Stochastik

Barth, Friedrich

München, [20]03

17. 3. Test bei zwei einfachen Hypothesen

gen, Urnen usw.) verschwindend gering, so daß es meist genügt, die einfachere Variante zu untersuchen.
Für den Statistiker ist die Frage wichtig, ob eine Stichprobe **repräsentativ** ist, d.h., ob sie eine genügend genaue Auskunft über die »Urne« geben kann, aus der sie stammt. Die Gewinnung einer repräsentativen Stichprobe gehört mit zu den schwierigsten Aufgaben der Beschreibenden Statistik. In der Markt- und Meinungsforschung nimmt man meist Stichproben der Größenordnung 2000.
Nach der Interpretationsregel für Wahrscheinlichkeiten und dem Gesetz der großen Zahlen kann man vermuten, daß genügend lange Stichproben im Sinne der Definition 335.1 auch repräsentativ sind. Eine genauere Auskunft hierüber gibt der 1933 von *Waleri Iwanowitsch Gliwenko* (1897–1940)* bewiesene

Hauptsatz der Mathematischen Statistik: Die mit Hilfe von Stichproben der Länge n gewonnenen empirischen Verteilungsfunktionen einer Zufallsgröße konvergieren mit Wahrscheinlichkeit 1 gleichmäßig gegen die wahre Verteilungsfunktion dieser Zufallsgröße, falls der Stichprobenumfang n gegen Unendlich strebt.

Dieser interessante Satz besagt also, daß die Aussagekraft einer Zufallsstichprobe von ihrer absoluten Länge abhängt und daß die Mächtigkeit der Grundgesamtheit, aus der sie gezogen wird, erstaunlicherweise keine Rolle spielt.

17.3. Test bei zwei einfachen Hypothesen

Das älteste Entscheidungsproblem der Wahrscheinlichkeitsrechnung ist *Blaise Pascal*s (1623–1662) *Infini-rien* (siehe Seite 343). Als ersten Test kann man *John Arbuthnot*s (1667–1735)** mathematischen Gottesbeweis *An Argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes* aus dem Jahre 1710 auffassen. (Siehe Aufgabe 369/**32.**) Auch die Versuche von *Daniel Bernoulli* (1700–1782) aus dem Jahre 1770, das wahre Geburtsverhältnis der Geschlechter zu finden, kann man als Test im heutigen Sinne deuten. Die eigentliche Testtheorie entwickelten jedoch erst in der ersten Hälfte des 20. Jahrhunderts *Jerzy Neyman* (1894–1981) und *Egon Sharpe Pearson* (1895–1980). Beginnend mit einem einfachen Beispiel wollen wir ihre Gedanken nachvollziehen.

Beispiel 1: An eine Werkstatt werden Schachteln mit Schrauben geliefert. Ein Teil davon enthält Erste Qualität, das sind Schrauben, von denen nur 15% die vorgeschriebenen Maßtoleranzen nicht einhalten. Die restlichen Schachteln enthalten Zweite Qualität, mit einem Ausschußanteil von 40%. Die Lieferfirma hat vergessen, die Schachteln nach ihrem Inhalt zu kennzeichnen. Für die Verarbeitung ist es aber wichtig, die Qualität der Schrauben zu kennen. Man braucht also ein **Entscheidungsverfahren**, mit dessen Hilfe man die Schachteln der jeweiligen Qualität zuordnet. Über die Qualität der Schrauben gibt es nur 2 Vermutungen, **Hypothesen***** genannt, nämlich

* Гливенко
** *John Arbuthnot* war Leibarzt der Königin *Anna*. Als politischer Satiriker schuf er die Figur des John Bull (1712).
*** ὑπόθεσις = das Untergelegte; die Annahme.

H_0: Der Anteil der defekten Schrauben in der Schachtel beträgt 0,15.
H_1: Der Anteil der defekten Schrauben in der Schachtel beträgt 0,4.

Bezeichnen wir den Anteil der defekten Schrauben in der Schachtel mit p, so lassen sich die beiden Hypothesen kurz wie folgt schreiben:

H_0: $p = 0{,}15$ bzw. H_1: $p = 0{,}4$.

Die beiden Hypothesen schließen einander aus; man nennt sie daher auch **Alternativen***, und da sie jeweils durch genau einen Wert für p beschrieben werden, nennt man sie **einfach**. Das Verfahren, das zur Entscheidung zwischen ihnen führt, heißt **Test****, hier genauer **Alternativtest.**
Jeder Test besteht zunächst in der Festlegung eines Zufallsexperiments. Man entschließt sich z. B., aus jeder Schachtel $n = 10$ Schrauben – zur Vereinfachung unserer Rechnung – mit Zurücklegen zu entnehmen und diese genau zu messen, d. h., man entnimmt jeder Schachtel eine Zufallsstichprobe der Länge 10. Die Zufallsgröße X_i ist die Qualität der i-ten entnommenen Schraube. Ein mögliches Stichprobenergebnis hat das Aussehen $(0|1|1|1|0|1|0|0|1|1)$, wobei 0 für »defekt« und 1 für »gut« stehen. Je nach der Anzahl Z der erhaltenen defekten Schrauben entscheidet man sich dann für eine der beiden Hypothesen. Diese Anzahl Z hängt natürlich vom Stichprobenergebnis ab; sie ist also eine Funktion $Z(X_1, X_2, \ldots, X_n)$ der Zufallsstichprobe $(X_1|X_2|\ldots|X_n)$, eine sog. **Stichprobenfunktion**. Als Funktion von Zufallsgrößen ist Z selbst wieder eine Zufallsgröße. Da von ihrem Ausfall die Entscheidung zwischen den Hypothesen abhängt, heißt Z **Prüffunktion** oder **Testgröße**.

Je nachdem, aus welcher Schachtel die Zufallsstichprobe entnommen wird, ergibt sich für die Zufallsgröße Z eine andere Wahrscheinlichkeitsverteilung, und zwar entweder $B(10; 0{,}15)$ oder $B(10; 0{,}4)$. Die Hypothesen H_0 und H_1 lassen sich daher als Hypothesen über die Wahrscheinlichkeitsverteilung der Testgröße Z formulieren:

H_0: »Z ist nach $B(10; 0{,}15)$ verteilt« bzw.
H_1: »Z ist nach $B(10; 0{,}4)$ verteilt«.

Die Wertemenge $\{0, 1, 2, \ldots, 10\}$ von Z nehmen wir als Ergebnisraum Ω unseres Zufallsexperiments. Für kleine Werte von Z wird man sich dann vernünftigerweise für H_0 entscheiden. Es fragt sich nur, bis zu welcher Grenze k diese Entscheidung für H_0 getroffen werden soll. Die Wahl dieser Grenze k, des sog. **kritischen Werts**, ist völlig willkürlich; sie muß jedoch *vor* der Ausführung des Zufallsexperiments erfolgen. Offensichtlich beeinflußt sie die Qualität des Urteils. Die Festlegung eines bestimmten kritischen Werts k führt zur **Entscheidungsregel**

$$\delta_k\colon \begin{cases} Z \leqq k \;\Rightarrow\; \text{Entscheidung für } H_0 \\ Z > k \;\Rightarrow\; \text{Entscheidung für } H_1 \end{cases}$$

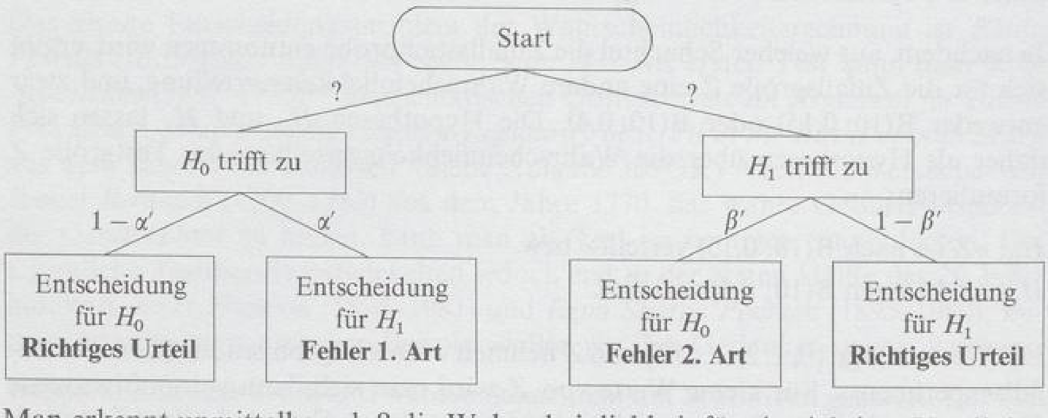
* alter (lat.) = der eine von zweien, der andere.
** Zur Herkunft des Wortes: lat. *testum*: Schüssel; altfranz. *test*: Tiegel für alchimistische Versuche; engl. *test*: Versuch, Prüfung.

Das Ereignis $A := »Z \leqq k« = \{0, 1, \ldots, k\}$ heißt **Annahmebereich** für die Hypothese H_0.
Entsprechend heißt das Ereignis $\bar{A} := »Z > k« = \{k+1, \ldots, n\}$ Annahmebereich für die Hypothese H_1.
Da man sich bei einem Alternativtest für eine der beiden Hypothesen entscheiden muß, muß der Annahmebereich für H_1 natürlich das Gegenereignis des Annahmebereichs A für H_0 sein.
Der Ausfall der Stichprobe ist zufallsbestimmt, also wird auch unser Urteil vom Zufall diktiert. Und wie bei jeder Entscheidung im Leben hat man auch hier die Möglichkeit, auf 2 Arten einen Fehler zu begehen.

Fehler 1. Art: Die Hypothese H_0 trifft tatsächlich zu, und $\bar{A}$ tritt ein. Wir entscheiden uns auf Grund der Entscheidungsregel δ_k aber für H_1 und begehen damit einen Fehler, den man üblicherweise »Fehler 1. Art« nennt. Die Wahrscheinlichkeit, einen solchen Fehler zu begehen, bezeichnen wir mit α'. Sie heißt auch **Irrtumswahrscheinlichkeit 1. Art.**

Fehler 2. Art: Die Hypothese H_1 trifft tatsächlich zu, und A tritt ein. Wir entscheiden uns auf Grund der Entscheidungsregel δ_k aber für H_0 und begehen damit einen Fehler, den man üblicherweise »Fehler 2. Art« nennt. Die Wahrscheinlichkeit, einen solchen Fehler zu begehen, bezeichnen wir mit β'. Sie heißt auch **Irrtumswahrscheinlichkeit 2. Art.***

Mittels eines Baumes läßt sich die Situation veranschaulichen:



Man erkennt unmittelbar, daß die Wahrscheinlichkeit für ein richtiges Urteil davon abhängt, welche der beiden Hypothesen in Wirklichkeit zutrifft. Dementsprechend heißen $1-\alpha'$ bzw. $1-\beta'$ **statistische Sicherheit des Urteils** bei Vorliegen von H_0 bzw. H_1.
Um andererseits die Qualität des Tests beurteilen zu können, muß man die Irrtumswahrscheinlichkeiten, d.h. die Wahrscheinlichkeiten für den Fehler 1. bzw. 2. Art,

* Die Bezeichnungen *Fehler 1. Art* und *Fehler 2. Art* suggerieren leider einen Qualitätsunterschied und können dadurch leicht falsche Vorstellungen hervorrufen. In Wahrheit sind die beiden Fehler von gleicher Art! *J. Neyman* und *E. S. Pearson* sprachen 1928 davon, welche Entscheidungsregel man auch aufstelle, »*two sources of error must arise*« – zwei Fehlerquellen müssen entstehen. Sie numerieren sie mit (1) und (2) und sprechen z. B. vom »*error of form* (1)«. In ihrer Arbeit aus dem Jahre 1932 rekapitulieren sie diese beiden »*sources of error*« und sprechen später von »*errors of the first kind referred to above*«, woraus dann »Fehler 1. Art« wurde.

berechnen. Durch jede der beiden Hypothesen wird, wie oben besprochen, auf dem Ergebnisraum Ω der Zufallsgröße Z eine Wahrscheinlichkeitsverteilung für Z festgelegt, die wir gelegentlich mit P_{H_0} bzw. P_{H_1} bezeichnen wollen. Handelt es sich bei einer solchen Wahrscheinlichkeitsverteilung um eine Binomialverteilung $B(n;p)$, dann schreibt man statt P_H gerne P_p^n. Da wir als Zufallsexperiment das Ziehen von 10 Kugeln mit Zurücklegen gewählt haben, gilt $P_{H_0} = B(10; 0{,}15) = P_{0,15}^{10}$ und $P_{H_1} = B(10; 0{,}4) = P_{0,4}^{10}$. Damit ergibt sich

$$\alpha' = P_{H_0}(\bar{A}) = P_{0,15}^{10}(\bar{A}) = \sum_{i=k+1}^{10} \binom{10}{i} 0{,}15^i \cdot 0{,}85^{10-i} = 1 - F_{0,15}^{10}(k) \text{ und}$$

$$\beta' = P_{H_1}(A) = P_{0,4}^{10}(A) = \sum_{i=0}^{k} \binom{10}{i} 0{,}4^i \cdot 0{,}6^{10-i} = F_{0,4}^{10}(k).$$

Figur 339.1 zeigt die Verhältnisse für den kritischen Wert $k = 3$. In diesem Fall erhält man

$$\alpha' = P_{0,15}^{10}(Z > 3) = 1 - F_{0,15}^{10}(3) =$$
$$= 0{,}04997 \approx 5\%$$

und

$$\beta' = P_{0,4}^{10}(Z \leq 3) = F_{0,4}^{10}(3)$$
$$= 0{,}38228 \approx 38{,}2\%.$$

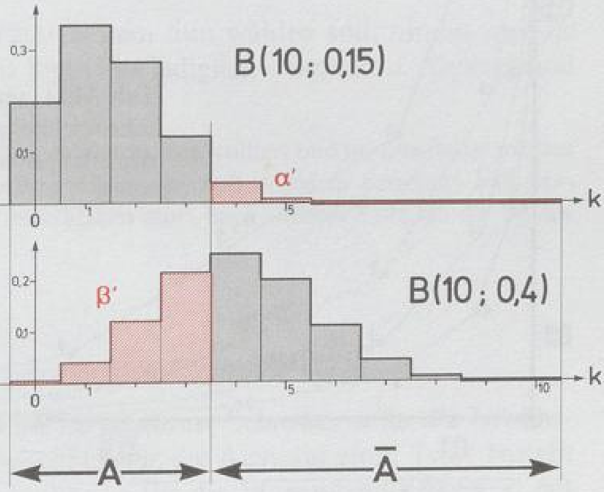
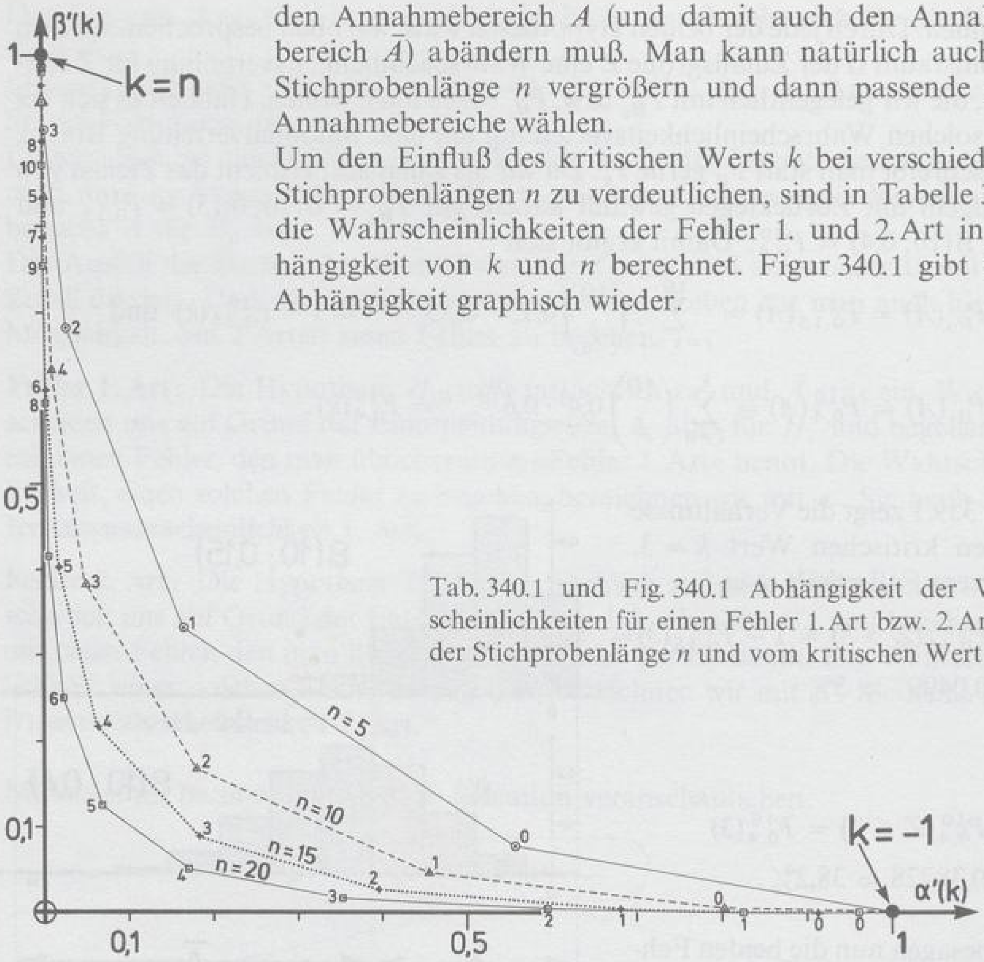


Fig. 339.1 Wahrscheinlichkeitsverteilung zur Hypothese 1 (Erste Qualität, oben) und Hypothese 2 (Zweite Qualität, unten) von Beispiel 1. grau: Wahrscheinlichkeit für ein richtiges Urteil. rot: Wahrscheinlichkeit für einen Fehler 1. Art bzw. 2. Art.

Was besagen nun die beiden Fehlerwahrscheinlichkeiten α' und β' für die Praxis? Hätte man sehr viele Schachteln mit Schrauben nach dem gegebenen Entscheidungsverfahren zu beurteilen, so würde man in etwa 95% der Fälle, in denen in Wirklichkeit Erste Qualität vorliegt ($p = 0{,}15$), dies aus der Stichprobe richtig erkennen und nur in etwa 5% der Fälle diese Schrauben irrtümlich für Zweite Qualität halten (Fehler 1. Art). Der andere mögliche Irrtum, nämlich Schachteln mit Schrauben Zweiter Qualität für besser zu halten, als sie in Wirklichkeit sind, wird aber in etwa 38% der Fälle vorkommen, in denen Schachteln mit Schrauben Zweiter Qualität untersucht werden (Fehler 2. Art). Unserem Test entspricht also eine recht optimistische Beurteilung der Ware. Es kann sein, daß dies erwünscht ist – daß man vor allem daran interessiert ist, die Erste Qualität nicht irrtümlich für Zweite zu halten. Dann ist der Test brauchbar. Andernfalls muß er geändert werden. Dies geschieht dadurch, daß man eine neue Entscheidungsregel δ_k festlegt. Will man die Stichprobenlänge n unverändert lassen, so heißt dies, daß man den Annahmebereich A (und damit auch den Annahmebereich $\bar{A}$) abändern muß. Man kann natürlich auch die Stichprobenlänge n vergrößern und dann passende neue Annahmebereiche wählen.

Um den Einfluß des kritischen Werts k bei verschiedenen Stichprobenlängen n zu verdeutlichen, sind in Tabelle 340.1 die Wahrscheinlichkeiten der Fehler 1. und 2. Art in Abhängigkeit von k und n berechnet. Figur 340.1 gibt diese Abhängigkeit graphisch wieder.



Tab. 340.1 und Fig. 340.1 Abhängigkeit der Wahrscheinlichkeiten für einen Fehler 1. Art bzw. 2. Art von der Stichprobenlänge n und vom kritischen Wert k

n	k	$\alpha'(k)$	$\beta'(k)$
5	−1	1	0
	0	0,55629	0,07776
	1	0,16479	0,33696
	2	0,02661	0,68256
	3	0,00223	0,91296
	4	0,00008	0,98976
	5	0	1
10	−1	1	0
	0	0,80313	0,00605
	1	0,45570	0,04636
	2	0,17980	0,16729
	3	0,04997	0,38228
	4	0,00987	0,63310
	5	0,00138	0,83376
	6	0,00013	0,94524
	7	0,00001	0,98771
	8	0,00000	0,99832
	9	0,00000	0,99990
	10	0	1
15	−1	1	0
	0	0,91265	0,00047
	1	0,68141	0,00517
	2	0,39577	0,02711
	3	0,17734	0,09050
	4	0,06171	0,21728
	5	0,01681	0,40322
	6	0,00361	0,60981
	7	0,00061	0,78690
	8	0,00008	0,90495
	9	0,00001	0,96617
	10	0,00000	0,99065
	11	0,00000	0,99807
	12	0,00000	0,99972
	13	0,00000	0,99997
	14	0,00000	1,00000
	15	0	1
20	−1	1	0
	0	0,96124	0,00004
	1	0,82444	0,00052
	2	0,59510	0,00361
	3	0,35227	0,01596
	4	0,17015	0,05095
	5	0,06731	0,12560
	6	0,02194	0,25001
	7	0,00592	0,41589
	8	0,00133	0,59560
	9	0,00025	0,75534
	10	0,00004	0,87248
	11	0,00000	0,94347
	12	0,00000	0,97897
	13	0,00000	0,99353
	14	0,00000	0,99839
	15	0,00000	0,99968
	16	0,00000	0,99995
	17	0,00000	0,99999
	18	0,00000	1,00000
	19	0,00000	1,00000
	20	0	1

Will man also die Stichprobenlänge $n = 10$ beibehalten, so wird man zur Verkleinerung von β' als kritischen Wert $k = 2$ wählen. Dann ist

$$\alpha' = P^{10}_{0,15}(Z > 2) = 0{,}17980 \qquad \text{und} \qquad \beta' = P^{10}_{0,4}(Z \leqq 2) = 0{,}16729\,.$$

Die Gefahr, zu viele schlechte Schachteln für gut zu halten, ist gebannt (Wahrscheinlichkeit für den Fehler 2. Art $\approx 17\%$); dafür werden aber nun ca. 18% aller guten Schachteln für schlecht gehalten. Ist man auch mit diesem Resultat nicht zufrieden, so bleibt nur noch der Ausweg, die Stichprobe zu vergrößern. Wenn Zeit und Kosten für die Prüfung der Stücke keine große Rolle spielen, wird man das von vornherein tun. Bei einer Stichprobenlänge von $n = 20$ und einem kritischen Wert $k = 5$ z.B. ergäbe sich dann

$$\alpha' = P^{20}_{0,15}(Z > 5) = 0{,}06731 \approx 6{,}7\% \qquad \text{und}$$
$$\beta' = P^{20}_{0,4}(Z \leqq 5) = 0{,}12560 \approx 12{,}6\%\,.$$

Die Entscheidung, welches Testverfahren man nun wählen soll, nimmt uns die mathematische Theorie nicht ab. Sie kann uns lediglich – um mit *J. Neyman* und *E.S. Pearson* zu sprechen –

»zeigen, wie die Risiken, die durch die Fehler entstehen, kontrolliert und minimalisiert werden können. Die Anwendung dieses statistischen Rüstzeugs muß in jedem einzelnen Fall dem Untersuchenden überlassen bleiben, der entscheiden muß, nach welcher Seite hin die Waage ausschlagen soll.«*

Damit erhebt sich die Frage:

Wie konstruiert man einen Test mit gewünschten Eigenschaften?

Beispiel 2: Konstruktion eines Tests bei vorgegebener Schranke α für die Irrtumswahrscheinlichkeit 1. Art. Eine Möglichkeit für die Auswahl eines Tests besteht darin, daß man sich eine obere Schranke α für die Wahrscheinlichkeit α' des Fehlers 1. Art vorgibt. Nehmen wir z.B. $\alpha = 1\%$, dann müssen wir die Bedingung $\alpha' \leqq 0{,}01$ erfüllen. In der Situation von Beispiel 1 bedeutet dies, daß wir $1 - F^{10}_{0,15}(k) \leqq 0{,}01$ erfüllen müssen. Aus Tabelle 340.1 entnehmen wir, daß diese Bedingung für $k \geqq 4$ erfüllt ist. Je größer wir den kritischen Wert k wählen, desto kleiner wird die Wahrscheinlichkeit α' für den Fehler 1. Art. Wir können sie sogar auf Null drücken, wenn wir $k = 10$ wählen. In diesem Extremfall entscheidet man sich unabhängig vom Ergebnis der Stichprobe immer für H_0, also im obigen Beispiel 1 für 1. Qualität. Es handelt sich aber dann eigentlich nicht mehr um einen Test; denn der Zufall spielt keine Rolle mehr.
Bei der Verkleinerung von α' muß man jedoch bedenken, daß dabei unvermeidlicherweise die Wahrscheinlichkeit β' für einen Fehler 2. Art wächst, wie man sich leicht an Figur 339.1 klarmacht.
Nach *Jerzy Neyman* und *Egon Sharpe Pearson*, den Vätern der Testtheorie, wählt man daher bei vorgegebener oberer Schranke α für die Irrtumswahrscheinlichkeit 1. Art α' denjenigen Wert k als **besten kritischen Wert**, für den die Irrtumswahrscheinlichkeit 2. Art β' minimal wird.

* *On the problem of the most efficient tests of statistical hypotheses* (1932) in Philosophical Transactions of the Royal Society of London, **A 231** (1933).

Stellt man also z. B. die Bedingung $\alpha' \leqq 0{,}01$, dann wird man als besten kritischen Wert die kleinstmögliche Zahl k, also $k = 4$ wählen. Damit ergibt sich

$$\alpha' = 1 - F_{0,15}^{10}(4) = 0{,}00987 \approx 1{,}0\% \quad \text{und} \quad \beta' = F_{0,4}^{10}(4) = 0{,}63310 \approx 63{,}3\%.$$

Bei diesem Test werden Schachteln Erster Qualität mit einer Sicherheit von 99% erkannt. Dagegen werden Schachteln Zweiter Qualität mit einer Wahrscheinlichkeit von 63,3% für Schachteln Erster Qualität gehalten. Dieser Wert ist erschreckend hoch. Behält man die Schranke α für α' bei, dann läßt sich β' nur verkleinern, wenn man die Länge der Stichprobe n vergrößert.
Wählt man z. B. $n = 50$, dann erhält man aus $\alpha' \leqq 0{,}01$ die Bedingung $k \geqq 14$. Nimmt man nun 14 als kritischen Wert, dann ist

$\alpha' = 1 - F_{0,15}^{50}(14) = 0{,}00529 \approx 0{,}5\%$ und
$\beta' = F_{0,4}^{50}(14) = 0{,}05396 \approx 5{,}4\%$. (Vgl. Figur 342.1.)

Die Gefahr, schlechte Schachteln für gute zu halten, ist jetzt weitgehend gebannt. Man muß allerdings bedenken, daß die Prüfung von 50 Schrauben mehr Zeit und damit auch mehr Geld kostet als die Prüfung von 10 Stück.

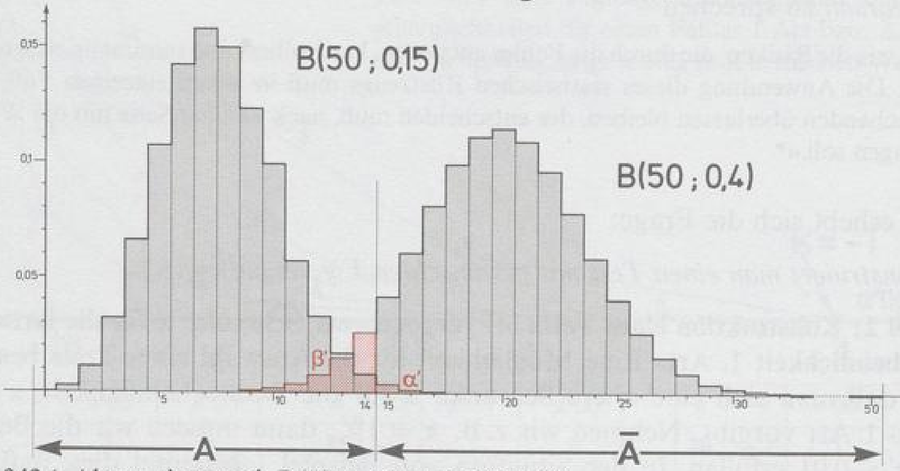
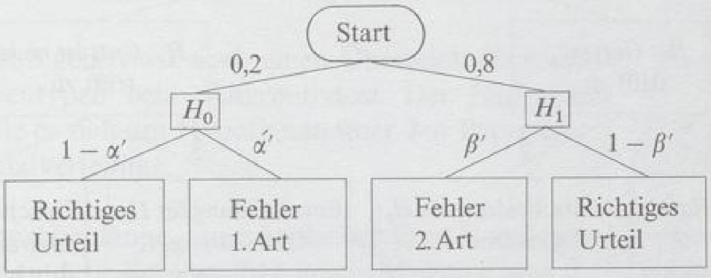


Fig. 342.1 Alternativtest mit B(50; 0,15) und B(50; 0,4)

In der Praxis spielen die finanziellen Folgen eines Tests natürlich eine beherrschende Rolle. Abgesehen von den Prüfkosten verursacht nämlich *jeder* Fehler Unkosten.

Beispiel 3: Test zur Minimierung des Schadens. Nehmen wir an, ein Fehler 1. Art verursacht beim Test des Beispiels 1 (Seite 336) einen Schaden von 3 DM pro Schachtel (weil man gute Schrauben verwendet, wo es auch weniger gute getan hätten), während ein Fehler 2. Art einen Schaden von 5 DM pro Schachtel erzeugt (weil die Verwendung dieser Schrauben mehr Reparaturen bedingt). Nimmt man zusätzlich an, daß unter 100 gelieferten Schachteln 80 von Zweiter Qualität und 20 von Erster Qualität waren, dann ist die Entscheidungsregel δ_k des Tests natürlich so zu wählen, daß der zu erwartende Schaden minimal wird. Betrachten wir dazu die Zufallsgröße $S :=$ Schaden pro Schachtel. Für ihre Wahrscheinlichkeitsverteilung gilt, wie man dem nachstehenden Baumdiagramm leicht entnimmt:

s	3	5	0
$W(s)$	$0{,}2\,\alpha'$	$0{,}8\,\beta'$	$1-(0{,}2\,\alpha'+0{,}8\,\beta')$



Für den erwartenden Schaden erhält man also, gemessen in DM,

$$\mathscr{E}S = 3\cdot 0{,}2\alpha' + 5\cdot 0{,}8\,\beta' = \\ = 0{,}6\alpha' + 4\beta'\,.$$

In Tabelle 343.1 stellen wir die Abhängigkeit des zu erwartenden Schadens $\mathscr{E}S$ bei der Stichprobenlänge 10 in Abhängigkeit vom kritischen Wert k dar.
Im vorliegenden Fall wählt man also δ_1, d.h., man hält die betreffende Schachtel für 1. Qualität, wenn unter den 10 mit Zurücklegen entnommenen Schrauben höchstens 1 Schraube defekt war.
Da man die a-priori-Wahrscheinlichkeiten für das Vorliegen von H_0 und H_1 kennt, kann man in diesem Fall auch die Wahrscheinlichkeit für eine Fehlentscheidung angeben; sie beträgt $0{,}2\,\alpha' + 0{,}8\,\beta'$.

k	$\alpha'(k)$	$\beta'(k)$	$\mathscr{E}S$
−1	1	0	0,60
0	0,80313	0,00605	0,51
1	0,45570	0,04636	0,46
2	0,17980	0,16729	0,78
3	0,04997	0,38228	1,56
4	0,00987	0,63310	2,54
5	0,00138	0,83376	3,34
6	0,00013	0,94524	3,78
7	0,00001	0,98771	3,95
8	0,00000	0,99832	3,99
9	0,00000	0,99990	4,00
10	0	1	4,00

Tab. 343.1 Der zu erwartende Schaden $\mathscr{E}S$ beim Test der Hypothese »$p = 0{,}15$« gegen die Alternative »$p = 0{,}4$« bei der Stichprobenlänge $n = 10$ in Abhängigkeit von k

Das Verfahren, die Entscheidung zwischen 2 Alternativen durch Minimierung des Schadens bzw. Maximierung des Gewinns herbeizuführen, stammt gewissermaßen aus der Geburtsstunde der Stochastik. *Blaise Pascal* (1623–1662) zeigt nämlich damit im Artikel *Infini–rien* – Das Unendliche–Das Nichts – aus den *Pensées*, niedergeschrieben vermutlich 1657, daß es sinnvoll ist, sich für die Existenz des christlichen Gottes mit all seinen Konsequenzen zu entscheiden.*

»Dieu est, ou il n'est pas«

heißen seine Alternativen. Die Wahrscheinlichkeit p für H_0: *Gott ist* möge nahezu unendlich klein sein. Da der menschliche Verstand keine Entscheidung für oder gegen die Existenz Gottes zu leisten imstande ist, spielt man pile ou face, d.h., man wirft eine (ideale) Münze und läßt Wappen oder Zahl entscheiden.

* *Denis Diderot* (1713–1784) bemerkte dazu, ein Imam könnte ebenso argumentieren.

Le Pari – Die Wette

$0 < p \ll 1$ | $1 - p$

H_0: *Gott ist* trifft zu		H_1: *Gott ist nicht* trifft zu	
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Entscheidung für H_0: Gewinn $= +\infty$, d. h. ein unendliches Leben voll unendlichen Glücks	Entscheidung für H_1: Gewinn $= 0$	Entscheidung für H_0: Gewinn $= g_0$. $g_0 \geqq 0$, falls religiöses Leben Glück ist. $g_0 \leqq 0$, falls Verzicht auf weltliche Freuden als Verlust empfunden wird.	Entscheidung für H_1: Gewinn $= g_1 < \infty$, d. h. höchstens endlich viele Güter dieser Welt

»Pesons le gain et la perte«: Setze ich auf Gott, so erwartet mich der Gewinn

$\mathscr{E}G = +\infty \cdot \frac{1}{2} \cdot p + g_0 \cdot \frac{1}{2} \cdot (1-p) = +\infty$.

Setze ich auf die Nichtexistenz Gottes, so erwartet mich der Gewinn

$\mathscr{E}G = 0 \cdot \frac{1}{2} \cdot p + g_1 \cdot \frac{1}{2} \cdot (1-p) < +\infty$.

»Setzen Sie also, ohne zu zögern, auf die Existenz Gottes.«
»Gagez donc, qu'il est, sans hésiter.«

Zusammenfassung: Für den Begriff des Tests ist es unwesentlich, daß eine Stichprobe mit Zurücklegen gezogen wird. Es kann sich statt dessen um irgendein anderes Zufallsexperiment handeln, auf dessen Ergebnisraum Ω eine Testgröße Z gewählt wird. Daher ist in der folgenden Definition nicht von der Binomialverteilung die Rede.

Definition 344.1: Für eine Testgröße Z gibt es über ein- und demselben Ergebnisraum Ω die beiden einfachen Hypothesen
$H_0 :=$ »Z besitzt die Wahrscheinlichkeitsverteilung P_0« und
$H_1 :=$ »Z besitzt die Wahrscheinlichkeitsverteilung P_1«.
Ferner sei A ein Ereignis in Ω; es heißt Annahmebereich für H_0.
Ω, A, P_0 und P_1 bestimmen einen **Alternativtest für zwei einfache Hypothesen** mit folgender Entscheidungsregel:

$$\delta\colon \begin{cases} A \text{ tritt ein} \Rightarrow \text{Entscheidung für } H_0 \\ \bar{A} \text{ tritt ein} \Rightarrow \text{Entscheidung für } H_1 \end{cases}$$

Die Wahrscheinlichkeiten für eine Fehlentscheidung sind

$$\alpha' = P_0(\bar{A}), \text{ falls } H_0 \text{ vorliegt,}$$
$$\beta' = P_1(A), \text{ falls } H_1 \text{ vorliegt.}$$

Bei Vorliegen von H_0 beträgt die Sicherheit des Urteils $1 - \alpha'$,
bei Vorliegen von H_1 beträgt die Sicherheit des Urteils $1 - \beta'$.

Figur 345.1 zeigt in einer vereinfachten Darstellung die Fehlerwahrscheinlichkeiten und die Sicherheiten, je nachdem, welche der beiden Hypothesen vorliegt.

Zum Abschluß geben wir noch einen Überblick über wichtige Aufgabentypen beim Alternativtest. Der Einfachheit halber handle es sich um Hypothesen über den Parameter p einer Binomialverteilung.

Typ 1: Stichprobenlänge n und kritischer Wert k sind gegeben; gesucht sind die Fehlerwahrscheinlichkeiten α' und β'.

Typ 2: Gegeben sind die Stichprobenlänge n und eine obere Schranke α für die Wahrscheinlichkeit α', einen Fehler 1. Art zu begehen. Gesucht ist der sog. beste kritische Wert k, für den α' höchstens α und β' möglichst klein werden.

Typ 3: Gegeben ist je eine obere Schranke α bzw. β für die Fehlerwahrscheinlichkeiten α' bzw. β'. Gesucht ist eine möglichst kleine Stichprobenlänge n und ein dazu passender kritischer Wert k. (Oft wird sich keine eindeutige Lösung ergeben.)

Typ 4: Gegeben sind die Stichprobenlänge n, die jeweiligen Schäden bei den Fehlern 1. bzw. 2. Art und die Wahrscheinlichkeiten für das tatsächliche Vorliegen der beiden Hypothesen. Gesucht ist derjenige kritische Wert k, für den der zu erwartende Schaden minimal wird.

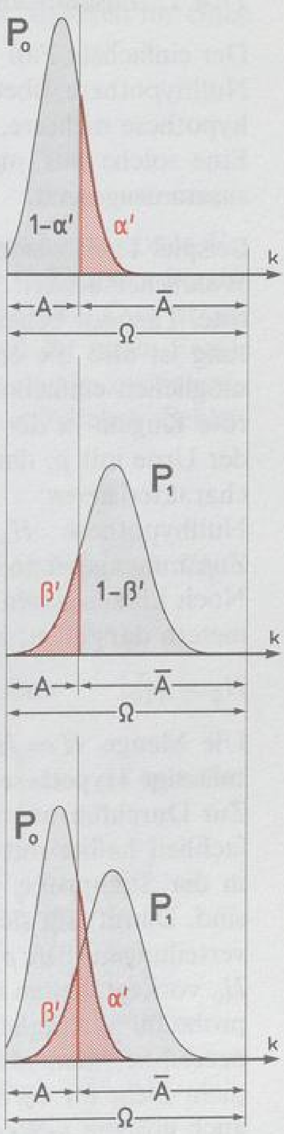


Fig. 345.1 Schematische Skizze für die Wahrscheinlichkeiten der Fehler und der Sicherheiten beim Alternativtest.

17.4. Signifikanztest

Die Situation eines Alternativtests, sich zwischen zwei einfachen Hypothesen entscheiden zu müssen, kommt in der Praxis selten vor, weil die Welt um uns dafür zu kompliziert ist. Sehr viel häufiger stellt sich einem jedoch das folgende **Problem:** Auf Grund irgendwelcher Erfahrungen oder Überlegungen hegt man *eine* Vermutung, die nun durch einen Test, den sog. Signifikanztest, entweder bestätigt oder widerlegt werden soll. Für diese Vermutung prägte *R. A. Fisher* (1890–1962) den Ausdruck **Nullhypothese.** Der Signifikanztest dient – wie sich zeigen wird – dazu, die Frage zu beantworten, ob man mit gutem Grund eine solche Nullhypothese ablehnen kann oder nicht.