Stochastik

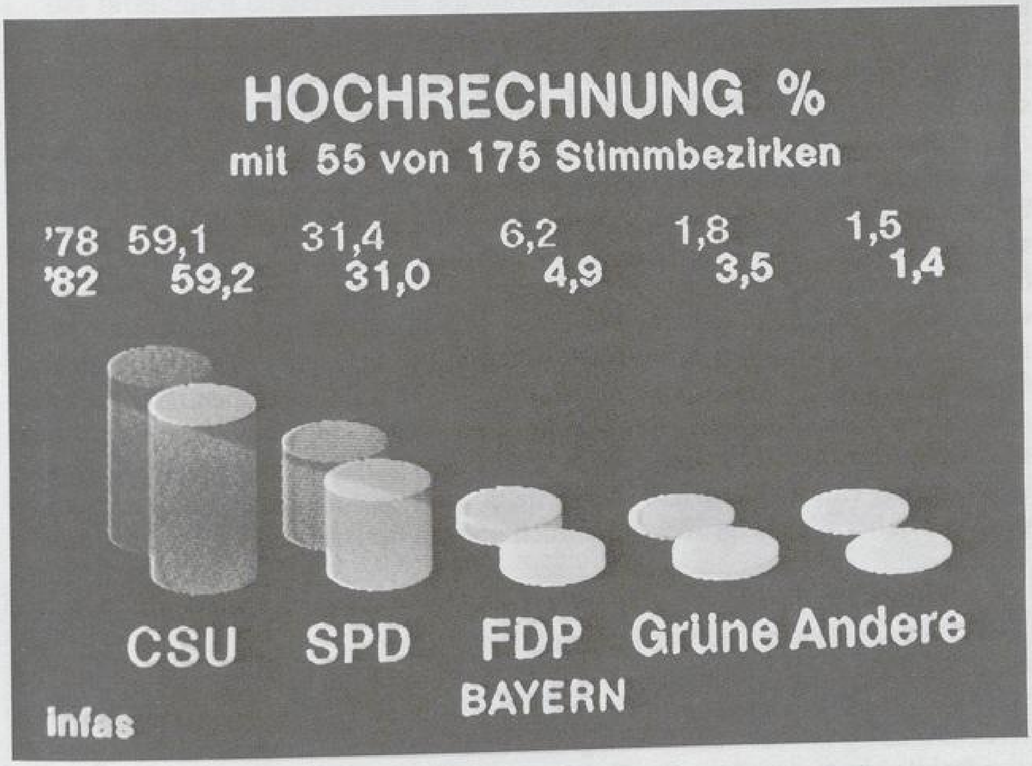
Barth, Friedrich

München, [20]03

18. Parameterschätzung

urn:nbn:de:hbz:466:1-83580

*18. Parameterschätzung



Landtagswahl in Bayern am 10. Oktober 1982 – Sendung der ARD

*18. Parameterschätzung

18.1. Problemstellung

Im vorausgegangenen Kapitel haben wir dargestellt, wie man Testprobleme lösen kann. Wir wenden uns nun der anderen typischen Fragestellung der Mathematischen Statistik zu, dem Schätzproblem, das wir auch nur für einfache Fälle angehen wollen.
Im einfachsten Fall handelt es sich darum, auf Grund eines Stichprobenergebnisses die Wahrscheinlichkeit p eines Ereignisses zu schätzen. Im Urnenmodell bedeutet dies, den Anteil p einer Kugelsorte zu schätzen. Als erster hat *Thomas Bayes* (1702–1761) diese Aufgabe in seiner erst 1763 erschienenen berühmten Schrift *An Essay towards solving a problem in the Doctrine of Chances* gestellt und unter der Voraussetzung, daß alle Werte von p aus $[0; 1]$ gleichwahrscheinlich in Frage kommen, durch eine **Intervallschätzung** für p gelöst. Erst 1934 gelang es *Jerzy Neyman* (1894–1981), das Problem allgemein durch Einführung der Konfidenzintervalle, wie wir sie in **14.8.** beschrieben haben, zu lösen.
Allgemeiner geht es darum, auf Grund von Stichprobenergebnissen gewisse Parameter der Wahrscheinlichkeitsverteilung einer Zufallsgröße zu schätzen. Solche Parameter sind beispielsweise der Parameter p einer Binomialverteilung, der Erwartungswert μ, der Median und die Quantile, die Standardabweichung und die Schiefe einer irgendwie gearteten Verteilung, ja sogar der Umfang der Grundgesamtheit.
Sei nun ϑ ein solcher zu schätzender Parameter der Wahrscheinlichkeitsverteilung einer Zufallsgröße X. Zu seiner Schätzung ziehen wir aus X eine Zufallsstichprobe $(X_1 | X_2 | \dots | X_n)$. Sie liefere das Stichprobenergebnis $(a_1 | a_2 | \dots | a_n)$. Aus diesen Stichprobenwerten a_i soll nun kein Intervall für ϑ, sondern durch eine geeignete Formel ein Näherungswert $\hat{\vartheta}$, eben ein Schätzwert, für den unbekannten Parameter ϑ errechnet werden. Man spricht dann von einer **Punktschätzung** für ϑ. Dieser Schätzwert ist somit eine Funktion der a_i; es gilt also $\hat{\vartheta} = T_n(a_1, a_2, \dots, a_n)$. Dabei soll der Index n anzeigen, daß eine Stichprobe der Länge n gezogen wurde. Der auf diese Weise errechnete Schätzwert $\hat{\vartheta}$ hängt natürlich vom Zufall ab; denn a_i ist ja der zufällige Wert aus der Wertemenge $\mathfrak{S} = \{x_1, x_2, \dots, x_s\}$ von X, den die Zufallsgröße X_i angenommen hat. $\hat{\vartheta}$ ist also aufzufassen als ein beobachteter Wert der aus den n Zufallsgrößen X_i gebildeten Stichprobenfunktion $T_n(X_1, X_2, \dots, X_n)$, die als Funktion von Zufallsgrößen selbst wieder eine Zufallsgröße ist. Sie heißt im Zusammenhang mit dem Schätzproblem daher »Schätzgröße« oder auch »Schätzfunktion«. Wir merken uns

Definition 376.1: Ist $(X_1 | X_2 | \dots | X_n)$ eine Stichprobe aus der Zufallsgröße X, dann heißt jede reellwertige Funktion

$$T_n\colon \quad (X_1 | X_2 | \dots | X_n) \mapsto T_n(X_1, X_2, \dots, X_n)$$

Schätzfunktion oder auch **Schätzgröße** für den reellen Parameter ϑ der Wahrscheinlichkeitsverteilung der Zufallsgröße X.

Eine so weit gefaßte Definition gibt uns keine Hilfe, wie man zu zweckmäßigen Schätzfunktionen gelangt. Und sie sagt uns erst recht nicht, welcher Schätzfunktion wir den Vorzug geben sollen, falls wir gar mehrere Schätzfunktionen gefunden haben.

18.2. Das Maximum-Likelihood-Prinzip

Ein besonders brauchbares Verfahren zur Gewinnung von Schätzgrößen führte 1760 *Johann Heinrich Lambert* (1728–1777) und unabhängig davon 1777 *Daniel Bernoulli* (1700–1782) in die Wahrscheinlichkeitsrechnung ein. *Carl Friedrich Gauß* (1777–1855) benützte es mehrfach, so z.B. 1798 zu seinem ersten Beweis der Methode der kleinsten Quadrate. Verallgemeinert hat das Verfahren aber erst 1912 *Ronald Aylmer Fisher* (1890–1962) zum

Maximum-Likelihood-Prinzip oder **Prinzip der maximalen Mutmaßlichkeit:** Eine Zufallsstichprobe $(X_1 | X_2 | \ldots | X_n)$ aus der Zufallsgröße X, deren Verteilung vom Parameter ϑ abhängt, zeitigte das Stichprobenergebnis $(a_1 | a_2 | \ldots | a_n)$. Als Schätzwert für den Parameter ϑ dient dann jeder Wert ϑ, für den die Wahrscheinlichkeit

$$P_\vartheta(X_1 = a_1 \wedge X_2 = a_2 \wedge \ldots \wedge X_n = a_n)$$

des tatsächlich eingetretenen Stichprobenergebnisses maximal wird.

Jedem möglichen Wert des Parameters ϑ wird also bei bekanntem Stichprobenergebnis eine Wahrscheinlichkeit zugeordnet. Diese Zuordnung

$$L\colon \vartheta \mapsto P(X_1 = a_1 \wedge X_2 = a_2 \wedge \ldots \wedge X_n = a_n)$$

heißt **Likelihood-Funktion** L. Eine Maximumstelle dieser Funktion L muß nicht notwendig existieren; andererseits kann es auch mehrere solcher Stellen geben. Ein nach dem Maximum-Likelihood-Prinzip bestimmter Schätzwert $\hat{\vartheta}(a_1, a_2, \ldots, a_n)$ heißt **Maximum-Likelihood-Schätzwert**, die zugehörige Zufallsgröße $\hat{\vartheta}(X_1, X_2, \ldots, X_n)$ dann **Maximum-Likelihood-Schätzgröße.**
Betrachten wir zum besseren Verständnis den besonders einfachen Fall, daß wir als Parameter ϑ den Anteil p einer Kugelsorte in einer Urne schätzen wollen. (Es sei $0 < p < 1$.) Wir entnehmen der Urne eine Kugel und betrachten die Zufallsgröße

$$X := \begin{cases} 1, & \text{falls die Kugel der Sorte angehört,} \\ 0 & \text{sonst.} \end{cases}$$

X ist nach $B(1; p)$ verteilt. Eine Stichprobe $(X_1 | X_2 | \ldots | X_n)$ der Länge n aus der Zufallsgröße X besteht im n-maligen Ziehen einer Kugel aus der Urne mit Zurücklegen. Die Wahrscheinlichkeit, daß bei einem gegebenen Kugelanteil p sich das Stichprobenergebnis $(a_1 | a_2 | \ldots | a_n)$, $a_i \in \{0; 1\}$, einstellt, wird durch die Likelihood-Funktion L in Abhängigkeit von p ausgedrückt:

$$L(p) := P_p(X_1 = a_1 \wedge X_2 = a_2 \wedge \ldots \wedge X_n = a_n).$$

Da in der Stichprobe die X_i stochastisch unabhängig sind, erhält man

$$L(p) = P_p(X_1 = a_1) \cdot P_p(X_2 = a_2) \cdot \ldots \cdot P_p(X_n = a_n).$$

Betrachten wir nun ein spezielles Stichprobenergebnis mit genau k Einsen, dann gilt

$$L(p) = p^k(1-p)^{n-k}.$$

Das Maximum dieser Wahrscheinlichkeit finden wir durch Differenzieren von $L(p)$ nach p. Für $0 < k < n$ erhalten wir – die Fälle $k = 0$ und $k = n$ erledigt man analog –

$$\begin{aligned}\frac{\mathrm{d}L(p)}{\mathrm{d}p} &= kp^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{n-k-1} = \\ &= -p^{k-1}(1-p)^{n-k-1}(np-k).\end{aligned}$$

Als Nullstelle ergibt sich $p = \frac{k}{n}$; der Vorzeichenwechsel von $\frac{\mathrm{d}L(p)}{\mathrm{d}p}$ zeigt, daß es sich um eine Maximumstelle handelt.

$\hat{p}(a_1, a_2, \ldots, a_n) = \frac{k}{n} = \frac{1}{n} \sum_{i=1}^{n} a_i$ ist der Maximum-Likelihood-Schätzwert für den Kugelanteil p in der Urne. Die zugehörige Maximum-Likelihood-Schätzgröße $\hat{p}(X_1, X_2, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$ ist aber nichts anderes als die uns längst bekannte Zufallsgröße relative Häufigkeit H_n.

Es ist erfreulich, daß auch das Maximum-Likelihood-Prinzip die relative Häufigkeit eines Ereignisses als brauchbare Schätzgröße für die Wahrscheinlichkeit eines Ereignisses liefert. Auf Grund der Interpretationsregel für Wahrscheinlichkeiten, die durch die Gesetze der großen Zahlen wissenschaftlich abgesichert ist, war die relative Häufigkeit immer schon ein brauchbarer »Meßwert« für die Wahrscheinlichkeit eines Ereignisses.

18.3. Beurteilungskriterien für Schätzfunktionen

Das Maximum-Likelihood-Prinzip ist ein Verfahren zur Gewinnung von Schätzfunktionen. Wie sollen wir uns aber entscheiden, wenn wir uns durch verschiedene Betrachtungsweisen mehrere Schätzfunktionen verschafft haben? Eine Festlegung auf eine Schätzfunktion ist nicht eindeutig möglich, da die Eignung einer Stichprobenfunktion zur Schätzung eines Parameters nach sehr unterschiedlichen Gesichtspunkten beurteilt werden kann. In den Jahren 1921 und 1925 hat *Ronald Aylmer Fisher* (1890–1962) vier Kriterien zur Beurteilung von Schätzfunktionen entwickelt.

1. Von einer Schätzgröße T_n für den Parameter ϑ wird man erwarten, daß ihre Werte, d.h. also die Schätzwerte, nach beiden Seiten um den unbekannten Wert

ϑ streuen, und zwar so, daß der Erwartungswert $\mathscr{E}T_n$ der Zufallsgröße Schätzgröße T_n gleich dem unbekannten Parameter ϑ ist. Schätzfunktionen, die diese Bedingung nicht erfüllen, weisen im Mittel einen systematischen Fehler, eine Tendenz nach einer Seite auf. Es lohnt sich daher

Definition 379.1: Eine Schätzgröße T_n für den Parameter ϑ einer Wahrscheinlichkeitsverteilung einer Zufallsgröße X heißt **erwartungstreu**, wenn $\mathscr{E}T_n = \vartheta$ ist.

Statt erwartungstreu findet man auch die Termini **unverzerrt**, **biasfrei** oder **unbias(s)ed.***

2. Der allgemeine Wunsch, daß mit wachsendem Stichprobenumfang n die Wahrscheinlichkeit dafür, daß der aus der Stichprobe gewonnene Schätzwert in der unmittelbaren Umgebung des zu schätzenden Parameters ϑ liegt, schlägt sich nieder in

Definition 379.2: Eine Folge von Schätzgrößen T_n für den Parameter ϑ heißt **konsistent**** oder **asymptotisch zutreffend** bezüglich ϑ, wenn für $\varepsilon > 0$ gilt:

$$\lim_{n \to \infty} P(|T_n - \vartheta| \geqq \varepsilon) = 0$$

3. Bekanntlich ist die Varianz einer Zufallsgröße ein Maß dafür, wie stark die Werte der Zufallsgröße um ihren Erwartungswert streuen. Man wird daher unter den erwartungstreuen Schätzgrößen diejenigen bevorzugen, die eine kleine Varianz besitzen. Mit ihnen wird man den gesuchten Parameter »genauer« treffen. *Fisher* nannte eine erwartungstreue Schätzfunktion **effizient** oder **wirksamst**, wenn es keine andere erwartungstreue Schätzfunktion gibt, deren Varianz noch kleiner ist.

In der Praxis nimmt man oft in Kauf, daß eine Schätzgröße nicht erwartungstreu ist, wenn sie dafür eine sehr kleine Varianz besitzt und ihr Erwartungswert nahe genug beim zu schätzenden Parameter liegt; dann kann nämlich der mittlere Fehler kleiner sein als der bei einer erwartungstreuen, aber sehr weit gestreuten Schätzfunktion.

4. Auf den Begriff der **Suffizienz** wollen wir nicht eingehen.

In den folgenden Abschnitten wenden wir die gewonnenen Kriterien auf Schätzgrößen für die Parameter p, μ und σ^2 an.

18.4. Die relative Häufigkeit H_n als Schätzgröße

Nach dem Maximum-Likelihood-Prinzip ist die relative Häufigkeit H_n eine brauchbare Schätzgröße für den Parameter Wahrscheinlichkeit p eines Ereignisses. Ist H_n erwartungstreu und konsistent?

* Das englische *bias* [gesprochen: baiəs] = *Neigung, Schräge, Tendenz* stammt vermutlich vom lateinischen *bifax* = *doppelblickend, schielend* ab, das aus *bis* (= *zweierlei*) und *facies* (= *Gesicht*) entstanden sein soll.

** consistere = sich hinstellen, an einer Stelle zum Stehen kommen.

1. Die Schätzgröße H_n hat als Stichprobenfunktion die Gestalt

$$H_n = H_n(X_1, X_2, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i .$$

Sie ist also ein spezielles arithmetisches Mittel, nämlich das der sämtlich nach $B(1; p)$ verteilten Zufallsgrößen X_i, die den Erwartungswert $\mathscr{E} X_i = p$ und Varianz $\operatorname{Var} X_i = pq$ besitzen, wie in **14.5.** gezeigt wurde. Damit erhalten wir nach Satz 212.1

$$\mathscr{E} H_n = p .$$

Da wir p schätzen, ist also H_n erwartungstreu bezüglich p.

2. Zum Nachweis der Konsistenz von H_n schätzen wir den zu untersuchenden Grenzwert mit Hilfe der *Bienaymé-Tschebyschow*-Ungleichung ab. Für die in dieser Ungleichung auftretende $\operatorname{Var} H_n$ gilt nach Satz 212.2: $\operatorname{Var} H_n = \frac{pq}{n}$. Damit erhalten wir

$$0 \leqq \lim_{n \to \infty} P(|H_n - p| \geqq \varepsilon) \leqq \lim_{n \to \infty} \frac{pq}{n\varepsilon^2} = 0 ,$$

was zu zeigen war.

Ohne Beweis teilen wir mit, daß H_n auch effizient ist.

18.5. Das Stichprobenmittel

Eine Zufallsgröße X besitze den Erwartungswert μ. Zu seiner Schätzung zieht man aus X eine Stichprobe $(X_1 | X_2 | \ldots | X_n)$ und bildet als Stichprobenfunktion das arithmetische Mittel $\bar{X}_n = \bar{X}_n(X_1, X_2, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$, das in diesem Zusammenhang auch »Stichprobenmittel« heißt. Da die X_i die gleiche Verteilung wie X haben, lassen sich wieder die Sätze 212.1 und 212.2 anwenden, und wir erhalten $\mathscr{E} \bar{X}_n = \mu$ und

$$0 \leqq \lim_{n \to \infty} P(|\bar{X}_n - \mu| \geqq \varepsilon) \leqq \lim_{n \to \infty} \frac{\operatorname{Var} \bar{X}_n}{\varepsilon^2} = \lim_{n \to \infty} \frac{\operatorname{Var} X}{n\varepsilon^2} = 0 .$$

Ohne Beweis teilen wir mit, daß das Stichprobenmittel auch effizient ist, und halten fest

> **Satz 380.1:** Ist $(X_1 | X_2 | \ldots | X_n)$ eine Stichprobe aus der Zufallsgröße X, dann ist das **Stichprobenmittel** $\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$ eine erwartungstreue, konsistente und effiziente Schätzgröße für den Erwartungswert μ der Verteilung von X.

Satz 380.1 beinhaltet auch die Erkenntnisse aus **18.4.**; denn die relative Häufigkeit H_n ist ein spezielles Stichprobenmittel und schätzt den Erwartungswert p der nach $B(1; p)$ verteilten Zufallsgröße X.

Auf Grund der bei den Beweisen verwendeten Sätze 212.1 und 212.2 gilt die Erwartungstreue von $\bar{X}_n$ auch für Stichproben, bei denen die X_i nicht unabhängig sind, und die Konsistenz von $\bar{X}_n$ für Stichproben, bei denen die X_i nur paarweise unabhängig sind.

18.6. Die Stichprobenvarianz

Wir wollen nun eine Schätzgröße für die Varianz σ^2 einer Zufallsgröße X ausfindig machen. Wenn möglich, soll sie erwartungstreu sein. Dazu ziehen wir aus X eine Stichprobe $(X_1 | X_2 | \ldots | X_n)$. Die X_i sind Kopien von X; also ist $\mathscr{E} X_i = \mathscr{E} X = \mu$ und $\sigma^2 = \operatorname{Var} X = \operatorname{Var} X_i = \mathscr{E}[(X_i - \mu)^2]$.
Da wir in **18.5.** unbekannte Erwartungswerte durch das Stichprobenmittel geschätzt haben, ist es naheliegend, als Schätzfunktion für σ^2 das Stichprobenmittel $V := \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$ der Zufallsgrößen $(X_i - \mu)^2$ zu verwenden. Wie in Aufgabe 384/**4** gezeigt werden soll, ist V eine erwartungstreue Schätzgröße für σ^2. Darüber hinaus ist sie konsistent und effizient.
In den meisten Fällen ist aber auch μ nicht bekannt. Wir schätzen dann μ gemäß **18.5.** durch das Stichprobenmittel $\bar{X}_n$, das wir an Stelle von μ in den Ausdruck für V einführen. Wir erhalten somit als Schätzfunktion für σ^2 die Stichprobenfunktion

$$\tilde{S}^2 := \frac{1}{n} [(X_1 - \bar{X}_n)^2 + \ldots + (X_n - \bar{X}_n)^2] = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 .$$

Zur Vereinfachung der Schreibweise läßt man üblicherweise beim Stichprobenmittel $\bar{X}_n$ den Index n weg und schreibt

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 .$$

Um zu prüfen, ob $\tilde{S}^2$ eine erwartungstreue Schätzgröße ist, formen wir die Summe um. Wir gehen dabei genauso vor wie bei der Herleitung des Verschiebungssatzes 207.1:

$$\begin{aligned}
\tilde{S}^2 &= \frac{1}{n} \sum_{i=1}^{n} [(X_i - \mu) - (\bar{X} - \mu)]^2 = \\
&= \frac{1}{n} \sum_{i=1}^{n} [(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(\bar{X} - \mu)(X_i - \mu)] = \\
&= \frac{1}{n} \left[\sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{i=1}^{n} (\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^{n} (X_i - \mu) \right] = \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 + \frac{1}{n} \cdot n(\bar{X} - \mu)^2 - \frac{2}{n} \cdot (\bar{X} - \mu)(n\bar{X} - n\mu) = \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(\bar{X} - \mu)^2 = \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - (\bar{X} - \mu)^2 .
\end{aligned}$$

Wir prüfen $\tilde{S}^2$ auf Erwartungstreue:

$$\mathscr{E}(\tilde{S}^2) = \mathscr{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-(\bar{X}-\mu)^2\right] =$$

$$= \frac{1}{n}\cdot\mathscr{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-\mathscr{E}[(\bar{X}-\mu)^2] =$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n}\mathscr{E}[(X_i-\mu)^2]-\mathscr{E}[(\bar{X}-\mu)^2] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\operatorname{Var}X_i-\mathscr{E}[(\bar{X}-\mu)^2] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sigma^2-\mathscr{E}[(\bar{X}-\mu)^2] =$$

$$= \sigma^2-\mathscr{E}[(\bar{X}-\mu)^2].$$

Da $\mu = \mathscr{E}\bar{X}$, ist der Subtrahend nichts anderes als $\operatorname{Var}\bar{X}$, wofür nach Satz 212.2 gilt $\operatorname{Var}\bar{X} = \frac{1}{n}\operatorname{Var}X = \frac{1}{n}\sigma^2$. Damit wird

$$\mathscr{E}(\tilde{S}^2) = \sigma^2-\frac{1}{n}\sigma^2 =$$

$$= \frac{n-1}{n}\sigma^2.$$

Wegen des Faktors $\frac{n-1}{n}$ ist $\tilde{S}^2$ keine erwartungstreue Schätzgröße für σ^2. Man würde σ^2 stets zu klein schätzen, und dies um so mehr, je kleiner die Länge n der Stichprobe ist. Der Mangel läßt sich aber für $n > 1$ leicht beheben: Wir multiplizieren $\tilde{S}^2$ mit $\frac{n}{n-1}$ und erhalten eine erwartungstreue Schätzgröße für σ^2, die sogenannte Stichprobenvarianz S^2.
Wir halten fest

Satz 382.1 Ist $(X_1 | X_2 | \ldots | X_n)$ eine Stichprobe aus der Zufallsgröße X mit $\bar{X} := \frac{1}{n}\cdot\sum_{i=1}^{n}X_i$, dann ist für $n \geqq 2$ die **Stichprobenvarianz**

$$S^2 := \frac{1}{n-1}\cdot\sum_{i=1}^{n}(X_i-\bar{X})^2$$

eine erwartungstreue Schätzgröße für die Varianz σ^2 der Verteilung von X.

Für große n ist der Unterschied zwischen S^2 und $\tilde{S}^2$ natürlich unerheblich. Sowohl die Stichprobenvarianz S^2 wie auch $\tilde{S}^2$ sind konsistent (vgl. Aufgabe 385/**14**), aber nicht effizient, was wir ohne Beweis mitteilen.
Es liegt nun die Vermutung nahe, man könne die **Stichprobenstreuung** $S := \sqrt{S^2}$

als erwartungstreue Schätzgröße für die Standardabweichung σ benützen. *Dies ist leider falsch,* wie wir an einem Beispiel zeigen:
Es sei $X :=$ »Zahl der Adler beim Wurf einer evtl. unsymmetrischen Münze«. X ist nach $\mathrm{B}(1; p)$ verteilt. $(X_1 \mid X_2)$ sei eine Stichprobe der Länge 2 aus X. Mit $\bar{X} = \frac{1}{2}(X_1 + X_2)$ erhält man

$$S^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 = \tfrac{1}{2}(X_1 - X_2)^2,$$
$$S \;= \tfrac{1}{2}\sqrt{2}\,|X_1 - X_2|.$$

S nimmt offensichtlich nur die Werte 0 und 1 an. Damit errechnet man

$$\mathscr{E}S = \tfrac{1}{2}\sqrt{2}\,[P(X_1 = 1 \wedge X_2 = 0) + P(X_1 = 0 \wedge X_2 = 1)] =$$
$$= p(1-p)\sqrt{2}.$$

Dagegen ist $\sigma(X) = \sqrt{p(1-p)}$.
Die Verschiedenheit von $\mathscr{E}S$ und σ läßt sich nicht so wie die von $\mathscr{E}\tilde{S}^2$ und σ^2 durch einen für alle p gültigen Faktor beseitigen.

Die Berechnung des Wertes von S^2. Bei mehrfacher Messung einer (z. B. physikalischen) Größe unter gleichen Bedingungen ist es üblich, aus der Meßreihe die Werte $\bar{x}$ und s von $\bar{X}$ bzw. S zu bestimmen und das Meßresultat in der Form

$\bar{x} \pm s$ zu schreiben.

So findet man z. B. für das Wirkungsquantum h die Angabe

$$\frac{h}{2\pi} = (1{,}05443 \pm 0{,}00004) \cdot 10^{-34}\,\mathrm{Js}.$$

s ist meist sehr mühsam zu berechnen, wenn man die in Satz 382.1 enthaltene Definition von S^2 direkt benützt, da $\bar{x}$ im allgemeinen keine glatte Zahl ist. Man hat viele unbequeme Subtraktionen auszuführen. Die folgende Umformung erleichtert die Arbeit. Sie entspricht genau der oben vorgeführten Umformung von $\tilde{S}^2$.

$$(n-1)\cdot s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 =$$
$$= \sum_{i=1}^{n} (x_i - a)^2 + 2\sum_{i=1}^{n} (x_i - a)(a - \bar{x}) + n(a - \bar{x})^2.$$

Der zweite Term ist gleich

$$2(a - \bar{x}) \sum_{i=1}^{n} (x_i - a) = 2(a - \bar{x})(n\bar{x} - na) = -2n(a - \bar{x})^2.$$

Also erhält man schließlich mit $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\boxed{s^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} (x_i - a)^2 - n(\bar{x} - a)^2\right)}$$

Diese Formel gilt für eine beliebige Zahl a. Man wählt a als glatte Zahl in der Nähe von $\bar{x}$ und kann damit die rechte Seite verhältnismäßig einfach ausrechnen. Ein Beispiel zeigt Tabelle 384.1.

Tab. 384.1 Punktbewertungen x_i von 10 Personen bei einem Gedächtnistest. Die Hilfszahl $a = 35$ wird bereits zur Berechnung von $\bar{x}$ mit Vorteil verwendet. (Vgl. Aufgabe 215/**20**)

x_i	$x_i - a$	$(x_i - a)^2$
12	−23	529
21	−14	196
28	− 7	49
30	− 5	25
34	− 1	1
37	+ 2	4
39	+ 4	16
39	+ 4	16
49	+14	196
56	+21	441
Summe:	− 5	1473

$\bar{x} = 35 - \frac{5}{10} = 34{,}5;$
$s^2 = \frac{1}{9}(1473 - 10 \cdot 0{,}5^2);$
$s = 12{,}8.$

Aufgaben

Zu 18.3.–18.6.

1. Bestimme bei folgenden Stichproben aus einer Zufallsgröße Schätzwerte für den Erwartungswert und die Varianz:
a) 3; 5; 3; 6; 9. **b)** 0,3; 0,7; −0,4; 0,8; −0,2.
c) 300; 700; −400; 800; −200 **d)** 1; 1; 1; 1; 1.

2. Bestimme Schätzwerte für Erwartungswert und Varianz der Zufallsgröße »Bremsweg bei einer Geschwindigkeit von 50 km/h«, wenn sich folgende Meßwerte in m ergaben:
15,5 14,0 14,1 14,9 13,4 15,0 14,4 14,4 15,8 15,9.

3. Welchen Erwartungswert und welche Varianz erhält man bei n Messungen für das Stichprobenmittel $\bar{X}$, wenn für die Zufallsgröße X gilt:
a) $\mathscr{E}X = 2{,}71$; $\operatorname{Var}X = 1{,}5$; $n = 10$. **b)** $\mathscr{E}X = 1$; $\operatorname{Var}X = 1$; $n = 100$.
c) $\mathscr{E}X = 1$; $\operatorname{Var}X = 1$; $n = 1000$. **d)** $\mathscr{E}X = 0$; $\operatorname{Var}X = 1$; $n = 1000$.

4. Zeige, daß $V := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ eine erwartungstreue Schätzfunktion für σ_X^2 ist.

5. **a)** Eine Zufallsgröße X ist binomial nach $B(1; 0{,}25)$ verteilt. Bestimme Erwartungswert, Varianz und Standardabweichung von X. Berechne für eine Stichprobe der Länge 3 die Wahrscheinlichkeitsverteilungen und die Erwartungswerte des Stichprobenmittels $\bar{X}_3$, der Stichprobenvarianz S^2 und der Stichprobenstreuung S.
b) Löse **a)** für eine Stichprobe der Länge 2, wenn X nach $B(2; 0{,}25)$ verteilt ist.

6. Eine Zufallsgröße X besitze die Verteilung $P(X = x_j) =: p_j$ für $j = 1, \ldots, s$. $(X_1|X_2|\ldots|X_n)$ sei eine Stichprobe der Länge n, ferner sei $N_j :=$ Anzahl der X_i, welche den Wert x_j annehmen. A_j bedeute das Ereignis »$\left|\frac{1}{n}N_j - p_j\right| \geqq a$«.
a) Formuliere für $P(A_j)$ die *Tschebyschow*-Ungleichung.
b) Formuliere mit den A_j das Ereignis $A :=$ »Die Stichprobenverteilung weicht nirgends um a oder mehr von der Wahrscheinlichkeitsverteilung ab«.
•**c)** Finde für $P(\bar{A})$ eine Abschätzung, die zeigt, daß das Ereignis A für genügend große n eine beliebig nahe an 1 gelegene Wahrscheinlichkeit hat. *Die Stichprobenverteilung konvergiert »in Wahrscheinlichkeit« gegen die Wahrscheinlichkeitsverteilung.*

7. Eine Zufallsgröße sei nach $B(100; 0{,}4)$ verteilt. Wie lang muß eine Stichprobe mindestens sein, damit die Varianz des Stichprobenmittels kleiner als 0,01 ist?

8. Die Wahrscheinlichkeit für eine Knabengeburt ist 0,514. Die Zufallsgröße X nehme den Wert 1 an, wenn ein Knabe geboren wird, sonst habe sie den Wert 0.

Berechne Erwartungswert und Standardabweichung des Stichprobenmittels für eine Stichprobe der Länge 100.

9. a) Man wirft eine Münze 1mal und nimmt $Z :=$ »Anzahl der Adler« als Schätzgröße für den Parameter $p := P$(»Adler«). Welche Werte kommen bei der Schätzung also stets heraus? Z scheint keine sehr gute Schätzgröße zu sein. Man zeige, daß sie aber erwartungstreu ist!

b) Untersuche, ob $T := tX_1 + (1-t)X_2$ eine erwartungstreue Schätzfunktion für $\mathscr{E}X$ ist, wenn $(X_1 | X_2)$ eine Stichprobe der Länge 2 aus der Zufallsgröße X ist.

10. Eine Schätzgröße Y sei nicht erwartungstreu bezüglich des Parameters ϑ. Warum kann man nicht einfach $Y_1 := Y - \mathscr{E}Y + \vartheta$ als neue, und zwar erwartungstreue Schätzgröße benützen, oder auch $Y_2 := \dfrac{Y \cdot \vartheta}{\mathscr{E}Y}$?

11. Der »Idealwert« np für die Anzahl des Eintretens eines Ereignisses der Wahrscheinlichkeit p bei n unabhängigen Versuchen wird durch die Anzahl Z des wirklichen Eintretens geschätzt. Ist diese Schätzung erwartungstreu?

•12. Die Kugeln in einer Urne sind von 1 an fortlaufend numeriert. Es werden n Kugeln *ohne* Zurücklegen gezogen und ihre Nummern aufgeschrieben. Es soll die unbekannte Anzahl τ aller Kugeln in der Urne geschätzt werden.*

a) Als Schätzgröße bietet sich $G :=$ »Größte gezogene Kugelnummer« an. Es gilt nach Aufgabe 192/**31** $\mathscr{E}G = \dfrac{n(\tau+1)}{n+1}$.
G ist also nicht erwartungstreu. Konstruiere aus G eine erwartungstreue Schätzgröße $\tilde{G}$.

b) Es gilt: $\operatorname{Var} G = \dfrac{n(\tau+1)(\tau-n)}{(n+1)^2(n+2)}$.
Berechne $\operatorname{Var}\tilde{G}$.

c) Eine andere plausible Schätzgröße für τ ist das doppelte arithmetische Mittel $2\bar{X}$ aus allen gezogenen Kugelnummern. Zeige durch Berechnung des Erwartungswertes, daß auch diese Schätzgröße erst nach einer kleinen Abänderung erwartungstreu ist.

d) Eine längere Rechnung (Durchführung nicht verlangt) ergibt:

$$\operatorname{Var}\bar{X} = \frac{(\tau+1)(\tau-n)}{12n}.$$

Berechne hieraus die Varianz der erwartungstreuen Schätzgröße aus **c)** und vergleiche sie mit $\operatorname{Var}\tilde{G}$ aus **b)**. Sind die beiden Schätzgrößen für τ gleichwertig?

•13. Aus der Urne der vorigen Aufgabe wird die Stichprobe *mit* Zurücklegen gezogen. Zeige, daß $\operatorname{Var}\bar{X} = \dfrac{\tau^2-1}{12n}$ gilt und daß die mit $\bar{X}$ gebildete erwartungstreue Schätzgröße stärker streut als die von Aufgabe **12. c)**.

•14. Zeige, daß S^2 und $\tilde{S}^2$ konsistente Schätzgrößen für σ^2 sind.
Hinweis: Zerlege die in der *Tschebyschow*-Ungleichung auftretende $\operatorname{Var} S^2$ in zwei Teile, so daß in einem dieser Teile nur über lauter verschiedene Indizes summiert wird.

* Dieses Schätzproblem spielte im 2. Weltkrieg eine Rolle, als man aus den Seriennummern von erbeuteten Waffen auf den Umfang der Waffenproduktion schließen wollte.