the low-noise parameter h has for CW testification. For this purpose, we compare the $\mathcal{A}^{\text{NTS}}$ with $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ when started with parameter $\gamma = 0.05$ on $\mathbf{Q}_1(x) \in \mathcal{Q}_m(1)$ and $\mathbf{Q}_2(x) \in \mathcal{Q}_m(\neg\text{CW})$ defined via

$$\mathbf{Q}_1(x) := \begin{pmatrix} - & x & x & x & x \\ & - & x & x & x \\ & & - & x & x \\ & & & - & x \\ & & & & x \end{pmatrix} \quad \text{and} \quad \mathbf{Q}_2(x) := \begin{pmatrix} - & x & 1-x & x & x \\ & - & x & 1-x & x \\ & & - & x & x \\ & & & - & x \\ & & & & x \end{pmatrix}$$

for different values of $x \in (1/2, 1]$. Table 4.2 shows the observed accuracy and sample complexity (with the corresponding standard error in brackets) averaged over 1000 repetitions. Throughout, $\mathcal{A}^{\text{NTS}}$ obtained as additional parameter $h = 0.1$. In any case, both $\mathcal{A}^{\text{NTS}}$ with $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ achieved an accuracy ≥ 0.95 as desired. With regard to the strong optimality guarantee of the SPRT (Prop. 2.17) as a solution to $\mathcal{P}^{0.6,0.4;\gamma}_{\text{Coin}}$, it is not surprising that $\mathcal{A}^{\text{NTS}}$ empirically outperforms $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ on $\mathbf{Q}_1(0.6)$ and $\mathbf{Q}_2(0.6)$. Even though this superiority of $\mathcal{A}^{\text{NTS}}$ over $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ also holds for the other considered choices of x, this effect seems to be decreasing in x.

Table 4.2.: Comparison of $\mathcal{A}^{\text{NTS}}$ and $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ on $\mathbf{Q}_1(x)$ and $\mathbf{Q}_2(x)$ for different x.

	$T^{\mathcal{A}}$		Accuracy	
$\mathbf{Q}$	$\mathcal{A}^{\text{NTS}}$	$\mathcal{A}^{\text{NTS}}_{\text{PPR}}$	$\mathcal{A}^{\text{NTS}}$	$\mathcal{A}^{\text{NTS}}_{\text{PPR}}$
$\mathbf{Q}_1(0.9)$	**85** (0.2)	92 (0.5)	1.00	1.00
$\mathbf{Q}_1(0.8)$	**111** (0.5)	166 (1.1)	0.99	1.00
$\mathbf{Q}_1(0.7)$	**166** (0.9)	386 (3.0)	1.00	1.00
$\mathbf{Q}_1(0.6)$	**332** (2.8)	1662 (14.3)	1.00	1.00
$\mathbf{Q}_2(0.9)$	**85** (0.2)	93 (0.5)	0.99	0.99
$\mathbf{Q}_2(0.8)$	**112** (0.4)	166 (1.2)	1.00	1.00
$\mathbf{Q}_2(0.7)$	**164** (0.9)	387 (3.2)	1.00	1.00
$\mathbf{Q}_2(0.6)$	**329** (2.8)	1669 (14.1)	1.00	1.00

As a further experiment, we evaluated $\mathcal{A}^{\text{NTS}}$ (with parameter $h = 0.05$) and $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ started with $\gamma = 0.05$ on 1000 instances $\mathbf{Q}$ that have been drawn uniformly at random from a set $\mathcal{Q}_m' \subsetneq \mathcal{Q}_m$. We restrict ourselves in our choice of $\mathcal{Q}_m'$ to $\mathcal{Q}_m^{h'}(\text{CW})$ and $\mathcal{Q}_m^{h'}(\neg\text{CW})$ for different values of h'. In Table 4.3 we collected the observed accuracy and average termination time, again with the corresponding standard errors in brackets. Both algorithms achieve in any considerd scenario as desired an accuracy, which is at least 0.95. In contrast to the results shown in Fig. 7 and Fig. 8, it does not seem to make a difference for this particular experiment whether the sampled instances $\mathbf{Q}$ have or do not have a CW, the observed termination times are almost the same. Moreover, we see that $\mathcal{A}^{\text{NTS}}$ clearly outperforms $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ when $\mathcal{Q}_m = \mathcal{Q}_m^{h'}((\neg)\text{CW})$ for small values of h', and this effect decreases with h', which is overall consistent with our observations from Table 4.2. Surprisingly, for $h' = 0.25 > 0.05 = h$, $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ even performs better than $\mathcal{A}^{\text{NTS}}$, and this may appear at first sight contradictory to Prop. 2.17; but recalling that Prop. 2.17 only states that the SPRT is that solution to $\mathcal{P}^{h,\gamma}_{\text{Coin}}$, which is w.r.t. the worst-case expected sample complexity optimal, there is actually no contradiction and if $h' - h$ is large, the entries $q_{i,j}$ from $\mathbf{Q} \in \mathcal{Q}_m^{h'}((\neg)\text{CW})$ are far from the "worst case", which is $1/2 \pm h$.

Table 4.3.: Comparison of $\mathcal{A}^{\text{NTS}}$ and $\mathcal{A}^{\text{NTS}}_{\text{PPR}}$ on different sets $\mathcal{Q}_m' \subseteq \mathcal{Q}_m$.

	$T^{\mathcal{A}}$		Accuracy	
$\mathcal{Q}_m'$	$\mathcal{A}^{\text{NTS}}$	$\mathcal{A}^{\text{NTS}}_{\text{PPR}}$	$\mathcal{A}^{\text{NTS}}$	$\mathcal{A}^{\text{NTS}}_{\text{PPR}}$
$\mathcal{Q}_m^{0.25}(\neg\text{CW})$	183 (1.0)	**128** (1.2)	1.00	1.00
$\mathcal{Q}_m^{0.10}(\neg\text{CW})$	**262** (2.3)	335 (6.4)	1.00	1.00
$\mathcal{Q}_m^{0.05}(\neg\text{CW})$	**325** (4.2)	729 (22.9)	1.00	0.99
$\mathcal{Q}_m^{0.25}(\text{CW})$	182 (0.9)	**127** (1.2)	1.00	1.00
$\mathcal{Q}_m^{0.10}(\text{CW})$	**264** (2.4)	331 (6.5)	1.00	1.00
$\mathcal{Q}_m^{0.05}(\text{CW})$	**331** (4.2)	688 (18.9)	1.00	1.00

4.6.2. The Passive Scenario

Finally, we demonstrate how the passive setting described in Sec. 4.2.3 can be utilized in order to justify the usage of dueling bandits algorithms focusing on alternative best arm concepts for the goal of regret minimization, if the test component detects a violation of the CW assumption. For this purpose, we consider two sampling strategies:

- *Relative Upper Confidence Bound* (RUCB) from Zoghi et al. [2014b], which is a dueling bandit algorithm based on the CW assumption.
- *Double Thompson Sampling* (DTS) from Wu and Liu [2016], which is a dueling bandit algorithm not relying on the CW assumption, but instead focusing on the set of Copeland winners (cf. Sec. 1.4 or [Bengs et al., 2021]).

Similar as Zoghi et al. [2015a], we consider the regret based on the difference in the normalized Copeland scores of the Copeland winner and the two chosen arms. It is well known that RUCB can achieve linear regret in case no CW exists, while DTS provably only suffers sub-linear regret (with respect to the Copeland scores) in such cases.

In light of this, we consider a two-staged algorithm (denoted by RUCB→DTS), which executes in its first stage Alg. 15 (with parameters $\gamma = 0.1$ and $h = 0.1$) instantiated with RUCB as its sampling strategy, and in its second stage simply DTS in case that no CW exists in the ground truth relation and otherwise RUCB. In other words, RUCB→DTS switches from the CW-based sampling strategy RUCB to the Copeland winner based DTS procedure if the CW assumption is likely violated. Additionally, we also consider RUCB→DTS(PPR), which is that modification of RUCB→DTS which uses PPR-Bernoulli (Alg. 4) instead of the SPRT as $\mathcal{A}_{\mathrm{Coin}}$ in Alg. 15 and thus obtains as parameter only $\gamma = 0.1$ (and not $h = 0.1$). For evaluating the algorithms we choose as the underlying ground-truth preference relation the reciprocal relation

$$\mathbf{Q}_{\mathrm{Hudry}} := \frac{1}{10}\begin{pmatrix} 5 & 1 & 1 & 1 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 \\ & 5 & 9 & 1 & 1 & 1 & 1 & 9 & 9 & 9 & 9 & 9 & 9 \\ & & 5 & 9 & 9 & 9 & 9 & 1 & 1 & 1 & 9 & 9 & 9 \\ & & & 5 & 9 & 9 & 9 & 9 & 9 & 9 & 1 & 1 & 1 \\ & & & & 5 & 9 & 9 & 9 & 9 & 9 & 1 & 1 & 1 \\ & & & & & 5 & 9 & 9 & 9 & 9 & 1 & 1 & 1 \\ & & & & & & 5 & 9 & 9 & 9 & 1 & 1 & 1 \\ & & & & & & & 5 & 9 & 9 & 9 & 9 & 9 \\ & & & & & & & & 5 & 9 & 9 & 9 & 9 \\ & & & & & & & & & 5 & 9 & 9 & 9 \\ & & & & & & & & & & 5 & 9 & 9 \\ & & & & & & & & & & & 5 & 9 \\ & & & & & & & & & & & & 5 \end{pmatrix} \in \mathcal{Q}_{13},$$

which does not have a CW, and 4 as the unique Copeland winner [Ramamohan et al., 2016].

Fig. 9 illustrates the benefit of changing from RUCB to DTS with regard to the regret. In particular, RUCB→DTS does not suffer the linear regret of RUCB but instead its cumulative regret appears only by a constant term larger than that of DTS. Even though RUCB→DTS(PPR) does not have knowledge of h, it performs almost exactly like RUCB→DTS. Since all non-diagonal entries of $\mathbf{Q}_{\mathrm{Hudry}}$ are in $\{0.1, 0.9\}$, this seems consistent with our observations in Sec. 2.4.1, where we saw that testing with desired confidence 0.95 for the mode of a coin $p \in \{0.1, 0.9\}$ via PPR-Bernoulli requires almost the same sample complexity as solving this task via the SPRT with additional information $|p - 1/2| = 0.4$, cf. Tables 2.3 and 2.4.

4.7. Discussion and Related Work

The notion of the Condorcet winner (CW) dates back to the 18th century [Caritat, 1785] and also appears in the social choice literature [Fishburn, 1974, Fishburn and Gehrlein, 1976], where the data is typically assumed to be available in the form of a list containing total rankings over all alternatives from different voters. In dueling bandits, the CW is arguably the most natural choice of the “best arm” one may think of, and identifying it is a prevalent goal [Bengs et al., 2021]. Unfortunately, it is not guaranteed to exist, neither in theory nor in real-life scenarios [Zoghi et al., 2015a]. In a variety of papers, this issue is circumvented by simply assuming its existence, either explicitly [Urvoy et al., 2013, Zoghi et al., 2014b, Komiyama et al., 2015, Karnin, 2016, Chen and Frazier, 2017, Li et al., 2020] or implicitly as a consequence of stronger assumptions on the underlying preference relation, such as a total order of arms or some kind of stochastic transitivity [Yue and Joachims, 2011, Yue et al., 2012, Falahatgar et al., 2017b,a, 2018, Mohajer et al., 2017], by assuming latent utilities [Yue and Joachims, 2009, Ailon et al., 2014b, Kumagai, 2017, Maystre and Grossglauser, 2017] or an underlying statistical ranking model [Busa-Fekete et al., 2014a, Szörényi et al., 2015]. Other papers, instead, focused on alternative best arm

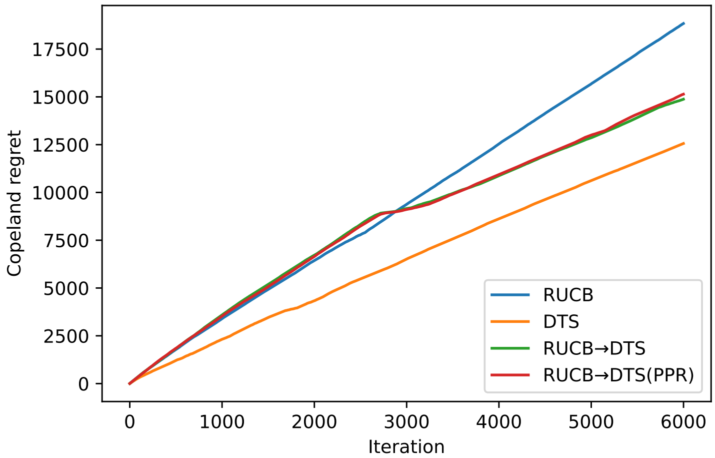


Figure 9.: Copeland regret of DTS, RUCB, RUCB→DTS and RUCB→DTS(PPR) on $\mathbf{Q}_{\text{Hudry}}$.

concepts that always exist, such as the Copeland winner [Zoghi et al., 2015a, Komiyama et al., 2016, Wu and Liu, 2016], the Borda winner [Jamieson et al., 2015, Lin and Lu, 2018], or more general tournament solutions [Ramamohan et al., 2016].

Falahatgar et al. [2017b,a, 2018] restricted themselves to the identification of a weakened notion of the CW in form of an ε-best arm for some $\varepsilon > 0$, i.e., an arm i that fulfills $q_{i,j} \geq -\varepsilon$ for all $j \in [m] \setminus \{i\}$. Similar as for the CW, the existence of such an ε-best arm is not guaranteed in theory, and analoga of CW testification, CW checking and CW verification for this notion may be formulated and of interest. The parameter ε resembles our low-noise parameter h to some extent, but in fact these scenarios are different, as e.g., in contrast to the CW, an ε-best arm of $\mathbf{Q} \in \mathcal{Q}_m^h$ does not have to be unique if existent. Hence, our results are not directly applicable to this setting.

This chapter contains to large parts results from [Haddenhorst et al., 2021a] but also extends upon these in the sense that we also consider the CW-related problems on $\mathcal{Q}_m^h$ for $h = 0$ here. We introduced CW testification as combined testing and verification of the CW in the passive and active scenario. In contrast to the quite prominent best-arm-identification problem [Mohajer et al., 2017, Ren et al., 2020, Bengs et al., 2021], actually testing validity of the underlying assumption that the CW exists has not been discussed so far. Here, our passive solution is of huge interest as it allows for passively checking the validity of an underlying CW assumption.

In the active scenario, we presented instance-dependent sample complexity lower and upper bounds for solutions to the problem, which match in the worst-case up to logarithmic factors. We saw that CW testification on $\mathcal{Q}_m^h$ for γ requires roughly the same sample complexity as any of CW verification, CW identification and CW checking, namely $\widetilde{\Theta}(\frac{m}{h^2} \ln \frac{1}{\gamma})$. As a main ingredient for most of these lower bounds, we used the optimality of the SPRT for solving a coin tossing problem (cf. Thm. 2.29 and Prop. 2.11); an alternative approach, which is actually underlying the lower bound for $\mathcal{P}_{\text{CWv}}^{m,h,\gamma}(\text{CW})$ from Thm. 4.4, would be to use the change-of-measure argument presented in Sec. 2.5.

Thm. 4.3 shows that any solution to CW testification on $\mathcal{Q}_m^0$ for γ necessarily fulfills $\limsup_{h\searrow 0} \sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} > 0$ and – even though explicitly stated – this result can supposedly also be transferred to CW checking, CW identification and CW verification. This indicates the necessity of an additional $\ln\ln\frac{1}{h}$ factor in the required sample complexity when comparing $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$ to $\mathcal{P}_{\mathrm{CWt}}^{m,0,\gamma}$, but it does only capture the dependence on h and neither on γ nor on m. In fact, when trying to show a lower bound of order $\Omega_{\sup}(\frac{m}{h^2}(\ln\ln\frac{1}{h})\ln\frac{1}{\gamma})$, we faced obstacles that we were not yet able to resolve; cf. for this our related discussions in Sec. 2.6 and 5.7.

In Ch. 5 below, we tackle WST testing via a likelihood-ratio based approach, resulting in a suboptimal solution for WST testing on $\mathcal{Q}_m^h$ for γ as well as asymptotic size-α tests for WST resp. $\neg$WST. We suppose that these results might be transferrable to the CW related problems from this chapter; however, we do not expect a better solution for CW testification (or related problems) on $\mathcal{Q}_m^h$ for γ with this and such asymptotic size-α tests are not the focus of this work, hence we did not go further into this direction.

In Sec. 4.5, we restricted ourselves to formulate the guarantees of Sticky-Track-and-Stop-solutions for the problems $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$ and $\mathcal{P}_{\mathrm{CWc}}^{m,h,\gamma}$. However, this framework is so general that it presumably also yields analogous guarantees for the remaining problems $\mathcal{P}_{\mathrm{CWv}}^{m,h,\gamma}$, $\mathcal{P}_{\mathrm{CWi}}^{m,h,\gamma}(\mathrm{CW})$ and $\mathcal{P}_{\mathrm{CWv}}^{m,h,\gamma}(\mathrm{CW})$.

In addition to the above mentioned open questions and conjectures, there are multiple possible directions for future research.

The sample complexity of our proposed solution *Noisy Tournament Sampling* for CW testification on $\mathcal{Q}_m^h$ depends to a large extend on $\mathcal{A}_{\mathrm{Coin}}$ as a parameter thereof. In case $h > 0$, the SPRT seems to be, due to its strong optimality guarantees, the best choice for $\mathcal{A}_{\mathrm{Coin}}$. If $h = 0$, however, the picture is not so clear as we have seen in Sec. 2.2. Instead, there are several reasonable choices with different types of guarantees, and it would thus be of interest to compare different of these in additional experiments. So far, we have restricted ourselves in the experiments to the solution PPR-Bernoulli to $\mathcal{P}_{\mathrm{Coin}}^{\gamma}$ as a choice of $\mathcal{A}_{\mathrm{Coin}}$ in Alg. 15 and Alg. 16, because it empirically outperformed the other solutions in Sec. 2.4.1.

Over and above the theoretical sample complexity upper and lower bounds for CW testification proven so far, one could also try to show further bounds of distinct type, e.g. an upper bound on the expected sample complexity or a high-probability lower bound. Alternatively, one might think of replacing the low-noise assumption by potentially stronger ones and e.g. discuss CW testification whilst assuming the stochastic triangle inequality or WST to hold. There are several modifications one might think of in this regard, and addressing any of these only seems reasonable if that learning scenario is of interest for either an existent algorithmic procedure or a particular theoretical result. In addition, one could further investigate the instance-wise sample complexity necessary to solve CW testification: Our bounds were merely shown to be sharp in a worst-case asymptotic sense, but how large is the true gap between the necessary and sufficient sample complexity of a solution $\mathcal{A}$ to $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$ when executed on some instance $\mathbf{Q} \in \mathcal{Q}_m$?

Apart from that, also the generalization of the learning problem could lead to interesting research questions, e.g., one could formulate for the best-k arm identification problem [Braverman et al., 2016, Mohajer et al., 2017, Ren et al., 2020], which coincides with CW identification for $k = 1$, possibly also a testification variant and theoretically analyze this more general problem.

Our empirical evaluation is quite limited and could be extended in several ways. One could further investigate our proposed active solution with regard to other performance measures such as worst-case termination time on subsets $\mathcal{Q}'_m \subseteq \mathcal{Q}_m$ which are not of the form $\mathcal{Q}'_m = \mathcal{Q}^h_m$. Yet another path to take could be to apply our passive CW testification procedure in further real-life scenarios. We have already seen in an example that the possibility to detect violations of the CW assumption may allow for a shift of strategies and prevent in this way from suffering linear regret in one particular scenario, and it could be that such an approach might also be fruitful for other scenarios, e.g., for pure exploration tasks. Last but not least, the mere idea of testification might be transferred to other machine learning problems as well in order to be more robust against violations of falsely made assumptions. For example, instead of only identifying the underlying ranking in dueling bandits whilst assuming its existence, one could try do so if it exists and detect non-existence if apparent.

5. Testing for Stochastic Transitivity

This chapter discusses the testing of yet another type of coherence, which is a common statistical assumption in the dueling bandits scenario, namely that of *stochastic transitivity*. Recall that the concept of transitivity is usually understood to be a property of the form "If i is better than j, and j is better than k, then i is also better than k". As mentioned in Ch. 1, for the case of reciprocal relations, which model the feedback in dueling bandits, it is not uniquely determined what "a is better than b" means. In fact, there are different notions of *stochastic* transitivity [Fishburn, 1973, Haddenhorst et al., 2020], which play a role in the dueling bandits literature, either as explicitly made assumptions [Yue et al., 2012, Yue and Joachims, 2011, Mohajer et al., 2017, Falahatgar et al., 2017a,b, 2018] or implicitly as consequence of even stronger assumptions [Szörényi et al., 2015, Maystre and Grossglauser, 2017], cf. Sec. 5.7 for slightly more details on this.

Somewhat surprisingly, the problem of testing the validity of transitivity assumptions underlying various algorithms has not been considered in the dueling bandits scenario before. Needless to say, this would be important to guarantee the meaningfulness of the results produced by algorithms, which formally require transitivity. In fact, if the assumptions made by an algorithm are violated by the data-generating process in a concrete application, then neither its prediction nor any of its guarantees can be trusted anymore. Before formalizing our testing problem and giving an overview over the results obtained in this chapter, we define the transitivity conditions of interest, two of which have already been mentioned in Ch. 1, and emphasize the importance of transitivity testing by means of a particular example.

Notions of Stochastic Transitivity A reciprocal relation $\mathbf{Q} = (q_{i,j})_{1 \le i,j \le m} \in \mathcal{Q}_m$ is said to satisfy

- *weak stochastic transitivity* (WST) if

$$\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge 1/2\,,$$

- *moderate stochastic transitivity* (MST) if

$$\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge \min\{q_{i,j}, q_{j,k}\}\,,$$

- *ν-relaxed stochastic transitivity* (νRST) for some $\nu \in (0,1)$ if

$$\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge \nu \max\{q_{i,j}, q_{j,k}\} + (1-\nu)/2\,,$$

- *strong stochastic transitivity* (SST) if[1]

$$\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge \max\{q_{i,j}, q_{j,k}\}\,,$$

[1]The notion of SST is originally due to Davidson and Marschak [1958].

- λ-*stochastic transitivity* (λST) for some $\lambda \in (0,1)$ if

$$\big(q_{i,j} \geq 1/2 \wedge q_{j,k} \geq 1/2\big) \Rightarrow q_{i,k} \geq \lambda \max\{q_{i,j}, q_{j,k}\} + (1-\lambda)\min\{q_{i,j}, q_{j,k}\},$$

where all previous conditions must hold for all distinct $i, j, k \in [m]$, respectively. Given a transitivity type XST $\in$ {WST, MST, νRST, λST, SST}, we write $\mathcal{Q}_m(\text{XST})$ for the set of all $\mathbf{Q} \in \mathcal{Q}_m$, which fulfill XST, and abbreviate $\mathcal{Q}_m(\neg\text{XST}) \coloneqq \mathcal{Q}_m \setminus \mathcal{Q}_m(\text{XST})$. The following relationships hold between the different types of stochastic transitivities [Haddenhorst et al., 2020]:

$$\mathcal{Q}_m(\text{SST}) \subsetneq \mathcal{Q}_m(\text{XST}) \subsetneq \mathcal{Q}_m(\text{WST}) \quad \text{for XST} \in \{\text{MST}, \nu\text{RST}, \lambda\text{ST}\},$$

but neither

$$\mathcal{Q}_m(\text{XST}) \subseteq \mathcal{Q}_m(\text{MST}) \quad \text{nor} \quad \mathcal{Q}_m(\text{MST}) \subseteq \mathcal{Q}_m(\text{XST})$$

for XST $\in$ {λST, νRST}.

In order to demonstrate the relevance of the WST assumption and the undesirable consequences in case of its violation, we provide an illustrative toy example. For this, let us introduce some more notation: Given $\mathbf{Q} \in \mathcal{Q}^0_m$, we write $\sigma_{\mathbf{Q}}$ for a permutation on $[m]$, which fulfills $q_{\sigma_{\mathbf{Q}}(i),\sigma_{\mathbf{Q}}(i+1)} > 1/2$ for every $i \in [m]$. We have seen in Lem. 3.3 that $\sigma_{\mathbf{Q}}$ exists for every $\mathbf{Q} \in \mathcal{Q}^0_m$. In case $\mathbf{Q} \in \mathcal{Q}^0_m(\text{WST})$, $\sigma_{\mathbf{Q}}$ is the underlying ground-truth ranking of $\mathbf{Q}$, and permuting rows and columns according to $\sigma_{\mathbf{Q}}$ results in a reciprocal relation with entries $> 1/2$ above the diagonal.

Now, suppose we want to identify in a DB scenario the ranking $\sigma_{\mathbf{Q}}$ for the underlying reciprocal relation $\mathbf{Q}$, and that $\mathbf{Q} \in \mathcal{Q}_3(\neg\text{WST})$ is given by $q_{1,2} = 0.9, q_{1,3} = 0.1$ and $q_{2,3} = 0.8$. Next, consider the (noisy) sorting-based ranking algorithm by Maystre and Grossglauser [2017], which terminates in any case and returns a ranking of the arms. This approach basically employs multiple runs of Quicksort yielding in turn multiple noisy rankings over the arms, which are then combined into a single final ranking via Copeland aggregation: The arms are ranked in an increasing order according to their Copeland score, that is, the number of other arms beaten in a majority of the noisy rankings, while ties are broken arbitrarily. Due to this aggregation step, the final ranking will likely correspond to the Copeland ranking, which, however, is any possible permutation of the three arms for $\mathbf{Q}$ as above. In particular, the final ranking will be arbitrary to a great extent, and since $\mathbf{Q} \in \mathcal{Q}_3(\neg\text{WST})$, no ranking (permutation) σ is coherent with $\mathbf{Q}$ in the sense that $q_{\sigma(i),\sigma(j)} \geq \frac{1}{2}$ holds for every $1 \leq i < j \leq m$.

Thus, without an appropriate test of WST, this algorithm would never notice the violation of WST during its learning phase, and consequently cannot ensure the trustworthiness for its final output. However, having a test component running in parallel to the learner, it may allow to either give a warning that the learned ranking is probably incoherent with the preference relation, or even intervene and interrupt the ranking algorithm. The other way around, the information that WST is not violated would increase the trustworthiness of the ranking predicted by the online learner.

In dependence of a low-noise parameter $h \in [0, 1/2)$, error probabilities $\alpha, \beta \in (0,1)$, the number $m \in \mathbb{N}$ of arms, we define the XST testing problem as follows: A (possibly probabilistic, sequential) algorithm $\mathcal{A}$ *solves* XST *testing on* $\mathcal{Q}^h_m$ *for* α *and* β (short:

$\mathcal{P}_{\text{XST}}^{m,h,\alpha,\beta}$) if $T^{\mathcal{A}}$ is a.s. finite on any instance $\mathbf{Q} \in \mathcal{Q}_m^h$ and both

$$\begin{aligned} &\inf\nolimits_{\mathbf{Q}\in\mathcal{Q}_m^h(\text{XST})} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \text{XST}) \geq 1-\alpha \\ \text{and} \quad &\inf\nolimits_{\mathbf{Q}\in\mathcal{Q}_m^h(\neg\text{XST})} \mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \neg\,\text{XST}) \geq 1-\beta \end{aligned} \tag{5.1}$$

are fulfilled.

Outline of The Chapter In the further course of this chapter, we discuss $\mathcal{P}_{\text{XST}}^{m,h,\alpha,\beta}$ for several choices of parameters. For XST $\neq$ WST, we point out that XST testing on $\mathcal{Q}_m^h$ for α and β is impossible in the sense that any solution $\mathcal{A}$ to $\mathcal{P}_{\text{XST}}^{m,h,\alpha,\beta}$ fulfills $\sup_{\mathbf{Q}\in\mathcal{Q}_m(\text{XST})} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$. For this reason, we focus on the case of WST from that point on.

Similarly as for the CW-related problems, WST testing may be reduced to pure exploration multi-armed bandits and results from Degenne and Koolen [2019] translate into a solution $\mathcal{A}(\gamma)$ to $\mathcal{P}_{\text{WST}}^{m,h,\gamma} := \mathcal{P}_{\text{WST}}^{m,h,\gamma,\gamma}$, which is optimal with respect to the quantity

$$\sup\nolimits_{\mathbf{Q}\in\mathcal{Q}_m^h} \lim_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln\frac{1}{\gamma}}.$$

Moreover, for this optimal $\mathcal{A}(\gamma)$, this quantity is of the order $\Theta(\frac{m^2}{h^2}\ln\frac{1}{\gamma})$. But again, this does neither result in a solution to WST testing on $\mathcal{Q}_m^h$ for fixed confidences $1-\alpha$ and $1-\beta$, nor is it applicable in the passive scenario. As this approach is not our main interest, we postpone its discussion to Sec. 5.5.

With the help of the lower bounds from Sec. 2.2 we infer expected sample complexity lower bounds for solutions to $\mathcal{P}_{\text{WST}}^{m,h,\alpha,\beta}$. More precisely, for any $\mathbf{Q} \in \mathcal{Q}_m^h(\text{WST})$, we obtain an instance-wise lower bound for $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}]$ in terms of the gaps $\bar{q}_{i,j} = |q_{i,j} - 1/2|$ of $\mathbf{Q}$. In the symmetric case $\alpha = \beta = \gamma$, this results in a worst-case bound of order $\Omega(\frac{m^2}{h^2}\ln\frac{1}{\gamma})$.

Afterwards, we follow two distinct approaches for tackling the WST testing problem: The first one is similar to that followed in Ch. 4 for solving $\mathcal{P}_{\text{CWt}}^{m,h,\alpha,\beta}$, it basically uses for each pair $\{i,j\}$ a binomial test to decide whether $q_{i,j} > 1/2$ holds or not, and aggregates this information into a final decision. The second one instead is based on two likelihood ratio test statistics, which indicate fulfillment resp. violation of WST.

As we will see, for $h > 0$, the first approach yields without much effort a "naive" solution to $\mathcal{P}_{\text{WST}}^{m,h,\alpha,\beta}$ that is w.r.t. its worst-case sample complexity already asymptotically optimal up to logarithmic factors. However, the incorporation of graph-theoretical considerations from Sec. 3 allows us to construct a more sophisticated solution, which outperforms the naive one both empirically and theoretically. This solution is applicable in the passive as well as in the active scenario, and in the latter one we provide instance-wise sample complexity bounds, which are up to logarithmic factors asymptotically optimal in a worst-case sense. Moreover, we solve $\mathcal{P}_{\text{WST}}^{m,0,\alpha,\beta}$ and obtain, compared to the solution of $\mathcal{P}_{\text{WST}}^{m,h,\alpha,\beta}$, an additional factor of $\ln\ln\frac{1}{h}$ in the – w.r.t. $\mathcal{Q}_m^h$ – worst-case sample complexity bound. A corresponding lower bound result indicates that this factor is indispensable.

In Sec. 5.4, we develop an alternative solution to $\mathcal{P}_{\text{WST}}^{m,h,\alpha,\beta}$ by following a likelihood-ratio test approach. Its main idea is to decide based on the LRT statistic λ_t of the hypotheses $\mathbf{H}_0 : \mathbf{Q} \in \mathcal{Q}_m(\text{WST})$ versus $\mathbf{H}_1 : \mathbf{Q} \in \mathcal{Q}_m(\neg\text{WST})$ and the LRT statistic μ_t of the corresponding interchanged hypotheses. Simple modifications thereof lead to test statistics $\tilde{\lambda}_t$ and $\tilde{\mu}_t$, which are the basis for a passive solution to $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{WST}}$ with merely mild assumptions on the underlying sampling strategy π. A particular choice of π results in a corresponding active solution, which has in the symmetric case $\alpha = \beta = \gamma$ a worst-case expected sample complexity of order $\mathcal{O}(\frac{m^{2\kappa}}{h^4} \ln \frac{1}{\gamma})$ w.r.t. $\mathcal{Q}^h_m$-instances for some parameter $\kappa > 1$ and is thus far from optimal. Nevertheless, we have included it for the sake of completeness. Additionally, we elaborate in Sec. 5.4.6 on the asymptotic behaviour of the test statistics and obtain results similar to those from [Iverson and Falmagne, 1985] in the more general dueling bandits scenario: We show that, under mild assumptions on π, the LRT statistics fulfill

$$\sup_{\mathbf{Q}\in\overline{\mathcal{Q}_m(\neg\mathrm{WST})}} \limsup_{t\to\infty} \mathbb{P}\left(\mu_t > l\right) \leq 2^{-M} \sum_{a=0}^{M} \binom{M}{a} \mathbb{P}_{\boldsymbol{\theta}}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right),$$

$$\sup_{\mathbf{Q}\in\overline{\mathcal{Q}_m(\mathrm{WST})}} \limsup_{t\to\infty} \mathbb{P}\left(\lambda_t > l\right) \leq 2^{-M} \sum_{a=0}^{M} \binom{M}{a} \mathbb{P}_{\boldsymbol{\theta}}\left(\chi^2_{(a\wedge(M-2))} > l\right),$$

where $M = \binom{m}{2}$, $\overline{A}$ is the closure of $A \subseteq \mathcal{Q}_m \cong \mathbb{R}^M$ and $\chi^2_{(k)}$ denotes the χ^2-distribution with k degrees of freedom. This allows the formulation of asymptotic size-α tests for testing WST and $\neg$WST, respectively.

After a discussion on the above mentioned implications of [Degenne and Koolen, 2019] in Sec. 5.5 and an empirical evaluation in Sec 5.6, we conclude this chapter in Sec. 5.7 with remarks on possible further research questions and the related literature.

5.1. Impossibility Results for Several Types of Stochastic Transitivity

We start our analysis of XST testing with the statement of negative results for the case XST $\neq$ WST. Similarly as Cor. 2.50, Cor. 2.48 allows us to infer without much effort the following.

Corollary 5.1. *Suppose* XST $\in \{\mathrm{SST}, \mathrm{MST}, \lambda\mathrm{ST}, \nu\mathrm{RST}\}$ *and* $h \in (0, 1/2)$, *with* $\nu > 2h$ *if* XST $= \nu$RST. *If* $\alpha, \beta \in (0, \frac{1}{2.4})$ *and* $\mathcal{A}$ *solves* $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{XST}}$, *then*

$$\sup_{\mathbf{Q}\in\mathcal{Q}^h_m(\mathrm{XST})} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \sup_{\mathbf{Q}\in\mathcal{Q}^h_m(\neg\mathrm{XST})} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty.$$

Proof. As in the proof of Cor. 2.50, with a slight abuse of notation, we may replace h by $h + \delta$ for some small $\delta > 0$ and then suppose w.l.o.g. $\mathcal{Q}^h_m = \{\mathbf{Q} \in \mathcal{Q}_m \,|\, \forall (i,j) \in (m)_2 : |q_{i,j} - 1/2| \geq h\}$ throughout this proof. Suppose XST $\in \{\mathrm{SST}, \mathrm{MST}, \lambda\mathrm{ST}, \nu\mathrm{RST}\}$ to be arbitrary but fixed. The reciprocal relation $\mathbf{Q} \in \mathcal{Q}_m$ defined via $q_{i,j} := 1/2 + h$ for all $1 \leq i < j \leq m$ fulfills $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{SST}) \cap \mathcal{Q}^{\clubsuit}_m \cap \mathcal{U}^m_2$, and $\mathcal{Q}_m(\mathrm{SST}) \subseteq \mathcal{Q}_m(\mathrm{XST})$ implies $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{XST}) \cap \mathcal{Q}^{\clubsuit}_m \cap \mathcal{U}^m_2$ with $\mathcal{Q}^{\clubsuit}_m$ and $\mathcal{U}^m_2$ defined as in Sec. 2.5.1.

If XST $\in \{\mathrm{SST}, \mathrm{MST}, \lambda\mathrm{ST}\}$, let $\varepsilon \in (0, 1/4 - h/2)$ and define $\mathbf{Q}' \in \mathcal{Q}_m$ via

$$q'_{i,j} := \begin{cases} 1 - 2\varepsilon, & \text{if } (i,j) = (1,m), \\ 1 - \varepsilon, & \text{otherwise} \end{cases}$$

for any $1 \leq i < j \leq m$. Then, $\mathbf{Q}' \in \mathcal{Q}^h_m \cap \mathcal{Q}^{\perp}_m \cap \mathcal{U}^m_2$ holds and $q'_{1,m} = 1 - 2\varepsilon < 1 - \varepsilon = \min(q'_{1,m-1}, q'_{m-1,m})$ assures $\mathbf{Q}' \in \mathcal{Q}_m(\neg\mathrm{MST})$. As $\mathcal{Q}_m(\mathrm{SST}) \cup \mathcal{Q}_m(\lambda\mathrm{ST}) \subseteq \mathcal{Q}_m(\mathrm{MST})$

holds, we thus have $\mathbf{Q'} \in \mathcal{Q}_m^h(\neg\text{XST}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$.

Now, we tackle the case XST = νRST, in which $\nu > 2h$ holds by assumption. Fix $\varepsilon \in (0, \min\{1/2 - h, 1 - 2h/\nu\})$ and define $\mathbf{Q'} \in \mathcal{Q}_m$ via

$$q'_{i,j} := \begin{cases} \nu(1-\varepsilon/2) + (1-\nu)/2, & \text{if } (i,j) = (1,m), \\ 1-\varepsilon, & \text{otherwise} \end{cases}$$

for any $1 \leq i < j \leq m$ and note that $\varepsilon > 0 > (\nu - 1)/(2\nu)$ assures $q'_{1,m} \in (1/2, 1)$, i.e., $\mathbf{Q'}$ is well-defined and an element in $\mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$. Note that $\varepsilon < 1/2 - h$ resp. $1 - 2h/\nu$ assure $1 - \varepsilon > 1/2 + h$ resp. $\nu(1-\varepsilon/2) + (1-\nu)/2 > 1/2 + h$ and thus $\mathbf{Q'} \in \mathcal{Q}_m^h$. Moreover,

$$q'_{1,m} = \nu(1-\varepsilon/2) + (1-\nu)/2 < \nu(1-\varepsilon) + (1-\nu)/2 = \nu \max(q_{i,j}, q_{j,k}) + (1-\nu)/2$$

yields $\mathbf{Q'} \in \mathcal{Q}_m(\neg\nu\text{RST})$.

To conclude, for any XST $\in \{\text{SST}, \text{MST}, \lambda\text{ST}, \nu\text{RST}\}$, we have given reciprocal relations $\mathbf{Q} \in \mathcal{Q}_m^h(\text{XST}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$ and $\mathbf{Q'} \in \mathcal{Q}_m^h(\neg\text{XST}) \cap \mathcal{Q}_m^{\clubsuit} \cap \mathcal{U}_2^m$. Consequently, the statement follows directly from Cor. 2.48. □

The technical assumption $\nu > 2/h$ made in Cor. 5.1 assures $\mathcal{Q}_m^h(\neg\nu\text{RST}) \cap \mathcal{Q}_m^{\clubsuit} \neq \emptyset$. In fact, since $\mathcal{Q}_m^{\clubsuit} \cap \mathcal{Q}_m^h \subseteq \mathcal{Q}_m^h(\text{WST})$ holds, the necessity of this assumption is shown with the following lemma.

Lemma 5.2. *If $\nu \leq 2h$, then $\mathcal{Q}_m^h(\text{WST}) \subseteq \mathcal{Q}_m(\nu\text{RST})$.*

Proof. Let $\mathbf{Q} \in \mathcal{Q}_m^h \cap \mathcal{Q}_m(\text{WST})$ and $i, j, k \in [m]$ be distinct with $q_{i,j}, q_{j,k} \geq 1/2$. As $\mathbf{Q}$ fulfills WST, we obtain $q_{i,k} \geq 1/2$, i.e., $q_{i,j}, q_{j,k}, q_{i,k} > 1/2 + h$ holds with regard to $\mathbf{Q} \in \mathcal{Q}_m^h$. Hence,

$$q_{i,k} - \nu \max(q_{i,j}, q_{j,k}) - \frac{1-\nu}{2} > \frac{1}{2} + h - \nu - \frac{1-\nu}{2} = h - \frac{\nu}{2} \geq 0$$

holds and $\mathbf{Q} \in \mathcal{Q}_m(\nu\text{RST})$ follows. □

By applying Thm. 2.44 instead of Cor. 2.48, one could obtain stronger results in the setting of Cor. 5.1, namely $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$ for any $\mathbf{Q} \in \mathcal{Q}_m^h(\text{XST}) \cap \partial\mathcal{Q}_m^h(\neg\text{XST}) \cap \mathcal{U}_2^m$. We did not state this here, because via an alternative approach based on Lem. 2.15, we additionally obtain such a negative result also for $\mathbf{Q} \in (\mathcal{Q}_m^h(\text{XST}) \cap \partial\mathcal{Q}_m^h(\neg\text{XST})) \setminus \mathcal{U}_2^m$. The following proposition has also been given in [Haddenhorst et al., 2021b]; here we also treat XST = λST and state $\nu > 2h$ as an additional assumption for XST = νRST.

Proposition 5.3. *Let $h, \alpha, \beta \in (0, 1/2)$, $m \in \mathbb{N}_{\geq 3}$ and XST $\in \{\text{MST}, \text{SST}, \nu\text{RST}, \lambda\text{ST}\}$ be fixed, where $\nu > 2h$ if XST = νRST. If $\mathcal{A}$ solves $\mathcal{P}_{\text{XST}}^{m,h,\alpha,\beta}$, then $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$ for any $\mathbf{Q} \in \mathcal{Q}_m^h(\text{XST}) \cap \partial\mathcal{Q}_m^h(\neg\text{XST}) \neq \emptyset$. In particular, we have $\sup_{\mathbf{Q} \in \mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$.*

Proof of Prop. 5.3. Let XST $\in \{\text{MST}, \text{SST}, \lambda\text{ST}, \nu\text{RST}\}$ be fixed and suppose $\mathcal{A}$ solves $\mathcal{P}_{\text{XST}}^{m,h,\alpha,\beta}$. Then, $\mathcal{A}$ solves in particular $\mathcal{P}_{\text{XST}}^{m,h,\gamma}$ with $\gamma := \max\{\alpha, \beta\}$. With $g_{\text{XST}} : [0,1]^2 \to [0,1]$ defined via

$$g_{\text{XST}}(x,y) := \begin{cases} \min(x,y), & \text{if XST = MST}, \\ \max(x,y), & \text{if XST = SST}, \\ \nu \max(x,y) + (1-\nu)/2, & \text{if XST} = \nu\text{RST}, \\ \lambda \max(x,y) + (1-\lambda)\min(x,y), & \text{if XST} = \lambda\text{ST}, \end{cases}$$

a look at the definition of XST reveals that $\mathcal{Q}_m(\text{XST})$ is the set

$$\{\mathbf{Q} \in \mathcal{Q}_m \,|\, \forall \text{ distinct } i,j,k \in [m] : (q_{i,j} \geq 1/2 \wedge q_{j,k} \geq 1/2) \Rightarrow q_{i,k} \geq g_{\text{XST}}(q_{i,j}, q_{j,k})\}. \tag{5.2}$$

We showed in the proof of Cor. 5.1 that none of the sets $\mathcal{Q}_m^h(\text{XST}) \cap \mathcal{Q}_m^{\clubsuit}$ and $\mathcal{Q}_m^h(\neg\text{XST}) \cap \mathcal{Q}_m^{\clubsuit}$ is empty, hence closedness of $\mathcal{Q}_m(\text{XST})$ and connectedness of $\mathcal{Q}_m^{\clubsuit}$ guarantee us that $\mathcal{Q}_m^h(\text{XST}) \cap \partial\mathcal{Q}_m^h(\neg\text{XST}) \neq \emptyset$. Now, suppose $\mathbf{Q} \in \mathcal{Q}_m^h(\text{XST}) \cap \partial\mathcal{Q}_m^h(\neg\text{XST})$ to be arbitrary but fixed. By Lem. 3.3 there exists $\sigma \in \mathbb{S}_m$ such that $q_{\sigma(i),\sigma(i+1)} > 1/2$ for all $i \in [m-1]$, and $\mathbf{Q} \in \mathcal{Q}_m^h(\text{XST}) \subseteq \mathcal{Q}_m^h(\text{WST})$ lets us infer that $q_{\sigma(i),\sigma(j)} > 1/2$ holds for any $(i,j) \in (m)_2$. By considering $(q_{\sigma(i),\sigma(j)})_{1 \leq i,j \leq m}$ instead of $\mathbf{Q}$, we may suppose from now on w.l.o.g. $\sigma = \text{id}$, i.e., $q_{i,j} > 1/2$ for all $(i,j) \in (m)_2$.

According to (5.2), $\mathbf{Q} \in \partial\mathcal{Q}_m^h(\neg\text{XST})$ assures the existence of a pair $(i,k) \in (m)_2$ s.t. $q_{i,k} = g_{\text{XST}}(q_{i,j}, q_{j,k})$ for at least one $j \in \{i+1, \ldots, k-1\}$. Since $\mathbf{Q} \in \mathcal{Q}_m^h$, we have $q_{i,k} > 1/2+h$. Hence, for any $r \in [0, q_{i,k} - (1/2+h))$ the relation $\mathbf{Q}(r) = (q(r)_{i',j'})_{1 \leq i',j' \leq m} \in \mathcal{Q}_m$, defined via

$$q(r)_{i',j'} := \begin{cases} q_{i',j'} - r, & \text{if } (i',j') = (i,k), \\ q_{i',j'}, & \text{otherwise} \end{cases}$$

for any $(i',j') \in (m)_2$, fulfills $\mathbf{Q}(r) \in \mathcal{Q}_m^h$. We have $\mathbf{Q}(0) = \mathbf{Q}$, which is an element of $\mathcal{Q}_m^h(\text{XST})$, whereas $q(r)_{i,j} = q_{i,j} > 1/2$, $q(r)_{j,k} > 1/2$ and

$$q(r)_{i,k} = q_{i,k} - r = g_{\text{XST}}(q_{i,j}, q_{j,k}) - r = g_{\text{XST}}(q(r)_{i,j}, q(r)_{j,k}) - r$$

reveal that $\mathbf{Q}(r) \in \mathcal{Q}_m^h(\neg\text{XST})$ for any $r \in (0, q_{i,k} - (1/2+h))$.

To show $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$, suppose we are given a coin C with unknown head probability $p \in \{q_{i,k}, q_{i,k} - r\}$, where $r \in (0, q_{i,k} - (1/2+h))$ is fixed for the moment. Define $\tilde{\mathcal{A}}$ to be the algorithm, which simulates $\mathcal{A}$ in the following way: Whenever $\mathcal{A}$ makes a query $\{i',j'\}$ for some $(i',j') \in (m)_2 \setminus \{(i,k)\}$, then we provide as feedback a sample $\sim \text{Ber}(q(r)_{i',j'})$, which is independent of the rest. If $\mathcal{A}$ queries $\{i,k\}$, we throw coin C and provide its outcome as feedback. Then, we terminate as soon as $\mathcal{A}$ terminates, and output 0 in case $\mathbf{D}(\mathcal{A}) = \text{XST}$ and 1 if $\mathbf{D}(\mathcal{A}) = \neg\text{XST}$. Since $\mathcal{A}$ solves $\mathcal{P}_{\text{XST}}^{m,h,\gamma,\gamma}$, $\tilde{\mathcal{A}}$ is able to test

$$\mathbf{H}_0 : p = q_{i,k} \qquad \mathbf{H}_1 : p = q_{i,k} - r$$

with type I/II errors of at most γ. According to Cor. 2.15, for small r, $\tilde{\mathcal{A}}$ has to throw the coin C in any case (i.e., if $p = q_{i,k}$ and if $p = q_{i,k} - r$) in expectation at least $c(4-2r)^2/r^2 \ln \frac{1}{\gamma}$ times for some constant $c > 0$, which does not depend on r. We obtain

$$\min\left\{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}], \mathbb{E}_{\mathbf{Q}(r)}[T^{\mathcal{A}}]\right\} = \min_{p \in \{q_{i,k}, q_{i,k}-r\}} \mathbb{E}_p\left[T^{\tilde{\mathcal{A}}}\right] \geq \frac{c(4-2r)^2}{r^2} \ln \frac{1}{\gamma}.$$

As this holds for any sufficiently small $r \in (0, q_{i,k} - (1/2+h))$, taking the limit $r \searrow 0$ reveals $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$. □

For XST ≠ WST, the negative results above indicate that XST testing under the low-noise assumption is in the worst-case infeasible to some extent. For this reason, we will mainly focus on WST testing throughout the rest of this chapter.

5.2. Lower Bounds for WST Testing

In this section, we provide lower bounds on the expected termination time of any algorithm solving $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$. Similar to Prop. 5.3, these results are obtained by reducing a testing problem for the biases of independent coins to $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$. A sample complexity analysis of the latter testing problem results in the bounds stated below, where we write as usual $\bar{q}_{i,j} = |q_{i,j} - 1/2|$.

Theorem 5.4. *Let $h_0, \gamma_0 \in (0, 1/2)$ be fixed, $h \in (0, h_0)$, $\alpha, \beta \in (0, \gamma_0)$ and $m \in \mathbb{N}_{\geq 3}$. Suppose $\mathcal{A}$ is an algorithm that solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$, and let $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{WST})$ be arbitrary. Define $\gamma := \max\{\alpha, \beta\}$ and $\sigma = \sigma_{\mathbf{Q}}$. Then, there exists a constant $c = c(h_0, \gamma_0) > 0$ such that*

$$\begin{aligned} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] &\geq \sum\nolimits_{1 \leq i < j-1 < m} \frac{1-2\gamma}{2\bar{q}_{\sigma(i),\sigma(j)}} \left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln\left((1/2+\bar{q}_{\sigma(i),\sigma(j)})/(1/2-\bar{q}_{\sigma(i),\sigma(j)})\right)} \right\rceil \\ &\geq c \sum\nolimits_{1 \leq i < j-1 < m} \frac{1}{\bar{q}^2_{\sigma(i),\sigma(j)}} \ln \frac{1}{\gamma} \geq \frac{c}{h^2} \binom{m-1}{2} \ln \frac{1}{\gamma}. \end{aligned} \tag{5.3}$$

Thus, $\sup_{\mathbf{Q} \in \mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}]$ is in $\Omega(\frac{m^2}{h^2} \ln \frac{1}{\gamma})$.

Proof. As any (probabilistic) algorithm $\mathcal{A}$, which solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$ also solves the easier problem $\mathcal{P}^{m,h,\gamma,\gamma}_{\mathrm{WST}}$, we may suppose w.l.o.g. $\alpha = \beta = \gamma$ from now on. Moreover, by replacing $\mathbf{Q}$ with $(q_{\sigma(i),\sigma(j)})_{1 \leq i,j \leq m}$ we may suppose w.l.o.g. $\sigma = \mathrm{id}$ in the following, i.e., $q_{i,i+1} > 1/2$ for any $i \in [m-1]$. Suppose for the moment $\mathbf{Q}' = (q'_{i,j})_{1 \leq i,j \leq m} \in \mathcal{Q}^h_m$ to be arbitrary with $q'_{i,i+1} = q_{i,i+1}$ for all $i \in [m-1]$. As by assumption $q_{i,i+1} > 1/2$ for all $i \in [m-1]$, we obtain by the definition of WST that

$$\mathbf{Q}' \in \mathcal{Q}^h_m(\mathrm{WST}) \quad \Leftrightarrow \quad q'_{i,j} > 1/2 \text{ for all } 1 \leq i < j-1 \leq m-1.$$

Hence, in particular $\mathbf{Q}$ fulfills $q_{i,j} > 1/2$ for all $1 \leq i < j-1 \leq m-1$. Since $\mathcal{A}$ solves $\mathcal{P}^{m,h,\gamma,\gamma}_{\mathrm{WST}}$, it is able to decide

$$\begin{aligned} \mathbf{H}'_0 &: \forall 1 \leq i < j-1 \leq m-1 : q'_{i,j} \geq 1/2 \\ \mathbf{H}'_1 &: \exists 1 \leq i < j-1 \leq m-1 : q'_{i,j} < 1/2 \end{aligned}$$

for any

$$(q'_{i,j})_{1 \leq i < j-1 \leq m-1} \in \prod_{1 \leq i < j-1 \leq m-1} \{1/2 \pm \bar{q}_{i,j}\}$$

with an error probability of at most γ in any case.[2] Regarding that $q'_{i,j} = 1/2 + \bar{q}_{i,j}$ for all $1 \leq i < j-1 \leq m-1$ implies $\mathbf{Q}' = \mathbf{Q}$, Thm. 2.29 thus yields

$$\begin{aligned} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] &\geq \sum_{1 \leq i < j-1 \leq m-1} \frac{1-2\gamma}{2\bar{q}_{i,j}} \left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln\left((1/2+\bar{q}_{i,j})/(1/2-\bar{q}_{i,j})\right)} \right\rceil \\ &\geq c \sum_{1 \leq i < j-1 \leq m-1} \frac{1}{\bar{q}^2_{i,j}} \ln \frac{1}{\gamma} \end{aligned}$$

[2] More precisely, define $\mathcal{A}'$ to be the algorithm, which is given sample access to $\mathrm{Ber}(q'_{i,j}), 1 \leq i < j-1 \leq m-1$, runs $\mathcal{A}$ on $\mathbf{Q}'$ and terminates with the output of $\mathcal{A}$ as soon as $\mathcal{A}$ terminates. Then, this algorithm is able to decide with an error probability of at most γ whether $\mathbf{H}'_0$ or $\mathbf{H}'_1$ is fulfilled.

with $c = c(h_0, \gamma_0)$ as in Thm. 2.29. The assumption $\mathbf{Q} \in \mathcal{Q}_m^h$ assures that this quantity is bounded from below by $\frac{c}{h^2}\binom{m-1}{2}\frac{1}{\gamma}$, which is in $\Omega(\frac{m^2}{h^2}\ln\frac{1}{\gamma})$ as $\max\{m, \frac{1}{h}, \frac{1}{\gamma}\} \to \infty$. $\square$

Note that the right-hand side of (5.3) is of the order $\frac{m^2}{h^2}\ln\frac{1}{\gamma}$, which is coherent with the results shown later on in Sec. 5.5. The fact that the instance-wise bound only depends on $\binom{m-1}{2}$ instead of all $\binom{m}{2}$ entries of $\mathbf{Q}$ is due to our proof technique, which is nonetheless w.r.t. m asymptotically of the same order.

Let us now consider the more complex case $h = 0$. As any solution to $\mathcal{P}_{\mathrm{WST}}^{m,0,\alpha,\beta}$ is also a solution to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$ for any $h \in (0, 1/2)$, Thm. 5.4 is applicable in this case. However, we can slightly improve upon this. For this, recall the notion of $\Omega_{\sup}$ from Sec. 1.5.

Theorem 5.5. *Let $\alpha, \beta \in (0, 1/2)$ be fixed and suppose $\mathcal{A}$ to be an algorithm that solves $\mathcal{P}_{\mathrm{WST}}^{m,0,\alpha,\beta}$. Then, the following holds:*

(a) $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$ *for any* $\mathbf{Q}$ *in a set* $\emptyset \neq \mathcal{Q}_m^{\dagger} \subsetneq \partial\mathcal{Q}_m(\mathrm{WST}) \cap \partial\mathcal{Q}_m(\neg\mathrm{WST})$*,*

(b) $\sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \Omega(\frac{m^2}{h^2}) \cap \Omega_{\sup}(\frac{1}{h^2}\ln\ln\frac{1}{h})$ *as* $\max\{m, \frac{1}{h}\} \to \infty$.

As we point out in the proof of this theorem the set $\mathcal{Q}_m^{\dagger}$ in (a) can be chosen as the set of all $\mathbf{Q} \in \mathcal{Q}_m$, for which some permutation σ on $[m]$ exists such that the following conditions are fulfilled:

$$\begin{aligned} \forall 1 \le i < j \le m : \; & q_{\sigma(i),\sigma(j)} \ge 1/2, \\ \forall i \in [m-1] : \; & q_{\sigma(i),\sigma(i+1)} > 1/2, \\ \exists 1 \le i' < j' - 1 \le m - 1 : \; & q_{\sigma(i'),\sigma(j')} = 1/2. \end{aligned}$$

In the proof of the theorem, to make (b) more explicit, we provide several examples for a family $\{\mathbf{Q}(h)\}_{h\in(0,1/2)} \subseteq \mathcal{Q}_m^h(\mathrm{WST})$, for which

$$\limsup_{h\searrow 0} \frac{\mathbb{E}_{\mathbf{Q}(h)}[T^{\mathcal{A}}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \ge \frac{1-2\gamma}{2}.$$

Regarding the occurrence of the limes superior in Prop. 2.13, this is the best we may infer from Prop. 2.13.

Proof of Thm. 5.5. Similarly as in the proof of Thm. 5.4, we suppose w.l.o.g. $\alpha = \beta = \gamma$ for convenience. We start with the proof of (b). For this, suppose $h_{1,2}, \ldots, h_{m-1,m} \in (\frac{1}{2}, 1)$ to be fixed and define

$$\tilde{\mathbf{Q}}(h) \coloneqq ({}^1\!/_2)_{1\le i,j\le m} + \begin{pmatrix} - & h_{1,2} & h & \cdots & \cdots & h \\ & - & h_{2,3} & h & \cdots & \vdots \\ & & \ddots & \ddots & \ddots & \vdots \\ & & & - & h_{m-2,m-1} & h \\ & & & & - & h_{m-1,m} \\ & & & & & - \end{pmatrix} \in \mathcal{Q}_m$$

and for $1 \le i < j - 1 \le m - 1$ the relation $\tilde{\mathbf{Q}}^{[i,j]}(h) = (\tilde{q}^{[i,j]}(h)_{i',j'})_{1\le i',j'\le m} \in \mathcal{Q}_m$ via

$$\forall 1 \le i' < j' \le m : \tilde{q}^{[i,j]}(h)_{i',j'} = \begin{cases} 1/2 + h, & \text{if } (i', j') = (i, j), \\ 1/2 + h_{i',j'}, & \text{otherwise.} \end{cases}$$

We will show in the following that

$$\limsup_{h\searrow 0} \frac{\mathbb{E}_{\tilde{\mathbf{Q}}(h)}[T^{\mathcal{A}}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \geq \frac{1-2\gamma}{2} \quad \text{and} \quad \limsup_{h\searrow 0} \frac{\mathbb{E}_{\tilde{\mathbf{Q}}^{[i,j]}(h)}[T^{\mathcal{A}}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \geq \frac{1-2\gamma}{2}$$

holds for any $1 \leq i < j-1 \leq m-1$, which is a stronger result than (b) in the statement of this theorem.

Let $c(\gamma) := \frac{1}{2}(1-2\gamma) > 0$. At first, let $(h_{i,j})_{1\leq i<j\leq m} \in (1/2,1]^{\binom{m}{2}}$ and $1 \leq i < j-1 \leq m-1$ be fixed and $\tilde{\mathbf{Q}}^{[i,j]}(h)$ be defined as above. Regarding the definition of weak stochastic transitivity, we have the equivalence

$$\tilde{\mathbf{Q}}^{[i,j]}(h) \in \mathcal{Q}_m(\mathrm{WST}) \quad \Leftrightarrow \quad 1/2+h > 1/2.$$

As $\mathcal{A}$ solves $\mathcal{P}^{m,0,\gamma,\gamma}_{\mathrm{WST}}$, $\mathcal{A}$ can be used[3] to decide whether the (unknown) bias p of a coin C fulfills $p > 1/2$ or $p < 1/2$ with an error probability of at most γ in any case. As Prop. 2.13 reveals, it has to throw the coin sufficiently often for this and we obtain

$$\limsup_{h\searrow 0} \frac{\mathbb{E}_{\tilde{\mathbf{Q}}^{[i,j]}(h)}[T^{\mathcal{A}}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \geq c(\gamma).$$

Since $\tilde{\mathbf{Q}}^{[i,j]}(h) \in \mathcal{Q}_m^{\tilde{h}}$ holds for any $\tilde{h} \in (0, \min\{h, h_{1,2},\ldots,h_{m-1,m}\})$, we infer the bound $\sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \Omega_{\sup}(\frac{1}{h^2}\ln\ln\frac{1}{h})$. Hence, in combination with the lower bound from Thm. 5.4 this shows $\sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \Omega(\frac{m^2}{h^2}) \cap \Omega_{\sup}(\frac{1}{h^2}\ln\ln\frac{1}{h})$ as $\max\{m, \frac{1}{h}\} \to \infty$.

To see the asymptotic behaviour on instances $\tilde{\mathbf{Q}}(h)$, note at first that, similarly as above,

$$\tilde{\mathbf{Q}}(h) \in \mathcal{Q}_m(\mathrm{WST}) \quad \Leftrightarrow \quad 1/2+h > 1/2 \tag{5.4}$$

follows from the definition of weak stochastic transitivity. Suppose now we are given sample access to iid coins $C_{i,j} \sim \mathrm{Ber}(p)$, $1 \leq i < j-1 \leq m-1$, where the bias p is unknown. Define $\mathcal{A}'$ to be the algorithm, which simulates $\mathcal{A}$ on $\tilde{\mathbf{Q}}(h)$ in the following way: Whenever $\mathcal{A}$ makes a query of the form $\{i, i+1\}$ it obtains a sample $\sim \mathrm{Ber}(1/2 + h_{i,i+1})$ as answer and whenever it makes a query $\{i,j\}$, $1 \leq i < j-1 \leq m-1$, it obtains as feedback a sample of the coin $C_{i,j}$. Then, $\mathcal{A}'$ terminates as soon as $\mathcal{A}$ terminates and returns 0 if $\mathcal{A}$ outputs WST and returns 1 if $\mathcal{A}$ outputs $\neg$WST. As $\mathcal{A}$ solves $\mathcal{P}^{m,0,\gamma,\gamma}_{\mathrm{WST}}$ and (5.4) holds, $\mathcal{A}$ is able to test whether $p > 1/2$ or $p < 1/2$ with error at most γ. Moreover, as the coins are iid and throws of one coin are independent, the behaviour of $\mathcal{A}'$ (i.e., its type I/II errors and its termination time) is the same if we replace each throw of a coin $C_{i,j}$ by a throw of a single coin $C \sim \mathrm{Ber}(p)$. This modification of $\mathcal{A}'$, denoted by $\mathcal{A}''$, tests with an error probability of at most γ whether one coin $C \sim \mathrm{Ber}(p)$ with unknown bias p fulfills $p > 1/2$ or $p < 1/2$. Consequently, Prop. 2.13 assures us that

$$\limsup_{h\searrow 0} \frac{\mathbb{E}_{\tilde{\mathbf{Q}}(h)}[T^{\mathcal{A}}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} = \limsup_{h\searrow 0} \frac{\mathbb{E}_{1/2+h}[T^{\mathcal{A}''}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \geq c(\gamma),$$

[3] More precisely: Given sample access to a coin $C \sim \mathrm{Ber}(p)$ with $p = 1/2 + h$, simulate $\mathcal{A}$ on $\tilde{\mathbf{Q}}^{[i,j]}(h)$. I.e., whenever $\{i,j\}$ is queried, throw coin C, and if another query $\{i',j'\}$ is made, draw a sample $\sim \mathrm{Ber}(1/2 + h_{i',j'})$ instead. Then, terminate as soon as $\mathcal{A}$ terminates and return 0 if $\mathbf{D}(\mathcal{A}) = \mathrm{WST}$, and 1 if $\mathbf{D}(\mathcal{A}) = \neg\mathrm{WST}$.

which shows (b).

To prove (a), recall that $\mathbb{S}_m$ is the set of all permutations on $[m]$ and define

$$\mathcal{Q}_m^\dagger := \big\{ \mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST}) \,|\, \exists \sigma \in \mathbb{S}_m \text{ s.t. } \forall 1 \le i < j \le m : q_{\sigma(i),\sigma(j)} \ge 1/2 \\ \text{and } \forall i \in [m-1] : q_{\sigma(i),\sigma(i+1)} > 1/2 \\ \text{and } \exists 1 \le i' < j'-1 \le m-1 : q_{\sigma(i'),\sigma(j')} = 1/2 \big\}.$$

Suppose $\mathbf{Q} \in \mathcal{Q}_m^\dagger$ and let $\sigma \in \mathbb{S}_m$ be such that $\forall 1 \le i < j \le m : q_{\sigma(i),\sigma(j)} \ge 1/2$ and $\forall i \in [m-1] : q_{\sigma(i),\sigma(i+1)} > 1/2$ and $\exists 1 \le i' < j'-1 \le m-1 : q_{\sigma(i'),\sigma(j')} = 1/2$ are fulfilled. By considering $(q_{\sigma(i),\sigma(j)})_{1\le i,j\le m}$ instead of $\mathbf{Q}$, we may suppose w.l.o.g. $\sigma = \mathrm{id}$.

At first, let us note that $\mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST})$ holds: Indeed, assuming weak stochastic transitivity of $\mathbf{Q}$, the (in-)equalities

$$q_{j',i'} = 1/2, \quad q_{i',i'+1} > 1/2, \quad q_{i'+1,i'+2} > 1/2 \quad \ldots \quad q_{j'-2,j'-1} > 1/2$$

would assure that $q_{j',j'-1} \ge 1/2$ holds, which clearly contradicts $q_{j'-1,j'} > 1/2$.

Now, let us define the set J of all $(i,j) \in (m)_2$, for which $q_{i,j} = 1/2$. By our assumption on $\mathbf{Q}$, we have $|J| \ge 1$ and $(i,i+1) \notin J$ for any $i \in [m-1]$. For any $h \in [0,1/2)$ let us define the relation $\mathbf{Q}(h) = (q(h)_{i,j})_{1\le i,j\le m} \in \mathcal{Q}_m$ via

$$q(h)_{i,j} := \begin{cases} q_{i,j}, & \text{if } (i,j) \in (m)_2 \setminus J, \\ 1/2 + h, & \text{if } (i,j) \in J \end{cases}$$

for any $(i,j) \in (m)_2$. In case $h > 0$, $q(h)_{i,j} > 1/2$ for any $(i,j) \in (m)_2$ shows that $\mathbf{Q}(h) \in \mathcal{Q}_m(\mathrm{WST})$. Since $[0,1/2) \to \mathcal{Q}_m, h \mapsto \mathbf{Q}(h)$ is continuous with $\mathbf{Q}(0) = \mathbf{Q}$, we infer $\mathbf{Q} \in \partial\mathcal{Q}_m(\mathrm{WST}) \cap \partial\mathcal{Q}_m(\neg\mathrm{WST})$.

Suppose we have a coin C with unknown head probability $p \in \{1/2, 1/2+h\}$, where $h \in (0,1/2)$ is fixed for the moment. Let $\tilde{\mathcal{A}}$ be the algorithm, which simulates $\mathcal{A}$ in the following way: Whenever $\mathcal{A}$ makes a query $\{i,j\}$ for some $(i,j) \in (m)_2 \setminus J$, then provide as feedback a sample $\sim \mathrm{Ber}(q(h)_{i,j})$, which is independent of the rest. If $\mathcal{A}$ makes a query $\{i,j\}$ for some $(i,j) \in J$ instead, then throw coin C and provide its outcome as feedback. Terminate as soon as $\mathcal{A}$ terminates, and output 0 in case $\mathbf{D}(\mathcal{A}) = \mathrm{WST}$ and 1 in case $\mathbf{D}(\mathcal{A}) = \neg\mathrm{WST}$. As $\mathcal{A}$ solves $\mathcal{P}^{m,0,\gamma,\gamma}_{\mathrm{WST}}$, $\tilde{\mathcal{A}}$ is able to test

$$\mathbf{H}_0 : p = 1/2 \qquad \mathbf{H}_1 : p = 1/2 + h$$

with type I/II errors of at most γ. From Cor. 2.15 we infer that, for any $h < 1/4$, $\tilde{\mathcal{A}}$ has to throw the coin C in expectation at least $c\,(h/(4-2h))^{-2}\ln(\gamma^{-1})$ times in any of the cases $p \in \{1/2, 1/2+h\}$, where $c = c(\varepsilon_0, \gamma_0) > 0$ is the constant from Cor. 2.15 with arbitrary but fixed $\varepsilon_0 \in (1/2, 1)$ and $\gamma_0 \in (\gamma, 1/2)$. As the throws of the coins as well as the feedback generated by the arm pairs in our setting are independent, we thus obtain

$$\min\big\{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}], \mathbb{E}_{\mathbf{Q}(h)}[T^{\mathcal{A}}]\big\} \ge \min\nolimits_{p\in\{1/2,1/2+h\}} \mathbb{E}_p[T^{\tilde{\mathcal{A}}}] \ge \frac{c(4-2h)^2}{h^2} \ln\frac{1}{\gamma}.$$

Since this holds for any $h < 1/4$, taking the limit $h \searrow 0$ yields $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$. □

5.3. Approach I: Conducting Multiple Binomial Tests

Guided by our findings in Sec. 5.2, we now focus on the WST testing problem in the framework developed in Sec. 1.1. Note that WST is in any case of particular interest for the ranking problem in dueling bandits, as it is both a sufficient and a necessary condition for the existence of a ranking over the arms consistent with the preference relation $\mathbf{Q}$, in the sense that an arm i is preferred over an arm j if and only if $q_{i,j} \geq 1/2$.

Algorithm 20 $\mathcal{A}_{\text{naive}}$

Parameters: m, a sampling strategy π, a testing algorithm $\mathcal{A}_{\text{Coin}} \in \mathfrak{A}_{\text{Coin}}$.
Initialization: For any $(i,j) \in (m)_2$ let $\mathcal{A}^{i,j}_{\text{Coin}}$ be an instance of $\mathcal{A}_{\text{Coin}}$
$\boldsymbol{n}_0 \leftarrow (0)_{1 \leq i,j \leq m}$, $\boldsymbol{w}_0 \leftarrow (0)_{1 \leq i,j \leq m}$, $\mathbf{Q}' \leftarrow (1/2)_{1 \leq i,j \leq m} \in \mathcal{Q}_m$, $\hat{E}_0 \leftarrow \emptyset$

for $t \in \mathbb{N}$ **do**
    $\{i,j\} \sim \pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0 \leq t' \leq t-1})$, w.l.o.g. $i < j$
    $\hat{E}_t \leftarrow \hat{E}_{t-1}$
    Observe $X^{[t]}_{i,j} \sim \text{Ber}(q_{i,j})$
    Update $\boldsymbol{n}_t$ and $\boldsymbol{w}_t$ according to (4.8)
    Reveal $X^{[t]}_{i,j}$ to $\mathcal{A}^{i,j}_{\text{Coin}}$
    if $\mathcal{A}^{i,j}_{\text{Coin}}$ terminated with $\mathbf{D}(\mathcal{A}^{i,j}_{\text{Coin}}) = 0$ **then**
        $q'_{i,j} \leftarrow 1$, $q'_{j,i} \leftarrow 0$, $\hat{E}_t \leftarrow \hat{E}_t \cup \{(i,j)\}$
    if $\mathcal{A}^{i,j}_{\text{Coin}}$ terminated with $\mathbf{D}(\mathcal{A}^{i,j}_{\text{Coin}}) = 1$ **then**
        $q'_{i,j} \leftarrow 0$, $q'_{j,i} \leftarrow 1$, $\hat{E}_t \leftarrow \hat{E}_t \cup \{(j,i)\}$
    if $|\hat{E}_t| = \binom{m}{2}$ and $\mathbf{Q}' \in \mathcal{Q}_m(\text{WST})$ **then return** WST
    else if $|\hat{E}_t| = \binom{m}{2}$ and $\mathbf{Q}' \in \mathcal{Q}_m(\neg\text{WST})$ **then return** $\neg$WST

A first naive approach for a testing component for the passive scenario (cf. Sec. 1) is Alg. 20, which does the following: Terminate as soon as we can decide, *for every* $(i,j) \in (m)_2$, each with error probability at most $\gamma' = \min\{\alpha, \beta\}/\binom{m}{2}$, whether $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ holds, and output WST if an auxiliary relation $\mathbf{Q}'$ generated during runtime is WST, and $\neg$WST otherwise. To construct $\mathbf{Q}'$, the value $q'_{i,j}$ is set to 1 resp. 0 whenever we are sure enough (for the first time) that $q_{i,j} > 1/2$ resp. $q_{i,j} < 1/2$ holds. Here, testing the sign of $q_{i,j} - 1/2$ with confidence level $1 - \gamma'$ may be done by means of an instance $\mathcal{A}^{i,j}_{\text{Coin}}$ of a testing algorithm $\mathcal{A}_{\text{Coin}} \in \mathfrak{A}_{\text{Coin}}$ that solves $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$.

As the following theorem shows, this simple approach results in a solution to $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$. In its proof, we may make use of the notation $\mathbf{D}(\mathcal{A}_{\text{Coin}}, t)$ introduced in (4.10).

Theorem 5.6. *Let* $m \in \mathbb{N}_{\geq 3}$*,* $\alpha, \beta \in (0,1)$*, and* $h \in [0, 1/2)$ *be fixed, define* $\gamma' := \min\{\alpha, \beta\}/\binom{m}{2}$ *and let* $\mathcal{A}_{\text{Coin}} \in \mathfrak{A}_{\text{Coin}}$ *be a solution to* $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$*. For any* $\pi \in \Pi_\infty$*,* $\mathcal{A}$ *:=Alg. 20 instantiated with parameters* m*,* π *and* $\mathcal{A}_{\text{Coin}}$ *is a solution to* $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$*.*

Proof. Abbreviate $T := T^{\mathcal{A}}$ and $\{i(t), j(t)\}$, $i(t) < j(t)$, for the query sampled from π at the t-th iteration of the loop. The testing algorithm $\mathcal{A}^{i,j}_{\text{Coin}}$ observes the sample $X^{[t]}_{i,j} \sim \text{Ber}(q_{i,j})$ iff $(i(t), j(t)) = (i,j)$, and after time t it has observed exactly $(\boldsymbol{n}_t)_{i,j}$ of these samples. Let $\mathbf{Q} \in \mathcal{Q}^h_m$ be fixed for the moment and recall $\gamma' = \frac{\alpha}{m} \wedge \frac{\beta}{m-1}$. We split the remaining proof into two parts.

Part 1: Almost sure finiteness of T
As in the proof of Thm. 4.6, we may suppose for the proof of Part 1 w.l.o.g. that $\mathcal{A}$ does not terminate before all of the internal testing algorithms $\mathcal{A}^{i,j}_{\text{Coin}}$ have terminated. By assumption on π, we have $(\boldsymbol{n}_t)_{i,j} \to \infty$ almost surely as $t \to \infty$ for each $(i,j) \in (m)_2$. For any $(i,j) \in (m)_2$, $\mathcal{A}^{i,j}_{\text{Coin}}$ solves by assumption $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$ and hence the stopping time

$$T_{i,j} = \min\left\{t \in \mathbb{N} \,:\, \mathbf{D}\left(\mathcal{A}^{i,j}_{\text{Coin}}, (\boldsymbol{n}_t)_{i,j}\right) \neq \text{“N/A”}\right\}$$

is a.s. finite. In particular,

$$T' := \max\nolimits_{(i,j)\in(m)_2} T_{i,j}$$

is a.s. finite. Since an edge between i and j is added to $\hat{E}_t$ iff $t = T_{i,j}$, we have in particular $|\hat{E}_{T'}| = \binom{m}{2}$. Regarding that either $\mathbf{Q}' \in \mathcal{Q}_m(\text{WST})$ or $\mathbf{Q}' \in \mathcal{Q}_m(\neg\text{WST})$ holds, this shows $T = T'$, i.e., $\mathcal{A}$ terminates a.s. This completes the proof of Part 1.

Part 2: Correctness of $\mathcal{A}$
Let $(i,j) \in (m)_2$ be arbitrary. As $\mathcal{A}^{i,j}_{\text{Coin}}$ solves $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$, we obtain in case $q_{i,j} > 1/2 + h$ that

$$\mathbb{P}_{\mathbf{Q}}(q'_{i,j} = 0) = \mathbb{P}_{\mathbf{Q}}\left(\mathbf{D}\left(\mathcal{A}^{i,j}_{\text{Coin}}\right) = 1\right) \leq \gamma'$$

and similarly in case $q_{i,j} < 1/2 - h$ that

$$\mathbb{P}_{\mathbf{Q}}(q'_{i,j} = 1) = \mathbb{P}_{\mathbf{Q}}\left(\mathbf{D}\left(\mathcal{A}^{i,j}_{\text{Coin}}\right) = 0\right) \leq \gamma'.$$

Consequently, we have with probability $\geq 1 - \binom{m}{2}\gamma'$ that

$$\forall (i,j) \in (m)_2 : q_{i,j} > 1/2 \Leftrightarrow q'_{i,j} > 1/2,$$

which is a sufficient condition for $\mathbf{Q} \in \mathcal{Q}_m(\text{WST}) \Leftrightarrow \mathbf{Q}' \in \mathcal{Q}_m(\text{WST})$. Regarding the definition of γ', $\mathcal{A}$ thus fulfills

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \text{WST}) \geq 1 - \alpha \quad \text{if } \mathbf{Q} \in \mathcal{Q}^h_m(\text{WST})$$

and

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) = \neg\text{WST}) \geq 1 - \beta \quad \text{if } \mathbf{Q} \in \mathcal{Q}^h_m(\neg\text{WST}),$$

which completes the proof. □

By construction, the sample complexity of Alg. 20 is exactly the number of iterations that are required for testing the signs of all $q_{i,j} - 1/2$, $(i,j) \in (m)_2$. By choosing $\mathcal{A}_{\text{Coin}}$ as the non-sequential solution to $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$ from Lem. 2.10, which we denote by $\mathcal{A}^{\text{Hoeffding}}_{\text{Coin}}(h,\gamma')$ throughout this thesis, testing the sign of $q_{i,j} - 1/2$ requires in any case exactly $N := \lceil \frac{1}{2h^2} \ln \frac{1}{\gamma} \rceil$ iid samples governed by $\text{Ber}(q_{i,j})$. However, the explicit time at which a pair has been sampled at least N times highly depends on the underlying sampling strategy π, so that an analysis of the sample complexity of $\mathcal{A}_{\text{naive}}$ can only be done w.r.t. the corresponding sampling strategy π. As the testing component is working in parallel to π in the passive setting, i.e., it has no influence on the behavior of π, the minimum requirement for a test component in the passive online test seems to be consistency in terms of an a.s. finite termination time and the adherence to predefined error bounds for a general class of sampling strategies. Both requirements are met by the test underlying $\mathcal{A}_{\text{naive}}$ by Thm. 5.6 for the class Π_∞ if $\mathcal{A}_{\text{naive}}$ is instantiated with a solution to $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$.

In the passive online testing scenario, i.e., the sampling strategy π is instantiated in a black-box fashion by some dueling bandits algorithm based on a transitivity assumption (such as those by Falahatgar et al. [2017a, 2018]), it might happen that π terminates before the testing algorithm came to a decision, and in particular that π is not defined any more. In this case, if one is still interested in whether transitivity was fulfilled in hindsight, one may continue sampling according to the strategy $\hat{\pi}$, which picks each query $\{i,j\} \in [m]_2$ with probability $1/\binom{m}{2}$. The other way around, if the testing algorithm came to a positive decision ($\mathbf{D}(\mathcal{A}) = \text{XST}$), although the online ranking algorithm has not yet terminated, one can simply continue the sampling strategy without the testing component. In case of a negative decision ($\mathbf{D}(\mathcal{A}) = \neg\text{XST}$), the online ranking algorithm should be interrupted due to violating the assumptions.

In the active online testing scenario, on the other side, we have the possibility to choose π in a favorable way and consequently analyze the sample complexity of Alg. 20. For this purpose, we consider a sampling strategy π_{WST}, which chooses its queries from the time-dependent set consisting of all pairs $\{i,j\}$, for which it is (according to the internal testing algorithm $\mathcal{A}^{i,j}_{\text{Coin}}$ of Alg. 20) not yet sure with confidence level $1-\gamma'$ whether $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ holds. Formally, the following set is considered:

$$U(t) \coloneqq \left\{ \{i,j\} \in [m]_2 \,|\, i < j \text{ and } \mathbf{D}\left(\mathcal{A}^{i,j}_{\text{Coin}}, (\boldsymbol{n}_{t-1})_{i,j}\right) \neq \text{“N/A”} \right\}$$

In each time t, the sampling strategy π_{WST} queries $\{i,j\} \in [m]_2$ uniformly at random from $U(t)$, if $U(t)$ is non-empty, and otherwise queries $\{i,j\} \in [m]_2$ uniformly at random from $[m]_2$. Note that the second case (i.e., $U(t)$ is empty) is only defined in order to ensure that $\pi_{\text{WST}} \in \Pi_\infty$, which in turn allows for applying Thm. 5.6. In light of this, we obtain the following corollary.

Corollary 5.7. *Let $m \in \mathbb{N}_{\geq 3}$, $h \in [0, 1/2)$, $\alpha, \beta \in (0, \gamma_0)$ for some $\gamma_0 \in (0,1)$, choose $\gamma' \coloneqq \min\{\alpha,\beta\}/\binom{m}{2}$ and let $\mathcal{A}_{\text{Coin}} \in \mathfrak{A}_{\text{Coin}}$ be a solution to $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$. Let π_{WST} be the sampling strategy from above and suppose $\mathcal{A}$ to be Alg. 20 called with parameters m, π_{WST} and $\mathcal{A}_{\text{Coin}}$. Then, $\mathcal{A}$ solves $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$ and fulfills*

$$\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] = \binom{m}{2} \sum\nolimits_{(i,j)\in(m)_2} \mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}^{i,j}_{\text{Coin}}}\right] = \binom{m}{2} \sum\nolimits_{(i,j)\in(m)_2} \mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}_{\text{Coin}}}\right]$$

In case $h > 0$ and $\mathcal{A}_{\text{Coin}} = \mathcal{A}^{\text{Hoeffding}}_{\text{Coin}}(h,\gamma')$, we obtain

$$T^{\mathcal{A}} = \binom{m}{2} \left\lceil \frac{1}{2h^2} \ln\left(\frac{m(m-1)}{2(\alpha \wedge \beta)}\right) \right\rceil \quad \mathbb{P}_{\mathbf{Q}}\text{-almost surely for all } \mathbf{Q} \in \mathcal{Q}^h_m,$$

i.e., if $\gamma = \alpha \wedge \beta$, we have

$$\sup\nolimits_{\mathbf{Q}\in\mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] \in \mathcal{O}\left(\frac{m^2 \ln m}{h^2} \ln \frac{1}{\gamma}\right).$$

Proof of Cor. 5.7. Throughout this proof, we will use for convenience the notation $T \coloneqq T^{\mathcal{A}}$. Suppose $\mathbf{Q} \in \mathcal{Q}^h_m$ to be fixed. With

$$U'(t) \coloneqq \begin{cases} U(t), & \text{if } U(t) \neq \emptyset, \\ [m]_2, & \text{otherwise,} \end{cases}$$

the sampling strategy π_{WST} may formally be defined via

$$\mathbb{P}_{\mathbf{Q}}(\pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0 \leq t' \leq t-1}) = \{i,j\}) = \begin{cases} \frac{1}{|U'(t)|}, & \text{if } \{i,j\} \in U'_C(t), \\ 0, & \text{otherwise.} \end{cases}$$

Since $\mathcal{A}_{\text{Coin}}$ solves $\mathcal{P}^{h,\gamma'}_{\text{Coin}}$, eventually $\{i,j\} \notin U(t)$ holds for all $(i,j) \in (m)_2$ for large values of t, and from then on π picks its queries $\{i,j\}$ uniformly at random from $(m)_2$. Thus, π_{WST} is an element of Π_∞ and Thm. 5.6 shows that $\mathcal{A}$ solves $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$.

With $T_{i,j}$ defined as in the proof of Thm. 5.6, $\mathcal{A}^{i,j}_{\text{Coin}}$ terminates after having seen exactly $(\boldsymbol{n}_{T_{i,j}})_{i,j}$ samples. The algorithm $\mathcal{A}$ terminates as soon as its internal statistic $\hat{E}_t$ (cf. Alg. 20) contains $\binom{m}{2}$ edges, which is equivalent to $U(t) = \emptyset$. Until then, π_{WST} does not make any query $\{i,j\} \in [m]_2 \setminus U(t)$; in other words, it queries until T the pair $\{i,j\}$ so often as is necessary for $\mathcal{A}^{i,j}_{\text{Coin}}$ to terminate, that is we have $(\boldsymbol{n}_{T_{i,j}})_{i,j} = (\boldsymbol{n}_T)_{i,j}$. Consequently, $T = \sum_{(i,j)\in(m)_2} (\boldsymbol{n}_T)_{i,j} = \sum_{(i,j)\in(m)_2} (\boldsymbol{n}_{T_{i,j}})_{i,j}$ holds and we obtain

$$\mathbb{E}_{\mathbf{Q}}[T] = \sum_{(i,j)\in(m)_2} \mathbb{E}_{\mathbf{Q}}\left[(\boldsymbol{n}_{T_{i,j}})_{i,j}\right] = \sum_{(i,j)\in(m)_2} \mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}^{i,j}_{\text{Coin}}}\right] = \sum_{(i,j)\in(m)_2} \mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}_{\text{Coin}}}\right],$$

where we have used in the last step that each $\mathcal{A}^{i,j}_{\text{Coin}}$ is an instance of $\mathcal{A}_{\text{Coin}}$.
If $h > 0$ and $\mathcal{A}_{\text{Coin}} = \mathcal{A}^{\text{Hoeffding}}_{\text{Coin}}(h,\gamma') \in \mathfrak{A}_{\text{Coin}}$ is the corresponding non-sequential testing algorithm from Lem. 2.10, each testing algorithm $\mathcal{A}^{i,j}_{\text{Coin}}$ terminates a.s. after having seen exactly $\left\lceil \frac{1}{2h^2} \ln\left(\frac{m(m-1)}{2(\alpha\wedge\beta)}\right)\right\rceil$ samples. Therefore, we have a.s.

$$T = \sum_{(i,j)\in(m)_2} (\boldsymbol{n}_T)_{i,j} = \binom{m}{2} \left\lceil \frac{1}{2h^2} \ln\left(\frac{m(m-1)}{2(\alpha\wedge\beta)}\right)\right\rceil.$$

□

With regard to Thm. 5.4, the testing algorithm from Cor. 5.7 is a solution to $\mathcal{P}^{m,h,\gamma}_{\text{WST}}$ that is w.r.t. the worst-case sample complexity already asymptotically optimal up to logarithmic factors. Nevertheless, one may ask, firstly, whether termination is only possible as soon as being sure about the signs of $q_{i,j} - 1/2$ of *all* the $\binom{m}{2}$ many $\{i,j\} \in [m]_2$, and secondly, if the rough correction term in the error probability (i.e., $\binom{m}{2}$) for the sign test of any $q_{i,j} - 1/2$, is optimal. In the following, we answer both questions negatively, giving rise to more sophisticated testing procedures. Moreover, we also present instance-wise upper bounds for $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$.

5.3.1. Enhanced Online WST Testing

In this section, we will improve upon the algorithm from Cor. 5.7 by exploiting the graph theoretical results from Sec. 3. Recall that we have defined for $\mathbf{Q} \in \mathcal{Q}_m$ the tournament $G(\mathbf{Q}) \in \overline{\mathcal{G}}_m$ via

$$i \to j \text{ in } G(\mathbf{Q}) \quad \Leftrightarrow \quad q_{i,j} > 1/2.$$

As seen in Prop. 3.1, any $\mathbf{Q} \in \mathcal{Q}^0_m$ is WST iff $G(\mathbf{Q})$ is WST. Thus, to test whether such $\mathbf{Q}$ is WST or not, one may initialize $\hat{E} = \emptyset$, and, whenever sure enough that $q_{i,j} > 1/2$, add the edge (i,j) to $\hat{E}$. Provided all edges in $\hat{E}$ are correct, $([m],\hat{E})$ is a subgraph of $G(\mathbf{Q})$. In case $([m],\hat{E})$ is acyclic in extension resp. non-acyclic in extension, we could thus correctly conclude that $G(\mathbf{Q})$ was acyclic resp. non-acyclic, which implies $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{WST})$ resp. $\mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST})$. By Prop. 3.22, acyclicity in extension resp. non-acyclicity of $([m],\hat{E})$ would be determined by at most $\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ resp. m of its edges, hence for an overall type I/II error of α/β it would be sufficient to be sure about each edge in $\hat{E}$ with confidence $1 - \min\{\frac{\alpha}{m}, \beta(\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor)^{-1}\}$.

Following these ideas results in Alg. 21, which we suggest as a testing procedure for $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$. In addition to $\mathcal{A}^{\mathrm{Hoeffding}}_{\mathrm{Coin}}$ and $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{Coin}}$, we will also consider the optimal solution to $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ from Prop. 2.22, denoted by $\mathcal{A}^{\mathrm{Farrell}}_{\mathrm{Coin}}(\gamma)$, as a particular choice for $\mathcal{A}_{\mathrm{Coin}}$ in the analysis of Alg. 21. In the next theorem, we verify that this algorithm has in fact the desired theoretical guarantees; its proof is deferred to Sec. 5.3.3. In contrast to Thm. 8.4 in [Haddenhorst et al., 2021b], Thm. 5.8 also allows $h = 0$.

Algorithm 21 : $\mathcal{A}_{\mathrm{improved}}$

Parameters: m, a sampling strategy π, a testing algorithm $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$
Initialization: For any $(i,j) \in (m)_2$ let $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ be an instance of $\mathcal{A}_{\mathrm{Coin}}$,
$\boldsymbol{n}_0 \leftarrow (0)_{1\le i,j\le m}$, $\boldsymbol{w}_0 \leftarrow (0)_{1\le i,j\le m}$, $\hat{E}_0 \leftarrow \emptyset$

```
for t ∈ ℕ do
    {i,j} ~ π(t, (n_{t'}, w_{t'})_{0≤t'≤t−1}), w.l.o.g. i < j
    Observe X^{[t]}_{i,j} ~ Ber(q_{i,j})
    Update n_t and w_t according to (4.8)
    Reveal X^{[t]}_{i,j} to A^{i,j}_Coin
    Ê_t ← Ê_{t−1}
    if A^{i,j}_Coin terminated with D(A^{i,j}_Coin) = 0 then        ▷ q_{i,j} > 1/2 w.h.p.
        Ê_t ← Ê_t ∪ {(i,j)}
    else if A^{i,j}_Coin terminated with D(A^{i,j}_Coin) = 1 then   ▷ q_{i,j} < 1/2 w.h.p.
        Ê_t ← Ê_t ∪ {(j,i)}
    if ([m], Ê_t) is acyclic in extension then return WST
    else if ([m], Ê_t) contains a cycle then return ¬WST
```

Theorem 5.8. *Let $\pi \in \Pi_\infty$, $\alpha,\beta \in (0,1)$ and $h \in [0,1/2)$ be fixed, define $\gamma' := \min\{\frac{\alpha}{m}, \beta(\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor)^{-1}\}$ and let $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$ be a solution to $\mathcal{P}^{h,\gamma'}_{\mathrm{Coin}}$. Write $\mathcal{A}$ for Alg. 21 called with parameters m, π and $\mathcal{A}_{\mathrm{Coin}}$. Then, $\mathcal{A}$ solves $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$. In case $\mathcal{A}_{\mathrm{Coin}} = \mathcal{A}^{\mathrm{X}}_{\mathrm{Coin}}[h,\gamma']$ for $\mathrm{X} \in \{\mathrm{Hoeffding}, \mathrm{SPRT}, \mathrm{Farrell}\}$ and $\tilde{\mathcal{A}}$ is Alg. 20 called with parameters m, π and $\mathcal{A}_{\mathrm{Coin}} = \mathcal{A}^{\mathrm{X}}_{\mathrm{Coin}}(h,\tilde{\gamma})$ with $\tilde{\gamma} := \min\{\alpha,\beta\}/\binom{m}{2}$ (as suggested by Thm. 5.6), $T^{\mathcal{A}} \le T^{\tilde{\mathcal{A}}}$ holds almost surely w.r.t. $\mathbb{P}_{\mathbf{Q}}$ for any $\mathbf{Q} \in \mathcal{Q}^0_m$.*

Prop. 3.25 indicates that one cannot expect to choose a correction term smaller than $\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ for the desired type II error within the choice of γ in Alg. 21. Furthermore, the fact that the graph $G \in \mathcal{G}_m$ with edges $1 \to 2 \to \ldots \to m \to 1$ contains a cycle, unlike any of its proper subgraphs, demonstrates optimality of the correction term m for the desired type I error within the choice of γ. As a direct consequence of Thm. 5.8, we obtain in case $h > 0$ a result analogous to the one stated in Cor. 5.7 for Alg. 21 called with m, the sampling strategy π from Cor. 5.7, and $\mathcal{A}_{\mathrm{Coin}} = \mathcal{A}^{\mathrm{SPRT}}_{\mathrm{Coin}}(h,\gamma')$ with

$\gamma' = \min\{\frac{\alpha}{m}, \beta(\binom{m}{2} - \lfloor\frac{m+1}{3}\rfloor)^{-1}\}$, so that it achieves an optimal worst-case runtime (up to a logarithmic term of m) in the active online testing scenario as well.

5.3.2. Instance-wise Upper Bounds and Exploiting Negligibility of Edges

Next, we turn to more sophisticated solutions to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$ in the active setting, that also take into account that those queries $\{i,j\}$, which are with high probability negligible (cf. Sec. 3.6), are superfluous and should be avoided. To this end, we define the sampling strategy π^*_{WST} as the sampling strategy which, similarly to the sampling strategies π_{WST} considered in Cor. 5.7, keeps track of a specific subset of $[m]_2$ consisting of all $\{i,j\}$ for which $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ can be decided with enough confidence (with regard to the internal testing algorithm $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ of Alg. 21) at time t. In contrast to the latter, the subset used by π^*_{WST} takes also the negligibility of edges into account. Formally, π^*_{WST} considers the following set at time t:

$$U^*(t) \coloneqq \big\{\{i,j\} \in [m]_2 \,\big|\, (i,j),(j,i) \notin \hat{E}_{t-1} \text{ and } \{i,j\} \text{ is not negligible for } ([m],\hat{E}_{t-1})\big\}.$$

Now, let π^*_{WST} be defined just like π_{WST} where $U(t)$ is replaced by $U^*(t)$, i.e., π^*_{WST} chooses its t-th query uniformly at random from $U^*(t)$ if $U^*(t) \neq \emptyset$, and uniformly at random from $[m]_2$ if $U^*(t) = \emptyset$.

Note that the set $\hat{E}_{t-1}$ may be defined as the set of all $\{i,j\} \in (m)_2$, for which

$$\exists s \in [t-1] : \mathbf{D}\left(\mathcal{A}^{i,j}_{\mathrm{Coin}}, (\boldsymbol{n}_s)_{i,j}\right) = 0 \text{ and } \forall s' \in [s-1] : \mathbf{D}\left(\mathcal{A}^{i,j}_{\mathrm{Coin}}, (\boldsymbol{n}_{s'})_{i,j}\right) = \text{“N/A”}$$

This condition only depends on the randomness of Alg. 21 up to time $t-1$ (and with this for $1 \leq i < j \leq m$ also on that of $\mathcal{A}^{i,j}_{\mathrm{Coin}}$ up to time $(\boldsymbol{n}_{t-1})_{i,j}$) and thus only on $\boldsymbol{n}_0, \boldsymbol{w}_0, \dots, \boldsymbol{n}_{t-1}, \boldsymbol{w}_{t-1}$ but not on $(\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})$ for $t' \geq t$. Hence, π^*_{WST} is in fact a sampling strategy as stipulated in Sec. 1.1.

From Thm. 5.8, we immediately obtain that Alg. 21 called with parameters m, π^*_{WST} and the solution $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$ to $\mathcal{P}^{h,\gamma'}_{\mathrm{Coin}}$, which π^*_{WST} formally depends on, is a solution to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$. But even if this guarantee holds for any solution $\mathcal{A}_{\mathrm{Coin}}$ to $\mathcal{P}^{h,\gamma'}_{\mathrm{Coin}}$, it is desirable to choose $\mathcal{A}_{\mathrm{Coin}}$ in such a way that the sample complexity of the corresponding algorithm is low. According to Prop. 2.17, Prop. 2.22, and Prop. 2.13, the choices $\mathcal{A}_{\mathrm{Coin}} = \mathcal{A}^{\mathrm{SPRT}}_{\mathrm{Coin}}(h,\gamma')$ resp. $\mathcal{A}_{\mathrm{Coin}} = \mathcal{A}^{\mathrm{Farrell}}_{\mathrm{Coin}}(\gamma')$ are to some extent optimal in this regard for the cases $h > 0$ resp. $h = 0$. With these, we obtain the following instance-wise upper bounds on the expected termination time for solutions to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$. They show that the gaps $\bar{q}_{i,j} = |q_{i,j} - 1/2|$ determine the complexity of testing whether $\mathbf{Q}$ is weakly stochastic transitive or not. In comparison to the lower bound stated in Thm. 5.4, our instance-wise upper bounds depend on all $\binom{m}{2}$ instead of only $\binom{m-1}{2}$ entries of $\mathbf{Q}$. Needless to say, in terms of the asymptotic behavior as $m \to \infty$, this difference is negligible.

Theorem 5.9. *Suppose $m \in \mathbb{N}_{\geq 3}$, $\alpha, \beta \in (0,1/2)$, $h \in [0,1/2)$, let $\mathcal{A}_{\mathrm{Coin}} \in \mathfrak{A}_{\mathrm{Coin}}$ be arbitrary and define $\gamma' \coloneqq \min\{\frac{\alpha}{m}, \beta(\binom{m}{2} - \lfloor\frac{m+1}{3}\rfloor)^{-1}\}$. Write $\mathcal{A}$ for Alg. 21 called with parameters m, the sampling strategy π^*_{WST} and $\mathcal{A}_{\mathrm{Coin}}$ as testing algorithm.*

(i) If $\mathcal{A}_{\mathrm{Coin}}$ solves $\mathcal{P}^{h,\gamma'}_{\mathrm{Coin}}$, then $\mathcal{A}$ solves $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$.

(ii) *If $h > 0$ and $\mathcal{A}_{\text{Coin}} = \mathcal{A}_{\text{Coin}}^{\text{SPRT}}(h,\gamma')$ is the corresponding solution to $\mathcal{P}_{\text{Coin}}^{h,\gamma'}$ from Prop. 2.17, the following holds: Suppose $\mathbf{Q} \in \mathcal{Q}_m^h$ is fixed. Then, with $e(h,\gamma') := \left\lceil \frac{\ln((1-\gamma')/\gamma')}{\ln((1/2+h)/(1/2-h))} \right\rceil$, we have that $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}]$ is bounded from above by*

$$\sum\nolimits_{(i,j)\in(m)_2} \frac{e(h,\gamma')}{2\bar{q}_{i,j}} \left| 1 - 2\left(1 + (1/2+\bar{q}_{i,j})^{e(h,\gamma')}(1/2-\bar{q}_{i,j})^{-e(h,\gamma')}\right)^{-1} \right|. \quad (5.5)$$

(iii) *If $h = 0$ and $\mathcal{A}_{\text{Coin}} = \mathcal{A}_{\text{Coin}}^{\text{Farrell}}(\gamma')$ is the corresponding solution to $\mathcal{P}_{\text{Coin}}^{h,\gamma'}$ from Prop. 2.22, then $\mathcal{A}$ solves $\mathcal{P}_{\text{WST}}^{m,0,\alpha,\beta}$ and there exists $h_0 \in (0,1/2)$ such that*

$$\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \leq \frac{1}{2}\sum\nolimits_{(i,j)\in(m)_2} \frac{1}{\bar{q}_{i,j}^2} \ln\ln\frac{1}{\bar{q}_{i,j}}$$

holds for any $\mathbf{Q} \in \mathcal{Q}_m^0$, for which $\bar{q}_{i,j} \leq h_0$ for all distinct $i,j \in [m]$.

Proof of Thm. 5.9. Throughout this proof, we will write for convenience $T := T^{\mathcal{A}}$.

(i) Let $\mathbf{Q} \in \mathcal{Q}_m^h$ be fixed and suppose at first $\mathcal{A}_{\text{Coin}} \in \mathfrak{A}_{\text{Coin}}$ to be an arbitrary solution to $\mathcal{P}_{\text{Coin}}^{h,\gamma'}$. Then, each $\mathcal{A}_{\text{Coin}}^{i,j}$ terminates a.s. by assumption, and thus $|\hat{E}_t| \to \binom{m}{2}$ holds a.s. as $t \to \infty$. From this we infer $(\boldsymbol{n}_t)_{i,j} \to \infty$ a.s. for any $(i,j) \in (m)_2$, i.e. $\pi \in \Pi_\infty$. Consequently, Thm. 5.8 is applicable and yields that $\mathcal{A}$ solves $\mathcal{P}_{\text{WST}}^{m,h,\alpha,\beta}$.

(ii) Suppose $h > 0$ and $\mathcal{A}_{\text{Coin}} = \mathcal{A}_{\text{Coin}}^{\text{SPRT}}(h,\gamma')$. As $\mathcal{A}_{\text{Coin}}$ solves $\mathcal{P}_{\text{Coin}}^{h,\gamma'}$ by Prop. 2.17, (i) assures that $\mathcal{A}$ solves $\mathcal{P}_{\text{WST}}^{m,h,\alpha,\beta}$. For arbitrary $\mathbf{Q} \in \mathcal{Q}_m^h$, Prop. 2.17 also yields

$$\mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}_{\text{Coin}}^{i,j}}\right] = \frac{e(h,\gamma')}{2\bar{q}_{i,j}} \left| 1 - 2\left(1 + (1/2+\bar{q}_{i,j})^{e(h,\gamma')}(1/2-\bar{q}_{i,j})^{-e(h,\gamma')}\right)^{-1} \right| \quad (5.6)$$

for any $(i,j) \in (m)_2$. A look at Alg. 21 and the choice of π^*_{WST} shows that $\mathcal{A}$ does not query $\{i,j\}$ after $\mathcal{A}_{\text{Coin}}^{i,j}$ has terminated. Hence, we have $\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] \leq \sum_{(i,j)\in(m)_2} \mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}_{\text{Coin}}^{i,j}}\right]$ and combining this with (5.6) completes the proof of (ii).

(iii) If $h = 0$ and $\mathcal{A}_{\text{Coin}} = \mathcal{A}_{\text{Coin}}^{\text{Farrell}}(\gamma')$, Prop. 2.22 assures that $\mathcal{A}_{\text{Coin}}$ solves $\mathcal{P}_{\text{Coin}}^{h,\gamma'}$ and thus $\mathcal{A}$ solves $\mathcal{P}_{\text{WST}}^{m,h,\alpha,\beta}$ according to (i). Moreover, Prop. 2.22 allows us to choose an $h_0 \in (0,1/2)$ such that

$$\frac{\mathbb{E}_{1/2\pm h}\left[T^{\mathcal{A}_{\text{Coin}}}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} \leq \frac{1}{2}$$

holds for any $h \leq h_0$. As each $\mathcal{A}_{\text{Coin}}^{i,j}$ is an instance of $\mathcal{A}_{\text{Coin}}$, we have for any $\mathbf{Q} \in \mathcal{Q}_m^0$ with $\bar{q}_{i,j} \leq h_0$ for all $(i,j) \in (m)_2$ the estimate

$$\mathbb{E}_{q_{i,j}}\left[T^{\mathcal{A}_{\text{Coin}}^{i,j}}\right] \leq \frac{1}{2\bar{q}_{i,j}^2}\ln\ln\frac{1}{\bar{q}_{i,j}}$$

for any $(i,j) \in (m)_2$. Thus, a similar argumentation as in (ii) proves the desired statement.

□

By means of Prop. 2.17, it immediately follows that the algorithm $\mathcal{A}$ from Thm. 5.9(ii) fulfills $\sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \mathcal{O}(\frac{m^2 \ln m}{h^2} \ln \frac{1}{\gamma})$, i.e., it is asymptotically optimal up to a $\ln m$-factor. In order to compare the result of Thm. 5.9 with the instance-wise lower bound from Thm. 5.4 more thoroughly, suppose $\mathbf{Q} \in \mathcal{Q}_m^h(\text{WST})$ and $(i,j) \in (m)_2$ with $|\sigma_{\mathbf{Q}}(i) - \sigma_{\mathbf{Q}}(j)| > 1$ to be fixed for the moment and let $\alpha = \beta = \gamma$ for simplicity. Due to $e(h,\gamma') \in \Theta(\frac{1}{h})$ as $h \searrow 0$, the dependency of (5.5) on the (i,j)-entry of $\mathbf{Q}$ is approximately $\frac{1}{\bar{q}_{i,j} h}$, whereas this dependency in (5.3) is of the form $\frac{1}{\bar{q}_{i,j}^2}$. This suggests, that the two bounds are closest in case $h \approx \bar{q}_{i,j}$. Considering that the choice $\mathcal{A}_{\text{Coin}} = \mathcal{A}_{\text{Coin}}^{\text{SPRT}}(h,\gamma')$ assures optimal early detection of $\text{sign}(q_{i,j} - 1/2)$ only in case $|q_{i,j} - 1/2| = h$, the appearance of $\frac{1}{h}$ in (5.5) may not come as a surprise. Moreover, the scaling $\gamma' \approx \gamma/m^2$ leads to an additional factor of $2 \ln m$ in (5.5) compared to (5.3).

Before giving a detailed proof of Thm. 5.8, let us briefly discuss the computational complexity of the presented WST testing solution required in any time step $t \in \mathbb{N}$. If $\mathcal{A}_{\text{improved}}$ observes the feedback for its t-th query $\{i(t), j(t)\}$, updating $\boldsymbol{n}_t$ and $\boldsymbol{w}_t$ can be done in $\mathcal{O}(1)$ and the cost for updating $\hat{E}_t$ is basically the cost of updating $\mathcal{A}_{\text{Coin}}^{i(t),j(t)}$, which is also $\mathcal{O}(1)$ for all choices of $\mathcal{A}_{\text{Coin}}$ considered in Thm. 5.9. Thus, the overall cost is dominated by (a) the cost for updating $U^*(t)$ and (b) the cost of deciding whether $\hat{E}_t$ is acyclic in extension or contains a cycle.

If $\{i,j\}$ is negligible for $([m], \hat{E}_t)$, then it is either negligible for $([m], \hat{E}_{t-1})$ or $\{i,j\} \cap \{i(t), j(t)\} \neq \emptyset$, and the same holds for non-negligibility. Hence, to calculate $U^*(t)$ from $U^*(t-1)$ one has to check negligibility of the $\mathcal{O}(m)$ many sets $\{\{i,j\} \,|\, j \in \{i(t), j(t)\}\}$, and naively checking negligibility (according to its definition) of any such $\{i,j\}$ can be done in $\mathcal{O}(m)$. Therefore, the costs for (a) are at most $\mathcal{O}(m^2)$. For (b), one only has to test (a)cyclicity in extension of $([m], \hat{E}_t)$ if $\hat{E}_t \neq \hat{E}_{t-1}$. In this case, $\hat{E}_t$ contains exactly those edges as $\hat{E}_{t-1}$ and in addition either $(i(t), j(t))$ or $(j(t), i(t))$, one only has to check whether this additional edge leads to a cycle in $\hat{E}_t$ containing $i(t)$ and $j(t)$, which may e.g. be done by means of a breadth-first search and requires at most $\mathcal{O}(m^2)$ operations. Moreover, Prop. 3.19 assures that acyclicity in extension of $([m], \hat{E}_t)$ can be checked with knowledge of $U^*(t)$ in $\mathcal{O}(m^2)$. Thus, (b) can be done in $\mathcal{O}(m^2)$ and the overall computational complexity for time step t is $\mathcal{O}(m^2)$.

5.3.3. Proof of Thm. 5.8

Proof of Thm. 5.8. Similarly as in the proof of Thm. 5.6, we abbreviate $T \coloneqq T^{\mathcal{A}}$ and write $\{i(t), j(t)\}$, $i(t) < j(t)$, for the query sampled from π at the t-th iteration of the loop. The testing algorithm $\mathcal{A}_{\text{Coin}}^{i,j}$ observes the sample $X_{i,j}^{[t]} \sim \text{Ber}(q_{i,j})$ iff $(i(t), j(t)) = (i,j)$, and after time t it has observed exactly $(\boldsymbol{n}_t)_{i,j}$ of these samples. For the sake of convenience, we write $\overline{G}$ for the set $\{\{i,j\} \,|\, (i,j) \in E_G \text{ or } (j,i) \in E_G\}$ for $G \in \mathcal{G}_m$. We denote by $\hat{E}_t$ the internal statistic of Alg. 21 *at the end* of the t-th iteration. Moreover, we abbreviate $\hat{G}_T \coloneqq ([m], \hat{E}_T)$. Prop. 3.22 allows us to fix a function $\mathfrak{l}_{\text{acyclic}} : \mathcal{G}_m \to \mathcal{G}_m$ such that for all $G \in \mathcal{G}_m$ we have $E_{\mathfrak{l}_{\text{acyclic}}(G)} \subseteq E_G$, $|E_{\mathfrak{l}_{\text{acyclic}}(G)}| \leq \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ as well as

$$G \in \mathcal{G}_m(\text{acyclic}) \quad \Leftrightarrow \quad \mathfrak{l}_{\text{acyclic}}(G) \in \mathcal{G}_m(\text{acyclic}). \tag{5.7}$$

Recall $\mathcal{G}_m(\neg\text{acyclic}) = \{G \in \mathcal{G}_m \,|\, G \text{ contains a cycle}\}$ and that $\mathbf{Q} \in \mathcal{Q}_m(\neg\text{WST})$ iff $G(\mathbf{Q})$ contains a cycle. As every cycle in any graph on $[m]$ contains at most m edges, it is straightforward to define a mapping $\iota_{\neg\text{acyclic}} : \mathcal{G}_m \to \mathcal{G}_m$ such that $E_{\iota_{\neg\text{acyclic}}(G)} \subseteq E_G$, $|E_{\iota_{\neg\text{acyclic}}(G)}| \leq m$ and

$$G \in \mathcal{G}_m(\neg\text{acyclic}) \quad \Leftrightarrow \quad \iota_{\neg\text{acyclic}}(G) \in \mathcal{G}_m(\neg\text{acyclic})$$

hold for all $G \in \mathcal{G}_m$. For $\mathbf{Q} \in \mathcal{Q}_m$, $E \subseteq [m] \times [m]$ and $\{i,j\} \in [m]_2$ we say that $\{i,j\}$ *is assigned incorrectly (ass. inc.) in* E *w.r.t.* $\mathbf{Q}$ if

$$(i,j) \in E \text{ and } q_{i,j} < 1/2 \quad \text{or} \quad (j,i) \in E \text{ and } q_{i,j} > 1/2$$

holds, where we may omit the term "w.r.t. $\mathbf{Q}$" in case $\mathbf{Q}$ is clear from the context. We split the proof into five parts.

Part 1: $\mathcal{A}$ terminates a.s. for any $\mathbf{Q} \in \mathcal{Q}_m^h$
As in the proof of Thm. 4.6, we may suppose for the proof of Part 1 w.l.o.g. that $\mathcal{A}$ does not terminate before all of the internal testing algorithms $\mathcal{A}_{\text{Coin}}^{i,j}$ have terminated. By the assumption $\pi \in \Pi_\infty$, we have $(\boldsymbol{n}_t)_{i,j} \to \infty$ almost surely as $t \to \infty$ for each $(i,j) \in (m)_2$. For any $(i,j) \in (m)_2$, $\mathcal{A}_{\text{Coin}}^{i,j}$ solves by assumption $\mathcal{P}_{\text{Coin}}^{h,\gamma'}$ and hence the stopping time

$$T_{i,j} = \min\left\{t \in \mathbb{N} \,:\, \mathbf{D}\left(\mathcal{A}_{\text{Coin}}^{i,j}, (\boldsymbol{n}_t)_{i,j}\right) \neq \text{“N/A”}\right\} \tag{5.8}$$

is a.s. finite. In particular,

$$T' := \max_{(i,j)\in(m)_2} T_{i,j}$$

is a.s. finite. Regarding the definitions of $\mathcal{A}$ and T' we see that $\hat{G}_{T'}$ is almost surely an element of $\overline{\mathcal{G}}_m = \overline{\mathcal{G}}_m(\text{acyclic}) \cup \overline{\mathcal{G}}_m(\neg\text{acyclic})$. Consequently, we obtain

$$T = \min\left\{t \in \mathbb{N} \,:\, \hat{G}_t \in \mathcal{G}_m(\text{acyclic}) \text{ or } \hat{G}_t \in \mathcal{G}_m(\neg\text{acyclic})\right\} \leq T' < \infty \quad \text{a.s.},$$

which completes the proof of Part 1. ■

Part 2: Showing $\mathbb{P}_{\mathbf{Q}}(\{i,j\}$ is ass. inc. in $\hat{E}_T) \leq \gamma'$ in case $|q_{i,j} - 1/2| > h$
A look at lines 6–10 of Alg. 21 reveals $\hat{E}_{t-1} \subseteq \hat{E}_t$ for all $t \leq T$ and moreover

$$(i,j) \in \hat{E}_T \quad \Leftrightarrow \quad T_{i,j} \leq T \text{ and } \mathbf{D}\left(\mathcal{A}_{\text{Coin}}^{i,j}\right) = \mathbf{D}\left(\mathcal{A}_{\text{Coin}}^{i,j}, T\right) = 1.$$

As $\mathcal{A}_{\text{Coin}}^{i,j}$ solves $\mathcal{P}_{\text{Coin}}^{h,\gamma'}$, we infer in case $q_{i,j} < 1/2 - h$ that

$$\begin{aligned}\mathbb{P}_{\mathbf{Q}}\left(\{i,j\} \text{ is assigned incorrectly in } \hat{E}_T\right) &= \mathbb{P}_{\mathbf{Q}}((i,j) \in \hat{E}_T) \\ &\leq \mathbb{P}_{\mathbf{Q}}\left(\mathbf{D}\left(\mathcal{A}_{\text{Coin}}^{i,j}\right) = 0\right) \leq \gamma',\end{aligned}$$

and in case $q_{j,i} < 1/2 - h$ we similarly obtain

$$\begin{aligned}\mathbb{P}_{\mathbf{Q}}(\{i,j\} \text{ is assigned incorrectly in } \hat{E}_T) &= \mathbb{P}_{\mathbf{Q}}((j,i) \in \hat{E}_T) \\ &\leq \mathbb{P}_{\mathbf{Q}}\left(\mathbf{D}\left(\mathcal{A}_{\text{Coin}}^{i,j}\right) = 1\right) \leq \gamma'.\end{aligned}$$

This shows the assertion of Part 2. ■

Part 3: Bounding the type I error

Let us first consider the case $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{WST})$. According to the choice of $\mathfrak{l}_{\neg\mathrm{WST}}$ we have

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq \mathrm{WST}) = \mathbb{P}_{\mathbf{Q}}(\hat{G}_T \in \mathcal{G}_m(\neg\mathrm{acyclic})) = \mathbb{P}_{\mathbf{Q}}(\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\neg\mathrm{acyclic})). \tag{5.9}$$

By Part 2 we have

$$\mathbb{P}_{\mathbf{Q}}\left(\{i,j\} \text{ is ass. inc. in } E_{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} \;\middle|\; \overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right) \leq \begin{cases} 0, & \text{if } \{i,j\} \notin \overline{G}, \\ \gamma', & \text{if } \{i,j\} \in \overline{G}. \end{cases}$$

If no $\{i,j\} \in \overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)}$ was assigned incorrectly in $E_{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)}$, then $\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\neg\mathrm{acyclic})$ would imply $\mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST})$. Consequently, $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{WST})$ lets us infer that $\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\neg\mathrm{acyclic})$ is only possible if there exists some $\{i,j\} \in \overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)}$, which is assigned incorrectly in $E_{\neg\mathfrak{l}_{\mathrm{acyclic}}(\hat{G}_T)}$. Regarding that $|\overline{G}| = |E_G|$, we thus get

$$\begin{aligned}
&\mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\neg\mathrm{acyclic}) \text{ and } \overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \mathbb{P}_{\mathbf{Q}}\left(\exists\{i,j\} \in \overline{G}, \text{ which is ass. inc. in } E_{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} \text{ and } \overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \sum\nolimits_{\{i,j\}\in\overline{G}} \mathbb{P}_{\mathbf{Q}}\left(\{i,j\} \text{ is ass. inc. in } E_{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} \text{ and } \overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \gamma' |E_G| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right)
\end{aligned}$$

for every $G \in \mathcal{G}_m$. Together with (5.9) and $\mathbb{P}_{\mathbf{Q}}(|E_{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)}| \leq m) = 1$, which holds according to the choice of $\mathfrak{l}_{\neg\mathrm{acyclic}}$, we infer

$$\begin{aligned}
&\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq \mathrm{WST}) = \mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\neg\mathrm{acyclic})\right)\\
&= \sum\nolimits_{\overline{G}:G\in\mathcal{G}_m} \mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\neg\mathrm{acyclic}) \text{ and } \overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \sum\nolimits_{\overline{G}:G\in\mathcal{G}_m} \mathbb{P}_{\mathbf{Q}}\left(\exists\{i,j\} \in \overline{G}, \text{ which is ass. inc. in } E_{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} \text{ and } \overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right)\\
&\leq \gamma' \sum\nolimits_{\overline{G}:G\in\mathcal{G}_m} |E_G| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right) \leq \gamma' m \leq \alpha,
\end{aligned}$$

where we have used that $\sum_{\overline{G}:G\in\mathcal{G}_m} \mathbb{P}_{\mathbf{Q}}(\overline{\mathfrak{l}_{\neg\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}) = 1$ holds trivially. ■

Part 4: Bounding the type II error
Now, we consider the case $\mathbf{Q} \in \mathcal{Q}_m^h(\neg\mathrm{WST})$. Similarly as above in Part 3, the choice of $\mathfrak{l}_{\mathrm{acyclic}}$ yields

$$\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq \neg\mathrm{WST}) = \mathbb{P}_{\mathbf{Q}}(\hat{G}_T \in \mathcal{G}_m(\mathrm{acyclic})) = \mathbb{P}_{\mathbf{Q}}(\mathfrak{l}_{\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\mathrm{acyclic})). \tag{5.10}$$

An analogue argumentation as above shows that $\mathfrak{l}_{\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\mathrm{acyclic})$ is only possible if there exists some $\{i,j\} \in \overline{\mathfrak{l}_{\mathrm{acyclic}}(\hat{G}_T)}$, which is assigned incorrectly in $E_{\mathfrak{l}_{\mathrm{acyclic}}(\hat{G}_T)}$. From this and Part 2 we can infer that

$$\mathbb{P}_{\mathbf{Q}}\left(\mathfrak{l}_{\mathrm{acyclic}}(\hat{G}_T) \in \mathcal{G}_m(\mathrm{acyclic}) \text{ and } \overline{\mathfrak{l}_{\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right) \leq \gamma' |E_G| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\mathrm{acyclic}}(\hat{G}_T)} = \overline{G}\right) \tag{5.11}$$

is fulfilled for every $G \in \mathcal{G}_m$. According to the choice of $\mathfrak{l}_{\text{acyclic}}$ we have $\mathbb{P}_{\mathbf{Q}}(|E_{\mathfrak{l}_{\text{acyclic}}(\hat{G}_T)}| \leq \binom{m}{2} - \lfloor (m+1)/3 \rfloor) = 1$, so that combining this with (5.10) and (5.11) yields

$$\begin{aligned}\mathbb{P}_{\mathbf{Q}}(\mathbf{D}(\mathcal{A}) \neq \neg\text{WST}) &\leq \gamma' \sum\nolimits_{\overline{G}:G\in\mathcal{G}_m} |E_G| \mathbb{P}_{\mathbf{Q}}\left(\overline{\mathfrak{l}_{\text{acyclic}}(\hat{G}_T)} = \overline{G}\right) \\ &\leq \gamma' \left(\binom{m}{2} - \left\lfloor \frac{m+1}{3} \right\rfloor \right) \leq \beta.\end{aligned}$$

This completes the proof of Part 4. ■

Part 5: Comparing Alg. 21 to Alg. 20
Now, suppose $\mathrm{X} \in \{\text{Hoeffding}, \text{SPRT}, \text{Farrell}\}$ and let $\mathcal{A} =: \mathcal{A}_1$ and $\tilde{A} =: \mathcal{A}_2$ as specified in the statement of this theorem above, i.e., $\mathcal{A}_1$ is called with $\mathcal{A}^{\mathrm{X}}_{\text{Coin}}[h,\gamma']$ whereas $\mathcal{A}_2$ is called with $\mathcal{A}^{\mathrm{X}}_{\text{Coin}}[h,\tilde{\gamma}]$ as input. Further, suppose $\mathbf{Q} \in \mathcal{Q}^0_m$ to be fixed and let $T^{(l)}_{i,j}$ for $l \in \{1,2\}$ be defined as in (5.8) by using the statistics $\boldsymbol{n}^{\mathcal{A}_1}_t = ((\boldsymbol{n}^{\mathcal{A}_1}_t)_{i,j})_{1\leq i,j\leq m}$ resp. $\boldsymbol{n}^{\mathcal{A}_2}_t = ((\boldsymbol{n}^{\mathcal{A}_2}_t)_{i,j})_{1\leq i,j\leq m}$ of $\mathcal{A}_1$ resp. $\mathcal{A}_2$. According to Ex. 2.20 and Prop. 2.22, $\mathcal{A}^{\mathrm{X}}_{\text{Coin}}[h,\gamma']$ resp. $\mathcal{A}^{\mathrm{X}}_{\text{Coin}}[h,\tilde{\gamma}]$ are symmetric GSPRTs with some barriers B' resp. $\tilde{B}$, that fulfill $B' \leq \tilde{B}$ due to $\gamma' > \tilde{\gamma}$. Consequently, $\mathcal{A}^{\mathrm{X}}_{\text{Coin}}(h,\gamma')$ terminates a.s. not later than $\mathcal{A}^{\mathrm{X}}_{\text{Coin}}(h,\tilde{\gamma})$, which shows $T^{(1)}_{i,j} \leq T^{(2)}_{i,j}$. As $(i,j) \in (m)_2$ was arbitrary, we obtain by construction of $\mathcal{A}_1$ and $\mathcal{A}_2$ that

$$T^{\mathcal{A}_1} \leq \max\nolimits_{(i,j)\in(m)_2} T^{(1)}_{i,j} \leq \max\nolimits_{(i,j)\in(m)_2} T^{(2)}_{i,j} \leq T^{\mathcal{A}_2} \quad \text{a.s. w.r.t. } \mathbb{P}_{\mathbf{Q}}.$$

□

5.4. Approach II: A Likelihood-ratio Based Approach

In this section, we approach WST testing from another direction. For the sake of convenience, we assume $(\boldsymbol{n}_0)_{i,j} = 1$ for each distinct $i,j \in [m]$, i.e., we suppose that each of the queries $\{i,j\} \in [m]_2$ has already been made exactly once at time $t = 1$. Moreover, if not explicitly stated differently, we suppose throughout this section $\pi \in \Pi_\infty$. Recall that we are interested in sequentially testing the hypotheses

$$\mathbf{H}^{\text{WST}}_0 : \mathbf{Q} \in \mathcal{Q}_m(\text{WST}) \qquad \mathbf{H}^{\text{WST}}_1 : \mathbf{Q} \in \mathcal{Q}_m(\neg\text{WST}). \tag{5.12}$$

Here, to assure a type I error bound for our testing algorithm, we follow a likelihood-ratio test (LRT) approach for (5.12), while we control the type II error by following an LRT approach for the corresponding interchanged hypotheses

$$\widetilde{\mathbf{H}}^{\text{WST}}_0 : \mathbf{Q} \in \mathcal{Q}_m(\neg\text{WST}) \qquad \widetilde{\mathbf{H}}^{\text{WST}}_1 : \mathbf{Q} \in \mathcal{Q}_m(\text{WST}). \tag{5.13}$$

In combination, this will allow us to construct a solution to $\mathcal{P}^{m,0,\alpha,\beta}_{\text{WST}}$. Our approach is an extension of the LRT for WST by Iverson and Falmagne [1985] to the online testing setting, for which finite sample properties instead of asymptotic ones are necessary. Again, we will treat both the passive and the active testing scenario. In the passive one, we restrict ourselves to sampling strategies $\pi \in \Pi_\infty$ that fulfill further constraints as specified in the following.

Definition 5.10. *Recall the definition of a sampling strategy from Sec. 1.1 and write* $\Pi_\infty^{\ln\ln}$ *for the set of sampling strategies which fulfill* $\lim_{t\to\infty} \frac{(\boldsymbol{n}_t)_{i,j}}{\ln\ln t} = \infty$ *a.s. as* $t \to \infty$ *for any* $(i,j) \in (m)_2$.

If $\pi \in \Pi \setminus \Pi_\infty$, a sampling strategy $\hat{\pi} \in \Pi$ that chooses $\pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0\le t'\le t-1}) = \hat{\pi}(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0\le t'\le t-1})$ with probability $1 - \frac{1}{\sqrt{t}}$, and otherwise samples uniformly at random from $[m]_2$ with probability $\frac{1}{\sqrt{t}}$, fulfills $\hat{\pi} \in \Pi_\infty^{\ln\ln}$ (Lem. 2.8) and

$$\mathbb{P}(\pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0\le t'\le t-1}) \neq \hat{\pi}(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0\le t'\le t-1})) \le \frac{1}{\sqrt{t}} \to 0$$

as $t \to \infty$. Thus, $\hat{\pi}$ and π behave similarly in the limit. This shows that the assumption $\pi \in \Pi_\infty^{\ln\ln}$, which is required for theoretical results in our framework, is rather mild.

We start this section with the introduction of the LRT statistics λ_t and μ_t for (5.12) and (5.13) as well as conveniently modified statistics $\tilde{\lambda}_t$ and $\tilde{\mu}_t$. These statistics allow us to construct passive and active solutions to $\mathcal{P}_{\mathrm{WST}}^{m,0,\alpha,\beta}$. We provide a theoretical analysis of the used statistics and also sophisticated update formulas for the calcuation of these. Afterwards, we turn to the asymptotic behaviour of λ_t and μ_t and formulate asymptotic size-α tests for testing WST resp. ¬WST. These latter results extend upon the work by Iverson and Falmagne [1985].

5.4.1. The Likelihood-ratio Test Statistics

Let us suppose for the moment to be in the passive scenario, where the query $\{i, j\}$ made at time t is chosen according to an arbitrary but fixed sampling strategy $\pi \in \Pi_\infty$, and we want to test WST of $\mathbf{Q}$ based on the statistics $(\boldsymbol{n}_t)_{i,j}$, $(\boldsymbol{w}_t)_{i,j}$ for $(i,j) \in (m)_2$, $t \in \mathbb{N}$. Recall that $(\boldsymbol{w}_t)_{i,j} \sim \mathrm{Bin}((\boldsymbol{n}_t)_{i,j}, q_{i,j})$ holds for any time step t. Hence, a naive approach for testing (5.12) with a (sequential) LRT would be based on the binomial distribution of $\boldsymbol{w}_t$, leading to the test statistic

$$\lambda_t := -2\ln\left(\frac{\sup_{\mathbf{Q}\in\mathcal{Q}_m(\mathrm{WST})} \mathcal{L}(\mathbf{Q}|\boldsymbol{w}_t,\boldsymbol{n}_t)}{\sup_{\mathbf{Q}\in\mathcal{Q}_m} \mathcal{L}(\mathbf{Q}|\boldsymbol{w}_t,\boldsymbol{n}_t)}\right),$$

where

$$\mathcal{L}(\mathbf{Q}|\boldsymbol{w}_t, \boldsymbol{n}_t) = \prod_{i<j} \mathbb{P}\big(\mathrm{Bin}((\boldsymbol{n}_t)_{i,j}, q_{i,j}) = (\boldsymbol{w}_t)_{i,j}\big)$$

is the likelihood for $\mathbf{Q}$ given the observed outcomes represented by $\boldsymbol{w}_t$ and $\boldsymbol{n}_t$. However, the specific form of the log-likelihood is analytically unwieldy and thus makes the finite sample analysis cumbersome. A remedy can be found by using a suitable transformation of the parameter space and an accompanying normal approximation of the underlying binomial distribution. For this, let us define the monotonically increasing transformation

$$\phi : [0,1] \to [-\pi/2, \pi/2],\ \phi(x) := 2\arcsin(\sqrt{x}) - \pi/2.$$

From the equivalence

$$(q_{i,j}, q_{j,k} \le 1/2 \ \Rightarrow\ q_{i,k} \le 1/2) \quad \Leftrightarrow \quad (\ q_{k,j}, q_{j,i} \ge 1/2 \ \Rightarrow\ q_{k,i} \ge 1/2),$$

which holds for every distinct $i, j, k \in [m]$, we can directly infer that any $\mathbf{Q} \in \mathcal{Q}_m$ is WST iff

$$q_{i,j}, q_{j,k} \leq 1/2 \;\;\Rightarrow\;\; q_{i,k} \leq 1/2$$

is fulfilled for every distinct $i, j, k \in [m]$. Consequently, writing $\theta_{i,j} := \phi(q_{i,j})$ as well as $\boldsymbol{\theta} = \phi(\mathbf{Q}) = (\theta_{i,j})_{1 \leq i,j \leq m}$, it is easy to see that some $\mathbf{Q} \in \mathcal{Q}_m$ is WST iff

$$\theta_{i,j}, \theta_{j,k} \leq 0 \;\;\Rightarrow\;\; \theta_{i,k} \leq 0$$

for any distinct $i, j, k \in [m]$. Thus, instead of (5.12), we may equivalently consider testing

$$\mathbf{H}_0 : \boldsymbol{\theta} \in \Theta_m(\mathrm{WST}) \qquad \mathbf{H}_1 : \boldsymbol{\theta} \in \Theta_m \setminus \Theta_m(\mathrm{WST}), \tag{5.14}$$

with $\Theta_m := \phi(\mathcal{Q}_m) = [-\pi/2, \pi/2]^{\binom{m}{2}}$ and $\Theta_m(\mathrm{WST}) := \phi(\overline{\mathcal{Q}_m(\mathrm{WST})}) = \overline{\phi(\mathcal{Q}_m(\mathrm{WST}))}$ being the closure of $\phi(\mathcal{Q}_m(\mathrm{WST}))$ in Θ_m; here and throughout, we regard $\Theta_m \subseteq \mathbb{R}^{m(m-1)/2}$ to be equipped with the corresponding topology, which is induced by the standard topology on $\mathbb{R}^{m(m-1)/2}$, i.e., $U' \subseteq \Theta_m$ is open iff $U' = \Theta_m \cap U$ for an open $U \subseteq \mathbb{R}^{m(m-1)/2}$. In particular, the boundary $\partial\Theta_m(\mathrm{WST})$ of $\Theta_m(\mathrm{WST})$ is a subset of $\{\boldsymbol{\theta} \in : \exists (i,j) \in (m)_2 : \theta_{i,j} = 0\}$ and e.g. $(\pi/2)_{1 \leq i < j \leq m} \in \Theta_m(\mathrm{WST}) \setminus \partial\Theta_m(\mathrm{WST})$ holds. Due to $\phi(\mathcal{Q}_m(\mathrm{WST})) \subseteq \Theta_m(\mathrm{WST})$, the type I error of any test for (5.12) is at most as large as its type I error for (5.14). Note that the choice of $\Theta_m(\mathrm{WST})$ instead of $\phi(\mathcal{Q}_m(\mathrm{WST}))$ as the null hypothesis ensures the existence of the maximum likelihood estimator (MLE) on $\Theta_m(\mathrm{WST})$, as $\Theta_m(\mathrm{WST})$ is compact.

Let us define $\boldsymbol{z}_t \in [-\pi/2, \pi/2]^{\binom{m}{2}}$ via $(\boldsymbol{z}_t)_{i,j} := \phi\big((\boldsymbol{w}_t)_{i,j}/(\boldsymbol{n}_t)_{i,j}\big)$ for every $(i,j) \in (m)_2$. Whenever $q_{i,j} \in (0,1)$, the *delta method* [Van der Vaart, 2000] yields[4] $\sqrt{(\boldsymbol{n}_t)_{i,j}}((\boldsymbol{z}_t)_{i,j} - \theta_{i,j}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ as $t \to \infty$, which means that $(\boldsymbol{z}_t)_{i,j}$ is approximately distributed as $\mathcal{N}(\theta_{i,j}, 1/(\boldsymbol{n}_t)_{i,j})$ for sufficiently large t (cf. Lem. 5.12 below). This motivates the usage of

$$\tilde{\lambda}_t := -2 \ln \left(\frac{\max_{\boldsymbol{\theta} \in \Theta_m(\mathrm{WST})} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\max_{\boldsymbol{\theta} \in \Theta_m} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)} \right) \tag{5.15}$$

as the LRT statistic, where

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t) := \prod\nolimits_{i<j} \frac{\sqrt{(\boldsymbol{n}_t)_{i,j}}}{\sqrt{2\pi}} \exp\Big(- \frac{((\boldsymbol{z}_t)_{i,j} - \theta_{i,j})^2}{2(\boldsymbol{n}_t)_{i,j}^{-1}} \Big).$$

Due to

$$\sup\nolimits_{\boldsymbol{\theta} \in \Theta_m} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t) = \tilde{\mathcal{L}}(\boldsymbol{z}_t|\boldsymbol{z}_t) = \prod\nolimits_{i<j} \sqrt{(\boldsymbol{n}_t)_{i,j}}/\sqrt{2\pi}$$

we obtain the identity

$$\begin{aligned} \tilde{\lambda}_t &= -2 \ln \Big(\sqrt{2\pi} \tilde{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{z}_t) \Big/ \prod\nolimits_{i<j} \sqrt{(\boldsymbol{n}_t)_{i,j}} \Big) \\ &= -2 \ln \Big(\prod\nolimits_{i<j} \exp(-((\boldsymbol{z}_t)_{i,j} - \hat{\theta}_{i,j})^2 (\boldsymbol{n}_t)_{i,j}) \Big) \\ &= \sum\nolimits_{i<j} (\boldsymbol{n}_t)_{i,j} ((\boldsymbol{z}_t)_{i,j} - \hat{\theta}_{i,j})^2 =: \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \hat{\boldsymbol{\theta}}), \end{aligned}$$

[4] This convergence result serves only as motivation for the choice of $(\boldsymbol{z}_t)_{i,j}$, it will neither be an argument in the construction of our passive and active solutions to WST testing nor for the tail bounds on μ_t and ν_t that we will achieve below. Nevertheless, let us note that a *uniform* delta method could be used to show that $\sqrt{(\boldsymbol{n}_t)_{i,j}}((\boldsymbol{z}_t)_{i,j} - \theta_{i,j}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ holds even (in an appropriate sense) uniformly for all $\boldsymbol{\theta}$ in a compact subset $\Theta'_m \subseteq \Theta_m$.

with $\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}\in\Theta_m(\mathrm{WST})}\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)$ as the MLE of $\boldsymbol{\theta}$ in $\Theta_m(\mathrm{WST})$ and $\mathrm{d}_{\boldsymbol{n}_t}$ the weighted Euclidean distance in $\Theta_m \subseteq \mathbb{R}^{m(m-1)/2}$ with weights $\boldsymbol{n}_t$. By definition of $\tilde{\mathcal{L}}$, it holds that $\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t) \leq \tilde{\mathcal{L}}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}_t)$ if and only if $\mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \boldsymbol{\theta}) \geq \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \tilde{\boldsymbol{\theta}})$, so that we get the representation

$$\tilde{\lambda}_t = \min_{\boldsymbol{\theta}\in\Theta_m(\mathrm{WST})} \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \boldsymbol{\theta})$$

and $\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}\in\Theta_m(\mathrm{WST})}\mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \boldsymbol{\theta})$. In other words, $\hat{\boldsymbol{\theta}}$ is the point in $\Theta_m(\mathrm{WST})$ closest to $\boldsymbol{z}_t$ with respect to the weighted Euclidean distance $\mathrm{d}_{\boldsymbol{n}_t}$, and $\tilde{\lambda}_t = \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \hat{\boldsymbol{\theta}})$.

The classical sequential probability ratio test approach [Ghosh and Sen, 1991] determines a stopping rule (a termination criterion) on the test statistic by deriving a lower resp. an upper bound triggering a stop of the test procedure in case of a shortfall resp. an exceedance. In case $\pi \in \Pi_\infty$, $(\boldsymbol{z}_t)_{i,j}$ converges due to the law of large numbers almost surely to $\theta_{i,j}$ as $t \to \infty$ and the asymptotical behavior of $\tilde{\lambda}_t$ is

$$\lim_{t\to\infty} \tilde{\lambda}_t = \begin{cases} 0, & \text{for } \mathbf{Q} \in \mathcal{Q}_m^0(\mathrm{WST}), \\ \infty, & \text{for } \mathbf{Q} \in \mathcal{Q}_m^0(\neg\mathrm{WST}). \end{cases} \tag{5.16}$$

For convenience, we defer the theoretical justification of this to Prop. 5.11 below.

This demonstrates that large values of $\tilde{\lambda}_t$ are an indicator for violations of WST. Since $\boldsymbol{z}_t \in \Theta_m(\mathrm{WST})$ implies $\tilde{\lambda}_t = 0$, it might happen by chance that the former is fulfilled for the current data at time step t, although $\mathbf{Q}$ is not WST. Therefore, the classical sequential probability ratio test approach with a lower bound for the test statistic $\tilde{\lambda}_t$ does not appear suitable for the detection of WST in this framework, as the type II error cannot be bounded appropriately.

To circumvent this disadvantage, we also consider the test in (5.13) and incorporate the considerations as before to test

$$\tilde{\mathbf{H}}_0 : \boldsymbol{\theta} \in \Theta_m(\neg\mathrm{WST}) \qquad \tilde{\mathbf{H}}_1 : \boldsymbol{\theta} \in \Theta_m \setminus \Theta_m(\neg\mathrm{WST}) \tag{5.17}$$

with $\Theta_m(\neg\mathrm{WST}) := \overline{\phi(\mathcal{Q}_m(\neg\mathrm{WST}))} = \overline{\Theta_m \setminus \Theta_m(\mathrm{WST})}$. Note here that $\Theta_m(\neg\mathrm{WST}) \neq \Theta_m \setminus \Theta_m(\mathrm{WST})$ but instead $\Theta_m(\mathrm{WST}) \cap \Theta_m(\neg\mathrm{WST}) = \partial\Theta_m(\mathrm{WST}) = \partial\Theta_m(\neg\mathrm{WST})$ holds. Similarly as above, we consider

$$\tilde{\mu}_t := -2\ln\left(\frac{\max_{\boldsymbol{\theta}\in\Theta_m(\neg\mathrm{WST})} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\max_{\boldsymbol{\theta}\in\Theta_m} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right), \tag{5.18}$$

with $\tilde{\mathcal{L}}$ as above, as LRT statistic and note that $\tilde{\mu}_t = \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \hat{\boldsymbol{\theta}})$ with $\hat{\boldsymbol{\theta}}$ being the point in $\Theta_m(\neg\mathrm{WST})$ that is closest to $\boldsymbol{z}_t$ with respect to $\mathrm{d}_{\boldsymbol{n}_t}$. Furthermore, an asymptotic property analogous to (5.16) holds for $\tilde{\mu}_t$ as well by interchanging the conditions, cf. Prop. 5.11 below. Therefore, large values of $\tilde{\mu}_t$ seem to indicate that $\mathbf{Q}$ is WST. To adequately state the following proposition, let us write for convenience

$$\Theta_m^0 := \phi(\mathcal{Q}_m^0) = \{\boldsymbol{\theta} \in \Theta_m \,|\, \forall\,(i,j) \in (m)_2 : \theta_{i,j} \neq 0\} \tag{5.19}$$

and abbreviate $\Theta_m^0(\mathrm{WST}) := \Theta_m^0 \cap \Theta_m(\mathrm{WST})$ and $\Theta_m^0(\neg\mathrm{WST}) := \Theta_m^0 \cap \Theta_m(\neg\mathrm{WST})$.

Proposition 5.11. *(i) If $\boldsymbol{\theta} \in \Theta_m^0(\mathrm{WST})$, then $\tilde{\lambda}_t \to 0$ and $\tilde{\mu}_t \to \infty$ almost surely as $t \to \infty$. Provided $\pi \in \Pi_\infty^{\ln\ln}$, we even have $\tilde{\mu}_t/\ln\ln t \to \infty$ almost surely as $t \to \infty$.*

(ii) Similarly, if $\boldsymbol{\theta} \in \Theta_m^0(\neg\mathrm{WST})$, then $\tilde{\mu}_t \to 0$ and $\tilde{\lambda}_t \to \infty$ almost surely as $t \to \infty$. In case $\pi \in \Pi_\infty^{\ln\ln}$ we have $\tilde{\lambda}_t/\ln\ln t \to \infty$ almost surely as $t \to \infty$.

Proof of Prop. 5.11. We only show (i), as (ii) follows by an analogous argumentation. Let

$$\varepsilon \coloneqq \min_{1\le i<j\le m} |q_{i,j} - 1/2| > 0$$

and write $\hat{\mathbf{Q}}(t) = (\hat{q}_{i,j}(t))_{1\le i,j\le m} = \left(\frac{(\boldsymbol{w}_t)_{i,j}}{(\boldsymbol{n}_t)_{i,j}}\right)_{1\le i,j\le m}$. The law of large numbers guarantees $\hat{q}_{i,j}(t) \to q_{i,j}$ a.s. as $t \to \infty$ for every distinct $i,j \in [m]$, in particular there exists an a.s. finite stopping time T_0 with $|\hat{q}_{i,j}(t) - q_{i,j}| \le \varepsilon/\left(2\sqrt{\binom{m}{2}}\right)$ for every $t \ge T_0$. In, other words, $\hat{\mathbf{Q}}(t) \in B_{\frac{\varepsilon}{2}}(\mathbf{Q}) \subseteq \mathcal{Q}_m(\mathrm{WST})$ and thus $\boldsymbol{z}_t \in \Theta_m^0(\mathrm{WST})$ holds for every $t \ge T_0$, wherein $B_{\frac{\varepsilon}{2}}(\mathbf{Q})$ denotes the ball with radius $\frac{\varepsilon}{2}$ in $\mathbb{R}^{\binom{m}{2}}$ equipped with the Euclidean distance. Consequently, $\tilde{\lambda}_t = \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \Theta_m(\mathrm{WST})) = 0$ holds for every $t \ge T_0$ and the first statement in (i) already follows. Note that

$$|\hat{q}_{i,j}(t) - 1/2| \ge |q_{i,j} - 1/2| - |\hat{q}_{i,j}(t) - q_{i,j}| \ge \varepsilon - \frac{\varepsilon}{2\sqrt{\binom{m}{2}}} \ge \frac{\varepsilon}{2}$$

holds for every $t \ge T_0$. Since $\phi(p) = -\phi(1-p)$ holds for every $p \in [0,1]$ and ϕ is strictly monotonically increasing on $[1/2, 1]$, we thus obtain the estimate

$$(\boldsymbol{z}_t)_{i,j}^2 = \phi^2\left(\hat{q}_{i,j}(t)\right) \ge \phi^2\left(\frac{1}{2} + \frac{\varepsilon}{2}\right) > 0$$

for every $t \ge T_0$ and every distinct $i,j \in [m]$. Together with $\boldsymbol{z}_t \notin \Theta_m(\neg\mathrm{WST})$ for $t \ge T_0$ and regarding the geometry of $\partial\Theta_m(\neg\mathrm{WST})$ we conclude that

$$\begin{aligned}
\lim_{t\to\infty,\, t\ge T_0} \tilde{\mu}_t &= \lim_{t\to\infty,\, t\ge T_0} \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \Theta_m(\neg\mathrm{WST})) \\
&\ge \lim_{t\to\infty,\, t\ge T_0} \min_{i<j} (\boldsymbol{n}_t)_{i,j} (\boldsymbol{z}_t)_{i,j}^2 \\
&\ge \phi^2\left(\frac{1}{2} + \frac{\varepsilon}{2}\right) \lim_{t\to\infty,\, t\ge T_0} \min_{i<j} (\boldsymbol{n}_t)_{i,j} = \infty
\end{aligned}$$

holds a.s., where we have used that $\pi \in \Pi_\infty$ assures $(\boldsymbol{n}_t)_{i,j} \to \infty$ a.s. as $t \to \infty$ for any $(i,j) \in (m)_2$. In the case $\pi \in \Pi_\infty^{\ln\ln}$ we similarly obtain the stronger result

$$\lim_{t\to\infty,\, t\ge T_0} \frac{\tilde{\mu}_t}{\ln\ln t} \ge \phi^2\left(\frac{1}{2} + \frac{\varepsilon}{2}\right) \lim_{t\to\infty,\, t\ge T_0} \frac{\min_{i<j}(\boldsymbol{n}_t)_{i,j}}{\ln\ln t} = \infty \quad \text{a.s.}$$

□

Lemma 5.12. *Let $p \in (0,1)$ be fixed and let $\{X_k^{(p)}\}_{k\in\mathbb{N}}$ be a sequence of i.i.d. random variables $X_k^{(p)} \sim \mathrm{Ber}(p)$ and $\phi : [0,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ given by $\phi(x) = 2\arcsin(\sqrt{x}) - \pi/2$ as above. Then,*

$$\sqrt{n}\left(\phi\left(\frac{1}{n}\sum_{k=1}^n X_k^{(p)}\right) - \phi(p)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Proof. Since $X_k^{(p)}$ has expectation p and variance $p(1-p)$, the central limit theorem ensures

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^n X_k^{(p)} - p\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, p(1-p)).$$

As ϕ is differentiable, the *delta method* [Van der Vaart, 2000] thus yields

$$\sqrt{n}\left(\phi\left(\frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(p)}\right) - \phi(p)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, p(1-p)\phi'(p)^2).$$

Due to $\phi'(x) = \frac{1}{\sqrt{-(x-1)x}}$, straightforward calculations show $p(1-p)\phi'(p)^2 = 1$. $\square$

5.4.2. Passive WST Testing

Algorithm 22 LRT-based solution to $\mathcal{P}_{\mathrm{WST}}^{m,0,\alpha,\beta}$

Input: m, α, β, κ, c, a sampling strategy π.

Initialization: Let $n' \in \mathbb{N}$ be such that $\sum_{n \geq n'} \frac{1}{n^\kappa} \leq 1$.

Choose $d := 3(1+\frac{1}{\sqrt{2}})^2$, $w := (60\binom{m}{2})^{\frac{2}{3}}$ and

$$n'' := \frac{2d}{c^2}\ln\left(\frac{2w\ln\left(\frac{3dw}{2c^2(\alpha\wedge\beta)^{\frac{2}{3}}\ln(3/2)}\right)}{(\alpha\wedge\beta)^{\frac{2}{3}}\ln(3/2)}\right)+1, \quad n''' := \frac{64}{(1-2c)^4}\left(\ln\left(\frac{6\binom{m}{2}}{\alpha\wedge\beta}\right)+\kappa\right).$$

Fix $q := \exp\left(-{(1-2c)^2}/{4(2-2c)^2}\right)$ and $\tilde{n} := \left\lceil \max\left\{n', n'', n''', \log_q\left({(1-q)(\alpha\wedge\beta)}/{3\binom{m}{2}}\right)\right\}\right\rceil$.

Define $L := L(c) := \sup_{x\in[c/2,1-c/2]} |\phi'(x)| = 1/\sqrt{\frac{c}{2}(1-\frac{c}{2})}$

for $t \in \mathbb{N}$ **do**

  Choose $\{i,j\} \sim \pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0\leq t'\leq t-1})$, w.l.o.g. $i<j$.

  Observe $X_{i,j}^{[t]} \sim \mathrm{Ber}(q_{i,j})$.

  Update $\boldsymbol{n}_t$ and $\boldsymbol{w}_t$ accordingly.

  Let $\hat{c}_{i,j} := \hat{c}_{i,j}(t) := \sqrt{\frac{1}{2(\boldsymbol{n}_t)_{i,j}}\ln\left(\frac{6\binom{m}{2}(\boldsymbol{n}_t)_{i,j}^\kappa}{\alpha\wedge\beta}\right)}$ for all $(i,j) \in (m)_2$.

  Let $\hat{E}_t := \left\{(i,j)\in(m)_2 \,\middle|\, \frac{(\boldsymbol{w}_t)_{i,j}}{(\boldsymbol{n}_t)_{i,j}} \in [0, c+\hat{c}_{i,j}) \cup (1-c-\hat{c}_{i,j}, 1]\right\}$ and $\hat{K}_t := (m)_2 \setminus \hat{E}_t$.

  Let $l_{\mathrm{WST}}(t) := f(\alpha, t)$ and $l_{\neg\mathrm{WST}}(t) := f(\beta, t)$ with

$$f(\gamma, t) := \frac{3L^2|\hat{K}_t|}{4}\left[1+\frac{1}{\sqrt{2}}\right]^2 \ln\left[\frac{(60|\hat{K}_t|)^{\frac{2}{3}}\ln\frac{3t}{2}}{\gamma^{\frac{2}{3}}\ln\frac{3}{2}}\right].$$

  Calculate $\tilde{\lambda}_t$ and $\tilde{\mu}_t$ from (5.15) and (5.18)

  if $\min_{i<j}(\boldsymbol{n}_t)_{i,j} \geq \tilde{n}$ and $\tilde{\lambda}_t > l_{\mathrm{WST}}(t)$ **then**

    return WST.

  if $\min_{i<j}(\boldsymbol{n}_t)_{i,j} \geq \tilde{n}$ and $\tilde{\mu}_t > l_{\neg\mathrm{WST}}(t)$ **then**

    return $\neg$WST.

With the deliberations above, we obtain with Alg. 22 an online WST testing algorithm which guarantees reasonable type I and type II error bounds by choosing appropriate critical values for $\tilde{\lambda}_t$ and $\tilde{\mu}_t$. The identity $L(c) = 1/\sqrt{-\frac{c}{2}(1-\frac{c}{2})}$ stated in Alg. 22 can be shown straight-forwardly and is verified in detail later on directly before the proof of Thm. 5.15.

Theorem 5.13. *For any $m \in \mathbb{N}_{\geq 3}$, $\alpha, \beta \in (0,1)$, $c \in (0, 1/2)$, $\pi \in \Pi_\infty^{\ln\ln}$ and $\kappa > 1$, Alg. 22 called with these parameters solves $\mathcal{P}_{\mathrm{WST}}^{m,0,\alpha,\beta}$.*

For the sake of convenience, we defer a detailed proof of Thm. 5.13 to Sec. 5.4.3 below. Before, let us explain at this point the rough idea behind the obtained Alg. 22, which we denote by $\mathcal{A}$ and suppose to be initialized with appropriate parameters m, α, β, c, π and κ in the following. Naturally, whether some $\mathbf{Q} \in \mathcal{Q}_m^0$ is WST or not depends exclusively on the set $\{(i,j) \in (m)_2 \,|\, q_{i,j} > 1/2\}$, and in particular the closer $q_{i,j}$ is to $1/2$, the more difficult it is to decide whether (i,j) is in this set or not. With this in mind, the parameter c serves as a threshold at which $\mathcal{A}$ classifies pairs (i,j) as difficult, namely it estimates the set

$$E \coloneqq \{(i',j') \in (m)_2 \,|\, q_{i',j'} \in [0,c) \cup (1-c, 1]\}$$

via $\hat{E}_t$ and regards $\hat{K}_t$ as the set of *difficult* pairs of winning probabilities[5]. The value $\tilde{n}$ is chosen large enough so that, as soon as $\mathcal{A}$ has queried every $\{i,j\} \in [m]_2$ at least $\tilde{n}$ times each, we have:

- $\mathcal{A}$ is confident enough whether $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ holds for each non-difficult pair $(i,j) \in E$,
- the deviation $\left|\phi((\boldsymbol{w}_t)_{i,j}/(\boldsymbol{n}_t)_{i,j}) - \phi(q_{i,j})\right|$ is small enough with high confidence for all difficult pairs $(i,j) \in (m)_2 \setminus E$.

The technical requirements on $\tilde{n}$ made in Alg. 22 suffice to ensure satisfactory error guarantees for $\mathcal{A}$. More precisely,

- $\tilde{n} \geq n'$ allows to control via Lem. 2.4 the probability that the algorithms estimation $\hat{E}_t$ of E is not contained in E,
- $\tilde{n} \geq n''$ helps us to bound with Lem. 2.7 the error probability in case all pairs $(i,j) \in (m)_2$ are classified correctly in some appropriate sense,
- and $\tilde{n} \geq \log_q \left({}^{(1-q)(\alpha \wedge \beta)}/{}_{3\binom{m}{2}} \right)$ allows to bound by means of Lem. 2.5 the probability in case a corresponding misclassification happens.

As the proof of Thm. 5.13 below reveals, choosing $\tilde{n} = \lceil \max\{n', n'', \log_q \left({}^{(1-q)(\alpha\wedge\beta)}/{}_{3\binom{m}{2}} \right)\} \rceil$ would actually already suffice for the results obtained in the passive scenario; the additional assumption $\tilde{n} \geq n'''$ is only used to achieve the corresponding sample complexity bound in the active scenario (Thm. 5.15), but for the sake of convenience, we have incorporated n''' already in the statement of Alg. 22.

The confidence length terms $\hat{c}_{i,j}(t)$ as well as the choices of the decision boundaries $l_{\mathrm{WST}}(t)$ and $l_{\neg\mathrm{WST}}(t)$ are mainly due to Lem. 2.7. It is worth noting that $l_{\mathrm{WST}}(t)$ and $l_{\neg\mathrm{WST}}(t)$ are decreasing with the cardinality of $\hat{E}_t$, so that the easier the problem instance $\mathbf{Q}$ (corresponding to a large cardinality of $|E|$), the sooner the termination of $\mathcal{A}$. The remainder of this section is dedicated to prepare and present the proof of Thm. 5.13.

5.4.3. Proof of Thm. 5.13

We prepared the proof of Thm. 5.13 already with Lem. 2.7, which allows us to control the tail probabilities of the summands occuring in the test statistics $\tilde{\lambda}_t$ resp. $\tilde{\mu}_t$. As further preparation, the following lemma shows that changing specific components $\boldsymbol{z}_t$ in an appropriate way if their sign is not coherent with the underlying component of the transformed parameter $\boldsymbol{\theta} = (\theta_{i,j})_{1 \leq i,j \leq m}$ leads to a modified vector, which is an element of the parameter space of $\boldsymbol{\theta}$, that is, of $\Theta_m(\mathrm{WST})$ or $\Theta_m(\neg\mathrm{WST})$.

[5]Note here, that $\pi \in \Pi_\infty$ and the law of large numbers ensure $\hat{E}_t \to E$ as $t \to \infty$.

Lemma 5.14. *If* $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$*,* $E \subseteq (m)_2$*,* $\boldsymbol{\theta} \in \Theta_m(\mathrm{X})$ *and* $\boldsymbol{z} \in \Theta_m$ *are s.t.*

$$\theta_{i,j} z_{i,j} \geq 0 \quad \text{and} \quad \theta_{i,j} \neq 0 \quad \text{for every } (i,j),(j,i) \in E,$$

then $\boldsymbol{z}' = (z'_{i,j})_{1 \leq i,j \leq m} \in \Theta_m$ *defined via*

$$z'_{i,j} := \begin{cases} z_{i,j}, & \text{if } (i,j) \in E \text{ or } (j,i) \in E, \\ \theta_{i,j}, & \text{otherwise,} \end{cases}$$

fulfills $\boldsymbol{z}' \in \Theta_m(\mathrm{X})$.

Proof of Lem. 5.14. Let $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$, $\boldsymbol{\theta} \in \Theta_m(\mathrm{X})$ and $\boldsymbol{z} \in \Theta_m$ be fixed and such that they fulfill the above mentioned constraints. Let $(i,j) \in (m)_2$ be arbitrary. If $(i,j) \in E$ or $(j,i) \in E$ holds, we have due to $\theta_{i,j} z'_{i,j} = \theta_{i,j} z_{i,j} \geq 0$ as well as $\theta_{i,j} \neq 0$ the implications

$$z'_{i,j} < 0 \Rightarrow \theta_{i,j} < 0 \text{ and } z'_{i,j} > 0 \Rightarrow \theta_{i,j} > 0. \tag{5.20}$$

In the other case $(i,j),(j,i) \notin E$ the definition of $\boldsymbol{z}'$ reveals $z'_{i,j} = \theta_{i,j}$, and thus (5.20) holds true as well.

We continue with a case distinction and start with the case $\mathrm{X} = \mathrm{WST}$. For this, assume $\boldsymbol{z}' \notin \Theta_m(\mathrm{WST})$. Then there exists some distinct $i', j', k' \in [m]$ with $z'_{i',j'}, z'_{j',k'} < 0$ and $z'_{i',k'} > 0$. As (5.20) holds for every $(i,j) \in (m)_2$ we thus obtain $\theta_{i',j'}, \theta_{j',k'}, \theta_{k',i'} > 0$. This contradicts $\boldsymbol{\theta} \in \Theta_m(\mathrm{WST})$.$\lightning$

Now, consider the case $\mathrm{X} = \neg\mathrm{WST}$. Assuming $\boldsymbol{z}' \notin \Theta_m(\neg\mathrm{WST})$ there does not exist any distinct $i', j', k' \in [m]$ with $z'_{i',j'}, z'_{j',k'}, z'_{k',i'} \leq 0$. In other words, for every distinct $i', j', k' \in [m]$ there exists some $(i'', j'') \in \{(i',j'),(j',k'),(k',i')\}$ with $z'_{i'',j''} > 0$, which implies $\theta_{i'',j''} > 0$ according to (5.20). Thus, there does not exist any distinct $i', j', k' \in [m]$ with $\theta_{i',j'}, \theta_{j',k'}, \theta_{k',i'} \leq 0$, which contradicts $\boldsymbol{\theta} \in \Theta_m(\neg\mathrm{WST})$.$\lightning$ □

In the proof of Thm. 5.13, we assume w.l.o.g. that a family $\{X_{i,j}^{[n]}\}_{n \in \mathbb{N}, 1 \leq i < j \leq m}$ of independent random variables $X_{i,j}^{[n]} \sim \mathrm{Ber}(p_{i,j})$ exists such that $(\boldsymbol{w}_t)_{i,j} = \sum_{n=1}^{(\boldsymbol{n}_t)_{i,j}} X_{i,j}^{[n]}$ holds for every $(i,j) \in (m)_2$. Recall that we assume throughout Sec. 5.4 that $(\boldsymbol{n}_0)_{i,j} = 1$ for every $(i,j) \in (m)_2$. If $\pi \in \Pi_\infty$ is fixed, we define $\tilde{\lambda}_t$ and $\tilde{\mu}_t$ as in Sec. 5.4.1. Moreover, for $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$, we write $\mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}, \Theta_m(\mathrm{X}))$ for the distance of $\boldsymbol{z} \in \Theta_m$ to $\Theta_m(\mathrm{X})$, i.e., $\mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}, \Theta_m(\mathrm{X})) = \min_{\boldsymbol{\theta} \in \Theta_m(\mathrm{X})} \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}, \Theta_m(\mathrm{X}))$.

Proof of Thm. 5.13. Suppose $m \in \mathbb{N}$, $\alpha, \beta \in (0,1)$, $c \in (0, 1/2)$, $\pi \in \Pi_\infty^{\ln\ln}$ and $\kappa > 1$ to be fixed and write $\mathcal{A}$ for Alg. 22 called with these parameters. For convenience, we abbreviate $T := T^{\mathcal{A}}$. We split the proof into five steps.

Step 1: Almost sure finiteness of T

Since $\pi \in \Pi_\infty^{\ln\ln} \subseteq \Pi_\infty$ holds by assumption, the stopping time

$$T_{\tilde{n}} := \inf\{t \in \mathbb{N} \,|\, (\boldsymbol{n}_t)_{i,j} \geq \tilde{n} \text{ for every distinct } i,j \in [m]\}$$

is a.s. finite and we have $T_{\tilde{n}} \leq T$. Thus, using $\max\{\tilde{\lambda}_t, \tilde{\mu}_t\} \in \{\tilde{\lambda}_t, \tilde{\mu}_t\}$, we can estimate

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\theta}}(T = \infty) &= \mathbb{P}_{\boldsymbol{\theta}}(\forall t \geq T_{\tilde{n}} : \tilde{\lambda}_t \leq l_{\mathrm{WST}}(t) \text{ and } \tilde{\mu}_t \leq l_{\neg\mathrm{WST}}(t)) \\ &\leq \mathbb{P}_{\boldsymbol{\theta}}(\forall t \geq T_{\tilde{n}} : \max\{\tilde{\lambda}_t, \tilde{\mu}_t\} \leq \max\{l_{\mathrm{WST}}(t), l_{\neg\mathrm{WST}}(t)\}) \\ &= 1 - \mathbb{P}_{\boldsymbol{\theta}}(\exists t \geq T_{\tilde{n}} : \max\{\tilde{\lambda}_t, \tilde{\mu}_t\} > \max\{l_{\mathrm{WST}}(t), l_{\neg\mathrm{WST}}(t)\}). \end{aligned} \tag{5.21}$$

As $l_{\text{WST}}(t), l_{\neg\text{WST}}(t) \in \mathcal{O}(\ln \ln t)$ as $t \to \infty$, Prop. 5.11 lets us infer that

$$\frac{\max\{\tilde{\lambda}_t, \tilde{\mu}_t\}}{\max\{l_{\text{WST}}(t), l_{\neg\text{WST}}(t)\}} \longrightarrow \infty$$

almost surely as $t \to \infty$, consequently for every $t_1 \in \mathbb{N}$ we obtain

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\exists t \geq t_1 : \max\{\tilde{\lambda}_t, \tilde{\mu}_t\} > \max\{l_{\text{WST}}(t), l_{\neg\text{WST}}(t)\}\right) = 1.$$

Therefore, abbreviating $t_1 \wedge T_{\tilde{n}} = \min\{t_1, T_{\tilde{n}}\}$, an application of Lebesgue's dominated convergence theorem yields

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(\exists t \geq T_{\tilde{n}} : \max\{\tilde{\lambda}_t, \tilde{\mu}_t\} > \max\{l_{\text{WST}}(t), l_{\neg\text{WST}}(t)\}\right) \\
&= \lim_{t_1 \to \infty} \mathbb{P}_{\boldsymbol{\theta}}\left(\exists t \geq t_1 \wedge T_{\tilde{n}} : \frac{\max\{\tilde{\lambda}_t, \tilde{\mu}_t\}}{\max\{l_{\text{WST}}(t), l_{\neg\text{WST}}(t)\}} > 1\right) \\
&= 1,
\end{aligned}$$

and thus $T < \infty$ almost surely follows from (5.21).

Step 2: Decomposition of the weighted projection tail probabilities
Let $\boldsymbol{\theta} = \phi(\mathbf{Q}) \in \Theta_m$. Due to $\Theta_m = \Theta_m(\text{WST}) \cup \Theta_m(\neg\text{WST})$ there exists some $\text{X} \in \{\text{WST}, \neg\text{WST}\}$ with $\boldsymbol{\theta} \in \Theta_m(\text{X})$. Regarding the definitions of $\tilde{\lambda}_t$ and $\tilde{\mu}_t$ and taking the arbitrariness of $\boldsymbol{\theta}$ into account it is sufficient for the theoretical guarantees of this test to show

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\mathrm{d}_{\boldsymbol{n}_T}(\boldsymbol{z}_T, \Theta_m(\text{X})) > l_{\text{X}}(T)\right) \leq \begin{cases} \alpha, & \text{if X} = \text{WST}, \\ \beta, & \text{if X} = \neg\text{WST}. \end{cases}$$

We call some $(i,j) \in (m)_2$ *classified correctly* if $(\boldsymbol{z}_T)_{i,j}\theta_{i,j} \geq 0$ holds, otherwise we say that (i,j) is *misclassified*. Moreover, we write

$$A_{i,j} \coloneqq \{(\boldsymbol{z}_T)_{i,j}\theta_{i,j} < 0\}$$

for the event that (i,j) is misclassified, i.e., on $A_{i,j}$ either $(\boldsymbol{z}_T)_{i,j} > 0$ and $\theta_{i,j} < 0$ or $(\boldsymbol{z}_T)_{i,j} < 0$ and $\theta_{i,j} > 0$ hold. Define the limit set of $\hat{E}_t$ as

$$E \coloneqq \{(i,j) \in (m)_2 \,|\, q_{i,j} \in [0,c) \cup (1-c,1]\}$$

and further set

$$\mathcal{E} \coloneqq \left\{\hat{E}_T \subseteq E\right\}^c,$$

which is the event that, at its termination, the algorithms estimation of E is not contained in E. With this, we obtain

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(\mathrm{d}_{\boldsymbol{n}_T}(\boldsymbol{z}_T, \Theta_m(\text{X})) > l_{\text{X}}(T)\right) \\
&\leq \mathbb{P}_{\boldsymbol{\theta}}\left(\left\{\mathrm{d}_{\boldsymbol{n}_T}(\boldsymbol{z}_T, \Theta_m(\text{X})) > l_{\text{X}}(T)\right\} \cap \bigcup\nolimits_{(i,j)\in E} A_{i,j} \cap \mathcal{E}^c\right) \\
&\quad + \mathbb{P}_{\boldsymbol{\theta}}\left(\left\{\mathrm{d}_{\boldsymbol{n}_T}(\boldsymbol{z}_T, \Theta_m(\text{X})) > l_{\text{X}}(T)\right\} \cap \bigcap\nolimits_{(i,j)\in E} A^c_{i,j} \cap \mathcal{E}^c\right) \\
&\quad + \mathbb{P}_{\boldsymbol{\theta}}\left(\mathcal{E}\right).
\end{aligned} \tag{5.22}$$

In the following three steps of the proof we will bound each summand in (5.22) separately by $\frac{\min\{\alpha,\beta\}}{3}$.

Step 3: Bounding the tail probabilities of the misclassified pairs
To obtain an estimate of the first summand in (5.22) note that for every distinct $i, j \in [m]$ the equality

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\theta}}(A_{i,j}) &= \mathbb{P}_{\boldsymbol{\theta}}((\boldsymbol{z}_T)_{i,j}\theta_{i,j} < 0)\\ &= \mathbb{P}\left(\phi\left(\frac{1}{(\boldsymbol{n}_T)_{i,j}}\sum\nolimits_{k=1}^{(\boldsymbol{n}_T)_{i,j}} X_k^{(q_{i,j})}\right)\phi(q_{i,j}) < 0\right)\end{aligned}$$

holds for a sequence $\{X_k^{(q_{i,j})}\}_{k\in\mathbb{N}}$ of i.i.d. random variables $X_k^{(q_{i,j})} \sim \mathrm{Ber}(q_{i,j})$ due to the definition of $(\boldsymbol{z}_T)_{i,j}$. Therefore, using $(\boldsymbol{n}_T)_{i,j} \geq \tilde{n}$ for every distinct $(i,j) \in E$, Lem. 2.5 yields

$$\mathbb{P}_{\boldsymbol{\theta}}(A_{i,j}) \leq \mathbb{P}\left(\exists n \geq \tilde{n} : \phi\left(\frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(q_{i,j})}\right)\phi(q_{i,j}) < 0\right) \leq \frac{\min\{\alpha,\beta\}}{3\binom{m}{2}}.$$

Thus, by means of the union bound we infer that

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\theta}}\left(\left\{\mathrm{d}_{\boldsymbol{n}_T}(\boldsymbol{z}_T, \Theta_m(\mathrm{X})) > l_{\mathrm{X}}(T)\right\} \cap \bigcup\nolimits_{(i,j)\in E} A_{i,j} \cap \mathcal{E}^c\right)\\ &\leq \sum\nolimits_{(i,j)\in E} \mathbb{P}_{\boldsymbol{\theta}}(A_{i,j}) \leq \frac{\min\{\alpha,\beta\}}{3}.\end{aligned} \tag{5.23}$$

Step 4: Bounding the tail probabilities of the correctly classified pairs
Next, we show

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\left\{\mathrm{d}_{\boldsymbol{n}_T}(\boldsymbol{z}_T, \Theta_m(\mathrm{X})) > l_{\mathrm{X}}(T)\right\} \cap \bigcap\nolimits_{(i,j)\in E} A_{i,j}^c \cap \mathcal{E}^c\right) \leq \frac{1}{3}\begin{cases}\alpha, & \text{if X} = \text{WST},\\ \beta, & \text{if X} = \neg\text{WST}.\end{cases} \tag{5.24}$$

For this, let $K := (m)_2 \setminus E$ and define $\boldsymbol{z'} \in \Theta$ via

$$z'_{i,j} := \begin{cases}(\boldsymbol{z}_T)_{i,j}, & (i,j) \in E \text{ or } (j,i) \in E,\\ \theta_{i,j}, & \text{otherwise},\end{cases}$$

for every distinct $i, j \in [m]$. Observe that $\theta_{i',j'}(\boldsymbol{z}_T)_{i',j'} = \phi(q_{i',j'})(\boldsymbol{z}_T)_{i',j'} \geq 0$ holds for every $(i',j') \in E$ on the event $A' := \bigcap_{(i',j')\in E} A_{i',j'}^c$. Since $c < \frac{1}{2}$ ensures $\theta_{i,j} \neq 0$ for every $(i,j) \in E$ and $\boldsymbol{\theta} \in \Theta_m(\mathrm{X})$ holds by assumption, Lem. 5.14 implies $\boldsymbol{z'} \in \Theta_m(\mathrm{X})$ on the event A'. We obtain the estimate

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\theta}}\left(\left\{\mathrm{d}_{\boldsymbol{n}_T}(\boldsymbol{z}_T, \Theta_m(\mathrm{X})) > l_{\mathrm{X}}(T)\right\} \cap \bigcap\nolimits_{(i,j)\in E} A_{i,j}^c \cap \mathcal{E}^c\right)\\ &\leq \mathbb{P}_{\boldsymbol{\theta}}\left(\left\{\mathrm{d}_{\boldsymbol{n}_T}(\boldsymbol{z}_T, \boldsymbol{z'}) > l_{\mathrm{X}}(T)\right\} \cap \mathcal{E}^c\right)\\ &= \mathbb{P}_{\boldsymbol{\theta}}\left(\left\{\sum\nolimits_{(k,l)\in K} (\boldsymbol{n}_T)_{k,l}((\boldsymbol{z}_T)_{k,l} - \theta_{k,l})^2 > l_{\mathrm{X}}(T)\right\} \cap \mathcal{E}^c\right)\\ &\leq \sum\nolimits_{(k,l)\in K} \mathbb{P}_{\boldsymbol{\theta}}\left(\left\{(\boldsymbol{n}_T)_{k,l}((\boldsymbol{z}_T)_{k,l} - \theta_{k,l})^2 > \frac{l_{\mathrm{X}}(T)}{|K|}\right\} \cap \mathcal{E}^c\right).\end{aligned} \tag{5.25}$$

In order to bound this further, we may suppose without loss of generality $|K| \geq 1$ in the following and write $\gamma_{\mathrm{WST}} = \alpha$ and $\gamma_{\neg\mathrm{WST}} = \beta$ for the moment. Choosing $\varepsilon' := \frac{1}{2}$ and

$$\delta'_{\mathrm{X}} := \left(\frac{\gamma_{\mathrm{X}}}{60|K|}\right)^{\frac{2}{3}} \ln\frac{3}{2} \in (0, {}^{\ln(1+\varepsilon')}/_{\exp(1)})$$

we have $\frac{\gamma_{\mathrm{X}}}{12|K|} = 5\left(\frac{\delta'_{\mathrm{X}}}{\ln(3/2)}\right)^{\frac{3}{2}} = \frac{2+\varepsilon'}{\varepsilon'}\left(\frac{\delta'_{\mathrm{X}}}{\ln(1+\varepsilon')}\right)^{1+\varepsilon'}$. As $\hat{E}_T \subset E$ and thus $|K| \leq |\hat{K}_T|$ holds on $\mathcal{E}^c$, we have with regard to the definition of $l_{\mathrm{X}}(T)$ on this event

$$\begin{aligned}
\frac{l_{\mathrm{X}}(T)}{|K|} &\geq \frac{3L^2}{4}\left[1+\frac{1}{\sqrt{2}}\right]^2 \ln\left[\frac{\ln(3t/2)(60|\hat{K}_T|)^{2/3})}{\gamma_{\mathrm{X}}^{2/3}\ln(3/2)}\right] \\
&\geq \frac{3L^2}{4}\left[1+\frac{1}{\sqrt{2}}\right]^2 \ln\left[\frac{\ln(3t/2)(60|K|)^{2/3})}{\gamma_{\mathrm{X}}^{2/3}\ln(3/2)}\right] \\
&= \frac{1}{2}L^2(1+\sqrt{\varepsilon'})^2(1+\varepsilon')\ln\left(\frac{\ln((1+\varepsilon')t)}{\delta'_{\mathrm{X}}}\right).
\end{aligned}$$

Using $|K| \leq \binom{m}{2}$ and $\gamma_{\mathrm{X}} \geq \min\{\alpha, \beta\}$ a straightforward calculation reveals

$$\tilde{n} \geq n'' \geq \frac{d'}{c^2}\ln\left(\frac{2\ln\left((1+\varepsilon')d'/(c^2\delta'_{\mathrm{X}})\right)}{\delta'_{\mathrm{X}}}\right) + 1$$

with $d' = 2(1+\sqrt{\varepsilon'})^2(1+\varepsilon')$. Note that $T \geq \binom{m}{2}\tilde{n}$ and in particular for each $(k,l) \in (m)_2$ we have $T \geq (\boldsymbol{n}_T)_{k,l} \geq \tilde{n}$ and thus $l_{\mathrm{X}}(T) \geq l_{\mathrm{X}}((\boldsymbol{n}_T)_{k,l})$. Hence, an application of Lem. 2.7 guarantees that for each $(k,l) \in K$

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(\left\{(\boldsymbol{n}_T)_{k,l}((\boldsymbol{z}_T)_{k,l} - \theta_{k,l})^2 > \frac{l_{\mathrm{X}}(T)}{|K|}\right\} \cap \mathcal{E}^c\right) \\
&\leq \mathbb{P}_{\boldsymbol{\theta}}\left(\left\{(\boldsymbol{n}_T)_{k,l}((\boldsymbol{z}_T)_{k,l} - \theta_{k,l})^2 > \frac{l_{\mathrm{X}}((\boldsymbol{n}_T)_{k,l})}{|K|}\right\} \cap \mathcal{E}^c\right) \\
&\leq \frac{4\gamma_{\mathrm{X}}}{12|K|} = \frac{1}{3|K|}\begin{cases}\alpha, & \text{if X} = \text{WST}, \\ \beta, & \text{if X} = \neg\text{WST}.\end{cases}
\end{aligned} \tag{5.26}$$

Combining (5.25) with (5.26) yields (5.24).

Step 5: Negligibility of the event $\mathcal{E}$

Regarding (5.22) and the already shown inequalities (5.23) and (5.24) it is sufficient for the theoretical guarantees on the type I/II errors of $\mathcal{A}$ to show

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathcal{E}) \leq \frac{\min\{\alpha, \beta\}}{3}. \tag{5.27}$$

Note that

$$
\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}(\mathcal{E}) = \mathbb{P}_{\boldsymbol{\theta}}(\exists\,(i,j) \in \hat{E}_T \cap K) \\
&\leq \sum\nolimits_{(i,j)\in K} \mathbb{P}_{\boldsymbol{\theta}}\left(\frac{(\boldsymbol{w}_T)_{i,j}}{(\boldsymbol{n}_T)_{i,j}} < c - \hat{c}_{i,j}(T)\right) + \sum\nolimits_{(i,j)\in K} \mathbb{P}_{\boldsymbol{\theta}}\left(\frac{(\boldsymbol{w}_T)_{i,j}}{(\boldsymbol{n}_T)_{i,j}} > 1 - c + \hat{c}_{i,j}(T)\right) \\
&\leq \sum\nolimits_{(i,j)\in K} \mathbb{P}\left(\exists n \geq \tilde{n} : \frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(q_{i,j})} < c - \hat{c}_{i,j}(n)\right) \\
&\quad + \sum\nolimits_{(i,j)\in K} \mathbb{P}\left(\exists n \geq \tilde{n} : \frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(q_{i,j})} > 1 - c + \hat{c}_{i,j}(n)\right).
\end{aligned}
$$

As $(\boldsymbol{n}_T)_{i,j}\tilde{n} \geq n'$ for all and $c < q_{i,j} < 1-c$ hold for every $(i,j) \in K$, we can infer from this by means of Lem. 2.4 with $\gamma = {}^{\min\{\alpha,\beta\}}/6\binom{m}{2}$ that

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(\mathcal{E}) &\leq \sum\nolimits_{(i,j)\in K} \mathbb{P}\left(\exists n \geq n' : \frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(q_{i,j})} - q_{i,j} < -\hat{c}_{i,j}(n)\right) \\
&\quad + \sum\nolimits_{(i,j)\in K} \mathbb{P}\left(\exists n \geq n' : \frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(q_{i,j})} - q_{i,j} > \hat{c}_{i,j}(n)\right) \\
&\leq \binom{m}{2}\left(\frac{\min\{\alpha,\beta\}}{6\binom{m}{2}} + \frac{\min\{\alpha,\beta\}}{6\binom{m}{2}}\right) = \frac{\min\{\alpha,\beta\}}{3}.
\end{aligned}
$$

This shows (5.27).

□

5.4.4. Active Online Testing

As already discussed in Sec. 5.3, in the passive online testing scenario the algorithm has no influence on the sampling strategy π. Since the explicit termination time of any test component highly depends on π, deriving general sample complexity bounds is cumbersome. Hence, we restricted the analysis of the test from Thm. 5.13 to proving a.s. finite termination time and adherence to predefined error bounds in case $\pi \in \Pi_{\infty}^{\ln\ln}$.

In the active online testing scenario, however, where the sampling strategy and the testing component are more interleaved, one can specify a sampling strategy to derive results on the expected termination time of the test. Needless to say, the relation $\mathbf{Q} \in \mathcal{Q}_m^0$ the algorithm is started with plays a crucial role for the termination time. As mentioned above, the closer its entries $q_{i,j}$ are to $\frac{1}{2}$, the closer $\phi(\mathbf{Q})$ is to the boundary $\partial\Theta_m(\text{WST}) = \partial\Theta_m(\neg\text{WST})$ and thus the harder it becomes to decide correctly whether $\phi(\mathbf{Q}) \in \Theta_m(\text{WST})$ or $\phi(\mathbf{Q}) \in \Theta_m(\neg\text{WST})$ holds. Encouraged by this idea, we analyze the worst-case sample complexity of our algorithm only on instances $\mathbf{Q} \in \mathcal{Q}_m^h$ for some arbitrary but fixed $h \in (0, 1/2)$. In the following corollary, we do this analysis for the sampling strategy $\pi =$ RoundRobin, which iterates through the set of all pairs in a deterministic way, i.e., it repeatedly queries $\{1,2\},\ldots,\{1,m\},\{2,3\},\ldots,\{m-1,m\}$. This sampling strategy ensures that, for all $t \in \mathbb{N}$,

$$
\max\nolimits_{(i,j)\in(m)_2}(\boldsymbol{n}_t)_{i,j} - \min\nolimits_{(i,j)\in(m)_2}(\boldsymbol{n}_t)_{i,j} \leq 1
$$

is fulfilled and thus arrives as fast as possible at that time t', where $(\boldsymbol{n}_{t'})_{i,j} \geq \tilde{n}$ holds for every $(i,j) \in (m)_2$. For the sake of convenience we only analyze the case $\alpha = \beta$.

Theorem 5.15. *Let $\kappa > 1$ be fixed. For any $m \in \mathbb{N}_{\geq 3}$, $\gamma \in (0,1)$ and $h \in (0, 1/4)$, $\mathcal{A}$:=Alg. 22 called with parameters m, $\alpha = \gamma$, $\beta = \gamma$, κ, $c = 1/2 - h/2$ and $\pi =$* RoundRobin *solves $\mathcal{P}^{m,0,\gamma}_{\mathrm{WST}}$ and fulfills*

$$\sup_{\mathbf{Q} \in \mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}} \left[T^{\mathcal{A}} \right] \in \mathcal{O} \left(\frac{m^{2\kappa}}{h^4} \ln \frac{1}{\gamma} \right) \tag{5.28}$$

as $\max\{m, \frac{1}{h}, \frac{1}{\gamma}\} \to \infty$.

Before proving Thm. 5.15, we further analyze the transformation $\phi : [0,1] \to [-\pi/2, \pi/2]$, $x \mapsto 2 \arcsin(\sqrt{x}) - \frac{\pi}{2}$, which we have used for the definition of the test statistics $\tilde{\lambda}_t$ and $\tilde{\mu}_t$. Its derivative

$$\phi'(x) = \frac{1}{\sqrt{-(x-1)x}}, \quad x \in (0,1),$$

is monotonically decreasing on $(0, 1/2)$, strictly positive on $(0,1)$ and symmetric around $1/2$. Thus, the function $L : (0, 1/2) \to (0, \infty), L(c) := \sup_{x \in [c/2, 1-c/2]} |\phi'(x)|$ defined in Alg. 22 is decreasing and fulfills $L(c) = \phi'(c/2)$ as well as $\lim_{c \nearrow 1/2} L(c) = 2$. Moreover, ϕ is bijective and its inverse $\phi^{-1} : [-\pi/2, \pi/2] \to [0,1]$, $\phi^{-1}(x) = \sin^2\left(\frac{2x+\pi}{4}\right)$ is strictly monotonically increasing.

Proof of Thm. 5.15. Let $\kappa > 1$ be fixed. Suppose $m \in \mathbb{N}_{\geq 3}$, $h \in (0, 1/4)$ and $\gamma \in (0,1)$ to be arbitrary and $\mathcal{A}$ as stated in the theorem. Since $\pi =$ RoundRobin $\in \Pi^{\ln \ln}_{\infty}$, Thm. 5.13 assures that $\mathcal{A}$ solves $\mathcal{P}^{m,0,\gamma}_{\mathrm{WST}}$. According to its definition, the termination time $T^{\mathcal{A}}$ is given as

$$\inf \left\{ t \in \mathbb{N} \,|\, (\tilde{\lambda}_t > l(t) \text{ or } \tilde{\mu}_t > l(t)) \text{ and } (\boldsymbol{n}_t)_{i,j} \geq \tilde{n} \; \forall (i,j) \in (m)_2 \right\}$$

with $\tilde{n} = \left\lceil \max\left\{ n', n'', n''', \log_q \left((1-q) \min\{\alpha,\beta\} / 3\binom{m}{2} \right) \right\} \right\rceil$ and

$$l(t) = \frac{3L^2 |\hat{K}_t|}{4} \left[1 + \frac{1}{\sqrt{2}} \right]^2 \ln \left[\frac{(60|\hat{K}_t|)^{\frac{2}{3}} \ln \frac{3t}{2}}{\gamma^{\frac{2}{3}} \ln \frac{3}{2}} \right],$$

with L as in Alg. 22 and in Lem. 2.7. Note here that $\tilde{n}$ directly depends on m and γ and via $n'' = n''(c)$ and $c = c(h) := 1/2 - h/2$ also on h, whereas L depends via c on h, i.e., we may write $\tilde{n} = \tilde{n}(m, h, \gamma)$ and $L = L(h)$. Started with some arbitrary $\mathbf{Q} \in \mathcal{Q}^h_m$, the termination time $T^{\mathcal{A}}$ of $\mathcal{A}$ can be bounded as

$$\mathbb{E}_{\mathbf{Q}} \left[T^{\mathcal{A}} \right] \leq 1 + \binom{m}{2} \tilde{n}(m,h,\gamma) + \sum_{t \geq \binom{m}{2} \tilde{n}(m,h,\gamma)} \mathbb{P}_{\mathbf{Q}} \left(T^{\mathcal{A}} > t \right). \tag{5.29}$$

In the following, we treat the second and third summand in (5.29) separately.

Part 1: We have

$$\binom{m}{2} \tilde{n}(m,h,\gamma) \in \mathcal{O} \left(\frac{m^2 \ln m}{h^4} \ln \frac{1}{\gamma} \right) \tag{5.30}$$

as $\max\{m, \frac{1}{h}, \frac{1}{\gamma}\} \to \infty$.

For $h \in (0,1/4)$ let $q(h) := e^{-\frac{(1-2c(h))^2}{4(2-2c(h))^2}} = e^{-\frac{1}{4}\frac{h^2}{(1+h)^2}} \in (0,1)$. Using that $\ln z \leq \frac{z-1}{\sqrt{z}}$ for $z > 1$ and $e^z - 1 \geq z$ for any $z \geq 0$ are fulfilled, we can estimate

$$\begin{aligned}\ln \frac{1}{1-q(h)} &\leq \frac{1/(1-q(h)) - 1}{\sqrt{1/(1-q(h))}} = \frac{q(h)}{\sqrt{1-q(h)}} = \frac{e^{-\frac{1}{4}\frac{h^2}{(1+h)^2}}}{\sqrt{1-e^{-\frac{1}{4}\frac{h^2}{(1+h)^2}}}} \\ &= \frac{e^{-\frac{1}{4}\frac{h^2}{(1+h)^2}}}{\sqrt{e^{-\frac{1}{4}\frac{h^2}{(1+h)^2}}}} \cdot \frac{1}{\sqrt{e^{\frac{1}{4}\frac{h^2}{(1+h)^2}} - 1}} \leq \frac{1}{\sqrt{\frac{h^2}{4(1+h)^2}}} = \frac{2(1+h)}{h}\end{aligned}$$

and hence a look at the definition of $q(h)$ reveals

$$\frac{\ln(1-q(h))}{\ln q(h)} \leq \frac{8(1+h)^3}{h^3} \text{ for all } h \in (0,1/4).$$

Consequently,

$$\begin{aligned}\log_{q(h)} \frac{(1-q(h))\gamma}{3\binom{m}{2}} &= \frac{\ln(\frac{\gamma}{3}) - \ln\binom{m}{2} + \ln(1-q(h))}{\ln q(h)} \\ &\leq \frac{4(1+h)^2(\ln(3/\gamma) + \ln\binom{m}{2})}{h^2} + \frac{8(1+h)^3}{h^3}\end{aligned}$$

holds, which is in $\mathcal{O}\big(\frac{\ln m}{h^3} \ln \frac{1}{\gamma}\big)$ as $\max\{m, \frac{1}{h}, \frac{1}{\gamma}\} \to \infty$. Moreover, the expressions

$$n''' = \frac{64}{(1-2c(h))^4}\left(\ln\left(\frac{6\binom{m}{2}}{\gamma}\right) + \kappa\right)$$

and

$$n'' = \frac{2d}{c(h)^2} \ln\left(\frac{2(60\binom{m}{2})^{\frac{2}{3}} \ln\left(\frac{3d(60\binom{m}{2})^{\frac{2}{3}}}{2c(h)^2\gamma^{\frac{2}{3}}\ln(3/2)}\right)}{\gamma^{\frac{2}{3}}\ln(3/2)}\right) + 1$$

with $d = 3(1+\frac{1}{\sqrt{2}})^2$ are in $\mathcal{O}\big(\frac{\ln m}{h^4} \ln \frac{1}{\gamma}\big)$ resp. $\mathcal{O}\big(\ln m \ln \frac{1}{\gamma}\big)$ as $\max\{m, \frac{1}{h}, \frac{1}{\gamma}\} \to \infty$. We assume w.l.o.g. that Alg. 22 chooses $n' \in \mathbb{N}$ minimal s.t. $\sum_{n \geq n'} \frac{1}{n^\kappa} \leq 1$ holds. Since κ is fixed, n' is constant. Hence, according to the definition of $\tilde{n}(m,h,\gamma)$, (5.30) follows.

Part 2: The term $\sum_{t \geq \binom{m}{2}\tilde{n}(m,h,\gamma)} \mathbb{P}_{\mathbf{Q}}\left(T^{\mathcal{A}} > t\right)$ is in $\mathcal{O}(m^{2\kappa})$ as $\max\{m, \frac{1}{h}, \frac{1}{\gamma}\} \to \infty$. As $\mathcal{A}$ uses RoundRobin as its sampling strategy and $(\boldsymbol{n}_0)_{i,j} = 1$ holds for every $(i,j) \in (m)_2$ by assumption, we have for all $t \in \mathbb{N}$ and $(i,j) \in (m)_2$

$$1 + \lfloor t/\tbinom{m}{2} \rfloor \geq (\boldsymbol{n}_t)_{i,j} \geq \lfloor t/\tbinom{m}{2} \rfloor \geq t/\tbinom{m}{2} - 1. \tag{5.31}$$

In particular, $\min_{i<j}(\boldsymbol{n}_t)_{i,j} \geq \tilde{n}(m,h,\gamma)$ is fulfilled as soon as $t \geq \binom{m}{2}\tilde{n}(m,h,\gamma)$. By the assumption of the theorem it holds that the limit set of $\hat{E}_t$ is

$$\begin{aligned}E &:= \{(i,j) \in (m)_2 \mid q_{i,j} \in [0, c(h)) \cup (1-c(h), 1]\} \\ &= \{(i,j) \in (m)_2 \mid q_{i,j} \in [0, 1-h/2) \cup (1/2+h/2, 1]\} \\ &= (m)_2,\end{aligned}$$

since $\mathbf{Q} \in \mathcal{Q}_m^h$. Define $\mathcal{E}_t := \{\hat{K}_t \neq \varnothing\}$ and note that

$$\mathcal{E}_t = \{\exists (i,j) \in (m)_2 \, : \, (i,j) \notin \hat{E}_t\} \quad \text{and}$$
$$\mathcal{E}_t^c = \{\hat{K}_t = \varnothing\} = \{\hat{E}_t = E = (m)_2\}.$$

Writing $\boldsymbol{\theta} = \phi(\mathbf{Q})$ we have, according to the definitions of $\mathcal{A}$ and its test statistics $\tilde{\lambda}_t$ and $\tilde{\mu}_t$,

$$\begin{aligned}\mathbb{P}_{\mathbf{Q}}\left(T^{\mathcal{A}} > t\right) &\le \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t \le l(t) \text{ and } \tilde{\mu}_t \le l(t)\right) \\ &\le \mathbb{P}_{\boldsymbol{\theta}}\left(\{\tilde{\lambda}_t \le l(t) \text{ and } \tilde{\mu}_t \le l(t)\} \cap \mathcal{E}_t^c\right) + \mathbb{P}_{\boldsymbol{\theta}}(\mathcal{E}_t). \qquad (5.32)\end{aligned}$$

On $\mathcal{E}_t^c$ it holds that $l(t) = 0$. Using that $\tilde{\lambda}_t$ and $\tilde{\mu}_t$ cannot be positive at the same time and also that $\tilde{\lambda}_t = 0 = \tilde{\mu}_t$ is only possible in case $\boldsymbol{z}_t \in \partial\Theta_m(\mathrm{WST})$, i.e., if $(\boldsymbol{z}_t)_{i,j} = 0$ for at least one $(i,j) \in (m)_2$, we can estimate the first term on the right-hand side of (5.32) via

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\{\tilde{\lambda}_t \le l(t) \text{ and } \tilde{\mu}_t \le l(t)\} \cap \mathcal{E}_t^c\right) \le \mathbb{P}_{\boldsymbol{\theta}}(\exists (i,j) \in (m)_2 \, : \, (\boldsymbol{z}_t)_{i,j} = 0)$$

and thus Lem. 2.3 ensures

$$\begin{aligned}&\sum\nolimits_{t \ge \binom{m}{2}\tilde{n}(m,h,\gamma)} \mathbb{P}_{\boldsymbol{\theta}}\left(\left\{\tilde{\lambda}_t \le l(t) \text{ and } \tilde{\mu}_t \le l(t)\right\} \cap \mathcal{E}_t^c\right) \\ &\le \sum\nolimits_{(i,j)\in(m)_2} \sum\nolimits_{t\in\mathbb{N}} \mathbb{P}_{\boldsymbol{\theta}}((\boldsymbol{z}_t)_{i,j} = 0) \le \binom{m}{2} h^{-2}. \qquad (5.33)\end{aligned}$$

The second term on the right-hand side of (5.32) can be bounded as follows[6]

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\theta}}(\mathcal{E}_t) &\le \sum\nolimits_{(i,j)\in(m)_2} \mathbb{P}_{\boldsymbol{\theta}}\left(\frac{(\boldsymbol{w}_t)_{i,j}}{(\boldsymbol{n}_t)_{i,j}} \ge c(h) - \hat{c}_{i,j}(t)\right) \mathbf{1}_{\{q_{i,j}\in[0,1/2-h)\}} \\ &\quad + \sum\nolimits_{(i,j)\in(m)_2} \mathbb{P}_{\boldsymbol{\theta}}\left(\frac{(\boldsymbol{w}_t)_{i,j}}{(\boldsymbol{n}_t)_{i,j}} \le 1 - c(h) + \hat{c}_{i,j}(t)\right) \mathbf{1}_{\{q_{i,j}\in(1/2+h,1]\}} \\ &\le \sum\nolimits_{(i,j)\in(m)_2} \mathbb{P}_{\boldsymbol{\theta}}\left(\frac{(\boldsymbol{w}_t)_{i,j}}{(\boldsymbol{n}_t)_{i,j}} - q_{i,j} \ge h/2 - \hat{c}_{i,j}(t)\right) \mathbf{1}_{\{q_{i,j}\in[0,1/2-h)\}} \\ &\quad + \sum\nolimits_{(i,j)\in(m)_2} \mathbb{P}_{\boldsymbol{\theta}}\left(\frac{(\boldsymbol{w}_t)_{i,j}}{(\boldsymbol{n}_t)_{i,j}} - q_{i,j} \le -h/2 + \hat{c}_{i,j}(t)\right) \mathbf{1}_{\{q_{i,j}\in(1/2+h,1]\}} \qquad (5.34)\end{aligned}$$

Using $\min_{i<j}(\boldsymbol{n}_t)_{i,j} \ge \tilde{n}(m,h,\gamma) \ge n'''$ for every $t \ge \binom{m}{2}\tilde{n}(m,h,\gamma)$, we obtain with regard to the definition of n''' that

$$\sqrt{(\boldsymbol{n}_t)_{i,j}} \ge \frac{8}{h^2}\left(\ln\left(^{6\binom{m}{2}}/_{\gamma}\right) + \kappa\right)$$

and together with $2\kappa\sqrt{(\boldsymbol{n}_t)_{i,j}} \ge 2\kappa \ln(\sqrt{(\boldsymbol{n}_t)_{i,j}}) = \ln((\boldsymbol{n}_t)_{i,j}^{\kappa})$ we see that

$$\frac{h^2}{4} \ge \frac{2}{\sqrt{(\boldsymbol{n}_t)_{i,j}}}\left(\ln\left(^{6\binom{m}{2}}/_{\gamma}\right) + \kappa\right) \ge \frac{2}{(\boldsymbol{n}_t)_{i,j}}\left(\ln\left(^{6\binom{m}{2}}/_{\gamma}\right) + \ln((\boldsymbol{n}_t)_{i,j}^{\kappa})\right)$$

[6]Here, we use the notation $x\mathbf{1}_A = \begin{cases} x, & \text{if } A \text{ holds,} \\ 0, & \text{otherwise.}\end{cases}$

holds, i.e., $\frac{h}{2} \geq 2\hat{c}_{i,j}(t)$ holds for every $t \geq \binom{m}{2}\tilde{n}(m,h,\gamma)$. With regard to (5.31), we can estimate (5.34) by means of Lem. 2.4 for every $t \geq \binom{m}{2}\tilde{n}(m,h,\gamma)$ further as

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\mathcal{E}_t\right) \leq \binom{m}{2}^{\kappa} \frac{\min\{\alpha,\beta\}}{3\,(t-\binom{m}{2})^{\kappa}},$$

where we also used that $(\boldsymbol{n}_0)_{i,j} = 1$ for all $(i,j) \in (m)_2$. Using that $\binom{m}{2}\tilde{n}(m,h,\gamma) - \binom{m}{2} \geq n'$ holds with n' s.t. $\sum_{n \geq n'} n^{-\kappa} \leq 1$, we obtain

$$\begin{aligned}\sum\nolimits_{t\geq\binom{m}{2}\tilde{n}(m,h,\gamma)} \mathbb{P}_{\boldsymbol{\theta}}\left(\mathcal{E}_t\right) &\leq \binom{m}{2}^{\kappa}\frac{\min\{\alpha,\beta\}}{3}\sum\nolimits_{t\geq\binom{m}{2}\tilde{n}(m,h,\gamma)} \frac{1}{(t-\binom{m}{2})^{\kappa}} \\ &\leq \binom{m}{2}^{\kappa}\frac{\min\{\alpha,\beta\}}{3}.\end{aligned}$$

Together with (5.32) and (5.33) this completes the proof of Part 2. Finally, combining (5.29), Part 1 and Part 2 completes the proof. □

5.4.5. Sequential Update Formulas for the LRT Statistics

Owing to the dynamic aspect of the dueling bandit framework, it is tempting to have sequential update formulas for the test statistics at hand. Recall the definition of $\mathcal{R}_m$ from Sec. 1.2. For the sake of convenience, abbreviate $\mathcal{R}_m^0 \coloneqq \mathcal{R}_m(\mathrm{WST})$ and $\mathcal{R}_m^1 \coloneqq \mathcal{R}_m(\neg\mathrm{WST}) = \mathcal{R}_m \setminus \mathcal{R}_m^0$, i.e., some $\mathbf{R} \in \mathcal{R}_m$ is in $\mathcal{R}_m^0$ iff it is transitive in the sense that, whenever $r_{i,j} = r_{i,j} = 1$, then $r_{i,k} = 1$. For any $\mathbf{R} \in \mathcal{R}_m$, let $\Theta_m(\mathbf{R}) \subsetneq \Theta_m$ be the set of all parameters $\boldsymbol{\theta} \in \Theta_m$ which fulfill $\theta_{i,j} \geq 0$ whenever $r_{i,j} = 1$ holds. It is easy to see that $\Theta_m = \bigcup_{\mathbf{R}\in\mathcal{R}_m} \Theta_m(\mathbf{R})$, $\Theta_m(\mathrm{WST}) = \bigcup_{\mathbf{R}\in\mathcal{R}_m^0} \Theta_m(\mathbf{R})$ and $\Theta_m(\neg\mathrm{WST}) = \bigcup_{\mathbf{R}\in\mathcal{R}_m^1} \Theta_m(\mathbf{R})$. Thus, we directly obtain $\tilde{\lambda}_t = \min_{\mathbf{R}\in\mathcal{R}_m^0} \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \Theta_m(\mathbf{R}))$, as well as $\tilde{\mu}_t = \min_{\mathbf{R}\in\mathcal{R}_m^1} \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \Theta_m(\mathbf{R}))$.

Now, let $\mathbf{R} = (r_{i,j})_{1\leq i,j\leq m} \in \mathcal{R}_m$ be fixed for the time being. As the term $\mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \boldsymbol{\theta}) = \sum_{i<j} (\boldsymbol{n}_t)_{i,j}((\boldsymbol{z}_t)_{i,j} - \theta_{i,j})^2$ is monotonically increasing in each of its summands, a minimizer $\hat{\boldsymbol{\theta}}^t(\mathbf{R}) = (\hat{\theta}^t_{i,j}(\mathbf{R}))_{1\leq i,j\leq m} \in \mathrm{argmin}_{\boldsymbol{\theta}\in\Theta_m(\mathbf{R})} \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \boldsymbol{\theta})$ is given by

$$\hat{\theta}^t_{i,j}(\mathbf{R}) = \begin{cases} (\boldsymbol{z}_t)_{i,j}, & \text{if } (r_{i,j} = 1 \text{ and } (\boldsymbol{z}_t)_{i,j} > 0) \text{ or } (r_{i,j} = 0 \text{ and } (\boldsymbol{z}_t)_{i,j} < 0), \\ 0, & \text{otherwise.} \end{cases}$$

Note that the latter can be written in a more compact way as

$$\hat{\theta}^t_{i,j}(\mathbf{R}) = (\boldsymbol{z}_t)_{i,j} \mathbf{1}_{\{(r_{i,j}-1/2)(\boldsymbol{z}_t)_{i,j} > 0\}}\,. \tag{5.35}$$

Thus, to calculate $\tilde{\lambda}_t$ and $\tilde{\mu}_t$, we may at first calculate $\hat{\boldsymbol{\theta}}^t(\mathbf{R}) \in \Theta_m(\mathbf{R})$ for each $\mathbf{R} \in \mathcal{R}_m$ via (5.35) and then obtain

$$\begin{aligned}\tilde{\lambda}_t &= \min\nolimits_{\mathbf{R}\in\mathcal{R}_m^0} \mathrm{d}_{\boldsymbol{n}_t}\left(\boldsymbol{z}_t, \hat{\boldsymbol{\theta}}^t(\mathbf{R})\right), \\ \tilde{\mu}_t &= \min\nolimits_{\mathbf{R}\in\mathcal{R}_m^1} \mathrm{d}_{\boldsymbol{n}_t}\left(\boldsymbol{z}_t, \hat{\boldsymbol{\theta}}^t(\mathbf{R})\right).\end{aligned} \tag{5.36}$$

Let us denote $\{k,l\} = \pi(t+1, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0\leq t'\leq t})$ for the moment and assume w.l.o.g. $k < l$. Then, $(\boldsymbol{n}_{t+1})_{i,j} = (\boldsymbol{n}_t)_{i,j}$ as well as $(\boldsymbol{w}_{t+1})_{i,j} = (\boldsymbol{w}_t)_{i,j}$ for every (i,j) with $i < j$ such that

$\{i,j\} \neq \{k,l\}$ and $(\boldsymbol{n}_{t+1})_{k,l} = (\boldsymbol{n}_t)_{k,l} + 1$ as well as $(\boldsymbol{w}_{t+1})_{k,l} = (\boldsymbol{w}_t)_{k,l} + \mathbf{1}_{\{X_{k,l}^{[t+1]}=1\}}$ hold. This shows

$$\begin{aligned} &\mathrm{d}_{\boldsymbol{n}_{t+1}}\left(\boldsymbol{z}_{t+1}, \hat{\boldsymbol{\theta}}^{t+1}(\mathbf{R})\right) - \mathrm{d}_{\boldsymbol{n}_t}\left(\boldsymbol{z}_t, \hat{\boldsymbol{\theta}}^{t}(\mathbf{R})\right) \\ &= ((\boldsymbol{n}_t)_{k,l}+1)(\boldsymbol{z}_{t+1})_{k,l}^2 \mathbf{1}_{\{(r_{k,l}-1/2)(\boldsymbol{z}_{t+1})_{k,l}\leq 0\}} - (\boldsymbol{n}_t)_{k,l}(\boldsymbol{z}_t)_{k,l}^2 \mathbf{1}_{\{(r_{k,l}-1/2)(\boldsymbol{z}_t)_{k,l}\leq 0\}} . \quad (5.37) \end{aligned}$$

Abbreviating $d(t,\mathbf{R}) \coloneqq \mathrm{d}_{\boldsymbol{n}_t}\big(\boldsymbol{z}_t, \hat{\boldsymbol{\theta}}^t(\mathbf{R})\big)$, the values $\{d(t,\mathbf{R})\}_{\mathbf{R}\in\mathcal{R}_m}$ can thus be updated sequentially, and these values are sufficient to compute the test statistics $\tilde{\lambda}_t$ and $\tilde{\mu}_t$, cf. (5.36). Nevertheless, updating all of these $|\mathcal{R}_m| = 2^{m(m-1)/2}$ values in each time step appears to be infeasible for large m. Fortunately, (5.37) also lets us infer that the calculation of $\tilde{\lambda}_t$ and $\tilde{\mu}_t$ are sometimes much less expensive. To be more specific, if it is known that $\hat{\mathbf{R}} \in \arg\min_{\mathbf{R}\in\mathcal{R}_m^0} d(t,\mathbf{R})$ and $d(t+1,\hat{\mathbf{R}}) < d(t,\hat{\mathbf{R}})$ or $d(t+1,\hat{\mathbf{R}}) = 0$ holds, then $\tilde{\lambda}_{t+1} = d(t+1,\hat{\mathbf{R}})$ is fulfilled, and an analogous result holds for $\tilde{\mu}_t$ (cf. Prop. 5.16 below). For sake of brevity, we will write $\eta_t^0 \coloneqq \tilde{\lambda}_t$ and $\eta_t^1 \coloneqq \tilde{\mu}_t$ in the following proposition.

Proposition 5.16. *Let $b \in \{0,1\}$ be fixed. If $\mathbf{R} = (r_{i,j})_{1\leq i,j\leq m} \in \mathcal{R}_m^b$ is such that $\eta_t^b = d(t,\mathbf{R})$ and $d(t+1,\mathbf{R}) < d(t,\mathbf{R})$ hold, we have $\eta_{t+1}^b = d(t+1,\mathbf{R})$.*

Proof. Suppose $\mathbf{R} = (r_{i,j})_{1\leq i,j\leq m} \in \mathcal{R}_m^b$ fulfills both $\eta_t^b = d(t,\mathbf{R})$ and $d(t+1,\mathbf{R}) < d(t,\mathbf{R})$. Furthermore, fix $\mathbf{R}' = (r'_{i,j})_{1\leq i,j\leq m} \in \mathcal{R}_m^b$ with

$$d(t+1,\mathbf{R}') = \eta_{t+1}^b = \min_{\tilde{\mathbf{R}}\in\mathcal{R}_m^b} d(t+1,\tilde{\mathbf{R}}).$$

We have to show $d(t+1,\mathbf{R}) \leq d(t+1,\mathbf{R}')$. In the case $d(t+1,\mathbf{R}') - d(t,\mathbf{R}') \geq 0$ this follows due to

$$d(t+1,\mathbf{R}) < d(t,\mathbf{R}) \leq d(t,\mathbf{R}') \leq d(t+1,\mathbf{R}').$$

In the remaining case $d(t+1,\mathbf{R}') - d(t,\mathbf{R}') < 0$ a first look at equation (5.37) shows $(\boldsymbol{z}_t)_{k,l} \neq 0$ with $\{k,l\} = \pi(t+1, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0\leq t'\leq t})$. Now, a second look at (5.37) reveals

$$r'_{k,l} = \begin{cases} 1, & \text{if } (\boldsymbol{z}_t)_{k,l} > 0, \\ 0, & \text{if } (\boldsymbol{z}_t)_{k,l} < 0. \end{cases}$$

Since by assumption $d(t+1,\mathbf{R}) - d(t,\mathbf{R}) < 0$ holds, we similarly obtain

$$r_{k,l} = \begin{cases} 1, & \text{if } (\boldsymbol{z}_t)_{k,l} > 0, \\ 0, & \text{if } (\boldsymbol{z}_t)_{k,l} < 0, \end{cases}$$

which shows $r'_{k,l} = r_{k,l}$. Consequently, according to (5.37), we have $d(t+1,\mathbf{R}) - d(t,\mathbf{R}) = d(t+1,\mathbf{R}') - d(t,\mathbf{R}')$ and together with $d(t,\mathbf{R}) \leq d(t,\mathbf{R}')$ this yields $d(t+1,\mathbf{R}) \leq d(t+1,\mathbf{R}')$. □

5.4.6. Asymptotic Behaviour of the LRT Statistics

Next, we discuss the asymptotic behaviour of the test statistics μ_t, λ_t, $\tilde{\mu}_t$ and $\tilde{\lambda}_t$. More precisely, we derive non-trivial upper bounds for asymptotic expressions like $\mathbb{P}_{\boldsymbol{\theta}}(\mu_t > l)$ etc. as $t \to \infty$. This allows us to formulate one-sided hypothesis tests for WST resp. $\neg$WST with asymptotic guarantees on the type I error. Our results generalize to some extent those from Iverson and Falmagne [1985], and we discuss on this on page 170. Moreover, provided π fulfills a rather mild condition, the asymptotics of $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l)$ and $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\lambda}_t > l)$ can be bounded *uniformly* with respect to $\boldsymbol{\theta}$ in a restricted set

$$\Theta_m^v := \{\boldsymbol{\theta} \in \Theta_m \,|\, \forall (i,j) \in (m)_2 : \theta_{i,j} = 0 \text{ or } |\theta_{i,j}| > v\} \subsetneq \Theta_m$$

for some $v \in (0, \pi/2)$. This restriction is related to the low-noise assumption in the sense that in particular $\phi\left(\mathcal{Q}_m^h\right) \subsetneq \Theta_m^{\phi(h)}$ for any $h \in (0, 1/2)$. Note here that $\Theta_m^0 \subsetneq \bigcup_{v>0} \Theta_m^v$ with Θ_m^0 as in (5.19), e.g. we have $(0)_{1\leq i,j\leq m} \in \Theta_m^v \setminus \Theta_m^0$ for any $v \in (0, \pi/2)$.

For $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$ write $\Theta_m(\mathrm{X})^\circ$ for the interior and $\partial\Theta_m(\mathrm{X})$ for the boundary of $\Theta_m(\mathrm{X}) \subsetneq \Theta_m$. Recalling that the considered topology on Θ_m is induced by the standard topology on $\mathbb{R}^{m(m-1)/2}$, $\Theta_m(\mathrm{X})^\circ$ is the set of all $\boldsymbol{\theta} \in \Theta_m(\mathrm{X})$ with $\inf_{\boldsymbol{\theta}' \in \Theta_m(\neg \mathrm{X})} \left|\left|\boldsymbol{\theta} - \boldsymbol{\theta}'\right|\right|_2 > 0$, where $||\cdot||_2$ denotes the standard Euclidean norm on $\mathbb{R}^{m(m-1)/2}$. Since $\Theta_m(\neg\mathrm{WST})^\circ \supsetneq \Theta_m^0(\neg\mathrm{WST})$, the notion of $\Theta_m(\mathrm{X})^\circ$ is not redundant to that of $\Theta_m^0(\mathrm{X})$ from p. 152.

According to Prop. 5.11, the limits of $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l)$ and $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\lambda}_t > l)$ (as $t \to \infty$) are trivial in case $\boldsymbol{\theta} \in \Theta_m^0 \subseteq \Theta_m(\mathrm{WST})^\circ \cup \Theta_m(\neg\mathrm{WST})^\circ$. For $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{WST}) = \partial\Theta_m(\neg\mathrm{WST})$ it may appear reasonable to conjecture this limit to depend on the number $\psi(\boldsymbol{\theta})$ of all $(i,j) \in (m)_2$ with $\theta_{i,j} = 0$: The larger $\psi(\boldsymbol{\theta})$, the more the property $\boldsymbol{\theta} \in \Theta_m(\mathrm{WST})^\circ \cup \Theta_m(\neg\mathrm{WST})^\circ$ appears to be violated. In fact, we derive in the upcoming theorem asymptotic bounds for $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l)$ and $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\lambda}_t > l)$ (as $t \to \infty$), which explicitly depend on $\psi(\boldsymbol{\theta})$. For the uniform convergence result in the theorem we assume that, for any $t' \in \mathbb{N}$, $\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t')$ converges uniformly on $\partial\Theta_m(\mathrm{WST}) \cap \Theta_m^v$ to 1. That this assumption is rather mild can be seen exactly as the mildness of the assumption $\pi \in \Pi_\infty$, which was discussed directly after introducing the term *sampling strategy* in Sec. 1.1. If $\pi \in \Pi$ violates this assumption, the sampling strategy $\hat{\pi} \in \Pi$ defined there behaves in the limit similar to π but at the same time it not only fulfills $\hat{\pi} \in \Pi_\infty$ but also the above uniform convergence assumption. In Thm. 5.17 we bound the limes superior of expressions like $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l)$, since it is – without further assumptions on $\pi \in \Pi_\infty$ – not clear whether these expressions actually converge for any $\boldsymbol{\theta}$ of interest. For example, if π was (for a fixed $\boldsymbol{\theta} \in \Theta_m$) adversarially tailored towards creating a strictly increasing sequence $\{t_n\}_{n\in\mathbb{N}} \in \mathbb{N}$ with $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_{t_{2n-1}} > l) < \varepsilon$ and $\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_{t_{2n}} > l) > \varepsilon'$ for some $0 < \varepsilon < \varepsilon' < 1$ and all $n \in \mathbb{N}$ the limit of $\{\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l)\}_{t\in\mathbb{N}}$ would not exist.

Theorem 5.17. *Suppose* $\pi \in \Pi_\infty$. *For any* $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{WST}) = \partial\Theta_m(\neg\mathrm{WST})$ *we have*

$$\limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l) \leq 2^{-\psi(\boldsymbol{\theta})} \sum_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right), \tag{5.38}$$

$$\limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\tilde{\lambda}_t > l) \leq 2^{-\psi(\boldsymbol{\theta})} \sum_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\binom{m}{2}-2))} > l\right). \tag{5.39}$$

If $v \in (0, \pi/2)$ *and* $\lim_{t\to\infty} \inf_{\boldsymbol{\theta}\in\partial\Theta_m(\mathrm{WST})\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t') = 1$ *for any* $t' \in \mathbb{N}$, *then we* *even have*

$$\limsup_{t\to\infty} \sup_{\boldsymbol{\theta}\in\partial\Theta_m(\neg\mathrm{WST})\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l) - 2^{-\psi(\boldsymbol{\theta})} \sum\nolimits_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right) \leq 0, \tag{5.40}$$

$$\limsup_{t\to\infty} \sup_{\boldsymbol{\theta}\in\partial\Theta_m(\mathrm{WST})\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}(\tilde{\lambda}_t > l) - 2^{-\psi(\boldsymbol{\theta})} \sum\nolimits_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\binom{m}{2}-2))} > l\right) \leq 0. \tag{5.41}$$

According to Lem. 2.9, the right-hand sides of (5.38) and (5.39) are monotonically increasing in $\psi(\boldsymbol{\theta})$, respectively. In this way, $\psi(\boldsymbol{\theta})$ can be seen as a hardness parameter for testing for WST resp. ¬WST based on $\tilde{\lambda}_t$ and $\tilde{\mu}_t$. A detailed proof of Thm. 5.17 is technical and requires some preparations. For this reason, we defer it to Sec. 5.4.7 and restrict ourselves at this point to a proof sketch for the bound on $\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l)$. Before, let us introduce some convenient notation. Recall the definition of $\Theta_m(\mathbf{R})$ from Sec. 5.4.5 and define

$$\Theta_m(\boldsymbol{\theta}) \coloneqq \{\boldsymbol{y} \in \Theta_m \,|\, \theta_{i,j} < 0 \Rightarrow y_{i,j} < 0 \text{ for every distinct } i,j \in [m]\}$$

and $\Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}) \coloneqq \Theta_m(\boldsymbol{\theta}) \cap \Theta_m(\mathbf{R})$ for any $\boldsymbol{\theta} \in \Theta_m$, $\mathbf{R} \in \mathcal{R}_m$. For $\boldsymbol{\theta} \in \Theta_m$ let $\mathcal{R}_m(\boldsymbol{\theta})$ denote the set of all $\mathbf{R} \in \mathcal{R}_m$ with $\boldsymbol{\theta} \in \Theta_m(\mathbf{R})$, i.e.,

$$\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta}) \quad \Leftrightarrow \quad \forall \text{ distinct } i,j \in [m] : (\theta_{i,j} < 0 \Rightarrow r_{i,j} = 0)$$

holds. Then, we have

$$\Theta_m(\boldsymbol{\theta}) = \bigcup\nolimits_{\mathbf{R}\in\mathcal{R}_m} \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}) = \bigcup\nolimits_{\mathbf{R}\in\mathcal{R}_m(\boldsymbol{\theta})} \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}). \tag{5.42}$$

For $\mathbf{R}, \tilde{\mathbf{R}} \in \mathcal{R}_m$ write

$$\Delta(\mathbf{R}, \tilde{\mathbf{R}}) = |\{(i,j) \in (m)_2 | \, r_{i,j} \neq \tilde{r}_{i,j}\}|$$

for the number of pairs $(i,j) \in (m)_2$, on which $\mathbf{R}$ and $\tilde{\mathbf{R}}$ disagree. For every $\mathbf{R} \in \mathcal{R}_m$ and every subset $K \subseteq (m)_2$ of size $|K| = \tau$ there exist exactly $\binom{\tau}{a}$ relations $\tilde{\mathbf{R}} \in \mathcal{R}_m$ with

$$r_{i,j} = \tilde{r}_{i,j} \text{ for every } (i,j) \in (m)_2 \setminus K \text{ and } \Delta(\mathbf{R}, \tilde{\mathbf{R}}) = a.$$

As an example we have for

$$\boldsymbol{\theta} = \begin{pmatrix} - & 0.6 & 0.6 & 0 \\ & - & 0.6 & 0 \\ & & - & 0.6 \\ & & & - \end{pmatrix}$$

that

$$\mathcal{R}_4(\boldsymbol{\theta}) = \left\{ \tilde{\mathbf{R}}_{x,y} \coloneqq \begin{pmatrix} - & 1 & 1 & x \\ & - & 1 & y \\ & & - & 1 \\ & & & - \end{pmatrix} : x, y \in \{0,1\} \right\},$$

$\Delta(\tilde{\mathbf{R}}_{0,0}, \tilde{\mathbf{R}}_{0,1}) = 1 = \Delta(\tilde{\mathbf{R}}_{0,1}, \tilde{\mathbf{R}}_{1,1})$ and $\Delta(\tilde{\mathbf{R}}_{0,0}, \tilde{\mathbf{R}}_{1,1}) = 2$.

Proof Sketch for Thm. 5.17. Let $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{WST}) = \partial\Theta_m(\neg\mathrm{WST})$ and fix a relation $\mathbf{R}_{\neg\mathrm{WST}} \in \mathcal{R}_m(\neg\mathrm{WST})$ with $\boldsymbol{\theta} \in \Theta_m(\mathbf{R}_{\neg\mathrm{WST}})$. By means of Hoeffding's inequality and the convergence $\sqrt{(\boldsymbol{n}_t)_{i,j}}(\boldsymbol{z}_t)_{i,j} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ we can choose $t_0 \in \mathbb{N}$ such that for any $t' \geq t_0$

(i) the approximation $\mathbb{P}_{\boldsymbol{\theta}}\left(\sum_{(i,j)\in K}(\boldsymbol{n}_t)_{i,j}(\boldsymbol{z}_t)_{i,j}^2 > l \middle| \boldsymbol{n}_t \geq t'\right) \approx \mathbb{P}\left(\chi^2_{(|K|)} > l\right)$ is good enough for any $\emptyset \neq K \subseteq (m)_2$ with $\theta_{i,j} = 0$ for all $(i,j) \in (m)_2$,

(ii) the probability $\mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t \notin \Theta_m(\boldsymbol{\theta}) \middle| \boldsymbol{n}_t \geq t'\right)$ is small enough.

In case $\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}\wedge\mathbf{R})$ for some $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$, setting at most $\Delta(\mathbf{R}, \mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta}))\wedge(\lfloor m/2\rfloor+1)$ of its entries (say, those indexed by $K_{\neg\mathrm{WST}} \subsetneq (m)_2$) to 0 results in a point $\boldsymbol{z}'_t \in \Theta_m(\neg\mathrm{WST})$. Then,

$$\tilde{\mu}_t \leq \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \boldsymbol{z}'_t) = \sum\nolimits_{(i,j)\in K_{\neg\mathrm{WST}}} (\boldsymbol{n}_t)_{i,j}(\boldsymbol{z}_t)_{i,j}^2$$

holds and (i) assures us

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}\wedge\mathbf{R}) \wedge \boldsymbol{n}_t \geq t_0\right) \lesssim \mathbb{P}\left(\chi^2_{(\Delta(\mathbf{R},\mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta}))\wedge(\lfloor m/2\rfloor+1))} > l\right).$$

Using that $(\boldsymbol{z}_t)_{i,j}$ is symmetric in case $\theta_{i,j} = 0$, one can show

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}\wedge\mathbf{R}) \,\middle|\, \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t'\right) = 2^{-\psi(\boldsymbol{\theta})}.$$

Since there are exactly $\binom{\psi(\boldsymbol{\theta})}{a}$ possible choices of $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ with $\Delta(\mathbf{R}, \mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta})) = a$, the above approximate inequality allows us to infer

$$\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l | \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2) \lesssim 2^{-\psi(\boldsymbol{\theta})} \sum\nolimits_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right).$$

Taking into account (ii) and the assumption $\pi \in \Pi_\infty$, we conclude for large values of t

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l\right) \\
&\leq \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2\right) + \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t \notin \Theta_m(\boldsymbol{\theta}) \middle| \boldsymbol{n}_t \geq t_2\right) + 2\mathbb{P}_{\boldsymbol{\theta}}\left(\neg(\boldsymbol{n}_t \geq t_2)\right) \\
&\lesssim 2^{-\psi(\boldsymbol{\theta})} \sum\nolimits_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right).
\end{aligned}$$

For the uniform variant, we choose $\mathbf{R}_{\neg\mathrm{WST}} = \mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta})$ and $K_{\neg\mathrm{WST}} = K_{\neg\mathrm{WST}}(\boldsymbol{\theta}, \mathbf{R})$ in an appropriate uniform way. □

For further extending Thm. 5.17, the following proposition will be of use. It can be seen as a uniform variant of results in Prop. 5.11, and its proof is deferred to Sec. 5.4.8

Proposition 5.18. *Let $v \in (0, \pi/2)$ and $\pi \in \Pi_\infty$ be s.t. $\lim_{t\to\infty} \inf_{\boldsymbol{\theta}\in\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{n}_t \geq t'\right) = 1$ for all $t' \in \mathbb{N}$. For any $\varepsilon > 0$ there exists $t_1 \in \mathbb{N}$ with*

$$\sup_{t\geq t_1} \sup_{\boldsymbol{\theta}\in\Theta_m(\mathrm{WST})^\circ\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > 0\right) \leq \varepsilon,\ \sup_{t\geq t_1} \sup_{\boldsymbol{\theta}\in\Theta_m(\neg\mathrm{WST})^\circ\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > 0\right) \leq \varepsilon.$$

By means of Prop. 5.18, we can extend the uniform result from Thm. 5.17 as follows.

Corollary 5.19. *If $\pi \in \Pi_\infty$ and $v \in (0, \pi/2)$ with $\lim_{t\to\infty} \inf_{\boldsymbol{\theta}\in\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t') = 1$ for all $t' \in \mathbb{N}$, then we have with $M := \binom{m}{2}$ that*

$$\limsup_{t\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_m(\neg\mathrm{WST})\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l\right) \leq 2^{-M} \sum_{a=0}^{M} \binom{M}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right),$$

$$\limsup_{t\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_m(\mathrm{WST})\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > l\right) \leq 2^{-M} \sum_{a=0}^{M} \binom{M}{a} \mathbb{P}\left(\chi^2_{(a\wedge(M-2))} > l\right).$$

Proof. For any $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$ we have $\Theta_m(\mathrm{X}) \cap \Theta_m^v \subseteq (\Theta_m(\mathrm{X})^\circ \cap \Theta_m^v) \cup (\partial\Theta_m(\mathrm{X}) \cap \Theta_m^v)$ and trivially also $\psi(\boldsymbol{\theta}) \leq M$ for all $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{X})$. Hence, the statement follows directly from combining Thm. 5.17, Lem. 2.9 and Prop. 5.18. □

So far, we have restricted ourselves to the asymptotics of $\tilde{\mu}_t$ and $\tilde{\lambda}_t$. The following lemma guarantees that, for any $\boldsymbol{\theta} \in \Theta_m$, $\mu_t - \tilde{\mu}_t$ and $\lambda_t - \tilde{\lambda}_t$ converge in probability to 0 as $t \to \infty$. This allows us to transfer the results from Thm. 5.17 and Cor. 5.19 to uniform bounds on the asymptotics of the tails of μ_t and λ_t as $t \to \infty$.

Lemma 5.20. *For every $\varepsilon > 0$ we have*

$$\sup_{\boldsymbol{\theta}\in\Theta_m} \lim_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(|\lambda_t - \tilde{\lambda}_t| > \varepsilon\right) = 0, \quad \sup_{\boldsymbol{\theta}\in\Theta_m} \lim_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(|\mu_t - \tilde{\mu}_t| > \varepsilon\right) = 0.$$

For the sake of readability, the proof of Lem. 5.20 is given in Sec. 5.4.8.

Corollary 5.21. *Suppose $\pi \in \Pi_\infty$ and $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{WST}) = \partial\Theta_m(\neg\mathrm{WST})$. Then,* (5.38) *and* (5.39) *also hold for μ_t and λ_t instead of $\tilde{\mu}_t$ and $\tilde{\lambda}_t$. In particular, if π is such that $\lim_{t\to\infty} \inf_{\boldsymbol{\theta}\in\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t') = 1$ for all $t' \in \mathbb{N}$, we obtain with $M := \binom{m}{2}$ that*

$$\sup_{\boldsymbol{\theta}\in\Theta_m(\neg\mathrm{WST})} \limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(\mu_t > l\right) \leq 2^{-M} \sum_{a=0}^{M} \binom{M}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right),$$

$$\sup_{\boldsymbol{\theta}\in\Theta_m(\mathrm{WST})} \limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(\lambda_t > l\right) \leq 2^{-M} \sum_{a=0}^{M} \binom{M}{a} \mathbb{P}\left(\chi^2_{(a\wedge(M-2))} > l\right).$$

Proof of Cor. 5.21. We only show the results for μ_t, those concerning λ_t can be seen similarly. To see that (5.38) also holds with $\tilde{\mu}_t$ replaced by μ_t, suppose $\boldsymbol{\theta} \in \partial\Theta_m(\neg\mathrm{WST})$ and $\varepsilon > 0$ to be arbitrary but fixed. Using $\psi(\boldsymbol{\theta}) \leq M$, Lem. 5.20 allows us to infer

$$\begin{aligned}\limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(\mu_t > l + \varepsilon\right) &\leq \limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(|\mu_t - \tilde{\mu}_t| > \varepsilon\right) + \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l\right) \\ &= \limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l\right) \\ &\leq 2^{-\psi(\boldsymbol{\theta})} \sum_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right).\end{aligned}$$

Hence, the statement follows via $\varepsilon \searrow 0$. Using that

$$\Theta_m(\neg\mathrm{WST}) \cap \Theta_m^v \subseteq (\Theta_m(\neg\mathrm{WST})^\circ \cap \Theta_m^v) \cup (\partial\Theta_m(\neg\mathrm{WST}) \cap \Theta_m^v)$$

and $\psi(\boldsymbol{\theta}) \leq M$, combining this with Prop. 5.18 yields

$$\sup_{\boldsymbol{\theta}\in\Theta_m(\neg\mathrm{WST})} \limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(\mu_t > l\right) \leq 2^{-M} \sum_{a=0}^{M} \binom{M}{a} \mathbb{P}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right).$$

□

With the help of the obtained bounds we can provide the following asymptotic hypothesis tests for WST and $\neg$WST.

Corollary 5.22 (Asymptotic level-α tests for testing WST resp. $\neg$WST)**.**
Let $\alpha \in (0,1)$ and $\pi \in \Pi_\infty$ be fixed. Choose $l_{\mathrm{WST}}, l_{\neg\mathrm{WST}} > 0$ such that

$$\frac{1}{2^M}\sum_{a=0}^{M}\binom{M}{a}\mathbb{P}\left(\chi^2_{(a\wedge(M-2))} > l_{\mathrm{WST}}\right) = \alpha,\ \frac{1}{2^M}\sum_{a=0}^{M}\binom{M}{a}\mathbb{P}\left(\chi^2_{\left(a\wedge(\lfloor\frac{m}{2}\rfloor+1)\right)} > l_{\neg\mathrm{WST}}\right) = \alpha$$

with $M = \binom{m}{2}$.

- *Let $\mathcal{A}_0(t)$ be the testing algorithm, which chooses its queries until time t according to π and outputs $\neg$WST if $\lambda_t > l_{\mathrm{WST}}$ and WST otherwise. Then, we have for any $\boldsymbol{\theta} \in \Theta_m(\mathrm{WST})$ that $\limsup_{t\to\infty}\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{D}(\mathcal{A}_0(t)) = \neg\mathrm{WST}) \leq \alpha$.*
- *Let $\mathcal{A}_1(t)$ be the testing algorithm, which chooses its queries until time t according to π and outputs WST if $\mu_t > l_{\neg\mathrm{WST}}$ and $\neg$WST otherwise. Then, we have for any $\boldsymbol{\theta} \in \Theta_m(\neg\mathrm{WST})$ that $\limsup_{t\to\infty}\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{D}(\mathcal{A}_1(t)) = \mathrm{WST}) \leq \alpha$.*

The same guarantees are valid when λ_t and μ_t are replaced in the definitions of $\mathcal{A}_0(t)$ and $\mathcal{A}_1(t)$ by $\tilde{\lambda}_t$ and $\tilde{\mu}_t$.

Despite its asymptotic guarantees on the type I error, the tests from Cor. 5.22 do not tell us when to stop sampling and decide to terminate. For this reason, they are apparently less applicable in real-world scenarios than solutions to $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$.

Comparison to Related Work Iverson and Falmagne [1985] have analyzed the convergence of the asymptotic behaviour of the WST test statistics λ_t and $\tilde{\lambda}_t$ under the assumption that at each time t, every pair $\{i,j\}$ has been queried exactly t times. By choosing $\pi = $ RoundRobin and restriction to the time steps $M, 2M, 3M, \ldots$ with $M = \binom{m}{2}$ (instead of $1, 2, 3, \ldots$), this can be seen as a special case of our setting, in which at each time step only one query is made. With this particular choice, the authors have shown that

$$\sup_{\boldsymbol{\theta}\in\Theta_m(\mathrm{WST})}\limsup_{t\to\infty}\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_{tM} > l\right) \leq 2^{-M'}\sum_{a=0}^{M'}\binom{M'}{a}\mathbb{P}\left(\chi^2_{(a)} > l\right) \tag{5.43}$$

with $M' = \binom{m-1}{2}$. Regarding Lem. 2.9, this bound is apparently stronger than ours, which depends on $\binom{m}{2} > \binom{m-1}{2}$. This is due to the benefits this particular choice of π provides for the analysis: With it, we have $(\boldsymbol{n}_{tM})_{i,j} = t$ for all $(i,j) \in (m)_2$ and thus

$$\tilde{\lambda}_{tM} = \mathrm{d}_{\boldsymbol{n}_{tM}}\left(\boldsymbol{z}_{tM}, \Theta_m(\mathrm{WST})\right) = t\min_{\boldsymbol{\theta}\in\Theta_m(\mathrm{WST})}\sum_{(i,j)\in(m)_2}\left((\boldsymbol{z}_{tM})_{i,j} - \theta_{i,j}\right)^2$$

holds, i.e. $\tilde{\lambda}_{tM}/t$ is the *unweighted* Euclidean distance from $\boldsymbol{z}_{tM}$ to $\Theta_m(\mathrm{WST})$. In particular, the standard Euclidean basis is an orthogonal basis w.r.t. the underlying inner product. Regarding that $\partial\Theta_m(\mathrm{WST})$ is the union of appropriate spans of these basis vectors, the orthogonal projections of $\boldsymbol{z}_{tM}$ onto $\Theta_m(\mathrm{WST})$ are thus rather easy to handle, and geometric arguments allow Iverson and Falmagne [1985] to prove (5.43).
In our general case, where $\pi \in \Pi_\infty$ is arbitrary, $\tilde{\lambda}_t$ is a *weighted* Euclidean distance from $\boldsymbol{z}_t$ to $\Theta_m(\mathrm{WST})$, where the weights are $\boldsymbol{n}_t$ and thus dependent on π. Therefore, the geometric arguments from [Iverson and Falmagne, 1985] do not seem to be transferrable to our setting and the obtained bounds are weaker than those for their special case.

5.4.7. Proof of Thm. 5.17

We prepare the proof of Thm. 5.17 with the following technical lemma. It will enable us to choose $\mathbf{R}_{\neg\mathrm{WST}}$ and $K_{\neg\mathrm{WST}}$ from the proof sketch of (5.38) in an appropriate uniform way. For this, let us write $\mathcal{P}((m)_2) := \{A : A \subseteq (m)_2\}$ and define for convenience

$$\Psi(\boldsymbol{\theta}) = \{(i,j) \in (m)_2 \,:\, \theta_{i,j} = 0\}$$

for every $\boldsymbol{\theta} \in \Theta_m$. Then, we have $\psi(\boldsymbol{\theta}) := |\Psi(\boldsymbol{\theta})|$ and regarding the definition of $\mathcal{R}_m(\boldsymbol{\theta})$ also

$$(i,j) \in \Psi(\boldsymbol{\theta}) \quad \Leftrightarrow \quad \exists \mathbf{R}, \tilde{\mathbf{R}} \in \mathcal{R}_m(\boldsymbol{\theta}) \,:\, r_{i,j} \neq \tilde{r}_{i,j}$$

for any $\boldsymbol{\theta} \in \Theta_m$ and $(i,j) \in (m)_2$.

Lemma 5.23. *Given $K \subseteq (m)_2$ and $\boldsymbol{y} \in \Theta_m$, write for convenience $\boldsymbol{y}_{\setminus K}$ for the element $\boldsymbol{y}' = (y'_{i,j})_{1 \le i,j \le m} \in \Theta_m$ defined via*

$$y'_{i,j} := \begin{cases} 0, & \text{if } (i,j) \in K, \\ y_{i,j}, & \text{if } (i,j) \in (m)_2 \setminus K. \end{cases}$$

For $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$ there exist mappings

$$\mathbf{R}_{\mathrm{X}} : \partial\Theta_m(\mathrm{X}) \to \mathcal{R}_m(\mathrm{X}), \quad K_{\mathrm{X}} : \partial\Theta_m(\mathrm{X}) \times \mathcal{R}_m \to \mathcal{P}((m)_2)$$

s.t. for all $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{WST}) = \partial\Theta_m(\neg\mathrm{WST})$ and $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ the following holds:

(i) $|K_{\mathrm{WST}}(\boldsymbol{\theta}, \mathbf{R})| \le (\binom{m}{2} - 2) \wedge \Delta(\mathbf{R}, \mathbf{R}_{\mathrm{WST}}(\boldsymbol{\theta}))$

(ii) $|K_{\neg\mathrm{WST}}(\boldsymbol{\theta}, \mathbf{R})| \le (\lfloor m/2 \rfloor + 1) \wedge \Delta(\mathbf{R}, \mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta}))$

(iii) $K_{\mathrm{X}}(\boldsymbol{\theta}, \mathbf{R}) \subseteq \Psi(\boldsymbol{\theta})$ *and* $\boldsymbol{y}_{\setminus K_{\mathrm{X}}(\boldsymbol{\theta}, \mathbf{R})} \in \Theta_m(\mathrm{X})$ *for any* $\boldsymbol{y} \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R})$ *and* $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$.

The proof of Lem. 5.23 is based on the following two corollaries, which are rather direct consequences of the graph theoretical considerations made in Sec. 3. In their proof, we use the notation $G(\mathbf{R})$ as introduced in Sec. 3.1 and also write $\mathbf{R}(G)$ for the relation $\mathbf{R} \in \mathcal{R}_m$ defined via $r_{i,j} = 1$ iff $(i,j) \in E_G$.

Corollary 5.24. *Let $\mathbf{R} \in \mathcal{R}_m$. Then, for any subset $K \subseteq (m)_2$ of size $|K| > \lfloor m/2 \rfloor$ there exists $\tilde{\mathbf{R}} \in \mathcal{R}_m(\neg\mathrm{WST})$ with*

$$\tilde{r}_{k,l} = r_{k,l} \quad \text{for all } (k,l) \in (m)_2 \setminus K. \tag{5.44}$$

This bound for $|K|$ is optimal in the sense that there exists $\mathbf{R} \in \mathcal{R}_m(\mathrm{WST})$ and a particular choice of K with $|K| = \lfloor \frac{m}{2} \rfloor$ such that every $\tilde{\mathbf{R}} \in \mathcal{R}_m$ fulfilling (5.44) is in $\mathcal{R}_m(\mathrm{WST})$ as well.

Proof of Cor. 5.24. To prove the first statement, suppose $\mathbf{R} \in \mathcal{R}_m$ and $K \subseteq (m)_2$ of size $|K| > \lfloor m/2 \rfloor$ to be fixed. Abbreviate $G := G(\mathbf{R}) \in \overline{\mathcal{G}}_m$, i.e. $(i,j) \in E_G$ iff $r_{i,j} = 1$, and define $G' = ([m], E_{G'}) \in \mathcal{G}_m$ via

$$(i,j) \in E_{G'} \quad \text{iff} \quad (i,j) \in E_G \text{ and } \{(i,j),(j,i)\} \cap K = \emptyset.$$

According to Prop. 3.18, $|E_{G'}| = |E_G| - |K| < \binom{m}{2} - \lfloor m/2 \rfloor$ implies that G' has a supergraph $\tilde{G} \in \overline{\mathcal{G}}_m(\neg\text{acyclic})$. Hence, the statement follows with $\tilde{\mathbf{R}} := \mathbf{R}(\tilde{G})$.

To validate second statement, note that Prop. 3.18 allows us to fix a $G \in \mathcal{G}_m(\text{acyclic})$ with $|E_G| = \binom{m}{2} - \lfloor m/2 \rfloor$. Let $G' \in \overline{\mathcal{G}}_m(\text{acyclic})$ be any extension of G. Now, $\mathbf{R} := \mathbf{R}(G')$ has the desired properties. □

Corollary 5.25. *Let $\mathbf{R} \in \mathcal{R}_m$. Then, for any subset $K \subseteq (m)_2$ of size $|K| > \binom{m}{2} - 3$ there exists a relation $\tilde{\mathbf{R}} \in \mathcal{R}_m(\text{WST})$ that fulfills*

$$\tilde{r}_{k,l} = r_{k,l} \quad \text{for all } (k,l) \in (m)_2 \setminus K. \tag{5.45}$$

This bound for $|K|$ is optimal in the sense that there exists some $\mathbf{R} \in \mathcal{R}_m(\neg\text{WST})$ and a particular choice of K with $|K| = \binom{m}{2} - 3$ such that every $\tilde{\mathbf{R}} \in \mathcal{R}_m$ fulfilling (5.45) is an element of $\mathcal{R}_m(\neg\text{WST})$ as well.

Proof of Cor. 5.25. The first statement follows from the fact that every graph $G \in \mathcal{G}_m$ with at most 2 edges has a supergraph $G' \in \overline{\mathcal{G}}_m(\text{acyclic})$. To see the second statement, choose an arbitrary $\mathbf{R} \in \mathcal{R}_m$ with $r_{1,2} = r_{2,3} = r_{3,1} = 1$ and let $K = (m)_2 \setminus \{(1,2),(2,3),(1,3)\}$. Then, $\mathbf{R} \in \mathcal{R}_m(\neg\text{WST})$ and every relation $\tilde{\mathbf{R}} \in \mathcal{R}_m$ satisfying (5.45) fulfills $\tilde{r}_{1,2} = \tilde{r}_{2,3} = \tilde{r}_{3,1} = 1$ and thus $\tilde{\mathbf{R}} \in \mathcal{R}_m(\neg\text{WST})$. □

Proof of Lem. 5.23. We start with the definition of $\mathbf{R}_{\neg\text{WST}}$ and $K_{\neg\text{WST}}$. Note that we can fix a mapping $\mathbf{R}_{\neg\text{WST}}$ such that $\mathbf{R}_{\neg\text{WST}}(\boldsymbol{\theta}) \in \mathcal{R}_m(\boldsymbol{\theta}) \cap \mathcal{R}_m(\neg\text{WST})$ holds for any $\boldsymbol{\theta} \in \partial\Theta_m(\neg\text{WST})$. For the sake of completeness, define $K_{\neg\text{WST}}(\boldsymbol{\theta}, \mathbf{R}) := \emptyset$ if $\mathbf{R} \notin \mathcal{R}_m(\boldsymbol{\theta})$, and suppose in the following $\boldsymbol{\theta} \in \partial\Theta_m(\neg\text{WST})$ and $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ to be arbitrary but fixed. From $\mathbf{R}, \mathbf{R}_{\neg\text{WST}}(\boldsymbol{\theta}) \in \mathcal{R}_m(\boldsymbol{\theta})$ we infer that

$$K'(\boldsymbol{\theta}, \mathbf{R}) := \{(i,j) \in (m)_2 : r_{i,j} \neq (\mathbf{R}_{\neg\text{WST}}(\boldsymbol{\theta}))_{i,j}\}$$

is a subset of $\Psi(\boldsymbol{\theta})$, i.e., $\Delta(\mathbf{R}, \mathbf{R}_{\neg\text{WST}}(\boldsymbol{\theta})) = |K'(\boldsymbol{\theta}, \mathbf{R})| \leq \psi(\boldsymbol{\theta})$. If $|K'(\boldsymbol{\theta}, \mathbf{R})| > \lfloor m/2 \rfloor + 1$, we can fix a subset $K''(\boldsymbol{\theta}, \mathbf{R}) \subseteq K'(\boldsymbol{\theta}, \mathbf{R})$ of size $\lfloor m/2 \rfloor + 1$ and according to Cor. 5.24 a relation $\tilde{\mathbf{R}} \in \mathcal{R}_m(\neg\text{WST})$ such that[7]

$$\tilde{r}_{i,j} = r_{i,j} \quad \text{for all } (i,j) \in (m)_2 \setminus K''(\boldsymbol{\theta}, \mathbf{R}).$$

Define

$$K_{\neg\text{WST}}(\boldsymbol{\theta}, \mathbf{R}) := \begin{cases} \{(i,j) \in (m)_2 \,:\, r_{i,j} \neq \tilde{r}_{i,j}\}, & \text{if } |K'(\boldsymbol{\theta}, \mathbf{R})| > \lfloor m/2 \rfloor + 1, \\ K'(\boldsymbol{\theta}, \mathbf{R}), & \text{otherwise,} \end{cases}$$

and note $K_{\neg\text{WST}}(\boldsymbol{\theta}, \mathbf{R}) \subseteq K'(\boldsymbol{\theta}, \mathbf{R}) \subseteq \Psi(\boldsymbol{\theta})$. If $\Delta(\mathbf{R}, \mathbf{R}_{\neg\text{WST}}(\boldsymbol{\theta})) = |K'(\boldsymbol{\theta}, \mathbf{R})| > \lfloor m/2 \rfloor + 1$, we have $|K_{\neg\text{WST}}(\boldsymbol{\theta}, \mathbf{R})| \leq |K''(\boldsymbol{\theta}, \mathbf{R})| = \lfloor m/2 \rfloor + 1$, and if $\Delta(\mathbf{R}, \mathbf{R}_{\neg\text{WST}}(\boldsymbol{\theta})) \leq \lfloor m/2 \rfloor + 1$, then $|K_{\neg\text{WST}}(\boldsymbol{\theta}, \mathbf{R})| = |K'(\boldsymbol{\theta}, \mathbf{R})| = \Delta(\mathbf{R}, \mathbf{R}_{\neg\text{WST}}(\boldsymbol{\theta}))$. Hence, $|K_{\neg\text{WST}}(\boldsymbol{\theta}, \mathbf{R})| \leq (\lfloor m/2 \rfloor + 1) \wedge \Delta(\mathbf{R}, \mathbf{R}_{\neg\text{WST}}(\boldsymbol{\theta}))$ holds in any case.

[7] In order enforce uniqueness of the choices of $\tilde{\mathbf{R}}$ and $K''(\boldsymbol{\theta}, \mathbf{R})$ we may equip $\mathcal{R}_m$ resp. $(m)_2$ with a unique ordering $\succ$ and choose $\tilde{\mathbf{R}}$ resp. $K''(\boldsymbol{\theta}, \mathbf{R})$ as the (w.r.t. $\succ$) highest ordered element with the desired property.

Now, suppose $\boldsymbol{y} \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R})$ to be fixed. In case $|K'(\boldsymbol{\theta}, \mathbf{R})| \leq \lfloor m/2 \rfloor + 1$ the point $\boldsymbol{y}'$ is given by

$$y'_{i,j} = \begin{cases} 0, & \text{if } r_{i,j} \neq (\mathbf{R}_{\neg \mathrm{WST}}(\boldsymbol{\theta}))_{i,j}, \\ y_{i,j}, & \text{otherwise} \end{cases}$$

and thus $\boldsymbol{y}' \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}_{\neg \mathrm{WST}}(\boldsymbol{\theta})) \subseteq \Theta_m(\neg \mathrm{WST})$. In the remaining case $|K'(\boldsymbol{\theta}, \mathbf{R})| > \lfloor m/2 \rfloor + 1$ we have

$$y'_{i,j} = \begin{cases} 0, & \text{if } r_{i,j} \neq \tilde{r}_{i,j}, \\ y_{i,j}, & \text{otherwise.} \end{cases}$$

To see $\boldsymbol{y}' \in \Theta_m(\boldsymbol{\theta} \wedge \tilde{\mathbf{R}}) \subseteq \Theta_m(\neg \mathrm{WST})$, let $(i,j) \in (m)_2$ with $r_{i,j} = 0$ be arbitrary. In case $\tilde{r}_{i,j} \neq r_{i,j}$ we have $y'_{i,j} \leq 0$ by definition of $\boldsymbol{y}$. If $\tilde{r}_{i,j} = r_{i,j} = 0$, then $\boldsymbol{y} \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R})$ assures $y'_{i,j} = y_{i,j} \leq 0$. In particular, $\boldsymbol{y}' \in \Theta_m(\neg \mathrm{WST})$ holds in any case.
The mappings $\mathbf{R}_{\mathrm{WST}}$ and K_{WST} can be defined analogously by using Cor. 5.25 instead of Cor. 5.24. □

We proceed with the proof of Thm. 5.17.

Proof of Thm. 5.17. We start this proof with three auxiliary results. Then, we prove (5.40) and (5.41) separately, and finally we show (5.38) and (5.39).

Claim 1: For any $\varepsilon > 0$ there is $t_0(\varepsilon) \in \mathbb{N}$ s.t. for any $\boldsymbol{\theta} \in \partial \Theta_m(\mathrm{WST})$ and any $t'_0 \geq t_0(\varepsilon)$

$$\forall \emptyset \neq K \subseteq \Psi(\boldsymbol{\theta}) : \left| \mathbb{P}_{\boldsymbol{\theta}} \left(\sum\nolimits_{(i,j) \in K} (\boldsymbol{n}_t)_{i,j} (\boldsymbol{z}_t)^2_{i,j} > l \middle| \boldsymbol{n}_t \geq t'_0 \right) - \mathbb{P}\left(\chi^2_{(|K|)} > l \right) \right| \leq \varepsilon.$$

Proof: This follows from the fact that $\pi \in \Pi_\infty$ assures $\sqrt{(\boldsymbol{n}_t)_{i,j}} (\boldsymbol{z}_t)_{i,j} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ as $t \to \infty$ for any $(i,j) \in \Psi(\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \partial \Theta_m(\mathrm{WST})$. ■

Claim 2: For any $\varepsilon > 0$ there exists $t_1(\varepsilon) \in \mathbb{N}$ s.t. for any $\boldsymbol{\theta} \in \Theta^v_m$ and any $t'_1 \geq t_1(\varepsilon)$

$$\mathbb{P}_{\boldsymbol{\theta}} \left(\boldsymbol{z}_t \notin \Theta_m(\boldsymbol{\theta}) \middle| \boldsymbol{n}_t \geq t'_1 \right) \leq \varepsilon.$$

Proof: The transformation ϕ is bijective, its inverse function $\phi^{-1} : [-\pi/2, \pi/2] \to [0,1]$ is given as $\phi^{-1}(x) = \sin^2((2x + \pi)/4)$ and strictly monotonically increasing. In particular, $c := \phi^{-1}(0) - \phi^{-1}(-v)$ is positive. Since $\sum_{s=t}^\infty (e^{-2c})^s \to 0$ as $t \to \infty$, we can choose $t_1 \in \mathbb{N}$ with $\sum_{s=t_1}^\infty (e^{-2c})^s \leq \varepsilon / \binom{m}{2}$. Using again monotonicity of ϕ^{-1}, we obtain by means of Hoeffding's inequality for any $\boldsymbol{\theta} \in \Theta^v_m$ and any $t'_1 \geq t_1$ the estimate

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{\theta}} \left(\boldsymbol{z}_t \notin \Theta_m(\boldsymbol{\theta}) \,\middle|\, \boldsymbol{n}_t \geq t'_1 \right) \\ &= \mathbb{P}_{\boldsymbol{\theta}} \left(\exists \text{ distinct } i,j \in [m] \text{ with } q_{i,j} \leq \phi^{-1}(-v) \,:\, (\boldsymbol{w}_t)_{i,j} / (\boldsymbol{n}_t)_{i,j} \geq \phi^{-1}(0) \,\middle|\, \boldsymbol{n}_t \geq t'_1 \right) \\ &\leq \sum\nolimits_{(i,j) \in (m)_2 : \theta_{i,j} \neq 0} \mathbb{P}_{\boldsymbol{\theta}} \left(|(\boldsymbol{w}_t)_{i,j} / (\boldsymbol{n}_t)_{i,j} - q_{i,j}| \geq c \,\middle|\, \boldsymbol{n}_t \geq t'_1 \right) \\ &\leq \binom{m}{2} \sum\nolimits_{s=t_1}^\infty e^{-2cs} \leq \varepsilon. \end{aligned}$$

■

Claim 3: For any $\boldsymbol{\theta} \in \partial \Theta_m(\mathrm{WST})$, $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ and any $t' \in \mathbb{N}$ we have

$$\mathbb{P}_{\boldsymbol{\theta}} \left(\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}) \,\middle|\, \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t' \right) = 2^{-\psi(\boldsymbol{\theta})}.$$

Proof: Let $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{WST})$, $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ and $t' \in \mathbb{N}$ be fixed and suppose $\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta})$ for the moment. For $\{i,j\} \in [m]_2$ with $(i,j),(j,i) \notin \Psi(\boldsymbol{\theta})$ we have $\theta_{i,j} < 0$ or $\theta_{i,j} > 0$. In the first case, $\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta})$ resp. $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ assure $(\boldsymbol{z}_t)_{i,j} < 0$ resp. $r_{i,j} = 0$, and in the other case an analogous argumentation yields $(\boldsymbol{z}_t)_{i,j} > 0$ and $r_{i,j} = 1$. Consequently, we have in particular

$$\forall \{i,j\} \in [m]_2 \text{ with } (i,j),(j,i) \notin \Psi(\boldsymbol{\theta}) \,:\, r_{i,j} = 1 \;\Rightarrow\; (\boldsymbol{z}_t)_{i,j} \geq 0,$$

and, as $\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta})$ holds by assumption, this implies

$$\begin{aligned}
&\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}) \\
\Leftrightarrow\ &\boldsymbol{z}_t \in \Theta_m(\mathbf{R}) \\
\Leftrightarrow\ &\forall \{i,j\} \in [m]_2 \,:\, r_{i,j} = 1 \;\Rightarrow\; (\boldsymbol{z}_t)_{i,j} \geq 0 \\
\Leftrightarrow\ &\forall \{i,j\} \in [m]_2 \text{ with } (i,j) \in \Psi(\boldsymbol{\theta}) \text{ or } (j,i) \in \Psi(\boldsymbol{\theta}) \,:\, r_{i,j} = 1 \;\Rightarrow\; (\boldsymbol{z}_t)_{i,j} \geq 0.
\end{aligned}$$

In other words, conditioned on $\{\boldsymbol{z}_t \in \Theta_m, \boldsymbol{n}_t \geq t'\}$, $\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R})$ holds iff the $\psi(\boldsymbol{\theta}) = |\Psi(\boldsymbol{\theta})|$ many signs of $\{(\boldsymbol{z}_t)_{i,j}\}_{(i,j)\in\Psi(\boldsymbol{\theta})}$ are consistent with the corresponding entries $\{r_{i,j}\}_{(i,j)\in\Psi(\boldsymbol{\theta})}$ of $\mathbf{R}$. Recall that $(\boldsymbol{z}_t)_{i,j} \sim \phi\left(\frac{1}{(\boldsymbol{n}_t)_{i,j}} \sum_{s=1}^{(\boldsymbol{n}_t)_{i,j}} (\boldsymbol{x}_s)_{i,j}\right)$ for any $(i,j) \in (m)_2$. If $(i,j) \in \Psi(\boldsymbol{\theta})$, then $\theta_{i,j} = 0$ and thus $(\boldsymbol{x}_s)_{i,j} \sim \mathrm{Ber}(1/2)$ for any $s \in \mathbb{N}$. Using that $\phi(1/2 + x) = \phi(1/2 - x)$ we thus obtain $(\boldsymbol{z}_t)_{i,j} \sim -(\boldsymbol{z}_t)_{i,j}$ and the statement follows. ■

Claim 4: (5.40) is fulfilled.
Proof: Let $\varepsilon > 0$ be arbitrary but fixed and $\mathbf{R}_{\neg\mathrm{WST}} : \partial\Theta_m(\neg\mathrm{WST}) \to \mathcal{R}_m(\neg\mathrm{WST})$ and $K_{\neg\mathrm{WST}} : \partial\Theta_m(\neg\mathrm{WST}) \times \mathcal{R}_m \to \mathcal{P}((m)_2)$ be as postulated in Lem. 5.23. Let $t_0 = t_0(\varepsilon)$ resp. $t_1 = t_1(\varepsilon)$ be as in Claims 1 resp. 2 and abbreviate $t_2 \coloneqq \max\{t_0, t_1\}$. Let $\boldsymbol{\theta} \in \partial\Theta_m(\neg\mathrm{WST})$ and $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ be fixed and suppose $\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R})$ and $\boldsymbol{n}_t \geq t_2$. By Lem. 5.23, $\boldsymbol{z}'_t = (\boldsymbol{z}_t)_{\setminus K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})} \in \Theta_m$ defined via

$$(\boldsymbol{z}'_t)_{i,j} \coloneqq \begin{cases} 0, & \text{if } (i,j) \in K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R}), \\ (\boldsymbol{z}_t)_{i,j}, & \text{if } (i,j) \in (m)_2 \setminus K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R}) \end{cases}$$

fulfills $\boldsymbol{z}'_t \in \Theta_m(\neg\mathrm{WST})$. Hence,

$$\begin{aligned}
\tilde{\mu}_t = \min\nolimits_{\tilde{\boldsymbol{\theta}}\in\Theta_m(\neg\mathrm{WST})} \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \tilde{\boldsymbol{\theta}}) &\leq \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \boldsymbol{z}'_t) \\
&= \sum\nolimits_{(i,j)\in K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})} (\boldsymbol{n}_t)_{i,j} \left((\boldsymbol{z}_t)_{i,j} - (\boldsymbol{z}'_t)_{i,j}\right)^2 \\
&= \sum\nolimits_{(i,j)\in K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})} (\boldsymbol{n}_t)_{i,j} (\boldsymbol{z}_t)^2_{i,j}
\end{aligned}$$

and thus

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}) \wedge \boldsymbol{n}_t \geq t_2\right) \\
&\leq \mathbb{P}_{\boldsymbol{\theta}}\left(\sum\nolimits_{(i,j)\in K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})} (\boldsymbol{n}_t)_{i,j} (\boldsymbol{z}_t)^2_{i,j} > l \middle| \boldsymbol{n}_t \geq t_2\right).
\end{aligned}$$

By means of Claim 3 we can thus estimate for each $\boldsymbol{\theta} \in \partial\Theta_m(\neg\mathrm{WST})$

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l | \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2\right) \\
&\leq \sum\nolimits_{\mathbf{R}\in\mathcal{R}_m(\boldsymbol{\theta})} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}) \wedge \boldsymbol{n}_t \geq t_2\right) \\
&\quad \times \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}) \,|\, \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2\right) \\
&\leq 2^{-\psi(\boldsymbol{\theta})} \sum_{a=0}^{\psi(\boldsymbol{\theta})} \sum_{\substack{\mathbf{R}\in\mathcal{R}_m(\boldsymbol{\theta}): \\ \Delta(\mathbf{R},\mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta}))=a}} \mathbb{P}_{\boldsymbol{\theta}}\left(\sum\nolimits_{(i,j)\in K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})} (\boldsymbol{n}_t)_{i,j}(\boldsymbol{z}_t)_{i,j}^2 > l \middle| \boldsymbol{n}_t \geq t_2\right).
\end{aligned} \tag{5.46}$$

For any $\boldsymbol{\theta} \in \partial\Theta_m(\neg\mathrm{WST})$, $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ and $(i,j) \in K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R}) \subseteq \Psi(\boldsymbol{\theta})$ we have $|K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})| \leq \Delta(\mathbf{R},\mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta})) \wedge (\lfloor m/2 \rfloor + 1)$, thus we can estimate with Claim 1

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(\sum\nolimits_{(i,j)\in K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})} (\boldsymbol{n}_t)_{i,j}(\boldsymbol{z}_t)_{i,j}^2 > l \middle| \boldsymbol{n}_t \geq t_2\right) \\
&\leq \mathbb{P}\left(\chi^2_{(|K_{\neg\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})|)} > l\right) + \varepsilon \\
&\leq \mathbb{P}\left(\chi^2_{(\Delta(\mathbf{R},\mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta}))\wedge(\lfloor m/2\rfloor+1))} > l\right) + \varepsilon.
\end{aligned}$$

Using that $\mathcal{R}_m(\boldsymbol{\theta})$ contains exactly $\binom{\psi(\boldsymbol{\theta})}{a}$ relations $\mathbf{R}$ with $\Delta(\mathbf{R},\mathbf{R}_{\neg\mathrm{WST}}(\boldsymbol{\theta})) = a$, we can combine this estimate with (5.46) and obtain

$$\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l | \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2) - \varepsilon \leq 2^{-\psi(\boldsymbol{\theta})} \sum_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}_{\boldsymbol{\theta}}\left(\chi^2_{(a\wedge(\lfloor m/2\rfloor+1))} > l\right) =: f(\boldsymbol{\theta})$$

for all $\boldsymbol{\theta} \in \partial\Theta_m(\neg\mathrm{WST}) \cap \Theta_m^v$. Since Claim 2 ensures

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l\right) \\
&= \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2\right) \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2\right) \\
&\quad + \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t \notin \Theta_m(\boldsymbol{\theta}) \vee \neg(\boldsymbol{n}_t \geq t_2)\right) \\
&\leq \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2\right) + \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t \notin \Theta_m(\boldsymbol{\theta}) \middle| \boldsymbol{n}_t \geq t_2\right) \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{n}_t \geq t_2\right) \\
&\quad + 2\mathbb{P}_{\boldsymbol{\theta}}\left(\neg(\boldsymbol{n}_t \geq t_2)\right) \\
&\leq f(\boldsymbol{\theta}) + \varepsilon + 2(1 - \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{n}_t \geq t_2\right))
\end{aligned}$$

for all $\boldsymbol{\theta} \in \partial\Theta_m(\neg\mathrm{WST}) \cap \Theta_m^v$, we obtain

$$\begin{aligned}
&\limsup\nolimits_{t\to\infty} \sup\nolimits_{\boldsymbol{\theta}\in\partial\Theta_m(\neg\mathrm{WST})\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l\right) - f(\boldsymbol{\theta}) \\
&\leq \varepsilon + 2\left(1 - \liminf\nolimits_{t\to\infty} \inf\nolimits_{\boldsymbol{\theta}\in\partial\Theta_m(\neg\mathrm{WST})\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t_2)\right) \\
&\leq \varepsilon.
\end{aligned}$$

Hence, (5.40) follows due to arbitrariness of ε. ■

Claim 5: (5.41) is fulfilled.
Proof: Suppose again $\varepsilon > 0$ to be arbitrary but fixed, let t_0, t_1, t_2 be as in the proof of Claim 4 and let $\mathbf{R}_{\mathrm{WST}} : \partial\Theta_m(\mathrm{WST}) \to \mathcal{R}_m(\mathrm{WST})$ and $K_{\mathrm{WST}} : \partial\Theta_m(\mathrm{WST}) \times \mathcal{R}_m \to$

$\mathcal{P}((m)_2)$ be as in Lem. 5.23. Let $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{WST})$ and $\mathbf{R} \in \mathcal{R}_m(\boldsymbol{\theta})$ be fixed. Similarly as in the proof of Claim 4 we see that, provided $\boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R})$, $\boldsymbol{z}_t' = (\boldsymbol{z}_t)_{\setminus K_{\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})}$ is an element in $\Theta_m(\mathrm{WST})$ and thus

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta} \wedge \mathbf{R}) \wedge \boldsymbol{n}_t \geq t_2\right)\\ &\leq \mathbb{P}_{\boldsymbol{\theta}}\left(\sum\nolimits_{(i,j)\in K_{\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})} (\boldsymbol{n}_t)_{i,j}(\boldsymbol{z}_t)_{i,j}^2 > l \middle| \boldsymbol{n}_t \geq t_2\right).\end{aligned}$$

Using $K_{\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R}) \subseteq \Psi(\boldsymbol{\theta})$, $|K_{\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})| \leq \Delta(\mathbf{R},\mathbf{R}_{\mathrm{WST}}(\boldsymbol{\theta})) \wedge \left(\binom{m}{2} - 2\right)$ and Claim 1, we obtain analogously as above

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2\right)\\ &\leq 2^{-\psi(\boldsymbol{\theta})} \sum\nolimits_{a=0}^{\psi(\boldsymbol{\theta})} \sum\nolimits_{\substack{\mathbf{R}\in\mathcal{R}_m(\boldsymbol{\theta}):\\ \Delta(\mathbf{R},\mathbf{R}_{\mathrm{WST}}(\boldsymbol{\theta}))=a}} \mathbb{P}_{\boldsymbol{\theta}}\left(\sum\nolimits_{(i,j)\in K_{\mathrm{WST}}(\boldsymbol{\theta},\mathbf{R})} (\boldsymbol{n}_t)_{i,j}(\boldsymbol{z}_t)_{i,j}^2 > l \middle| \boldsymbol{n}_t \geq t_2\right)\\ &\leq 2^{-\psi(\boldsymbol{\theta})} \sum\nolimits_{a=0}^{\psi(\boldsymbol{\theta})} \sum\nolimits_{\substack{\mathbf{R}\in\mathcal{R}_m(\boldsymbol{\theta}):\\ \Delta(\mathbf{R},\mathbf{R}_{\mathrm{WST}}(\boldsymbol{\theta}))=a}} \mathbb{P}\left(\chi^2_{\left(\Delta(\mathbf{R},\mathbf{R}_{\mathrm{WST}}(\boldsymbol{\theta}))\wedge\left(\binom{m}{2}-2\right)\right)} > l\right) + \varepsilon.\end{aligned}$$

By repeating exactly the same argumentation as in the proof of Claim 4 we thus get

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > l\right)\\ &\leq 2^{-\psi(\boldsymbol{\theta})} \sum\nolimits_{a=0}^{\psi(\boldsymbol{\theta})} \binom{\psi(\boldsymbol{\theta})}{a} \mathbb{P}\left(\chi^2_{\left(\Delta(\mathbf{R},\mathbf{R}_{\mathrm{WST}}(\boldsymbol{\theta}))\wedge\left(\binom{m}{2}-2\right)\right)} > l\right) + \varepsilon + 2\left((1 - \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t_2)\right).\end{aligned}$$

Since $\lim_{t\to\infty} \inf_{\boldsymbol{\theta}\in\partial\Theta_m(\mathrm{WST})\cap\Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t_2) = 1$ holds by assumption, (5.41) follows due to arbitrariness of ε. ■

Claim 6: (5.38) and (5.39) are fulfilled.
Proof: In Claim 4 we have seen that for arbitrary but fixed $\varepsilon > 0$ and any $\boldsymbol{\theta} \in \partial\Theta_m(\mathrm{WST})$

$$\mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l | \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2) - \varepsilon \leq f(\boldsymbol{\theta})$$

is fulfilled, where t_2 and $f(\boldsymbol{\theta})$ are as in Claim 4. As $\pi \in \Pi_\infty$, $\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t_2) = 1$ holds. The identity $\Theta_m = \bigcup_{v>0} \Theta_m^v$ allows us to choose $v > 0$ with $\boldsymbol{\theta} \in \Theta_m^v$, and thus $\limsup_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{z}_t \notin \Theta_m(\boldsymbol{\theta})) \leq \varepsilon$ is fulfilled according to Claim 2. Consequently,

$$\begin{aligned}&\limsup\nolimits_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l)\\ &\leq \limsup\nolimits_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\tilde{\mu}_t > l \wedge \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2) + \limsup\nolimits_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{z}_t \notin \Theta_m(\boldsymbol{\theta}))\\ &\leq \limsup\nolimits_{t\to\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\mu}_t > l \middle| \boldsymbol{z}_t \in \Theta_m(\boldsymbol{\theta}) \wedge \boldsymbol{n}_t \geq t_2\right) + \varepsilon\\ &\leq f(\boldsymbol{\theta}) + \varepsilon\end{aligned}$$

holds and (5.38) follows via $\varepsilon \searrow 0$. (5.39) can be seen similarly.

□

5.4.8. Proofs of Proposition 5.18 and Lemma 5.20

Proof of Prop. 5.18. As in the proof of Thm. 5.17 (in the proof of Claim 2) choose $c := \phi^{-1}(0) - \phi^{-1}(-v) > 0$ and a $t_0 \in \mathbb{N}$ with $\sum_{s=t_0}^{\infty} (e^{-2c})^s \leq \frac{1}{2}\varepsilon / \binom{m}{2}$. Let $\boldsymbol{\theta} \in \Theta_m(\mathrm{WST})^\circ \cap \Theta_m^v$

and $t_0' \geq t_0$ be arbitrary but fixed for the moment. Due to $\tilde{\lambda}_t = 0 \Leftrightarrow \boldsymbol{z}_t \in \Theta_m(\text{WST})$ we have

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > 0 \middle| \boldsymbol{n}_t \geq t_0'\right) = \mathbb{P}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t \in \Theta_m(\neg\text{WST})^\circ \middle| \boldsymbol{n}_t \geq t_0'\right)$$

for any $t_0' \geq t_0$. If $\boldsymbol{z}_t \in \Theta_m(\neg\text{WST})^\circ$, then

$$\forall \text{ distinct } i,j \in [m] : (\boldsymbol{z}_t)_{i,j} \geq 0 \;\;\Rightarrow\;\; \theta_{i,j} \geq 0$$

would imply $\boldsymbol{\theta} \in \Theta_m(\neg\text{WST})$ and contradict $\boldsymbol{\theta} \in \Theta_m(\text{WST})^\circ$. Using $\boldsymbol{\theta} \in \Theta_m^v$, monotonicity of ϕ^{-1} and Hoeffding's inequality, we can thus estimate as in the proof of Thm. 5.17 for any $t_0' \geq t_0$

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > 0 \middle| \boldsymbol{n}_t \geq t_0'\right) &\leq \mathbb{P}_{\boldsymbol{\theta}}\left(\exists \text{ distinct } i,j \in [m] : (\boldsymbol{z}_t)_{i,j} \geq 0 \text{ and } \theta_{i,j} \leq -v \middle| \boldsymbol{n}_t \geq t_0'\right) \\
&= \mathbb{P}_{\boldsymbol{\theta}}\left(\exists \text{ distinct } i,j \in [m] : |(\boldsymbol{w}_t)_{i,j}/(\boldsymbol{n}_t)_{i,j} - q_{i,j}| \geq c \middle| \boldsymbol{n}_t \geq t_0'\right) \\
&\leq \binom{m}{2} \sum_{s=t_0}^{\infty} e^{-2cs} \leq \varepsilon/2.
\end{aligned}$$

By assumption on π there exists $t_1 \geq t_0$ with $\inf_{\boldsymbol{\theta} \in \Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{n}_t \geq t_0) \geq 1 - \varepsilon/2$ for all $t \geq t_1$. Combining these estimates yields

$$\begin{aligned}
&\sup_{t \geq t_1} \sup_{\boldsymbol{\theta} \in \Theta_m(\text{WST})^\circ \cap \Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > 0\right) \\
&\leq \sup_{t \geq t_1} \sup_{\boldsymbol{\theta} \in \Theta_m(\text{WST})^\circ \cap \Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\tilde{\lambda}_t > 0 \middle| \boldsymbol{n}_t \geq t_0\right) + \sup_{t \geq t_1} \sup_{\boldsymbol{\theta} \in \Theta_m^v} \mathbb{P}_{\boldsymbol{\theta}}\left(\neg(\boldsymbol{n}_t \geq t_0)\right) \\
&\leq \varepsilon/2 + \varepsilon/2 = \varepsilon.
\end{aligned}$$

The statement for $\tilde{\mu}_t$ can be proven analogously. □

We prepare the proof of Lem. 5.20 with the following result.

Lemma 5.26. *If $\pi \in \Pi_\infty$, then we have for any $\boldsymbol{\theta}' \in \Theta_m$*

$$\mathbb{P}_{\boldsymbol{\theta}'}\left(\lim_{t \to \infty} \sup_{\boldsymbol{\theta} \in \Theta_m} \left|\ln(\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)) - \ln(\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t))\right| = 0\right) = 1.$$

Proof. Suppose $\boldsymbol{\theta}' \in \Theta_m$ as well as a sample $(\boldsymbol{n}_t, \boldsymbol{w}_t)_{t \in \mathbb{N}}$ to be fixed. According to [Bagui and Mehra, 2017, pp. 403ff.] we have for any $(i,j) \in (m)_2$ that

$$\begin{aligned}
C(q_{i,j}, (\boldsymbol{n}_t)_{i,j}) &:= \ln\left(\mathbb{P}(\text{Bin}((\boldsymbol{n}_t)_{i,j}, q_{i,j}) = (\boldsymbol{w}_t)_{i,j})\right) \\
&\quad - \ln\left(\frac{\sqrt{(\boldsymbol{n}_t)_{i,j}}}{\sqrt{2\pi}} \exp\left(-(\boldsymbol{n}_t)_{i,j}\left(\phi\left(\frac{(\boldsymbol{w}_t)_{i,j}}{(\boldsymbol{n}_t)_{i,j}}\right) - \phi(q_{i,j})\right)^2\right)\right)
\end{aligned}$$

fulfills $C(q_{i,j}, (\boldsymbol{n}_t)_{i,j}) \to 0$ as $t \to \infty$ (due to $(\boldsymbol{n}_t)_{i,j} \to \infty$ as $t \to \infty$). Furthermore, it is shown in [Bagui and Mehra, 2017, pp. 404ff.] that C is a polynomial in $q_{i,j}$ and thus we may conclude

$$\sup_{q_{i,j} \in [0,1]} \lim_{t \to \infty} |C(q_{i,j}, (\boldsymbol{n}_t)_{i,j})| \to 0 \quad \text{ as } t \to \infty.$$

Taking a look at the definitions of $\mathcal{L}$ and $\tilde{\mathcal{L}}$ we can infer

$$\begin{aligned}
\sup_{\boldsymbol{\theta} \in \Theta_m} |\ln(\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)) - \ln(\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t))| &= \sup_{\mathbf{Q} \in \mathcal{Q}_m} \prod_{i<j} |C(q_{i,j}, (\boldsymbol{n}_t)_{i,j})| \\
&= \prod_{i<j} \sup_{q_{i,j} \in [0,1]} |C(q_{i,j}, (\boldsymbol{n}_t)_{i,j})| \to 0
\end{aligned}$$

as $t \to \infty$. □

Proof of Lem. 5.20. Nonnegativity of $\mathcal{L}$ and $\tilde{\mathcal{L}}$ imply for $\widehat{\Theta}_m \in \{\Theta_m(\text{WST}), \Theta_m\}$ on the one hand

$$\frac{\sup_{\boldsymbol{\theta}\in\widehat{\Theta}_m} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\sup_{\tilde{\boldsymbol{\theta}}\in\widehat{\Theta}_m} \tilde{\mathcal{L}}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}_t)} \leq \sup_{\boldsymbol{\theta}\in\widehat{\Theta}_m} \frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}$$

and on the other hand

$$\frac{\sup_{\boldsymbol{\theta}\in\widehat{\Theta}_m} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\sup_{\tilde{\boldsymbol{\theta}}\in\widehat{\Theta}_m} \tilde{\mathcal{L}}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}_t)} = \sup_{\boldsymbol{\theta}\in\widehat{\Theta}_m} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t) \inf_{\tilde{\boldsymbol{\theta}}\in\widehat{\Theta}_m} \frac{1}{\tilde{\mathcal{L}}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}_t)} \geq \inf_{\boldsymbol{\theta}\in\widehat{\Theta}_m} \frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}.$$

Thus, monotonicity of the logarithm implies

$$\left|\ln \frac{\sup_{\boldsymbol{\theta}\in\widehat{\Theta}_m} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\sup_{\tilde{\boldsymbol{\theta}}\in\widehat{\Theta}_m} \tilde{\mathcal{L}}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}_t)}\right| \leq \max\left\{\sup_{\boldsymbol{\theta}\in\widehat{\Theta}_m} \left|\ln \frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right|, \inf_{\boldsymbol{\theta}\in\widehat{\Theta}_m} \left|\ln \frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right|\right\}$$

for any $\widehat{\Theta} \in \{\Theta_m(\text{WST}), \Theta\}$. Recalling that

$$\lambda_t = -2\ln\left(\frac{\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\sup_{\boldsymbol{\theta}\in\Theta_m} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right) \quad \text{and} \quad \tilde{\lambda}_t = -2\ln\left(\frac{\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\sup_{\boldsymbol{\theta}\in\Theta_m} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right)$$

we estimate for any $\boldsymbol{\theta}' \in \Theta_m$

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}'}(|\lambda_t - \tilde{\lambda}_t| > \varepsilon) \\
&\quad \leq \mathbb{P}_{\boldsymbol{\theta}'}\left(\left|\ln\left(\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)\right) - \ln\left(\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)\right)\right| > \frac{\varepsilon}{4}\right) \\
&\qquad + \mathbb{P}_{\boldsymbol{\theta}'}\left(\left|\ln\left(\sup_{\boldsymbol{\theta}\in\Theta_m} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)\right) - \ln\left(\sup_{\boldsymbol{\theta}\in\Theta_m} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)\right)\right| > \frac{\varepsilon}{4}\right). \qquad (5.47)
\end{aligned}$$

The first summand fulfills

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}'}\left(\left|\ln\left(\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)\right) - \ln\left(\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)\right)\right| > \frac{\varepsilon}{4}\right) \\
&\quad = \mathbb{P}_{\boldsymbol{\theta}'}\left(\left|\ln\left(\frac{\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right)\right| > \frac{\varepsilon}{4}\right) \\
&\quad \leq \mathbb{P}_{\boldsymbol{\theta}'}\left(\max\left\{\sup_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \left|\ln \frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right|, \inf_{\boldsymbol{\theta}\in\Theta_m(\text{WST})} \left|\ln \frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right|\right\} > \frac{\varepsilon}{4}\right) \\
&\quad \longrightarrow 0
\end{aligned}$$

as $t \to \infty$, wherein the convergence follows from Lem. 5.26. Similarly, we see

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}'}\left(\left|\ln\left(\sup_{\boldsymbol{\theta}\in\Theta_m} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)\right) - \ln\left(\sup_{\boldsymbol{\theta}\in\Theta_m} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)\right)\right| > \frac{\varepsilon}{4}\right) \\
&\quad \leq \mathbb{P}_{\boldsymbol{\theta}'}\left(\max\left\{\sup_{\boldsymbol{\theta}\in\Theta_m} \left|\ln \frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right|, \inf_{\boldsymbol{\theta}\in\Theta_m} \left|\ln \frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}\right|\right\} > \frac{\varepsilon}{4}\right) \\
&\quad \longrightarrow 0
\end{aligned}$$

as $t \to \infty$. Combining these convergence results with (5.47) completes the proof of the first statement. The second one can be seen analogously. □

5.4.9. Notes on Other Types of Stochastic Transitivity

Now, let us briefly discuss the possibility as well as the difficulties arising by applying the LRT approach for testing other types of stochastic transitivity than WST. Let us focus on the case of strong stochastic transitivity (SST) first, and let π be again some fixed but arbitrary sampling strategy in Π_∞. Writing $\boldsymbol{\theta} = \phi(\mathbf{Q})$ as above, the relation $\mathbf{Q} \in \mathcal{Q}_m$ is SST iff

$$\theta_{i,j}, \theta_{j,k} \leq 0 \quad \Rightarrow \quad \theta_{i,k} \leq \min\{\theta_{i,j}, \theta_{j,k}\} \tag{5.48}$$

is fulfilled for every distinct $i, j, k \in [m]$. Writing $\Theta_m(\mathrm{SST}) := \overline{\phi(\mathcal{Q}_m(\mathrm{SST}))}$ and further $\Theta_m(\neg\mathrm{SST}) := \overline{\phi(\mathcal{Q}_m(\neg\mathrm{SST}))} = \Theta_m \setminus \Theta_m(\mathrm{SST})$, we might define the test statistics

$$\tilde{\lambda}_t^{\mathrm{SST}} := -2 \ln \frac{\max_{\boldsymbol{\theta} \in \Theta_m(\mathrm{SST})} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\max_{\boldsymbol{\theta} \in \Theta_m} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}, \qquad \tilde{\mu}_t^{\mathrm{SST}} := -2 \ln \frac{\max_{\boldsymbol{\theta} \in \Theta_m(\neg\mathrm{SST})} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}{\max_{\boldsymbol{\theta} \in \Theta} \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{z}_t)}$$

with $\tilde{\mathcal{L}}$ defined as in Sec. 5.4.1. With these, we may obtain analogous results as in Thm. 5.13 for the case of SST. But as (5.48) might already indicate, the geometry of $\partial\Theta_m(\mathrm{SST})$ is much more complicated than that of $\partial\Theta_m(\mathrm{WST})$. In fact, $\partial\Theta_m(\mathrm{WST})$ is equal to $[-\frac{\pi}{2}, \frac{\pi}{2}]^{\binom{m}{2}} \cap A$, where $A \subsetneq \mathbb{R}^{m(m-1)/2}$ is a union of hyperplanes of the form $\mathrm{span}\{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_l}\}$ with $\mathbf{e}_1, \ldots, \mathbf{e}_{m(m-1)/2}$ being the standard basis of $\mathbb{R}^{m(m-1)/2}$. Thus, the projection of $\boldsymbol{z}_t \in \Theta_m$ onto $\partial\Theta_m(\mathrm{WST})$, which equals $\hat{\boldsymbol{\theta}}$, is realized by simply replacing some of its entries (say those in $I \subseteq (m)_2$) by 0 and maintaining the others; we can infer

$$\tilde{\lambda}_t = \mathrm{d}_{\boldsymbol{n}_t}(\boldsymbol{z}_t, \Theta_m(\mathrm{WST})) = \sum\nolimits_{i<j} (\boldsymbol{n}_t)_{i,j}((\boldsymbol{z}_t)_{i,j} - \hat{\theta}_{i,j})^2 = \sum\nolimits_{(i,j)\in I} (\boldsymbol{n}_t)_{i,j} (\boldsymbol{z}_t)_{i,j}^2.$$

In contrast to this, $\partial\Theta_m(\mathrm{SST})$ has geometrically a much more complicated form than $\partial\Theta_m(\mathrm{WST})$ and consequently projections of $\boldsymbol{z} \in \Theta_m$ onto $\partial\Theta_m(\mathrm{SST})$ do not permit a straight-forward closed-form computation formula. Hence, the calculation as well as the analysis of sequential update formulas for $\tilde{\lambda}_t^{\mathrm{SST}}$ is much more cumbersome than for $\tilde{\lambda}_t$. Since the corresponding analogous definitions of MST, λST and νRST in terms of the parameter $\boldsymbol{\theta} = \phi(\mathbf{Q})$ resemble (5.48), similar feasibility issues arise for these types of transitivity as well.

5.5. A Reduction to Pure Exploration Multi-Armed Bandits

Similar as $\mathcal{P}_{\mathrm{CWt}}^{m,h,\gamma}$, $\mathcal{P}_{\mathrm{CWc}}^{m,h,\gamma}$ and $\mathcal{P}_{\mathrm{CWv}}^{m,h,\gamma}$, the problem $\mathcal{P}_{\mathrm{WST}}^{m,h,\gamma}$ can be reduced to the pure exploration multi-armed bandits scenario with multiple correct answers as presented by Degenne and Koolen [2019], and in fact, we obtain similar results as for the CW-related problems. In this section, we discuss the corresponding guarantees obtained for WST testing and compare them with our previously stated solutions to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$.

For this purpose, let $\Delta_{(m)_2}$, $\Delta_{(m)_2}^{\varepsilon}$, $D(\mathbf{v}, \mathbf{Q}, \mathbf{Q}')$ and $D(\mathbf{v}, \mathbf{Q}, \mathcal{Q}'_m)$ be defined as in Sec. 4.5.1 and define, for $h \in [0, 1/2)$ and $\mathbf{Q} \in \mathcal{Q}_m^h$, the complexity term $D_{\mathrm{WST}}^{m,h}(\mathbf{Q})$ via

$$D_{\mathrm{WST}}^{m,h}(\mathbf{Q}) \coloneqq \begin{cases} \sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v},\mathbf{Q},\mathcal{Q}_m^h(\neg\mathrm{WST})), & \text{if } \mathbf{Q}\in\mathcal{Q}_m^h(\mathrm{WST}), \\ \sup_{\mathbf{v}\in\Delta_{(m)_2}} D(\mathbf{v},\mathbf{Q},\mathcal{Q}_m^h(\mathrm{WST})), & \text{if } \mathbf{Q}\in\mathcal{Q}_m^h(\neg\mathrm{WST}). \end{cases}$$

$$= \begin{cases} \sup_{\mathbf{v}\in\Delta_{(m)_2}} \inf_{\mathbf{Q}'\in\mathcal{Q}_m^h(\neg\mathrm{WST})} \sum_{(i,j)\in(m)_2} v_{i,j}\mathrm{kl}\left(q_{i,j},q'_{i,j}\right), & \text{if } \mathbf{Q}\in\mathcal{Q}_m^h(\mathrm{WST}), \\ \sup_{\mathbf{v}\in\Delta_{(m)_2}} \inf_{\mathbf{Q}'\in\mathcal{Q}_m^h(\mathrm{WST})} \sum_{(i,j)\in(m)_2} v_{i,j}\mathrm{kl}\left(q_{i,j},q'_{i,j}\right), & \text{if } \mathbf{Q}\in\mathcal{Q}_m^h(\neg\mathrm{WST}). \end{cases}$$

In our setting, the Sticky Track-and-Stop algorithm from [Degenne and Koolen, 2019] can be stated as Alg. 23. Before discussing the theoretical guarantees and sample

Algorithm 23 : Sticky Track-and-Stop for WST testing

Input: $\gamma\in(0,1)$, $h\in[0,1/2)$, a sequence $(\varepsilon_t)_{t\in\mathbb{N}}$, functions $t\mapsto f(t)$ and $(t,\gamma)\mapsto\beta(t,\gamma)$
Initialization: $t\leftarrow 1$, $\hat{\boldsymbol{Q}}_0\leftarrow(0)_{1\le i<j\le m}$, $\boldsymbol{n}_0\leftarrow 0$.

while True **do**
  Let $\mathcal{C}_t\leftarrow\{\mathbf{Q}'\in\mathcal{Q}_m^h : D(\boldsymbol{n}_{t-1}/(t-1),\hat{\mathbf{Q}}_{t-1},\mathbf{Q}')\le\ln(f(t-1))\}$
  Compute $I_t=\{\mathrm{X}\in\{\mathrm{WST},\neg\mathrm{WST}\}\,|\,\exists\mathbf{Q}'\in\mathcal{Q}_m^h(\mathrm{X})\cap\mathcal{C}_t\}$
  Choose an element X from I_t, prefer WST over ¬WST
  Compute that weight $\mathbf{v}_t\in\Delta_{(m)_2}$, which maximizes $D(\mathbf{v},\hat{\mathbf{Q}}_{t-1},\mathcal{Q}_m^h(\neg\mathrm{X}))$
  Compute the projection $\mathbf{v}_t^{\varepsilon_t}$ of $\mathbf{v}_t$ onto $\Delta_{(m)_2}^{\varepsilon_t}$
  Pull $(i,j)=\mathrm{argmin}_{(i',j')\in(m)_2}(\boldsymbol{n}_t)_{i',j'}-\sum_{s=1}^{t}(\mathbf{v}_s^{\varepsilon_s})_{i',j'}$, observe $X_{i,j}\sim\mathrm{Ber}(q_{i,j})$
  Update $\boldsymbol{w}_t$ via $(\boldsymbol{w}_t)_{k,l}\leftarrow(\boldsymbol{w}_{t-1})_{k,l}+\mathbf{1}_{\{\{k,l\}=\{i,j\}\text{ and }X_{k,l}=1\}}$ $\forall 1\le k,l\le m$
  Update $\boldsymbol{n}_t$ via $(\boldsymbol{n}_t)_{k,l}\leftarrow(\boldsymbol{n}_{t-1})_{k,l}+\mathbf{1}_{\{\{k,l\}=\{i,j\}\}}$ $\forall 1\le k,l\le m$
  Update $\hat{\boldsymbol{Q}}_t\leftarrow\frac{\boldsymbol{w}_t}{\boldsymbol{n}_t}$.
  Let $\mathcal{D}_t\leftarrow\{\mathbf{Q}'\in\mathcal{Q}_m^h : D(\boldsymbol{n}_t/t,\hat{\mathbf{Q}}_t,\mathbf{Q}')\le\beta(t,\gamma)\}$
  if $\exists\mathrm{X}\in\{\mathrm{WST},\neg\mathrm{WST}\}$ with $\mathcal{D}_t\cap\mathcal{Q}_m^h(\neg\mathrm{X})=\emptyset$ **then**
    return X
  Update $t\leftarrow t+1$

complexity of Alg. 23, note that running this algorithm in practice requires – similarly as its CW-related counterparts from Sec. 4.5 – a large computational complexity: For fixed $\mathbf{v}\in\Delta_{(m)_2}$, calculating $D(\mathbf{v},\mathbf{Q},\mathcal{Q}_m^h(\mathrm{X}))$ is computationally costly for X = WST resp. X = ¬WST, because $\mathcal{Q}_m^h(\mathrm{WST})$ resp. $\mathcal{Q}_m^h(\neg\mathrm{WST})$ is a union of $m!$ resp. $2^{\binom{m}{2}}-m!$ hypercubes (one for for each permutation of m elements resp. one for each non-transitive deterministic reciprocal relation with m alternatives) and in particular non-convex. Hence, estimating $D(\mathbf{v},\mathbf{Q},\mathcal{Q}_m^h(\mathrm{X}))$ via a grid search with $L(m)$ relations per hypercube results in computational costs of $L(m)m!$ for X = WST and $L(m)(2^{\binom{m}{2}}-m!)$ for X = ¬WST. These considerations suggest that any iteration step of Alg. 23 is computationally costly, and running the algorithm until termination is seemingly infeasible in practice. Therefore, we focus only on the theoretical analysis of it and do not incorporate it in our simulation study in Sec. 5.6.

From [Degenne and Koolen, 2019] we obtain the following results regarding the asymptotics (w.r.t. γ) of the expected sample complexities of solutions to $\mathcal{P}_{\mathrm{WST}}^{m,h,\gamma}$.

Theorem 5.27. *Let $h\in[0,1/2)$ and $m\in\mathbb{N}_{\ge 3}$ be fixed. If $\mathcal{A}(\gamma)$ solves $\mathcal{P}_{\mathrm{WST}}^{m,h,\gamma}$ for any*

$\gamma > 0$, *then*

$$\liminf_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} \geq \frac{1}{D^{m,h}_{\mathrm{WST}}(\mathbf{Q})}.$$

for any $\mathbf{Q} \in \mathcal{Q}^h_m$.

Proof. Confer Thm. 1 in [Degenne and Koolen, 2019]. □

We obtain the following analogon to the second statement in Prop. 4.16.

Theorem 5.28. *Let* $h \in [0, 1/2)$ *and* $m \in \mathbb{N}_{\geq 3}$ *be fixed. Choose* $C > 0$ *and* $\varepsilon_t, f(t)$ *and* $\beta(\gamma, t)$ *for any* $t \in \mathbb{N}$ *as in (ii) of Prop. 4.16. Write* $\mathcal{A}(\gamma)$ *for Alg. 23 called with parameters* $\gamma, h, (\varepsilon_t)_t, f$ *and* β. *Then,* $\mathcal{A}(\gamma)$ *solves* $\mathcal{P}^{m,h,\gamma}_{\mathrm{WST}}$ *and fulfills*

$$\lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}} = \frac{1}{D^{m,h}_{\mathrm{WST}}(\mathbf{Q})}$$

for any $\mathbf{Q} \in \mathcal{Q}^h_m$.

Proof. The statement corresponds to Thm. 11 in [Degenne and Koolen, 2019]. For the choice of ε_t confer p. 7 in [Garivier and Kaufmann, 2016], and for $f(t)$ resp. $\beta(t, \gamma)$ see Lem. 14 on p. 9 resp. Thm. 10 on p. 8 in [Degenne and Koolen, 2019]. □

The next definition will be helpful for analyzing $1/D^{m,h}_{\mathrm{WST}}(\mathbf{Q})$.

Definition 5.29. *For* $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$ *and* $\mathbf{Q} \in \mathcal{Q}^0_m(\mathrm{X})$ *write* $\rho(\mathbf{Q})$ *for the minimum number of entries* $q_{i,j}$, $(i,j) \in (m)_2$, *which have to be modified via* $q_{i,j} \rightsquigarrow 1 - q_{i,j}$ *such that the resulting relation* $\mathbf{Q}'$ *is in* $\mathcal{Q}^0_m(\neg\mathrm{X})$, *i.e., formally* $\rho(\mathbf{Q})$ *is given as*

$$\min\{k \in \mathbb{N} \,|\, \exists \text{ distinct } (i_1, j_1), \ldots, (i_k, j_k) \text{ s.t. } \mathbf{Q}((i_1, j_1), \ldots, (i_k, j_k)) \in \mathcal{Q}^h_m(\neg\mathrm{X})\},$$

where $\mathbf{Q}((i_1, j_1), \ldots, (i_k, j_k)) =: \mathbf{Q}' = (q'_{i,j})_{1 \leq i,j \leq m}$ *is defined for all* $(i,j) \in (m)_2$ *via*

$$q'_{i,j} = \begin{cases} 1 - q_{i,j}, & \text{if } (i,j) \in \{(i_1, j_1), \ldots, (i_k, j_k)\}, \\ q_{i,j}, & \text{otherwise.} \end{cases}$$

As the term $\rho(\mathbf{Q})$ will play a crucial role in our lower bounds on $D^{m,h}_{\mathrm{WST}}(\mathbf{Q})$, let us briefly discuss some of its properties. It is straight-forward to show that $\rho(\mathbf{Q}) = 1$ for any $\mathbf{Q} \in \mathcal{Q}^0_m(\mathrm{WST})$, for the sake of completeness we prove this in Lem. 5.30 below. In case $\mathbf{Q} \in \mathcal{Q}^0_m(\neg\mathrm{WST})$, $\rho(\mathbf{Q})$ may take any value in $\{1, \ldots, f(m)\}$ where $f(m) := \max_{\mathbf{Q} \in \mathcal{Q}^0_m(\neg\mathrm{WST})} \rho(\mathbf{Q})$. Moreover, in this case, $\rho(\mathbf{Q})$ is the minimum number of edges in $G(\mathbf{Q})$, which have to be flipped to obtain an acyclic tournament, or in other words, a tournament containing no cycles. This value is also known as the *Slater index* of $G(\mathbf{Q})$ [Slater, 1961, Bermond, 1972]. It is a well known fact (see [Bermond, 1972] and references therein for the proofs) that

$$\left\lfloor \frac{m}{3} \left\lfloor \frac{m-1}{2} \right\rfloor \right\rfloor \leq f(m) \leq \left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{m-3}{2} \right\rfloor$$

for all $m \in \mathbb{N}$ and also that

$$\forall \varepsilon > 0 \, \exists m' \in \mathbb{N} \, \forall m \geq m' : f(m) > \binom{m}{2}(1 - \varepsilon).$$

Lemma 5.30. *For any* $\mathbf{Q} \in \mathcal{Q}_m^0(\mathrm{WST})$ *we have* $\rho(\mathbf{Q}) = 1$.

Proof. Let $\mathbf{Q} \in \mathcal{Q}_m^0(\mathrm{WST})$ be fixed. According to Lem. 3.3, there exists a permutation σ on $[m]$ s.t. $q_{\sigma(i),\sigma(i+1)} > 1/2$ holds for all $i \in [m-1]$. By replacing $\mathbf{Q}$ with $(q_{\sigma(i),\sigma(j)})_{1 \le i,j \le m}$ we may suppose w.l.o.g. $\sigma = \mathrm{id}$ in the following, i.e., $q_{i,i+1} > 1/2$ for any $i \in [m-1]$. As $\mathbf{Q}$ is weakly stochastic transitive, $q_{1,2}, q_{2,3} > 1/2$ lets us infer $q_{1,3} > 1/2$. Thus, $\mathbf{Q}' \coloneqq \mathbf{Q}((1,3))$ fulfills $q'_{1,2}, q'_{2,3}, q'_{3,1} > 1/2$, which shows $\mathbf{Q}((1,3)) \in \mathcal{Q}_m^0(\neg\mathrm{WST})$. We infer $\rho(\mathbf{Q}) \le 1$. Moreover, $\rho(\mathbf{Q}) \ge 1$ holds trivially. □

For proving lower and upper bounds on $1/D^{m,h}_{\mathrm{WST}}(\mathbf{Q})$, the following lemma will be of use.

Lemma 5.31. *For every* $m \in \mathbb{N}_{\ge 4}$ *and* $\mathbf{v} = (v_{i,j})_{1 \le i < j \le m} \in \Delta_{(m)_2}$ *there exists some* $(i,j), (j,k) \in (m)_2$ *such that* $v_{i,j}, v_{j,k} \le 24/\binom{m}{2}$.

Proof. At first, let

$$I \coloneqq \left\{(i,j,k) \in [m]^3 \,:\, 1 \le i < j < k \le m\right\}.$$

Note that $|I| = \binom{m}{3} = \frac{1}{6}m(m-1)(m-2) \ge \frac{1}{24}m^3$ holds due to $m \ge 4$. Moreover, $\mathbf{v} \in \Delta_{(m)_2}$ implies

$$\sum\nolimits_{(i,j,k)\in I} v_{i,j} + v_{j,k} \le 2 \sum\nolimits_{i \in [m]} \sum\nolimits_{(j,k) \in (m)_2} v_{j,k} \le 2m.$$

Combining this with $|I| \ge \frac{1}{24}m^3$ and $\binom{m}{2} \le \frac{m^2}{2}$, we conclude that there exists some $(i,j,k) \in I$ with $\max\{v_{i,j}, v_{j,k}\} \le v_{i,j} + v_{j,k} \le \frac{48}{m^2} \le 24/\binom{m}{2}$. □

To establish bounds on the term $1/D^{m,h}_{\mathrm{WST}}(\mathbf{Q})$, we make again use of Lem. 4.13.

Lemma 5.32. *Let* $m \in \mathbb{N}_{\ge 4}$ *and* $h \in [0, 1/2)$ *be fixed.*

(i) For any $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{WST})$ *we have*

$$D^{m,h}_{\mathrm{WST}}(\mathbf{Q}) \le \frac{48 d_h(\mathbf{Q})}{\binom{m}{2}}$$

with $d_h(\mathbf{Q}) \coloneqq \max_{(i,j)\in(m)_2} \max\{\mathrm{kl}\,(q_{i,j}, 1/2+h), \mathrm{kl}\,(q_{i,j}, 1/2-h)\}$.

(ii) For any $\tilde{h} \in [h, 1/2) \setminus \{0\}$ *and any* $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}$ *we have*

$$D^{m,h}_{\mathrm{WST}}(\mathbf{Q}) \ge \frac{\rho(\mathbf{Q})\mathrm{kl}\left(1/2+h, 1/2-\tilde{h}\right)}{\binom{m}{2}} \ge \frac{2\rho(\mathbf{Q})(h+\tilde{h})^2}{\binom{m}{2}}.$$

Proof. (i) Let $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{WST})$ be fixed, and suppose $\mathbf{v} \in \Delta_{(m)_2}$ to be fixed for the moment. According to Lem. 5.31 there exist $(i,j), (j,k) \in (m)_2$ such that $v_{i,j}, v_{j,k} \le 24/\binom{m}{2}$. Now, define $\mathbf{Q}' \in \mathcal{Q}_m^h$ via

$$q'_{r,s} \coloneqq \begin{cases} 1/2 - (h+\delta), & \text{if } (r,s) \in \{(i,j),(j,k)\} \text{ and } q_{i,k} > 1/2, \\ 1/2 + (h+\delta), & \text{if } (r,s) \in \{(i,j),(j,k)\} \text{ and } q_{i,k} < 1/2, \\ q_{r,s}, & \text{otherwise} \end{cases}$$

for each $(r,s) \in (m)_2$ and some small $\delta \in (0, 1/2-h)$. The definition of $\mathbf{Q}'$ assures that

$$\begin{aligned}\sum\nolimits_{(r,s)\in(m)_2} v_{r,s}\mathrm{kl}\left(q_{r,s}, q'_{r,s}\right) &\le v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right) + v_{j,k}\mathrm{kl}\left(q_{j,k}, q'_{j,k}\right)\\ &\le \frac{48}{\binom{m}{2}} \max\nolimits_{(r,s)\in(m)_2} \max\left\{\mathrm{kl}\left(q_{r,s}, 1/2+h+\delta\right), \mathrm{kl}\left(q_{r,s}, 1/2-h-\delta\right)\right\}.\end{aligned}$$

As either $q'_{i,j}, q'_{j,k}, q'_{k,i} > 1/2$ or $q'_{i,j}, q'_{j,k}, q'_{k,i} < 1/2$ holds, we have $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\mathrm{WST})$, which allows us to infer via taking the limit $\delta \searrow 0$ that

$$\inf\nolimits_{\mathbf{Q}'\in\mathcal{Q}_m^h(\neg\mathrm{WST})} \sum\nolimits_{(i,j)\in(m)_2} v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right) \le \frac{48 d_h(\mathbf{Q})}{\binom{m}{2}}.$$

As this estimate is obtained for any $\mathbf{v} \in \Delta_{(m)_2}$, we can conclude that

$$D_{\mathrm{WST}}^{m,h}(\mathbf{Q}) = \sup\nolimits_{\mathbf{v}\in\Delta_{(m)_2}} \inf\nolimits_{\mathbf{Q}\in\mathcal{Q}_m^h(\neg\mathrm{WST})} \sum\nolimits_{(i,j)\in(m)_2} v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right) \le \frac{48 d_h(\mathbf{Q})}{\binom{m}{2}}.$$

(ii) Let $\tilde{h} \in [h, 1/2)\setminus\{0\}$ and $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}$ be fixed and let $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$ be such that $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}(\mathrm{X})$. For any $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\mathrm{X})$ there exist distinct $(i_1,j_1),\ldots,(i_{\rho(\mathbf{Q})}, j_{\rho(\mathbf{Q})}) \in (m)_2$ such that $(q_{i_l,j_l} - 1/2)(q'_{i_l,j_l} - 1/2) < 0$ for every $1 \le l \le \rho(\mathbf{Q})$. Due to $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}$ and $\mathbf{Q}' \in \mathcal{Q}_m^h$ we thus obtain

$$\sum\nolimits_{(i,j)\in(m)_2} \mathrm{kl}\left(q_{i,j}, q'_{i,j}\right) \ge \sum\nolimits_{l=1}^{\rho(\mathbf{Q})} \mathrm{kl}\left(q_{i_l,j_l}, q'_{i_l,j_l}\right) \ge \rho(\mathbf{Q})\mathrm{kl}\left(1/2+\tilde{h}, 1/2-h\right).$$

As $\mathbf{Q}' \in \mathcal{Q}_m^h(\neg\mathrm{X})$ was arbitrary, choosing $\mathbf{v} = (1/\binom{m}{2})_{(i,j)\in(m)_2}$ lets us infer

$$\begin{aligned}D_{\mathrm{WST}}^{m,h}(\mathbf{Q}) &= \sup\nolimits_{\mathbf{v}\in\Delta_{(m)_2}} \inf\nolimits_{\mathbf{Q}'\in\mathcal{Q}_m^h(\neg\mathrm{X})} \sum\nolimits_{(i,j)\in(m)_2} v_{i,j}\mathrm{kl}\left(q_{i,j}, q'_{i,j}\right)\\ &\ge \frac{\rho(\mathbf{Q})}{\binom{m}{2}}\mathrm{kl}\left(1/2+\tilde{h}, 1/2-h\right) \ge \frac{2\rho(\mathbf{Q})}{\binom{m}{2}}(h+\tilde{h})^2,\end{aligned}$$

where we have used Lem. 4.13 in the last step.

□

Note that the bounds on $D_{\mathrm{WST}}^{m,h}(\mathbf{Q})$ above are instance-wise, since they depend on $d_h(\mathbf{Q})$ resp. $\rho(\mathbf{Q})$. For our analysis of the worst-case value of $1/D_{\mathrm{WST}}^{m,h}(\mathbf{Q})$, we estimate $d_h(\mathbf{Q})$ according to Lem. 4.13 and use the simple bound $\rho(\mathbf{Q}) \ge 1$.

Corollary 5.33. *Let $m \in \mathbb{N}_{\ge 4}$ be fixed.*

(i) Under the assumptions of Thm. 5.27 we get in case $h \in (0, 1/2)$

$$\sup_{\mathbf{Q}\in\mathcal{Q}_m^h(\mathrm{WST})} \liminf\nolimits_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln\frac{1}{\gamma}} \ge \sup_{\mathbf{Q}\in\mathcal{Q}_m^h(\mathrm{WST})} \frac{1}{D_{\mathrm{WST}}^{m,h}(\mathbf{Q})} \ge \frac{1/4-h^2}{192h^2}\binom{m}{2}$$

and in case $h = 0$ for each $\tilde{h} \in (0, 1/2)$

$$\sup_{\mathbf{Q}\in\mathcal{Q}_m^{\tilde{h}}(\mathrm{WST})} \liminf\nolimits_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln\frac{1}{\gamma}} \ge \sup_{\mathbf{Q}\in\mathcal{Q}_m^{\tilde{h}}(\mathrm{WST})} \frac{1}{D_{\mathrm{WST}}^{m,0}(\mathbf{Q})} \ge \frac{1}{192\tilde{h}^2}\binom{m}{2}.$$

(ii) *Under the assumptions of Thm. 5.28 we obtain in case* $h \in (0, 1/2)$

$$\sup_{\mathbf{Q} \in \mathcal{Q}_m^h} \lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln \frac{1}{\gamma}} \leq \sup_{\mathbf{Q} \in \mathcal{Q}_m^h} \frac{1}{D_{\mathrm{WST}}^{m,h}(\mathbf{Q})} \leq \frac{1}{8h^2}\binom{m}{2}$$

and in case $h = 0$

$$\sup_{\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}} \lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}(\gamma)}\right]}{\ln \frac{1}{\gamma}} \leq \sup_{\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}} \frac{1}{D_{\mathrm{WST}}^{m,0}(\mathbf{Q})} \leq \frac{1}{2\tilde{h}^2}\binom{m}{2}$$

for any $\tilde{h} \in (0, 1/2)$.

Similarly as Thm. 4.3, Part (ii) of Cor. 5.33 may at first appear contradictory to Part (b) of Thm. 5.5, which involves an additional $\ln \ln \frac{1}{h}$-factor. But actually, there is no contradiction, since Cor. 5.33 only yields an upper bound on the worst-case of the *asymptotic* of $\frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln \frac{1}{\gamma}}$ as $\gamma \searrow 0$, whereas the lower bound from Thm. 5.5 holds for any fixed γ.

Proof of Cor. 5.33. (i) At first, let us consider the case $h > 0$. For $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{WST})$ with $q_{i,j} \in \{1/2 \pm (h + \delta)\}$ for all $(i,j) \in (m)_2$ and an arbitrarily small $\delta \in (0, 1/2 - h)$ we have

$$\begin{aligned} d_h(\mathbf{Q}) &= \max_{(i,j) \in (m)_2} \max\left\{\mathrm{kl}\left(q_{i,j}, 1/2 + h\right), \mathrm{kl}\left(q_{i,j}, 1/2 - h\right)\right\} \\ &= \mathrm{kl}\left(1/2 + (h + \delta), 1/2 - h\right) \leq \frac{4(h + \delta/2)^2}{1/4 - h^2}, \end{aligned}$$

where we have used Lem. 4.13 in the last step. Thus, the statement follows from Thm. 5.27 and part (i) of Lem. 5.32 via taking the limit $\delta \searrow 0$.

Now, suppose $h = 0$. Let $\tilde{h} \in (0, 1/2)$ and $\mathbf{Q} \in \mathcal{Q}_m^{\tilde{h}}$ with entries $q_{i,j} \in \{1/2 \pm (\tilde{h} + \delta)\}$ for all $(i,j) \in (m)_2$ and some $\delta \in (0, 1/2 - \tilde{h})$ be fixed. As

$$d_0(\mathbf{Q}) = \max_{(i,j) \in (m)_2} \mathrm{kl}\left(q_{i,j}, 1/2\right) = \mathrm{kl}\Big(1/2 + (\tilde{h} + \delta), 1/2\Big) \leq 4(\tilde{h} + \delta)^2$$

is assured by Lem. 4.13, the statement follows again from Thm. 5.27 and part (i) of Lem. 5.32 by taking the limit $\delta \searrow 0$.

(ii) As $\rho(\mathbf{Q}) \geq 1$ for every $\mathbf{Q} \in \mathcal{Q}_m^h$ is guaranteed by Lem. 5.30, the statement follows from Thm. 5.28 and Part (ii) of Lem. 5.32.

□

5.6. Empirical Evaluation

In this section, we compare the WST testing procedures from Theorems 5.6 and 5.8. Since the solution obtained by Degenne and Koolen [2019] appears infeasible in practice as indicated in Sec. 5.5, we do not consider it in our experiments.

For the sake of convenience, we restrict ourselves to the symmetric case $\alpha = \beta = 0.05 =: \gamma$ throughout this section. Given a sampling strategy $\pi \in \Pi_\infty$ and $h \in (0, 1/2)$, we abbreviate $\gamma' := \gamma/\binom{m}{2}$ and $\gamma'' := \gamma/(\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor)$ and write

	WST			¬WST		
	$\mathcal{A}_{\text{naive}}$	$\mathcal{A}^{\text{SPRT}}_{\text{improved}}$	$\mathcal{A}^{\text{PPR}}_{\text{improved}}$	$\mathcal{A}_{\text{naive}}$	$\mathcal{A}^{\text{SPRT}}_{\text{improved}}$	$\mathcal{A}^{\text{PPR}}_{\text{improved}}$
$m=4$	5346 (299.8)	4081 (254.7)	**2062** (238.1)	5075 (284.0)	3022 (205.3)	**1558** (266.1)
$m=5$	12579 (651.9)	10502 (645.2)	**7393** (899.8)	13789 (624.3)	4371 (348.1)	**1776** (336.8)
$m=6$	21014 (1036.5)	17614 (958.3)	**14019** (1413.5)	20374 (921.3)	4903 (235.6)	**1216** (139.6)
$m=7$	33922 (1605.2)	31064 (1616.5)	**26686** (2407.4)	35261 (1535.9)	6203 (342.1)	**1549** (272.8)
$m=8$	53275 (2152.6)	48502 (1965.7)	**42545** (2891.6)	55728 (1910.8)	7067 (191.8)	**1039** (724.7)

Table 5.1.: Comparison of $\mathcal{A}_{\text{naive}}$, $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ on $\mathcal{Q}^{0.05}_m(\text{WST})$ and $\mathcal{Q}^{0.05}_m(\neg\text{WST})$.

- $\mathcal{A}_{\text{naive}}(h,\pi)$ for Alg. 20 called with the parameters m, π and $\mathcal{A}_{\text{Coin}} = \mathcal{A}^{\text{SPRT}}_{\text{Coin}}(h,\gamma')$,
- $\mathcal{A}^{\text{SPRT}}_{\text{improved}}(h,\pi)$ for Alg. 21 called with parameters m, π and $\mathcal{A}_{\text{Coin}} = \mathcal{A}^{\text{SPRT}}_{\text{Coin}}(h,\gamma'')$,
- $\mathcal{A}^{\text{PPR}}_{\text{improved}}(\pi)$ for Alg. 21 called with m, π and $\mathcal{A}_{\text{Coin}} = \mathcal{A}_{\text{PPR-Ber}}(\gamma'')$.

According to Thm. 5.8, $\mathcal{A}_{\text{naive}}(h,\pi)$ and $\mathcal{A}^{\text{SPRT}}_{\text{improved}}(h,\pi)$ solve $\mathcal{P}^{m,h,\gamma}_{\text{WST}}$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}(\pi)$ solves $\mathcal{P}^{m,0,\gamma}_{\text{WST}}$. Here, we have chosen $\mathcal{A}_{\text{Coin}}$ due to its optimal behavior w.r.t. the expected runtime on some instances as stated in Prop. 2.17, and PPR-Bernoulli due to its good empirical performance observed in Sec. 2.4.1.

The Passive Scenario

For the sake of simplicity, we restrict ourselves in the passive testing scenario to that sampling strategy $\pi = \pi_{\text{Random}} \in \Pi_\infty$, which chooses its queries at each time step uniformly at random from $[m]_2$, fix $h = 0.01$ in the following and simply write $\mathcal{A}_{\text{naive}}$, $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ for $\mathcal{A}_{\text{naive}}(h,\pi)$, $\mathcal{A}^{\text{SPRT}}_{\text{improved}}(h,\pi)$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}(\pi)$, respectively. Note here that $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ does not need and does not obtain h as a parameter.

In the first experiment, we investigate the termination time of $\mathcal{A}_{\text{naive}}$, $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ for preference relations in $\mathcal{Q}^{0.05}_m(\text{WST})$ or $\mathcal{Q}^{0.05}_m(\neg\text{WST})$. To this end, we sample $\mathbf{Q}$ uniformly at random from $\mathcal{Q}^{0.05}_m(\text{WST})$ (resp. $\mathcal{Q}^{0.05}_m(\neg\text{WST})$), run the test algorithms until termination, respectively, and repeat this process for 100 times. When started with such $\mathbf{Q}$, all algorithms observe the same duel chosen by π in each time step, as well as the same outcome of the duel. As stated in Thm. 5.8, $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ may thus terminate earlier than $\mathcal{A}_{\text{naive}}$ in any case.

Table 5.1 reports the obtained average termination times (and the corresponding standard error in brackets) for varying values of m. The results reveal that $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ throughout need significantly less samples for checking WST than $\mathcal{A}_{\text{naive}}$, and the effect is strongest if $\mathbf{Q}$ is not WST and m is large. If the underlying preference relation is not WST, the termination time of $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ resp. $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ is mostly decreasing resp. slightly increasing with the number of available arms, while the termination time of $\mathcal{A}_{\text{naive}}$, on the other side, increases rapidly with the number of arms. Moreover, the three testing algorithms did not make any error in deciding whether WST holds or not for the underlying preference relation $\mathbf{Q}$, i.e., the observed accuracy of all testing algorithms was 100% throughout. Last but not least, it is worth mentioning that $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ (as well as $\mathcal{A}_{\text{naive}}$) terminates for each problem scenario much earlier than the derived worst-case upper bound $\frac{1}{2h}\left\lceil\frac{\ln((1-\gamma'')/\gamma'')}{\ln((1/2+h)/(1/2-h))}\right\rceil(1-2\gamma'')\binom{m}{2}$, which is $\geq 4370\binom{m}{2}$ for any $m\geq 3$ (cf. Thm. 5.9). Moreover, $\mathcal{A}^{\mathrm{PPR}}_{\mathrm{improved}}$ apparently even outperforms $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$ in any case .

Next, we analyze the impact of the degree of violation of WST within a preference relation $\mathbf{Q}$ – measured by the number of cycles[8] in the identifying tournament $G(\mathbf{Q})$ – on the sample complexities of $\mathcal{A}_{\mathrm{naive}}$, $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$ and $\mathcal{A}^{\mathrm{PPR}}_{\mathrm{improved}}$, respectively. For this purpose, we choose $\mathbf{Q}_1 := (x)_{1\leq i<j\leq 6} \in \mathcal{Q}_6$ and $\mathbf{Q}_2$,$\mathbf{Q}_3$ and $\mathbf{Q}_4$ as

$$\begin{pmatrix} - & x & y & x & x & x \\ & - & x & x & x & x \\ & & - & x & x & x \\ & & & - & x & x \\ & & & & - & x \\ & & & & & - \end{pmatrix}, \begin{pmatrix} - & x & y & x & y & x \\ & - & x & y & x & x \\ & & - & x & x & x \\ & & & - & x & x \\ & & & & - & x \\ & & & & & - \end{pmatrix} \text{ and } \begin{pmatrix} - & x & y & x & y & x \\ & - & x & y & x & x \\ & & - & x & x & y \\ & & & - & x & x \\ & & & & - & x \\ & & & & & - \end{pmatrix},$$

respectively, where $x := 0.6$ and $y := 0.4$. Table 5.2 shows the number of cycles in $G(\mathbf{Q}_i)$ together with the average runtimes (as well as the empirical standard errors in brackets) of the three testing procedures, if started with $\mathbf{Q}_i$, over 100 runs. We also added the average elapsed time T_{elapsed} (in seconds) per run as an indicator of the computational costs of the algorithms.

These results support the following conclusions: Firstly, the larger the number of cycles in

		$\mathcal{A}_{\mathrm{naive}}$		$\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$		$\mathcal{A}^{\mathrm{PPR}}_{\mathrm{improved}}$	
i	# cycles in $G(\mathbf{Q}_i)$	$T^{\mathcal{A}}$	T_{elapsed}	$T^{\mathcal{A}}$	T_{elapsed}	$T^{\mathcal{A}}$	T_{elapsed}
1	0	25919 (332.3)	0.60	25639 (340.8)	2.16	**18811** (560.8)	2.64
2	1	25170 (296.4)	0.58	10609 (187.4)	0.44	**6963** (302.1)	0.83
3	9	25599 (366.1)	0.60	8988 (110.3)	0.31	**4518** (155.3)	0.50
4	28	26014 (355.7)	0.60	9063 (110.7)	0.31	**4716** (121.7)	0.53

Table 5.2.: Comparison of $\mathcal{A}_{\mathrm{naive}}$, $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$ and $\mathcal{A}^{\mathrm{PPR}}_{\mathrm{improved}}$ on $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_3$ and $\mathbf{Q}_4$.

the identifying tournament $G(\mathbf{Q}_i)$ of the underlying preference relation $\mathbf{Q}_i$ (i.e., the more severe the WST property is violated), the lower the sample complexity of $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$ and $\mathcal{A}^{\mathrm{PPR}}_{\mathrm{improved}}$ is on average. Secondly, the latter effect reveals an "elbow" dependency in the sense that the decrease of the termination time is rapidly declining with the number of cycles, with the strongest decline if at least one cycle is present. Thirdly, $\mathcal{A}_{\mathrm{naive}}$ does not seem to benefit from stronger violations of WST and in fact does not exploit structural properties of the current estimated preference relation for an early termination such as $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$ and $\mathcal{A}^{\mathrm{PPR}}_{\mathrm{improved}}$ do. Finally, the results for $\mathbf{Q}_1$ with regard to the averaged elapsed time demonstrate that checking acyclicity in extension of the internal graph maintained by $\mathcal{A}_{\mathrm{naive}}$ (i.e., line 7 in Alg. 21) increases the computational cost per iteration step by a factor of $\approx \frac{2.16}{25639}\frac{25919}{0.6} \approx 3.64$. However, the superiority of $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$ and $\mathcal{A}^{\mathrm{PPR}}_{\mathrm{improved}}$ over $\mathcal{A}_{\mathrm{naive}}$ in terms of sample complexity is so strong, that they outperform $\mathcal{A}_{\mathrm{naive}}$ even with regard to computational costs on $\mathbf{Q}_3$ and $\mathbf{Q}_4$, and for $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$ this holds also on $\mathbf{Q}_2$.

In summary, the experiments empirically confirm our theoretical results on the superiority of the enhanced testing algorithm $\mathcal{A}^{\mathrm{SPRT}}_{\mathrm{improved}}$ compared to $\mathcal{A}_{\mathrm{naive}}$.

[8] Recall that, according to our definition above, any cycle is of the form $i_1 \to \dots \to i_k \to i_1$, where $i_1, \dots, i_k$ are *distinct*.

The Active Scenario

For arbitrary but fixed $h \in (0, 1/2)$ write $\mathcal{A}_{\text{naive}}$ for $\mathcal{A}_{\text{naive}}(h, \pi_{\text{Random}})$ with $\pi_{\text{Random}} \in \Pi_\infty$ as before, i.e., π_{Random} chooses its queries at each time step uniformly at random from $[m]_2$. Recall the definition of γ'' from the beginning of this section and let π^*_{SPRT} be the improved sampling strategy π^*_{WST} from Sec. 5.3.2 with[9] $\mathcal{A}_{\text{Coin}} = \mathcal{A}^{\text{SPRT}}_{\text{Coin}}(h, \gamma'')$, and similarly write π^*_{PPR} for that version of π^*_{WST} with $\mathcal{A}_{\text{Coin}} = \mathcal{A}_{\text{PPR-Ber}}(\gamma'')$. With this, let us abbreviate $\mathcal{A}^{\text{SPRT}}_{\text{improved}} := \mathcal{A}^{\text{SPRT}}_{\text{improved}}(h, \pi^*_{\text{SPRT}})$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}} := \mathcal{A}^{\text{PPR}}_{\text{improved}}(\pi^*_{\text{PPR}})$. By Thm. 5.8 and Thm. 5.9, any of $\mathcal{A}_{\text{naive}}$, $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ solves $\mathcal{P}^{m,h,\gamma}_{\text{WST}}$. We ran these three algorithms for 100 runs each on $\mathbf{Q}_1$ and $\mathbf{Q}_3$ from above and repeated this experiment for different values of h. Apart from the fact that $\mathcal{A}_{\text{naive}}$ resp. $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ achieved an accuracy of 0.99 resp. 0.97 on $\mathbf{Q}_1$ for $h = 0.1$, all algorithms achieved throughout an accuracy of 1.00. Hence, we report in Table 5.3 only the observed estimated sample complexity together with the corresponding standard error in brackets.

As already indicated by Thm. 5.8 and Thm. 5.9, $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ outperforms $\mathcal{A}_{\text{naive}}$ in any case. Since $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ does not depend on h, its sample complexity on a fixed instance is constant in h, whereas that of $\mathcal{A}_{\text{naive}}$ and $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ are decreasing in h, and this is consistent with the observations on the underlying algorithm $\mathcal{A}^{\text{SPRT}}_{\text{Coin}}$ made in Sec. 2.2. As a result, we see that $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ outperforms $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ for large values of h, whereas $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ is better than $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ when h is small.

$\mathbf{Q}$	h	$\mathcal{A}_{\text{naive}}$	$\mathcal{A}^{\text{PPR}}_{\text{improved}}$	$\mathcal{A}^{\text{SPRT}}_{\text{improved}}$
$\mathbf{Q}_1$	0.10	3575 (108.1)	5769 (84.6)	**1097** (14.7)
$\mathbf{Q}_1$	0.05	6339 (173.4)	5769 (84.6)	**2150** (22.4)
$\mathbf{Q}_1$	0.02	13503 (269.1)	5769 (84.6)	**5321** (352.6)
$\mathbf{Q}_1$	0.01	25705 (363.5)	**5769** (84.6)	10457 (51.8)
$\mathbf{Q}_3$	0.10	3575 (108.1)	4564 (91.2)	**829** (15.3)
$\mathbf{Q}_3$	0.05	6340 (173.4)	4564 (91.2)	**1826** (29.7)
$\mathbf{Q}_3$	0.02	13505 (269.1)	**4564** (91.2)	4819 (47.4)
$\mathbf{Q}_3$	0.01	25705 (363.5)	**4564** (91.2)	9699 (58.7)

Table 5.3.: Comparison of $\mathcal{A}_{\text{naive}}$, $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ on $\mathbf{Q}_1$ and $\mathbf{Q}_3$.

In Sec. 5.4 we have developed with Alg. 22 yet another solution to $\mathcal{P}^{m,0,\gamma}_{\text{WST}}$. If initiated with $\pi = $ RoundRobin as sampling strategy, an analysis of its expected sample complexity resulted in a worst-case sample complexity bound for instances in $\mathcal{Q}^h_m$ that is of order $\mathcal{O}\left(\frac{m^{2\kappa}}{h^4} \ln \frac{1}{\gamma}\right)$ (cf. Thm. 5.15) and thus with regard to Thm. 5.9 far from optimal (as a solution to $\mathcal{P}^{m,h,\gamma}_{\text{WST}}$). As we will see in the following, $\mathcal{A} :=$Alg. 22 (with $\pi = $ RoundRobin) does not only seem to be suboptimal to $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ and $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ from a theoretical perspective but also with regard to its performance in practice.

If we choose $m = 6$, $h = 0.1$, $\alpha = \beta = 0.05$ and $c = \frac{1}{2} - \frac{h}{2} = 0.45$ as suggested by

[9] Recall that π^*_{WST} depends via $\hat{E}_t$ on $\mathcal{A}_{\text{Coin}}$.

Thm. 5.15, $\mathcal{A}$ has to query – regardless of the choice of $\kappa > 1$ – each pair in $[m]_2$ at least

$$\tilde{n} > \frac{64}{(1-2c)^4}\left(\ln\left(\frac{6\binom{m}{2}}{\alpha \wedge \beta}\right) + 1\right) > 5437146$$

times before it terminates, i.e., for any $\mathbf{Q} \in \mathcal{Q}_m$, $\mathcal{A}$ would fulfill $T^{\mathcal{A}} \geq 5437146 \cdot \binom{6}{2} = 81557190$ a.s. w.r.t. $\mathbb{P}_{\mathbf{Q}}$. As a result, the observed average sample complexity for $\mathbf{Q}_1$ and $\mathbf{Q}_3$ for $\mathcal{A}$ would be by far larger than those of $\mathcal{A}_{\text{naive}}$, $\mathcal{A}^{\text{PPR}}_{\text{improved}}$ and $\mathcal{A}^{\text{SPRT}}_{\text{improved}}$ shown in Table 5.3.

Nevertheless, we want to validate that $\mathcal{A}$ fulfills its desired guarantees. For this reason, we executed it for 100 times for $m = 4$, $h = 0.1$, $\alpha = \beta = 0.05$, $c = \frac{1}{2} - \frac{h}{2} = 0.45$ and $\kappa = 2$ on the two instances

$$\mathbf{Q} = \begin{pmatrix} - & 0.65 & 0.65 & 0.65 \\ & - & 0.65 & 0.65 \\ & & - & 0.65 \\ & & & - \end{pmatrix} \quad \text{and} \quad \mathbf{Q}' = \begin{pmatrix} - & 0.65 & 0.35 & 0.65 \\ & - & 0.65 & 0.65 \\ & & - & 0.65 \\ & & & - \end{pmatrix}$$

from $\mathcal{Q}_4^h$. On all of these 100 repetitions, $\mathcal{A}$ correctly classified $\mathbf{Q}$ as WST and $\mathbf{Q}'$ as $\neg$WST and terminated after exactly 32944326 queries, i.e., after having queried each pair in $[4]_2$ exactly $\tilde{n} = 5490721$ times. Of course, this experimental evaluation of $\mathcal{A}$ is far from extensive, but as the stated results already indicate a huge suboptimality of $\mathcal{A}$ as a solution to $\mathcal{P}^{m,0,\alpha,\beta}_{\text{WST}}$, we kept it to a minimum and did not investigate $\mathcal{A}$ further.

5.7. Discussion and Related Work

The literature on testing transitivity conditions is primarily rooted in the social sciences, psychology, and economics, with a special focus on experimental studies for real data. The works by McNamara and Diwadkar [1997] and Waite [2001] carry out multiple binomial tests to test weak stochastic transitivity (WST) of preferences in different field studies, while Cavagnaro and Davis-Stober [2014] suggest the use of Bayes factors for the testing of stochastic transitivity. A rigorous mathematical treatment of testing for WST is provided by Iverson and Falmagne [1985]. In particular, they derive an asymptotic likelihood ratio test (LRT) for WST and apply their method to empirical data. Ballinger and Wilcox [1997] empirically carry out LRTs for strong stochastic transitivity conditions. Yet, all these works are settled in classical hypothesis testing, assuming all the data to be available beforehand. In contrast to these works, we are interested in testing the validity of stochastic transitivity assumptions in the dueling bandits scenario; thus, we consider hypothesis testing in an online manner, assuming that data arrives sequentially and test decisions should be taken as quickly as possible, while maintaining a predefined level of confidence.

In dueling bandits, testing for stochastic transitivity is of particular interest as this type of transitivity plays a crucial role for several dueling bandits algorithms: Yue et al. [2012] resp. Yue and Joachims [2011] considered regret minimization under strong stochastic transitivity (SST) resp. ν-relaxed stochastic transitivity (νRST), Mohajer et al. [2017] investigated the best-arm identification problem as well as the (top-k-)ranking of arms under WST, whereas Agarwal et al. [2022] resp. [Falahatgar et al., 2017a,b, 2018] analyzed top-k identification under SST resp. the impact of various transitivity assumptions on these latter two goals in an online (ε, γ)-PAC setting. In these works, the transitivity assumption is explicitly required for the theoretical guarantees, while other approaches assume transitivity properties in a more indirect way, for instance by considering probabilistic models for the feedback process. This includes the Plackett-Luce model [Luce, 1959, Plackett, 1975] resp. Bradley-Terry model [Bradley and Terry, 1952] considered by Szörényi et al. [2015] resp. Maystre and Grossglauser [2017], as well as the Mallows model [Mallows, 1957] studied by Busa-Fekete et al. [2014a]; in the notation of Sec. 2.5.1, we have $\mathcal{Q}_m(\mathrm{PL}) \cup \mathcal{Q}_m(\mathrm{Mal}) \subseteq \mathcal{Q}_m(\mathrm{SST})$ [Haddenhorst et al., 2020]. Finally, transitivity assumptions were also analyzed in batch learning scenarios, for example to estimate the underlying pairwise preference relation [Shah et al., 2016], or for the purpose of rank aggregation [Korba et al., 2017].

In this chapter, we analyzed XST testing on $\mathcal{Q}_m^h$ for α, β from several perspectives. We saw that testing for XST $\neq$ WST is impossible to some extent, and, similarly as those from Sec. 2.5.1, these results should be used with caution: They do not show impossibility per se but only with respect to finite worst-case expected sample complexity on $\mathcal{Q}_m^h$, and testing might still be possible in a different manner. For example, we suppose that SST testing may actually be doable via the general Sticky-Track-and-Stop procedure from Degenne and Koolen [2019], but as the geometry of $\mathcal{Q}_m(\mathrm{SST})$ is far more complex than that of $\mathcal{Q}_m(\mathrm{WST})$, this approach appears even less computationally feasible for SST testing than for WST testing. Moreover, one could try to approach SST testing via corresponding analoga of the ideas for STI testing that we gave in Sec. 2.6.

Note that we actually provided two approaches for showing these impossibility results: Cor. 5.1 was a rather direct consequence of Cor. 2.48 and thus relies on the change-of-measure argument from Sec. 2.5, whereas Prop. 5.3 mainly exploits the optimality of the SPRT stated in Prop. 2.17. Even though these results are very similar, there are slight differences, as briefly commented on before the statement of Prop. 5.3, and thus we included both into this thesis for the sake of completeness.

For WST testing on $\mathcal{Q}_m^h$ for α and β, we presented instance-wise sample complexity upper and lower bounds, which asymptotically coincide in the symmetric case $\alpha = \beta = \gamma$ in a worst-case sense up to logarithmic factors and are then of order $\widetilde{\Theta}(\frac{m^2}{h^2} \ln \frac{1}{\gamma})$. In contrast to the CW-related problems, their dependence on m is of order $\widetilde{\Theta}(m^2)$ instead of $\widetilde{\Theta}(m)$, which indicates that basically all of the $\Theta(m^2)$ entries of $\mathbf{Q}$ have to be taken into account for WST testing, and hence a naive testing procedure is already optimal in this regard. This difference in terms of the asymptotics w.r.t. m coincides with the results from Ch. 3 for deterministic CW testification and WST testing. However, we saw that incorporating the graph-theoretical concept of acyclicity in extension (as well as negligibility of edges for the active scenario) potentially allows for earlier termination and results in a more sophisticated WST testing procedure that outperforms the naive one both theoretically and in experiments. This variant is not guaranteed to terminate before being sure about all $\binom{m}{2}$ entries of $\mathbf{Q}$, but if $m \geq 4$ is even, Thm. 3.26 would allow the construction of a solution $\mathcal{A}$ to $\mathcal{P}_{\mathrm{WST}}^{m,h,\gamma}$ with worst-case sample complexity $\widetilde{\mathcal{O}}\big(\frac{m^2}{h^2} \ln \frac{1}{\gamma}\big)$, which queries until termination a.s. only $\binom{m}{2} - 1$ different pairs $\{i, j\} \in [m]_2$. However, we do not expect this solution to be better than that from Thm. 5.9(ii) and thus we did not state it here.

In Part (ii) of Thm. 5.5 we saw that any solution $\mathcal{A}$ to $\mathcal{P}_{\mathrm{WST}}^{m,0,\alpha,\beta}$ necessarily fulfills $\sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \Omega_{\sup}\left(\frac{1}{h^2}\ln\ln\frac{1}{h}\right)$, which is w.r.t. h by a factor of $\ln\ln\frac{1}{h}$ larger than the corresponding lower bound for solutions to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$, and this is consistent with our lower bounds for CW testification stated in Ch. 4. Unfortunately, this bound only captures the asymptotic behaviour w.r.t. h for fixed m, α and β. We suppose that even $\sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \Omega_{\sup}\left(\frac{m^2}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\alpha\vee\beta}\right)$ holds, but we were not able to show it, because the limes superior in Prop. 2.13 seemingly does not allow to prove an appropriate sample complexity lower bound of order $\Omega_{\sup}\left(\frac{|J|}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\alpha\vee\beta}\right)$ for the multiple coin tossing problem (2.14); cf. our related discussion in Sec. 2.6.

In Sec. 5.4 we approached the WST testing problem via the LRT statistics for $\mathbf{H}_0 : \mathbf{Q} \in \mathcal{Q}_m(\mathrm{WST})$ and $\mathbf{H}_1 : \mathbf{Q} \in \mathcal{Q}_m(\neg\mathrm{WST})$ and the corresponding interchanged hypotheses. Our efforts resulted in an alternative solution (Alg. 22) to $\mathcal{P}_{\mathrm{WST}}^{m,0,\gamma}$ and thus also to $\mathcal{P}_{\mathrm{WST}}^{m,h,\gamma}$, which is applicable in both the passive and active scenario. For the latter one, we proved a sample complexity bound of order $\mathcal{O}\left(\frac{m^{2\kappa}}{h^4}\ln\frac{1}{\gamma}\right)$ that is with regard to Thm. 5.9 up to a factor $\widetilde{\Theta}\left(\frac{m^{2\kappa-2}}{h^2}\right)$ suboptimal. In our experiments, Alg. 22 performed much worse than Alg. 21; nevertheless, we incorporated this solution for the sake of completeness.

Further investigation of the above mentioned LRT statistics allowed us in Sec. 5.4.6 to formulate asymptotic size-α tests for WST and $\neg$WST, which are of the same fashion as corresponding results in [Iverson and Falmagne, 1985]. There, the authors did not consider the general DB setting but restricted themselves to an offline scenario, in which the analysis is easier and sharper bounds on the asymptotics of the tail probabilities are obtained, cf. our discussion at the end of Sec. 5.4.6. Even though these asymptotic results are not the focus of this thesis, we added them for the sake of completeness.

As already mentioned, some of the results presented in this chapter have been published in [Haddenhorst et al., 2021b], but some others are novel: In contrast to Thm. 4.2 in the paper, we showed in Sec. 5.1 also impossibility of XST testing on $\mathcal{Q}_m^h$ for the case XST = λST and deduced some of our results from Cor. 2.48. Whereas the content of Sec. 5.2, Sec. 5.3 and 5.5 has basically already been contained in [Haddenhorst et al., 2021b], all of Sec. 5.4 has not been published so far. And for the empirical evaluation in Sec. 5.6, we added here $\mathcal{A}_{\mathrm{improved}}^{\mathrm{PPR}}$ as a further solution to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$ and also briefly discussed Alg. 22.

There appear to be multiple aspects that could potentially be of interest for further research. In comparison to Ch. 4, we merely focused on WST *testing*. As any $\mathbf{Q} \in \mathcal{Q}_m^0(\mathrm{WST})$ admits an underlying ranking $\sigma \in \mathbb{S}_m$ such that $q_{i,j} > 1/2$ iff $\sigma(i) < \sigma(j)$, one may also consider the problem of *testification* for this underlying ranking. Since verifying the true ranking can presumably only be done when being sure enough about all entries of $\mathbf{Q}$, an early termination appears only possible in case one decides for $\neg$WST. For this reason, we suppose the worst-case sample complexity of testification of this ranking on $\mathcal{Q}_m^h$ to be of the order $\widetilde{\Theta}\left(\frac{m^2}{h^2}\ln\frac{1}{\gamma}\right)$. This would basically coincide with the sample complexity required to identify such a ranking whilst assuming its existence [Falahatgar et al., 2017b,a, 2018, Ren et al., 2019a, Jamieson and Nowak, 2011]. Related to this is the problem to learn the *top-k ranking*, a ranking over the best k arms according to $\mathbf{Q}$, as e.g. discussed in [Mohajer et al., 2017], and for this problem one might also formulate and investigate a testification variant.

Apart from that, further investigation of other types of transitivity could be of interest. Recall that the impossibility results for testing XST $\neq$ WST from Sec. 5.1 merely show that XST testing is to some extent impossible *in a worst-case sense on* $\mathcal{Q}_m^h$. One could still investigate XST testing with guarantees for all $\mathbf{Q}$ in an appropriate parameter space $\mathcal{Q}'_m \neq \mathcal{Q}_m^h$ or ask for more explicit instance-wise sample complexity bounds. Last but not least, one might analyze XST testing with respect to alternative performance measures: Our main focus has been to prove upper and lower bounds for the expected sample complexity of solutions to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$, but of course one could also ask for high-probability bounds instead. For such purpose, one would presumably require lower bounds for solutions to $\mathcal{P}_{\mathrm{Coin}}^{h,\gamma}$ and $\mathcal{P}_{\mathrm{Coin}}^{\gamma}$ that hold with high probability; though we restricted ourselves in Ch. 2 to lower expected sample complexity bounds, such kind of bounds have already been shown in related testing scenarios [Daskalakis and Kawase, 2017] and might potentially be transferrable for this purpose.

Part III.

Learning Problems in Multi-Dueling Bandits

6. Identification of the GCW in Multi-Dueling Bandits

In this chapter, we move from the dueling to the more general multi-dueling bandits (MDB) scenario, which has recently been introduced [Brost et al., 2016, Sui et al., 2017, Saha and Gopalan, 2018] as a generalization of the former. It comes with multiple practically relevant applications, such as algorithm configuration [El Mesaoudi-Paul et al., 2020] or online retrieval evaluation [Schuth et al., 2016]. Instead of pairs of arms, in this generalization a set consisting of $k \geq 2$ arms can be chosen in each time step. These arms compete against each other and determine a single winner, which is observed as feedback by the learner. The outcomes of the (multi-)duels in the MDB scenario are typically assumed to be of time-stationary stochastic nature in the sense that whenever arms $1, \ldots, k$ compete against each other, then i wins with some underlying (unknown) ground-truth probability $\mathbf{P}(i|\{1, \ldots, k\})$.

We will analyze the problem to identify the "best" of all arms in MDB, which is an often targeted learning task in the context of multi-armed bandits (MAB) and its variants. Similarly as for dueling bandits, the term "best arm" is not uniquely defined in the field of MDB, but instead there are several possible notions prevalent in the literature. We will focus on the generalized Condorcet winner (GCW), which we introduced in Sec. 1.2, as the notion for the "best arm", and analyze the best-arm identification problem in a γ-PAC scenario. In Sec. 6.4, we briefly address alternative notions and learning scenarios in the literature.

More precisely, we analyze the sample complexity of (probabilistic) algorithms that are able to identify the GCW with high probability under the assumption of mere existence as well as more restrictive assumptions. We provide upper and lower bounds on the sample complexity for this task, which depend on the desired confidence, the total number m of alternatives, the size k of allowed query sets as well as the underlying unknown preference probabilities $\mathbf{P}(i|S)$. With only a few exceptions that we briefly comment on in Sec. 6.4, all results presented in this chapter have already been published in [Haddenhorst et al., 2021c].

Parameter Classes As a gentle start, we fix the notation used in the further course, some of which has already been introduced in Ch. 1 and Sec. 2.3, but is restated here for the sake of convenience. If not explicitly stated otherwise, we suppose throughout this chapter the total number of arms m, the query set size $k \in \{2, \ldots, m\}$, a desired confidence $1 - \gamma \in (0, 1)$ and a complexity parameter $h \in (0, 1)$ to be arbitrary but fixed. Recall $[m]_k = \{S \subseteq [m] \,|\, |S| = k\}$. For any subset of size k, i.e., $S \in [m]_k$, define $\Delta_S := \{\mathbf{p} = (p_j)_{j \in S} \in [0, 1]^{|S|} \,|\, \sum_{j \in S} p_j = 1\}$ as the set of all possible parameters for a categorical random variable $X \sim \mathrm{Cat}((p_j)_{j \in S})$, i.e., $\mathbb{P}(X = j) = p_j$ for any $j \in S$. For $\mathbf{p} \in \Delta_S$, we write $\mathrm{mode}(\mathbf{p}) := \mathrm{argmax}_{j \in S} p_j$ and in case $|\mathrm{mode}(\mathbf{p})| = 1$ we denote by $\mathrm{mode}(\mathbf{p})$ — with a slight abuse of notation — also the unique element in $\mathrm{mode}(\mathbf{p})$. Let us define for $h \in (0,1]$ the sets

$$\Delta_S^h := \left\{ \mathbf{p} \in \Delta_S \,|\, \exists i \in S \text{ s.t. } p_i \geq \max_{j \in S \setminus \{i\}} p_j + h \right\},$$

and with this $\Delta_S^0 := \bigcup_{h \in (0,1)} \Delta_S^h = \{\mathbf{p} \in \Delta_S \,|\, \exists i \in S \text{ s.t. } p_i > \max_{j \in S \setminus \{i\}} p_j\}$. These sets are nested in the sense that $\Delta_S^h \subseteq \Delta_S^{h'} \Leftrightarrow h \geq h'$. If $\mathbf{p} \in \Delta_S$ is fixed, the value $h(\mathbf{p}) := \max\{h \in [0,1] \,|\, \mathbf{p} \in \Delta_S^h\}$ is well-defined and we have $\mathbf{p} \in \Delta_S^h$ iff $h \leq h(\mathbf{p})$. Obviously, the equivalence $|\text{mode}(\mathbf{p})| = 1 \Leftrightarrow \mathbf{p} \in \Delta_S^0$ holds for all $\mathbf{p} \in \Delta_S$.

Recall that we called a family $\mathbf{P} = \{\mathbf{P}(\cdot \,|\, S)\}_{S \in [m]_k}$ of parameters $\mathbf{P}(\cdot \,|\, S) \in \Delta_S$, $S \in [m]_k$, a *probability model* (PM) *on* $[m]_k$, and write PM_k^m for the set of all probability models on $[m]_k$. We have defined the particular subsets

$$PM_k^m(\Delta^0) := \{\mathbf{P} = \{\mathbf{P}(\cdot \,|\, S)\}_{S \in [m]_k} \,\big|\, \forall\, S \in [m]_k \,:\, \mathbf{P}(\cdot \,|\, S) \in \Delta_S^0\},$$
$$PM_k^m(\Delta^h) := \{\mathbf{P} = \{\mathbf{P}(\cdot \,|\, S)\}_{S \in [m]_k} \,\big|\, \forall S \in [m]_k \,:\, \mathbf{P}(\cdot \,|\, S) \in \Delta_S^h \},$$
$$PM_k^m(\text{PL}) := \{\{\mathbf{P}(\cdot \,|\, S)\}_{S \in [m]_k} \,\big|\, \exists \boldsymbol{\theta} \in (0,\infty)^m \,\forall\, S \in [m]_k \,:\, \mathbf{P}(i \,|\, S) = \theta_i / (\textstyle\sum_{j \in S} \theta_j)\}$$

of PM_k^m, where $PM_k^m(\text{PL})$ is the set of all probability models $\mathbf{P}$ coherent with a Plackett-Luce (PL) model [Plackett, 1975, Luce, 1959]. Writing

$$h(\mathbf{P}) := \max_{h \in [0,1]}\{\mathbf{P} \in PM_k^m(\Delta^h)\} = \min_{S \in [m]_k} h(\mathbf{P}(\cdot \,|\, S)),$$

it is easy to see that $\mathbf{P} \in PM_k^m(\Delta^h)$ iff $h \leq h(\mathbf{P})$. As already defined in Ch. 1, an element $i \in [m]$ is called a *generalized Condorcet winner* (GCW) of $\mathbf{P}$ if

$$\forall S \in [m]_k \text{ with } i \in S, \forall j \in S : \mathbf{P}(i \,|\, S) - \mathbf{P}(j \,|\, S) \geq 0$$

and we write $\text{GCW}(\mathbf{P})$ for the set of all GCWs of $\mathbf{P}$. With this, we have defined

$$PM_k^m(\text{GCW}) := \{\mathbf{P} = \{\mathbf{P}(\cdot \,|\, S)\}_{S \in [m]_k} \,\big|\, \text{GCW}(\mathbf{P}) \neq \emptyset\},$$
$$PM_k^m(\text{GCW}^*) := \{\mathbf{P} = \{\mathbf{P}(\cdot \,|\, S)\}_{S \in [m]_k} \,\big|\, |\text{GCW}(\mathbf{P})| = 1\}$$

and further $PM_k^m(h\text{GCW})$ as the set

$$\{\{\mathbf{P}(\cdot \,|\, S)\}_{S \in [m]_k} \,\big|\, \exists i : \forall S \in [m]_k \text{ with } i \in S, \forall j \in S \setminus \{i\} : \mathbf{P}(i \,|\, S) - \mathbf{P}(j \,|\, S) \geq h\}.$$

Note that $PM_k^m(\text{GCW}^*) = \bigcup_{h > 0} PM_k^m(h\text{GCW})$ and every $\mathbf{P} \in PM_k^m(\text{GCW})$ has at least one GCW, whereas any $\mathbf{P} \in PM_k^m(\text{GCW}^*)$ has a unique GCW.

The following lemma reveals the relationships consisting between the different introduced classes of probability models; in addition these results have been illustrated as a Venn diagram in Fig. 2, which we restate for convenience as Fig. 10. For simplicity, we simply refer to these classes as *assumptions*.

Lemma 6.1. *For any $k, m \in \mathbb{N}$ and $h \in (0,1)$ we have the implications*

$$PM_k^m(h\text{GCW}) \subsetneq PM_k^m(\text{GCW}^*) \subsetneq PM_k^m(\text{GCW}),$$
$$PM_k^m(\Delta^h) \subsetneq PM_k^m(\Delta^0),$$
$$PM_k^m(\text{PL}) \subsetneq PM_k^m(\text{GCW}),$$
$$PM_k^m(\Delta^h) \cap PM_k^m(\text{GCW}) \subseteq PM_k^m(h\text{GCW}),$$
$$PM_k^m(\Delta^0) \cap PM_k^m(\text{GCW}) \subseteq PM_k^m(\text{GCW}^*).$$

If $k = m$, the last two statements are correct with $=$ instead of $\subseteq$, and otherwise with $\subsetneq$.

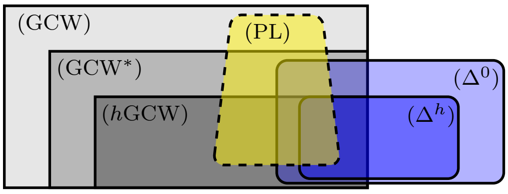


Figure 10.: Problem classes for GCW identification in multi-dueling bandits; here, we simply write (X) instead of $PM_k^m(\mathrm{X})$ for X in $\{\mathrm{GCW}, \mathrm{GCW}^*, h\mathrm{GCW}, \Delta^0, \Delta^h, \mathrm{PL}\}$.

Proof. This is a direct consequence of the definitions. □

Also, let us briefly recap the notion of a multi-dueling bandits algorithm. At any time step $t \in \mathbb{N}$, such an algorithm is allowed to choose one *query set* $S_t \in [m]_k$, for which it then observes a winner X_{t,S_t}. Here, we suppose the family $\{X_{t,S}\}_{t\in\mathbb{N},S\in[m]_k}$ to be an independent family of categorical random variables $X_{t,S} \sim \mathrm{Cat}(\mathbf{P}(\cdot|S)) \in S$, i.e., the observed feedback is supposed to be independent across time and query sets. At some time, $\mathcal{A}$ may decide to make no more queries and output a decision $\mathbf{D}(\mathcal{A})$, which can e.g. be seen in a GCW identification scenario as a guess for the GCW of $\mathbf{P}$. As usual, $T^{\mathcal{A}} \in \mathbb{N} \cup \{\infty\}$ denotes the sample complexity of $\mathcal{A}$, i.e., the total number of queries made by $\mathcal{A}$ before termination, and both $\mathbf{D}(\mathcal{A})$ and $T^{\mathcal{A}}$ are apparently random variables w.r.t. the sigma-algebra generated by the stochastic feedback mechanism. We write $\mathbb{P}_{\mathbf{P}}$ for the probability measure corresponding to the stochastic feedback mechanism if the unknown ground-truth probability model is $\mathbf{P}$.

If an assumption X with $PM_k^m(\mathrm{X}) \subseteq PM_k^m(\exists\mathrm{GCW})$, $m, k \in \mathbb{N}$ with $2 \leq k \leq m$ and $\gamma \in (0,1)$ are given, we say that such an algorithm $\mathcal{A}$ solves *GCW identification on* $PM_k^m(\mathrm{X})$ (short: $\mathcal{P}_{\mathrm{GCWi}}^{m,k,\gamma}$) if

$$\mathbb{P}_{\mathbf{P}}\big(\mathbf{D}(\mathcal{A}) \in \mathrm{GCW}(\mathbf{P})\big) \geq 1-\gamma \quad \text{for all } \mathbf{P} \in PM_k^m(\mathrm{X}).$$

To state the yet missing argument in the proof of Prop. 4.4 from above, we also take a look at a generalization of $\mathcal{P}_{\mathrm{CWv}}^{m,h,\gamma}(\mathrm{CW})$. Formally, if k, m, γ and X are as above, $\mathcal{A}$ is said to solve *GCW verification on* $PM_k^m(\mathrm{X})$ (short: $\mathcal{P}_{\mathrm{GCWv}}^{m,k,\gamma}(\mathrm{X})$) if

$$\forall i \in [m]\, \forall \mathbf{P} \in PM_k^m(\mathrm{X}) : \mathbb{P}_{\mathbf{P}}\left(\mathbf{D}(\mathcal{A}(i)) = \mathbf{1}_{\{\mathrm{GCW}(\mathbf{P})=i\}}\right) \geq 1-\gamma,$$

where we have as usual written $\mathcal{A}(i)$ for $\mathcal{A}$ started with input i. In other words, whenever started with a parameter $i \in [m]$, it correctly decides with error probability $\leq \gamma$ for any $\mathbf{P} \in PM_k^m(\mathrm{X})$ whether $\mathrm{GCW}(\mathbf{P}) = i$ is true or not.

Regarding that $[k]_k = \{[k]\}$ and $PM_k^k = \Delta_k$, $PM_k^k(\Delta^h) = PM_k^k(h\mathrm{GCW}) = \Delta_k^h$, $PM_k^k(\mathrm{GCW}^*) = \Delta_k^0$ and $h(\mathbf{P}) = h(\mathbf{p})$ hold, we have the identity $\mathcal{P}_{\mathrm{GCWi}}^{k,k,\gamma}(\Delta^h \wedge \mathrm{GCW}) = \mathcal{P}_{\mathrm{GCWi}}^{k,k,\gamma}(h\mathrm{GCW}) = \mathcal{P}_{\mathrm{Die}}^{k,h,\gamma}$ for any $h \in [0,1]$, i.e., the case $m = k$ of GCW identification coincides with the mode identification problem of a die, which has already been discussed in Sec. 2.3.

Outline of The Chapter In this chapter, we treat the combinatorically more challenging case $m > k$ and show the worst-case bounds presented in Table 1.1, which we restate below for convenience as Table 6.1. Recall from Sec. 1.3 that the bounds in the table are sample complexity lower and upper bounds for solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{X})$, which are valid in the worst-case w.r.t. $PM^m_k(\mathrm{X} \wedge \mathrm{Y})$. In fact, for most of the bounds therein, we also obtain more sophisticated but technical instance-wise versions below. We start with several sample complexity lower bounds for $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{X})$ under different assumptions X (Thm. 6.3, Thm. 6.4 and Thm. 6.8) as well as the promised lower bound for $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWv}}(\Delta^h \wedge \mathrm{GCW})$, that we state as Thm. 6.9. Then, we develop a solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$, which we call *Dvoretzky-Kiefer-Wolfowitz tournament* (DKWT), since it is a knockout procedure and based on the famous Dvoretzky-Kiefer-Wolfowitz inequality (Thm. 6.12). For the easier problem $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(h\mathrm{GCW} \wedge \Delta^0)$ we construct a slightly modified version of DKWT that comes with appealing theoretical guarantees (Thm. 6.13). A look at the corresponding lower bounds will indicate (up to logarithmic factors) asymptotic optimality of both solutions in a worst-case sense, showing that both problems require basically $\widetilde{\Theta}(\frac{m}{kh^2} \ln \frac{1}{\gamma})$ samples to be solved. By translating results from [Saha and Gopalan, 2020b] into our setting, we will see that an additional Plackett-Luce assumption simplifies the GCW identification problem by a factor $\frac{1}{k} + h$ w.r.t. the worst-case asymptotic required sample size.

Similarly as in the previous chapters, we try to improve the readability of this chapter by deferring technical proofs and also some technical instance-wise bounds to separate sections.

Table 6.1.: Sample complexity bounds of solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{X})$.

(X)	(Y)	Type	Asymptotic bounds	Reference
(PL)	(hGCW)	in exp.	$\Omega(\frac{m}{h^2 k}(\frac{1}{k} + h) \ln \frac{1}{\gamma})$	Thm. 6.3
($\Delta^h \wedge$ GCW)	(Δ^h)	in exp.	$\Omega(\frac{m}{h^2 k} \ln \frac{1}{\gamma})$	Thm. 6.4
(PL $\wedge$ GCW*)	(hGCW)	w.h.p.	$\mathcal{O}(\frac{m}{h^2 k}(\frac{1}{k} + h) \ln(\frac{k}{\gamma} \ln \frac{1}{h}))$	Thm. 6.11
(GCW $\wedge \Delta^0$)	(Δ^h)	w.h.p.	$\mathcal{O}(\frac{m}{h^2 k} \ln(\frac{m}{k})(\ln\ln \frac{1}{h} + \ln \frac{1}{\gamma}))$	Thm. 6.12
(hGCW $\wedge \Delta^0$)	(hGCW)	a.s.	$\mathcal{O}(\frac{m}{h^2 k} \ln(\frac{m}{k\gamma}))$	Thm. 6.13

6.1. Lower Bounds on the GCW Identification Problem

In this section we provide sample complexity lower bounds for solutions to the GCW identification problem for arbitrary $2 \leq k \leq m$. We start with bounds for solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{PL})$. For this purpose, recall the notion of $\mathbf{P}(\boldsymbol{\theta}) \in PM^m_k(\mathrm{PL})$ introduced in Sec. 2.5.1, i.e., for $\boldsymbol{\theta} \in (0,\infty)^m$, $\mathbf{P}(\boldsymbol{\theta}) = \{\mathbf{P}(\boldsymbol{\theta})(\cdot|S)\}_{S \in [m]_k}$ is given via

$$\mathbf{P}(\boldsymbol{\theta})(i|S) \coloneqq \frac{\theta_i}{\sum_{a \in S} \theta_a} \quad \text{for any } S \in [m]_k \text{ and } i \in S.$$

As $\mathbf{P}(x\boldsymbol{\theta}) = \mathbf{P}(\boldsymbol{\theta})$ holds for any $x > 0$ and $\boldsymbol{\theta} \in (0,\infty)^m$, we may restrict ourselves w.l.o.g. to those $\boldsymbol{\theta}$ with $\max_{i \in [m]} \theta_i = 1$ throughout the rest of this chapter.

Saha and Gopalan [2020b] gave the following instance-wise sample complexity lower bound for solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{PL})$ in dependence on the underlying Plackett-Luce parameter.

Theorem 6.2. *Any solution $\mathcal{A}$ to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{PL})$ fulfills*

$$\mathbb{E}_{\mathbf{P}(\boldsymbol{\theta})}\left[T^{\mathcal{A}}\right] \in \Omega\left(\max\left(\sum\nolimits_{j=2}^{m} \frac{\theta_j}{(1-\theta_j)^2} \ln\frac{1}{\gamma}, \frac{m}{k}\ln\frac{1}{\gamma}\right)\right)$$

for any $\boldsymbol{\theta} \in (0,1]^m$ with $1 = \theta_1 > \max_{j\geq 2}\theta_j$.

Proof. Confer Thm. 7 in [Saha and Gopalan, 2020b]. □

By means of the characterization of $PM^m_k(\mathrm{PL}\wedge h\mathrm{GCW})$ made in Lem. 2.45, this bound translates to a lower bound for solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{PL}\wedge h\mathrm{GCW})$ as follows.

Theorem 6.3. *Any solution $\mathcal{A}$ to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{PL})$ fulfills*

$$\sup\nolimits_{\mathbf{P}\in PM^m_k(\mathrm{PL}\wedge h\mathrm{GCW})} \mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}}\right] \in \Omega\left(\frac{m\left(\frac{1}{k}+h\right)\ln\frac{1}{\gamma}}{kh^2}\right). \tag{6.1}$$

Proof of Thm. 6.3. Define $\boldsymbol{\theta} \in (0,1]^m$ via $\theta_1 := 1$ and $\theta_j := \frac{1-h}{h(k-1)+1}$ for $2 \leq j \leq m$. Then,

$$\begin{aligned} h\sum\nolimits_{j=1}^{k}\theta_j + \theta_2 - \theta_1 &= h\left(1+\frac{(k-1)(1-h)}{h(k-1)+1}\right) + \frac{1-h-h(k-1)-1}{h(k-1)+1} \\ &= \frac{h(h(k-1)+1+(k-1)(1-h))-hk}{h(k-1)+1} = 0 \end{aligned}$$

shows with regard to Lem. 2.45 that $\mathbf{P}(\boldsymbol{\theta}) \in PM^m_k(h\mathrm{GCW})$ is fulfilled. Moreover, for $j \in \{2,\ldots,m\}$ we have $1-\theta_j = \frac{hk}{h(k-1)+1}$ and thus

$$\frac{\theta_j}{(1-\theta_j)^2} = \frac{(h(k-1)+1)(1-h)}{h^2k^2} = \frac{hk(1-h)+(1-h)^2}{h^2k^2},$$

which is in $\Theta(\frac{1}{hk}+\frac{1}{h^2k^2}) = \Theta\left(\frac{1}{kh^2}\left(\frac{1}{k}+h\right)\right)$, since $1-h \in \Theta(1)$ as $h \searrow 0$. In particular,

$$\sum_{j=2}^{m}\frac{\theta_j}{(1-\theta_j)^2} \in \Theta\left(\frac{m}{kh^2}\left(\frac{1}{k}+h\right)\right)$$

and thus the statement follows from Thm. 6.2. □

One of the key ingredients for Thm. 6.2 is the change-of-measure argument by Kaufmann et al. [2016], which we already stated as Lem. 2.42. By means of the latter technique, we are also able to show the following instance-based as well as worst-case lower bounds for any solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\Delta^h\wedge\mathrm{GCW})$.

Theorem 6.4. *Suppose $\mathcal{A}$ solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\Delta^h\wedge\mathrm{GCW})$ and let $\mathbf{P} \in PM^m_k(\Delta^h\wedge\mathrm{GCW})$ be arbitrary with $\min_{S\in[m]_k}\min_{j\in S}\mathbf{P}(j|S) > 0$. For $S \in [m]_k$ write $m_S := \mathrm{mode}(\mathbf{P}(\cdot|S))$ and for any $l \in S\setminus\{m_S\}$ define $\mathbf{P}^{[l]}(\cdot|S) \in \Delta_S$ via*

$$\mathbf{P}^{[l]}(l|S) := \mathbf{P}(m_S|S), \quad \mathbf{P}^{[l]}(m_S|S) := \mathbf{P}(l|S), \quad \forall j\in S\setminus\{l,m_S\}: \mathbf{P}^{[l]}(j|S) := \mathbf{P}(j|S).$$

Then,

$$\mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}}\right] \geq \frac{\ln \frac{1}{2.4\gamma}}{k-1} \sum\nolimits_{l\in[m]\setminus\{\mathrm{GCW}(\mathbf{P})\}} \min\nolimits_{S\in[m]_k: l\in S\setminus\{m_S\}} \frac{1}{\mathrm{KL}(\mathbf{P}(\cdot|S),\mathbf{P}^{[l]}(\cdot|S))},$$

where $\mathrm{KL}(\mathbf{P}(\cdot|S),\mathbf{P}^{[l]}(\cdot|S))$ *denotes the Kullback-Leibler divergence between two categorical distributions* $X\sim\mathrm{Cat}(\mathbf{P}(\cdot|S))$ *and* $Y\sim\mathrm{Cat}(\mathbf{P}^{[l]}(\cdot|S))$*. Moreover, we have*

$$\sup\nolimits_{\mathbf{P}\in PM_k^m(\Delta^h\wedge\mathrm{GCW})} \mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}}\right] \geq \frac{m(1-h^2)\ln\frac{1}{2.4\gamma}}{4kh^2}.$$

Proof of Thm. 6.4. We prove the instance-wise and asymptotic worst-case lower bound separately.

Part 1: Proof of the instance-wise bound

After relabeling the items in $[m]$, we may suppose w.l.o.g. $\mathrm{GCW}(\mathbf{P})=1$ throughout the proof. Write for convenience $\mathbf{P}^{[1]}:=\mathbf{P}$, recall that $m_S=\mathrm{mode}(\mathbf{P}^{[1]}(\cdot|S))$ for any $S\in[m]_k$ and define $\mathbf{P}^{[l]}\in PM_k^m(\Delta^h)$ for each $l\in\{2,\ldots,m\}$ via

$$\begin{aligned}\mathbf{P}^{[l]}(l|S) &:= \mathbf{P}^{[1]}(m_S|S), \quad \mathbf{P}^{[l]}(m_S|S) := \mathbf{P}^{[1]}(l|S),\\ \mathbf{P}^{[l]}(j|S) &:= \mathbf{P}^{[1]}(j|S) \text{ for all } j\in S\setminus\{l,m_S\}\end{aligned} \tag{6.2}$$

for any $S\in[m]_k$ with $l\in S$ and

$$\mathbf{P}^{[l]}(j|S) := \mathbf{P}^{[1]}(j|S) \text{ for all } j\in S$$

for any $S\in[m]_k$ with $l\notin S$. Abbreviating $\mathbf{P}_S^{[r]}:=\mathbf{P}^{[r]}(\cdot|S)$ we have $\mathrm{KL}\left(\mathbf{P}_S^{[1]},\mathbf{P}_S^{[l]}\right)=0$ whenever $S\notin[m]_k^{(l)}:=\{S\in[m]_k\,|\,l\in S \text{ and } l\neq m_S\}$. Define

$$\Sigma(l) := \sum\nolimits_{S\in[m]_k^{(l)}} \mathbb{E}_{\mathbf{P}^{[1]}}\left[T_S^{\mathcal{A}}\right]$$

for each $l\in\{2,\ldots,m\}$. Now, suppose l to be fixed for the moment and note that $\mathrm{GCW}(\mathbf{P}^{[l]})=l$ holds by construction of $\mathbf{P}^{[l]}$. As $\mathcal{A}$ solves $\mathcal{P}_{\mathrm{GCWi}}^{m,k,\gamma}(\Delta^h\wedge\mathrm{GCW})$, the event $\mathcal{E}:=\{\mathbf{D}(\mathcal{A})=1\}\in\mathcal{F}_{T^{\mathcal{A}}}$ fulfills $\mathbb{P}_{\mathbf{P}^{[1]}}(\mathcal{E})\geq 1-\gamma$ and $\mathbb{P}_{\mathbf{P}^{[l]}}(\mathcal{E})\leq\gamma$. Consequently, by applying Part (ii) of Lem. 2.43 and Lem. 2.42, we obtain

$$\begin{aligned}\ln\frac{1}{2.4\gamma} &\leq \mathrm{kl}\left(\mathbb{P}_{\mathbf{P}^{[1]}}(\mathcal{E}),\mathbb{P}_{\mathbf{P}^{[l]}}(\mathcal{E})\right)\\ &\leq \sum\nolimits_{S\in[m]_k} \mathbb{E}_{\mathbf{P}^{[1]}}\left[T_S^{\mathcal{A}}\right]\mathrm{KL}\left(\mathbf{P}_S^{[1]},\mathbf{P}_S^{[l]}\right)\\ &= \sum\nolimits_{S\in[m]_k^{(l)}} \mathbb{E}_{\mathbf{P}^{[1]}}\left[T_S^{\mathcal{A}}\right]\mathrm{KL}\left(\mathbf{P}_S^{[1]},\mathbf{P}_S^{[l]}\right),\end{aligned}$$

that is,

$$\Sigma(l) \geq \ln\left(\frac{1}{2.4\gamma}\right)\min\nolimits_{S\in[m]_k^{(l)}} \frac{1}{\mathrm{KL}\left(\mathbf{P}_S^{[1]},\mathbf{P}_S^{[l]}\right)}. \tag{6.3}$$

For any $S = \{i_1, \dots, i_k\} \in [m]_k$ with $i_1 := m_S$ the term $\mathbb{E}_{\mathbf{P}^{[1]}}\left[T_S^{\mathcal{A}}\right]$ appears exactly $k-1$ times as a summand in

$$\Sigma(2) + \cdots + \Sigma(m) = \sum\nolimits_{l=2}^{m} \sum\nolimits_{S \in [m]_k : m_S \neq l \in S} \mathbb{E}_{\mathbf{P}^{[1]}}\left[T_S^{\mathcal{A}}\right],$$

namely as one summand in $\Sigma(i_2), \dots, \Sigma(i_k)$ each. Hence, (6.3) lets us infer

$$\begin{aligned}(k-1)\mathbb{E}_{\mathbf{P}^{[1]}}\left[T^{\mathcal{A}}\right] &= \sum\nolimits_{S \in [m]_k} (k-1)\mathbb{E}_{\mathbf{P}^{[1]}}\left[T_S^{\mathcal{A}}\right] \\ &\geq \Sigma(2) + \cdots + \Sigma(m) \\ &\geq \ln\left(\frac{1}{2.4\gamma}\right) \sum\nolimits_{l=2}^{m} \min\nolimits_{S \in [m]_k^{(l)}} \frac{1}{\mathrm{KL}\left(\mathbf{P}_S^{[1]}, \mathbf{P}_S^{[l]}\right)}.\end{aligned}$$

This completes our proof of the instance-wise bound. ■

Part 2: Proof of the worst-case bound

Since the statement is trivial for $h = 1$, we may assume w.l.o.g. $h \in (0,1)$ in the following. Let us abbreviate $\Delta_{[m]_k} := \{\mathbf{w} = (w_S)_{S \in [m]_k} \in [0,1]^{[m]_k} \mid \sum_{S \in [m]_k} w_S = 1\}$. For $S \in [m]_k$, write $S = \{S_{(1)}, \dots, S_{(k)}\}$ with $S_{(1)} < \cdots < S_{(k)}$. Suppose $\varepsilon \in (0, 1/2 \wedge (1-h))$ to be arbitrary but fixed for the moment and define $\mathbf{P}^{[1,\varepsilon]} \in PM_k^m(\mathrm{GCW} \wedge \Delta^h)$ via

$$\mathbf{P}^{[1,\varepsilon]}(S_{(1)}|S) := \frac{1+h-\varepsilon}{2}, \quad \mathbf{P}^{[1,\varepsilon]}(S_{(2)}|S) := \frac{1-h-\varepsilon}{2}$$

and

$$\forall j \in \{3, \dots, k\} : \mathbf{P}^{[1,\varepsilon]}(S_{(j)}|S) := \frac{\varepsilon}{k-2}$$

for any $S \in [m]_k$. For $l \in \{2, \dots, m\}$ let $\mathbf{P}^{[l,\varepsilon]}$ be as $\mathbf{P}^{[1,\varepsilon]}$ with $[m]$ being relabeled via the l-shift $\nu_l : [m] \to [m]$ given by

$$1 \mapsto l, \quad 2 \mapsto l+1, \quad \dots \quad m-l-1 \mapsto m, \quad m-l \mapsto 1, \quad \dots \quad m \mapsto l-1,$$

i.e., $\mathbf{P}^{[l,\varepsilon]}(\nu_l(i_r)|\{\nu_l(i_1), \dots, \nu_l(i_k)\}) = \mathbf{P}^{[1,\varepsilon]}(i_r|\{i_1, \dots, i_k\})$ for any $\{i_1, \dots, i_k\} \in [m]_k$ and $r \in [k]$. Then, $\mathbf{P}^{[l,\varepsilon]} \in PM_k^m(\mathrm{GCW} \wedge \Delta^h)$ and $\mathrm{GCW}(\mathbf{P}^{[l,\varepsilon]}) = l$ hold for any $l \in [m]$. Write

$$\mathfrak{P}^*(\varepsilon) := \left\{\mathbf{P}^{[1,\varepsilon]}, \mathbf{P}^{[2,\varepsilon]}, \dots \mathbf{P}^{[m,\varepsilon]}\right\}$$

and define $\mathfrak{P}_*(\neg l)$ as the set

$$\left\{\mathbf{P} \in PM_k^m(\mathrm{GCW} \wedge \Delta^h) \,|\, \mathrm{GCW}(\mathbf{P}) \neq l \text{ and } \forall S \in [m]_k : \min\nolimits_{j \in S} \mathbf{P}(j|S) > 0\right\}.$$

For any $\mathbf{P}, \mathbf{P}' \in PM_k^m(\mathrm{GCW} \wedge \Delta^h)$ fulfilling $\min_{S \in [m]_k} \min_{j \in S} \mathbf{P}(j|S) > 0$ as well as $\min_{S \in [m]_k} \min_{j \in S} \mathbf{P}'(j|S) > 0$ and $\mathrm{GCW}(\mathbf{P}) \neq \mathrm{GCW}(\mathbf{P}')$ Lem. 2.42 guarantees similarly as above

$$\ln \frac{1}{2.4\gamma} \leq \sum\nolimits_{S \in [m]_k} \mathbb{E}_{\mathbf{P}}\left[T_S^{\mathcal{A}}\right] \mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right),$$

where we have written $\mathbf{P}_S$ resp. $\mathbf{P}'_S$ for $\mathbf{P}(\cdot|S)$ resp. $\mathbf{P}'(\cdot|S)$. Regarding arbitrariness of $\mathbf{P}$ and $\mathbf{P}'$ therein and using that $\mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] > 0$ and $\left(\mathbb{E}_{\mathbf{P}}[T_S^{\mathcal{A}}]/\mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}]\right)_{S \in [m]_k} \in \Delta_{[m]_k}$ hold trivially for any $\mathbf{P} \in PM_k^m$, we may follow an idea from Garivier and Kaufmann [2016] (cf. the proof of Thm. 1 therein) and estimate

$$\begin{aligned}
&\ln \frac{1}{2.4\gamma} \leq \min_{\mathbf{P}\in\mathfrak{P}^*(\varepsilon)} \inf_{\mathbf{P}'\in\mathfrak{P}_*(\neg\mathrm{GCW}(\mathbf{P}))} \sum_{S\in[m]_k} \mathbb{E}_{\mathbf{P}}\left[T_S^{\mathcal{A}}\right] \mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right) \\
&\leq \min_{\mathbf{P}\in\mathfrak{P}^*(\varepsilon)} \mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] \inf_{\mathbf{P}'\in\mathfrak{P}_*(\neg\mathrm{GCW}(\mathbf{P}))} \sum_{S\in[m]_k} \frac{\mathbb{E}_{\mathbf{P}}\left[T_S^{\mathcal{A}}\right]}{\mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}]} \mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right) \\
&\leq \sup_{\mathbf{w}\in\Delta_{[m]_k}} \min_{\mathbf{P}\in\mathfrak{P}^*(\varepsilon)} \mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] \inf_{\mathbf{P}'\in\mathfrak{P}_*(\neg\mathrm{GCW}(\mathbf{P}))} \sum_{S\in[m]_k} w_S \mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right). \quad (6.4)
\end{aligned}$$

Suppose $\mathbf{w} \in \Delta_{[m]_k}$ to be arbitrary but fixed for the moment. The identity

$$k = k\sum_{S\in[m]_k} w_S = \sum_{l\in[m]} \sum_{S\in[m]_k : l\in S} w_S$$

assures the existence of some $l = l(\mathbf{w}) \in [m]$ with $\sum_{S\in[m]_k:l\in S} w_S \leq \frac{k}{m}$. Abbreviate $\mathbf{P} \coloneqq \mathbf{P}^{[l,\varepsilon]}$. After relabeling $[m]$ via ν_l^{-1}, we may assume w.l.o.g. $l = 1$ in the following, i.e. $\mathbf{P} = \mathbf{P}^{[1,\varepsilon]} \in \mathfrak{P}^*(\varepsilon)$. Define $\mathbf{P}' \in PM_k^m$ via

$$\mathbf{P}'(2|S) \coloneqq \frac{1+h-\varepsilon}{2}, \quad \mathbf{P}'(\min S \setminus \{2\} \,|\, S) \coloneqq \frac{1-h-\varepsilon}{2} \quad \text{and} \quad \mathbf{P}'(j|S) \coloneqq \frac{\varepsilon}{k-2}$$

for any $j \in S \setminus \{2, \min(S \setminus \{2\})\}$, if $2 \in S$, and

$$\mathbf{P}'(j|S) \coloneqq \mathbf{P}(j|S)$$

for any $j \in S$, if $2 \notin S$. From $\mathbf{P}' \in PM_k^m(\mathrm{GCW} \wedge \Delta^h)$ and $\mathrm{GCW}(\mathbf{P}') = 2 \neq 1 = \mathrm{GCW}(\mathbf{P})$ we infer $\mathbf{P}' \in \mathfrak{P}_*(\neg\mathrm{GCW}(\mathbf{P}))$. In case $\{1,2\} \not\subseteq S$, we have $\mathbf{P}(j|S) = \mathbf{P}'(j|S)$ for any $j \in S$ and thus $\mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right) = 0$. In the remaining case $\{1,2\} \subseteq S$ Lem. 2.43 allows us to estimate

$$\begin{aligned}
&\mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right) \\
&= \mathrm{KL}\left(\left(\frac{1+h-\varepsilon}{2}, \frac{1-h-\varepsilon}{2}, \frac{\varepsilon}{k-2}, \dots\right), \left(\frac{1-h-\varepsilon}{2}, \frac{1+h-\varepsilon}{2}, \frac{\varepsilon}{k-2}, \dots\right)\right) \\
&\leq h^2 \left(\frac{2}{1-h-\varepsilon} + \frac{2}{1+h-\varepsilon}\right) = \frac{4h^2(1-\varepsilon)}{(1-h-\varepsilon)(1+h-\varepsilon)}.
\end{aligned}$$

Regarding the choice of $l = 1$ we infer

$$\begin{aligned}
\sum_{S\in[m]_k} w_S \mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right) &= \sum_{S\in[m]_k:\{1,2\}\subseteq S} w_S \mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right) \\
&\leq \frac{4h^2(1-\varepsilon)}{(1-h-\varepsilon)(1+h-\varepsilon)} \sum_{S\in[m]_k:1\in S} w_S \\
&\leq \frac{4kh^2(1-\varepsilon)}{m(1-h-\varepsilon)(1+h-\varepsilon)}
\end{aligned}$$

and thus clearly

$$\mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] \sum_{S\in[m]_k} w_S \mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right) \leq \frac{4kh^2(1-\varepsilon)}{m(1-h-\varepsilon)(1+h-\varepsilon)} \mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}].$$

Since $\mathbf{w}$ was arbitrary and $\mathbf{P} = \mathbf{P}^{[l(\mathbf{w}),\varepsilon]}$, combining this with (6.4) yields

$$\begin{aligned}
&\ln \frac{1}{2.4\gamma} \\
&\le \sup_{\mathbf{w}\in\Delta_{[m]_k}} \min_{\mathbf{P}\in\mathfrak{P}^*(\varepsilon)} \mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}] \inf_{\mathbf{P}'\in\mathfrak{P}_*(\neg\mathrm{GCW}(\mathbf{P}))} \sum\nolimits_{S\in[m]_k} w_S \mathrm{KL}\left(\mathbf{P}_S, \mathbf{P}'_S\right) \\
&\le \sup_{\mathbf{w}\in\Delta_{[m]_k}} \mathbb{E}_{\mathbf{P}^{[l(\mathbf{w}),\varepsilon]}}[T^{\mathcal{A}}] \inf_{\mathbf{P}'\in\mathfrak{P}_*(\neg\mathrm{GCW}(\mathbf{P}^{[l(\mathbf{w}),\varepsilon]}))} \sum\nolimits_{S\in[m]_k} w_S \mathrm{KL}\left(\mathbf{P}_S^{[l(\mathbf{w}),\varepsilon]}, \mathbf{P}'_S\right) \\
&\le \frac{4kh^2(1-\varepsilon)}{m(1-h-\varepsilon)(1+h-\varepsilon)} \sup_{\mathbf{w}\in\Delta_{[m]_k}} \mathbb{E}_{\mathbf{P}^{[l(\mathbf{w}),\varepsilon]}}[T^{\mathcal{A}}] \\
&\le \frac{4kh^2(1-\varepsilon)}{m(1-h-\varepsilon)(1+h-\varepsilon)} \max_{l\in[m]} \mathbb{E}_{\mathbf{P}^{[l,\varepsilon]}}[T^{\mathcal{A}}].
\end{aligned}$$

As $\varepsilon \in (0, 1/2 \wedge (1-h))$ was arbitrary, we finally conclude

$$\begin{aligned}
\sup_{\mathbf{P}\in PM_k^m(\mathrm{GCW}\wedge\Delta^h)} \mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}}\right] &\ge \sup_{\varepsilon\in(0,1/2\wedge(1-h))} \max_{l\in[m]} \mathbb{E}_{\mathbf{P}^{[l,\varepsilon]}}\left[T^{\mathcal{A}}\right] \\
&\ge \sup_{\varepsilon\in(0,1/2\wedge(1-h))} \frac{m(1-h-\varepsilon)(1+h-\varepsilon)}{4kh^2(1-\varepsilon)} \ln\frac{1}{2.4\gamma} \\
&\ge \frac{m(1-h^2)}{4kh^2} \ln \frac{1}{2.4\gamma}.
\end{aligned}$$

□

In the case of dueling bandits ($k = 2$), the instance-dependent bound from Thm. 6.4 has apparently been a novel result when published in [Haddenhorst et al., 2021c][1]. It reduces to

$$\begin{aligned}
\mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}}\right] &\ge \ln\frac{1}{2.4\gamma} \sum\nolimits_{l\in[m]\setminus\{i\}} \frac{1}{\mathrm{KL}(\mathbf{P}(\cdot|\{i,l\}), \mathbf{P}^{[l]}(\cdot|\{i,l\}))} \\
&= \ln\frac{1}{2.4\gamma} \sum\nolimits_{l\in[m]\setminus\{i\}} \frac{1}{\mathrm{kl}(\mathbf{P}(i|\{i,l\}), \mathbf{P}(l|\{i,l\}))}
\end{aligned}$$

for any $\mathbf{P} \in PM_2^m(\mathrm{GCW} \wedge \Delta^h)$ with $\mathrm{GCW}(\mathbf{P}) = i$ and any solution $\mathcal{A}$ to $\mathcal{P}^{m,2,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^h)$. By means of this, we obtain the following worst-case sample complexity lower bound, which is by a factor $\frac{2(m-1)}{m}$ larger than the worst-case one stated in Thm. 6.4.

Corollary 6.5. *If $\mathcal{A}$ solves $\mathcal{P}^{m,2,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^h)$, then*

$$\sup_{\mathbf{P}\in PM_2^m(\mathrm{GCW}\wedge\Delta^h)} \mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}}\right] \ge \frac{(m-1)(1-h^2)}{4h^2} \ln\frac{1}{2.4\gamma}.$$

Proof. Define $\mathbf{P} \in PM_2^m(\mathrm{GCW} \wedge \Delta^h)$ via $\mathbf{P}(i|\{i,j\}) := \frac{1+h}{2}$ for any $1 \le i < j \le m$.

[1] So far, existing sample complexity lower bounds for solutions to $\mathcal{P}_2^{m,\gamma}(\Delta^h \wedge \mathrm{GCW})$ are either restricted to worst-case scenarios [Braverman and Mossel, 2008] or to the special case where $\mathbf{P}$ belongs to a Thurstone model [Ren et al., 2020] or a Plackett-Luce model [Saha and Gopalan, 2020b].

Thm. 6.4 and Lem. 2.43 allow us to infer

$$\begin{aligned}\mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}}\right] &\geq \frac{m-1}{\mathrm{kl}((1+h)/2,(1-h)/2)} \ln\frac{1}{2.4\gamma} \\ &\geq (m-1)\left(\frac{2h^2}{1-h}+\frac{2h^2}{1+h}\right)^{-1}\ln\frac{1}{2.4\gamma} \\ &= \frac{(m-1)(1-h^2)}{4h^2}\ln\frac{1}{2.4\gamma}.\end{aligned}$$

□

For $k \geq 3$, the worst-case bound in Thm. 6.4 is not a consequence of the instance-wise version, instead it requires a more involved proof than the latter. This is made formal in the next remark.

Remark 6.6. *The instance-wise bound in Thm. 6.4 appears to be maximal on an instance* $\mathbf{P} \in PM_k^m$ *defined via*

$$\mathbf{P}(m_S|S) := \frac{1-h+hk}{k} \quad \text{and} \quad \mathbf{P}(j|S) := \frac{1-h}{k} \text{ for each } j \in S \setminus \{m_S\}.$$

with $m_S := \min S$ *for each* $S \in [m]_k$*. Note that* $\mathbf{P}(m_S|S) = \mathbf{P}(j|S) + h$ *is fulfilled for each* $S \in [m]_k$, $j \in S \setminus \{m_S\}$*. Regarding the definition of* m_S *we thus have* $\mathbf{P} \in PM_k^m(\Delta^h)$ *with* $\mathrm{GCW}(\mathbf{P}) = 1$*. With* $\mathbf{P}^{[l]}(\cdot|S)$ *defined as in Thm. 6.4 we can estimate for each* $l \in \{2,\ldots,m\}$ *and* $S \in [m]_k$ *with* $l \in S \setminus \{m_S\}$ *via Lem. 2.43*

$$\begin{aligned}\mathrm{KL}\left(\mathbf{P}(\cdot|S),\mathbf{P}^{[l]}(\cdot|S)\right) &\leq \sum\nolimits_{j\in S}\frac{(\mathbf{P}(j|S)-\mathbf{P}^{[l]}(j|S))^2}{\mathbf{P}(j|S)} \\ &= \frac{(\mathbf{P}(m_S|S)-\mathbf{P}^{[l]}(m_S|S))^2}{\mathbf{P}^{[l]}(m_S|S)} + \frac{(\mathbf{P}(l|S)-\mathbf{P}^{[l]}(l|S))^2}{\mathbf{P}^{[l]}(l|S)} \\ &= \frac{(\mathbf{P}(m_S|S)-\mathbf{P}(l|S))^2}{\mathbf{P}(l|S)} + \frac{(\mathbf{P}(l|S)-\mathbf{P}(m_S|S))^2}{\mathbf{P}(m_S|S)} \\ &= \frac{(\mathbf{P}(m_S|S)-\mathbf{P}(l|S))^2(\mathbf{P}(m_S|S)+\mathbf{P}(l|S))}{\mathbf{P}(m_S|S)\mathbf{P}(l|S)} \\ &= \frac{h^2k(1-h+hk+1-h)}{(1-h+hk)(1-h)} \leq \frac{2kh^2}{1-h},\end{aligned}$$

where we have used $hk \geq 0$ *in the last step. Consequently, the instance-wise bound from Thm. 6.4 yields*

$$\mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}}\right] \geq \frac{(m-1)(1-h)}{2h^2k(k-1)}\ln\frac{1}{2.4\gamma} \in \Omega\left(\frac{m}{k^2h^2}\ln\frac{1}{\gamma}\right),$$

which is by a factor $\frac{1}{k}$ *asymptotically smaller than the worst-case bound from Thm. 6.4.*

As the next remark illustrates, for $m = k$ the instance-wise lower bound underlying Prop. 2.30 is apparently larger than that of Thm. 6.4. The reason is that the proof for the instance-wise bound in Thm. 6.4 is tailored to the problem class $PM_k^m(\Delta^h \wedge \mathrm{GCW})$ and consequently has to deal with combinatorial issues arising in case $k < m$.

Remark 6.7. *Suppose $\mathcal{A}$ solves $\mathcal{P}^{k,k,\gamma}_{\mathrm{GCWi}}(\Delta^h)$, let $\mathbf{p} \in \Delta^h_k$ and write $i := \mathrm{mode}(\mathbf{p})$. According to Prop. 2.30 we have*

$$\mathbb{E}_{\mathbf{p}}\left[T^{\mathcal{A}}\right] \geq \max_{l\in[m]\setminus\{i\}} \frac{1-2\gamma}{2\phi_{l,i}(\mathbf{p})(p_l+p_i)} \left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+\phi_{l,i}(\mathbf{p}))/(1/2-\phi_{l,i}(\mathbf{p})))} \right\rceil =: \mathrm{LB}_1(\mathbf{p},\gamma)$$

with $\phi_{l,i}(\mathbf{p}) := \frac{p_i - p_l}{2(p_l+p_i)}$, and Thm. 6.4 guarantees

$$\mathbb{E}_{\mathbf{p}}\left[T^{\mathcal{A}}\right] \geq \frac{\ln \frac{1}{2.4\gamma}}{k-1} \sum_{l\in[k]\setminus\{i\}} \left(p_l \ln\left(\frac{p_l}{p_i}\right) + p_i \ln\left(\frac{p_i}{p_l}\right)\right)^{-1} =: \mathrm{LB}_2(\mathbf{p},\gamma).$$

In an empirical study we observed $\mathrm{LB}_1(\mathbf{p},\gamma) > \mathrm{LB}_2(\mathbf{p},\gamma)$ for 1000 parameters $\mathbf{p}$ sampled iid and uniformly at random from Δ^0_k, for any $(k,\gamma) \in \{5,10,15\} \times \{0.01, 0.05, 0.1\}$. For example, we have

$$\mathrm{LB}_1((0.2,0.2,0.15,0.2,0.25),0.05) \approx 252 > 152.9 \approx \mathrm{LB}_2((0.2,0.2,0.15,0.2,0.25),0.05).$$

This indicates that the instance-wise lower bound of Prop. 2.30 is larger than that from Thm. 6.4.

Prop. 2.33 indicated that for the special case $m = k$, a $\ln\ln\frac{1}{h}$-factor is indispensable. The next theorem formally shows that an analogon is true in case $m \geq k$.

Theorem 6.8. *Suppose m, k and γ to be fixed. There is a family $\{\mathbf{P}^h\}_{h\in(0,1)}$ with $\mathbf{P}^h \in PM^m_k(\Delta^h)$ for all $h \in (0,1)$ such that any solution $\mathcal{A}$ to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\Delta^0)$ fulfills*

$$\limsup_{h\searrow 0} \frac{\mathbb{E}_{\mathbf{P}^h}[T^{\mathcal{A}}]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} > 0.$$

Proof of Thm. 6.8. Define $\mathbf{P}^h = \{\mathbf{P}^h(\cdot|S)\}_{S\in[m]_k} \in PM^m_k$ for any $h \in (-1,1)$ via

$$\mathbf{P}^h(i|S) := \begin{cases} 1/2 + h/2, & \text{if } i = \max S, \\ 1/2 - h/2, & \text{if } i = \min S, \\ 0, & \text{otherwise.} \end{cases}$$

and note that $\mathbf{P}^h \in PM^m_k(\Delta^{|h|})$ with $\mathrm{GCW}(\mathbf{P}^h) = m$ if $h > 0$ and $\mathrm{GCW}(\mathbf{P}^h) = 1$ if $h < 0$. Let $\mathcal{A}$ be any solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\Delta^0)$. Suppose $a \in [0,1]$ to be arbitrary and let $C \sim \mathrm{Ber}(a)$. Define $\mathcal{A}'$ to be that algorithm, which simulates $\mathcal{A}$ in the following way: If $\mathcal{A}$ makes its query $S_t \in [m]_k$, draw a sample C_t from the coin C and and provide as feedback

$$X_t = \begin{cases} \max S_t, & \text{if } C_t = 1, \\ \min S_t, & \text{if } C_t = 0, \end{cases}$$

and in case $\mathcal{A}$ terminates, terminate and output $\mathbf{D}(\mathcal{A}') = 0$ if $\mathbf{D}(\mathcal{A}) = m$ and $\mathbf{D}(\mathcal{A}') = 1$ otherwise.

In case $a = 1/2 \pm h/2$ for $h > 0$, the feedback observed by $\mathcal{A}$ is distributed as if generated by $\mathbf{P}^{\pm h} \in PM^m_k(\Delta^0)$. Since $\mathcal{A}$ solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\Delta^0)$, we have

$$\mathbb{P}_{1/2+h/2}(\mathbf{D}(\mathcal{A}') = 0) = \mathbb{P}_{\mathbf{P}^h}(\mathbf{D}(\mathcal{A}) = m) \geq 1-\gamma,$$
$$\mathbb{P}_{1/2-h/2}(\mathbf{D}(\mathcal{A}') = 1) = \mathbb{P}_{\mathbf{P}^{-h}}(\mathbf{D}(\mathcal{A}) \neq m) \geq \mathbb{P}_{\mathbf{P}^{-h}}(\mathbf{D}(\mathcal{A}) = 1) \geq 1-\gamma,$$

that is, $\mathcal{A}'$ is thus able to correctly decide $\mathbf{H}_0 : a > 1/2$ vs. $\mathbf{H}_1 : a < 1/2$ with error probability $\leq \gamma$ for any $a \neq 1/2$. Regarding that $\mathcal{A}'$ has by construction the same sample complexity as $\mathcal{A}'$,

$$\limsup_{h \searrow 0} \frac{\mathbb{E}_{\mathbf{P}^h}[T^{\mathcal{A}}]}{\frac{1}{h^2} \ln \ln \frac{1}{h}} = \limsup_{h \searrow 0} \frac{\mathbb{E}_{1/2+h/2}[T^{\mathcal{A}'}]}{\frac{1}{h^2} \ln \ln \frac{1}{h}} > 0$$

follows from Prop. 2.13. □

Note here that the lower bound from Thm. 6.8 is only valid for any fixed m and k, it does not provide a bound of order $\Omega(\frac{m}{kh^2} \ln \ln \frac{1}{h})$. We will briefly address this again in Sec. 6.4. We conclude this section with a lower bound for solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWv}}(\Delta^h \wedge \mathrm{GCW})$. Its proof is almost identical to the first part of the proof of Thm. 6.4.

Theorem 6.9. *Suppose $\mathcal{A}$ solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWv}}(\Delta^h \wedge \mathrm{GCW})$ and $\mathbf{P} \in PM^m_k(\Delta^h \wedge \mathrm{GCW})$ is fixed with $i = \mathrm{GCW}(\mathbf{P})$. For any $S \in [m]_k$ let m_S and $\mathbf{P}^{[l]}(\cdot|S)$ for $l \in [m] \setminus \{m_S\}$ be defined as in Thm. 6.4. Then,*

$$\mathbb{E}_{\mathbf{P}}\left[T^{\mathcal{A}(i)}\right] \geq \frac{\ln \frac{1}{2.4\gamma}}{k-1} \sum_{l \in [m] \setminus \{\mathrm{GCW}(\mathbf{P})\}} \min_{S \in [m]_k : l \in S \setminus \{m_S\}} \frac{1}{\mathrm{KL}(\mathbf{P}(\cdot|S), \mathbf{P}^{[l]}(\cdot|S))}.$$

Proof of Thm. 6.9. After relabeling $[m]$, we may suppose w.l.o.g. $\mathrm{GCW}(\mathbf{P}) = 1$ and that $\mathcal{A}$ is started with 1 as input. For convenience, write $\mathcal{A}$ for $\mathcal{A}(1)$. Let $\mathbf{P}^{[l]}$ be defined as in the proof of Thm. 6.4. Since $\mathcal{A}$ solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\Delta^h \wedge \mathrm{GCW})$ and $\mathrm{GCW}(\mathbf{P}^{[l]}) = 1$ iff $l = 1$, it correctly outputs $\mathbf{1}_{\{l=1\}}$ whenever started with $\mathbf{P}^{[l]}$. Consequently, the event $\mathcal{E} := \{\mathbf{D}(\mathcal{A}) = 1\} \in \mathcal{F}_{T^{\mathcal{A}}}$ fulfills

$$\mathbb{P}_{\mathbf{P}}(\mathcal{E}) \geq 1 - \gamma \quad \text{and} \quad \mathbb{P}_{\mathbf{P}}(\mathcal{E}) \leq \gamma.$$

Thus, the same argumentation as in the proof of Thm. 6.4 completes the proof. □

6.2. Upper Bounds on the GCW Identification Problem

Saha and Gopalan [2020b] introduced PAC-Wrapper, an algorithm able to identify the GCW under the Plackett-Luce assumption with (up to logarithmic terms) optimal instance-wise sample complexity. Instead of discussing it in detail, we restrict ourselves to state its theoretical guarantees.

Theorem 6.10. *The algorithm $\mathcal{A} :=$ PAC-Wrapper from [Saha and Gopalan, 2020b] solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{PL} \wedge \mathrm{GCW}^*)$ and fulfills for any $\boldsymbol{\theta} \in (0,1]^m$ with $1 = \theta_1 > \max_{j \geq 2} \theta_j$ the estimate*

$$\mathbb{P}_{\mathbf{P}(\boldsymbol{\theta})}\left(\mathbf{D}(\mathcal{A}) \in \mathrm{GCW}(\mathbf{P}) \text{ and } T^{\mathcal{A}} \leq t'(\boldsymbol{\theta}, k, \gamma)\right) \geq 1 - \gamma$$

with

$$t'(\boldsymbol{\theta}, k, \gamma) \in \mathcal{O}\left(\frac{\Theta_{[k]}}{k} \sum_{j=2}^{m} \frac{1}{(1-\theta_j)^2} \ln\left(\frac{k}{\gamma} \ln\left(\frac{1}{1-\theta_j}\right)\right)\right)$$

and $\Theta_{[k]} := \max_{S \in [m]_k} \sum_{a \in S} \theta_a$.

Proof. Confer Thm. 3 in [Saha and Gopalan, 2020b] and note that $\min_{j \geq 2} \frac{1}{(1-\theta_j)^2} \geq 1$ holds for any $\boldsymbol{\theta} \in (0,1]^m$ with $1 = \theta_1 > \max_{j \geq 2} \theta_j$. □

By translating the sample complexity result of PAC-Wrapper into our setting, we obtain the following result, which is also by Thm. 6.3 suggested to be optimal up to logarithmic factors. Its proof is a bit technical, but we provide it for the sake of completeness.

Theorem 6.11. *The solution* $\mathcal{A} := \text{PAC-Wrapper}$ *from [Saha and Gopalan, 2020b] solves* $\mathcal{P}^{m,k,\gamma}_{\text{GCWi}}(\text{PL} \wedge \text{GCW}^*)$ *s.t.*

$$\inf_{\mathbf{P} \in PM^m_k(\text{PL} \wedge h\text{GCW})} \mathbb{P}_{\mathbf{P}} \left(\mathbf{D}(\mathcal{A}) \in \text{GCW}(\mathbf{P}) \text{ and } T^{\mathcal{A}} \leq t'(m,h,k,\gamma) \right) \geq 1 - \gamma$$

holds with $t'(h,m,k,\gamma) \in \mathcal{O}\left(\frac{m\left(\frac{1}{k}+h\right)\ln\left(\frac{k}{\gamma}\ln\frac{1}{h}\right)}{kh^2} \right)$.

Proof of Thm. 6.11. Suppose $\gamma \in (0,1)$, $h \in (0,1)$ and $m,k \in \mathbb{N}_{\geq 2}$ with $k \leq m$ to be arbitrary but fixed for the moment and let $\mathcal{A} := \text{PAC-Wrapper}$ from [Saha and Gopalan, 2020b]. Recall the guarantees of $\mathcal{A}$ from Thm. 6.10. For $l \in \{2,\dots,k\}$ define $g_l : [0,1]^m \to \mathbb{R}$ via $g_l(\boldsymbol{\theta}) := h(1+\theta_2+\cdots+\theta_k)+\theta_l-1$ and denote by $\mathfrak{B}$ the set

$$\{\boldsymbol{\theta} \in (0,1]^m \,|\, 1 = \theta_1 > \theta_2 \geq \cdots \geq \theta_m \text{ and } \forall l \in \{2,\dots,k\} : g_l(\boldsymbol{\theta}) \leq 0\}.$$

Lem. 2.45 assures that any $\mathbf{P} \in PM^m_k(\text{PL})$ with $\text{GCW}(\mathbf{P}) = 1$ fulfills $\mathbf{P} \in PM^m_k(h\text{GCW})$ iff $\mathbf{P} = \mathbf{P}(\boldsymbol{\theta})$ for some $\boldsymbol{\theta} \in \mathfrak{B}$. Consequently, it is with regard to Thm. 6.10 sufficient to show that

$$\frac{\Theta_{[k]}}{k} \sum\nolimits_{j=2}^{m} \frac{1}{(1-\theta_j)^2} \ln\left(\frac{k}{\gamma}\ln\left(\frac{1}{1-\theta_j}\right)\right) \leq \frac{6m}{kh^2}\left(\frac{1}{k}+h\right)\ln\left(\frac{k}{\gamma}\ln\frac{1}{h}\right) \tag{6.5}$$

holds for any $\boldsymbol{\theta} \in \mathfrak{B}$. We prove this in several steps.

Claim 1: For any $\boldsymbol{\theta} \in \mathfrak{B}$ we have

$$\sum\nolimits_{j=2}^{k} \frac{1+\theta_2+\cdots+\theta_k}{(1-\theta_j)^2} \leq \frac{3(1+hk)}{h^2}. \tag{6.6}$$

Proof of Claim 1: Let $\mathfrak{B}'$ be the set of all $\boldsymbol{\theta} = (1,\theta_2,\dots,\theta_k)$ with $1 \geq \theta_2 \geq \cdots \geq \theta_k \geq 0$ and $g_l(\boldsymbol{\theta}) \leq 0$ for all $l \in \{2,\dots,k\}$. As $(1,\theta_2,\dots,\theta_k) \in \mathfrak{B}'$ holds for any $(1,\theta_2,\dots,\theta_m) \in \mathfrak{B}$, it is sufficient to show that (6.6) holds for any $\boldsymbol{\theta} = (1,\theta_2,\dots,\theta_k) \in \mathfrak{B}'$.

Claim 1a: For any $\boldsymbol{\theta} \in \mathfrak{B}'$ and $l \in \{2,\dots,k\}$ we have $\theta_l \leq 1-h$.
Proof: For $\boldsymbol{\theta} = (1,\theta_2,\dots,\theta_k) \in \mathfrak{B}'$ and $l \in \{2,\dots,k\}$ we have

$$0 \geq g_l(\boldsymbol{\theta}) = h(1+\theta_2+\cdots+\theta_k)+\theta_l-1 \geq h+\theta_l-1,$$

and thus $\theta_l \leq 1-h$. ♣

According to Claim 1a, $\mathfrak{B}'$ is a compact subset of $\{1\} \times [0,1-h]^{k-1}$. Consequently, the continuous function $f : \mathfrak{B}' \to \mathbb{R}$, $f(\boldsymbol{\theta}) := \sum_{j=2}^{k} \frac{1+\theta_2+\cdots+\theta_k}{(1-\theta_j)^2}$ is well-defined and takes its maximum on $\mathfrak{B}'$ in a point $\boldsymbol{\theta}^* \in \mathfrak{B}'$.

Claim 1b: There is some $j \in \{2,\dots,k\}$ s.t. $g_2(\boldsymbol{\theta}^*) = \cdots = g_j(\boldsymbol{\theta}^*) = 0$ and $\theta^*_{j+2} = \cdots = \theta^*_k = 0$.

Proof: To show indirectly the existence of some $j \in \{2, \dots, k\}$ with $g_j(\boldsymbol{\theta}^*) = 0$ assume on the contrary that $g_l(\boldsymbol{\theta}^*) < 0$ for any $l \in \{2, \dots, k\}$. Then, if $\varepsilon > 0$ is small enough, $\boldsymbol{\theta}_\varepsilon \coloneqq (1, \theta_2^* + \varepsilon, \theta_3^*, \dots, \theta_k^*)$ is an element of $\mathfrak{B}'$. Since

$$\frac{\partial f}{\partial \theta_2}(\boldsymbol{\theta}) = \frac{2\theta_2(1 + \theta_2 + \dots + \theta_k)}{(1 - \theta_2)^3} + \sum\nolimits_{l=2}^{k} \frac{1}{(1 - \theta_l)^2} > 0$$

holds for any $\boldsymbol{\theta}$ in the interior of $\mathfrak{B}'$, we would obtain $f(\boldsymbol{\theta}_\varepsilon) > f(\boldsymbol{\theta}^*)$ in contradiction to the optimality of $\boldsymbol{\theta}^*$. Hence, there has to be a $j \in \{2, \dots, k\}$ with $g_j(\boldsymbol{\theta}^*) = 0$. In case $j \geq 3$, we may infer from $g_{j-1}(\boldsymbol{\theta}^*) - g_j(\boldsymbol{\theta}^*) = \theta_{j-1}^* - \theta_j^* \geq 0$ inductively $0 = g_{j-1}(\boldsymbol{\theta}^*) = \dots = g_2(\boldsymbol{\theta}^*)$. It remains to prove $\theta_{j+2}^* = \dots = \theta_k^* = 0$. Assume this was not the case, i.e., $j \leq k - 2$ and $j' \coloneqq \max\{l \in \{2, \dots, k\} \,|\, \theta_l^* > 0\} \geq j + 2$. By definition of j we have $g_j(\boldsymbol{\theta}^*) < 0$. Consequently,

$$\boldsymbol{\theta}'_\varepsilon \coloneqq (1, \theta_2^*, \dots, \theta_j^*, \theta_{j+1}^* + \varepsilon, \theta_{j+2}^*, \dots, \theta_{j'}^* - \varepsilon, 0, \dots, 0)$$

is for small values of $\varepsilon \geq 0$ an element of $\mathfrak{B}'$. Using $\sum_{l=2}^{k} (\boldsymbol{\theta}'_\varepsilon)_k = \sum_{l=2}^{k} \theta_l^*$ we see that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} f(\boldsymbol{\theta}'_\varepsilon) = \frac{2}{(1 - \theta_{j+1}^* - \varepsilon)^3} - \frac{2}{(1 - \theta_{j'}^* + \varepsilon)^3},$$

which is due to $\theta_{j+1}^* \geq \theta_{j'}^*$ positive for small values of $\varepsilon > 0$. In particular, $f(\boldsymbol{\theta}'_\varepsilon) > f(\boldsymbol{\theta}'_0) = f(\boldsymbol{\theta}^*)$ holds for small $\varepsilon > 0$, which contradicts the optimality of $\boldsymbol{\theta}^*$. This completes the proof of Claim 1b. ♣

According to Claim 1b we may fix some $j \in \{2, \dots, k\}$ with $g_2(\boldsymbol{\theta}^*) = \dots = g_j(\boldsymbol{\theta}^*) = 0$ and $\theta_{j+2}^* = \dots = \theta_k^* = 0$. Since $g_l(\boldsymbol{\theta}^*) - g_{l'}(\boldsymbol{\theta}^*) = \theta_l^* - \theta_{l'}^* = 0$ holds for any $l, l' \in \{2, \dots, k\}$, we have $\theta_2^* = \dots = \theta_j^*$. From $0 \geq g_2(\boldsymbol{\theta}^*) \geq h(1 + (j - 1)\theta_2^*) + \theta_2^* - 1$ we infer

$$\theta_2^* = \dots = \theta_j^* \leq \frac{1 - h}{1 + (j - 1)h} = 1 - \frac{hj}{1 + h(j - 1)}.$$

Together with $\theta_j^* \geq \theta_{j+1}^* \geq 0 = \theta_{j+2}^* = \dots = \theta_k^*$ we obtain

$$\begin{aligned}
\frac{1 + \theta_2^* + \dots + \theta_k^*}{(1 - \theta_2^*)^2} &\leq \frac{1 + j\theta_2^*}{(1 - \theta_2^*)^2} \leq \frac{(1 + h(j - 1))^2}{h^2 j^2} \left(1 + \frac{j(1 - h)}{1 + h(j - 1)}\right) \\
&= \frac{(1 + h(j - 1))(1 - h + j)}{h^2 j^2} \leq 2 \left(\frac{1}{h^2 j} + \frac{h(j - 1)}{h^2 j}\right) \\
&\leq \frac{2}{h^2} \left(\frac{1}{j} + h\right),
\end{aligned}$$

where we have used that $1 - h + j \leq 2j$ holds trivially. Combining this with the fact that $g_2(\boldsymbol{\theta}^*) \leq 0$ implies $(1 + \theta_2^* + \dots + \theta_k^*) \leq \frac{1 - \theta_2^*}{h} \leq \frac{1}{h}$ yields

$$\begin{aligned}
f(\boldsymbol{\theta}^*) &= \sum\nolimits_{l=2}^{k} \frac{1 + \theta_2^* + \dots + \theta_k^*}{(1 - \theta_l^*)^2} \\
&\leq (1 + \theta_2^* + \dots + \theta_k^*) \left(\sum\nolimits_{l=2}^{j+1} \frac{1}{(1 - \theta_2^*)^2} + \sum\nolimits_{l=j+2}^{k} 1\right) \\
&\leq \frac{2j}{h^2} \left(\frac{1}{j} + h\right) + \frac{k - j - 1}{h} \leq \frac{3(1 + hk)}{h^2}.
\end{aligned}$$

Since $\boldsymbol{\theta}^*$ was a maximum point of f in $\mathfrak{B}'$, Claim 1 follows. ■

Claim 2: For any $\boldsymbol{\theta} \in \mathfrak{B}$ we have $\sum_{j=2}^{m} \frac{1}{(1-\theta_j)^2} \leq \frac{m-1}{k-1} \sum_{j=2}^{k} \frac{1}{(1-\theta_j)^2}$.
Proof of Claim 2: Using $1 \geq \theta_2 \geq \cdots \geq \theta_m$, this follows directly from comparing the $(m-1)(k-1)$ summands in $(k-1)\sum_{j=2}^{m} \frac{1}{(1-\theta_j)^2} = \sum_{j=2}^{m} \frac{1}{(1-\theta_j)^2} + \cdots + \sum_{j=2}^{m} \frac{1}{(1-\theta_j)^2}$ with those in $(m-1)\sum_{j=2}^{k} \frac{1}{(1-\theta_j)^2}$. ■

Claim 3: Inequality (6.5) holds for any $\boldsymbol{\theta} \in \mathfrak{B}$.
Proof of Claim 3: Let $\boldsymbol{\theta} \in \mathfrak{B}$ be fixed and note that $\Theta_{[k]} = 1 + \theta_2 + \cdots + \theta_k$ holds. From $1 \geq \theta_2 \geq \cdots \geq \theta_m \geq 0$ we get $\Theta_{[k]} \in [1, k]$. Together with $\frac{1-\theta_2}{\Theta_{[k]}} = \frac{h\Theta_{[k]} - g_2(\boldsymbol{\theta})}{\Theta_{[k]}} \geq h$ this shows $1 - \theta_j \geq 1 - \theta_2 \geq h$ and in particular $\ln \frac{1}{1-\theta_j} \leq \ln \frac{1}{h}$ for each $j \in \{2, \ldots, m\}$. In combination with Claims 1 and 2 this allows us to conclude

$$
\begin{aligned}
&\frac{\Theta_{[k]}}{k} \sum_{j=2}^{m} \frac{1}{(1-\theta_j)^2} \ln\left(\frac{k}{\gamma} \ln\left(\frac{1}{1-\theta_j}\right)\right) \\
&\leq \frac{1}{k} \ln\left(\frac{k}{\gamma} \ln \frac{1}{h}\right) \sum_{j=2}^{m} \frac{1 + \theta_2 + \cdots + \theta_k}{(1-\theta_j)^2} \\
&\leq \frac{m-1}{k(k-1)} \ln\left(\frac{k}{\gamma} \ln \frac{1}{h}\right) \sum_{j=2}^{k} \frac{1 + \theta_2 + \cdots + \theta_k}{(1-\theta_j)^2} \\
&\leq \frac{3(m-1)(1+hk)}{k(k-1)h^2} \ln\left(\frac{k}{\gamma} \ln \frac{1}{h}\right) \\
&\leq \frac{6m}{kh^2} \left(\frac{1}{k} + h\right) \ln\left(\frac{k}{\gamma} \ln \frac{1}{h}\right),
\end{aligned}
$$

where we have used that $\frac{m-1}{k-1} \leq \frac{2m}{k}$ holds due to $k \geq 2$. This completes the proof of Claim 3 and of the theorem. □

Algorithm 24 Dvoretzky–Kiefer–Wolfowitz tournament (DKWT) – Solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$

Input: $k, m \in \mathbb{N}$, $\gamma \in (0,1)$, sample access to $\mathbf{P} = \{\mathbf{P}(\cdot|S)\}_{S \in [m]_k}$
Initialization: $\mathcal{A}_{\mathrm{Die}} :=$ Alg. 7, choose $S_1 \in [m]_k$ arbitrary, $F_1 \leftarrow [m]$, $\gamma' \leftarrow \frac{\gamma}{\lceil m/(k-1) \rceil}$, $s \leftarrow 1$
▷ S_s : candidates in round s, F_s : remaining elements in round s, i_s : output of $\mathcal{A}_{\mathrm{Die}}$ in round s

```
while s ≤ ⌈m/(k−1)⌉ − 1 do
    i_s ← A_Die(γ', sample access to P(·|S_s))
    F_{s+1} ← F_s \ S_s
    Write F_{s+1} = {j_1, ..., j_{|F_{s+1}|}}.
    if |F_{s+1}| < k then
        Fix distinct j_{|F_{s+1}|+1}, ..., j_{k−1} ∈ [m] \ F_{s+1}.
    S_{s+1} ← {i_s, j_1, ..., j_{k−1}}
    s ← s + 1
i_s ← A_Die(γ', sample access to P(·|S_s))
return i_s
```

Next, we consider the problem $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$, for which we propose the Dvoretzky–Kiefer–Wolfowitz tournament (DKWT) algorithm (see Alg. 24). DKWT is a simple round-based procedure eliminating in each round those arms from a candidate set of possible GCWs that have been discarded by Alg. 7 with high confidence as being the GCW. In the following theorem we derive theoretical guarantees for DKWT.

Theorem 6.12. *Let $\mathcal{A}$ be* DKWT *(Alg. 24) called with the parameters $k, m \in \mathbb{N}$ with $k \le m$ and $\gamma \in (0,1)$. Then, $\mathcal{A}$ solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$ and fulfills*

$$\mathbb{P}_{\mathbf{P}}\left(\mathbf{D}(\mathcal{A}) = \mathrm{GCW}(\mathbf{P}) \text{ and } T^{\mathcal{A}} \le t'(\mathbf{P}, m, k, \gamma)\right) \ge 1 - \gamma$$

for any $\mathbf{P} = \{\mathbf{P}(\cdot|S)\}_{S \in [m]_k} \in PM^m_k(\mathrm{GCW} \wedge \Delta^0)$, where $t'(\mathbf{P}, m, k, \gamma)$ is given as

$$\max\left\{ \sum\nolimits_{s \le s'} t_0(h(\mathbf{P}(\cdot|B_s)), \gamma') \,:\, B_1, B_2, \ldots, B_{s'} \in [m]_k \text{ s.t. } \bigcup\nolimits_{s \le s'} B_s = [m] \right\} \quad (6.7)$$

with $s' := \lceil \frac{m}{k-1} \rceil$, $\gamma' := \frac{\gamma}{s'}$ and $t_0(h, \gamma)$ defined as in Prop. 2.40, i.e.,

$$t_0(h, \gamma) = \sum\nolimits_{s \le s_0(h)} \left\lceil \frac{8}{h_s^2} \ln \frac{4}{\gamma_s} \right\rceil \quad \text{with} \quad s_0(h) = \left\lceil \log_2 \frac{3}{h} \right\rceil - 1.$$

In particular, $\mathcal{A}$ fulfills

$$\mathbb{P}_{\mathbf{P}}\left(\mathbf{D}(\mathcal{A}) \in \mathrm{GCW}(\mathbf{P}) \text{ and } T^{\mathcal{A}} \le T'(h(\mathbf{P}), m, k, \gamma)\right) \ge 1 - \gamma$$

for all $\mathbf{P} \in PM^m_k(\mathrm{GCW} \wedge \Delta^0)$, where $T'(h, m, k, \gamma) \in \mathcal{O}\left(\frac{m}{kh^2}\left(\ln \frac{m}{k}\right)\left(\ln\ln \frac{1}{h} + \ln \frac{1}{\gamma}\right)\right)$.

Proof. Suppose $\mathbf{P} = \{\mathbf{P}(\cdot|S)\}_{S \in [m]_k} \in PM^m_k(\mathrm{GCW} \wedge \Delta^0)$ to be fixed and abbreviate $i := \mathrm{GCW}(\mathbf{P})$. Recall the internal values s, S_s and F_s of Alg. 24. If $\mathcal{A}$ terminates, then the value of s is $s' := \lceil \frac{m}{k-1} \rceil$. Let us write $\widetilde{\mathcal{A}}_s$ for the instance of Alg. 7, which is called with parameters m, γ' and sample access to $\mathbf{P}(\cdot|S_s)$ in Step 2 (or Step 9), i.e., we have $i_s = \mathbf{D}(\widetilde{\mathcal{A}}_s) \in S_s$ for each $s \le s'$. For $s \ge 2$, S_s and F_s depend on the outcome of $\widetilde{\mathcal{A}}_{s-1}$ and are thus random variables.

Claim 1: On the event $\{T^{\mathcal{A}} < \infty\}$ we have

(i) $F_{s'} = \emptyset$ and $\bigcup_{s \le s'} S_s = [m]$, i.e., $\sum_{s \le s'} t_0(\gamma', h(\mathbf{P}(\cdot|S_s))) \le t'(\mathbf{P}, m, k, \gamma)$ holds a.s.,

(ii) $\{\mathbf{D}(\mathcal{A}) \ne i\} \subseteq \bigcup_{s \le s'} \{\mathbf{D}(\widetilde{\mathcal{A}}_s) \ne \mathrm{mode}(\mathbf{P}(\cdot|S_s))\}$.

Proof of Claim 1: Suppose $T^{\mathcal{A}} < \infty$. Clearly, $|F_s|$ is monotonically decreasing in s. Whenever $|F_s| \ge k$, then $|S_s \cap F_s| \ge k-1$ and thus $|F_{s+1}| \le |F_s| - (k-1)$ are fulfilled. Hence, $|F_s| \le m - s(k-1)$ holds for any $s \le s'-1$. In particular, we have $|F_{s'-1}| \le k-1$, which implies $F_{s'} = \emptyset$.
From $[m] = F_0 \supseteq F_1 \supseteq \cdots \supseteq F_{s'} = \emptyset$ and $\forall s \le s' : F_{s+1} = F_s \setminus S_s$ we infer $\bigcup_{s \le s'} S_s = [m]$, which proves (i). Regarding that the implications

$$j \in S_s \setminus S_{s'} \;\Rightarrow\; \exists l \in \{0, \ldots, s'-s\} : j \in S_{s+l-1} \setminus S_{s+l}$$

and

$$j \in S_s \setminus S_{s+1} \;\Rightarrow\; j \ne i_s$$

are trivially fulfilled for all $j \in [m]$ and $s \in \{0, \dots, s'-1\}$, we obtain

$$\begin{aligned}\{i \notin S_{s'}\} &\subseteq \{\exists s < s' : i \in S_s \text{ and } i \notin S_{s+1}\} \\ &\subseteq \{\exists s < s' : i \in S_s \text{ and } i_s \neq i\}.\end{aligned}$$

Due to $\{i \in S_s \text{ and } i_s \neq i\} \subseteq \{\mathbf{D}(\widetilde{\mathcal{A}}_s) \neq \mathrm{mode}(\mathbf{P}(\cdot|S_s))\}$, this implies

$$\begin{aligned}\{\mathbf{D}(\mathcal{A}) \neq i\} &= \{i \in S_{s'} \text{ and } i \neq i_{s'}\} \cup \{i \notin S_{s'}\} \\ &\subseteq \bigcup\nolimits_{s \leq s'} \{i \in S_s \text{ and } i \neq i_s\} \\ &\subseteq \bigcup\nolimits_{s \leq s'} \{\mathbf{D}(\widetilde{\mathcal{A}}_s) \neq \mathrm{mode}(\mathbf{P}(\cdot|S_s))\}.\end{aligned}$$

■

Claim 2: We have the estimate

$$\mathbb{P}_{\mathbf{P}}\left(\exists s \leq s' : \mathbf{D}(\widetilde{\mathcal{A}}_s) \neq \mathrm{mode}(\mathbf{P}(\cdot|S_s))) \text{ or } T^{\widetilde{\mathcal{A}}_s} > t_0(\gamma', h(\mathbf{P}(\cdot|S_s)))\right) \leq \gamma.$$

Proof of Claim 2: For $s \leq s'$ let

$$E_s := \left\{\mathbf{D}(\widetilde{\mathcal{A}}_s) \neq \mathrm{mode}(\mathbf{P}(\cdot|S_s))) \text{ or } T^{\widetilde{\mathcal{A}}_s} > t_0(\gamma', h(\mathbf{P}(\cdot|S_s)))\right\}$$

denote the set, where $\mathcal{A}$ fails at round s in the sense that $\widetilde{\mathcal{A}}_s$ either makes an error in finding $\mathrm{mode}(\mathbf{P}(\cdot|S_s))$ or queries "too many" samples for this. For $B \in [m]_k$ and $s \leq s'-1$ with $\mathbb{P}_{\mathbf{P}}(\{S_s = B\} \cap \bigcap_{\tilde{s} \leq s-1} E_{\tilde{s}}^c) > 0$ we have with regard to Prop. 2.40

$$\begin{aligned}&\mathbb{P}_{\mathbf{P}}\left(E_s \,\middle|\, \{S_s = B\} \cap \bigcap\nolimits_{\tilde{s} \leq s-1} E_{\tilde{s}}^c\right) \\ &= \mathbb{P}_{\mathbf{P}(\cdot|B)}\left(\mathbf{D}(\widetilde{\mathcal{A}}_s) \neq \mathrm{mode}(\mathbf{P}(\cdot|B)) \text{ or } T^{\widetilde{\mathcal{A}}_s} > t_0(\gamma', h(\mathbf{P}(\cdot|B)))\right) \leq \gamma',\end{aligned}$$

where we have used that both $\bigcap_{\tilde{s} \leq s-1} E_{\tilde{s}}^c$ and the choice $\{S_s = B\}$ are independent of the samples observed by $\widetilde{\mathcal{A}}_s$. We conclude

$$\begin{aligned}&\mathbb{P}_{\mathbf{P}}\left(\bigcup\nolimits_{s \leq s'} E_s\right) = \mathbb{P}_{\mathbf{P}}\left(\bigcup\nolimits_{s \leq s'} E_s \setminus \left(\bigcup\nolimits_{\tilde{s} \leq s-1} E_{\tilde{s}}\right)\right) \\ &\leq \sum\nolimits_{s \leq s'} \sum\nolimits_{B \in [m]_k} \mathbb{P}_{\mathbf{P}}\left(E_s \cap \{S_s = B\} \cap \bigcap\nolimits_{\tilde{s} \leq s-1} E_{\tilde{s}}^c\right) \\ &= \sum\nolimits_{s \leq s'} \left[\sum\nolimits_B \mathbb{P}_{\mathbf{P}}\left(E_s \middle| \{S_s = B\} \cap \bigcap\nolimits_{\tilde{s} \leq s-1} E_{\tilde{s}}^c\right) \mathbb{P}_{\mathbf{P}}\left(\{S_s = B\} \cap \bigcap\nolimits_{\tilde{s} \leq s-1} E_{\tilde{s}}^c\right)\right] \\ &\leq \sum\nolimits_{s \leq s'} \gamma' \leq \gamma,\end{aligned}$$

where $\sum_B$ denotes the sum over all $B \in [m]_k$ with $\mathbb{P}_{\mathbf{P}}\left(\{S_s = B\} \cap \bigcap_{\tilde{s} \leq s-1} E_{\tilde{s}}^c\right) > 0$. ■

Now, let us define for $s \leq s'$ the events

$$\mathcal{R}_s := \left\{T^{\widetilde{\mathcal{A}}_s} \leq t_0(\gamma', h(\mathbf{P}(\cdot|S_s)))\right\}$$

and $\mathcal{R} := \bigcap_{s \leq s'} \mathcal{R}_s$. Due to $T^{\mathcal{A}} = \sum_{s \leq s'} T^{\widetilde{\mathcal{A}}_s}$ we have

$$\mathcal{R} \subseteq \left\{T^{\mathcal{A}} \leq \sum\nolimits_{s \leq s'} t_0(\gamma', h(\mathbf{P}(\cdot|S_s)))\right\} \subseteq \left\{T^{\mathcal{A}} < \infty\right\}.$$

The equality $\mathcal{R}^c = \bigcup_{s \le s'} \mathcal{R}_s^c$ together with Part (ii) of Claim 1 and Claim 2 let us infer

$$\begin{aligned}
&\mathbb{P}_{\mathbf{P}}\left(\{\mathbf{D}(\mathcal{A}) \neq i\} \cup \mathcal{R}^c\right) = \mathbb{P}_{\mathbf{P}}\left((\{\mathbf{D}(\mathcal{A}) \neq i\} \cap \mathcal{R}) \cup \mathcal{R}^c\right) \\
&\le \mathbb{P}_{\mathbf{P}}\left(\bigcup\nolimits_{s \le s'} \left\{\mathbf{D}(\widetilde{\mathcal{A}}_s) \neq \mathrm{mode}(\mathbf{P}(\cdot|S_s)))\right\} \cup \mathcal{R}_s^c\right) \\
&= \mathbb{P}_{\mathbf{P}}\left(\exists s \le s' : \mathbf{D}(\widetilde{\mathcal{A}}_s) \neq \mathrm{mode}(\mathbf{P}(\cdot|S_s))) \text{ or } T^{\widetilde{\mathcal{A}}_s} > t_0(\gamma', h(\mathbf{P}(\cdot|S_s)))\right) \\
&\le \gamma
\end{aligned}$$

and we can thus conclude with the help of Part (i) of Claim 1 that

$$\begin{aligned}
&\mathbb{P}_{\mathbf{P}}\left(\mathbf{D}(\mathcal{A}) = i \text{ and } T^{\mathcal{A}} \le t'(\mathbf{P}, m, k, \gamma)\right) \\
&\ge \mathbb{P}_{\mathbf{P}}\left(\mathbf{D}(\mathcal{A}) = i \text{ and } T^{\mathcal{A}} \le \sum\nolimits_{s \le s'} t_0(\gamma', h(\mathbf{P}(\cdot|S_s))\right) \\
&\ge \mathbb{P}_{\mathbf{P}}\left(\{\mathbf{D}(\mathcal{A}) = i\} \cap \mathcal{R}\right) \\
&\ge 1 - \gamma.
\end{aligned}$$

It remains to show the second statement of the theorem. By definition of $h(\mathbf{P})$ we have $h(\mathbf{P}(\cdot|S)) \ge h(\mathbf{P})$ for any $S \in [m]_k$, hence monotonicity of $t_0(h, \gamma)$ from Prop. 2.40 w.r.t. h shows us that $t_0(h(\mathbf{P}(\cdot|S)), \gamma) \ge t_0(h(\mathbf{P}), \gamma)$ for any $S \in [m]_k$. Thus, a look at (6.7) reveals that

$$t'(\mathbf{P}, m, k, \gamma) \le T'(h(\mathbf{P}), m, k, \gamma)$$

with $T'(h, m, k, \gamma) := \left\lceil \frac{m}{k-1} \right\rceil t_0\left(h, \frac{\gamma}{\lceil m/(k-1) \rceil}\right)$, which is according to Prop. 2.40 contained in $\mathcal{O}\left(\frac{m}{kh^2}\left(\ln \frac{m}{k}\right)\left(\ln\ln \frac{1}{h} + \ln \frac{1}{\gamma}\right)\right)$. $\square$

The result stated in Table 6.1 for (X) = (GCW $\wedge\, \Delta^0$) and (Y) = (Δ^h) follows from Thm. 6.12 by noting that $h(\mathbf{P}) \ge h$ holds for any $\mathbf{P} \in PM_k^m(h\mathrm{GCW} \wedge \Delta^h)$. Regarding Prop. 2.33, the additional factor $\ln\ln \frac{1}{h}$ in the upper bounds from Thm. 6.11 and Thm. 6.12 appears indispensable. Since $PM_k^m(\mathrm{PL} \wedge \mathrm{GCW}^*) \not\subseteq PM_k^m(\mathrm{GCW} \wedge \Delta^0)$ and $PM_k^m(\mathrm{PL} \wedge \mathrm{GCW}^*) \not\subseteq PM_k^m(\mathrm{GCW} \wedge \Delta^0)$ hold, a solution to $\mathcal{P}_{\mathrm{GCWi}}^{m,k,\gamma}(\mathrm{PL} \wedge \mathrm{GCW}^*)$ is in general not comparable with a solution to $\mathcal{P}_{\mathrm{GCWi}}^{m,k,\gamma}(\mathrm{GCW} \wedge \Delta^0)$, i.e., neither Thm. 6.11 nor Thm. 6.12 implies the other one.

Replacing $\mathrm{GCW} \wedge \Delta^0$ with the more restrictive assumption $h\mathrm{GCW} \wedge \Delta^0$ (as an assumption on $\mathbf{P}$) makes the GCW identification task much easier. This is similar to the case of $\mathcal{P}_{\mathrm{GCWi}}^{k,k,\gamma}(\Delta^h) = \mathcal{P}_{\mathrm{Die}}^{k,h,\gamma}$ and $\mathcal{P}_{\mathrm{GCWi}}^{k,k,\gamma}(\Delta^0) = \mathcal{P}_{\mathrm{Die}}^{k,0,\gamma}$ discussed in Sec. 2.3.2. For $\mathcal{P}_{\mathrm{GCWi}}^{m,k,\gamma}(h\mathrm{GCW} \wedge \Delta^0)$ we can modify Alg. 24 in order to incorporate the knowledge of h as follows: Choose in round s a query set $S_s \subseteq F_s$ (filled up with $|F_s| - k$ further elements from $[m] \setminus F_s$ if $|F_s| < k$) and execute Alg. 6 with parameters $\frac{h}{3}$, $\frac{\gamma}{\lceil m/(k-1) \rceil}$ and sample access to $\mathbf{P}(\cdot|S_s)$. In case Alg. 6 returns as decision an element $i \in S_s$, we let $F_{s+1} = F_s \setminus (S_s \setminus \{i\})$, and otherwise $F_{s+1} = F_s$. Then we proceed with the next round $s+1$. We repeat this procedure until $|F_s| = 1$, and return the unique element in F_s as the prediction for the GCW. Detailed pseudocode of this procedure is given as Alg. 25.

We conclude this section with a theorem on the theoretical guarantees of Alg. 25. For technical reasons, we consider in its proof also a modified version of Alg. 25, which we state as Alg. 26.

Algorithm 25 Solution to $\mathcal{P}^{m,k,\gamma}_{\text{GCWi}}(h\text{GCW} \wedge \Delta^0)$

Input: $k, m \in \mathbb{N}$, $\gamma \in (0,1)$, $h \in (0,1)$, sample access to $\mathbf{P} = \{\mathbf{P}(\cdot|S)\}_{S\in[m]_k}$,
Initialization: $\mathcal{A}_{\text{Die}} := \text{Alg. 6}$, $i_0 \leftarrow \text{UNSURE}$, $h' \leftarrow \frac{h}{3}$, $\gamma' \leftarrow \frac{\gamma}{\lceil m/(k-1)\rceil}$, let $S_1 \in [m]_k$ arbitrary, $F_1 \leftarrow [m]$, $s \leftarrow 1$

$\triangleright$ S_s : candidate set in round s, F_s : remaining elements in round s
$\triangleright$ $i_s \in S_s \cup \{\text{UNSURE}\}$: output of $\mathcal{A}_{\text{Die}}$ in round s

while $|F_s| > 0$ **do**
  $i_s \leftarrow \mathcal{A}_{\text{Die}}(h', \gamma', \text{sample access to } \mathbf{P}(\cdot|S_s))$
  $F_{s+1} \leftarrow F_s \setminus S_s$
  Write $F_{s+1} = \{j_1, \ldots, j_{|F_{s+1}|}\}$.
  if $|F_{s+1}| < k$ **then**
    Fix distinct $j_{|F_{s+1}|+1}, \ldots, j_k \in [m] \setminus (F_{s+1} \cup \{i_s\})$.
  if $i_s \in [m]$ **then** $S_{s+1} \leftarrow \{i_s, j_1, \ldots, j_{k-1}\}$
  else $S_{s+1} \leftarrow \{j_1, \ldots, j_k\}$
  $s \leftarrow s + 1$
$i_s \leftarrow \mathcal{A}_{\text{Die}}(h', \gamma', \text{sample access to } \mathbf{P}(\cdot|S_s))$
if $i_s \in [m]$ **then return** i_s
else return 1

Theorem 6.13. *Let $\mathcal{A}$ be Alg. 25 called with parameters $m, k \in \mathbb{N}$ with $k \leq m$ and $\gamma, h \in (0,1)$. Then, $\mathcal{A}$ solves $\mathcal{P}^{m,k,\gamma}_{\text{GCWi}}(h\text{GCW} \wedge \Delta^0)$ and terminates a.s. for any $\mathbf{P} \in PM^m_k(h\text{GCW} \wedge \Delta^0)$ before some time $t'(m,k,h,\gamma) \in \mathcal{O}\left(\frac{m}{kh^2} \ln \frac{m}{k\gamma}\right)$.*

Before proving Thm. 6.13, note that Thm. 6.4 shows that this solution to $\mathcal{P}^{m,k,\gamma}_{\text{GCWi}}(h\text{GCW} \wedge \Delta^0)$ is asymptotically optimal up to logarithmic factors in a worst-case sense w.r.t. $PM^m_k(h\text{GCW} \wedge \Delta^0)$.

Proof of Thm. 6.13. Let us define the random variable $s^{\mathcal{A}} := \min\{s \in \mathbb{N} \,|\, F_s = \emptyset\} \in \mathbb{N} \cup \{\infty\}$ and suppose $\mathbf{P} \in PM^m_k$ to be arbitrary but fixed for the moment.

Claim 1: We have $s^{\mathcal{A}} \leq s' := \lceil \frac{m}{k-1} \rceil$ a.s. w.r.t. $\mathbb{P}_{\mathbf{P}}$.
Proof of Claim 1: Assume on the contrary that $s^{\mathcal{A}} > s'$. Note that $|F_s|$ is monotonically decreasing in s. Whenever $|F_s| \geq k$, then $|S_s \cap F_s| \geq k-1$ and thus $|F_{s+1}| \leq |F_s| - (k-1)$ are fulfilled. Hence, $|F_s| \leq m - s(k-1)$ holds for any $s \leq s'-1$. In particular, we have $|F_{s'-1}| \leq k-1$, which implies $F_{s'} = \emptyset$, contradicting the assumption $s^{\mathcal{A}} > s'$. This proves that $s^{\mathcal{A}} \leq s'$ is fulfilled a.s. ■

Using that $\mathcal{A}$ makes exactly $s^{\mathcal{A}}$ calls of $\mathcal{A}_{\text{Die}}$ (i.e., Alg. 6) with parameters h', γ' and each such call is executed with a sample complexity of exactly $\left\lceil \frac{8}{h'^2} \ln \frac{4}{\gamma'} \right\rceil$, the total sample complexity of $\mathcal{A}$ is at most

$$s' \lceil 8 \ln(4/\gamma')/h'^2 \rceil = \left\lceil \frac{m}{k-1} \right\rceil \left\lceil \frac{72}{h^2} \ln \left(\frac{4\lceil m/(k-1)\rceil}{\gamma} \right) \right\rceil,$$

which is in $\mathcal{O}\left(\frac{m}{kh^2} \ln \frac{m}{k\gamma}\right)$ as $\max\{m, k, \frac{1}{h}, \frac{1}{\gamma}\} \to \infty$. It remains to prove correctness of $\mathcal{A}$. Write $\mathcal{A}'$ for Alg. 26 called with the same parameters as $\mathcal{A}$.

Claim 2: For any $\mathbf{P} \in PM_k^m$, we have

$$\mathbb{P}_{\mathbf{P}}\left(\mathbf{D}(\mathcal{A}) \neq \mathrm{GCW}(\mathbf{P})\right) = \mathbb{P}_{\mathbf{P}}\left(\mathbf{D}(\mathcal{A}') \neq \mathrm{GCW}(\mathbf{P})\right).$$

Proof of Claim 2: This follows directly from the fact that for any $S \in [m]_k$, different calls of $\mathcal{A}_{\mathrm{Die}}$ on $\mathbf{P}(\cdot|S)$ are by assumption executed on different samples of $\mathbf{P}(\cdot|S)$ and thus independent of each other. ■

This result shows that it is sufficient to prove correctness of $\mathcal{A}'$. In the following, we denote by s, i_s, F_s and S_s the internal statistics of $\mathcal{A}'$ and write $\widetilde{\mathcal{A}}_s$ for that instance of $\mathcal{A}_{\mathrm{Die}}$, which is executed in $\mathcal{A}'$ to determine i_s. Let $\mathbf{P} \in PM_k^m(h\mathrm{GCW} \wedge \Delta^0)$ be fixed and define $i := \mathrm{GCW}(\mathbf{P})$.

Claim 3: For all $s \leq s'$ we have

$$\mathbb{P}_{\mathbf{P}}\left(i \in S_s \text{ and } i_s \neq i\right) \leq \gamma'.$$

Proof of Claim 3: Suppose $B \in [m]_k$ with $i \in [m]$ and $\mathbb{P}_{\mathbf{P}}(S_s = B) > 0$ to be arbitrary but fixed for the moment. By assumption on $\mathbf{P}$ we have $\mathbf{P}(\cdot|B) \in \Delta_k^{3h'}$ and since $\widetilde{\mathcal{A}}_s$ is Alg. 6 executed with parameters h', γ' and sample access to $\mathbf{P}(\cdot|S_s)$ only, Lem. 2.37 assures

$$\begin{aligned}&\mathbb{P}_{\mathbf{P}}\left(i \in S_s \text{ and } i_s \neq i | S_s = B\right)\\ &= \mathbb{P}_{\mathbf{P}(\cdot|B)}(\text{Alg. 6 started with } h', \gamma' \text{ does not output } \mathrm{mode}(\mathbf{P}(\cdot|B))) \leq \gamma'.\end{aligned}$$

Claim 3 thus follows via summation over all such B. ■

On the event $\{T^{\mathcal{A}'} < \infty\}$, we infer from $[m] = F_0 \supseteq F_1 \supseteq \cdots \supseteq F_{s^{\mathcal{A}}} = \cdots = F_{s'} = \emptyset$ and $\forall s \leq s' : F_{s+1} = F_s \setminus S_s$ similarly as in the proof of Thm. 6.12

$$\{\mathbf{D}(\mathcal{A}') \neq i\} \subseteq \bigcup\nolimits_{s \leq s'} \{i \in S_s \text{ and } i_s \neq i\}.$$

As $T^{\mathcal{A}'} < \infty$ holds a.s. w.r.t. $\mathbb{P}_{\mathbf{P}}$, combining this with Claim 3 directly yields

$$\mathbb{P}_{\mathbf{P}}(\mathbf{D}(\mathcal{A}') \neq i) \leq \sum\nolimits_{s \leq s'} \gamma' = \gamma,$$

which completes the proof. □

6.3. Empirical Evaluation

In this section, we empirically investigate the GCW identification problem. We present experimental results on DKWT, which have already been published in [Haddenhorst et al., 2021c], and additionally also include as further solution to GCW identification a modification of Alg. 24, which uses as $\mathcal{A}_{\mathrm{Die}}$ – instead of Alg. 7 – the procedure PPR1v1 stated in Alg. 8. For the sake of completeness, we stated PPRT as Alg. 27. Since PPR1v1 solves $\mathcal{P}_{\mathrm{Die}}^{k,0,\gamma}$ (Prop. 2.41), a similar argumentation as in the proof of the correctness of DKWT (Thm. 6.12) reveals that PPRT solves $\mathcal{P}_{\mathrm{GCWi}}^{m,k,\gamma}(\mathrm{GCW} \wedge \Delta^0)$.
Throughout all experiments, if not specified differently in the pseudocode, every choice of an element within a specific set made by DKWT or PPRT is performed uniformly at random.

Algorithm 26 Modification of Alg. 25 for the proof of Thm. 6.13

Input: $k, m \in \mathbb{N}$, $\gamma \in (0,1)$, $h \in (0,1)$, sample access to $\mathbf{P} = \{\mathbf{P}(\cdot|S)\}_{S \in [m]_k}$,
Initialization: $\mathcal{A}_{\mathrm{Die}} \coloneqq$ Alg. 6, $i_0 \leftarrow$ UNSURE, $h' \leftarrow \frac{h}{3}$, $\gamma' \leftarrow \frac{\gamma}{\lceil m/(k-1) \rceil}$
$S_1 \leftarrow [k]$, $F_1 \leftarrow [m]$, $s \leftarrow 1$

```
Execute steps 1–8 of Alg. 25.
let s' ← ⌈m/(k−1)⌉
while s < s' do
    i_s ← A_Die(h', γ', sample access to P(·|S_s))
    F_{s+1} ← F_s, S_{s+1} ← S_s
    s ← s + 1
i_s ← A_Die(h', γ', sample access to P(·|S_s))
return i_s
```

Algorithm 27 PPRT – Solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$

Input: $k, m \in \mathbb{N}$, $\gamma \in (0,1)$, sample access to $\mathbf{P} = \{\mathbf{P}(\cdot|S)\}_{S \in [m]_k}$
Initialization: $\mathcal{A}_{\mathrm{Die}} \coloneqq$Alg. 8, and initialize S_1, F_1, γ' and s as in Alg. 24

```
Execute Steps 1–10 from Alg. 24.
```

6.3.1. Comparison of DKWT and PPRT with PAC-Wrapper

At first, we compare DKWT and PPRT with PAC-Wrapper (PW), which is the solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{PL})$ in [Saha and Gopalan, 2020b] underlying Thm. 6.11 and so far, to the best of our knowledge, the only solution in the literature for identifying the GCW in MDB with confidence $1-\gamma$ under some assumptions on $\mathbf{P}$. Table 6.2 resp. Table 6.3 show the results of both algorithms when started on an instance $\mathbf{P} \in PM^5_k(\mathrm{PL})$ with underlying PL-parameter $\boldsymbol{\theta} = (1, 0.8, 0.6, 0.4, 0.2)$ resp. $\boldsymbol{\theta} = (1, 2^{-1}, 2^{-2}, \ldots, 2^{-9})$ and $\gamma = 0.05$ resp. $\gamma = 0.1$, for different values of k. The observed termination time $T^{\mathcal{A}}$, the corresponding standard error (in brackets) and the accuracy are averaged over 10 repetitions. The fact that the observed sample complexities are not throughout decreasing in k is supposedly due to the large standard errors and the little number of repetitions.

All algorithms achieve the desired accuracy $\geq 95\%$ for Table 6.2 and $\geq 90\%$ for Table 6.3 in every case, respectively, but DKWT requires far less samples than PW to find the GCW, and PPRT requires even less samples. Note that the observed extremely large sample complexity of PW appears to be consistent with the experimental results in [Saha and Gopalan, 2020b] and is supposedly caused by multiple runs of a costly procedure PAC-Best-Item, which is based on applications of the Chernoff bound.

Next, we compare DKWT, PPRT and PW on the synthetic data considered in [Saha

Table 6.2.: Termination times of DKWT, PPRT and PW and on $\boldsymbol{\theta} = (1, 0.8, 0.6, 0.4, 0.2)$; all observed accuracies are 1.00.

k	DKWT	PW	PPRT
3	44293 (3695.6)	1631668498 (1453661392.0)	**2931** (452.1)
4	32427 (2516.2)	263543687 (127401593.7)	**3074** (357.3)

Table 6.3.: Termination times of DKWT, PPRT and PW on $\boldsymbol{\theta} = (1, 2^{-1}, 2^{-2}, \ldots, 2^{-9})$; all observed accuracies are 1.00.

k	DKWT	PW	PPRT
2	8310 (0.0)	2509460 (226634.0)	**290** (30.6)
3	4078 (348.9)	46277676 (30635546.4)	**229** (38.1)
4	3925 (1014.3)	775101 (108535.7)	**336** (47.1)
5	3397 (529.2)	6450264 (1363336.3)	**263** (56.0)
6	2213 (465.0)	130069344 (77405795.5)	**246** (28.8)
7	2856 (507.4)	253206333 (125199242.0)	**222** (23.7)
8	3817 (608.9)	27159632 (12458792.0)	**305** (65.1)
9	2855 (680.7)	146229360 (79427860.6)	**342** (32.2)

and Gopalan, 2020b], where PW has first been introduced. We restrict ourselves to $\boldsymbol{\theta}^{\text{arith}}, \boldsymbol{\theta}^{\text{geo}} \in [0,1]^{16}$ defined via

$$\theta_1^{\text{arith}} := 1, \quad \forall i \in [15] : \theta_{i+1}^{\text{arith}} := \theta_i^{\text{arith}} - 0.06,$$
$$\theta_1^{\text{geo}} := 1, \quad \forall i \in [15] : \theta_{i+1}^{\text{geo}} := \frac{4}{5} \cdot \theta_i^{\text{geo}},$$

because the other synthetic datasets considered in Fig. 2 of [Saha and Gopalan, 2020b] (i.e., **g1** and **b1**) are not in $PM_k^m(\text{GCW} \wedge \Delta^0)$, which is formally required for DKWT. For $\boldsymbol{\theta} \in \{\boldsymbol{\theta}^{\text{arith}}, \boldsymbol{\theta}^{\text{geo}}\}$ we execute DKWT and PPRT with $\gamma = 0.01$ for 1000 repetitions on feedback generated by $\mathbf{P}(\boldsymbol{\theta})$ and report the mean termination time (and standard error in brackets) as well as the observed accuracy in Table 6.4. A look at Fig. 2 of [Saha and Gopalan, 2020b] reveals that both DKWT and PPRT outperform PW on both datasets while still keeping its theoretical guarantees, and PPRT clearly outperforms DKWT.

Table 6.4.: Termination times of DKWT and PPRT on $\boldsymbol{\theta}^{\text{arith}}$ and $\boldsymbol{\theta}^{\text{geo}}$; all observed accuracies are 1.00.

	DKWT	PPRT
$\boldsymbol{\theta}^{\text{arith}}$	1277781 (22284.0)	71724 (917.8)
$\boldsymbol{\theta}^{\text{geo}}$	55132 (910.5)	4243 (61.7)

6.3.2. DKWT and PPRT versus SELECT, SEEBS and Explore-then-Verify

For $k = 2$, the GCW identification problem coincides with the CW identification problem in dueling bandits. Thus, we can compare DKWT and PPRT to state-of-the art solutions for finding the CW if it exists: SELECT [Mohajer et al., 2017], SEEBS [2] [Ren et al., 2020] and Explore-then-Verify (EtV) [Karnin, 2016]. Formally, SELECT requires $h \in (0,1)$ as a parameter as it solves $\mathcal{P}_{\text{GCWi}}^{m,2,\gamma}(\text{GCW} \wedge \Delta^h) = \mathcal{P}_{\text{CWi}}^{m,h,\gamma}(\text{CW})$, while DKWT, SEEBS and EtV solve the more challenging problem $\mathcal{P}_{\text{GCWi}}^{m,2,\gamma}(\text{GCW} \wedge \Delta^0) = \mathcal{P}_{\text{CWi}}^{m,0,\gamma}(\text{CW})$.

[2]We include SEEBS even though it techniqually requires $\mathbf{P} \in PM_2^m = \mathcal{Q}_m$ to fulfill SST and STI as defined in Sec. 1.2 and 2.5.1. In particular, SEEBS is only proven to identify the correct (G)CW with confidence $\geq 1 - \gamma$ for any $\mathbf{P}$ in a set $PM_2^m(\text{GCW} \wedge \Delta^0 \wedge \text{SST} \wedge \text{STI}) \subsetneq PM_2^m(\text{GCW} \wedge \Delta^0)$.

For this reason, we provide SELECT a priori the value of h, whereas DKWT, SEEBS, EtV and PPRT do not obtain this information.

Table 6.5 shows the observed termination times (and standard errors thereof in brackets) of the different algorithms compared on $PM_2^m(\text{GCW} \wedge \Delta^h)$ obtained for $\gamma = 0.05$ and different choices of m and h, the numbers are averaged over 100 repetitions for $m \in \{5, 10\}$ and over 10 repetitions for $m \in \{15, 20\}$. The accuracy of SELECT is 0.97, 0.99, 0.95, 0.98, 0.99 and 0.90 if (m, h) is $(5, 0.2)$, $(5, 0.1)$, $(10, 0.2)$, $(10, 0.15)$, $(10, 0.1)$ and $(20, 0.2)$, respectively, and all the other observed accuracies in the scenario of Table 6.5 are 1.00. DKWT clearly outperforms SEEBS and EtV in any case. Again, PPRT achieves an even better performance than DKWT, and it is for small values of h even better than SELECT, which is due to its unfair knowledge of h in any other case unsurprisingly superior to all other procedures. Overall, these results show that DKWT and PPRT are also well suited for the dueling bandit case.

Table 6.5.: Comparison of DKWT, SELECT, SEEBS and EtV and PPRT.

		$T^{\mathcal{A}}$				
m	h	DKWT	SELECT	SEEBS	EtV	PPRT
5	0.20	6010 (293.2)	**252** (4.2)	7305 (432.1)	8601 (589.2)	449 (26.8)
5	0.15	8874 (460.0)	**460** (7.3)	13393 (904.5)	11899 (986.9)	604 (45.6)
5	0.10	15769 (1457.1)	**989** (17.0)	19802 (1543.2)	260171 (210678.1)	1003 (83.4)
5	0.05	31454 (4127.4)	3924 (68.6)	36855 (3533.2)	156534 (115903.1)	**2080** (195.8)
10	0.20	14334 (492.8)	**565** (2.5)	16956 (617.9)	26115 (969.2)	963 (41.7)
10	0.15	18563 (734.5)	**1009** (4.2)	27527 (1126.7)	32548 (2514.6)	1433 (74.8)
10	0.10	33040 (1625.1)	**2245** (9.7)	47330 (2138.2)	68858 (11304.5)	2290 (124.3)
10	0.05	78660 (6517.2)	8971 (39.2)	83877 (5842.6)	220098 (92484.9)	**5442** (472.1)
15	0.20	21932 (1618.1)	**803** (13.9)	28605 (2161.5)	54197 (5307.3)	1596 (89.2)
15	0.15	27446 (2500.0)	**1436** (12.3)	38084 (4985.3)	78753 (27741.4)	2521 (408.6)
15	0.10	45737 (6709.6)	**3248** (20.7)	67383 (8117.1)	116014 (24282.2)	3324 (376.6)
15	0.05	114152 (18704.0)	12993 (82.7)	108738 (19780.4)	2804238 (2560594.1)	**11170** (1938.6)
20	0.20	32038 (1209.2)	**1154** (8.7)	40910 (2893.1)	78286 (3451.5)	2301 (161.2)
20	0.15	39792 (3923.6)	**2080** (12.6)	58793 (4828.0)	122582 (24065.7)	3144 (304.9)
20	0.10	87667 (13380.8)	**4616** (32.3)	105249 (13231.8)	631195 (281883.6)	5725 (426.5)
20	0.05	134628 (21743.3)	18375 (138.2)	164439 (30175.4)	2094505 (1694236.4)	**13924** (2197.2)

6.3.3. Comparison of DKWT with Alg. 25

Finally, we compare DKWT, PPRT and Alg. 25 by means of their average sample complexity and accuracy when executed on 1000 instances $\mathbf{P}$, which were drawn independently and uniformly at random from (a) $PM_k^5(\text{GCW} \wedge \Delta^h)$ and (b) $PM_k^5(h\text{GCW} \wedge \Delta^{0.01})$. We choose $\gamma = 0.05$, consider $h \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and restrict ourselves to $k \in \{2, 3, 4\}$, because $PM_5^5(\text{GCW} \wedge \Delta^h) = PM_5^5(h\text{GCW} \wedge \Delta^{0.01})$ holds for any $h \geq 0.01$. Similarly as in our comparison to SELECT, Alg. 25 is revealed the true value of h and started with this as parameter. The results are collected in (a) Table 6.6 and (b) Table 6.7. In any of the cases (a) and (b), DKWT apparently outperforms Alg. 25 if h is smaller than some threshold h_0, and the value of h_0 appears to be significantly larger for (a) than for (b). This indicates that Alg. 25 may be preferable over DKWT if $h(\mathbf{P})$ is small and $\mathbf{P} \in PM_k^m(\exists h'\text{GCW} \wedge \Delta^0)$ holds for some a priori known $h' \in (0, 1/2)$.

Moreover, PPRT clearly outperforms both competitors in (a), and in (b) it is in any case better than DKWT and also better than Alg. 25 except if h is large.

Table 6.6.: Termination times of DKWT, Alg. 25 and PPRT on $PM_k^5(\text{GCW} \wedge \Delta^h)$; all observed accuracies are 1.00.

k	h	DKWT	Alg. 25	PPRT
2	0.9	4155 (0.0)	2664 (0.0)	**63** (0.2)
2	0.7	4155 (0.0)	4405 (0.0)	**87** (0.6)
2	0.5	4155 (0.0)	8630 (0.0)	**135** (1.4)
2	0.3	4195 (12.7)	23970 (0.0)	**264** (3.9)
2	0.1	14729 (423.4)	215695 (0.0)	**1075** (29.7)
3	0.9	2298 (0.0)	1464 (0.0)	**40** (0.2)
3	0.7	2298 (0.0)	2418 (0.0)	**61** (0.5)
3	0.5	2298 (0.0)	4737 (0.0)	**100** (1.0)
3	0.3	2381 (17.5)	13155 (0.0)	**216** (3.2)
3	0.1	14933 (436.1)	118383 (0.0)	**1006** (26.5)
4	0.9	1428 (0.0)	1356 (0.0)	**28** (0.2)
4	0.7	1428 (0.0)	2238 (0.0)	**43** (0.4)
4	0.5	1428 (0.0)	4386 (0.0)	**75** (0.8)
4	0.3	1492 (15.0)	12183 (0.0)	**163** (2.2)
4	0.1	13449 (306.4)	109626 (0.0)	**786** (16.9)

Table 6.7.: Termination times of DKWT, Alg. 25 and PPRT on $PM_k^5(h\text{GCW} \wedge \Delta^{0.01})$; all observed accuracies are 1.00.

k	h	DKWT	Alg. 25	PPRT
2	0.9	53913 (7092.4)	**2477** (8.0)	4611 (651.9)
2	0.7	63647 (8322.8)	**4124** (13.0)	4388 (682)
2	0.5	54370 (6753.8)	8167 (24.2)	**4815** (776)
2	0.3	59488 (7738.0)	23275 (53.4)	**5024** (637.4)
2	0.1	60682 (7256.5)	214358 (236.4)	**4874** (512.9)
3	0.9	40359 (6188.7)	**1464** (0.0)	3296 (522.0)
3	0.7	27069 (3621.2)	**2418** (0.0)	3472 (558.3)
3	0.5	37362 (5774.2)	4737 (0.0)	**2057** (317.3)
3	0.3	31553 (4551.6)	13155 (0.0)	**3107** (440.3)
3	0.1	45929 (5277.3)	118383 (0.0)	**2766** (347.0)
4	0.9	24164 (4446.0)	**1356** (0.0)	1998 (373.9)
4	0.7	39088 (6293.2)	2238 (0.0)	**1986** (359.2)
4	0.5	31835 (5462.0)	4386 (0.0)	**2095** (342.4)
4	0.3	31796 (5131.8)	12183 (0.0)	**2775** (454.9)
4	0.1	48202 (5765.3)	109626 (0.0)	**2708** (391.0)

6.4. Discussion and Related Work

The multi-dueling bandits (MDB) setting has recently been introduced by Brost et al. [2016] and is used in several practically relevant applications such as algorithm configuration [El Mesaoudi-Paul et al., 2020] or online retrieval evaluation [Schuth et al., 2016]. Multiple works considered this framework under different names, e.g. Saha and Gopalan [2018] as *battling*, Agarwal et al. [2020] as *choice* and [Bengs and Hüllermeier, 2020] as *preselection bandits*, and Bengs et al. [2021] termed a generalization thereof as *preference-based multi-armed bandits* (PB-MAB).

Whereas for standard MABs, the canonical definition of the "best arm" is the arm with highest expected reward, the picture is less clear for its variants. A majority of papers on MDBs assume latent utility values for the arms and the feedback process to be coherent with a random utility model (RUM) initialized with these values [Ben-Akiva and Bierlaire, 1999]. This assumption quite naturally provides an underlying ordering over the arms – namely that one given by the utility values – and thus makes it easy to define an objective such as the best arm or the top-k arms. In such a scenario, the PB-MAB multi-dueling bandits problem was investigated with respect to various performance metrics such as the regret [Saha and Gopalan, 2019b, Bengs and Hüllermeier, 2020, Agarwal et al., 2020] or the sample complexity in an (ϵ, γ)-PAC setting [Saha and Gopalan, 2019c, 2020a,b]. In contrast to these works, we focused as Agarwal et al. [2020] on a generalized concept of the Condorcet winner (CW) from dueling bandits, the *generalized Condorcet winner* (GCW). If latent utility values for the arms and a RUM for the feedback process are assumed, the GCW coincides with that arm with highest utility. While in [Agarwal et al., 2020] the problem for finding this GCW is investigated in a regret minimization scenario, we are interested in the minimum sample complexity required to identify it with confidence $1 - \gamma$. In light of this, the work by Saha and Gopalan [2020b] is the most related to the setting of this chapter, although the authors assume a PL model, which is a special case of a RUM.

Apart from that, there are a number of similar problem scenarios, namely the stochastic click model (SCM) [Zoghi et al., 2017], the dynamic assortment problem (DAS) [Caro and Gallien, 2007] and the best-of-k-bandits [Simchowitz et al., 2016]. However, all these scenarios take into account other specific aspects in the modelling such as the order of the arms in the action subset (SCM), known revenues associated with the arms (DAS) or a so-called "no-choice option" (all three). Accordingly, these problem scenarios are fundamentally different from our learning scenario (see also Sec. 6.6 in [Bengs et al., 2021] for a more detailed discussion). The same is true for combinatorial bandits [Cesa-Bianchi and Lugosi, 2012], which also allow subsets of arms as actions, but differ fundamentally in the nature of feedback (quantitative vs. qualitative feedback).

The CW is also a prominent notion in the realm of social choice [Fishburn, 1974, Fishburn and Gehrlein, 1976], and there have been several generalizations of it in this field. For example, Saari [1992] introduced a weighted variant, where the weights control the relevance given to the ranking positions of the alternatives, while Meyers et al. [2014] defined the k-winner as an alternative that (in some appropriate sense) outperforms all other arms among any k alternatives. In contrast to our work, these papers focus on offline learning tasks and suppose full rankings over all alternatives to be given. The particular notion of the *generalized Condorcet winner* (GCW), which we introduced in Sec. 1.2 and that is used throughout this thesis, is basically taken from [Agarwal et al., 2020]. Regarding the DB setting as the MDB setting where the allowed multi-duels S are exactly those with $|S| = 2$, the GCW is indeed a generalization of the CW.

In this chapter, we solved multiple variants of the GCW identification problem which differ in the assumptions made on the feedback mechanism. We presented sophisticated instance-dependent sample complexity upper and lower bounds for several variants and presented the solution *Dvoretzky-Kiefer-Wolfowitz tournament* (DKWT) to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$, which is up to logarithmic factors asymptotically optimal in a worst-case sense w.r.t $PM^m_k(\mathrm{GCW} \wedge \Delta^h)$-instances. Its name comes from the fact that a major ingredient of DKWT is a solution $\mathcal{A}_{\mathrm{Die}}$ to $\mathcal{P}^{k,h,\gamma}_{\mathrm{Die}}$, whose correctness relies on the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality. When replacing $\mathcal{A}_{\mathrm{Die}}$ in the construction of DKWT by an alternative solution to $\mathcal{P}^{k,h,\gamma}_{\mathrm{Die}}$, e.g. one of those stated in [Shah et al., 2020] or [Jain et al., 2021], one obtains further solutions to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$. We saw in Sec. 2.3 and Sec. 6.3 that the particular choice $\mathcal{A}_{\mathrm{Die}} = \mathrm{PPR1v1}$ from [Jain et al., 2021] leads to better empirical results but comes with the cost of slightly weaker theoretical guarantees (in the form of larger sample complexity) for $\mathcal{P}^{k,h,\gamma}_{\mathrm{Die}}$ and thus also for GCW identification; more precisely, PPR1v1 comes with an additional factor $\ln k$ in its sample complexity bound whereas the DKW inequality allowed us to formulate a corresponding upper bound without this factor for Alg. 7, which is used as $\mathcal{A}_{\mathrm{Die}}$ in DKWT, cf. Prop. 2.40 and Prop. 2.41.

As DKWT solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW} \wedge \Delta^0)$, it also solves the easier problem $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(h\mathrm{GCW} \wedge \Delta^0)$. Nevertheless, modifications of DKWT allowed us to formulate yet another solution to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(h\mathrm{GCW} \wedge \Delta^0)$, for which we have proven a sample complexity upper bound that is valid in a worst-case sense w.r.t. $PM^m_k(h\mathrm{GCW} \wedge \Delta^0)$; the worst-case upper bound for DKWT only takes instances $\mathbf{P} \in PM^m_k(\mathrm{GCW} \wedge \Delta^h) \subsetneq PM^m_k(h\mathrm{GCW} \wedge \Delta^0)$ into account.

To informally summarize our results, one could say that GCW identification with confidence $1 - \gamma$ whilst assuming $\mathrm{GCW} \wedge \Delta^h$ or $h\mathrm{GCW} \wedge \Delta^0$ requires roughly $\widetilde{\Theta}(\frac{m}{kh^2} \ln \frac{1}{\gamma})$ samples in a worst-case sense. Interestingly, this problem apparently becomes (in a worst-case asymptotic sense) to a factor $\frac{1}{k} + h$ easier when including a Plackett-Luce assumption on the feedback mechanism. This exact term also appeared in Prop. 2.30 in a different context, but unfortunately, we were not able to explain this phenomenon. As already mentioned, almost all of the results presented in this chapter have been published in [Haddenhorst et al., 2021c] before. In fact, only the lower bound for GCW verification (Thm. 6.9) and the statement and empirical evaluation of PPRT (Alg. 27) are novel in this regard.

Ren et al. [2021] have analyzed GCW identification in a related MDB setting, where they allowed all queries in $\mathcal{S}' \coloneqq \{S \subseteq [m] \, : \, 2 \leq |S| \leq k\}$. Under the assumption that the underlying parameter $\mathbf{P} = \{\mathbf{P}(i|S)\}_{S \in \mathcal{S}', i \in S}$ fulfills, for some fixed $q \in (1/2, 1]$, $\max_{i \in S} \mathbf{P}(i|S) = q$ for all $S \in \mathcal{S}'$, they showed that identifying the GCW with confidence $1-\gamma$ based on winner feedback requires $\Omega(\frac{m}{k} \ln \frac{1}{\gamma})$ samples and can be done with $\mathcal{O}(\frac{m}{k} \ln \frac{\ln k}{\gamma})$ samples. For fixed $h \in (0, 1)$, our lower bound from Thm. 6.4 is also of order $\Omega(\frac{m}{k} \ln \frac{1}{\gamma})$ and thus basically coincides with their result, but our upper bound from Thm. 6.12 is of order $\mathcal{O}(\frac{m}{k} \ln \frac{m}{k} \ln \frac{1}{\gamma})$ and therefore slightly larger than theirs. This is supposedly due to the fact that the subtle differences in the learning scenario make the GCW identification problem easier in their setting than in ours.

Regarding Sec. 4.5 and Sec. 5.5, one may ask whether one could also reduce GCW identification to the pure exploration multi-armed bandits (PE-MAB) setting and obtain, similar as for CW testification and WST testing, a solution based on the general Sticky Track-and-Stop procedure from [Degenne and Koolen, 2019]. We suppose that this is not as easily possible as for the DB problems: The natural reduction from MDB to MAB is supposedly that one which models each query set $S \in [m]_k$ as an arm with corresponding reward distribution $\mathbf{P}(\cdot|S)$, which is a categorical distribution. As Degenne and Koolen [2019] restrict themselves to learning problems that can be defined in terms of the means of these reward distributions whereas the GCW is defined in terms of their modes, their results do not appear applicable for GCW identification in case $k \geq 3$.

To conclude this chapter, let us collect some suggestions how this line of research could be extended in the future. Similarly as in Ch. 4 and Ch. 5, a presumably natural question is whether a lower bound of order $\Omega_{\sup}(\frac{m}{kh^2}(\ln\ln\frac{1}{h})\ln\frac{1}{\gamma})$ might be shown for the quantity $\sup_{\mathbf{P}\in PM_k^m(\mathrm{GCW}\wedge\Delta^h)}\mathbb{E}[T^{\mathcal{A}}]$ if $\mathcal{A}$ solves $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW}\wedge\Delta^0)$. When trying to answer this question, we faced basically the same obstacles as discussed in Sec. 2.6 and Sec. 5.7.

We have already seen that and to which extend incorporating an additional Plackett-Luce assumption simplifies the GCW identification problem. In a similar fashion, one could investigate in which sense the incorporation of alternative assumptions has an influence on the required sample complexity. For example, one could ask whether assuming a more general random utility model (RUM) underlying the feedback already makes GCW identification easier, i.e., does there exist a solution $\mathcal{A}$ to $\mathcal{P}^{m,k,\gamma}_{\mathrm{GCWi}}(\mathrm{GCW}\wedge\Delta^0\wedge\mathrm{RUM})$ for which $\sup_{\mathbf{P}\in PM_k^m(\mathrm{GCW}\wedge\Delta^h\wedge\mathrm{RUM})}\mathbb{E}_{\mathbf{P}}[T^{\mathcal{A}}]$ is asymptotically smaller than $\widetilde{\Theta}(\frac{m}{kh^2}\ln\frac{1}{\gamma})$? Saha and Gopalan [2020a] have considered best-arm identification in MDB in an (ε,γ)-PAC scenario under such a RUM assumption, which might be a good starting point.

Moreover, the GCW identification problem itself may be modified in several ways. One could consider as the set $\mathcal{S}$ of all queries other choices than $[m]_k$ and e.g. allow all query sets $S \subseteq [m]$ of size $2 \leq |S| \leq k$; this particular variant has also been considered under the Plackett-Luce assumption in [Saha and Gopalan, 2019b, Agarwal et al., 2020] for regret-minimization or in [Saha and Gopalan, 2019c, Ren et al., 2021] in a PAC setting.

Above, we have already defined the problem GCW *verification*, and – similarly as for the CW in Ch. 4 – one could also formulate *checking*- and *testification*-variants for the GCW in MDB. Since verifying that $i \in [m]$ is the GCW of $\mathbf{P}$ formally requires verifying $i = \mathrm{argmax}_{j\in S}\mathbf{P}(j|S)$ for the $\binom{m-1}{k}$ many query sets $S \in [m]_k$ with $i \in S$, verification and testification of the GCW appear at first sight not efficiently solvable without further restrictions on the feedback mechanism.

Last but not least, one could address GCW identification and its counterparts under alternative restrictions on $\mathbf{P}$ that capture dependencies on the parameters $\mathbf{P}(\cdot|S)$ across the query sets S; for example, one might consider in this regard the *weak optimal set consistency* from [Yang et al., 2021], that has initially been defined in a regret minimization setting but is presumably also transferrable to γ-PAC learning, as well as further generalizations of the general RUM model, which have recently been introduced by Ghoshal and Saha [2022] and have already been analyzed in an (ε,γ)-PAC best-arm identification scenario.

7. Conclusion and Outlook

The dueling and multi-dueling variants of multi-armed bandits describe sequential learning scenarios, which attained increased attention in the recent years, and have possible real-world applications in the realm of algorithm configuration and online retrieval evaluation. Prevalent goals in this regard include the identification of a best arm, the top-k arms or even a ranking over all arms, and these are oftentimes achieved by means of algorithms that formally require some type of coherence of the underlying feedback mechanism to be fulfilled. Even though violations of such assumptions would result in the loss of the theoretical guarantees of the corresponding algorithm, the statistical testing of these coherences has not gained much attention so far. In this thesis, we discussed testing for the particular coherences of the existence of a best arm in form of a Condorcet winner (CW) and also for different types of stochastic transitivity in dueling bandits. Moreover, we analyzed the best-arm identification problem in multi-dueling bandits under several assumptions.

The CW is arguably the most intuitive notion of the best arm in dueling bandits, and many works tackle the CW identification task under the assumption that it exists. We introduced *CW testification* as combined testing for and verification of the CW: In case a CW exists, a solution to this problem shall find and return it, and otherwise it shall detect non-existence of it. Based on a deterministic sequential testing algorithm for the analogue deterministic problem, we developed *Noisy Tournament Sampling* as a solution with interleaved testing and verification of the CW under mild assumptions. A passive version thereof can be used to detect on-the-fly violations of the CW-assumption made by a dueling bandits algorithm, and an appropriate active version was shown to be optimal up to logarithmic terms in the worst-case sense and outperforms a naive two-stage approach consisting of a separated testing and verification phase. We provided instance-dependent sample complexity lower and upper bounds on solutions of CW testification and obtained via a reduction to pure-exploration multi-armed bandits (PE-MABs) further different but consistent bounds. In a similar fashion, we also discussed the related problems *CW checking*, *CW identification* and *CW verification* (the latter one both with and without assuming the existence of a CW).

Other prominent coherences apparent as modelling assumptions in the realm of dueling bandits are different types of stochastic transitivity. We showed that, from these, any other than weak stochastic transitivity (WST) is to some extent impossible to test for. Hence, we focused on WST and, by making use of insights on acyclicity testing of tournaments, we presented a sophisticated algorithm, which solves WST testing almost asymptotically optimal in a worst-case sense. Again, we gave instance-dependent sample complexity lower and upper bounds, compared our results to those obtainable from the literature on PE-MABs and also provided a passive testing procedure. Moreover, we approached WST testing via likelihood ratio test statistics for WST resp. $\neg$WST. This resulted on the one hand in suboptimal passive and active solutions to the problem, and on the other hand we derived asymptotic bounds on the tails of the LRT statistics that allowed us to formulate asymptotic size-α tests for WST resp. $\neg$WST.

Graph-theoretical considerations used for the construction of our dueling bandits algorithms led as by-product to further results on deterministic property testing of tournaments, which are of interest on their own. For example, we saw that testing acycilicity of a tournament with an even number of nodes in an online, sequential and deterministic manner can be done without querying all edges.

In the realm of multi-dueling bandits, we analyzed the problem to identify the generalized Condorcet winner (GCW) under several assumptions, all of which imply existence of the GCW. We restricted ourselves to the active scenario and stated instance-wise sample complexity lower and upper bounds, asymptotically (worst-case) almost optimal solutions and pointed out to which extent the problem gets easier when incorporating an additional Plackett-Luce assumption on the feedback. We prepared our analysis with a theoretical discussion on mode identification for categorical random variables, and some of the therein stated results may be of interest for themselves.

There are various possible directions for future work. As already mentioned in the discussion sections of the chapters, we formulated several apparently unsolved precise research questions such as "Which sample complexity is actually required for testing acyclicity of tournaments?". There, we also mentioned that some approaches are presumably also transferrable to other problems but have not been analyzed in full detail so far. This includes e.g. Sticky Track-and-Stop for CW verification, and testification of the underlying ranking in DB whilst assuming $\mathbf{Q} \in \mathcal{Q}_m^0$ to be WST. Even though testing for the validity of statistical assumptions on the data generating process (the environment) is important, since violations of these may easily lead to erroneaous outputs, it has received little attention in the field of (multi-)dueling bandits so far. In this regard there is much room for possibly interesting future work: Firstly, there are prominent assumptions such as a general underlying random utility model (RUM) or the existence of an ε-best arm that have not been tested in this thesis. Secondly, one could test combinations of different assumptions X and Y such as $\mathrm{X} \wedge \mathrm{Y}$ or $\mathrm{X} \vee \mathrm{Y}$ or "X whilst assuming Y to be true" – with $(\mathrm{X}, \mathrm{Y}) = (\mathrm{SST}, \mathrm{STI})$, the former variant would e.g. be of particular interest for validating whether the assumptions of SEEBS, which we compared our solution to in Sec. 6.3, are fulfilled. Thirdly, our impossibility results on testing for STI, SST, the assumption $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{Mal})$ etc. are mere worst-case statements w.r.t. the low-noise assumption and they do not show non-testibility per se, but there could still be the possibility to test these properties in a different manner, e.g. under different assumptions. And finally, the mere idea of testification could be generalized to "If a problem is solvable, solve it, otherwise detect non-solvability". In this way, several other testification problems may be considered in the dueling bandits scenario, e.g. testification of the top-k arms or testification of an ε-best arm, and testification might also be considered in the realm of multi-dueling bandits.

Last but not least, motivated by real-life scenarios, some works in the field of DB assume that multiple queries can be evaluated in parallel and then aim to minimize not only the total number of queries but also the number of parallel rounds. However, the notion of a *round* is not used consistently in the literature: Lin and Lu [2018] discuss Borda winner identification under the assumption that any pairwise distinct queries can be run in parallel, as it would e.g. be the case in a chess tournament. In contrast, Agarwal et al. [2022] tackle top-k identification in a scenario, where any arbitrary family of queries can be evaluated in parallel. It might be interesting to discuss the (M)DB problems treated in this thesis also under such kinds of parallelization assumptions, and at first sight both types of parallelization appear plausible and appealing for this purpose.

Bibliography

Arpit Agarwal, Nicholas Johnson, and Shivani Agarwal. Choice bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 18399–18410, 2020.

Arpit Agarwal, Sanjeev Khanna, and Prathamesh Patil. PAC top-k identification under SST in limited rounds. In *To appear in: Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

Nir Ailon, Kohei Hatano, and Eiji Takimoto. Bandit online optimization over the permutahedron. In *Proceedings of Conference on Algorithmic Learning Theory (ALT)*, pages 215–229, 2014a.

Nir Ailon, Zohar Karnin, and Thorsten Joachims. Reducing dueling bandits to cardinal bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 856–864, 2014b.

Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, 1999.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002a.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002b.

Ashwinkumar Badanidiyuru, Robert D. Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. In *Proceedings of IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 207–216, 2013.

Subhash Bagui and K. Mehra. Convergence of binomial to normal: Multiple proofs. *International Mathematical Forum*, 12:399–411, 2017.

Ramachandran Balasubramanian, Venkatesh Raman, and G. Srinivasaragavan. Finding scores in tournaments. *Journal of Algorithms*, 24(2):380–394, 1997.

T. Parker Ballinger and Nathaniel T. Wilcox. Decisions, error and heterogeneity. *The Economic Journal*, 107(443):1090–1105, 1997.

Akshay Balsubramani. Sharp finite-time iterated-logarithm martingale concentration. *CoRR*, abs/1405.2639, 2014.

Akshay Balsubramani, Zohar Karnin, Robert E. Schapire, and Masrour Zoghi. Instance-dependent regret bounds for dueling bandits. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 336–360, 2016.

Heinz Bauer and Robert B. Burckel. *Probability Theory*. De Gruyter Studies in Mathematics. De Gruyter, 1996.

Moshe Ben-Akiva and Michel Bierlaire. Discrete choice methods and their applications to short term travel decisions. In *Handbook of Transportation Science*, pages 5–33. Springer, 1999.

Viktor Bengs and Eyke Hüllermeier. Preselection bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 778–787, 2020.

Viktor Bengs, Róbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Preference-based online learning with dueling bandits: A survey. *Journal of Machine Learning Research*, 22(7):1–108, 2021.

Viktor Bengs, Aadirupa Saha, and Eyke Hüllermeier. Stochastic contextual dueling bandits under linear stochastic transitivity models. *CoRR*, abs/2202.04593, 2022.

Daniel Berend and Aryeh Kontorovich. On the convergence of the empirical distribution. *CoRR*, abs/1205.6711, 2012.

Jean-Claude Bermond. Ordres à distance minimum d'un tournoi et graphes partiels sans circuits maximaux. *Mathématiques et Sciences Humaines*, 37:5–25, 1972.

Donald A. Berry, Robert W. Chen, Alan Zame, David C. Heath, and Larry A. Shepp. Bandit problems with infinitely many arms. *The Annals of Statistics*, 25(5):2103–2116, 1997.

Béla Bollobás. *Extremal Graph Theory*. L.M.S. Monographs. Academic Press Inc., 1978.

Béla Bollobás and Stephen E. Eldridge. Packings of graphs and applications to computational complexity. *Journal of Combinatorial Theory, Series B*, 25(2):105–124, 1978.

Djallel Bouneffouf and Irina Rish. A survey on practical applications of multi-armed and contextual bandits. *CoRR*, abs/1904.10040, 2019.

Ralph Alan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

Jasmin Brandt, Björn Haddenhorst, Viktor Bengs, and Eyke Hüllermeier. Finding optimal arms in non-stochastic combinatorial bandits with semi-bandit feedback and finite budget. *CoRR*, abs/2202.04487, 2022.

Mark Braverman and Elchanan Mossel. Noisy sorting without resampling. In *Proceedings of Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 268–276, 2008.

Mark Braverman, Jieming Mao, and S. Matthew Weinberg. Parallel algorithms for select and partition with noisy comparisons. In *Proceedings of ACM Symposium on Theory of Computing*, pages 851–862, 2016.

Brian Brost, Yevgeny Seldin, Ingemar J. Cox, and Christina Lioma. Multi-dueling bandits and their application to online ranker evaluation. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2161–2166, 2016.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1): 1–122, 2012.

Róbert Busa-Fekete, Balázs Szörényi, Paul Weng, Weiwei Cheng, and Eyke Hüllermeier. Top-k selection based on adaptive sampling of noisy preferences. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1094–1102, 2013.

Róbert Busa-Fekete, Eyke Hüllermeier, and Balázs Szörényi. Preference-based rank elicitation using statistical models: The case of Mallows. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1071–1079, 2014a.

Róbert Busa-Fekete, Balázs Szörényi, and Eyke Hüllermeier. PAC rank elicitation through adaptive sampling of stochastic pairwise preferences. In *Proceedings of AAAI Conference on Artificial Intelligence*, pages 1701–1707, 2014b.

Róbert Busa-Fekete, Dimitris Fotakis, Balázs Szörényi, and Manolis Zampetakis. Identity testing for Mallows model. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Francesco Paolo Cantelli. Considerazioni sulla legge uniforme dei grandi numeri e sulla generalizzazione di un fondamentale teorema del sig. Paul Lévy. *Giornale dell'Istituto Italiano degli Attuari*, 4(3):327–350, 1933.

Marie Jean Antoine Nicolas Caritat. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. De l'Imprimerie royale, 1785.

Felipe Caro and Jérémie Gallien. Dynamic assortment with demand learning for seasonal consumer goods. *Management Science*, 53(2):276–292, 2007.

Alexandra Carpentier and Michal Valko. Simple regret for infinitely many armed bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1133–1141, 2015.

Daniel R. Cavagnaro and Clintin P. Davis-Stober. Transitive in our preferences, but transitive in different ways: An analysis of choice variability. *Decision*, 1(2):102, 2014.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Nicolò Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

Bangrui Chen and Peter I. Frazier. Dueling bandits with dependent arms. *CoRR*, abs/1605.08838, 2016.

Bangrui Chen and Peter I. Frazier. Dueling bandits with weak regret. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 731–739, 2017.

Lijie Chen and Jian Li. On the optimal sample complexity for best arm identification. *CoRR*, abs/1511.03774, 2015.

Lijie Chen, Jian Li, and Mingda Qiao. Towards instance optimal bounds for best arm identification. In *Proceedings of Conference on Learning Theory (COLT)*, volume 65, pages 535–592, 2017.

Xi Chen and Yining Wang. A note on a tight lower bound for capacitated MNL-bandit assortment selection models. *Operations Research Letters*, 46(5):534–537, 2018.

Xi Chen, Paul N. Bennett, Kevyn Collins-Thompson, and Eric Horvitz. Pairwise ranking aggregation in a crowdsourced setting. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*, pages 193–202, 2013.

Xi Chen, Yuanzhi Li, and Jieming Mao. A nearly instance optimal algorithm for top-k ranking under the Multinomial Logit Model. In *Proceedings of Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2504—-2522. Society for Industrial and Applied Mathematics, 2018.

Weiwei Cheng, Krzysztof Dembczyński, and Eyke Hüllermeier. Label ranking methods based on the Plackett-Luce model. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 215–222, 2010.

Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

Arthur H. Copeland. A reasonable social welfare function. *Seminar on Applications of Mathematics to Social Sciences. University of Michigan, Ann Arbor*, 1951.

Constantinos Daskalakis and Yasushi Kawase. Optimal stopping rules for sequential hypothesis testing. In *Proceedings of Annual European Symposium on Algorithms (ESA)*, volume 87, pages 32:1–32:14, 2017.

Donald Davidson and Jacob Marschak. Experimental tests of a stochastic decision theory. Technical report 17, Applied Mathematics and Statistics Laboratory, 1958. URL `https://web.stanford.edu/group/csli-suppes/techreports/IMSSS_17.pdf`.

Rémy Degenne and Wouter M. Koolen. Pure exploration with multiple correct answers. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

L. Devroye. The equivalence of weak, strong and complete convergence in l_1 for kernel density estimates. *Annals of Statistics*, 11:896–904, 1983.

Miroslav Dudík, Katja Hofmann, Robert E. Schapire, Aleksandrs Slivkins, and Masrour Zoghi. Contextual dueling bandits. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 563–587, 2015.

A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642–669, 1956.

Stephen G. Eick. The two-armed bandit with delayed responses. *The Annals of Statistics*, 16(1):254–264, 1988.

Adil El Mesaoudi-Paul, Dimitri Weiß, Viktor Bengs, Eyke Hüllermeier, and Kevin Tierney. Pool-based realtime algorithm configuration: A preselection bandit approach. In *Learning and Intelligent Optimization*, pages 216—-232. Springer, 2020.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. PAC bounds for multi-armed bandit and Markov decision processes. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 255–270, 2002.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(39):1079–1105, 2006.

Moein Falahatgar, Yi Hao, Alon Orlitsky, Venkatadheeraj Pichapati, and Vaishakh Ravindrakumar. Maxing and ranking with few assumptions. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 7060–7070, 2017a.

Moein Falahatgar, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh. Maximum selection and ranking under noisy comparisons. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1088–1096, 2017b.

Moein Falahatgar, Ayush Jain, Alon Orlitsky, Venkatadheeraj Pichapati, and Vaishakh Ravindrakumar. The limits of maxing, ranking, and preference learning. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1426–1435, 2018.

Roger H. Farrell. Asymptotic behavior of expected sample size in certain one sided tests. *The Annals of Mathematical Statistics*, 35(1):36–72, 1964.

Thomas S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Probability and Mathematical Statistics. Academic Press, 1967.

Peter C. Fishburn. Binary choice probabilities: On the varieties of stochastic transitivity. *Journal of Mathematical Psychology*, 10(4):327–352, 1973.

Peter C. Fishburn. Paradoxes of voting. *American Political Science Review*, 68(2):537–546, 1974.

Peter C. Fishburn and Jean-Claude Falmagne. Binary choice probabilities and rankings. *Economics Letters*, 31(2):113–117, 1989.

Peter C. Fishburn and William V. Gehrlein. Borda's rule, positional voting, and Condorcet's simple majority principle. *Public Choice*, 28:79–88, 1976.

Michael A. Fligner and Joseph S. Verducci. *Probability Models and Statistical Analyses for Ranking Data*. Springer, 1993.

Johannes Fürnkranz and Eyke Hüllermeier. *Preference Learning*. Springer, 2011.

Manegueu Anne Gael, Claire Vernade, Alexandra Carpentier, and Michal Valko. Stochastic bandits with arm-dependent delays. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 3348–3356, 2020.

Pratik Gajane and Tanguy Urvoy. Utility-based dueling bandits as a partial monitoring game. *CoRR*, abs/1507.02750, 2015.

Aurélien Garivier. Informational confidence bounds for self-normalized averages and applications. In *Proceedings of IEEE Information Theory Workshop (ITW)*, pages 1–5, 2013.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 998–1027, 2016.

Bhaskar Kumar Ghosh and Pranab Kumar Sen. *Handbook of sequential analysis.* CRC Press, 1991.

Suprovat Ghoshal and Aadirupa Saha. Exploiting correlation to achieve faster learning rates in low-rank preference bandits. *CoRR*, abs/2202.11795, 2022.

Shengbo Guo, Scott Sanner, Thore Graepel, and Wray Buntine. Score-based Bayesian skill learning. In *Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD)*, pages 106–121, 2012.

Shubham Gupta and Aadirupa Saha. Optimal and efficient dynamic regret algorithms for non-stationary dueling bandits. *CoRR*, abs/2111.03917, 2021.

Björn Haddenhorst, Eyke Hüllermeier, and Martin Kolb. Generalized transitivity: A systematic comparison of concepts with an application to preferences in the Babington Smith model. *International Journal of Approximate Reasoning*, 119:373–407, 2020.

Björn Haddenhorst, Viktor Bengs, Jasmin Brandt, and Eyke Hüllermeier. Testification of Condorcet winners in dueling bandits. In *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)*, 2021a.

Björn Haddenhorst, Viktor Bengs, and Eyke Hüllermeier. On testing transitivity in online preference learning. *Machine Learning*, page 2063–2084, 2021b.

Björn Haddenhorst, Viktor Bengs, and Eyke Hüllermeier. Identification of the generalized Condorcet winner in multi-dueling bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 25904–25916, 2021c.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Richard C. Holt and Edward M. Reingold. On the time required to detect cycles and connectivity in graphs. *Mathematical Systems Theory*, 6:103–106, 1972.

Steven R. Howard and Aaditya Ramdas. Sequential estimation of quantiles with applications to A/B-testing and best-arm identification. *CoRR*, abs/1906.09712, 2021.

Geoffrey Iverson and Jean-Claude Falmagne. Statistical issues in measurement. *Mathematical Social Sciences*, 10(2):131–153, 1985.

Shubham Jain, Rohan Shah, Sanit Gupta, Denil Mehta, Inderjeet Nair, Jian Vora, Sushil Khyalia, Sourav Das, Vinay J. Ribeiro, and Shivaram Kalyanakrishnan. PAC mode estimation using PPR martingale confidence sequences. *CoRR*, abs/2109.05047, 2021.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'UCB: An optimal exploration algorithm for multi-armed bandits. *Journal of Machine Learning Research*, 35, 2013.

Kevin Jamieson, Sumeet Katariya, Atul Deshpande, and Robert Nowak. Sparse dueling bandits. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 416–424, 2015.

Kevin G. Jamieson and Robert Nowak. Active ranking using pairwise comparisons. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 2240–2248, 2011.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 227—-234, Madison, WI, USA, 2012. Omnipress.

Zohar Karnin. Verification based solution for structured MAB problems. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 145–153, 2016.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1238–1246, 2013.

Richard M. Karp and Robert Kleinberg. Noisy binary search and its applications. In *Proceedings of Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 881—-890, 2007.

Emilie Kaufmann and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In *Proceedings of Annual Conference on Learning Theory (COLT)*, volume 30 of *Proceedings of Machine Learning Research*, pages 228–251, 2013.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

Donald Knuth. *The Art Of Computer Programming: Sorting And Searching*, volume 3. Addison-Wesley, 1973.

Patrick Kolpaczki, Viktor Bengs, and Eyke Hüllermeier. Non-stationary dueling bandits. *CoRR*, abs/2202.00935, 2022.

Junpei Komiyama, Junya Honda, Hisashi Kashima, and Hiroshi Nakagawa. Regret lower bound and optimal algorithm in dueling bandit problem. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 1141–1154, 2015.

Junpei Komiyama, Junya Honda, and Hiroshi Nakagawa. Copeland dueling bandit problem: Regret lower bound, optimal algorithm, and computationally efficient algorithm. In

Proceedings of International Conference on Machine Learning (ICML), pages 1235–1244, 2016.

Anna Korba, Stephan Clémençon, and Eric Sibony. A learning theory of ranking aggregation. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1001–1010, 2017.

Michael R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, USA, 2008.

Wataru Kumagai. Regret analysis for continuous dueling bandit. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 1488–1497, 2017.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvári. Tight regret bounds for stochastic combinatorial semi-bandits. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 38, pages 535–543, 2015.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Chang Li, Ilya Markov, Maarten De Rijke, and Masrour Zoghi. MergeDTS: A method for effective large-scale online ranker evaluation. *ACM Trans. Inf. Syst.*, 38(4):1–28, 2020.

Chuang-Chieh Lin and Chi-Jen Lu. Efficient mechanisms for peer grading and dueling bandits. In *Proceedings of Asian Conference on Machine Learning (ACML)*, Proceedings of Machine Learning Research, pages 740–755, 2018.

Robert Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

Colin L. Mallows. Non-null ranking models. *Biometrika*, 44(1):114–130, 1957.

Shie Mannor and John N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5:623–648, 2004.

P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990.

Pascal Massart. *Concentration inequalities and model selection*, volume 6. Springer, 2007.

Lucas Maystre and Matthias Grossglauser. Just sort it! A simple and effective approach to active preference learning. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 2344–2353, 2017.

Timothy P. McNamara and Vaibhav A. Diwadkar. Symmetry and asymmetry of human spatial memory. *Cognitive Psychology*, 34(2):160–190, 1997.

Aaron Meyers, Michael E. Orrison, Jennifer Townsend, Sarah Wolff, and Angela Wu. Generalized Condorcet winners. *Social Choice and Welfare*, 43(1):11–27, 2014.

Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis.* Cambridge University Press, 2017.

Soheil Mohajer, Changho Suh, and Adel Elmahdy. Active learning for top-k rank aggregation from noisy comparisons. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 2488–2497, 2017.

John W. Moon. *Topics on Tournaments in Graph Theory.* Dover Publications, 2015.

Guillermo Owen. *Game Theory.* Academic Press, 1982.

Erol Peköz, Sheldon M. Ross, and Zhengyu Zhang. Dueling bandit problems. *Probability in the Engineering and Informational Sciences*, pages 1—12, 2020.

Robert L. Plackett. The analysis of permutations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 24(1):193–202, 1975.

Ariel D. Procaccia. A note on the query complexity of the Condorcet winner problem. *Information Processing Letters*, 108(6):390–393, 2008.

Siddartha Ramamohan, Arun Rajkumar, and Shivani Agarwal. Dueling bandits: Beyond Condorcet winners to general tournament solutions. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 1253–1261, 2016.

Charvi Rastogi, Sivaraman Balakrishnan, Nihar Shah, and Aarti Singh. Two-sample testing on pairwise comparison data and the role of modeling assumptions. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, pages 1271–1276, 2020.

Wenbo Ren, Jia Liu, and Ness Shroff. On sample complexity upper and lower bounds for exact ranking from noisy comparisons. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 10014–10024, 2019a.

Wenbo Ren, Jia Liu, and Ness Shroff. On sample complexity upper and lower bounds for exact ranking from noisy comparisons. *CoRR*, abs/1909.03194v3, 2019b.

Wenbo Ren, Jia Liu, and Ness Shroff. The sample complexity of best-k items selection from pairwise comparisons. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 8051–8072, 2020.

Wenbo Ren, Jia Liu, and Ness Shroff. Sample complexity bounds for active ranking from multi-wise comparisons. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 4290–4300, 2021.

Ronald L. Rivest and Jean Vuillemin. A generalization and proof of the Aanderaa-Rosenberg conjecture. In *Proceedings of Annual ACM Symposium on Theory of Computing (STOC)*, pages 6—11, 1975.

Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

Sheldon M. Ross. *Stochastic Processes*. Wiley Series in Probability and Statistics: Probability and Statistics. Wiley, 1996.

Donald G. Saari. Symmetry extensions of "neutrality" I: Advantage to the Condorcet loser. *Social Choice and Welfare*, 9(4):307–336, 1992.

Horst Sachs. *Einführung in die Theorie der endlichen Graphen*. Hanser, 1971.

Aadirupa Saha. Optimal algorithms for stochastic contextual preference bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 30050–30062, 2021.

Aadirupa Saha and Aditya Gopalan. Battle of bandits. In *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)*, 2018.

Aadirupa Saha and Aditya Gopalan. Active ranking with subset-wise preferences. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 3312–3321, 2019a.

Aadirupa Saha and Aditya Gopalan. Combinatorial bandits with relative feedback. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 983–993, 2019b.

Aadirupa Saha and Aditya Gopalan. PAC battling bandits in the Plackett-Luce model. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, pages 700–737, 2019c.

Aadirupa Saha and Aditya Gopalan. Best-item learning in random utility models with subset choices. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 4281–4291, 2020a.

Aadirupa Saha and Aditya Gopalan. From PAC to instance-optimal sample complexity in the Plackett-Luce model. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 8367–8376, 2020b.

Anne Schuth, Harrie Oosterhuis, Shimon Whiteson, and Maarten de Rijke. Multileave gradient descent for fast online learning to rank. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*, pages 457—466. Association for Computing Machinery, 2016.

Dhruti Shah, Tuhinangshu Choudhury, Nikhil Karamchandani, and Aditya Gopalan. Sequential mode estimation with oracle queries. *Proceedings of AAAI Conference on Artificial Intelligence*, 34(04):5644–5651, 2020.

Nihar Shah, Sivaraman Balakrishnan, Adityanand Guntuboyina, and Martin J. Wainwright. Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 11–20, 2016.

David Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer Series in Statistics. Springer, 1985.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. Best-of-k bandits. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 1440–1489, 2016.

Patrick Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48:303–312, 1961.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *CoRR*, abs/1904.07272v7, 2022.

B. Babington Smith. Discussion on professor Ross's paper. *Journal of the Royal Statistical Society*, 12:153–162, 1950.

Yanan Sui, Vincent Zhuang, Joel W. Burdick, and Yisong Yue. Multi-dueling bandits with dependent arms. In *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)*, 2017.

Yanan Sui, Masrour Zoghi, Katja Hofmann, and Yisong Yue. Advancements in dueling bandits. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5502–5510, 2018.

Zbigniew Świtalski. General transitivity conditions for fuzzy reciprocal preference matrices. *Fuzzy Sets and Systems*, 137:85–100, 2003.

Balázs Szörényi, Róbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Online rank elicitation for Plackett-Luce: A dueling bandits approach. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 604–612, 2015.

Ambuj Tewari and Susan A. Murphy. *From Ads to Interventions: Contextual Bandits in Mobile Health*, pages 495–517. Springer International Publishing, 2017.

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

Tanguy Urvoy, Fabrice Clerot, Raphael Féraud, and Sami Naamane. Generic exploration and k-armed voting bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 91–99, 2013.

A. W. Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge University Press, 2000.

Claire Vernade, Alexandra Carpentier, Tor Lattimore, Giovanni Zappella, Beyza Ermis, and Michael Brückner. Linear bandits with stochastic delayed feedback. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 9712–9721, 2020.

Thomas A. Waite. Intransitive preferences in hoarding gray jays (perisoreus canadensis). *Behavioral Ecology and Sociobiology*, 50(2):116–121, 2001.

Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

Abraham Wald and Jacob Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, pages 326–339, 1948.

Yizao Wang, Jean-yves Audibert, and Rémi Munos. Algorithms for infinitely many-armed bandits. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, volume 21, 2009.

Ian Waudby-Smith and Aaditya Ramdas. Confidence sequences for sampling without replacement. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS 2020)*, volume 33, pages 20204–20214, 2020.

Huasen Wu and Xin Liu. Double Thompson sampling for dueling bandits. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 649–657, 2016.

Xinyi Yan, Chengxi Luo, Charles Clarke, Nick Craswell, Ellen Voorhees, and Pablo Castells. Human preferences as dueling bandits. In *To appear in: Proceedings of International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2022.

Shuo Yang, Tongzheng Ren, Inderjit S. Dhillon, and Sujay Sanghavi. Combinatorial bandits without total order for arms. *CoRR*, abs/2103.02741, 2021.

Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of Annual Symposium on Foundations of Computer Science (SFCS)*, pages 222–227, 1977.

Yisong Yue and Thorsten Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1201–1208, 2009.

Yisong Yue and Thorsten Joachims. Beat the mean bandit. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 241–248, 2011.

Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.

Shengjia Zhao, Enze Zhou, Ashish Sabharwal, and Stefano Ermon. Adaptive concentration inequalities for sequential decision problems. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, volume 29, 2016.

Julian Zimmert and Yevgeny Seldin. Factored bandits. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 2835–2844, 2018.

Julian Zimmert and Yevgeny Seldin. An optimal algorithm for stochastic and adversarial bandits. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 467–475, 2019.

Masrour Zoghi, Shimon Whiteson, Maarten de Rijke, and Remi Munos. Relative confidence sampling for efficient online ranker evaluation. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*, pages 73–82, 2014a.

Masrour Zoghi, Shimon Whiteson, Remi Munos, and Maarten de Rijke. Relative upper confidence bound for the k-armed dueling bandit problem. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 10–18, 2014b.

Masrour Zoghi, Zohar Karnin, Shimon Whiteson, and Maarten de Rijke. Copeland dueling bandits. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 307–315, 2015a.

Masrour Zoghi, Shimon Whiteson, and Maarten de Rijke. MergeRUCB: A method for large-scale online ranker evaluation. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*, pages 17–26, 2015b.

Masrour Zoghi, Tomas Tunys, Mohammad Ghavamzadeh, Branislav Kveton, Csaba Szepesvári, and Zheng Wen. Online learning to rank in stochastic click models. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 4199–4208, 2017.

Appendix

A. Remaining Proofs for Chapter 2

Lemma 2.3. *For every $h \in (0, \frac{1}{2})$ we have*

$$\sup_{p:|p-1/2|>h} \sum_{n\in\mathbb{N}} \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} = \frac{1}{2}\right) \leq \frac{1}{h^2}. \tag{2.1}$$

Proof of Lem. 2.3. Let $h \in (0, \frac{1}{2})$ be fixed and $p \in [0, \frac{1}{2} - h) \cup (\frac{1}{2} + h, 1]$ be fixed for the moment. With the help of Hoeffding's inequality we obtain in the case $p \in [0, \frac{1}{2} - h)$ that

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} = \frac{1}{2}\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p > h\right) \leq \mathbb{P}\left(\sum_{k=1}^{n} X_k^{(p)} - np > nh\right)$$
$$\leq e^{-2h^2 n}.$$

As $X_k^{(p)}$ is distributed as $1 - X_k^{(1-p)}$ for every $k \in \mathbb{N}$, we have in the case $p \in (\frac{1}{2} + h, 1]$

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} = \frac{1}{2}\right) = \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(1-p)} = \frac{1}{2}\right) \leq e^{-2h^2 n}.$$

Consequently, using that $1 - e^{-z} \geq \frac{z}{2}$ for every $z \in (0, 1)$ we obtain

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} = \frac{1}{2}\right) \leq \sum_{n\in\mathbb{N}} \left(e^{-2h^2}\right)^n = \frac{1}{1 - e^{-2h^2}} \leq \frac{1}{h^2}.$$

As this bound does not depend on the explicit value of p, we obtain (2.1). □

Lemma 2.4. *For $\kappa > 1$ and $U_\kappa(n, \gamma) := \sqrt{\frac{\ln\left(n^\kappa/\gamma\right)}{2n}}$ we have*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \geq U_\kappa(n, \gamma)\right) \leq \frac{\gamma}{n^\kappa},$$
$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \leq -U_\kappa(n, \gamma)\right) \leq \frac{\gamma}{n^\kappa}.$$

In particular, if $n' \in \mathbb{N}$ is such that $\sum_{n\geq n'} \frac{1}{n^\kappa} \leq 1$, then

$$\mathbb{P}\left(\exists n \geq n' : \frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \geq U_\kappa(n, \gamma)\right) \leq \gamma,$$
$$\mathbb{P}\left(\exists n \geq n' : \frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \leq -U_\kappa(n, \gamma)\right) \leq \gamma.$$

Proof of Lem. 2.4. For any $n \in \mathbb{N}$ an application of Hoeffding's inequality yields

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} - p \geq U_\kappa(n, \gamma)\right) = \mathbb{P}\left(\sum_{k=1}^{n} X_k^{(p)} - \mathbb{E}\left[X_k^{(p)}\right] \geq \sqrt{n \ln(n^\kappa/\gamma)}/\sqrt{2}\right)$$
$$\leq \exp\left(-\frac{n}{n}\ln\frac{n^\kappa}{\gamma}\right) = \frac{\gamma}{n^\kappa},$$

and due to $\sum_{n\geq n'} \frac{1}{n^\kappa} \leq 1$ summation over n shows

$$\mathbb{P}\left(\exists n \geq n' : \frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(p)} - p \geq U_\kappa(n,\gamma)\right) \leq \gamma.$$

Since $X_k^{(1-p)}$ is distributed as $1 - X_k^{(p)}$, the rest follows due to symmetry. □

Lemma 2.5. *Let $\phi : [0,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ be given as $\phi(x) := 2\arcsin(\sqrt{x}) - \pi/2$ and $Z_n^{(p)} := \phi(\frac{1}{n}\sum_{i=1}^{n} X_i^{(p)})$ for any $n \in \mathbb{N}, p \in [0,1]$. Then, for fixed $c \in (0, \frac{1}{2})$ and every $p \in [0,c) \cup (1-c, 1]$ we obtain*

$$\mathbb{P}\left(Z_n^{(p)}\phi(p) < 0\right) \leq \exp\left(-\frac{(1-2c)^2 n}{4(2-2c)^2}\right). \tag{2.2}$$

For $\gamma \in (0,1)$, $q := \exp\left(-\frac{(1-2c)^2}{4(2-2c)^2}\right)$ and $\tilde{n} := \lceil \log_q((1-q)\gamma)\rceil$ we thus have

$$\sup\nolimits_{p\in[0,c)\cup(1-c,1]} \mathbb{P}\left(\exists n \geq \tilde{n} : Z_n^{(p)}\phi(p) < 0\right) \leq \gamma. \tag{2.3}$$

Proof of Lem. 2.5. Suppose $c \in (0, \frac{1}{2})$ and $p \in [0,c) \cup (1-c, 1]$ to be fixed and note that we have the equivalence

$$Z_n^{(p)} = \phi\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} X_i^{(p)}\right) < 0 \Leftrightarrow \frac{1}{n}\sum\nolimits_{i=1}^{n} X_i^{(p)} < \frac{1}{2}. \tag{A.1}$$

Hence, in the case $p \in (1-c, 1]$ this equivalence together with $\phi(p) > 0$ imply the identity

$$\mathbb{P}\left(Z_n^{(p)}\phi(p) < 0\right) = \mathbb{P}\left(\sum\nolimits_{i=1}^{n} X_i^{(p)} < \frac{n}{2}\right) = \mathbb{P}\left(\sum\nolimits_{i=1}^{n} X_i^{(p)} < np(1-\delta)\right)$$

with $\delta := 1 - \frac{1}{2p} \in \left(\frac{1-2c}{2-2c}, \frac{1}{2}\right]$. Applying the Chernoff bound shows

$$\mathbb{P}\left(Z_n^{(p)}\phi(p) < 0\right) \leq \exp\left(-\frac{\delta^2 np}{2}\right) \leq \exp\left(-\frac{(1-2c)^2 n}{4(2-2c)^2}\right). \tag{A.2}$$

Let us now consider the case $p \in [0,c)$. Taking into account that

$$\mathbb{P}\left(\frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(p)} < \frac{1}{2}\right) = \mathbb{P}\left(\frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(1-p)} > \frac{1}{2}\right)$$

and (A.1) hold, $\phi(p) = -\phi(1-p) < 0$ implies

$$\begin{aligned}\mathbb{P}\left(Z_n^{(p)}\phi(p) < 0\right) &= \mathbb{P}\left(\frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(p)} > \frac{1}{2}\right) = \mathbb{P}\left(\frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(1-p)} < \frac{1}{2}\right)\\ &= \mathbb{P}\left(Z_n^{(1-p)}\phi(1-p) < 0\right).\end{aligned}$$

By means of (A.2) this can be bounded further as

$$\mathbb{P}\left(Z_n^{(1-p)}\phi(1-p) < 0\right) \leq \exp\left(-\frac{(1-2c)^2 n}{4(2-2c)^2}\right)$$

since $1-p \in (1-c, 1]$ is fulfilled. To see (2.3) note that (2.2), the definition of $\tilde{n}$ and $q \in (0,1)$ ensure for every $p \in [0,c) \cup (1-c, 1]$ that

$$\mathbb{P}\left(\exists n \geq \tilde{n} \,:\, Z_n^{(p)}\phi(p) < 0\right) \leq \sum\nolimits_{n\geq\tilde{n}} \mathbb{P}\left(Z_n^{(p)}\phi(p) < 0\right) \leq \sum\nolimits_{n\geq\tilde{n}} q^n = \frac{q^{\tilde{n}}}{1-q}$$
$$\leq \frac{q^{\log_q((1-q)\gamma)}}{1-q} = \gamma.$$

□

Lemma 2.6. *Let* $\varepsilon = \varepsilon(\gamma) \in (0,1)$ *and* $\delta = \delta(\gamma) \in \left(0, \frac{\ln(1+\varepsilon)}{e}\right)$ *be such that* $\gamma = \frac{2+\varepsilon}{\varepsilon}\left(\frac{\delta}{\ln(1+\varepsilon)}\right)^{1+\varepsilon}$ *holds. Define* $U_\gamma(n) \coloneqq U_{\varepsilon(\gamma),\delta(\gamma)}$ *as*

$$U_\gamma(n) \coloneqq \left(1+\sqrt{\varepsilon}\right)\sqrt{\frac{1}{2}(1+\varepsilon)n\ln\left(\frac{\ln((1+\varepsilon)n)}{\delta}\right)}.$$

Then, the centered sequence $Y_k^{(p)} \coloneqq X_k^{(p)} - p$, $k \in \mathbb{N}$, *fulfills*

$$\mathbb{P}\left(\forall n \in \mathbb{N} : \sum\nolimits_{k=1}^n Y_k^{(p)} \leq U_\gamma(n)\right) \geq 1-\gamma \tag{2.4}$$

as well as

$$\mathbb{P}\left(\forall n \in \mathbb{N} : \left|\sum\nolimits_{k=1}^n Y_k^{(p)}\right| \leq U_\gamma(n)\right) \geq 1-2\gamma. \tag{2.5}$$

Proof of Lem. 2.6. According to Hoeffding's Lemma (cf. Lemma 4.13 in [Mitzenmacher and Upfal, 2017]), for every $k \in \mathbb{N}$, the random variable Y_k is centered and sub-Gaussian with scale parameter $\sigma = 1/2$, i.e., $\mathbb{E}[Y_k] = 0$ and $\mathbb{E}[e^{tY_k}] \leq \exp\left(t^2\sigma^2/2\right)$ for every $t \in \mathbb{R}$ hold. Consequently, Lem. 3 in [Jamieson et al., 2013] implies (2.4). To see (2.5) note that (2.4) does not only hold for $\{Y_k\}_{k\in\mathbb{N}}$, but, due to symmetry, also for $\{Y_k'\}_{k\in\mathbb{N}}$ given by $Y_k' = -Y_k$, $k \in \mathbb{N}$. □

Lemma 2.7. *Let* $c \in (0, 1/2)$, $\gamma \in (0,1)$, $\gamma' \coloneqq \gamma/4$ *and* $\varepsilon' \in (0,1)$, $\delta' \in \left(0, \frac{\ln(1+\varepsilon')}{e}\right)$ *be such that* $\gamma' = \frac{2+\varepsilon'}{\varepsilon'}\left(\frac{\delta'}{\ln(1+\varepsilon')}\right)^{1+\varepsilon'}$. *As in Lem. 2.5 let* $\phi : [0,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\phi(x) \coloneqq 2\arcsin(\sqrt{x}) - \pi/2$ *and* $Z_n^{(p)} \coloneqq \phi\left(\frac{1}{n}\sum_{i=1}^n X_i^{(p)}\right)$ *for all* $n \in \mathbb{N}$ *and* $p \in [0,1]$ *and let*

$$L \coloneqq L(c) \coloneqq \sup\nolimits_{x\in[c/2,1-c/2]} |\phi'(x)|.$$

Define for $n \in \mathbb{N}$

$$l(n) \coloneqq \frac{1}{2}L^2\left(1+\sqrt{\varepsilon'}\right)^2(1+\varepsilon')\ln\left(\frac{\ln((1+\varepsilon')n)}{\delta'}\right)$$

and further

$$\tilde{n} \coloneqq \frac{d'}{c^2}\ln\left(\frac{2}{\delta'}\ln\frac{(1+\varepsilon')d'}{c^2\delta'}\right) + 1$$

with $d' \coloneqq 2(1+\sqrt{\varepsilon'})^2(1+\varepsilon')$. *Then, we have*

$$\sup\nolimits_{p\in[c,1-c]} \mathbb{P}\left(\exists\, n \geq \tilde{n} \,:\, n(Z_n^{(p)} - \phi(p))^2 > l(n)\right) \leq \gamma.$$

Proof of Lem. 2.7. Let $p \in [c, 1-c]$ be arbitrary but fixed. Defining the event

$$A := \left\{ \forall n \geq \tilde{n} : \left| \frac{1}{n} \sum\nolimits_{k=1}^{n} X_k^{(p)} - p \right| \leq \frac{c}{2} \right\}$$

we can estimate

$$\begin{aligned} &\mathbb{P}\left(\exists\, n \geq \tilde{n} \,:\, n(Z_n^{(p)} - \phi(p))^2 > l(n) \right) \\ &\leq \mathbb{P}\left(\left\{ \exists\, n \geq \tilde{n} \,:\, n(Z_n^{(p)} - \phi(p))^2 > l(n) \right\} \cap A \right) + \mathbb{P}(A^c) \end{aligned} \tag{A.3}$$

and it suffices to show that every summand therein is bounded by $2\gamma'$. Due to $p \in [c, 1-c]$ we have $\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)} \in \left[\frac{c}{2}, 1 - \frac{c}{2}\right]$ for every $n \geq \tilde{n}$ on the event A, and thus Lipschitz continuity of ϕ on $\left[\frac{c}{2}, 1-\frac{c}{2}\right]$ with Lipschitz constant L and an application of Lem. 2.6 lets us conclude with regard to the definition of $l(n)$ that

$$\begin{aligned} &\mathbb{P}\left(\left\{ \exists\, n \geq \tilde{n} : n(Z_n^{(p)} - \phi(p))^2 > l(n) \right\} \cap A \right) \\ &\leq \mathbb{P}\left(\exists\, n \geq \tilde{n} : \left| \frac{1}{n} \sum\nolimits_{k=1}^{n} X_k^{(p)} - p \right| > \frac{\sqrt{l(n)}}{L\sqrt{n}} \right) \\ &= \mathbb{P}\left(\exists\, n \geq \tilde{n} : \left| \sum\nolimits_{k=1}^{n} X_k^{(p)} - np \right| > U_{\gamma'}(n) \right) \\ &\leq 2\gamma'. \end{aligned} \tag{A.4}$$

Moreover, as Lem. 2.24 yields for every $n \geq \tilde{n}$

$$\frac{1}{n} \ln\left(\frac{\ln((1+\varepsilon')n)}{\delta'} \right) < \frac{c^2}{d'} = \frac{c^2}{2(1+\sqrt{\varepsilon'})^2(1+\varepsilon')}$$

and thus

$$\frac{U_{\gamma'}(n)}{n} = (1+\sqrt{\varepsilon'}) \sqrt{\frac{\frac{1}{2}(1+\varepsilon')\ln\left(\ln((1+\varepsilon')n)/\delta'\right)}{n}} < \frac{c}{2},$$

we can estimate with the help of Lem. 2.6 further

$$\begin{aligned} \mathbb{P}(A^c) &= \mathbb{P}\left(\exists\, n \geq \tilde{n} : \left| \frac{1}{n} \sum\nolimits_{k=1}^{n} X_k^{(p)} - p \right| > \frac{c}{2} \right) \\ &\leq \mathbb{P}\left(\exists\, n \geq \tilde{n} : \left| \sum\nolimits_{k=1}^{n} X_k^{(p)} - np \right| > U_{\gamma'}(n) \right) \\ &\leq 2\gamma'. \end{aligned} \tag{A.5}$$

As $p \in [c, 1-c]$ was arbitrary, combining (A.3), (A.4) and (A.5) completes the proof. □

Lemma 2.8. *Let $a \geq 1$, $m \in \mathbb{N}$ and suppose $f : \mathbb{N} \to [1, \infty)$ is monotonically increasing with $f(t) \in o(\frac{t}{\ln^a(t)})$ as $t \to \infty$. Let $\{Z_t\}_{t \in \mathbb{N}}$ be a family of independent random variables with $Z_t \sim \mathrm{Ber}(\frac{1}{f(t)m})$, $t \in \mathbb{N}$. Then,*

$$\frac{1}{\ln^a(t)} \sum\nolimits_{t'=1}^{t} Z_{t'} \to \infty \quad \textit{a.s. as } t \to \infty. \tag{2.6}$$

Proof of Lem. 2.8. Abbreviate $n(t) \coloneqq \sum_{t'=1}^{t} Z_{t'}$, let $C > 8$ be arbitrary but fixed and define for each $t \in \mathbb{N}$ the event $D_t \coloneqq \{n(t) \le C \ln^a(t)\}$. The assumption $f(t) \in o(t/\ln^a(t))$ as $t \to \infty$ ensures the existence of some $t_1 \in \mathbb{N}$ such that $2mCf(t)\ln^a(t) \le t$ holds for every $t \ge t_1$. As monotonicity of f allows us to estimate $\mathbb{E}[n(t)] \ge \frac{t}{f(t)m}$, an application of the Chernoff bound yields for each $t \ge t_1$

$$\begin{aligned}\mathbb{P}(D_t) = \mathbb{P}(n(t) \le C\ln^a(t)) &\le \mathbb{P}\left(n(t) \le \frac{1}{2}\frac{t}{f(t)m}\right) \\ &\le \mathbb{P}\left(n(t) \le \frac{1}{2}\mathbb{E}[n(t)]\right) \le e^{-\mathbb{E}[n(t)]/8} \\ &\le e^{-t/(8f(t)m)} \le e^{-C\ln^a(t)/4} \le e^{-2\ln(t)} = \frac{1}{t^2}.\end{aligned}$$

We obtain $\sum_{t\in\mathbb{N}} \mathbb{P}(D_t) \le t_1 + \sum_{t>t_1} \frac{1}{t^2}$ and thus the Borel-Cantelli lemma lets us infer $\mathbb{P}(\limsup_{t\to\infty} D_t) = 0$, i.e.

$$\liminf_{t\to\infty} \frac{n(t)}{\ln^a(t)} > C \quad \text{a.s.}$$

As $C > 8$ was arbitrary, we obtain (2.6). □

Lemma 2.9. *For fixed $c, l > 0$ the sequence $\{a_t\}_{t\in\mathbb{N}}$ given as*

$$a_t = \frac{1}{2^t}\sum\nolimits_{r=0}^{t}\binom{t}{r}\mathbb{P}\left(\chi^2_{(r\wedge c)} > l\right)$$

is monotonically increasing.

Proof of Lem. 2.9. Let $(Y_t)_{t\in\mathbb{N}}$ be a sequence of iid random variables $Y_t \sim \mathrm{Ber}(\frac{1}{2})$ and abbreviate $X_t \coloneqq Y_1 + \cdots + Y_t$. Note that $X_t \le X_{t+1}$ holds almost surely by construction and $X_t \sim \mathrm{Bin}(t, \frac{1}{2})$. Therefore, monotonicity of $r \mapsto \mathbb{P}\left(\chi^2_{(r\wedge c)} > l\right)$ lets us conclude

$$\begin{aligned}a_t &= \frac{1}{2^t}\sum\nolimits_{r=0}^{t}\binom{t}{r}\mathbb{P}\left(\chi^2_{(r\wedge c)} > l\right) = \mathbb{E}\left[\mathbb{P}\left(\chi^2_{(X_t\wedge c)} > l\right)\right] \\ &\le \mathbb{E}\left[\mathbb{P}\left(\chi^2_{(X_{t+1}\wedge c)} > l\right)\right] = a_{t+1}.\end{aligned}$$

□

Lemma 2.10. *Let $h \in (0, 1/2)$ be fixed.*

(i) Let $\gamma \in (0,1)$ be arbitrary. Choose $T \coloneqq \left\lceil \frac{1}{2h^2}\ln\frac{1}{\gamma}\right\rceil$ and define $f : \{0,1\}^T \to \{0,1\}$ via

$$f(x_1,\ldots,x_T) \coloneqq \begin{cases} 0, & \text{if } \frac{1}{T}\sum_{i=1}^{T} x_i < \frac{1}{2}, \\ 1, & \text{if } \frac{1}{T}\sum_{i=1}^{T} x_i \ge \frac{1}{2}. \end{cases}$$

The corresponding non-sequential testing algorithm $\mathcal{A}$ solves $\mathcal{P}^{h,\gamma}_{\text{Coin}}$ and $\mathcal{P}^{1/2+h;\gamma}_{\text{Coin}}$.

(ii) Let $\gamma \in (0, 1/4)$ and suppose $\mathcal{A}$ to be a non-sequential testing algorithm, which solves $\mathcal{P}^{h,\gamma}_{\text{Coin}}$. Then, we have a.s.

$$T^{\mathcal{A}} \ge \frac{1}{2}\left\lfloor \frac{1-4h^2}{h^2}\ln\left(\frac{1}{8\gamma(1-2\gamma)}\right)\right\rfloor.$$

In particular, $T^{\mathcal{A}} \in \Omega\left(\frac{1}{h^2}\ln\frac{1}{\gamma}\right)$ as $\min\{h,\gamma\} \to 0$.

Proof of Lem. 2.10. (i) Suppose $p \in [0,1]$ with $|p-1/2| \geq h$ to be fixed, let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence of random variables $X_n \sim \mathrm{Ber}(p)$. Due to symmetry we have

$$\mathbb{P}_p\left(f(X_1,\ldots,X_T)=1\right) = \mathbb{P}_{1-p}\left(f(X_1,\ldots,X_T)=0\right),$$

hence we may suppose w.l.o.g. $p \leq \frac{1}{2} - h$. Due to $\mathbb{E}_p[X_n] \leq \frac{1}{2} - h$ an application of Hoeffding's inequality lets us infer

$$\begin{aligned}
\mathbb{P}_p\left(f(X_1,\ldots,X_T)=1\right) &= \mathbb{P}_p\left(\frac{1}{T}\sum\nolimits_{i=1}^{T} X_i \geq \frac{1}{2}\right)\\
&\leq \mathbb{P}_p\left(\frac{1}{T}\sum\nolimits_{i=1}^{T} X_i - \frac{T(1/2-h)}{T} \geq h\right)\\
&\leq \mathbb{P}_p\left(\frac{1}{T}\sum\nolimits_{i=1}^{T} (X_i - \mathbb{E}[X_i]) \geq h\right)\\
&\leq e^{-2Th^2} \leq \gamma.
\end{aligned}$$

(ii) Confer Lem. 5.1 on p. 59 in [Anthony and Bartlett, 1999].

□

Proposition 2.12. *Let $\gamma \in \left(0, \frac{1}{40e^8}\right)$ and $h \in \left(0, \frac{1}{4}\right)$ be fixed. If $\mathcal{A}$ solves $\mathcal{P}_{\mathrm{Coin}}^{1/2+h;\gamma}$ resp. $\mathcal{P}_{\mathrm{Coin}}^{h,\gamma}$, then*

$$\max\nolimits_{p\in\{1/2\pm h\}} \mathbb{E}_p\left[T^{\mathcal{A}}\right] \geq \frac{c}{h^2}\ln\frac{1}{\gamma} \quad \textit{resp.} \quad \sup\nolimits_{p:|p-1/2|>h} \mathbb{E}_p\left[T^{\mathcal{A}}\right] \geq \frac{c}{h^2}\ln\frac{1}{\gamma},$$

where $c > 0$ is a universal constant, which does not depend on h or γ.

Proof of Prop. 2.12. This follows from the proof of Thm. 13 in [Mannor and Tsitsiklis, 2004]. More precisely, the problem $\mathcal{P}_{\mathrm{Coin}}^{1/2+h;\gamma}$ is the problem of testing (2.7) if $p \in \{\frac{1}{2}-h, \frac{1}{2}+h\}$ is known beforehand with error $\leq \gamma$, and this corresponds exactly to the problem Π_2 the authors define on p. 642. Reducing this to another problem Π_3 allows them to infer from Thm. 5 in [Mannor and Tsitsiklis, 2004] the lower bound on the sample complexity of Π_2, which is of the form $\max_{p\in\{1/2\pm h\}} \mathbb{E}_p\left[T^{\mathcal{A}}\right] \geq \frac{c}{h^2}\ln\frac{1}{\gamma}$. If $\mathcal{A}$ is a solution to $\mathcal{P}_{\mathrm{Coin}}^{h,\gamma}$, then it solves $\mathcal{P}_{\mathrm{Coin}}^{1/2+h+\varepsilon;\gamma}$ for small $\varepsilon > 0$, which leads to the bound $\max_{p\in\{1/2\pm(h+\varepsilon)\}} \mathbb{E}_p\left[T^{\mathcal{A}}\right] \geq \frac{c}{(h+\varepsilon)^2}\ln\frac{1}{\gamma}$. Taking $\varepsilon \searrow 0$ we see $\sup_{p:|p-1/2|>h} \mathbb{E}_p\left[T^{\mathcal{A}}\right] \geq \frac{c}{h^2}\ln\frac{1}{\gamma}$. □

Lemma 2.14. *Suppose $\{X^{(p)}\}_{p\in[0,1]}$ and $\{U^{(r)}\}_{r\in[0,1]}$ to be families of random variables $X^{(p)} \sim \mathrm{Ber}(p)$ and $U^{(r)} \sim \mathrm{Ber}(r)$ such that, for any $p, r \in [0,1]$, $X^{(p)}$ and $U^{(r)}$ are independent.*

(i) $Y^{(p,r)} := X^{(p)} + \mathbf{1}_{\{X^{(p)}=0\}}U^{(r)}$ fulfills $Y^{(p,r)} \sim \mathrm{Ber}(p+(1-p)r)$.

(ii) If $p_0, p_1 \in [0,1]$ with $p_1 > p_0$ and $p_0 + p_1 \geq 1$ are fixed, we obtain with the choices

$$h := \frac{p_1 - p_0}{2(2-p_1-p_0)} \quad \textit{and} \quad r' := \frac{p_0 - (1/2-h)}{1/2+h}$$

that $Y^{(1/2-h,r')} \sim \mathrm{Ber}(p_0)$ and $Y^{(1/2+h,r')} \sim \mathrm{Ber}(p_1)$.

Proof of Lem. 2.14.

(i) We have $Y^{(p,r)} \in \{0,1\}$ and

$$\mathbb{P}\left(Y^{(p,r)} = 1\right) = \mathbb{P}\left(X^{(p)} = 1\right) + \mathbb{P}\left(X^{(p)} = 0 \text{ and } U^{(r)} = 1\right) = p + (1-p)r.$$

(ii) The choices of p_0 and p_1 guarantee that $h \in (0, 1/2)$ and $r' \in [0,1]$. Moreover, a straight-forward calculation shows that

$$1/2 - h + (1/2 + h)\, r' = p_0 \quad \text{and} \quad 1/2 + h + (1/2 - h)\, r' = p_1,$$

hence the statement follows from (i).

□

Proposition 2.17. *Suppose* $0 < \gamma < \gamma_0 < 1/2$ *and* $0 < h < h_0 < 1/2$ *to be fixed.*

(i) The symmetric SPRT $\mathcal{A}$ *with barrier* $B := \left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+h)/(1/2-h))} \right\rceil$ *solves* $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$ *and* $\mathcal{P}_{\text{Coin}}^{h,\gamma}$*, i.e.,*

$$\forall p \geq 1/2 + h : \mathbb{P}_p(\mathbf{D}(\mathcal{A}) = 0) \geq 1 - \gamma \quad \text{and} \quad \forall p \leq 1/2 - h : \mathbb{P}_p(\mathbf{D}(\mathcal{A}) = 1) \geq 1 - \gamma.$$

Moreover, the termination time $T^{\mathcal{A}}$ *of* $\mathcal{A}$ *fulfills*

$$\sup_{p \in [0, 1/2-h] \cup [1/2+h, 1]} \mathbb{E}_p[T^{\mathcal{A}}] = \mathbb{E}_{1/2 \pm h}[T^{\mathcal{A}}] = \frac{1 - 2\gamma}{2h} \left\lceil \frac{\ln \frac{1-\gamma}{\gamma}}{\ln \frac{1/2+h}{1/2-h}} \right\rceil, \tag{2.8}$$

which is in $\mathcal{O}(\frac{1}{h^2} \ln \frac{1}{\gamma})$ *as* $\max\{\frac{1}{h}, \frac{1}{\gamma}\} \to \infty$.

(ii) The testing algorithm $\mathcal{A}$ *from (i) is w.r.t.* $\mathbb{E}_{1/2+h}[T^{\mathcal{A}}]$ *and* $\mathbb{E}_{1/2-h}[T^{\mathcal{A}}]$ *optimal among all solutions to* $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$*. In other words: If* $\mathcal{A}'$ *is an algorithm, which fulfills*

$$\mathbb{P}_{1/2+h}(\mathbf{D}(\mathcal{A}') = 0) \geq 1 - \gamma \quad \text{and} \quad \mathbb{P}_{1/2-h}(\mathbf{D}(\mathcal{A}') = 1) \geq 1 - \gamma,$$

then it fulfills

$$\mathbb{E}_{1/2 \pm h}[T^{\mathcal{A}'}] \geq \mathbb{E}_{1/2 \pm h}[T^{\mathcal{A}}] = \frac{1 - 2\gamma}{2h} \left\lceil \frac{\ln \frac{1-\gamma}{\gamma}}{\ln \frac{1/2+h}{1/2-h}} \right\rceil \geq \frac{c(h_0, \gamma_0)}{h^2} \ln \frac{1}{\gamma}$$

for some appropriate constant $c(h_0, \gamma_0)$*, which does not depend on* γ *or* h.

Proof of Prop. 2.17. (i) The test $\mathcal{A}$ from (i) is the sequential probability ratio test (SPRT) for the problem at hand and has its origins in [Wald, 1945]. Statement (i) can be inferred from p.10–15 in [Siegmund, 1985]. More precisely, equation (2.28) on p.15 in [Siegmund, 1985] shows that

$$\mathbb{P}_{1-p}\left(S_{T^{\mathcal{A}}} \leq -B\right) = \mathbb{P}_p\left(S_{T^{\mathcal{A}}} \geq B\right) = \left(1 + {(1-p)^B}/{p^B}\right)^{-1}$$

for every $p \neq 1/2$. Since B is chosen such that the right-hand side is $\leq \gamma$ if $p = \frac{1}{2} - h$, $\mathcal{A}$ solves $\mathcal{P}_{\mathrm{Coin}}^{1/2+h;\gamma}$. As $p \mapsto 1/(1+(1-p)^B/p^B)$ is monotonically increasing on $[\frac{1}{2}, 1]$, $\mathcal{A}$ decides $\mathbf{H}_0 : p > \frac{1}{2}$ versus $\mathbf{H}_1 : p < \frac{1}{2}$ with error probability at most γ also for every $p \in [0, \frac{1}{2} - h) \cup (\frac{1}{2} + h, 1]$, i.e., it solves $\mathcal{P}_{\mathrm{Coin}}^{h,\gamma}$. Moreover, equation (2.29) on p.15 in [Siegmund, 1985] shows that for each $h' \in (0, \frac{1}{2})$

$$\mathbb{E}_{1/2\pm h'}[T^{\mathcal{A}}] = \frac{1}{2h'} B \left| 1 - 2\left(1 + \left({}^{(1/2+h')^B}/_{(1/2-h')^B}\right)\right)^{-1} \right|, \tag{A.6}$$

which is continuous and decreasing in h' for $h' \in (0, \frac{1}{2})$. Consequently, (2.8) holds by the choice of B. Using that $\frac{x}{x+1} < \ln(1+x) < x$ holds for each $x > -1$ we see that $\ln({}^{(1/2+h)}/_{(1/2-h)}) \in \Theta(h)$ as $h \to 0$, and thus the right-hand side of (2.8) is in $\mathcal{O}\left(\frac{1}{h^2} \ln \frac{1}{\gamma}\right)$ as $\min\{h, \gamma\} \to 0$.

(ii) For the optimality of $\mathcal{A}$ stated in (ii) as a solution for deciding $\mathbf{H}_0 : p > \frac{1}{2}$ versus $\mathbf{H}_1 : p < \frac{1}{2}$ with error $\leq \gamma$ for $p \in \{\frac{1}{2} \pm h\}$ confer pages 19–22 in [Siegmund, 1985] or [Ferguson, 1967, Theorem 2, pp. 365] or the original proof from Wald and Wolfowitz [1948].

In order to conclude the lemma, we need the show a lower bound for the right-hand side of (2.8) of the form $\frac{c(h_0,\gamma_0)}{h^2} \ln \frac{1}{\gamma}$ for some appropriate constant $c(h_0, \gamma_0)$, which does not depend on γ or h. The function $f : (0,1) \to \mathbb{R}, \gamma \mapsto \frac{\ln((1-\gamma)/\gamma)\cdot(1-2\gamma)}{\ln(1/\gamma)}$ fulfills $f(1/2) = 0$ and

$$f'(\gamma) = \frac{(1-2\gamma)\ln\frac{1}{\gamma} - (\gamma-1)\ln\left(\frac{1}{\gamma}-1\right)\left(2\gamma + 2\gamma\ln\frac{1}{\gamma} - 1\right)}{(\gamma-1)\gamma\ln^2\frac{1}{\gamma}} < 0$$

for every $\gamma \in (0, \frac{1}{2})$. Consequently, there exists some $c'(\gamma_0) > 0$ with $\ln((1-\gamma)/\gamma)(1-2\gamma) \geq c'(\gamma_0)\ln\frac{1}{\gamma}$ for each $\gamma \in (0, \gamma_0)$. Moreover, as $\ln(1+x) < x$ for $x > -1$, we obtain for $h \in (0, h_0)$ the inequality

$$\ln\left(\frac{\frac{1}{2}+h}{\frac{1}{2}-h}\right) = \ln\left(1 + \frac{4h}{1-2h}\right) < \frac{4h}{1-2h} < \frac{4h}{1-2h_0}.$$

Combining these estimates, we obtain with $c(h_0, \gamma_0) := \frac{c'(\gamma_0)(1-2h_0)}{8}$ that

$$\frac{1-2\gamma}{2h}\left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln({}^{(1/2+h)}/_{(1/2-h)})} \right\rceil \geq \frac{c(h_0,\gamma_0)}{h^2}\ln\frac{1}{\gamma}.$$

As the SPRT is optimal (w.r.t. expected runtime) this shows, with regard to (2.8), that any such algorithm $\mathcal{A}'$ as in (ii) fulfills

$$\mathbb{E}_{1/2\pm h}\left[T^{\mathcal{A}'}\right] \geq \frac{1-2\gamma}{2h}\left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln({}^{(1/2+h)}/_{(1/2-h)})} \right\rceil \geq \frac{c(h_0,\gamma_0)}{h^2}\ln\frac{1}{\gamma}.$$

□

Lemma 2.18. *There does not exist a SPRT that solves* $\mathcal{P}_{\mathrm{Coin}}^{\gamma}$.

Proof of Lem. 2.18. Assume that $\mathcal{A}$ was a SPRT, which solves $\mathcal{P}_{\text{Coin}}^{\gamma}$ for some $\gamma < 1$, i.e., there exist some $A, B \in \mathbb{R}$ with $A < B$ such that $\mathcal{A}$ stops as soon as $S_n \notin (A, B)$ and outputs 0 if $S_n \geq B$ and 1 if $S_n \leq A$. As by assumption on $\mathcal{A}$, for any $p \neq 1/2$, both $\{\mathbf{D}(\mathcal{A}) = 0\}$ and $\{\mathbf{D}(\mathcal{A}) = 1\}$ have probability greater than zero, $A < 0 < B$ is fulfilled. Now, choose $h \in (0, 1/2)$ such that $\max\{-A, B\} + 1 < \left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+h)/(1/2-h))} \right\rceil =: C$ and denote by $\widetilde{\mathcal{A}}$ the corresponding SPRT from Prop. 2.17, which has boundaries $-C$ and C and solves $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$. As $\mathcal{A}$ is also a solution to $\mathcal{P}_{\text{Coin}}^{1/2+h;\gamma}$, Prop. 2.17 shows that $\mathbb{E}_{1/2+h}[T^{\widetilde{\mathcal{A}}}] \leq \mathbb{E}_{1/2+h}[T^{\mathcal{A}}]$. But at the same time, $-C + 1 < A < B < C - 1$ implies $\mathbb{E}_{1/2+h}[T^{\widetilde{\mathcal{A}}}] > \mathbb{E}_{1/2+h}[T^{\mathcal{A}}]$, a contradiction. □

Lemma 2.21. *If $\mathcal{A}$ is a symmetric GSPRT, then*

$$\forall 1/2 \leq p_1 \leq p_2, \forall n \in \mathbb{N} : \mathbb{P}_{p_1}\left(T^{\mathcal{A}} \leq n\right) \leq \mathbb{P}_{p_2}\left(T^{\mathcal{A}} \leq n\right).$$

In particular, the function $[\frac{1}{2}, 1] \to \mathbb{R} \cup \{\infty\}$, $h \mapsto \mathbb{E}_{\frac{1}{2}+h}[T^{\mathcal{A}}]$ is monotonically decreasing.

Proof of Lem. 2.21. Let $B : \mathbb{N} \to [0, \infty]$ be the barrier function for $\mathcal{A}$. Recall that $\mathcal{A}$ observes an iid sequence $\{X_n^{(p)}\}_{n\in\mathbb{N}}$ where $X_n^{(p)} \sim \text{Ber}(p)$ for some $p \in [0, 1]$ and that it terminates as soon as $M_n^{(p)} := |\sum_{i=1}^n Z_i^{(p)}| \geq B(n)$ holds, where $Z_i^{(p)} := 2X_i^{(p)} - 1$ for every $i \in \mathbb{N}$. According to [Ross, 1996, pp. 166-167], $M_n^{(p)}$ is a random walk on $\mathbb{N}_0$ with transition probabilities

$$\mathbb{P}\left(M_{n+1}^{(p)} = i \,\middle|\, M_n^{(p)} = j\right) = \begin{cases} 1, & \text{if } j = 0, i = 1, \\ a_i(p), & \text{if } j > 0, i = j + 1, \\ 1 - a_i(p), & \text{if } j > 0, i = j - 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $a_i(p) := \frac{p^{i+1}+(1-p)^{i+1}}{p^i+(1-p)^i}$. Writing $x := p - \frac{1}{2}$ we see that

$$a_i(p) = \frac{(\frac{1}{2}+x)^{i+1} + (\frac{1}{2}-x)^{i+1}}{(\frac{1}{2}+x)^i + (\frac{1}{2}-x)^i} = \frac{1}{2} + xb_i(x) \text{ with } b_i(x) = \frac{(\frac{1}{2}+x)^i - (\frac{1}{2}-x)^i}{(\frac{1}{2}+x)^i + (\frac{1}{2}-x)^i}.$$

Now, $\frac{\mathrm{d}}{\mathrm{d}x} b_i(x) = \frac{2i\left(1/4-x^2\right)^{i-1}}{((1/2-x)^i+(1/2+x)^i)^2} \geq 0$ for all $x \in (0, 1/2)$ and $i \in \mathbb{N}$ lets us infer that $x \mapsto b_i(x)$ is monotonically increasing on $[0, \frac{1}{2}]$, i.e., $p \mapsto a_i(p)$ is monotonically increasing on $[\frac{1}{2}, 1]$.

Moreover, for fixed $p \in [1/2, 1]$, we may extend $\mathbb{N} \ni i \mapsto a_i(p)$ to a differentiable function $(0, \infty) \to \mathbb{R}, i \mapsto a_i(p) := \frac{p^{i+1}+(1-p)^{i+1}}{p^i+(1-p)^i}$. Here, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}i} a_i(p) = -\frac{(2p-1)(-(p-1)p)^i(\ln(1-p) - \ln(p))}{(p^i + (1-p)^i)^2}.$$

As $p \in [\frac{1}{2}, 1]$ assures $2p - 1 \geq 0$ and $-(p-1)p \geq 0$ as well as $\ln(1-p) - \ln(p) \leq 0$, we have $\frac{\mathrm{d}}{\mathrm{d}i} a_i(p) \geq 0$ on $(0, \infty)$. In other words, $i \mapsto a_i(p)$ is monotonically increasing.

Now, suppose $\frac{1}{2} \le p_1 \le p_2$ to be fixed and assume that $X_n^{(p_1)}$ and $X_n^{(p_2)}$, $n \in \mathbb{N}$, are all defined on a common probability space. Let $\{U_n\}_{n\in\mathbb{N}}$ be a family of iid random variables $U_n \sim \mathcal{U}([0,1])$. Define for $k \in \{1,2\}$ the function

$$f_k : \mathbb{N}_0 \times [0,1] \to \mathbb{N}_0, \quad f_k(i,u) := \begin{cases} i+1, & \text{if } i>0, u \le a_i(p_k), \\ i-1, & \text{if } i>0, u > a_i(p_k), \\ 1, & \text{if } i=0. \end{cases}$$

Then, for $k \in \{1,2\}$, $M_0^k := 0$ and $M_{n+1}^k := f_k(M_n^k, U_{n+1})$ defines a random walk $\{M_n^k\}_{n\in\mathbb{N}}$ on $\mathbb{N}_0$ with $M_n^k \sim M_n^{(p_k)}$ for all $n \in \mathbb{N}$. We will show in the following that

$$M_n^1 \le M_n^2 \quad \text{almost surely.} \tag{A.7}$$

To see this inductively, note at first that $M_0^1 = 0 = M_0^2$ holds, and then suppose (A.7) to hold for some arbitrary but fixed $n \in \mathbb{N}$. In the case $M_{n+1}^1 - M_n^1 = -1$ we trivially have

$$M_{n+1}^1 = M_n^1 - 1 \le M_n^2 - 1 \le M_{n+1}^2 \quad \text{a.s.}$$

In the other case, i.e., $M_{n+1}^1 - M_n^1 = 1$, either (a) $M_n^1 = 0$ or (b) $U_{n+1} \le a_{M_n^1}(p_1)$ holds. As $M_0^k = 0$ and $|M_{\tilde{n}+1}^k - M_{\tilde{n}}^k| = 1$ for all $\tilde{n} \in \mathbb{N}$ hold for $k \in \{1,2\}$, M_n^1 is even iff M_n^2 is even. Consequently, in the case (a) either $M_n^2 \ge 2$ or $M_n^2 = 0$ is fulfilled, which directly assures $M_{n+1}^1 \le M_{n+1}^2$. If (b) is fulfilled, then the monotonicity properties of $a_i(p)$ shown above reveal $U_{n+1} \le a_{M_n^1}(p_1) \le a_{M_n^1}(p_2) \le a_{M_n^2}(p_2)$ a.s. Hence, $M_{n+1}^2 - M_n^2 = 1$ and thus $M_{n+1}^2 = 1 + M_n^2 \ge 1 + M_n^1 = M_{n+1}^1$ follows in this case.
From (A.7) we infer

$$\min\left\{n \in \mathbb{N} \,:\, M_n^1 \ge B(n)\right\} \ge \min\left\{n \in \mathbb{N} \,:\, M_n^2 \ge B(n)\right\} \text{ a.s.,}$$

hence the statement follows due to $M_n^k \sim M_n^{(p_k)}$. □

Proposition 2.22. *Let $\gamma \in (0,1/2)$ be fixed. Let S_n' be a symmetric random walk on $\mathbb{Z}$, i.e. $S_n' = \sum_{i=1}^n X_i'$ where $X_i' \sim \frac{1}{2}(\delta_1 + \delta_{-1})$ are iid. For arbitrary $c > 3$ the number*

$$n_0 := \min\left\{n \in \mathbb{N} \,\middle|\, \mathbb{P}\left(\exists \tilde{n} \ge n+1 \,:\, \left|S_{\tilde{n}}'\right| \ge \frac{\tilde{n}}{\sqrt{2}} \ln_2(\tilde{n}+e) + c\ln_3(\tilde{n}+e^e)\right) \le \gamma\right\}$$

is finite. The corresponding symmetric GSPRT $\mathcal{A}$ with the barrier $B_\gamma^{\text{Farrell}} : \mathbb{N} \to [0,\infty]$ given by

$$B_\gamma^{\text{Farrell}}(n) := \begin{cases} \sqrt{n\ln_2(n+e) + c\ln_3(n+e^e)}/\sqrt{2}, & \textit{if } n \ge n_0+1, \\ n, & \textit{otherwise} \end{cases}$$

solves $\mathcal{P}_{\text{Coin}}^\gamma$ and fulfills

$$\lim_{h\to 0} \frac{\mathbb{E}_{1/2\pm h}\left[T^{\mathcal{A}}\right]}{\frac{1}{h^2}\ln\ln\frac{1}{h}} = \frac{1}{2}\mathbb{P}_{1/2}(T^{\mathcal{A}} = \infty) > 0. \tag{2.9}$$

Proof of Prop. 2.22. In the proof of Thm. 1 in [Farrell, 1964] it is shown that $\mathcal{A}$ fulfills

$$\mathbb{P}_p(\mathbf{D}(\mathcal{A}) = 0) \geq 1 - \gamma \quad \text{and} \quad \mathbb{P}_{1-p}(\mathbf{D}(\mathcal{A}) = 1) \geq 1 - \gamma$$

for any $p > 1/2$ and also (2.9); more precisely, the construction of the test can be found on pp. 68f. For verifying (2.9) note that $\frac{1}{|\ln|\ln|h|||} = \frac{1}{\ln\ln\frac{1}{h}}$ holds for $h < \frac{1}{e}$.

To see that $\mathcal{A}$ terminates almost surely for any $p \neq \frac{1}{2}$, we may assume due to symmetry w.l.o.g. $p = \frac{1}{2} + h \in (\frac{1}{2}, 1]$. Note that (2.9) guarantees the existence of some $h' \in (0, h)$ with $\mathbb{E}_{1/2+h'}[T^{\mathcal{A}}] = \frac{1}{2}\mathbb{P}_{1/2}(T^{\mathcal{A}} = \infty)\left(\frac{1}{h'^2}\ln\ln\frac{1}{h'}\right) < \infty$. Therefore, we can infer from Lem. 2.21 that $\mathbb{E}_{1/2+h}[T^{\mathcal{A}}] \leq \mathbb{E}_{1/2+h'}[T^{\mathcal{A}}] < \infty$ holds, i.e., in particular $T^{\mathcal{A}} < \infty$ holds a.s. w.r.t $\mathbb{P}_p$. □

Proposition 2.23. *(i) For $h \in (0, 1/2)$, $\gamma \in (0, 1)$ let $\varepsilon = \varepsilon(\gamma) \in (0, 1)$ and $\delta = \delta(\gamma) \in \left(0, \frac{\ln(1+\varepsilon)}{e}\right)$ be such that $\gamma = \frac{2+\varepsilon}{\varepsilon}\left(\frac{\delta}{\ln(1+\varepsilon)}\right)^{1+\varepsilon}$. Then, the symmetric GSPRT with barrier $B^{\mathrm{LiL}}_{h,\gamma}$ defined via*

$$B^{\mathrm{LiL}}_{h,\gamma}(n) := \max\left\{0, (1+\sqrt{\varepsilon})\sqrt{\frac{1}{2n}(1+\varepsilon)\ln\left(\frac{\ln((1+\varepsilon)n)}{\delta}\right)} - h\right\}. \tag{2.10}$$

solves $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$. Moreover, $n \mapsto B^{\mathrm{LiL}}_{h,\gamma}(n)$ is monotonically decreasing.

(ii) Let $\gamma_0, \varepsilon_0 \in (0, 1)$ be such that $(\frac{\varepsilon_0}{2+\varepsilon_0})^{\frac{1}{1+\varepsilon_0}} < \frac{1}{e}$ holds[1]. With $\varepsilon(\gamma) := \varepsilon_0$ and $\delta(\gamma) := (\frac{\gamma\varepsilon_0}{2+\varepsilon_0})^{\frac{1}{1+\varepsilon_0}}\ln(1+\varepsilon_0) \in \left(0, \frac{\ln(1+\varepsilon_0)}{e}\right)$ we have $\gamma = \frac{2+\varepsilon(\gamma)}{\varepsilon(\gamma)}\left(\frac{\delta(\gamma)}{\ln(1+\varepsilon(\gamma))}\right)^{1+\varepsilon(\gamma)}$ and the symmetric GSPRT $\mathcal{A}$ from (i) started with h and γ fulfills

$$\inf_{p\in[0,1/2-h)\cup(1/2+h,1]} \mathbb{P}_p\left(T^{\mathcal{A}} \leq N_0(h,\gamma)\right) = 1,$$

where

$$N_0(h,\gamma) := \frac{d_0}{h^2}\ln\left(\frac{1}{\delta(\gamma)}\ln\frac{(1+\varepsilon_0)d_0}{h^2\delta(\gamma)}\right) + 2 \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$$

with $d_0 := \frac{1}{2}(1+\sqrt{\varepsilon_0})^2(1+\varepsilon_0)$.

Proof of Prop. 2.23. As $N \mapsto \ln\ln(cN)/N$ is monotonically decreasing for every $c > 0$, $B_{h,\gamma}(N) := B^{\mathrm{LiL}}_{h,\gamma}(N)$ is in fact monotonically decreasing. For the sake of convenience, suppose $p \in [0, \frac{1}{2} - h) \cup (\frac{1}{2} + h, 1]$, $h \in (0, \frac{1}{2})$ and $\gamma \in (0, 1)$ to be fixed for the moment. Recall that the termination time $T^{\mathcal{A}}$ of the symmetric GSPRT $\mathcal{A}$ with barrier $B_{h,\gamma}$ is given by

$$T^{\mathcal{A}} = \min\left\{N \in \mathbb{N} \,\middle|\, \frac{1}{N}\sum_{n=1}^{N} X^{(p)}_n \notin \left[\frac{1}{2} - B_{h,\gamma}(N), \frac{1}{2} + B_{h,\gamma}(N)\right]\right\},$$

where $\{X^{(p)}_t\}_{t\in\mathbb{N}}$ is an iid family of random variables $X^{(p)}_t \sim \mathrm{Ber}(p)$, and that $\mathcal{A}$ decides at termination for "$p > \frac{1}{2}$" if $\frac{1}{T^{\mathcal{A}}}\sum_{t\leq T^{\mathcal{A}}} X^{(p)}_t > \frac{1}{2} + B_{h,\gamma}(T^{\mathcal{A}})$ and for "$p < \frac{1}{2}$" otherwise. We split the proof into two parts.

[1] E.g., $\varepsilon_0 = 1/2$ works here.

Part 1: $\mathcal{A}$ fulfills its error guarantees.
We start with the case $p > \frac{1}{2} + h$. As $1 - X_n^{(1-p)} \sim \mathrm{Ber}(p) \sim X_n^{(p)}$ holds for every $n \in \mathbb{N}$, the inequality $h + B_{h,\gamma}(N) \geq \frac{1}{N} U_\gamma(N)$ and Lem. 2.6 yield

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T^{\mathcal{A}}}\sum\nolimits_{n=1}^{T^{\mathcal{A}}} X_n^{(p)} < \frac{1}{2} - B_{h,\gamma}(T^{\mathcal{A}})\right) \leq \mathbb{P}\left(\exists N \in \mathbb{N} \,:\, \frac{1}{N}\sum\nolimits_{n=1}^{N} X_n^{(p)} < \frac{1}{2} - B_{h,\gamma}(N)\right)\\
&= \mathbb{P}\left(\exists N \in \mathbb{N} \,:\, \frac{1}{N}\sum\nolimits_{n=1}^{N}\left(1 - X_n^{(1-p)}\right) < \frac{1}{2} - B_{h,\gamma}(N)\right)\\
&= \mathbb{P}\left(\exists N \in \mathbb{N} \,:\, \frac{1}{N}\sum\nolimits_{n=1}^{N}\left(X_n^{(1-p)} - (1-p)\right) > p - \frac{1}{2} + B_{h,\gamma}(N)\right)\\
&\leq \mathbb{P}\left(\exists N \in \mathbb{N} \,:\, \frac{1}{N}\sum\nolimits_{n=1}^{N}\left(X_n^{(1-p)} - (1-p)\right) > h + B_{h,\gamma}(N)\right)\\
&\leq \mathbb{P}\left(\exists N \in \mathbb{N} \,:\, \frac{1}{N}\sum\nolimits_{n=1}^{N}\left(X_n^{(1-p)} - (1-p)\right) > \frac{1}{N}U_\gamma(N)\right)\\
&= 1 - \mathbb{P}\left(\forall N \in \mathbb{N} \,:\, \sum\nolimits_{n=1}^{N}\left(X_n^{(1-p)} - (1-p)\right) \leq U_\gamma(N)\right) \leq 1 - (1-\gamma) = \gamma.
\end{aligned}$$

In the other case $p < \frac{1}{2} - h$ we have $1 - p > \frac{1}{2} + h$ and thus we can infer

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T^{\mathcal{A}}}\sum\nolimits_{n=1}^{T^{\mathcal{A}}} X_n^{(p)} > \frac{1}{2} + B_{h,\gamma}(T^{\mathcal{A}})\right)\\
&= \mathbb{P}\left(-\frac{1}{T^{\mathcal{A}}}\sum\nolimits_{n=1}^{T^{\mathcal{A}}} X_n^{(1-p)} > -\frac{1}{2} + B_{h,\gamma}(T^{\mathcal{A}})\right)\\
&= \mathbb{P}\left(\frac{1}{T^{\mathcal{A}}}\sum\nolimits_{n=1}^{T^{\mathcal{A}}} X_n^{(1-p)} < \frac{1}{2} - B_{h,\gamma}(T^{\mathcal{A}})\right) \leq \gamma.
\end{aligned}$$

Part 2: Bounding $T^{\mathcal{A}}$.
With $d \coloneqq \frac{(1+\sqrt{\varepsilon})^2(1+\varepsilon)}{2}$ and

$$N_0 \coloneqq \frac{d}{h^2}\ln\left(\frac{1}{\delta}\ln\frac{(1+\varepsilon)d}{h^2\delta}\right) \tag{A.8}$$

Lem. 2.24 implies that for every $N > N_0$

$$\frac{1}{N}\ln\left(\frac{\ln((1+\varepsilon)N)}{\delta}\right) < \frac{h^2}{d} = \frac{2h^2}{(1+\sqrt{\varepsilon})^2(1+\varepsilon)}$$

and thus

$$\frac{U_\gamma(N)}{N} = (1+\sqrt{\varepsilon})\sqrt{\frac{\frac{1+\varepsilon}{2}\ln\left(\ln((1+\varepsilon)N)/\delta\right)}{N}} < h$$

is fulfilled, with $U_\gamma(N)$ as in Lem. 2.6. In particular, $B_{h,\gamma}(N) = 0$ holds for every $N > N_0$. On the event $\{T^{\mathcal{A}} > N_0 + 2\}$ we have $|\frac{1}{N_0+2}\sum_{n=1}^{N_0+2} X_n^{(p)} - \frac{1}{2}| \leq B_{h,\gamma}(N_0+2) = 0$ and $|\frac{1}{N_0+1}\sum_{n=1}^{N_0+1} X_n^{(p)} - \frac{1}{2}| \leq B_{h,\gamma}(N_0+1) = 0$. But this implies $2\sum_{i=1}^{N_0+2} X_n^{(p)} = N_0 + 2$ as well as $2\sum_{i=1}^{N_0+1} X_n^{(p)} = N_0 + 1$, i.e., N_0 would have to be odd and even at the same time. This is not possible, hence $T^{\mathcal{A}} \leq N_0 + 2$ holds with probability 1.

Now, suppose $\varepsilon_0 \in (0,1)$ to be such that $(\frac{\varepsilon_0}{2+\varepsilon_0})^{\frac{1}{1+\varepsilon_0}} < \frac{1}{e}$ holds and let $\varepsilon(\gamma) = \varepsilon_0$ for all $\gamma \in (0,1)$. Then, $\delta(\gamma) := (\frac{\gamma\varepsilon_0}{2+\varepsilon_0})^{\frac{1}{1+\varepsilon_0}} \ln(1+\varepsilon_0) \in \left(0, \frac{\ln(1+\varepsilon_0)}{e}\right)$ fulfills $\gamma = \frac{2+\varepsilon(\gamma)}{\varepsilon(\gamma)}\big(\frac{\delta(\gamma)}{\ln(1+\varepsilon(\gamma))}\big)^{1+\varepsilon(\gamma)}$ for each $\gamma \in (0,1)$. With the argumentation from above we obtain for every $h \in (0, \frac{1}{2})$, $\gamma \in (0, \gamma_0)$ and $p \in [0, \frac{1}{2}-h) \cup (\frac{1}{2}+h, 1]$ that $\mathcal{A}$ started with parameters h and γ fulfills

$$T^{\mathcal{A}} \leq \frac{d_0}{h^2} \ln\left(\frac{1}{\delta(\gamma)} \ln \frac{(1+\varepsilon_0)d_0}{h^2\delta(\gamma)}\right) + 2 =: N_0(h,\gamma) \quad \text{a.s. w.r.t. } \mathbb{P}_p,$$

where $d_0 := \frac{(1+\sqrt{\varepsilon_0})^2(1+\varepsilon_0)}{2}$. Due to $\delta(\gamma) \in \mathcal{O}(\gamma^{1/(1+\varepsilon_0)})$ we have

$$N_0(h,\gamma) \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right) \quad \text{as } \max\left\{\frac{1}{h}, \frac{1}{\gamma}\right\} \to \infty. \tag{A.9}$$

□

Corollary 2.25. *For fixed $\gamma \in (0,1)$, $h \in (0,1/2)$ let ε and δ be as in Prop. 2.23 and define $B : \mathbb{N} \to [0,\infty]$ via*

$$B(n) := \begin{cases} B^{\mathrm{LIL}}_{h,\gamma}(n), & \text{if } n \leq \frac{1}{2h^2}\ln\frac{1}{\gamma} \\ 0, & \text{otherwise.} \end{cases}$$

with $B^{LIL}_{h,\gamma}$ as in Prop. 2.23. Then, the symmetric GSPRT $\mathcal{A}$ with barrier B solves $\mathcal{P}^{h,3\gamma}_{\mathrm{Coin}}$ and fulfills $P_p\left(T^{\mathcal{A}} \leq \frac{1}{2h^2}\ln\frac{1}{\gamma} + 2\right) = 1$ for any $p \in [0,1]$.

Proof of Cor. 2.25. As in the proof of Prop. 2.23 we see that $T^{\mathcal{A}} \leq \frac{1}{2h^2}\ln\frac{1}{\gamma} + 2$ has probability one for every $p \in [0,1]$. As $\mathcal{A}$ is symmetric, for proving that it solves $\mathcal{P}^{h,\gamma}_{\mathrm{Coin}}$ we only have to consider the case $p < 1/2 - h$. The probability that $\mathcal{A}$ makes an error in this case is given by

$$\begin{aligned}
&\mathbb{P}_p(\mathbf{D}(\mathcal{A}) = 1) \\
&= \mathbb{P}_p\left(\left\{\mathbf{D}(\mathcal{A}) = 1 \text{ and } T^{\mathcal{A}} \leq \frac{1}{2h^2}\ln\frac{1}{\gamma}\right\} \cup \left\{\mathbf{D}(\mathcal{A}) = 1 \text{ and } T^{\mathcal{A}} > \frac{1}{2h^2}\ln\frac{1}{\gamma}\right\}\right) \\
&\leq \mathbb{P}_p\left(\mathbf{D}(\mathcal{A}) = 1 \text{ and } T^{\mathcal{A}} \leq \frac{1}{2h^2}\ln\frac{1}{\gamma}\right) + \mathbb{P}_p\left(\mathbf{D}(\mathcal{A}) = 1 \text{ and } T^{\mathcal{A}} > \frac{1}{2h^2}\ln\frac{1}{\gamma}\right)
\end{aligned}$$

Until time $\lceil\frac{1}{2h^2}\ln\frac{1}{\gamma}\rceil$, $\mathcal{A}$ coincides with the solution $\widetilde{\mathcal{A}}$ (with barrier $B^{\mathrm{LIL}}_{h,\gamma}$) from Prop. 2.23 and thus correctness of the latter yields

$$\begin{aligned}
\mathbb{P}_p\left(\mathbf{D}(\mathcal{A}) = 1 \text{ and } T^{\mathcal{A}} \leq \frac{1}{2h^2}\ln\frac{1}{\gamma}\right) &= \mathbb{P}_p\left(\mathbf{D}(\widetilde{\mathcal{A}}) = 1 \text{ and } T^{\mathcal{A}} \leq \frac{1}{2h^2}\ln\frac{1}{\gamma}\right) \\
&\leq \mathbb{P}_p\left(\mathbf{D}(\widetilde{\mathcal{A}}) = 1\right) \leq \gamma.
\end{aligned}$$

As we have already seen $T^{\mathcal{A}} \leq \frac{1}{2h^2}\ln\frac{1}{\gamma} + 2$ a.s., $T^{\mathcal{A}} > \frac{1}{2h^2}\ln\frac{1}{\gamma}$ is with probability one only possible when $T^{\mathcal{A}} \in \{T+1, T+2\}$ where $T := \lceil\frac{1}{2h^2}\ln\frac{1}{\gamma}\rceil$. A look at its proof shows that

Lem. 2.10(i) also holds for $T+1$ and $T+2$ instead of T, hence we obtain

$$\begin{aligned}
&\mathbb{P}_p\left(\mathbf{D}(\mathcal{A}) = 1 \text{ and } T^{\mathcal{A}} > \frac{1}{2h^2}\ln\frac{1}{\gamma}\right)\\
&\leq \mathbb{P}_p\left(\mathbf{D}(\mathcal{A}) = 1 \text{ and } T^{\mathcal{A}} = T+1\right) + \mathbb{P}_p\left(\mathbf{D}(\mathcal{A}) = 1 \text{ and } T^{\mathcal{A}} = T+2\right)\\
&\leq \mathbb{P}_p\left(\frac{1}{T+1}\sum\nolimits_{t=1}^{T+1} X_t \geq 1/2\right) + \mathbb{P}_p\left(\frac{1}{T+2}\sum\nolimits_{t=1}^{T+2} X_t \geq 1/2\right) \leq \gamma+\gamma = 2\gamma
\end{aligned}$$

Consequently, $\mathcal{A}$ solves $\mathcal{P}_{\text{Coin}}^{h,3\gamma}$. □

Proposition 2.26. *Let $\varepsilon \in (0,1)$ be fixed, define $c_\varepsilon := \frac{2+\varepsilon}{\varepsilon}\left(\frac{1}{\ln(1+\varepsilon)}\right)^{1+\varepsilon}$. Let $\gamma \in (0,1)$ be such that $\delta := \frac{\gamma}{8c_\varepsilon} \in \left(0, \frac{\ln(1+\varepsilon)}{ec_\varepsilon}\right)$ holds and choose $\beta \in (0,3]$ arbitrarily. Then, there exists a constant $\lambda > 0$ with the following property: Denote by $\mathcal{A}$ Alg. 1 with sample access to $\{X_n\}_{n\in\mathbb{N}}$, where the black-box component $\mathcal{A}_{\text{MAB}}$ is* LIL'UCB *instantiated with $m=2$, $\delta, \varepsilon, \lambda$ and β. Then, $\mathcal{A}$ solves $\mathcal{P}_{\text{Coin}}^{\gamma}$ and fulfills*

$$\sup\nolimits_{p:|p-1/2|>h} \mathbb{P}_p\left(T^{\mathcal{A}} \leq C(h,\gamma)\right) \geq 1-\gamma \tag{2.11}$$

where $C(h,\gamma) \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ as $\min\{h,\gamma\} \to 0$. Moreover, for any $t' \in \mathbb{N}$ and $p \in [0,1]$ we have $\mathbb{P}_p(T^{\mathcal{A}} > t') > 0$.

Proof of Prop. 2.26. To see indirectly that $\mathcal{A}$ terminates almost surely for $p \neq 1/2$, assume $\mathcal{A}$ does not terminate, i.e., $\mathcal{A}_{\text{MAB}}$ does not terminate. According to Line 1 in Alg. 2 this means $n^{[1]}(t) \to \infty$ and $n^{[2]}(t) \to \infty$ as $t \to \infty$. But the law of large numbers ensures that

$$\frac{1}{n^{[i]}(t)}\sum\nolimits_{n'=1}^{n^{[i]}(t)} Z_{n'}^{[i]} + (1+\beta)(1+\sqrt{\varepsilon})\sqrt{\frac{(1+\varepsilon)\ln\left(\ln\left((1+\varepsilon)n^{[i]}(t)\right)/\delta\right)}{2n^{[i]}(t)}} \longrightarrow p^{[i]} \quad \text{a.s.}$$

as $n^{[i]}(t) \to \infty$. Therefore, $p^{[1]} = p \neq 1/2 = p^{[2]}$ and Line 2 let us infer that with probability 1 the best arm is pulled infinitely often, whereas the other suboptimal one is only pulled finitely often. According to Line 1 in Alg. 2 this is not possible. Consequently, $\mathcal{A}$ terminates a.s.

According to Thm. 2 in [Jamieson et al., 2013], with probability $\geq 1-4\sqrt{c_\varepsilon\delta}-4c_\varepsilon\delta \geq 1-\gamma$, $\mathcal{A}_{\text{MAB}}$ correctly identifies the best of the arms $a^{[1]}, a^{[2]}$ from Alg. 1 and stops after observing at most $\frac{c_1}{h^2}\ln\frac{1}{\delta} + \frac{c_3}{h^2}\max\left\{0, \ln\ln\frac{1}{h^2}\right\}$ samples from both arms, where $c_1, c_3 > 0$ are constant and only depend on ε and β. As $T^{\mathcal{A}}$ corresponds to the number of times arm $a^{[1]}$ has been pulled, we infer (2.11). As the output of $\mathcal{A}$ is correct iff $\mathcal{A}_{\text{MAB}}$ outputs the index of the best arm, $\mathcal{A}$ solves $\mathcal{P}_{\text{Coin}}^{\gamma}$.

Finally, suppose $\lambda > 1$ and let $t' \in \mathbb{N}$ and $p \in [0,1]$ be arbitrary but fixed. Writing $i^{\perp} := 3-i$, i.e., $1^{\perp} = 2$ and $2^{\perp} = 1$, we have

$$\begin{aligned}
\mathbb{P}_p\left(T^{\mathcal{A}} \geq t'\right) &= \mathbb{P}_p\left(\forall \tilde{t} \leq t'\, \forall i \in \{1,2\} : n^{[i]}(\tilde{t}) < 1 + \lambda n^{[i^{\perp}]}(\tilde{t})\right)\\
&\geq \mathbb{P}_p\left(\forall \tilde{t} \leq t'\, \forall i \in \{1,2\} : n^{[i]}(\tilde{t}) < 1 + n^{[i^{\perp}]}(\tilde{t})\right)\\
&= \mathbb{P}\left(\forall \tilde{t} \leq t' \,:\, |n^{[1]}(\tilde{t}) - n^{[2]}(\tilde{t})| \leq 1\right)
\end{aligned}$$

Since $Z_n^{[2]} \sim \mathrm{Ber}(\frac{1}{2})$ implies $\mathbb{P}(\forall n \in [N] : Z_n^{[2]} = z_n) > 0$ for every $(z_1, \dots, z_N) \in \{0,1\}^N$ and every $N \in \mathbb{N}$, this latter probability is positive and we conclude $\mathbb{P}_p\left(T^{\mathcal{A}} \geq t'\right) > 0$. □

Proposition 2.27. *Let $\gamma \in (0,1)$ be arbitrary. Write $\mathcal{A}$ for Alg. 1 with sample access to $\{X_n\}_{n\in\mathbb{N}}$ with $\mathcal{A}_{\mathrm{MAB}}$ chosen to be Alg. 3 called with γ. Then, $\mathcal{A}$ solves $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ and fulfills*

$$\sup\nolimits_{p:|p-1/2|>h} \mathbb{P}_p\left(T^{\mathcal{A}} \leq C(h,\gamma)\right) \geq 1-\gamma, \tag{2.12}$$

where $C(h,\gamma) \in \mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ as $\min\{h,\gamma\} \to 0$. Moreover, for any $t' \in \mathbb{N}$ and any $p \in [0,1]$ we have $\mathbb{P}_p(T^{\mathcal{A}} > t') > 0$.

Proof of Prop. 2.27. For $p \neq \frac{1}{2}$, the law of large numbers assures that the average reward over n pulls of $a^{[k]}$ converges to $p^{[k]}$ as $n \to \infty$, for $k \in \{1,2\}$, respectively. Together with $p^{[1]} = p \neq \frac{1}{2} = p^{[2]}$ this shows that both the inner and the outer while loops in Alg. 3 are left with probability 1, i.e., $\mathcal{A}$ terminates a.s. for any $p \neq \frac{1}{2}$.

According to Thm. 3.1 in [Karnin et al., 2013], $\mathcal{A}_{\mathrm{MAB}}$ outputs with probability $\geq 1-\gamma$ the index of the best arm with at most $\mathcal{O}\left(\frac{1}{h^2}\left(\ln\ln\frac{1}{h}\right)\ln\frac{1}{\gamma}\right)$ pulls of arms. Consequently, $\mathcal{A}$ solves $\mathcal{P}^{\gamma}_{\mathrm{Coin}}$ and fulfills (2.12).

Now, let $t' \in \mathbb{N}$ be arbitrary. At termination, both arms $a^{[1]}$ and $a^{[2]}$ have been pulled exactly the same number of times. With regard to Line 10 of Alg. 3 we thus get

$$\mathbb{P}_p(T^{\mathcal{A}} > t') \geq \mathbb{P}_p(\forall \tilde{t} \leq 2t' : Z_{\tilde{t}}^{[1]} = Z_{\tilde{t}}^{[2]}) > 0,$$

for every $p \neq \frac{1}{2}$, where the last inequality follows since $p^{[2]} = \frac{1}{2}$ assures $\mathbb{P}((Z_1^{[2]}, \dots, Z_{2t'}^{[2]}) = (z_1, \dots, z_{2t'})) > 0$ for every $(z_1, \dots, z_{2t'}) \in \{0,1\}^{2t'}$. □

Lemma 2.31. *Let $k \in \mathbb{N}$ and $(p_1, \dots, p_k) \in \Delta_k$ be fixed. Suppose $\{X_t\}_{t\in\mathbb{N}}$ to be an iid family of random variables $X_t \sim \mathrm{Cat}(p_1, \dots, p_k)$ on some joint probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\mathcal{F}_t\}_{t\in\mathbb{N}} \subseteq \mathcal{F}$ to be a filtration, such that $\{X_t\}_t$ is $\{\mathcal{F}_t\}_t$-adapted and $\forall t : X_t \perp\!\!\!\perp \mathcal{F}_{t-1}$, e.g. $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$. If τ is an $\{\mathcal{F}_t\}_t$-stopping time, then the random variables*

$$T_i(\tau) \coloneqq \sum\nolimits_{t\leq\tau} \mathbf{1}_{\{X_t=i\}}, \quad i \in [k],$$

fulfill $\mathbb{E}[T_i(\tau)] = p_i\mathbb{E}[\tau]$ for each $i \in [k]$. In particular, we obtain

$$\mathbb{E}[\tau] = \frac{\sum_{i\in I}\mathbb{E}[T_i(\tau)]}{\sum_{i\in I} p_i}$$

for any $I \subseteq [k]$ with $\sum_{i\in I} p_i > 0$.

Proof of Lem. 2.31. Since $\{t \leq \tau\} = \{t > \tau\}^c = \{\tau \leq t-1\}^c \in \mathcal{F}_{t-1}$ holds for any $t \in \mathbb{N}$ and $X_t \perp\!\!\!\perp \mathcal{F}_{t-1}$, we obtain

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\{X_t=i\}}\mathbf{1}_{\{t\leq\tau\}}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{X_t=i\}}\mathbf{1}_{\{t\leq\tau\}}\middle|\mathcal{F}_{t-1}\right]\right] \\ &= \mathbb{E}[\mathbf{1}_{\{t\leq\tau\}}\mathbb{E}[\mathbf{1}_{\{X_t=i\}}|\mathcal{F}_{t-1}]] = p_i\mathbb{E}[\mathbf{1}_{\{t\leq\tau\}}].\end{aligned}$$

Via an application of the monotone convergence theorem we infer

$$\begin{aligned}\mathbb{E}\left[T_i(\tau)\right] &= \lim\nolimits_{T\to\infty}\mathbb{E}[T_i(\tau\wedge T)] = \lim\nolimits_{T\to\infty}\sum\nolimits_{t\leq T}\mathbb{E}\left[\mathbf{1}_{\{X_t=i\}}\mathbf{1}_{\{t\leq\tau\}}\right] \\ &= p_i\lim\nolimits_{T\to\infty}\sum\nolimits_{t\leq T}\mathbb{E}\left[\mathbf{1}_{\{t\leq\tau\}}\right] = p_i\lim\nolimits_{T\to\infty}\mathbb{E}[\tau\wedge T] = p_i\mathbb{E}[\tau].\end{aligned}$$

and thus in particular $\sum_{i\in I}\mathbb{E}[T_i(\tau)] = \mathbb{E}[\tau]\sum_{i\in I} p_i$. □

Lemma 2.32. *Suppose* $\mathbf{p} \in \Delta_k^h \setminus \Delta_k^{\tilde{h}}$ *for some* $0 < h < \tilde{h} < 1$ *and let* $i := \mathrm{mode}(\mathbf{p})$ *and* $j \in \mathrm{argmax}_{l \in [k] \setminus \{i\}} p_l$. *Then, we have* $p_i + p_j \geq \frac{2+(k-2)h}{k}$ *and* $p_i - p_j < \tilde{h}$.

Proof of Lem. 2.32. From $\mathbf{p} \in \Delta_k^h$ and $\mathrm{mode}(\mathbf{p}) = i$ we infer that $p_l \leq p_i - h$ holds for each $l \in [k] \setminus \{i\}$. Thus,

$$1 = \sum\nolimits_{l \in [k]} p_l \leq p_i + \sum\nolimits_{l \neq i} (p_i - h) = kp_i - (k-1)h$$

shows us that $p_i = \frac{1+(k-1)h}{k} + \varepsilon$ for some $\varepsilon \geq 0$. Due to $\sum_{l \neq i} p_l = 1 - p_i$ and $p_j = \max_{l \in [k] \setminus \{i\}} p_l$, we have

$$p_j \geq \frac{1-p_i}{k-1} = \frac{1 - \frac{1+(k-1)h}{k} - \varepsilon}{k-1} = \frac{1-h}{k} - \frac{\varepsilon}{k-1}.$$

Consequently,

$$\begin{aligned} p_i + p_j &\geq \frac{1+(k-1)h}{k} + \varepsilon + \frac{1-h}{k} - \frac{\varepsilon}{k-1} \\ &= \frac{2+(k-2)h}{k} + \frac{(k-2)\varepsilon}{k-1} \\ &\geq \frac{2+(k-2)h}{k}. \end{aligned}$$

Moreover, $\mathbf{p} \notin \Delta_k^{\tilde{h}}$ assures the existence of some $j' \in [k] \setminus \{i\}$ with $p_i < p_{j'} + \tilde{h}$. Since the choice of j guarantees $p_{j'} + \tilde{h} \leq p_j + \tilde{h}$, this implies $p_i - p_j < \tilde{h}$. □

Lemma 2.34 (Dvoretzky-Kiefer-Wolfowitz inequality for categorical random variables)**.** *Suppose* $X_1, X_2, \ldots$ *to be iid random variables* $X_n \sim \mathrm{Cat}(\mathbf{p})$ *for some* $\mathbf{p} \in \Delta_k$. *For* $t \in \mathbb{N}$ *let* $\hat{\mathbf{p}}^t$ *be the corresponding empirical distribution after the* t *observations* $X_1, \ldots, X_t$, *i.e.,* $\hat{p}_i^t = \frac{1}{t}\sum_{s=1}^t \mathbf{1}_{\{X_s = i\}}$ *for all* $i \in [k]$. *Then, we have for any* $\varepsilon > 0$ *and* $t \in \mathbb{N}$ *the estimate*

$$\mathbb{P}\left(\left|\left|\hat{\mathbf{p}}^t - \mathbf{p}\right|\right|_\infty > \varepsilon\right) \leq 4e^{-t\varepsilon^2/2}.$$

Proof of Lem. 2.34. Confer [Dvoretzky et al., 1956, Massart, 1990] as well as Thm. 11.6 in [Kosorok, 2008]. Moreover, note that the cumulative distribution functions[2] F resp. $\hat{F}^t$ of $X_1 \sim \mathrm{Cat}(\mathbf{p})$ resp. $\hat{\mathbf{p}}^t$ fulfill $p_j = F(j) - F(j-1)$ and $\hat{p}_j^t = \hat{F}^t(j) - \hat{F}^t(j-1)$ and thus

$$|\hat{p}_j^t - p_j| \leq |\hat{F}^t(j) - F(j)| + |\hat{F}^t(j-1) - F(j-1)|.$$

for each $j \in [k]$. □

Lemma 2.35. *For* $h \in [0,1], \varepsilon \in (-h, 1], \mathbf{p} \in \Delta_k^h$ *and* $\tilde{\mathbf{p}} \in \Delta_k$ *we have*

$$(\exists i : \tilde{p}_i - \max_{j \neq i} \tilde{p}_j \geq \varepsilon \text{ and } p_i \neq \max_j p_j) \quad \Rightarrow \quad ||\tilde{\mathbf{p}} - \mathbf{p}||_\infty \geq \frac{h+\varepsilon}{2}.$$

[2]Recall that the cumulative distribution function of a real-valued random variable X is defined as $F(x) = \mathbb{P}(X \leq x)$ for $x \in \mathbb{R}$.

Proof of Lem. 2.35. Suppose there is some $i \in [k]$ s.t. $\tilde{p}_i - \max_{j \neq i} \tilde{p}_j \geq \varepsilon$ and $p_i \neq \max_j p_j$ hold. Then, there exists some $j \in [k] \setminus \{i\}$ with

$$p_j \geq p_i + h \quad \text{and} \quad \tilde{p}_i \geq \tilde{p}_j + \varepsilon$$

and we conclude

$$2\,||\tilde{\mathbf{p}} - \mathbf{p}||_\infty \geq |p_j - \tilde{p}_j| + |\tilde{p}_i - p_i| \geq (p_j - p_i) + (\tilde{p}_i - \tilde{p}_j) \geq h + \varepsilon.$$

□

Lemma 2.38. *Let $h > 0$, $\mathbf{p} \in \Delta_k^{3h}$ and $\tilde{\mathbf{p}} \in \Delta_k$ be fixed. Then,*

$$\forall i : \tilde{p}_i \leq \max_{j \neq i} \tilde{p}_j + h \quad \Rightarrow \quad ||\mathbf{p} - \tilde{\mathbf{p}}||_\infty \geq h.$$

Proof of Lem. 2.38. To prove the contraposition, we suppose $||\mathbf{p} - \tilde{\mathbf{p}}||_\infty < h$ to be fulfilled. Let $i := \mathrm{mode}(\mathbf{p}) \in [k]$ and fix some arbitrary $j \in [k] \setminus \{i\}$. Since $\mathbf{p} \in \Delta_k^{3h}$ assures $p_i \geq p_j + 3h$, we obtain

$$\begin{aligned}\tilde{p}_i - \tilde{p}_j &= p_i + (\tilde{p}_i - p_i) + (p_j - \tilde{p}_j) - p_j \geq p_i - p_j - 2\,||\mathbf{p} - \tilde{\mathbf{p}}||_\infty \\ &> p_i - p_j - 2h \geq h.\end{aligned}$$

As j was arbitrary, we conclude that $\tilde{p}_i > \max_{j \neq i} \tilde{p}_j + h$, which completes the proof. □

Lemma 2.39. *For any $h \in (0, \frac{1}{8})$, $\varepsilon \in (0, \frac{1}{3})$ and $k \in \mathbb{N}_{\geq 3}$ there exist $\mathbf{p} \in \Delta_k^{(3-\varepsilon)h}$ and $\tilde{\mathbf{p}} \in \Delta_k$ such that*

$$\forall i \in [k] : \tilde{p}_i \leq \max_{j \neq i} \tilde{p}_j + h \quad \text{and} \quad ||\mathbf{p} - \tilde{\mathbf{p}}||_\infty < h.$$

Proof of Lem. 2.39. Suppose $h \in (0, \frac{1}{8})$, $\varepsilon \in (0, \frac{1}{3})$ and $k \in \mathbb{N}_{\geq 3}$ to be fixed. Now, define $\mathbf{p} \in \Delta_k$ and $\tilde{\mathbf{p}} \in \Delta_k$ via

$$p_j := \begin{cases} \frac{1}{2} + h, & \text{if } j = 1, \\ \frac{1}{2} - (2 - \varepsilon)h, & \text{if } j = 2, \\ \frac{(1-\varepsilon)h}{k-2}, & \text{if } j \geq 3, \end{cases}$$

and

$$\tilde{p}_j := \begin{cases} p_1 - (1 - \frac{\varepsilon}{4})h = \frac{1}{2} + \frac{\varepsilon h}{4}, & \text{if } j = 1, \\ p_2 + (1 - \frac{\varepsilon}{4})h = \frac{1}{2} + (\frac{3\varepsilon}{4} - 1)h, & \text{if } j = 2, \\ \frac{(1-\varepsilon)h}{k-2}, & \text{if } j \geq 3. \end{cases}$$

From $h < \frac{1}{8}$ we infer $\frac{1}{2} - (2 - \varepsilon)h > \frac{1}{2} - 2h > \frac{1}{4}$ and thus

$$\forall j \geq 3 : \frac{(k-2)p_j}{p_2} = \frac{(1-\varepsilon)h}{\frac{1}{2} - (2-\varepsilon)h} < 4(1-\varepsilon)h < 4h < \frac{1}{2} < k - 2.$$

This shows $p_1 - (3 - \varepsilon)h = p_2 > \max_{j \geq 3} p_j$ and consequently $\mathbf{p} \in \Delta_k^{(3-\varepsilon)h}$. Since $\tilde{p}_j = p_j$ is fulfilled for each $j \geq 3$, we have $\tilde{p}_1 > \tilde{p}_2 > p_2 > \max_{j \geq 3} \tilde{p}_j$, and together with

$$\tilde{p}_1 - \tilde{p}_2 = \frac{\varepsilon h}{4} - \frac{3\varepsilon h}{4} + h = \left(1 - \frac{\varepsilon}{2}\right) h < h$$

we see that $\tilde{p}_i \leq \max_{j \neq i} \tilde{p}_j + h$ holds for each $i \in [k]$. Finally $||\mathbf{p} - \tilde{\mathbf{p}}||_\infty < h$ follows from $|p_1 - \tilde{p}_1| = (1 - \frac{\varepsilon}{4})h = |p_2 - \tilde{p}_2|$ as well as $p_j = \tilde{p}_j$ for all $j \geq 3$. □

Lemma 2.46. *For any $h \in (0,1)$ and $m, k \in \mathbb{N}$ with $k \leq m$ we have $PM_k^m(\mathrm{PL} \wedge h\mathrm{GCW}) \supseteq PM_k^m(\mathrm{PL} \wedge \Delta^h) \neq \emptyset$.*

Proof of Lem. 2.46. Note that $PM_k^m(\mathrm{PL} \wedge h\mathrm{GCW}) \supseteq PM_k^m(\mathrm{PL} \wedge \Delta^h)$ is a direct consequence from the definitions. To see $PM_k^m(\mathrm{PL} \wedge \Delta^h) \neq \emptyset$ we fix $x > 1$ with $h + \frac{h}{x} \leq 1$ and define $\boldsymbol{\theta} \in (0,1]^m$ via $\theta_j := \frac{h^j}{(kx)^j}$ for any $j \in [m]$. Then,

$$
\begin{aligned}
&h(\theta_i + \cdots + \theta_{i+k-1}) + \theta_{i+1} - \theta_i \\
&= \frac{h^{i+1}}{(kx)^i} + \left(\frac{h^{i+2}}{(kx)^{i+1}} + \cdots + \frac{h^{i+k}}{(kx)^{i+k-1}} + \frac{h^{i+1}}{(kx)^{i+1}} \right) - \frac{h^i}{(kx)^i} \\
&\leq \frac{h^{i+1}}{(kx)^i} + \frac{kh^{i+1}}{(kx)^{i+1}} - \frac{h^i}{(kx)^i} = \frac{h^i}{(kx)^i}\left(h + \frac{h}{x} - 1 \right) \leq 0
\end{aligned}
$$

holds for any $i \in [m-k]$ and thus $\mathbf{P}(\boldsymbol{\theta}) \in PM_k^m(\mathrm{PL} \wedge \Delta^h)$ follows from Lem. 2.45. □

B. Remaining Proofs for Chapter 3

Lemma 3.2. *Some $G \in \overline{\mathcal{G}}_m$ is acyclic iff it does not contain a 3-cycle.*

Proof of Lem. 3.2. If G is acyclic then it clearly does not contain a 3-cycle. Therefore, we might suppose that G contains a cycle. Then

$$k^* := \min\{k \in \mathbb{N} \,|\, G \text{ contains a } k\text{-cycle}\}$$

exists and there is a cycle $(i_1, \ldots, i_{k^*})$ in G. To show $k^* \leq 3$ indirectly, let us assume that $k^* \geq 4$ holds. Since G is complete, either $i_1 \to i_3$ or $i_3 \to i_1$ in G. In the first case, we would have the (k^*-1)-cycle $(i_1, i_3, i_4, \ldots, i_{k^*})$, and in the second one, we would have the 3-cycle (i_1, i_2, i_3), which is in both cases a contradiction to the minimality of k^*. ↯ Hence, $k^* = 3$ holds, which completes the proof. □

Lemma 3.3. *For any $\mathbf{Q} \in \mathcal{Q}_m^0$ there exists a permutation σ on $[m]$ s.t. $q_{\sigma(i),\sigma(i+1)} > \frac{1}{2}$ for every $i \in [m-1]$.*

Proof of Lem. 3.3. By replacing $q_{i,j}$ with 1 if $q_{i,j} > 1/2$ and with 0 if $q_{i,j} < 1/2$, we may assume w.l.o.g. $\mathbf{Q} \in \mathcal{R}_m$. The Theorem of Rédei (cf. e.g. [Sachs, 1971]) assures that $\Phi_m(\mathbf{Q})$ contains a Hamiltonian path, i.e., there exists a permutation σ on $[m]$ such that $(\sigma(i), \sigma(i+1)) \in E_G$ for every $i \in [m-1]$. Regarding the definition of Φ_m, this proves the statement. □

Lemma 3.6. *Let $\mathrm{X}, \mathrm{X}_1, \mathrm{X}_2$ and Y be possible properties of tournaments on $[m]$. Then, we have:*

(i) If $G \in \overline{\mathcal{G}}_m$ fulfills $G \in \overline{\mathcal{G}}_m(\mathrm{Y})$ and has a subgraph $\tilde{G} \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$, then $G \in \overline{\mathcal{G}}_m(\mathrm{X})$.

(ii) $\mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y}) \cap \overline{\mathcal{G}}_m = \overline{\mathcal{G}}_m(\mathrm{X} \wedge \mathrm{Y})$.

(iii) $\mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \subseteq \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$ iff $\overline{\mathcal{G}}_m(X_1 \wedge \mathrm{Y}) \subseteq \overline{\mathcal{G}}_m(\mathrm{X}_2 \wedge \mathrm{Y})$.

(iv) If $G \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$ and $G' \in \mathcal{G}_m(\mathrm{Y})$ with $E_G \subseteq E_{G'}$, then $G' \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$.

(v) $\mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y}) \subseteq \mathcal{G}_m \setminus \mathcal{G}_m(\neg \mathrm{Y})$ with equality iff $\overline{\mathcal{G}}_m(\mathrm{X}) = \overline{\mathcal{G}}_m(\mathrm{Y})$.

(vi) $\mathcal{G}_m(\mathrm{X}_1 \wedge \mathrm{X}_2 \,|\, \mathrm{Y}) = \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cap \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$.

(vii) $\mathcal{G}_m(\mathrm{X}_1 \vee \mathrm{X}_2 \,|\, \mathrm{Y}) \supseteq \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cup \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$.

(viii) If $\overline{\mathcal{G}}_m(\mathrm{X}_1) \cap \overline{\mathcal{G}}_m(\mathrm{X}_2) = \emptyset$, then $\mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cap \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y}) = \emptyset$.

(ix) If $\overline{\mathcal{G}}_m(\mathrm{Y}) \subseteq \overline{\mathcal{G}}_m(\mathrm{X})$, then $\mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y}) = \{G \in \mathcal{G}_m \,|\, \exists\, \mathrm{Y}\text{-extension of } G\}$.

Proof of Lem. 3.6. (i) Any Y-extension of a subgraph $\tilde{G} \in \mathcal{G}_m$ of $G \in \overline{\mathcal{G}}_m$ is also a Y-extension of G.

(ii) The only extension of $G \in \overline{\mathcal{G}}_m$ is itself.

(iii) The first implication "$\Rightarrow$" follows from (ii). To show the other one, suppose $\overline{\mathcal{G}}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \subseteq \overline{\mathcal{G}}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$. For $G \in \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y})$, any Y-extension is an element of $\overline{\mathcal{G}}_m(\mathrm{X}_1 \wedge \mathrm{Y}) \subseteq \overline{\mathcal{G}}_m(\mathrm{X}_2 \wedge \mathrm{Y})$, hence $G \in \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$.

(iv) Any Y-extension of G is also a Y-extension of any of its supergraphs.

(v) If $G \in \mathcal{G}_m(\neg\mathrm{Y})$, (iv) assures that each extension of it is in $\mathcal{G}_m(\neg\mathrm{Y}) \cap \overline{\mathcal{G}}_m = \overline{\mathcal{G}}_m(\neg\mathrm{Y})$. In other words, G has no Y-extension and thus $G \notin \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$.
To show the first part of the equivalence indirectly, suppose (possibly after replacing X with $\mathrm{X} \wedge \mathrm{Y}$) w.l.o.g. that some $G \in \overline{\mathcal{G}}_m(\mathrm{Y}) \setminus \overline{\mathcal{G}}_m(\mathrm{X})$ exists. Then, $G \in \overline{\mathcal{G}}_m(\neg\mathrm{Y}) \subseteq \mathcal{G}_m(\neg\mathrm{Y})$ holds and at the same time (ii) shows $G \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$.
For seeing the second part of the equivalence let $\overline{\mathcal{G}}_m(\mathrm{X}) = \overline{\mathcal{G}}_m(\mathrm{Y})$ and fix $G \in \mathcal{G}_m \setminus \mathcal{G}_m(\neg\mathrm{Y})$. Then, there exists a Y-extension of G and by assumption this has the property $\mathrm{X} = \mathrm{Y}$. Consequently, $G \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$.

(vi) For $G \in \mathcal{G}_m$ we have the equivalences

$$\begin{aligned} G \in \mathcal{G}_m(\mathrm{X}_1 \wedge \mathrm{X}_2 \,|\, \mathrm{Y}) \quad &\Leftrightarrow \quad G \text{ has a Y-extension and every Y-extension of G is in} \\ &\qquad \overline{\mathcal{G}}_m(\mathrm{X}_1 \wedge \mathrm{X}_2) = \overline{\mathcal{G}}_m(\mathrm{X}_1) \cap \overline{\mathcal{G}}_m(\mathrm{X}_2) \\ &\Leftrightarrow \quad G \in \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \wedge \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y}). \end{aligned}$$

(vii) Suppose $G \in \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y}) \cup \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$, and assume w.l.o.g. $G \in \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y})$. Any Y-extension of G is an element of $\overline{\mathcal{G}}_m(\mathrm{X}_1 \wedge \mathrm{Y}) \subseteq \overline{\mathcal{G}}_m((\mathrm{X}_1 \vee \mathrm{X}_2) \wedge \mathrm{Y})$, hence $G \in \mathcal{G}_m(\mathrm{X}_1 \vee \mathrm{X}_2 \,|\, \mathrm{Y})$ follows.

(viii) Let $G \in \mathcal{G}_m(\mathrm{X}_1 \,|\, \mathrm{Y})$. Any Y-extension G' of G is in $\overline{\mathcal{G}}_m(\mathrm{X}_1)$ and thus not in $\overline{\mathcal{G}}_m(\mathrm{X}_2)$. Therefore, $G \notin \mathcal{G}_m(\mathrm{X}_2 \,|\, \mathrm{Y})$.

(ix) By definition, any $G \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$ requires the existence of a Y-extension of G. On the other side, if some $G \in \mathcal{G}_m$ with a Y-extension is fixed, then each of its Y-extensions is an element of $\overline{\mathcal{G}}_m(\mathrm{Y}) \subseteq \overline{\mathcal{G}}_m(\mathrm{X})$, which implies $G \in \mathcal{G}_m(\mathrm{X} \,|\, \mathrm{Y})$.

□

Lemma 3.24. *Let $m \geq 3$ and $S \subseteq [m]$.*

(a) If $|S| > m - \lfloor \frac{m}{3} \rfloor$, there exists $i \in [m-2]$ with $i, i+1, i+2 \in S$.

(b) If $|S| > m - \lfloor \frac{m+2}{3} \rfloor$, then $\{1,2\} \subseteq S$ or $\{m-1, m\} \subseteq S$ or there exists $i \in [m-2]$ with $i, i+1, i+2 \in S$.

Proof of Lem. 3.24. (a) If $|S| > m - \lfloor \frac{m}{3} \rfloor$, the set $S^c = [m] \setminus S$ has $< \lfloor \frac{m}{3} \rfloor$ elements. In particular, at least one of the $\lfloor \frac{m}{3} \rfloor$ disjoint sets

$$A_l := S^c \cap \{3l+1, 3l+2, 3l+3\}, \quad 0 \leq l \leq \left\lfloor \frac{m}{3} \right\rfloor - 1,$$

is empty. With $l' \in \{0, \dots \lfloor \frac{m}{3} \rfloor - 1\}$ such that $A_{l'} = \emptyset$ we can thus conclude $\{3l'+1, 3l'+2, 3l'+3\} \subseteq S$.

(b) As $m = 3$ and $m = 4$ are trivial, suppose w.l.o.g. $m \geq 5$ and let $S \subseteq [m]$ with $|S| > m - \lfloor \frac{m+2}{3} \rfloor$, $\{1, 2\} \not\subseteq S$ and $\{m-1, m\} \not\subseteq S$. Then, $S' := S \setminus \{1, 2, m-1, m\}$ fulfills

$$|S'| \geq |S| - 2 > (m-2) - \left\lfloor \frac{m+2}{3} \right\rfloor = (m-4) - \left\lfloor \frac{m-4}{3} \right\rfloor.$$

By replacing $[m]$ with $[m] \setminus \{1, 2, m-1, m\}$, (a) lets us conclude that some $3 \leq i \leq m-3$ exists with $i, i+1, i+2 \in S' \subsetneq S$.

$\square$