Stochastik

Barth, Friedrich

München, [20]03

17. Das Testen von Hypothesen

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

17. Das Testen von Hypothesen

Das **Paris-Urteil** von *Joseph Hauber* (1766–1834) – Bayerische Staatsgemäldesammlungen. Der trojanische Prinz *Paris* hat auf dem Berg Ida zu entscheiden, welche der drei Göttinnen *Hera*, *Athene* und *Aphrodite* die schönste sei. Das von *Paris* angewandte Testverfahren, die Testgröße und die Entscheidungsregel sind nicht überliefert, lediglich der Ausfall des Tests: *Aphrodite* erhielt den mit der Aufschrift »Der Schönsten« versehenen goldenen Apfel der Zwietrachtgöttin *Eris* zugesprochen.

17. Das Testen von Hypothesen

17.1. Zur Geschichte und Aufgabe der Statistik

Στοχαστικὴ τέχνη, *Stochastik*, ins Lateinische übersetzt *ars conjectandi* (der Titel von *Jakob Bernoulli*s Buch über unseren Gegenstand) ist die Kunst, im Falle von Ungewißheit auf geschickte Weise Vermutungen anzustellen. Ursprünglich entwickelte sich die Stochastik aus dem Bedürfnis, die Gewinnchancen bei Glücksspielen in den Griff zu bekommen (Seite 71 ff.). Wenn dieser Gesichtspunkt auch heute noch interessant ist, so würde er allein es doch kaum rechtfertigen, daß Stochastik in der Schule gelehrt wird! Die wichtigste Anwendung findet die »Kunst des Vermutens« heute als *mathematische Statistik* in allen Zweigen der Wirtschaft, der Technik, der Politik und der Wissenschaften.

Verstand man Statistik schon immer in diesem Sinn? Nein; denn die mathematische Statistik entstand erst in diesem Jahrhundert und speist sich aus mehreren geschichtlichen Quellen.
Am Anfang steht die *Amtliche Statistik*, die bevölkerungsstatistischen Erhebungen. Überliefert sind uns Volkszählungen aus dem Alten Reich der Ägypter (um 2600 v. Chr.) und aus China. Der 6. König Roms, *Servius Tullius* (Regierungszeit 577–534), bestimmte in seiner Verfassung, alle 5 Jahre den census durchzuführen, eine Volkszählung, verbunden mit einer Erhebung über die Vermögensverhältnisse der Bürger und einer Einteilung für den Waffendienst. Eine solche Einteilung in Zensusklassen war auch in Griechenland üblich. Unter dem im Jahre 27 v. Chr. von *Augustus* (63 v. Chr. bis 14 n. Chr.) eingerichteten Prinzipat fanden die ersten Volkszählungen in den Provinzen des Römischen Reichs statt, so 27 v. Chr. in Gallien und 14 n. Chr. in Germanien. Die berühmteste Volkszählung ist wohl jener Provinzialcensus, der in Judäa im Jahre 6 n. Chr. durchgeführt wurde, als es römische Provinz wurde, und den *Lukas* in seinem Evangelium (2,1) irrtümlich für eine Reichszählung hält. Der 70. und letzte census fand 73 n. Chr. statt. Aber schon aus dem Alten Testament sind Volkszählungen bekannt. So künden das 2. Buch Mose (30,11) und das 4. Buch Mose (1), das auf lateinisch bezeichnenderweise *numeri* heißt, von einer von Gott angeordneten Volkszählung (um 1200 v. Chr.). König *David* (1004–965) hingegen verführte der Satan zu einer Volkszählung, wie im 2. Buch Samuel (24,2) und im 1. Buch der Chronik (21,2) berichtet wird. Für diesen Fürwitz wurde das Volk Israel mit der Pest bestraft. *Helmut Swoboda* meint:

> »Diese biblische Warnung bestimmte bis in die Neuzeit das Verhältnis zur statistischen Erhebung: Es war zweifellos sträfliche Neugier oder vorwitzige Vermessenheit, durch Volkszählungen oder gar durch systematische Beobachtungen von Geburten, Krankheiten und Todesfällen in die unerforschlichen Absichten Gottes Einsicht nehmen zu wollen.«

Die mittelalterlichen Erhebungen wie das karolingische *Capitulare de villis* und das *Domesday Book* (1086) *Wilhelms des Eroberers* sind daher fast ausschließlich Vermögens- oder Ständeerhebungen, so auch 1250 in Asti und 1288 in Mailand. Alle Seelen zählte erstmals Venedig 1422, und 1444 Straßburg, als eine Belagerung drohte. 1449 tat es Nürnberg.
Die Erweiterung des geographischen Horizonts, die wachsende Verflechtung der Staaten untereinander und die Ausweitung der Wirtschaftsbeziehungen zu Beginn der Neuzeit ließen eine weitere Quelle der heutigen Statistik entstehen, die *Staatskunde* als *Lehre von den Staatsmerkwürdigkeiten*, auch *Universitätsstatistik* genannt. So ist *Francesco Sansovinos* (1521–1586) Werk *Del governo et amministratione di diversi regni, et republiche*, [...] (1562) eine Sammlung von Staatsbeschreibungen. *Hermann Conring* (1606–1681) führte diese beschreibende Staatswissen-

schaft als Lehrfach an der Universität Helmstedt ein. *Gottfried Achenwall* (1719–1772) führte *Conring*s Arbeiten in Göttingen weiter. In seiner *Staatsverfassung** definierte er 1748 das Wort »Statistik« im Sinne von Staatskunde, wohl durch Rückgriff auf den lateinischen Begriff des *status rei publicae*, des Zustands des Staates. Er schreibt:

> »Der Inbegriff der wirklichen Staatsmerkwürdigkeiten eines Reichs, oder einer Republik, macht ihre Staatsverfassung im weitern Verstande aus: und die Lehre von der Staatsverfassung eines oder mehrerer einzelner Staaten, ist die Statistik [Staatskunde], oder Staatsbeschreibung.«

Er grenzt Statistik gegen die philosophische Staatslehre und gegen das Staatsrecht ab. Sein Schüler *August Ludwig von Schlözer* (1735–1809) in Göttingen und *Anton Friedrich Büsching* (1724–1793) in Berlin waren bedeutende Vertreter dieser Statistik.

Die dritte Quelle der modernen Statistik, die *Bevölkerungsstatistik* oder *Politische Arithmetik*, entsprang in England. Der Tuchhändler *John Graunt* (1620–1674) legte die Sterbelisten der Stadt London, beginnend mit dem Jahre 1603, seiner 1662 erschienenen Studie *Natural and Political Observations, mentioned in a Following Index, and Made upon the Bills of Mortality* zugrunde, dem ersten Werk über Bevölkerungsstatistik. Er wurde zum Begründer der Biometrie und der Bevölkerungsstatistik, die der Nationalökonom Sir *William Petty* (1623–1687) *Politische Arithmetik* nannte. Man sammelte bevölkerungsstatistische Massentatsachen und fragte nach ihren Ursachen und Regelmäßigkeiten. Eine erste Anwendung fanden solche Untersuchungen in der Ermittlung der Prämien für Leibrenten mittels einer Statistischen Mortalitätstheorie durch *Edmond Halley*** (1656–1742) in *An Estimate of the Degrees of Mortality of Mankind, drawn from curious Tables of the Births and Funerals at the City of Breslaw; with an Attempt to ascertain the Price of Annuities upon Lives* (1693). Sein Freund *Abraham de Moivre* (1667–1754) führte diese Untersuchungen weiter in seinen *Annuities on lives* (1725, 1743, 1750, 1752). *John Arbuthnot* (1667–1735) versuchte 1710 einen mathematischen Gottesbeweis auf statistischer Grundlage, ausgehend von der Tatsache der zahlenmäßigen Gleichheit der Geschlechter, obwohl in den letzten 82 Jahren in London fast konstant 18 Knabengeburten auf 17 Mädchengeburten kamen. Der bekannteste Vertreter der Politischen Arithmetik ist *Thomas Robert Malthus* (1766–1834). In Deutschland findet die Politische Arithmetik durch die Leistungen des Feldpredigers und späteren Oberkonsistorialrats *Johann Peter Süßmilch* (1707–1767) Anerkennung. Bevölkerungsstatistik dient auch bei ihm dem Nachweis, daß Gott die Welt weise eingerichtet hat, wie der Titel seines Werks zeigt: *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen* (1741).

Aber schon 1666 wurde die alte biblische Warnung in den Wind geschlagen; in La Nouvelle France (Quebec) fand die erste Volkszählung eines ganzen Landes in der Neuzeit statt. Deutsche Staaten begannen ab 1720 mit Volkszählungen. Schweden ordnete als erstes Land der Neuzeit 1749 regelmäßige Volkszählungen an; 1756 schuf es als erstes Land ein Statistisches Zentralamt, das sich mit der fortlaufenden Analyse der Bevölkerungszahlen beschäftigen sollte. 1790 begannen die USA mit regelmäßigen Volkszählungen, wie sie die Unionsverfassung als Grundlage für Wahlen verlangte. 1800 entstand in Paris das Bureau de Statistique, 1801 fanden erste Volkszählungen in Frankreich und Großbritannien (beschlossen bereits 1753) statt. Frankreich verwendete dabei Methoden, die *Laplace* vorgeschlagen hatte.

* Die 1. Auflage von 1748 trug den Titel *Vorbereitung zur Staatswissenschaft der Europäischen Reiche*. 1749 hieß sie dann *Abriß der Staatswissenschaft der Europäischen Reiche* und schließlich 1752 *Staatsverfassung der heutigen vornehmsten Europäischen Reiche und Völker im Grundrisse*.

** Gesprochen hæli. – In der zitierten Arbeit findet man neben analytischen Beweisen wahrscheinlichkeitstheoretischer Formeln zum ersten Mal auch geometrische Beweisverfahren, wie sie 1733 *Buffon* (1707–1788) verwendete (siehe dazu Anhang I). Als erster benützte *Newton* (1643–1727) geometrische Wahrscheinlichkeiten in einem Manuskript, geschrieben zwischen 1664 und 1666.

Die neue Amtliche Statistik, die Universitätsstatistik und die Politische Arithmetik verschmolzen im 19. Jahrhundert zur *Deskriptiven Statistik*. Diese untersucht eine Gesamtheit nach bestimmten, ihr wesenseigenen Merkmalen. Statistik in diesem Sinne ist also eine Kunst des geschickten Zählens und der Handhabung von Zählergebnissen. Von Vermutungen oder vom Zufall ist dabei nicht die Rede. Man rechnet im Gegenteil damit, daß durch das Erheben einer sehr großen Anzahl von Daten sich die Besonderheiten des Einzelfalls »herausmitteln« und dafür die allgemeinen Gesetzmäßigkeiten, der »Trend«, zutage treten.

Das Eindringen erster Vorstellungen aus der Wahrscheinlichkeitstheorie führte bei *Adolphe Quetelet* (1796–1874) zur Schaffung des statistischen Idealtyps, des *homme moyen*.* Sir *Francis Galton* (1822–1911) verfeinerte u.a. diese Begriffsbildung und begründete zusammen mit *Karl Pearson* (1857–1936) und Sir *Ronald Aylmer Fisher* (1890–1962) die biometrische Schule der Statistik.

Zu Beginn dieses Jahrhunderts zeichnete sich jedoch eine große Wende in der Statistik ab, die in den 30er Jahren zur Geburt der modernen Statistik, der *Mathematischen Statistik* oder auch der *Analytischen Statistik*, führte. Man erkannte, daß es vielfach unmöglich war, eine Gesamtheit durch eine Vollerhebung zu erfassen. Denken wir nur an die Qualitätskontrolle in der Industrie. Es wäre finanziell nicht tragbar und auch technisch oft unmöglich, *alle* Produkte einer Serienfertigung peinlich genau zu prüfen. Statt dessen schlug in den zwanziger Jahren *W.A. Shewhart* von den Bell Telephone Laboratories vor, eine *Zufallsstichprobe* von verhältnismäßig wenigen Stücken aus der laufenden Produktion zu entnehmen und diese um so sorgfältiger zu prüfen. Vom Prüfergebnis schließt man dann auf den Zustand der gesamten Ware und entscheidet, ob die Produktion weiterlaufen darf oder gestoppt werden muß. Dabei können natürlich Irrtümer vorkommen. Mit Hilfe der Mathematik ist es aber möglich, das Risiko des Irrtums zu kalkulieren und von vorneherein in gewünschten Grenzen zu halten**. Das Ziel der Mathematischen Statistik ist also nicht mehr die *Vollerhebung*. Statt ihrer sollen *Zufallsstichproben* Aufschluß geben über die Eigenschaften der Gesamtheit; Vermutungen, sog. *statistische Hypothesen*, sollen durch Stichproben entschieden werden. Die darauf basierenden Folgerungen heißen *statistische Schlüsse*, die natürlich im Sinne der klassischen Logik nie zwingend sein können. Unter Verwendung von Methoden der Höheren Mathematik entstand eine Vielfalt von Testverfahren zur Entscheidung von Hypothesen. Die von *R.A. Fisher* und anderen begründeten Verfahren wurden von *Egon Sharpe Pearson* (1895–1980) und *Jerzy Neyman* (1894–1981) zu einer Theorie der Stichproben ausgebaut. Während des 2. Weltkriegs entwarf *Abraham Wald* (1902–1950) die *Sequentialanalyse*, die als Kriegsgeheimnis galt und erst 1947 veröffentlicht werden konnte. Nach dem Kriege entwickelte er die *statistische Entscheidungstheorie*, die es erlaubt, auch in Situationen großer Ungewißheit noch vernünftig begründbare Entscheidungen zu fällen. Und so wird Statistik heute aufgefaßt, wenngleich die Amtliche Statistik immer noch das Material für viele Entscheidungen liefern muß.

Worin unterscheidet sich nun die Mathematische Statistik von der gewöhnlichen Wahrscheinlichkeitsrechnung, die wir bisher ausgiebig betrieben haben? Wir erläutern dies am wohlvertrauten Urnenbeispiel. Die Urne enthalte schwarze und andersfarbige Kugeln. In der Wahrscheinlichkeitsrechnung gehen wir davon aus, daß der Anteil p der schwarzen Kugeln *bekannt* ist. Man betrachtet ein Zufallsexperiment und *berechnet* die Wahrscheinlichkeiten dabei auftretender Ereignisse.

Anders ist jedoch die Ausgangslage in der Statistik. Nun ist der Anteil p der schwarzen Kugeln in der Urne *unbekannt*. Man führt ein Zufallsexperiment aus –

* Den Begriff des *homme moyen* prägte *Buffon* (1707–1788) in seinem *Essai d'arithmétique morale:* »[...] *l'homme moyen*, c'est-à-dire les hommes en général, bien portans ou malades, sains ou infirmes, vigoureux ou foibles.«

** Auf die große Bedeutung der Teilerhebung wies 1895 als erster der Norweger *Anders Nicolai Kiaer* (1838–1919) hin.

es handelt sich um das Ziehen einer Stichprobe – und *schließt* auf Grund des eingetretenen Ereignisses zurück auf den Anteil p der schwarzen Kugeln. Dabei unterscheidet man zwei Situationen.

1. Das Schätzproblem. Man hat keinerlei Vermutung über den Anteil p der schwarzen Kugeln in der Urne. In diesem Fall *schätzt* man den Anteil p auf Grund des eingetretenen Ereignisses (**Hochrechnung**). Man gibt als Schätzergebnis entweder einen einzigen Wert für p an (**Punktschätzung**) oder ein ganzes Intervall, in dem p liegen soll (**Intervallschätzung**). Das so abgegebene Urteil über den Anteil p ist mit einer gewissen Unsicherheit behaftet. Die Berechnung dieses Unsicherheitsgrades ist eine der wesentlichen Aufgaben der *Beurteilenden Statistik*.

2. Das Testproblem. Man hat von vornherein gewisse Vermutungen, Hypothesen genannt, über den Anteil p der schwarzen Kugeln in der Urne. Auf Grund des eingetretenen Ereignisses wird nun *entschieden*, welche dieser Hypothesen man beibehält oder verwirft. Auch hier ist es wesentlich, sich darüber klarzuwerden, mit welcher Sicherheit ein solches Urteil ausgesprochen werden kann.

Bevor wir uns diesen beiden Problemen zuwenden, wollen wir erst den Begriff der Stichprobe klären.

17.2. Stichproben

Die Grundlage aller Anwendungen der Stochastik ist die Möglichkeit, einen Versuch unter gleichen Bedingungen mehrmals zu wiederholen. Wollen wir z. B. über die Einkommensverteilung in einer Bevölkerung Ω etwas erfahren, so nützt es so gut wie nichts, wenn wir nur von einem zufällig ausgewählten Bürger ω das Einkommen wissen. Wir müssen eine Stichprobe von mehreren Personen ziehen. Dabei muß jede Person die gleiche Chance haben, in die Stichprobe aufgenommen zu werden. Man spricht dann von einer **Zufallsstichprobe*.**
Nun sei X die Zufallsgröße »Einkommen der ausgewählten Person in DM«. Sie habe die Wertemenge $\mathfrak{S} := \{x_1, \ldots, x_s\}$ und die Wahrscheinlichkeitsverteilung W mit

$$W(x_j) = \frac{\text{Zahl der Personen mit } x_j \text{ DM Einkommen}}{\text{Zahl aller Personen}} \qquad \text{für } j = 1, \ldots, s.$$

Um über W etwas zu erfahren, wählen wir n-mal eine Person aus der Gesamtbevölkerung aus. Wir erhalten als Ergebnis ein n-Tupel von Zahlen, die sämtlich der Wertemenge $\mathfrak{S}$ von X angehören. Diese n Zahlen hängen vom Zufall ab. Wir haben es also mit n verschiedenen Zufallsgrößen $X_1, \ldots, X_n$ zu tun:

$X_i :=$ »Einkommen der i-ten ausgewählten Person in DM«

mit $i = 1, \ldots, n$. Die Wertemengen aller X_i stimmen mit der von X überein; die

* Das Wort *Stichprobe* entstammt der Bergmannssprache. Die alten Schmelzöfen wurden angestochen, um die Schmelze auf ihren Zustand zu prüfen.

X_i haben sogar die *gleiche* Verteilung W wie X, da ihre Werte – vom Experiment her gesehen – unter den gleichen Bedingungen angenommen werden. Außerdem sind die X_i insgesamt *unabhängig*. In die mathematische Theorie des Stichprobenziehens gehen allein diese Eigenschaften der X_i ein. Wir halten daher fest:

Definition 335.1: Das n-Tupel $(X_1|X_2|\ldots|X_n)$ der Zufallsgrößen X_i heißt **Stichprobe der Länge n aus der Zufallsgröße X**, wenn gilt:

1. Die X_i sind stochastisch unabhängig.

2. Jedes X_i hat dieselbe Wahrscheinlichkeitsverteilung wie X.

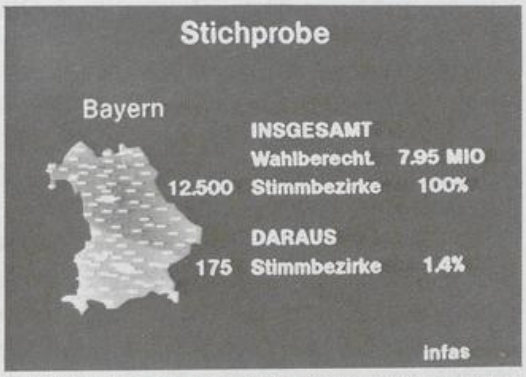


Bild 335.1 Landtagswahl in Bayern 1982 – ARD

Die Verwendung des Wortes »Stichprobe« ist in der Literatur nicht einheitlich. Oft nennt man auch das einzelne Werte-n-Tupel, das sich beim Stichprobenziehen ergibt, eine Stichprobe. Wir nennen gemäß Definition 335.1 das ganze Verfahren »Stichprobe«; das einzelne n-Tupel von Werten ist demgemäß **Stichprobenergebnis** zu nennen. Die Menge aller Stichprobenergebnisse kann als neuer Ergebnisraum genommen werden, der in diesem Zusammenhang auch manchmal **Stichprobenraum** (englisch: *sample space*) heißt. Ist $\mathfrak{S}$ die Wertemenge der Zufallsgröße X, so ist der Stichprobenraum das n-fache kartesische Produkt der Faktoren $\mathfrak{S}$, d. h.

$$\mathfrak{S}^n = \underbrace{\mathfrak{S} \times \mathfrak{S} \times \ldots \times \mathfrak{S}}_{n \text{ Faktoren}}$$

Beispiel: Die Zufallsgröße $X :=$ »Anzahl der Adler beim Wurf einer Münze« hat die Wertemenge $\mathfrak{S} = \{0; 1\}$. Eine Stichprobe der Länge n aus X ist das n-Tupel $(X_1|X_2|\ldots|X_n)$ mit $X_i :=$ »Anzahl der Adler beim i-ten Wurf«. Die X_i sind stochastisch unabhängig und besitzen dieselbe Wahrscheinlichkeitsverteilung wie X. Der Stichprobenraum $\mathfrak{S}^n$ ist die Menge aller n-Tupel, die aus 0 und 1 gebildet werden können.

Für verschiedene Fragestellungen der Praxis erweist es sich als zweckmäßig, auch dann noch von Stichproben zu sprechen, wenn die Zufallsgrößen $X_1, X_2, \ldots, X_n$ nicht mehr gleichverteilt oder nicht mehr stochastisch unabhängig sind. Das ist z. B. der Fall beim Ziehen *ohne Zurücklegen*. Man würde in unserem Beispiel alle n Personen auf einmal auswählen, so daß prinzipiell niemand die Chance hätte, zweimal gewählt zu werden. Die Zufallsgrößen $X_1, \ldots, X_n$ sind dann zwar noch gleichverteilt, aber nicht mehr unabhängig (siehe Aufgabe 223/**2**). Die für die Statistik wichtigen Formeln werden dadurch im allgemeinen komplizierter als beim Ziehen *mit Zurücklegen*. Glücklicherweise ist der Unterschied zwischen beiden Arten von Stichproben bei großen Grundgesamtheiten Ω (Bevölkerun-

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 Pascals* (1623–1662) *Infini-rien* (siehe Seite 343). Als ersten Test kann man *John Arbuthnots* (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**

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

* alter (lat.) = der eine von zweien, der andere.

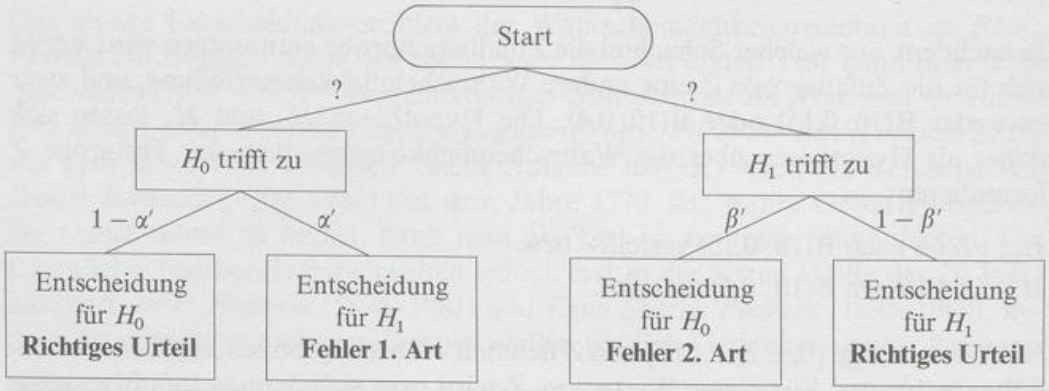
** Zur Herkunft des Wortes: lat. *testum*: Schüssel; altfranz. *test*: Tiegel für alchimistische Versuche; engl. *test*: Versuch, Prüfung.

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

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

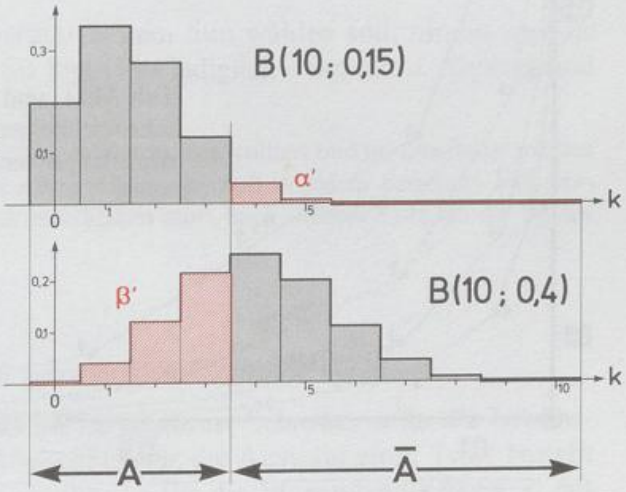
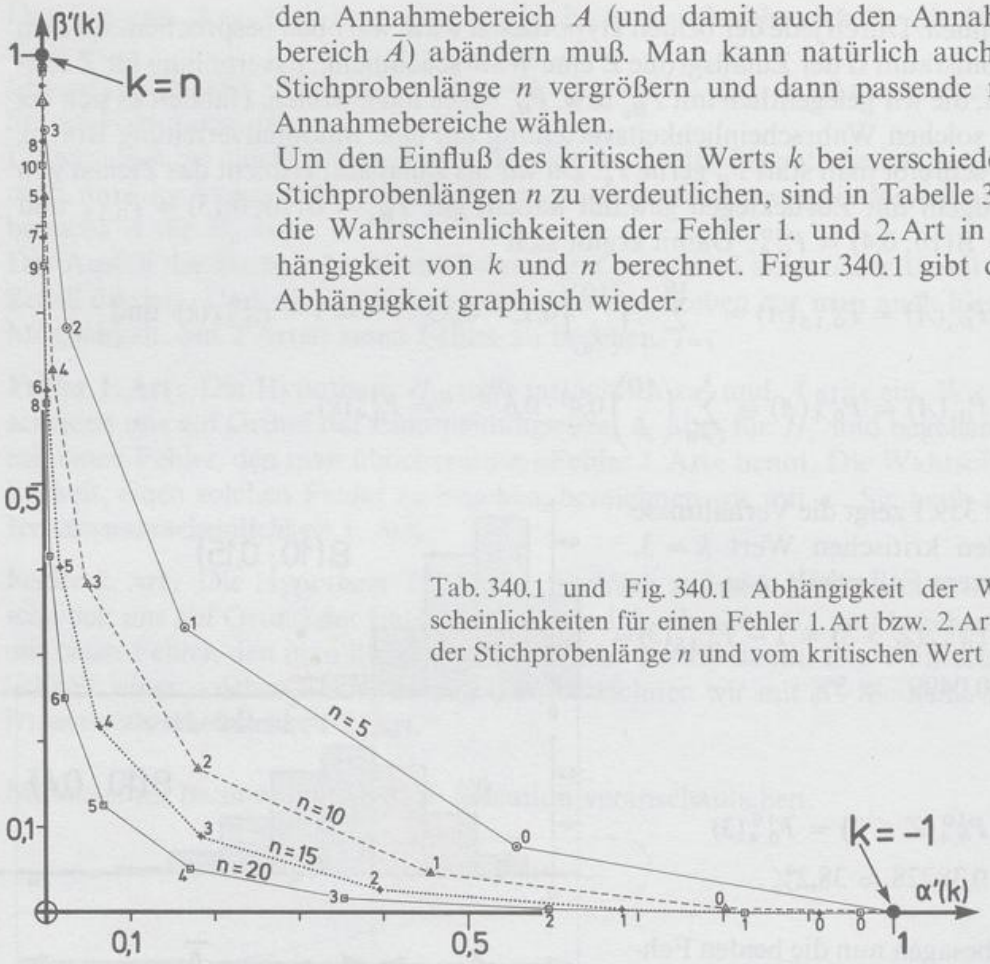


Fig. 339.1 Wahrscheinlichkeitsverteilung zur Hypothese 1 (Erste Qualität, oben) und Hypothese 2 (Zweite Qualität, unten) von Beispiel 1. grau: Wahrscheinlichkeit für ein richtiges Urteil. rot: Wahrscheinlichkeit für einen Fehler 1. Art bzw. 2. Art.

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

n	k	$\alpha'(k)$	$\beta'(k)$
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

n	k	$\alpha'(k)$	$\beta'(k)$
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

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

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

$$\alpha' = P_{0,15}^{20}(Z > 5) = 0{,}06731 \approx 6{,}7\% \qquad \text{und}$$

$$\beta' = P_{0,4}^{20}(Z \leqq 5) = 0{,}12560 \approx 12{,}6\%\,.$$

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 $1 - F_{0,15}^{10}(k) \leqq 0{,}01$ 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

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

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

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

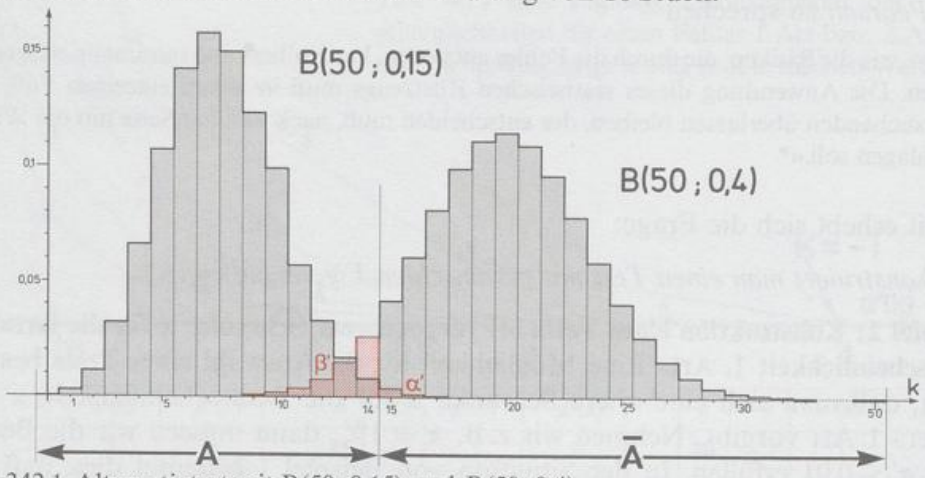
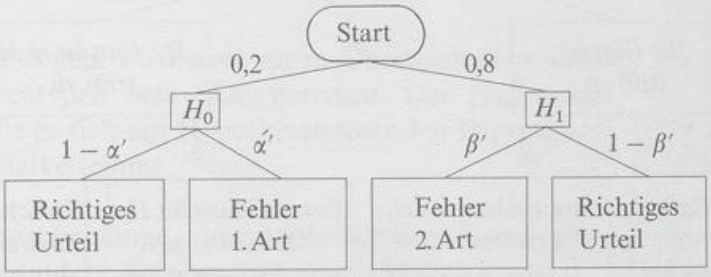


Fig. 342.1 Alternativtest mit B(50; 0,15) und B(50; 0,4)

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

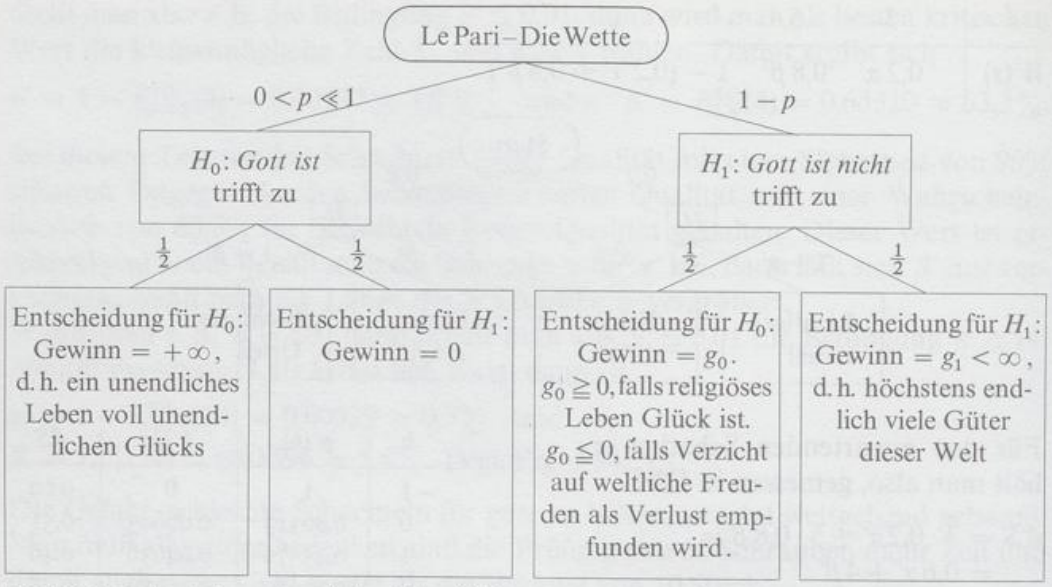
Tab. 343.1 Der zu erwartende Schaden $\mathscr{E}S$ beim Test der Hypothese »$p = 0{,}15$« gegen die Alternative »$p = 0{,}4$« bei der Stichprobenlänge $n = 10$ in Abhängigkeit von k

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.



»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:

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

Die Wahrscheinlichkeiten für eine Fehlentscheidung sind

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

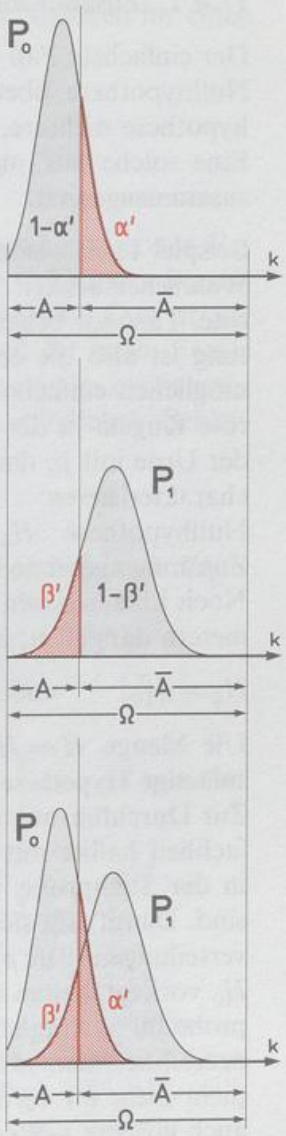


Fig. 345.1 Schematische Skizze für die Wahrscheinlichkeiten der Fehler und der Sicherheiten beim Alternativtest.

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.

17.4.1. Zusammengesetzte Hypothesen beim zweiseitigen Test

Der einfachste Fall eines Signifikanztests besteht zunächst einmal darin, daß die Nullhypothese, über die entschieden werden soll, einfach ist, wogegen als Gegenhypothese mehrere, meist sogar unendlich viele Hypothesen in Frage kommen. Eine solche aus mehreren einfachen Hypothesen bestehende Hypothese heißt **zusammengesetzt.**

Beispiel 1: Zweiseitiger Test einer einfachen Nullhypothese über eine unbekannte Wahrscheinlichkeit. Eine Urne enthalte 10 Kugeln, darunter womöglich auch rote. Theodor behauptet, die Urne enthalte genau 7 rote Kugeln. Diese Behauptung ist also die einfache Nullhypothese. Die Gegenhypothese besteht aus 10 möglichen einfachen Hypothesen; es können nämlich weniger oder mehr als 7 rote Kugeln in der Urne sein. Bezeichnet man den Anteil der roten Kugeln in der Urne mit p, dann kann man diese beiden Hypothesen folgendermaßen kurz charakterisieren:
Nullhypothese $H_0 : p = \frac{7}{10}$
Zusammengesetzte Gegenhypothese $H_1 : p \in \{0, \frac{1}{10}, \frac{2}{10}, \ldots, \frac{6}{10}, \frac{8}{10}, \frac{9}{10}, 1\}$
Noch kürzer lassen sich die beiden Hypothesen abstrakt als Mengen von Parametern darstellen; in unserem Fall

$$H_0 = \{\tfrac{7}{10}\} \qquad \text{und} \qquad H_1 = \{0, \tfrac{1}{10}, \tfrac{2}{10}, \ldots, \tfrac{6}{10}, \tfrac{8}{10}, \tfrac{9}{10}, 1\}.$$

Die Menge $H := H_0 \cup H_1$ ist die Menge aller zulässigen Parameter; sie heißt **zulässige Hypothese.**
Zur Durchführung des Tests ziehen wir eine Stichprobe von 6 Kugeln, der Einfachheit halber mit Zurücklegen. Testgröße Z ist die Anzahl der roten Kugeln in der Stichprobe, für die 11 Wahrscheinlichkeitsverteilungen $B(6;p)$ möglich sind. Damit läßt sich die zulässige Hypothese H auch als Menge aller Binomialverteilungen $B(6;p)$ mit $p \in \{0, \frac{1}{10}, \ldots, \frac{9}{10}, 1\}$ schreiben. Da $\mathscr{E}Z = 4{,}2$ ist, falls H_0 vorliegt, halten wir die Ergebnisse »4 rote« bzw. »5 rote Kugeln« in der Stichprobe für verträglich mit H_0. Größere Abweichungen vom Erwartungswert $\mathscr{E}Z$ bezeichnet man als **signifikante Abweichungen*.** Wir halten sie normalerweise nicht mehr für verträglich mit H_0. Da die Gegenhypothese sowohl kleinere als auch größere p-Werte als $\frac{7}{10}$ enthält, wird man als Annahmebereich für H_1 zwei getrennt liegende Intervalle wählen. Tests mit solchen Annahmebereichen heißen **zweiseitig.** In unserem Beispiel liegt somit folgende Entscheidungsregel nahe:

$$\delta: \begin{cases} Z \in \{0,1,2,3\} \cup \{6\} & \Rightarrow \text{Entscheidung für } H_1 \\ Z \in \{4,5\} & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

Wie beim Alternativtest haben wir auch hier 2 Möglichkeiten, Fehlentscheidungen zu treffen.

Fehler 1. Art: Die Nullhypothese H_0 trifft tatsächlich zu, aber $Z \in \{0,1,2,3,6\}$, d. h., es hat sich trotzdem eine signifikante Abweichung ergeben. Man würde

* significare (lat.) = anzeigen, verkünden.

sich also fälschlicherweise für H_1 entscheiden. Die Wahrscheinlichkeit für einen derartigen Fehler 1. Art ergibt sich zu

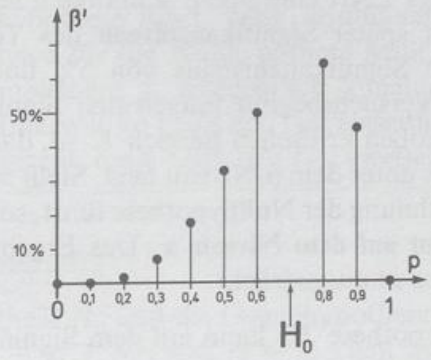
$$\begin{aligned}\alpha' = P^6_{0,7}(\{0,1,2,3,6\}) &= F^6_{0,7}(3) + \mathrm{B}(6;\tfrac{7}{10};6) = \\ &= 0{,}25569 + 0{,}11765 = \\ &= 0{,}37334 \approx 37{,}3\%.\end{aligned}$$

Fehler 2. Art: Eine der 10 einfachen Hypothesen aus der zusammengesetzten Gegenhypothese H_1 trifft tatsächlich zu, aber $Z \in \{4;5\}$. Man müßte sich für H_0 entscheiden. Und wie groß ist der Fehler, den man dann begeht? Das ist gar nicht so leicht zu beantworten! Denn die Wahrscheinlichkeit für einen Fehler 2. Art hängt nun davon ab, welche der einfachen Hypothesen, die die zusammengesetzte Hypothese H_1 bilden, tatsächlich vorliegt. Diese möglichen Fehlerwahrscheinlichkeiten β' hängen also von p ab:

$$\beta'(p) = P^6_p(\{4;5\}) = F^6_p(5) - F^6_p(3).$$

Eine leichte Rechnung liefert Tabelle 347.1, deren graphischer Ausdruck Figur 347.1 ist.

p	$\beta'(p)$
0	0
0,1	0,00127
0,2	0,01690
0,3	0,06974
0,4	0,17510
0,5	0,32813
0,6	0,49766
0,8	0,63898
0,9	0,45271
1	0



Tab. 347.1 und Fig. 347.1 Abhängigkeit der Wahrscheinlichkeit für einen Fehler 2. Art von der tatsächlich vorliegenden einfachen Gegenhypothese zur Nullhypothese »$p = 0{,}7$«

Weil man mit dem Schlimmsten rechnen muß, interessiert man sich für den Maximalwert der Wahrscheinlichkeit für einen Fehler 2. Art. In unserem Fall ist dies

$$\beta'(\tfrac{8}{10}) = 0{,}63898 \approx 63{,}9\%.$$

Dieser Wert ist so groß, daß man sich trotz der oben aufgestellten Entscheidungsregel guten Gewissens nicht für H_0 entscheiden kann. Dieses schlechte Gewissen bringt der Statistiker dadurch zum Ausdruck, daß er in diesem Fall sagt: »Man kann die Nullhypothese H_0 nicht ablehnen (nicht verwerfen).« *Ronald Aylmer Fisher* (1890–1962) schreibt dazu 1935 in *The Design of Experiments:*

»[...] it should be noted that the null hypothesis is never proved or established, but is possibly disproved in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis.«

Die Entscheidung eines Signifikanztests besteht also nicht in der Entscheidung für H_0 oder für H_1, sondern nur in der Ablehnung der Nullhypothese H_0. Eine solche Entscheidung fällt man genau dann, wenn die Testgröße Z einen der signifikanten Werte aus $\{0, 1, 2, 3, 6\}$ annimmt. Man nennt diesen Annahmebereich der Gegenhypothese den **kritischen Bereich** K. Wir müssen also die oben aufgestellte Entscheidungsregel revidieren! Bei einem Signifikanztest lautet sie

$$\delta: \begin{cases} Z \in K \Rightarrow \text{Nullhypothese } H_0 \text{ wird abgelehnt.} \\ Z \in \bar{K} \Rightarrow \text{Nullhypothese } H_0 \text{ kann nicht abgelehnt werden.} \end{cases}$$

In Worten: Ist der Ausfall der Stichprobe signifikant, so wird die Nullhypothese abgelehnt, andernfalls beibehalten.
Im Falle $Z \in \bar{K}$ fällt also eigentlich gar keine Entscheidung! Weil dem so ist, interessiert man sich beim Signifikanztest nur für den Fehler 1. Art, die Nullhypothese auf Grund eines signifikanten Ausfalls der Stichprobe zu verwerfen, obwohl sie zutrifft. Fußend auf den Erkenntnissen von *Poisson* (1781–1840) führte 1840 sein Schüler, der Arzt *Louis-Dominique-Jules Gavarret**, in seinem Werk *Principes généraux de statistique médicale* ein, für die Wahrscheinlichkeit α' dieses Fehlers 1. Art eine obere Schranke α festzulegen. Diese obere Schranke α nannte man später **Signifikanzniveau** des Tests. Die heute besonders häufig verwendeten Signifikanzniveaus von 5% und 1% führte *R. A. Fisher* ein. Zu einem vor Versuchsbeginn festgelegten Signifikanzniveau α wählt man einen möglichst großen kritischen Bereich K so, daß die Wahrscheinlichkeit für einen Fehler 1. Art unter dem α-Niveau liegt. Stellt sich dann ein Versuchsergebnis ein, das zur Ablehnung der Nullhypothese führt, so sagt man, dieses Versuchsergebnis sei **signifikant auf dem Niveau** α. Das Ergebnis des Tests wird in diesem Fall üblicherweise so ausgedrückt:

»Die Nullhypothese H_0 kann auf dem Signifikanzniveau α abgelehnt werden.«

Die statistische Sicherheit des Urteils hat dann mindestens den Wert $1 - \alpha$.

Versuchen wir nun zu $\alpha = 25\%$ einen kritischen Bereich K für Theodors Vermutung $H_0 = \{\frac{7}{10}\}$ bzw. $H_0 = »Z$ ist nach $B(6; \frac{7}{10})$ verteilt« zu konstruieren. Dem Problem angemessen setzt sich der kritische Bereich K aus zwei Intervallen $[0; k_1]$ und $[k_2; 6]$ zusammen. Es gäbe viele Möglichkeiten, die Fehlerwahrscheinlichkeit α' auf die beiden Teilintervalle aufzuteilen. Üblich ist es, k_1 und k_2 so zu bestimmen, daß in jedem Teilbereich die Fehlerwahrscheinlichkeiten höchstens $\frac{1}{2}\alpha$ sind. Das führt zu

$$\begin{aligned} & P_{H_0}(Z \leqq k_1) \leqq 12{,}5\% \quad \text{und} \quad P_{H_0}(Z \geqq k_2) \leqq 12{,}5\%. \\ \Leftrightarrow\ & F^6_{0,7}(k_1) \leqq 12{,}5\% \quad \text{und} \quad 1 - F^6_{0,7}(k_2 - 1) \leqq 12{,}5\%. \end{aligned}$$

Das ergibt mit Hilfe der *Stochastik-Tabellen* die Bedingungen

$$k_1 \leqq 2 \quad \text{und} \quad k_2 \geqq 6, \quad \text{also} \quad K = [0; 2] \cup [6; 6] = \{0, 1, 2, 6\}.$$

* 28. 1. 1809 Astaffort – 31. 8. 1890 Valmont. Vor seinem Medizinstudium Artillerie-Offizier; 1843 wurde er auf den Lehrstuhl für Physique médicale der Medizinischen Fakultät von Paris berufen.

Hätte Theodors Stichprobe beispielsweise 2 rote Kugeln geliefert, so könnte man seine Vermutung H_0, die Urne enthalte 7 rote Kugeln, auf dem 25%-Niveau ablehnen. Die Sicherheit des Urteils »Ablehnung von H_0« beträgt mindestens 75%.

Je niedriger das Signifikanzniveau, d.h., je kleiner α ist, desto schärfer ist der Test, aber desto seltener wird man H_0 verwerfen können. Dies entspricht der Erfahrung des täglichen Lebens: Klare Urteile kann man nur selten abgeben, verschwommene Aussagen (d.h. großes Signifikanzniveau!) sind hingegen sehr leicht zu machen.
Wir fassen die Erkenntnisse aus Beispiel 1 zusammen in

Definition 349.1:
Beschränkt man sich bei einem Test darauf, nur für die eine der beiden Hypothesen die Wahrscheinlichkeit α' der fälschlichen Ablehnung klein zu machen, so spricht man von einem **Signifikanztest**. Man nennt diese Hypothese dann **Nullhypothese**. Die gewählte obere Schranke α für die Irrtumswahrscheinlichkeit α' heißt auch **Signifikanzniveau**. Ein Versuchsergebnis, das zur Ablehnung der Nullhypothese führt, heißt **signifikant auf dem Niveau α**. Der Ablehnungsbereich für die Nullhypothese heißt **kritischer Bereich K** des Tests, sein Komplement $\overline{K}$ gelegentlich Annahmebereich. Besteht K aus einem einzigen Intervall, so heißt der Test **einseitig**. Wird K durch $\overline{K}$ in zwei Intervalle aufgeteilt, dann heißt der Test **zweiseitig**.

Wie konstruiert man einen Signifikanztest?

1. Man formuliert eine Nullhypothese H_0 und die Gegenhypothese H_1 bzw. die zulässige Hypothese H. Dabei – so *J. Neyman* 1939 in Genf auf einer vom Völkerbund veranstalteten Tagung –

   »hat sich mehr oder weniger eingebürgert, als Nullhypothese diejenige Hypothese zu wählen, bei der die Fehler 1. Art von größerer Bedeutung sind als die Fehler 2. Art.«

2. Man legt eine Testgröße Z fest.
3. Man legt das Signifikanzniveau α des Tests fest.
4. Man konstruiert einen möglichst großen kritischen Bereich K so, daß $P_{H_0}(Z \in K) \leqq \alpha$.
   Besteht K aus zwei Teilintervallen K_1 und K_2, dann bestimmt man sie so, daß $P(Z \in K_1) \leqq \frac{1}{2}\alpha$ und $P(Z \in K_2) \leqq \frac{1}{2}\alpha$ erfüllt sind.
5. Man entscheidet nach folgender Regel:

   $$\delta\colon \begin{cases} Z \in K \Rightarrow H_0 \text{ wird abgelehnt.} \\ Z \in \overline{K} \Rightarrow H_0 \text{ kann nicht abgelehnt werden.} \end{cases}$$

6. Sicherheit des Urteils:
   $1 - \alpha$ heißt **statistische Sicherheit** des Urteils »Ablehnung von H_0«, weil mindestens mit der Wahrscheinlichkeit $1 - \alpha$ das Vorliegen von H_0 erkannt würde.

Zur Veranschaulichung der statistischen Sicherheit stellen wir uns vor, daß n Urnen zum Testen vorliegen. n_0 dieser Urnen enthalten tatsächlich 7 rote Kugeln.

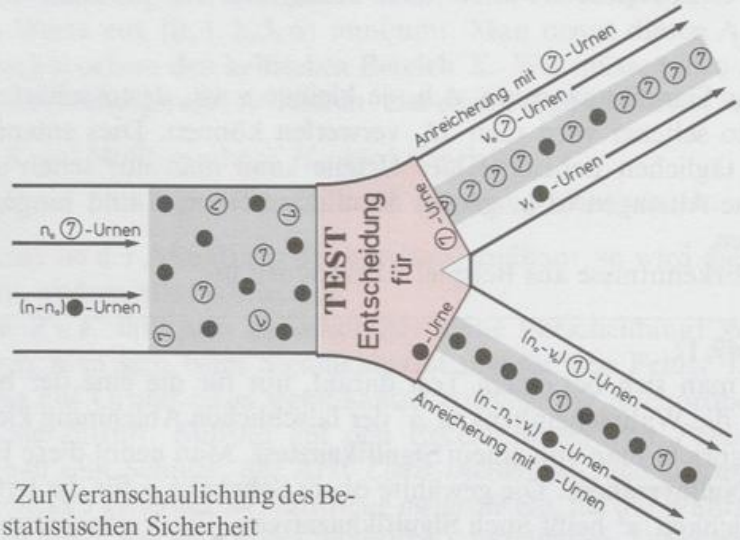


Fig. 350.1 Zur Veranschaulichung des Begriffs der statistischen Sicherheit

(In Figur 350.1 mit ⑦ gekennzeichnet.) Auf Grund der Interpretationsregel für Wahrscheinlichkeiten werden etwa $\alpha' = 37{,}3\%$ dieser Urnen falsch bezeichnet. Der Anteil der falsch bezeichneten Urnen des anderen Typs hängt davon ab, wie viele rote Kugeln die Urne jeweils enthält.

Natürlich ist ein Test kein todsicheres Verfahren zur Trennung der beiden Hypothesen; denn man muß immer Fehlermöglichkeiten in Kauf nehmen. Hören wir dazu *J. Neyman und E.S. Pearson:*

»The tests themselves give no final verdict, but as tools help the worker who is using them to form his final decision; [...]. What is of chief importance in order that a sound judgment may be formed is that the method adopted, its scope and its limitations, should be clearly understood«.*

17.4.2. Zusammengesetzte Hypothesen beim einseitigen Test

Beispiel 2: Einseitiger Test einer einfachen Nullhypothese über eine unbekannte Wahrscheinlichkeit. Der Teetassen-Test von *R. A. Fisher***: Lady X. behauptet, sie könne es am Geschmack erkennen, ob der Tee zuerst in der Tasse war und die Milch dazugegeben wurde oder ob man umgekehrt den Tee auf die Milch gegossen habe.
Wir glauben das nicht. Wir setzen, anders als *R. A. Fisher*, Lady X. 10 Tassen Tee mit Milch vor, die in beliebiger – uns bekannter – Weise gefüllt worden sind.

* *On the use and interpretation of certain test criteria for purposes of statistical inference.* Biometrika **20 A** (1928).

** *Sir Ronald Aylmer Fisher* (1890–1962) wählte in *The Design of Experiments* (1935) dieses Beispiel zur Einführung: »A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested.«

Lady X. probiert und macht 8mal eine richtige Angabe. Können wir Lady X. die von ihr behauptete geradezu übernatürliche Fähigkeit zugestehen?
Im Gegensatz zu Beispiel 1 aus **17.3.1.** ist das Ergebnis der Stichprobe bereits bekannt. Eine solche Situation ist in der Praxis auch oft anzutreffen. Man könnte nun zwar auch hier vorgehen wie in Beispiel 1, zu einem vorgegebenen Signifikanzniveau α einen kritischen Bereich bestimmen und überprüfen, ob das bekannte Ergebnis des Zufallsexperiments zur Ablehnung der Nullhypothese hinreicht. Statt dessen geht man oft anders vor und bestimmt zu dem eingetretenen Stichprobenergebnis das niedrigste Signifikanzniveau, auf dem man gerade noch die Nullhypothese ablehnen könnte. Wir wollen diese andere Art eines Signifikanztests hier weiter verfolgen. Dazu legen wir uns wieder ein mathematisches Modell für dieses reale Zufallsexperiment zurecht. Das Probieren der Tassen entspricht einer *Bernoulli*-Kette der Länge 10; Treffer beim i-ten Versuch ist das Ereignis »Lady X. beurteilt die i-te Tasse richtig«. Wenn Lady X. sich aufs bloße Raten verlegte, könnte sie genausogut mit einer Laplace-Münze werfen. In diesem Fall hätte also der Parameter der *Bernoulli*-Kette den Wert $\frac{1}{2}$. Besitzt Lady X. hingegen eine Begabung der behaupteten Art, so ist die Wahrscheinlichkeit p für einen Treffer verschieden von $\frac{1}{2}$. $p < \frac{1}{2}$ würde bedeuten, daß Lady X. den Sachverhalt zwar mit gewisser Sicherheit richtig erkennen kann, ihn aber verkehrt benennt. Das hätte sie wohl bei eigenen Versuchen längst selbst bemerkt. Es ist somit sinnvoll, als zulässige Hypothese die Menge $H := \{p \mid \frac{1}{2} \leqq p \leqq 1\}$ zu nehmen. Der Wert p ist also ein Maß für die Begabung von Lady X.; je größer p ist, um so begabter ist sie. Wir wählen als Nullhypothese »Lady X. hat keine Begabung«, kurz »Lady X. rät blind«, also $H_0 := \{\frac{1}{2}\}$, da uns hier ein Fehler 1. Art, nämlich eine unbegabte Dame für begabt zu halten, schlimmer erscheint als ein Fehler 2. Art, nämlich einer begabten Dame die Begabung abzusprechen. Nehmen wir als Testgröße Z die Anzahl der richtig geratenen Tassen, so besagt H_0, Z besitzt die Wahrscheinlichkeitsverteilung $B(10; \frac{1}{2})$. Die Gegenhypothese lautet »Lady X. ist begabt« also $H_1 := H \backslash H_0$. Sie läßt sich nicht mehr durch endlich viele Parameterwerte beschreiben; alle Zahlen $p \in]\frac{1}{2}; 1]$ sind möglich. Es gibt somit für die Zufallsgröße Z unendlich viele Wahrscheinlichkeitsverteilungen zu dieser Hypothese, nämlich alle $B(10; p)$ mit $p > \frac{1}{2}$. Da alle p-Werte der Gegenhypothese H_1 auf derselben Seite bezüglich der Nullhypothese »$p = \frac{1}{2}$« liegen, wählt man sinnvollerweise als kritischen Bereich ein Intervall $K := [k; 10]$, so daß das Ereignis »$Z \geqq k$« zur Ablehnung der Nullhypothese führt. Würde man nämlich als kritischen Bereich das Ereignis $K' := [0; k_1] \cup [k_2; 10]$ wählen, so würde man im Falle $Z \in K'$ die Nullhypothese ablehnen, also Lady X. auch dann Begabung bescheinigen, wenn sie nur wenige oder gar keine Tasse richtig benannt hat, was sicherlich nicht erwünscht ist. Da K aus einem einzigen Intervall besteht, handelt es sich also um einen einseitigen Test.
Unser Stichprobenergebnis lautet »$Z = 8$«. Wir müssen somit einen kritischen Bereich wählen, der 8 enthält. Ein möglichst niedriges Signifikanzniveau erzielt man, wenn man den kritischen Bereich möglichst klein wählt. Also entschließen wir uns zu $K := [8; 10]$. Für die Wahrscheinlichkeit α', einen Fehler 1. Art zu begehen, ergibt sich damit
$\alpha' = P_{H_0}(Z \in K) = P^{10}_{0,5}(Z \geqq 8) = 1 - F^{10}_{0,5}(7) \approx 5{,}5\%$.

Beim üblichen Signifikanzniveau 5% können wir die Nullhypothese »Lady X. rät blind« nicht ablehnen. Ist man jedoch mit einem Signifikanzniveau von 5,5% oder höher zufrieden, so kann man die Nullhypothese »Lady X. rät blind« ablehnen und der Dame Begabung bescheinigen. Die statistische Sicherheit unseres Urteils »Lady X. ist begabt« beträgt dann höchstens 94,5%. Was heißt das? Wenn viele Ladies sich unserer Prüfung unterzögen, attestierten wir ca. 5,5% dieser Damen fälschlicherweise eine gewisse Begabung, weil sie 8 oder mehr Tassen richtig benennen, obwohl sie blind raten.
Was ist aber mit den begabten Damen? Dieser Frage wollen wir im nächsten Abschnitt nachgehen.

17.4.3. Die Operationscharakteristik eines Tests

Beispiel 3: Dem Teetassentest aus Beispiel 2 stellt sich eine Lady, die tatsächlich über eine gewisse Begabung verfügt und mit der Wahrscheinlichkeit $p = 0{,}6$ die Tassen richtig benennt. Mit welcher Wahrscheinlichkeit wird man ihre Begabung verkennen, wenn wir wie in Beispiel 2 als kritischen Bereich die Menge $K = [8; 10]$ nehmen?
Die Wahrscheinlichkeit β', einen solchen Fehler 2. Art zu begehen, ergibt sich zu

$$\beta' = P^{10}_{0,6}(Z \in \bar{K}) = P^{10}_{0,6}(Z \leqq 7) = F^{10}_{0,6}(7) \approx 83{,}3\%.$$

Solchen schwach begabten Damen wird mit unserem Test also oft unrecht getan! Wäre die Begabung der Dame größer, z. B. $p = 0{,}9$, so würden wir sie auch besser erkennen; es ergäbe sich nämlich $\beta' = F^{10}_{0,9}(7) \approx 7{,}0\%$. Weil wir aber über die Begabung der Damen, die sich dem Test unterziehen, nichts wissen, müssen wir uns einen Überblick über alle Wahrscheinlichkeiten für einen Fehler 2. Art verschaffen. Da diese Wahrscheinlichkeiten offensichtlich von p abhängen, betrachten wir die Funktion

$$\beta' \colon p \mapsto P^{10}_p(Z \in \bar{K}),\ D_{\beta'} = \left]\tfrac{1}{2}; 1\right].$$

Mit Hilfe einer Wertetabelle können wir den Graphen dieser Funktion zeichnen (Tabelle 353.1 und Figur 353.1).

Man erkennt, daß die Wahrscheinlichkeit β' für einen Fehler 2. Art um so größer wird, je weniger sich die Begabung vom blinden Raten ($p = \frac{1}{2}$) unterscheidet. Da die Definitionsmenge $D_{\beta'}$ links offen ist, gibt es keine größte Irrtumswahrscheinlichkeit 2. Art. Als Ersatz dafür nimmt man das Supremum aller Irrtumswahrscheinlichkeiten 2. Art, also den Wert $1 - \alpha'$. Er ist in unserem Fall etwa 94,5%. Man riskiert also, mit einer Wahrscheinlichkeit bis zu 94,5% begabte – wenn auch sehr schwach begabte – Damen zu Unrecht für unbegabt zu halten. Wir können trotzdem zufrieden sein: Der unangenehme Fall, daß eine Dame nur flunkert und wir ihr dennoch hohe Sensibilität bescheinigen, tritt nur mit 5,5% Wahrscheinlichkeit ein. Daß wir andererseits u. U. einer wirklich begabten Dame ein Unrecht antun, nehmen wir in Kauf in der Gewißheit, daß sich das Genie so oder so eines Tages durchsetzen wird.

p	$\beta' = P_p^{10}(Z \leqq 7)$
0,51	0,94
55	90
60	83
65	74
70	62
75	47
80	32
85	18
90	07
95	01
99	0001
1	0

Tab. 353.1 Wahrscheinlichkeit β' für einen Fehler 2. Art beim kritischen Bereich $K = [8; 10]$

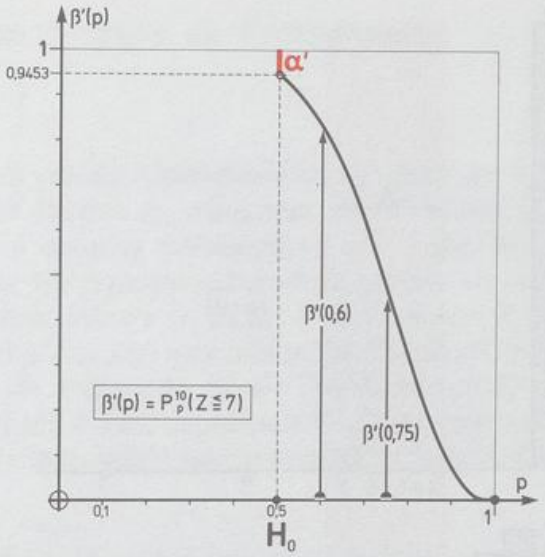


Fig. 353.1 Graph der Funktion $\beta' : p \mapsto P_p^{10}(Z \in \overline{K})$

Setzt sich die Gegenhypothese nur aus endlich vielen einfachen Hypothesen zusammen wie bei Theodors Urne in Beispiel 1 von Seite 346, dann besteht der Graph von β' nur aus diskreten Punkten, so wie ihn Figur 347.1 zeigt. In einem solchen Fall gibt es natürlich eine größte Irrtumswahrscheinlichkeit 2. Art.
Es hat sich in der Statistik eingebürgert, die auf der Gegenhypothese H_1 definierte Funktion $p \mapsto \beta'(p)$ auf die Menge *aller* beim Test betrachteten Hypothesen, d. h. auf die zulässige Hypothese $H := H_0 \cup H_1$ fortzusetzen. Diese Funktion heißt dann Operationscharakteristik des Tests, kurz OC des Tests.

Definition 353.1: Es sei auf dem Ergebnisraum Ω der Testgröße Z eine Menge von Wahrscheinlichkeitsverteilungen als zulässige Hypothese H gegeben. Diese Verteilungen lassen sich durch einen Parameter p kennzeichnen. $A \subset \Omega$ sei ein Ereignis. Dann heißt die Funktion

$$\text{OC}: p \mapsto P_p(A),\ D_{\text{OC}} = H$$

die **Operationscharakteristik des Ereignisses A bezüglich H.** Ihr Graph heißt **OC-Kurve.***

Bemerkung: Der Parameter p muß nicht unbedingt eine Wahrscheinlichkeit sein. So werden z. B. *Poisson*-Verteilungen durch den Parameter »Erwartungswert μ«, Normalverteilungen durch die Parameter μ und σ^2 gekennzeichnet.
Figur 354.1 veranschaulicht am Beispiel des Ereignisses $A := [4; 7]$ und an der Schar $B(16; p)$, $p \in [0; 1]$, als zulässiger Hypothese das Zustandekommen der

* In der Literatur verwendet man vielfach noch die ursprünglich von *Jerzy Neyman* und *E. S. Pearson* zur Kennzeichnung der Güte oder Macht eines Tests eingeführte *power function* = **Gütefunktion g**. Für sie gilt $g(p) := 1 - \text{OC}(p)$.

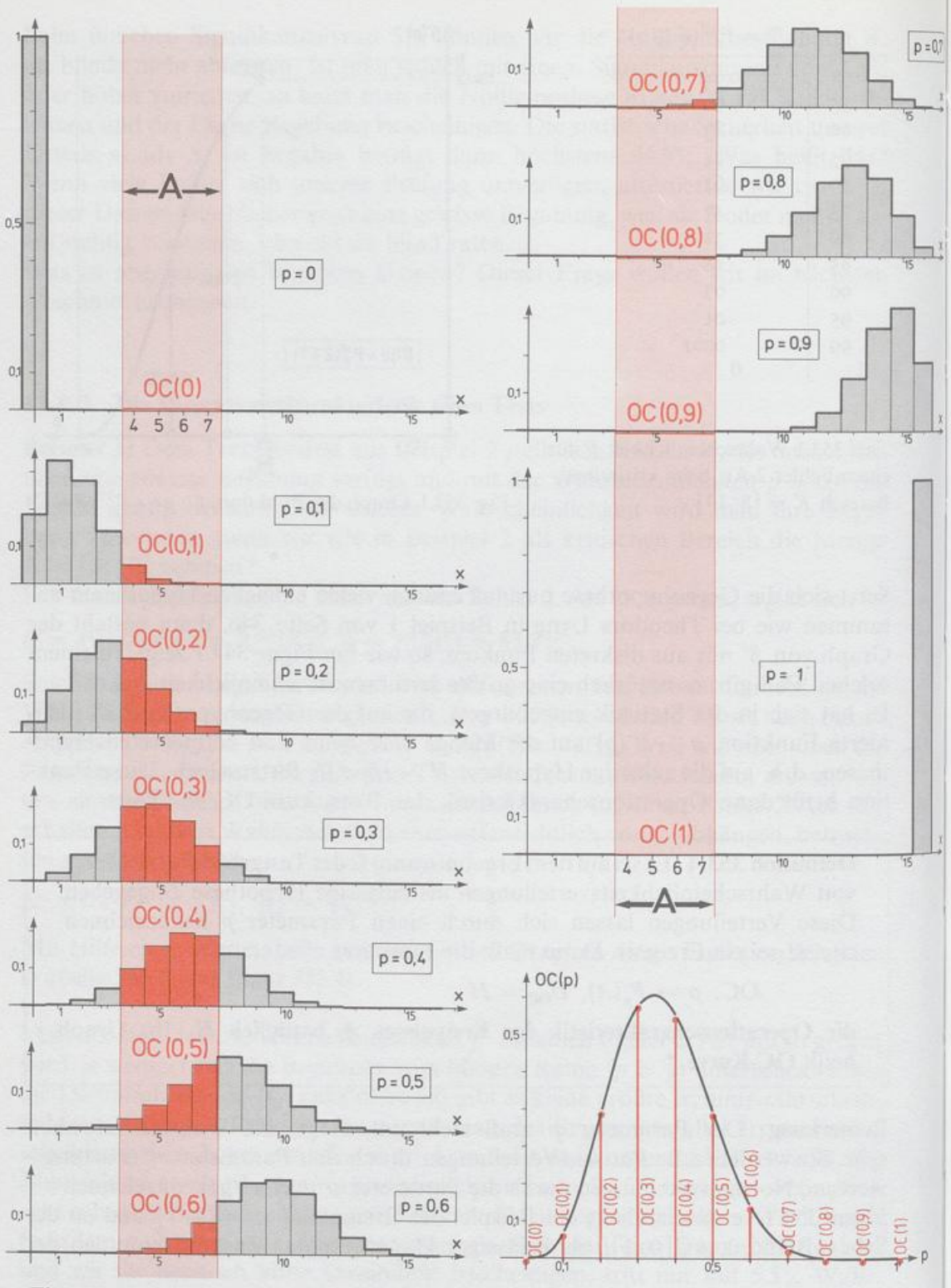


Fig. 354.1 Veranschaulichung der Entstehung der OC des Ereignisses $A := [4; 7]$ bezüglich der zulässigen Hypothese $H := \{B(16; p) | p \in [0; 1]\}$. Bedeutet f_p die Dichtefunktion der Binomialverteilung $B(16; p)$, so läßt sich die Operationscharakteristik mittels eines Integrals schreiben, nämlich OC: $p \mapsto \int_{3,5}^{7,5} f_p(t)\,dt$.

Operationscharakteristik: Zu jedem p gehört als Funktionswert

$$\mathrm{OC}(p) = P_p^{16}(Z \in A) = \sum_{i=4}^{7} \mathrm{B}(16; p; i).$$

Bei einem Signifikanztest spricht man von der Operationscharakteristik der Entscheidungsregel δ mit dem kritischen Bereich K, wenn man $A = \bar{K}$ wählt. Ihre Funktionswerte $\mathrm{OC}(p) = P_p(\bar{K})$ sind dann in Abhängigkeit von p die Wahrscheinlichkeiten, mit denen man die Nullhypothese beibehält, gleich, ob diese Entscheidung die richtige ist oder nicht. Für $p \in H_1$ ist der Funktionswert $P_p(\bar{K})$ jeweils die Irrtumswahrscheinlichkeit 2. Art, daß man nämlich die Nullhypothese nicht ablehnt, obwohl sie nicht zutrifft. Für $p \in H_0$ ist der Funktionswert $P_p(\bar{K})$ jeweils gleich der Sicherheit $1 - \alpha'(p)$, mit der die zutreffende Nullhypothese nicht abgelehnt wird. Dabei ist $\alpha'(p)$ die zu $p \in H_0$ gehörende Irrtumswahrscheinlichkeit 1. Art.

Übrigens kann auch die Nullhypothese H_0 selbst zusammengesetzt sein. Nehmen wir etwa im Teetassentest von Beispiel 2 (**17.4.2.**) als zulässige Hypothese $H := [0; 1]$ und als Nullhypothese $H_0 := [0; \frac{1}{2}]$, dann ergäbe sich als Operationscharakteristik des Ereignisses »$Z \leqq 7$« die Funktion OC: $p \mapsto F_p^{10}(7)$, $D_{\mathrm{OC}} = [0; 1]$, deren Graph Figur 355.1 wiedergibt. Nun gibt es auch unendlich viele Irrtumswahrscheinlichkeiten 1. Art. Zur Charakterisierung des Tests genügt es offenbar, die größte dieser Wahrscheinlichkeiten anzugeben.
Je nach Lage des kritischen Bereichs K haben die Graphen der Operationscharakteristik, kurz OC-Kurven genannt, eine typische Gestalt. Nehmen wir als zulässige Hypothese die Menge aller Binomialverteilungen $\mathrm{B}(n; p)$ mit $p \in [0; 1]$, so gibt es 4 besonders wichtige Typen. Der Nachweis der aufgeführten Eigenschaften wird Aufgabe 372/**48** vorbehalten.

1) $K := [0; k] \Rightarrow$ OC: $p \mapsto 1 - F_p^n(k)$
Ist K linksbündig, so ist die OC-Kurve echt monoton steigend.

2) $K := [k; n] \Rightarrow$ OC: $p \mapsto F_p^n(k-1)$
Ist K rechtsbündig, so ist die OC-Kurve echt monoton fallend.

3) $K := [0; k_1] \cup [k_2; n] \Rightarrow$
OC: $p \mapsto F_p^n(k_2 - 1) - F_p^n(k_1)$
Ist K getrennt, so hat die OC-Kurve einen inneren Hochpunkt.

4) $K := [k_1; k_2] \Rightarrow$
OC: $p \mapsto F_p^n(k_1 - 1) + 1 - F_p^n(k_2)$
Ist K ein inneres Intervall, so hat die OC-Kurve einen inneren Tiefpunkt.

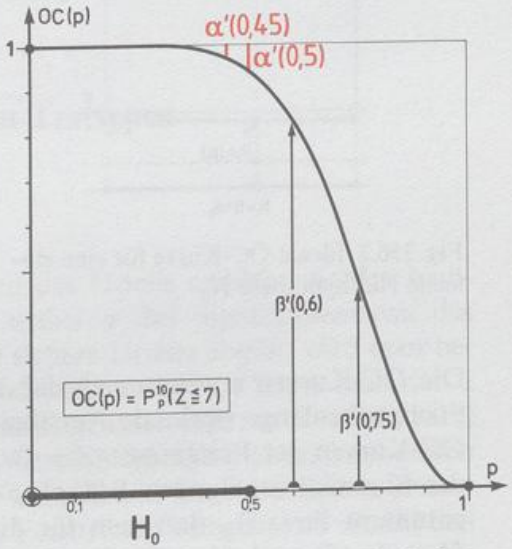


Fig. 355.1 Operationscharakteristik des Ereignisses »$Z \leq 7$« bezüglich $H = [0; 1]$. Vgl. Fig. 353.1

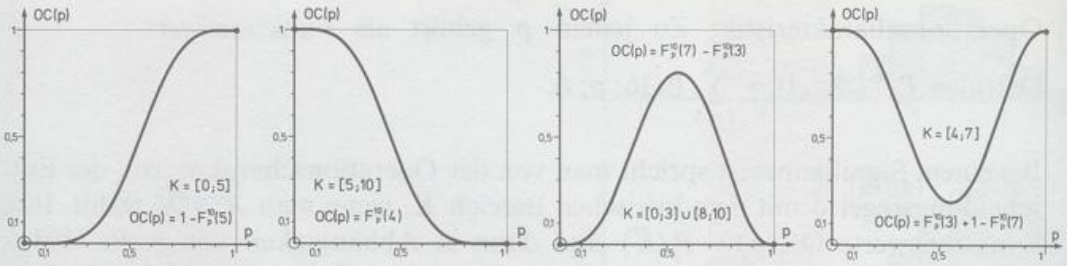


Fig. 356.1 Die 4 wichtigen Typen von OC-Kurven bezüglich $H = \{B(n;p) | p \in [0;1]\}$, veranschaulicht mittels Binomialverteilungen $B(10;p)$

Wie man sich leicht überlegt, sind diese 4 Operationscharakteristiken Polynome n-ten Grades in p. Figur 356.1 veranschaulicht sie für $n = 10$.

Die OC-Kurve gibt uns einen Hinweis auf die Güte des Tests. Je steiler sie nämlich in ihren Flanken ist, desto schneller werden die Irrtumswahrscheinlichkeiten 2. Art klein. Im Idealfall wären für jedes $p \in H_0$ die Irrtumswahrscheinlichkeit $\alpha'(p) = 0$ und für jedes $p \in H_1$ die Irrtumswahrscheinlichkeit $\beta'(p) = 0$. Dann würde man nur richtige Urteile abgeben! Die zugehörige OC-Kurve hätte über H_0 konstant den Wert 1 und über H_1 konstant den Wert 0. Figur 356.2 zeigt die ideale OC-Kurve für eine einfache Nullhypothese, Figur 356.3 für eine zusammengesetzte Nullhypothese.

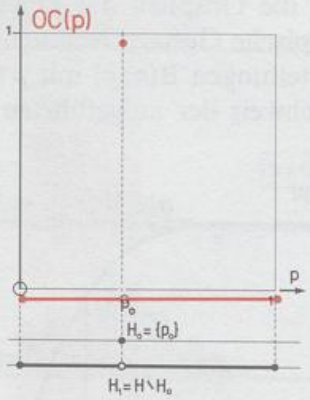


Fig. 356.2 Ideale OC-Kurve für eine einfache Nullhypothese H_0

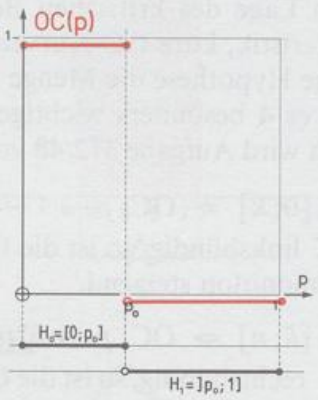


Fig. 356.3 Ideale OC-Kurve für eine zusammengesetzte Nullhypothese H_0

Die OC-Kurven erweisen sich daher als praktisches Hilfsmittel, bei gegebener Stichprobenlänge optimale Annahmebereiche zu finden. Figur 357.1 zeigt die OC-Kurven der Ereignisse »$Z = 0$«, »$Z \leqq 1$«, …, »$Z \leqq 5$« bezüglich der Schar der Binomialverteilungen $B(5;p)$, $p \in [0;1]$, als zulässiger Hypothese H. Man entnimmt ihr z. B., daß man für die Entscheidung zwischen den Hypothesen $H_0 = \{0{,}15\}$ und $H_1 = \{0{,}4\}$ am besten das Ereignis »$Z \leqq 2$« heranzieht, wenn die Wahrscheinlichkeit für einen Fehler 1. Art unter 5% liegen soll. Ohne diese Bedingung würde man sich für »$Z \leqq 1$« entscheiden, weil dann $\alpha' + \beta'$ minimal

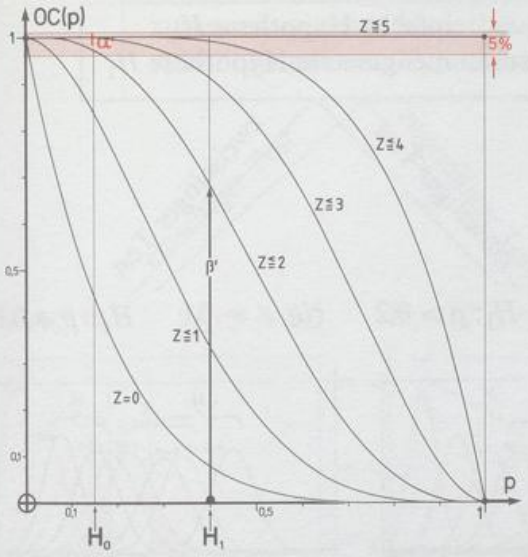


Fig. 357.1 Alternativtest für $H_0 = \{0{,}15\}$, $H_1 = \{0{,}4\}$ und $A = [0; k]$ mit $k \in \{0, 1, 2, 3, 4, 5\}$. Auswahl des optimalen Tests für die Schranke $\alpha = 5\%$

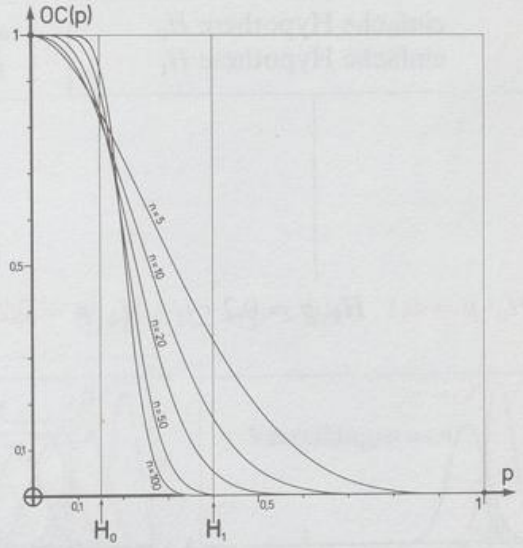


Fig. 357.2 Illustration des Einflusses der Stichprobenlänge n auf die Trennschärfe $H_0 = \{0{,}15\}$; $H_1 = \{0{,}4\}$; $A = [0; 0{,}2n]$; $n \in \{5; 10; 20; 50; 100\}$.

wird. Ein Ereignis ist desto besser für eine Entscheidungsregel geeignet, je stärker die OC-Kurve von dem einen der beiden in Frage kommenden p-Werte bis zum anderen abfällt. Andererseits läßt sich der Einfluß der Stichprobenlänge n auf die **Trennschärfe** des Tests an Hand der zugehörigen OC-Kurven beobachten (Figur 357.2). Wie erwartet fallen die OC-Kurven für größere n steiler von 1 auf 0 ab und trennen daher die Hypothesen besser. Für $n \to \infty$ hätte man einen idealen Test mit senkrecht abfallender OC-Kurve. Die Trennung ist perfekt, die Fehler haben die Wahrscheinlichkeit 0.

17.5. Überblick über die behandelten Testtypen

Siehe Seite 358 f.

17.6. Verfälschte Tests

Bei einem Signifikanztest hat die Sicherheit des Urteils »Ablehnung der Nullhypothese« mindestens den Wert $1 - \alpha$, wobei α das Signifikanzniveau des Tests ist. Da man natürlich gern möglichst sichere Urteile abgibt, wird man bestrebt sein, das Signifikanzniveau α möglichst klein zu halten. Wählt man nun α und damit auch den kritischen Bereich K sehr klein, dann muß man leider in Kauf nehmen, daß nur noch in seltenen Fällen die Nullhypothese abgelehnt werden kann; d. h., der Test wird sehr häufig kein brauchbares Ergebnis liefern. Dieser Sachverhalt könnte einen Tester nun in die Versuchung bringen, erst einmal den Ausfall der Stichprobe abzuwarten und dann den kritischen Bereich K möglichst eng um das Stichprobenergebnis herumzulegen und damit das Signifikanzniveau recht klein zu machen. Der Versuchsausgang erschiene dann in einem besonders

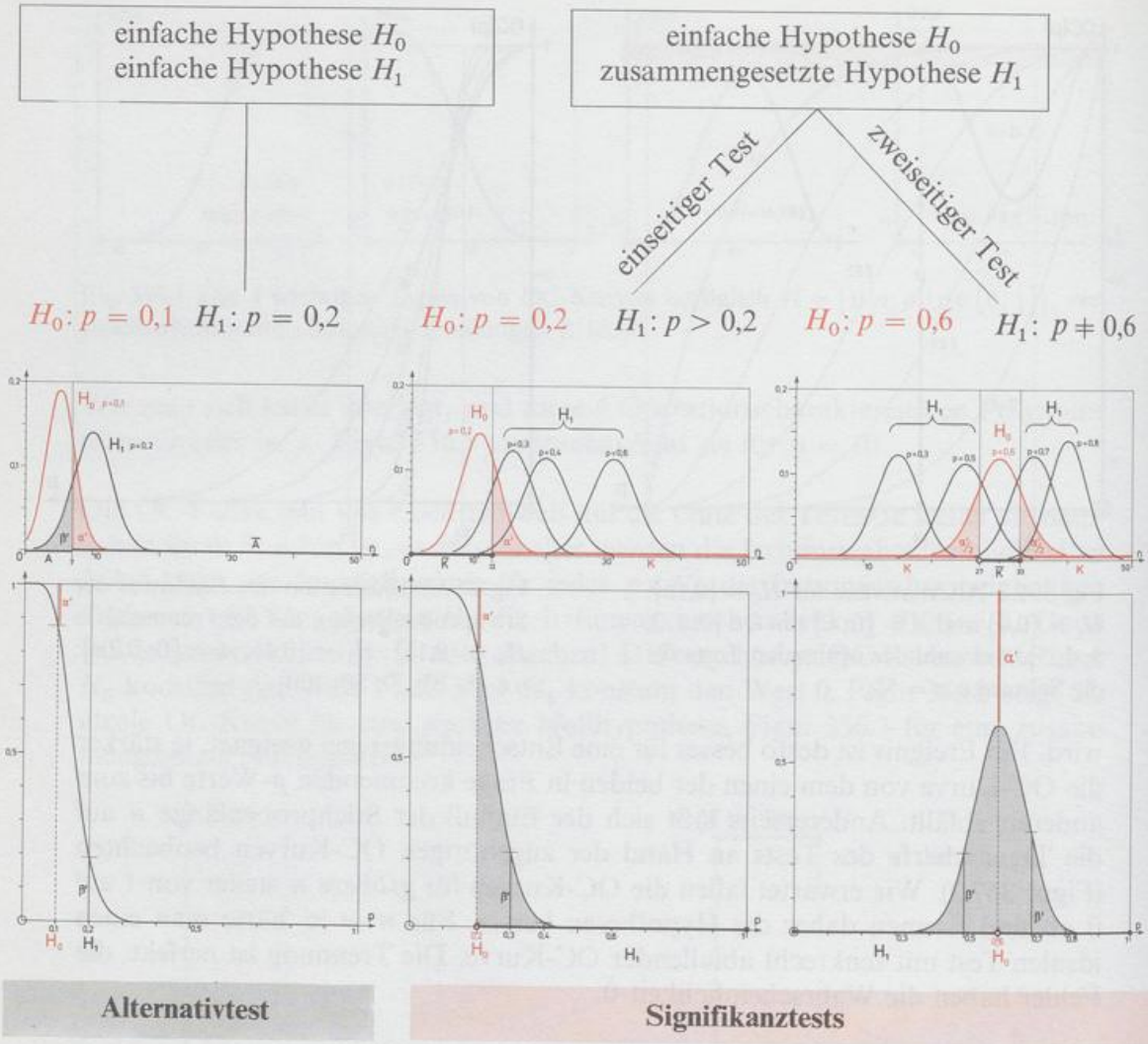
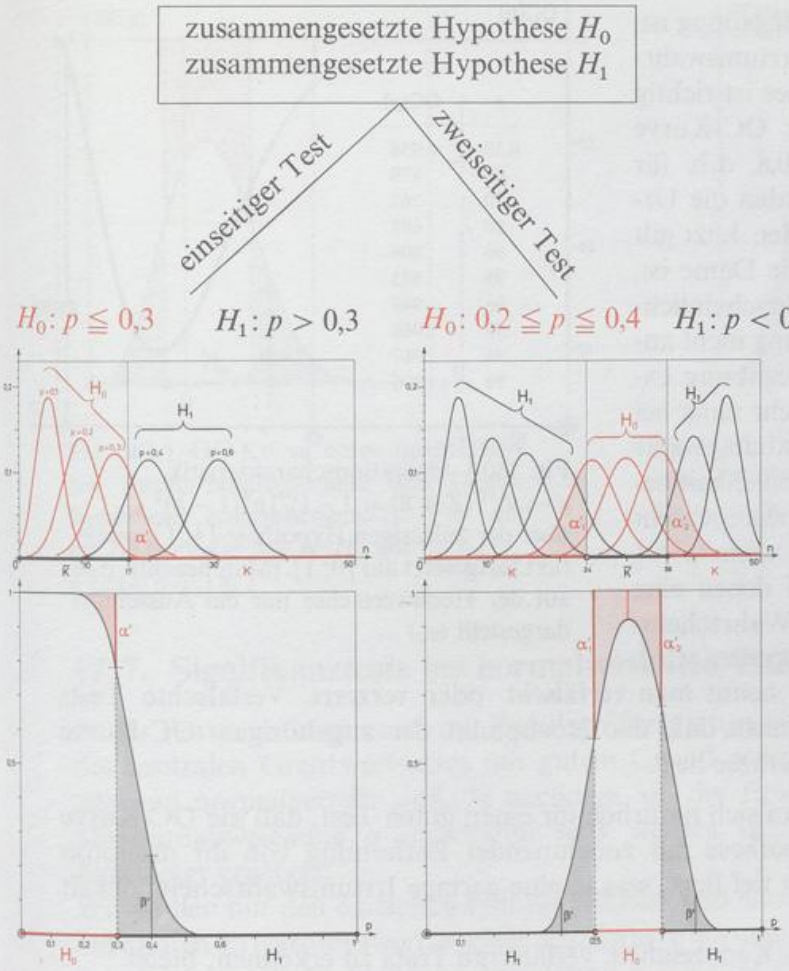


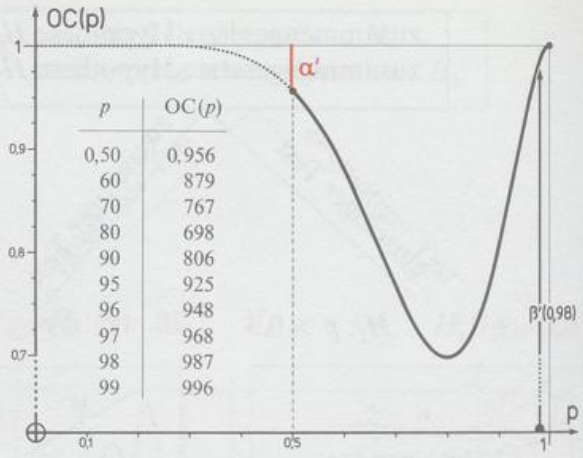
Fig. 358.1 Überblick über die behandelten Testtypen

günstigen Licht! Allerdings ist die Auswahl des Tests nach dem Ausgang der zufallsbedingten Stichprobe sehr gefährlich. Wählt man nämlich den Test nach Bedarf, also letzten Endes zufallsbestimmt, so verlieren die errechneten Wahrscheinlichkeiten jeden Sinn, und man kann sie dann auch nicht mehr als Beleg für irgendwelche Behauptungen anführen. Bei der Vielzahl verschiedener Tests, die man in einem Handbuch der mathematischen Statistik finden kann, ist natürlich die Versuchung groß, sich einen solchen auszusuchen, der irgendeine erwünschte Aussage am besten »bestätigt«. Vor einem derartigen Mißbrauch der Statistik muß daher ganz besonders gewarnt werden. Wir betrachten ein abschreckendes



Beispiel: Im Teetassentest (Seite 350) möchte Sir Y., ein leidenschaftlicher Verehrer von Lady X., ihr mit möglichst großer Sicherheit Begabung bescheinigen. Weil er weiß, daß Lady X. 8 Tassen richtig benannt hat, nimmt er einen kleinstmöglichen kritischen Bereich, bei dem er ihr noch Begabung attestieren kann, also $K = \{8\}$. Für den Fehler 1. Art ergibt sich die Wahrscheinlichkeit $\alpha' = P^{10}_{0,5}(Z = 8) \approx 4{,}4\%$. Sir Y. wird erläutern, daß nach seinem sehr scharfen Test nur in 4,4% der Fälle, in denen in Wirklichkeit $p = \frac{1}{2}$ vorliegt, auf »$p > \frac{1}{2}$« erkannt wird. Damit hat er zweifellos recht. Aber sehen wir uns die OC-Kurve des Ereignisses $\bar{K}$ an (Figur 360.1). Sie zeigt, daß für kleine Begabungen (d. h. p wenig größer als $\frac{1}{2}$) der Test durchaus brauchbar ist. Denn für solche Begabun-

gen gilt: Je größer die Begabung ist, desto kleiner ist die Irrtumswahrscheinlichkeit 2. Art. Dies ist richtig bis zum Tiefpunkt der OC-Kurve bei $p = 0{,}8$. Für $p > 0{,}8$, d. h. für große Begabungen, werden die Urteile aber immer absurder. Jetzt gilt nämlich: Je begabter die Dame ist, desto größer ist die Wahrscheinlichkeit β', ihr diese Begabung nicht anzuerkennen. Ist ihre Begabung extrem gut, d. h., liegt p sehr nahe bei 1, dann wird diese Begabung sogar mit größerer Wahrscheinlichkeit bestritten, als wenn sie überhaupt nicht vorhanden wäre, d. h., wenn $p = \frac{1}{2}$ wäre. Solche Tests, bei denen eine Hypothese mit größter Wahrscheinlichkeit dann angenommen wird, wenn sie nicht zutrifft, nennt man **verfälscht** oder **verzerrt**. Verfälschte Tests erkennt man offenbar daran, daß der Hochpunkt der zugehörigen OC-Kurve nicht über der Nullhypothese liegt.



p	OC(p)
0,50	0,956
60	879
70	767
80	698
90	806
95	925
96	948
97	968
98	987
99	996

Fig. 360.1 Operationscharakteristik $p \mapsto P_p^{10}(Z \neq 8) = 1 - \binom{10}{8} p^8 (1-p)^2$ über der zulässigen Hypothese $[\frac{1}{2}; 1]$, punktiert fortgesetzt auf $[0; 1]$. (Man beachte, daß auf der Hochwertachse nur ein Ausschnitt dargestellt ist.)

Im übrigen wünscht man sich natürlich für einen guten Test, daß die OC-Kurve außerhalb der Nullhypothese mit zunehmender Entfernung von ihr monoton abnimmt und möglichst tief liegt, was ja eine geringe Irrtumswahrscheinlichkeit 2. Art bedeutet.
Ein besonders einfaches Kennzeichen, verfälschte Tests zu erkennen, bietet

Definition 360.1: Ein Signifikanztest mit der zulässigen Hypothese H und der Nullhypothese H_0 heißt **unverfälscht**, wenn für jedes $p \in H_0$ und jedes $p_1 \in H \setminus H_0$

$$\alpha'(p) + \beta'(p_1) \leqq 1$$

gilt. Andernfalls heißt der Test **verfälscht**.

In dieser Definition kommt zum Ausdruck, daß bei einem unverfälschten Test das Maximum der Operationscharakteristik über der Nullhypothese H_0 angenommen werden muß. Die Figuren 361.1 und 361.2 veranschaulichen Definition 360.1.
Was bedeutet eigentlich die Bedingung $\alpha'(p) + \beta'(p_1) > 1$, die einen verfälschten Test kennzeichnet? Sie besagt, daß

$$\alpha'(p) > 1 - \beta'(p_1),$$

d. h.: Die Wahrscheinlichkeit, die Nullhypothese abzulehnen, falls sie zutrifft, ist größer als die Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn sie nicht zutrifft.

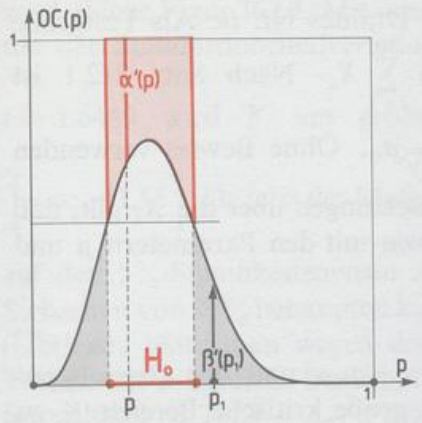


Fig. 361.1 OC-Kurve eines unverfälschten Tests. Nullhypothese und Gegenhypothese zusammengesetzt. Für alle $p \in H_0$ und alle $p_1 \in H \backslash H_0$ gilt: $\alpha'(p) + \beta'(p_1) \leqq 1$.

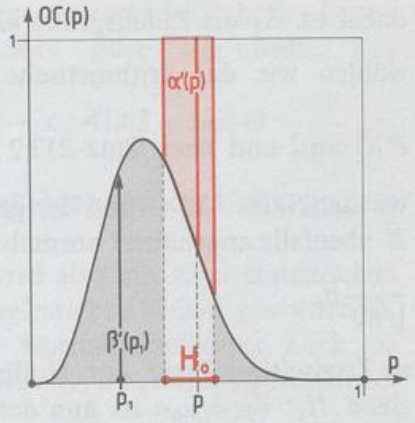


Fig. 361.2 OC-Kurve eines verfälschten Tests. Nullhypothese und Gegenhypothese zusammengesetzt. Es gibt $p \in H_0$ und $p_1 \in H \backslash H_0$, so daß $\alpha'(p) + \beta'(p_1) > 1$.

17.7. Signifikanztests bei normalverteilten Zufallsgrößen

In der Praxis hat man es oft mit Zufallsgrößen zu tun, von denen man auf Grund des zentralen Grenzwertsatzes mit gutem Grund annehmen kann, daß sie annähernd normalverteilt sind. Je nachdem, ob der Erwartungswert μ oder die Standardabweichung σ unbekannt sind, werden sich Hypothesen über diese Parameter ergeben.
Wir wollen nur den einfachen Fall besprechen, daß eine Hypothese über den unbekannten Erwartungswert μ einer angenähert normalverteilten Zufallsgröße zu testen ist, wobei die Standardabweichung σ bekannt ist. Ein solcher Test heißt ***Gaußtest.***

Beispiel: Die Untersuchung von Drähten einer bestimmten Legierung ergab für die Zugfestigkeit den Mittelwert 41,62 N/mm² und die Standardabweichung 0,60 N/mm². Wir dürfen annehmen, daß die Zufallsgröße »Zugfestigkeit eines Drahtes« angenähert normalverteilt ist mit $\mu_0 = 41{,}62\,\text{N/mm}^2$ und $\sigma_0 = 0{,}60\,\text{N/mm}^2$. Eine Versuchsserie an 80 Drähten mit einer etwas veränderten Legierung ergab eine mittlere Zugefstigkeit von 41,50 N/mm² bei gleicher Standardabweichung. Kann man auf dem 5%-Niveau die Hypothese »Die mittlere Zugfestigkeit hat sich nicht verändert« ablehnen?
Ob man einseitig oder zweiseitig testen wird, hängt davon ab, welche Alternativen man in Betracht ziehen will. Man kann sich auf den Standpunkt stellen, daß die Zugfestigkeit sowohl größer als auch kleiner geworden ist, und dann zweiseitig testen, oder man nimmt auf Grund des Stichprobenergebnisses an, daß die Zugfestigkeit höchstens kleiner geworden sein kann, und dann einseitig testen.

Lösung: Die neue Legierung besitze die mittlere Zugfestigkeit μ, die unbekannt ist. Die Zugfestigkeit eines Drahtes ist dann angenähert normalverteilt mit μ und σ_0^2. Im Zufallsexperiment wurde die Stichprobe $(X_1 | X_2 | \ldots | X_{80})$ bestimmt;

dabei ist X_i die Zufallsgröße »Zugfestigkeit des Drahtes Nr. i«. Als Testgröße wählen wir das arithmetische Mittel $\bar{X} = \frac{1}{80} \sum_{i=1}^{80} X_i$. Nach Satz 212.1 ist $\mathscr{E}\bar{X} = \mu$, und nach Satz 212.2 ist $\sigma(\bar{X}) = \frac{1}{\sqrt{80}} \sigma_0$. Ohne Beweis verwenden wir den **Satz:** Auf Grund der gemachten Voraussetzungen über die X_i gilt, daß $\bar{X}$ ebenfalls angenähert normalverteilt ist, und zwar mit den Parametern μ und $\frac{1}{\sqrt{80}} \sigma_0$.

a) Zweiseitiger Test. Zur Nullhypothese H_0: »$\mu = \mu_0$« und der Gegenhypothese H_1: »$\mu \neq \mu_0$« ist nun derjenige möglichst große kritische Bereich K zu bestimmen, so daß $\alpha' = P_{\mu_0}(\bar{X} \in K) \leqq \alpha$ wird, wobei α das vorgegebene Signifikanzniveau ist. Dabei wählen wir auf Grund der Symmetrie der Normalverteilung K so, daß $\bar{K} =]\mu_0 - t\sigma; \mu_0 + t\sigma[$ ist. Somit erhalten wir mit den vorgegebenen Werten die Bedingung $P\left(|\bar{X} - \mu_0| < t \frac{0{,}60}{\sqrt{80}}\right) \geqq 95\%$. (Siehe Figur 362.1.)
Wir entnehmen der Tabelle »σ-Bereiche bei normalverteilten Zufallsgrößen«, daß $t \geqq 1{,}96$ sein muß. Für $t = 1{,}96$ wird K am größten und $\bar{K}$ am kleinsten. Wir erhalten

$$\bar{K} = \left]41{,}62 - 1{,}96 \cdot \frac{0{,}60}{\sqrt{80}};\ 41{,}62 + 1{,}96 \cdot \frac{0{,}60}{\sqrt{80}}\right[=]41{,}49;\ 41{,}75[.$$

Da der Meßwert 41,50 nicht in den kritischen Bereich K fällt, kann man die Nullhypothese H_0 auf dem 5%-Niveau nicht ablehnen. Man wird also bei weiteren Überlegungen mit einer Sicherheit von 95% davon ausgehen, daß sich die Zugfestigkeit nicht verändert hat.

b) Einseitiger Test. Zur Nullhypothese H_0: »$\mu = \mu_0$« gehört nun die Gegenhypothese H_2: »$\mu < \mu_0$«. Der kritische Bereich K ist nun möglichst groß so zu bestimmen, daß $P_{\mu_0}(\bar{X} \in K) = P(\bar{X} \leqq \mu_0 - t\sigma) = P\left(\frac{\bar{X} - \mu_0}{\sigma} \leqq -t\right) = \Phi(-t) \leqq \alpha$

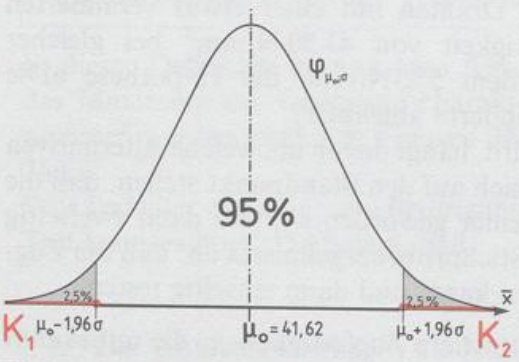


Fig. 362.1 Kritischer Bereich $K = K_1 \cup K_2$ des zweiseitigen Tests

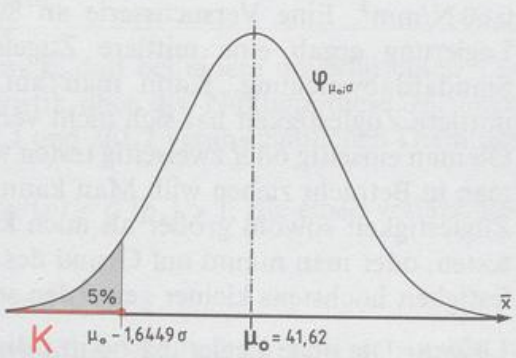


Fig. 362.2 Kritischer Bereich K des einseitigen Tests

wird. (Siehe Figur 362.2.)Mit unseren Werten erhalten wir aus der Tabelle »Quantile der Standardnormalverteilung« $-t \leqq -1{,}6449$ oder $t \geqq 1{,}6449$. Für $t = 1{,}6449$ wird K am größten, nämlich $\left]-\infty;\ 41{,}62 - 1{,}6449 \cdot \frac{0{,}60}{\sqrt{80}}\right] = \left]-\infty;\ 41{,}51\right]$. Da jetzt der Meßwert 41,50 in K liegt, kann man die Nullhypothese auf dem 5%-Signifikanzniveau ablehnen. Man wird also mit einer statistischen Sicherheit von 95% behaupten können, daß die Zugfestigkeit kleiner geworden ist. (Übrigens hätte man wegen der Symmetrie der Normalverteilungen auch die Tabelle der σ-Bereiche benützen können, wenn man die obige Bedingung umgeformt hätte zu $P(|\bar{X} - \mu_0| < t\sigma) \leqq 2\alpha = 10\%$.)

Die Ergebnisse von **a)** und **b)** scheinen sich zu widersprechen! Dem ist jedoch nicht so. Es handelt sich nämlich um Antworten auf verschiedene Fragestellungen. Bei der Hypothese H_2 wird nämlich schon berücksichtigt, daß auf Grund der physikalischen Versuchsergebnisse nur mit einer Verkleinerung der Zugfestigkeit gerechnet werden kann, während bei H_1 das physikalische Ergebnis nicht berücksichtigt wird, weil der Experimentator trotz der Verkleinerung des Mittelwerts eine Vergrößerung der Zugfestigkeit für möglich hält.

Wir merken uns: Man kann bei einem *Gauß*test bei angenähert normalverteilten Zufallsgrößen kritische Bereiche K zu vorgegebenem Signifikanzniveau α bei zweiseitigem Test der Tabelle »σ-Bereiche bei normalverteilten Zufallsgrößen«, bei einseitigem Test der Tabelle »Quantile der Standardnormalverteilung« leicht entnehmen. (Siehe *Stochastik-Tabellen*, Seite 44 und 45.)

Natürlich kann man auch zu einem *Gauß*test über den Erwartungswert μ einer Verteilung die Operationscharakteristik bestimmen. Wir erhalten für unser

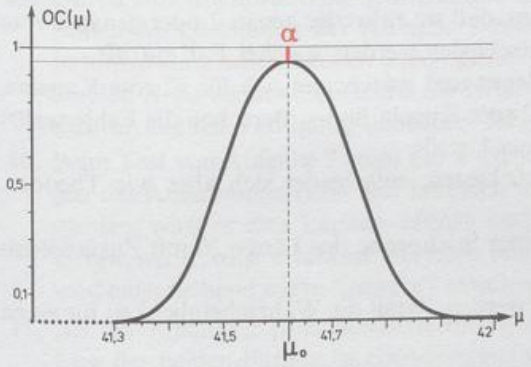


Fig. 363.1 Operationscharakteristik des Ereignisses $\bar{X} \in]41{,}49;\ 41{,}75[$ bezüglich $H = \mathbb{R}$

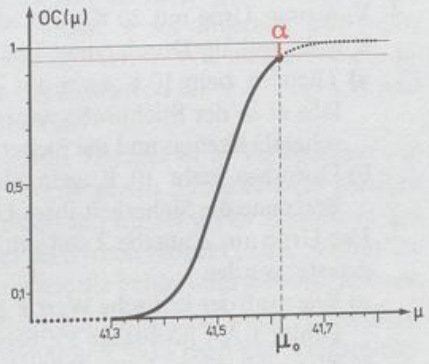


Fig. 363.2 Operationscharakteristik des Ereignisses $\bar{X} \in]41{,}51;\ \mu_0]$ bezüglich $H =]-\infty; \mu_0]$. Die Fortsetzung auf $H = \mathbb{R}$ ist punktiert gezeichnet.

Beispiel im

Fall a): OC: $\mu \mapsto P_\mu(\bar{X} \in \bar{K}) = \Phi\left(\frac{41{,}75-\mu}{\frac{0{,}60}{\sqrt{80}}}\right) - \Phi\left(\frac{41{,}49-\mu}{\frac{0{,}60}{\sqrt{80}}}\right), \quad \mu \in \mathbb{R}$

Fall b): OC: $\mu \mapsto P_\mu(\bar{X} \in \bar{K}) = 1 - \Phi\left(\frac{41{,}51-\mu}{\frac{0{,}60}{\sqrt{80}}}\right), \quad \mu \leqq \mu_0 .$

Die Figuren 363.1 und 363.2 zeigen die zugehörigen OC-Kurven.

Aufgaben

Zu 17.1.

Joseph Bertrand (1822–1900) wendet sich in seinem *Calcul des Probabilités* (1889) gegen den Begriff des *homme moyen* von *Quetelet*. Um zu zeigen, daß es keinen Menschen mit gleichzeitig durchschnittlicher Höhe, durchschnittlichem Gewicht usw. geben kann, betrachtet er 2 Kugeln aus gleichem Material vom Radius 1 bzw. Radius 3. Die Durchschnittskugel hat dann den Radius 2. Welche durchschnittliche Oberfläche, welches durchschnittliche Gewicht würde sie besitzen? Sind das ihre wirklichen Größen?

Zu 17.2.

In Bild 335.1 ist von einer Stichprobe die Rede. Das Titelbild zu Kapitel 18 (Seite 375) zeigt das Ergebnis einer Stichprobe aus dieser Stichprobe. Aus welcher Zufallsgröße X wurde die Stichprobe gezogen? Welche Länge hat sie? Was bedeuten die X_i? Sind sie unabhängig?

Zu 17.3.

1. Erläutere die Begriffe »Fehler 1. Art« und »Fehler 2. Art« an folgender Zeitungsüberschrift: »Sheriff hält Schnupftabak für Rauschgift«.
2. Von einer Urne mit 20 Kugeln ist bekannt, daß sie entweder genau 2 oder genau 6 rote Kugeln enthält. Durch einen Test soll entschieden werden, welcher Fall zutrifft.
   a) Theodor zieht 10 Kugeln mit Zurücklegen und entscheidet sich für »2 rote Kugeln«, falls er in der Stichprobe weniger als 2 rote Kugeln findet. Berechne die Fehlerwahrscheinlichkeiten und die Sicherheit seines Urteils.
   b) Dorothea zieht 10 Kugeln ohne Zurücklegen, entscheidet sich aber wie Theodor. Berechne die Sicherheit ihres Urteils.
3. Die Urne aus Aufgabe **2** soll mit Hilfe einer Stichprobe der Länge 20 mit Zurücklegen getestet werden.
   a) Wie muß der kritische Wert k gewählt werden, damit die Wahrscheinlichkeit für einen Fehler 1. Art höchstens 5% beträgt?
   b) Wie muß der kritische Wert gewählt werden, damit die Wahrscheinlichkeit für einen Fehler 2. Art höchstens 1% beträgt?
   c) Wie muß der kritische Wert gewählt werden, damit die Summe der beiden Fehlerwahrscheinlichkeiten minimal wird?
   d) Veranschauliche die Aufgaben in einem α'-β'-Diagramm ($1 \triangleq 10\,\text{cm}$).

4. **a)** Bestimme mit Hilfe der *Stochastik-Tabellen* für die Urne aus Aufgabe **2** eine möglichst kleine Stichprobenlänge n und einen dazu passenden kritischen Wert k so, daß die Wahrscheinlichkeit für einen Fehler 1. Art höchstens 5% und die für einen Fehler 2. Art höchstens 10% wird.

●**b)** Der in **a)** gefundene Wert für n ist vermutlich zu groß. Bestimme daher mit Hilfe der Normalverteilung ein besseres n und ein dazugehöriges k. Überprüfe, falls du Zugang zu einem programmierbaren Rechner besitzt, ob die so gefundenen Werte für n und k tatsächlich die gestellten Bedingungen erfüllen.

5. Die Wahrscheinlichkeit, daß die Urne aus Aufgabe **2** tatsächlich genau 2 rote Kugeln enthält, sei 10%. Wer diese Urne richtig erkennt, erhält eine Belohnung von 100 DM. Liegt aber die andere Urne vor und erkennt man diese, so erhält man 10 DM. Bestimme für eine Stichprobe der Länge 25 mit Zurücklegen einen kritischen Wert, so daß der Erwartungswert der Belohnung maximal wird.

6. **a)** Der Urne der Aufgabe **2** werde eine Stichprobe der Länge 10 mit Zurücklegen entnommen. Zeichne ein α'-β'-Diagramm zu den Annahmebereichen

$$A(k) := »Z \leqq k« \text{ für } k \in \{-1, 0{,}1, \ldots, 10\};\ 1 \mathrel{\hat{=}} 10\,\text{cm}.$$

b) Zeichne in das Diagramm von **a)** die Geraden $\alpha' = 10\%$ und $\beta' = 20\%$ ein und suche diejenigen Tests, für die

1) $\alpha' \leqq 10\%$, **2)** $\beta' \leqq 20\%$, **3)** $\alpha' \leqq 10\%$ und $\beta' \leqq 20\%$ gilt.

c) Bestimme aus dem Diagramm von **a)** denjenigen Test, für den die Summe der Fehlerwahrscheinlichkeiten minimal wird.

d) Bestimme aus dem Diagramm von **a)** denjenigen Test, für den $5\alpha' + 3\beta'$ minimal wird.

e) Zeichne in das α'-β'-Diagramm von **a)** denjenigen Punkt ein, der zum Annahmebereich $A := \{1, 2, 3\}$ gehört.

7. Bei einer Urne soll ermittelt werden, ob sie 6 rote und 4 grüne Kugeln oder umgekehrt 4 rote und 6 grüne Kugeln enthält. Es ist eine Stichprobe der Länge 5 mit Zurücklegen erlaubt.

a) Welches Entscheidungsverfahren erscheint als einziges vernünftig? Zu welchen Irrtumswahrscheinlichkeiten führt es?

b) 10 Personen haben den Test gemäß **a)** ausgeführt, und 6 haben falsch geurteilt. Kann man diese Abweichung vom »Ideal« noch als zufällig bezeichnen?

8. Jemand wählt beim Problem der vorigen Aufgabe das folgende Entscheidungsverfahren: Entscheidung für »6 rote Kugeln in der Urne« genau dann, wenn die ersten 3 gezogenen Kugeln rot sind. Berechne die Irrtumswahrscheinlichkeiten.

●9. Durch eine Stichprobe mit Zurücklegen der Länge n soll bei einer Urne zwischen den beiden Möglichkeiten der Aufgabe **7** entschieden werden. n sei ungerade, und beide Irrtumswahrscheinlichkeiten sollen gleich groß gemacht werden. Wie ist der Annahmebereich A zu wählen? Berechne die Irrtumswahrscheinlichkeiten für $n = 1, 3, \ldots$, soweit dies mit der zur Verfügung stehenden Tabelle möglich ist.

10. Beim Test von Aufgabe **7** seien nur 4 Ziehungen erlaubt. Jemand ist in Verlegenheit wegen des Annahmebereichs und hilft sich wie folgt: Wenn genau 2 rote Kugeln gezogen werden, wirft er eine Laplace-Münze und entscheidet sich für »4 grüne Kugeln in der Urne«, wenn Adler erscheint. Bei mehr oder weniger als 2 roten Kugeln in der Stichprobe wird entsprechend wie in Aufgabe **7** entschieden. Berechne die Fehlerwahrscheinlichkeiten.

11. Bei der Züchtung einer gewissen Blumensorte erhält man rote und weiße Exemplare. Eine der beiden Farben ist ein »dominantes« Merkmal und muß nach den Vererbungsgesetzen mit der Wahrscheinlichkeit $\frac{3}{4}$ auftreten. In einem Kreuzungsversuch ergeben sich 15 Nachkommen. Mit welcher Wahrscheinlichkeit irrt man sich, wenn man die dabei häufiger auftretende Farbe für dominant hält?

12. Aus einer Urne mit 7 Kugeln werden 3 Stück *ohne* Zurücklegen entnommen. Nach der

Zahl Z schwarzer Kugeln in dieser Stichprobe wird entschieden, ob in der Urne 2 oder 4 schwarze Kugeln sind.

a) Ermittle die beiden Wahrscheinlichkeitsverteilungen und stelle sie graphisch dar.

●**b)** Suche unter allen denkbaren Annahmebereichen für die Hypothese »2 schwarze Kugeln«, d. h. unter allen Teilmengen von $\{0, 1, 2, 3\}$, denjenigen aus, bei dem die Summe der Fehlerwahrscheinlichkeiten am kleinsten ist.

13. An eine Werkstatt werden Schachteln mit Schrauben geliefert. Ein Teil davon enthält Erste Qualität, das sind Schrauben, von denen nur 10% 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. Man entnimmt jeder Schachtel mit Zurücklegen 5 Schrauben. Sind alle Schrauben bis auf höchstens eine in Ordnung, so soll der Schachtelinhalt als Erste Qualität behandelt werden, andernfalls als Zweite Qualität. Bestimme die beiden Fehlerwahrscheinlichkeiten.

14. Für das Entscheidungsverfahren in Aufgabe **13** macht ein Mitarbeiter der Werkstatt folgenden Vorschlag: Es werden nacheinander Schrauben aus der gewählten Schachtel geprüft. Sind die ersten 3 Stück in Ordnung, so entscheidet man »$p = 0{,}1$«, andernfalls »$p = 0{,}4$«.

a) Wie groß sind die Fehlerwahrscheinlichkeiten α' und β' dieses Tests?

●**b)** Zeichne das α'-β'-Diagramm für die Regeln »Entscheidung für $p = 0{,}1$, wenn die ersten k Schrauben in Ordnung sind«, $k = 1, 2, \ldots, 5$.

●**c)** Trage in das Diagramm von **b)** auch die Regeln »Entscheidung für $p = 0{,}4$ genau dann, wenn die ersten k Schrauben Ausschuß sind« ein ($k = 1, 2, 3$).

15. In einer Schießbude gibt es sehr gute und mittelmäßige Gewehre (Treffwahrscheinlichkeiten 0,9 bzw. 0,7). Weil bei einem davon die geheime Kennzeichnung unleserlich geworden ist, macht der Besitzer mit ihm 20 Probeschüsse. Er weiß, daß ihm der Fehler, ein schlechtes Gewehr fälschlich für ein gutes zu halten, mehr Schaden bringt als der umgekehrte Irrtum (Verärgerung anspruchsvoller Kunden!). Er möchte daher die Wahrscheinlichkeit für diesen Fehler höchstens halb so groß machen wie die für den zweiten Fehler. Welche Entscheidungsregel muß er aufstellen?

16. Bei einer Prüfung werden n Fragen gestellt. Wir nehmen an, daß ein Prüfling alle Fragen unabhängig voneinander je mit der Wahrscheinlichkeit p richtig bearbeitet. Die geforderte Mindestzahl richtiger Antworten soll nun so gewählt werden, daß ein sehr gut vorbereiteter Prüfling ($p = 97\%$) mit einer Wahrscheinlichkeit von mindestens 97,5% die Prüfung besteht, ein schlecht vorbereiteter ($p = 75\%$) aber mit mindestens 90% Sicherheit durchfällt. Zeige, daß diese Bedingungen bei $n = 15$ nicht, bei $n = 20$ und $n = 50$ jedoch erfüllt werden können, und gib jeweils die möglichen Grenzen zwischen »bestanden« und »nicht bestanden« an.

17. Zu einem Ergebnisraum von 6 Elementen sind zwei Wahrscheinlichkeitsverteilungen gegeben:

ω	1	2	3	4	5	6
$P_1(\{\omega\})$	0,1	0,2	0	0,3	0,3	0,1
$P_2(\{\omega\})$	0,4	0,15	0,3	0,05	0,1	0

a) Stelle P_1 und P_2 analog zu Figur 342.1 graphisch dar. Jemand wählt als Annahmebereich für P_1 das Ereignis $\{4; 5\}$. Mit welchen Wahrscheinlichkeiten sind seine Urteile richtig?

b) Wähle einen Annahmebereich A für die Hypothese »P_1 liegt vor« so, daß das Vorliegen von P_1 mit 80% Sicherheit und das Vorliegen von P_2 mit möglichst großer Sicherheit erkannt wird.

●**c)** Ein Statistiker konstruiert den Annahmebereich A für P_1 nach folgendem Prinzip: $\omega \in A \Leftrightarrow P_1(\{\omega\}) > P_2(\{\omega\})$.
Welches A und welche Irrtumswahrscheinlichkeiten erhält er?

●**d)** Begründe, daß man nach dem Prinzip der Aufgabe **c)** den Test mit der kleinstmöglichen Summe $\alpha' + \beta'$ der Irrtumswahrscheinlichkeiten erhält! Zeige, daß der in Figur 342.1 dargestellte Test nicht das Minimum von $\alpha' + \beta'$ erreicht.

18. Ein Glücksspieler besitzt einen Laplace-Würfel und einen Würfel, bei dem die Sechs mit der Wahrscheinlichkeit 20% erscheint. Bei einer Razzia testet die Polizei die äußerlich ununterscheidbaren Würfel. Sie entscheidet sich nach 600 Würfen für die Hypothese »Laplace-Würfel«, falls höchstens 110 Sechser fallen.

a) Berechne die Fehlerwahrscheinlichkeiten mit Hilfe der Normalverteilung.

b) Bestimme mit Hilfe der Normalverteilung einen möglichst kleinen kritischen Wert k so, daß die Wahrscheinlichkeit für einen Fehler 1. Art höchstens 5% wird. Wie groß ist dann die Wahrscheinlichkeit für einen Fehler 2. Art?

●**c)** Bestimme mit Hilfe der Normalverteilung eine möglichst kleine Wurfzahl n und einen dazu passenden möglichst kleinen kritischen Wert k, so daß die beiden Fehlerwahrscheinlichkeiten jeweils unter 1% liegen.

19. Wie groß sind die Fehlerwahrscheinlichkeiten für einen Test zu Beispiel 1 (Seite 336) bei der Stichprobenlänge $n = 300$ und dem kritischen Wert $k = 82$

a) mit der *Tschebyschow*-Ungleichung abgeschätzt,

b) mit der Normalverteilung näherungsweise berechnet?

20. Bei dem einen von zwei Spielautomaten ist die Gewinnwahrscheinlichkeit auf 0,49 eingestellt, bei dem anderen versehentlich auf 0,51. Wie oft muß man spielen, bis man mit mindestens 90% Sicherheit den für den Spieler günstigeren Automaten benennen kann? Die Entscheidung wird danach getroffen, ob man mehr als die Hälfte der Spiele gewinnt oder nicht. Man verwende die Normalverteilung als Näherung für die Binomialverteilung.

21. Eine Lieferung besteht aus 70 Schachteln mit Schrauben Erster Qualität (10% Ausschuß) und 30 Schachteln mit Schrauben Zweiter Qualität (40% Ausschuß). Durch eine Stichprobe von 20 Stück soll bei einer beliebig ausgewählten Schachtel entschieden werden, zu welcher Sorte sie gehört. Urteilt man richtig, so entsteht kein Verlust. Werden die besseren Schrauben irrtümlich dort verwendet, wo es auch die schlechteren getan hätten, so verliert man pro Schachtel 10 DM wegen des höheren Preises der guten Ware. Werden umgekehrt schlechtere Schrauben dort verwendet, wo man gute braucht, so entstehen pro Schachtel 5 DM Kosten für die Nachbearbeitung von Werkstücken.
Wie muß der Annahmebereich für die Hypothese »Es liegt Erste Qualität vor« gewählt werden, um den mittleren Schaden pro Schachtel möglichst gering zu halten?

22. Berechne zu Aufgabe **21** die mittleren Kosten für die Stichprobenlängen 1 und 2 und alle jeweils möglichen Annahmegrenzen k.

23. Eine Lieferung enthält 70 Schachteln mit 10% Ausschuß und 30 Schachteln mit 40% Ausschuß. Eine gute Schachtel für schlecht zu halten kostet 10 DM, der entgegengesetzte Fehler kostet 5 DM.

a) Der Abnehmer verzichtet ganz auf den Test, weil die Prüfkosten zu hoch sind. Er wirft statt dessen vor der Verwendung einer Schachtel eine Münze und entscheidet für »schlechte Schachtel«, wenn der Adler oben liegt. Wie groß ist der mittlere Verlust pro Schachtel, wenn bei der Münze $P(\text{»Adler«}) = \gamma$ ist? Man optimiere dieses Entscheidungsverfahren durch geeignete Wahl von γ.

b) Von welchem Preis pro Prüfung an ist das Verfahren **a)** auf jeden Fall sparsamer als irgendein Prüfverfahren?

24. Eine Warenlieferung von 100 Stück enthält den Ausschußanteil p. Nimmt man die Lieferung an, so bringt jedes unbrauchbare Stück 0,5 DM Verlust. Lehnt man sie ab, so ent-

stehen 20 DM Spesen für die Rücksendung und für jedes brauchbare Stück noch 0,3 DM weitere Kosten.

a) Es werden alle Stücke geprüft. p ist also bekannt. Für welche p-Werte wird man die Lieferung annehmen bzw. ablehnen? Man zeichne den Verlust als Funktion von p für Annahme bzw. Ablehnung (RW: $1 \triangleq 10\,\text{cm}$; HW: $5\,\text{DM} \triangleq 1\,\text{cm}$).
Welcher p-Wert ist für den Abnehmer am ungünstigsten? Wie groß wäre in diesem Fall der Verlust?

b) Die Totalprüfung dauert zu lange. Daher werden nur 15 Stücke geprüft (Stichprobe mit Zurücklegen). Die Entscheidungsregel lautet: Annahme der Lieferung, wenn höchstens k Stücke schlecht sind, sonst Ablehnung. Berechne den Erwartungswert $\mathscr{E}V$ der Zufallsgröße Verlust V allgemein. Zeichne $\mathscr{E}V$ für $k = 5$ und $k = 11$ in Abhängigkeit von p in das Bild von **a)** ein. Bestimme graphisch das Maximum von $\mathscr{E}V$. (Durch geeignete Wahl von k kann man dieses Maximum möglichst klein machen – sog. **Minimax-Verfahren**.)

c) Begründe, warum die Kurve für $\mathscr{E}V$ stets zwischen den in **a)** gezeichneten Verlustkurven der unbedingten Annahme bzw. Ablehnung liegt.

25. Eine Lieferung von 100 Elektrogeräten enthalte d defekte Stücke. Über die Annahme entscheidet folgende Regel: Sobald im Laufe der Prüfung zwei gute Stücke aufgetreten sind – Annahme; sobald zwei schlechte Stücke aufgetreten sind – Ablehnung.
Ein Test dieser Art heißt **Sequentialtest** oder **Folgetest***.
Wie viele Stücke müssen höchstens geprüft werden? Berechne und zeichne die Ablehnungswahrscheinlichkeit und den Erwartungswert der Zufallsgröße Stichprobenlänge N in Abhängigkeit von d für eine Stichprobe mit Zurücklegen.

26. Löse die vorhergehende Aufgabe für Stichproben *ohne* Zurücklegen.

27. Ein Elektrohändler wendet bei allen Lieferungen, die er erhält, den Test von Aufgabe **25** an. Die Lieferungen enthalten 10% oder 30% Ausschuß, je mit der Wahrscheinlichkeit $\frac{1}{2}$. Bei Ablehnung einer guten Lieferung entstehen 200 DM, bei Annahme einer schlechten Lieferung 100 DM Schaden.

a) Wie hoch ist der mittlere Schaden infolge von Irrtümern beim Testen?

b) Auch das Prüfen der Geräte kommt teuer. Wie hoch darf der Preis für die Prüfung eines Geräts höchstens sein, damit sich das Testen überhaupt lohnt und der Händler nicht besser daran ist, die Lieferungen ungeprüft anzunehmen?

Zu 17.4.

28. Ein Würfel soll getestet werden, ob er die Sechs mit der Wahrscheinlichkeit $\frac{1}{6}$ bringt. Dazu wird er 30mal geworfen und die Anzahl der Sechser als Testgröße gewählt. Kritischer Bereich sei die Menge $K := [0; 2] \cup [8; 30]$.

a) Formuliere die zulässige Hypothese, die Nullhypothese und die Entscheidungsregel.

b) Berechne die Irrtumswahrscheinlichkeit 1. Art.

c) Berechne die Irrtumswahrscheinlichkeiten 2. Art, wenn der Würfel die Sechs tatsächlich mit 15% bzw. 20% Wahrscheinlichkeit bringt.

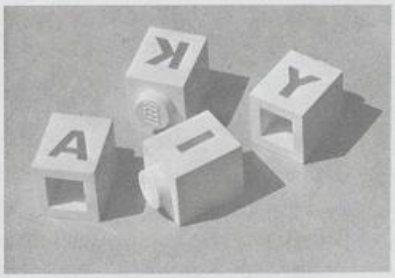
d) Fasse die 1200 Würfe von Tabelle 10.1 als 40 Tests auf und gib jedesmal das Urteil an. (Tabelle 32.1 erleichtert die Arbeit!)

e) Bestimme einen möglichst großen kritischen Bereich zum Signifikanzniveau 10%. Wie lauten nun die Urteile über den Würfel von Tabelle 10.1, wenn man wie in **d)** vorgeht?

29. a) Es gibt Lego-Steinchen, die auf einer von 4 gleichberechtigten Seitenflächen einen Buchstaben tragen. Sie bleiben immer auf einer dieser Seiten liegen. 50 Steinchen sind

* Sequentialtests gehören zu den modernsten statistischen Verfahren. Vor allem in der industriellen Qualitätskontrolle und in der Medizin haben sie große Bedeutung. Die USA hüteten sie während des 2. Weltkriegs, als *Abraham Wald* (1902–1950) sie entwickelt hatte, als wichtiges militärisches Geheimnis.

auf eine solche Fläche gefallen; 15 haben den Buchstaben oben liegen. Ist die Annahme der Symmetrie gerechtfertigt? Entscheide auf dem Signifikanzniveau 10%.



b) Was ergibt sich, wenn 500 Steinchen auf eine solche Fläche fallen und bei 150 der Buchstabe oben liegt? (Rechne mit Hilfe der Normalverteilung!)

30. a) Entwirf einen Test der Nullhypothese »Eine Münze ist symmetrisch«, der 50 (100, 200) Münzenwürfe benützt und ein Signifikanzniveau von 10% hat. Welche Wahrscheinlichkeit hat jeweils ein Fehler 2. Art bei einer Münze mit $P(\text{»Adler«}) = 0{,}6$? Zu welchen Entscheidungen führen die 3 Tests bei Tabelle 11.1, aufgefaßt als sechzehn 50fach-Würfe bzw. acht 100fach-Würfe bzw. vier 200fach-Würfe?

b) Als *Buffon* (1707–1788) das Petersburger Problem experimentell untersuchte, erhielt er bei 4040 Würfeln 2048mal Adler. (Vgl. Aufgabe 226/**22.b**.) *Poisson* (1781–1840) prüfte die Vermutung, daß *Buffons* Münze »Adler« mit größerer Wahrscheinlichkeit produziert hatte als »Zahl«. Auf welchem Signifikanzniveau konnte er die Hypothese »*Buffons* Münze war symmetrisch« ablehnen? Mit welcher Wahrscheinlichkeit hätte er die Münze für eine Laplace-Münze gehalten, obwohl sie »Adler« mit $p = 0{,}52$ brachte?

31. *Laplace* (1749–1827) behandelte 1780 in seinem *Mémoire sur les probabilités* die Frage, ob die Wahrscheinlichkeiten für eine Knaben- bzw. eine Mädchengeburt gleich groß sind. Als Material verwendete er

1) das Geburtsregister von Paris für die Jahre 1745–1770, das 251527 Knaben- und 241945 Mädchengeburten auswies,
2) das Geburtsregister von London für die Jahre 1664–1757, das 737629 Knaben- und 698958 Mädchengeburten auswies.*

a) Es sei $p := P(\text{»Knabengeburt«})$. Zeige, daß man in beiden Fällen die Hypothese $\text{»}p = \frac{1}{2}\text{«}$ praktisch auf jedem Signifikanzniveau ablehnen kann.

b) Langjährige statistische Beobachtungen legen für p den Wert 0,514 nahe. Untersuche, ob auf dem 10%-Niveau bzw. auf dem 5%-Niveau die Hypothese $\text{»}p = 0{,}514\text{«}$ mit den obigen Daten abgelehnt werden kann. Führe sowohl einen einseitigen wie auch einen zweiseitigen Test durch.

32. Bei der Untersuchung der Frage, ob Knabengeburten häufiger sind als Mädchengeburten, kann man nach *John Arbuthnot* (1667–1735) folgendermaßen vorgehen. Man vergleicht über einen längeren Zeitraum hinweg die jährliche Anzahl der Knabengeburten mit der der Mädchengeburten. Wäre die Wahrscheinlichkeit für die Geburt eines Knaben genausogroß wie die für die Geburt eines Mädchens, so müßte es auf lange Sicht gleich viele Jahre mit mehr Knaben wie Jahre mit mehr Mädchen geben. (Die Wahrscheinlichkeit, daß in einem Jahr genau gleich viel Knaben wie Mädchen auf die Welt kommen, ist praktisch Null.) Wir wählen somit als Nullhypothese $\text{»}P(\text{»Pro Jahr werden mehr Knaben als Mädchen geboren«}) = \frac{1}{2}\text{«}$. Bezeichnen wir diese Wahrscheinlichkeit mit p, so lautet die Gegenhypothese $\text{»}p > \frac{1}{2}\text{«}$. Gib einen kritischen Bereich für einen Test dieser

* In Paris begann man erst 1745 damit, die Taufregister getrennt nach Geschlechtern zu führen. – Den geringeren Anteil an Knaben in Paris gegenüber London (und auch Neapel und Petersburg) konnte *Laplace* in seinem *Essai philosophique sur les Probabilités* (1814) klären: In den Archiven des »Hospice des Enfants-Trouvés« wurden für die Jahre 1745 bis 1809 als Findelkinder 163499 Knaben und 159405 Mädchen registriert. Der Anteil der Knabengeburten war also noch kleiner als der für Paris. *Laplace* schloß daraus, daß die Bevölkerung der Umgebung mehr Mädchen als Knaben in Paris aussetzte. Nach Bereinigung der Pariser Zahlen durch die Findelkinder ergab sich schließlich für Paris derselbe Anteil von Knaben wie in den anderen Städten.

Nullhypothese auf dem $1‰$-Niveau an, wenn man über 50 Jahre hinweg die Geburten verfolgt. Was besagt das für *Arbuthnots* Folgerung,

»that it is Art, not Chance, that governs«,

aus seiner Feststellung, daß in den 82 Jahren von 1629 bis 1710 in London stets mehr Knaben als Mädchen zur Welt kamen?*

33. **a)** Theodor stellt nach 100 Würfen seiner 4 Astragali fest, daß 6mal »Aphrodite« erschienen ist. Die Erfahrungswahrscheinlichkeit für einen Aphrodite-Wurf ist 0,03. Kann er auf dem Signifikanzniveau von 5% annehmen, daß unter seinen Astragalen mindestens ein präparierter ist?**

b) Theodor hat einen Astragalus im Verdacht, daß die Seite mit dem Wert 6 mit einer Wahrscheinlichkeit fällt, die

1) von 7% verschieden ist, **2)** größer als 7% ist, **3)** kleiner als 7% ist.

Konstruiere jeweils einen Signifikanztest, damit Theodor mit 500 Würfen eine Entscheidung auf dem 1%-Niveau herbeiführen kann. Rechne sowohl mit der Normal- als auch mit der **Poisson*-Verteilung.

34. Lady X. konnte auf dem 5%-Niveau keine Begabung attestiert werden (Seite 352). Daraufhin wird der Test abgeändert: Lady X. bekommt 20 Tassen vorgesetzt. Sie beurteilt 16 davon richtig. Auf welchem Signifikanzniveau kann ihr nun eine Begabung attestiert werden?

35. Bei einer Prüfung werden einem Schüler 20 Aufgaben gestellt. Zu jeder Aufgabe werden 4 Lösungen angeboten, von denen genau eine richtig ist.

a) Angenommen, man wendet folgenden Notenschlüssel an:

Zahl der richtig angekreuzten Antworten	0 1 2 3 4	5 6 7 8	9 10 11	12 13 14	15 16 17	18 19 20
Note	6	5	4	3	2	1

Mit welcher Wahrscheinlichkeit erhält dann ein Schüler, der sich völlig aufs Raten verlegt, die Note 1 (2, ..., 6)?

b) Von welcher Anzahl richtig gelöster Aufgaben an können wir die Hypothese »Der Schüler rät blindlings« verwerfen, wenn wir höchstens 5% Wahrscheinlichkeit dafür riskieren wollen, daß wir ihm irrtümlich Wissen bescheinigen?

36. Um zu prüfen, ob ein eben ausgeschlüpftes Küken Formen unterscheiden kann, legt man ihm »Körner« aus Papier vor. Es sind zur Hälfte Dreiecke, zur Hälfte Kreise mit gleicher Fläche wie die Dreiecke. Man läßt es 20mal picken. Ergebnis: 10111001111011011111 (Dreieck 0, Kreis 1). Das Ergebnis scheint für angeborenen Formensinn zu sprechen. Welche Wahrscheinlichkeit hätte dieses oder ein noch »besseres« Ergebnis unter der Voraussetzung, daß das Küken keine Formen unterscheiden kann? – Gleiche Frage für das Ergebnis 111011101.

37. Vor der Wahl zum 10. Deutschen Bundestag am 6. März 1983 behauptete das »Institut für Demoskopie Allensbach« auf Grund einer Umfrage unter 2000 Bürgern, daß die Unionsparteien 47,0%, die Grünen 6,5% der abgegebenen Stimmen erhalten werden.

Theodor ist der Meinung, daß beide Schätzungen nicht zutreffen; Dorothea hingegen meint, daß beide Prozentzahlen zu hoch seien. Sie starten daher eine neue Umfrage unter 2000 Bürgern. Welche kritischen Bereiche müssen sie für das 5%-Signifikanz-Niveau verwenden?

* In heutiger Sprechweise wählte *Arbuthnot* als kritischen Bereich $K = \{82\}$ und berechnete $\alpha' = P_{0,5}^{82}(\{82\}) = 2^{-82}$ zu 1 : 4836000000000000000000000.

** Von solchen mit Blei beschwerten Astragali berichtet *Aristoteles* (384–322) in *Problemata*, XVI.

38. Zwei verschiedene Düngemittel X und Y sollen verglichen werden. 20 Versuchsfelder werden je zur Hälfte mit X und Y gedüngt. Auf 13 Feldern bringt X einen größeren Ertrag als Y, auf den übrigen ist es umgekehrt. Da weitere Anhaltspunkte fehlen, ist eine plausible Nullhypothese: »X und Y sind gleich wirksam«. Wir nehmen an, daß Abweichungen der Erträge in der einen oder anderen Richtung gleich wahrscheinlich sind. Entscheide auf dem Signifikanzniveau von 15%, ob man die Nullhypothese »Beide Düngemittel sind gleichwertig« ablehnen kann. Auf welchem Niveau könnte man gerade noch ablehnen? Wie groß wäre die Wahrscheinlichkeit für einen Fehler zweiter Art, wenn tatsächlich die Wahrscheinlichkeit dafür, daß der X-Ertrag größer ist als der Y-Ertrag, 70% wäre?
39. Der Hersteller behauptet, Dünger X sei besser als Dünger Y (vgl. Aufgabe **38**). Entscheide mit Hilfe eines einseitigen Tests, ob die Nullhypothese der Gleichwirksamkeit auf dem Signifikanzniveau von 15% abgelehnt werden kann. Auf welchem Niveau kann sie nach den Daten der Aufgabe **38** gerade noch abgelehnt werden?
40. Eine Firma behauptet, das von ihr hergestellte Haarwasser heile in mehr als 70% aller Fälle Kahlköpfigkeit. Man stelle für die Stichprobenlänge 20 eine Entscheidungsvorschrift für das Testen dieser Hypothese auf, und zwar so, daß die Wahrscheinlichkeit für den Fehler 1. Art, die Behauptung zu glauben, obwohl sie nicht stimmt, höchstens 5% ist. Warum muß unbedingt ein einseitiger Test gewählt werden?
41. Das neue Waschmittel Albil soll durch eine große Werbeaktion eingeführt werden. Wenn es der Werbeagentur gelingt, Albil bei mehr als 45% der Bevölkerung bekannt zu machen, erhält sie von den Albil-Werken eine besondere Prämie. Die Entscheidung soll auf Grund einer Befragung von 200 bzw. 2000 Personen getroffen werden. Wie muß die Entscheidungsregel lauten, wenn die Albil-Werke nur 0,5% Risiko dafür eingehen wollen, daß die Agentur zu unrecht die Prämie erhält? Wie hoch ist dann das Risiko für die Agentur, die Prämie nicht zu erhalten, obwohl 60% der Bevölkerung von Albil erfahren haben?
42. Der Kaufpreis für eine Sendung Äpfel wird unter der Annahme vereinbart, daß 15% des Obstes unbrauchbar sind. Sollte die Qualität wider Erwarten besser sein, so ist ein gewisser Preisaufschlag zu zahlen; ist sie schlechter, so wird ein Preisnachlaß gewährt. Die Entscheidung wird nach folgender Regel getroffen: Sind von 50 zufällig ausgewählten Äpfeln mehr als 11 faul oder wurmbefallen, dann Preisnachlaß. Sind weniger als 5 Stück unbrauchbar, dann Preisaufschlag. In allen anderen Fällen gilt der vereinbarte Preis.
    a) Wie groß ist das Risiko des Verkäufers, einen ungerechtfertigten Preisnachlaß hinnehmen zu müssen, im ungünstigsten Fall?
    b) Wie groß ist das Risiko des Käufers, einen ungerechtfertigten Preisaufschlag hinnehmen zu müssen, im ungünstigsten Fall?
    c) Bei gleicher Stichprobenlänge sollen die beiden Risiken aus **a)** und **b)** unter 5% gedrückt werden. Welches Entscheidungsverfahren kann man wählen?
    d) Der wahre Gehalt der Sendung an unbrauchbarem Obst sei 25%. Mit welcher Wahrscheinlichkeit wird beim ursprünglichen Entscheidungsverfahren Preisnachlaß bzw. Preisaufschlag erzielt?
43. In der Zeitung steht: »Die Hälfte unserer Erwerbspersonen verdient weniger als 1600 DM monatlich.« Wir wählen daraufhin 100 Personen mit Einkommen zufallsbestimmt aus und finden, daß nur 42 davon ein Monatseinkommen unter 1600 DM haben. Auf welchem Signifikanzniveau können wir die Zeitungsbehauptung ablehnen? (Zweiseitiger Test)
44. Die Glühlampen einer bestimmten Marke haben zu 25% eine Brenndauer unter 1000 Stunden. Die Konkurrenz bringt einen neuen Typ auf den Markt, bei dem dieser Anteil angeblich kleiner ist. Wie viele von 100 Lampen der neuen Sorte müssen mindestens 1000 Stunden brennen, wenn man der Behauptung bei nur 5% Fehlerrisiko glauben soll?
45. Ein Präparat zur Steigerung der Konzentrationsfähigkeit wird an 15 Personen ausprobiert. Sie lösen an einem Tag Denkaufgaben ohne vorherige Stärkung, an einem andern

Tag verwandte Aufgaben nach Einnahme des Mittels. Bei 10 von ihnen zeigt sich eine Leistungssteigerung, bei 5 ist es umgekehrt. Wie ist auf dem 30%-Niveau zu testen, wenn
a) eine Leistungsminderung durch das Präparat ausgeschlossen ist,
b) eine solche Leistungsminderung in Betracht gezogen wird?
Welche Entscheidung wird in jedem der Fälle getroffen?

46. Wir nennen beim Zahlenlotto eine Zahl »selten« bzw. »häufig«, wenn ihre Ziehungshäufigkeit (ohne Berücksichtigung als Zusatzzahl) im jeweiligen kritischen Bereich zum 1%-Signifikanz-Niveau liegt. Bestimme diese kritischen Bereiche für 1225 Ziehungen bei »6 aus 49« und entnimm dann der Tabelle von Seite 38 die seltenen und häufigen Zahlen. Überlege vorher, ob einseitig oder zweiseitig getestet werden soll.

47. Lady X. behauptet, Teebeuteltee von richtig frei gebrühtem Tee unterscheiden zu können. Bestimme bei folgenden Tests jeweils das niedrigste Signifikanzniveau, bei dem man Lady X. noch eine solche Begabung attestieren könnte.
Für alle Aufgaben gelte, daß Lady X. die Bedingungen, unter denen sie getestet wird, kennt.
a) Es werden je eine Tasse vorgesetzt. Sie benennt beide richtig.
b) Zwei zufällig gefüllte Tassen werden beide richtig benannt.
c) 5 Paare mit je zwei Tassen verschieden gebrühten Tees werden alle richtig benannt.
d) 10 zufällig gefüllte Tassen werden alle richtig benannt.
e) 5 Teebeuteltassen und 5 andere werden in zufälliger Reihenfolge alle richtig benannt.
f) 5 Teebeuteltassen und 5 andere werden von Lady X. in zwei Gruppen auseinandersortiert, ohne daß sie aber sagen kann, welche Gruppe die Teebeuteltassen sind.

48. **a)** Man beweise folgende Formeln über die Binomialverteilung:

1) $$\frac{\mathrm{d}F_p^n(k)}{\mathrm{d}p} = -n\binom{n-1}{k}p^k(1-p)^{n-1-k}$$

2) Für $l > k$ gilt:

$$\frac{\mathrm{d}}{\mathrm{d}p}\sum_{i=k+1}^{l} \mathrm{B}(n;p;i) = np^k(1-p)^{n-1-l}\cdot\left[\binom{n-1}{k}(1-p)^{l-k} - \binom{n-1}{l}p^{l-k}\right].$$

b) Man beweise die auf Seite 355 aufgestellten Behauptungen über die 4 verschiedenen Typen von OC-Kurven.

49. Zeichne die OC-Kurve zum Test von **a)** Aufgabe 368/**28**, **b)** Aufgabe 370/**33**, **c)** Aufgabe 371/**41**. Gib den Term der zugehörigen Polynomfunktion an und bestimme ihre Maximumstelle.

50. Eine Urne enthält 10 Kugeln; mindestens drei davon sind schwarz. Die Nullhypothese sei »Genau drei der Kugeln sind schwarz«. Man zieht sechs Kugeln mit Zurücklegen und verwendet als Testgröße die Anzahl Z der schwarzen Kugeln in der Stichprobe. Die Entscheidung falle gemäß

$$\delta_k: \begin{cases} Z \geqq k \Rightarrow \text{Ablehnung von } H_0 \\ Z < k \Rightarrow \text{Keine Ablehnung von } H_0. \end{cases}$$

Bestimme die Irrtumswahrscheinlichkeiten $\alpha'(\delta_k)$. Zeichne die OC-Kurven für alle möglichen Entscheidungsregeln δ_k. Gib auch jeweils den Term der zugehörigen Polynomfunktion an.

51. Ein Hersteller liefert Glühbirnen mit einem Ausschußanteil von 10%. Der Empfänger testet die Lieferung, indem er 100 herausgreift und prüft. Er akzeptiert die Lieferung, falls 14 oder weniger defekt sind. Zeichne die OC-Kurve des Ereignisses »Annahme der Lieferung« in Abhängigkeit von der tatsächlichen Ausschußquote. Wie groß ist die Wahrscheinlichkeit für einen Fehler erster Art, falls die Ausschußquote wirklich 10% ist? Wie

groß ist die Wahrscheinlichkeit für einen Fehler zweiter Art, falls die Ausschußquote wegen eines Maschinenschadens auf 15% gestiegen ist? Wie ist es bei 25%?

52. Die Nullhypothese, eine Binomialverteilung habe den Parameter $p = 0{,}5$, soll auf dem Signifikanzniveau 5% zweiseitig getestet werden. Der Annahmebereich $\bar{K}$ sei möglichst schmal. Bestimme $\bar{K}$, zeichne die OC-Kurven von $\bar{K}$ für $n = 15, 20, 50, 100$ und gib den zugehörigen OC-Term an.

53. Die CSP wünscht ihren Kandidaten Meier auf jeden Fall bei der nächsten Wahl durchzubringen. Sie beschließt, den Wahlkampf auf die augenblickliche Stimmung des Publikums einzustellen. Sind mindestens 60% der Wähler zur Zeit für Meier, so genügt ein normaler Wahlkampf. Sind es weniger als 60%, so muß die sehr harte und kostspielige Variante des Wahlkampfes geführt werden.

a) Eine Umfrage bei 20 zufällig ausgesuchten Wählern soll die Entscheidung bringen. Gib ein Entscheidungsverfahren an, mit dem auf dem 10%-Niveau entschieden werden kann, ob ein normaler Wahlkampf genügt. Zeichne die OC-Kurve.

b) Wie lautet die Entscheidungsregel, falls 2000 Personen befragt werden? (Normalverteilung!) Zeichne die zugehörige OC-Kurve. (Die interessanten Werte liegen im Intervall $[0{,}55; 0{,}65]$.)

54. Z sei nach $B(5; p)$ verteilt. Es soll die Nullhypothese »$p \leqq 0{,}4$« gegen die Alternative »$p > 0{,}4$« getestet werden. Gibt es einen Annahmebereich $\bar{K} = »Z \leqq k«$, bei dem die Wahrscheinlichkeit für einen Fehler 1. Art stets $\leqq 1\%$ ist? Benütze Figur 357.1.

55. X sei nach $B(10; p)$ verteilt. Zeichne jeweils die OC-Kurve für das angegebene Ereignis, gib den Funktionsterm und die Monotoniebereiche an.

a) $[0; 6]$ **b)** $[4; 10]$ **c)** $[0; 4] \cup [7; 10]$ **d)** $[5; 6]$ **e)** $[0; 10]$ **f)** $\emptyset$

•56. **a)** Die Qualitätskontrolle klinisch-chemischer Analysen dient zur Überwachung der verwendeten Methode. Man analysiert dabei Proben, bei denen die Konzentration des zu bestimmenden Stoffes bekannt ist. Man sagt, »die Methode ist außer Kontrolle«, wenn eines der folgenden Kriterien zutrifft.

1) 7 aufeinanderfolgende Meßwerte liegen auf derselben Seite des Mittelwerts μ.

2) 7 aufeinanderfolgende Werte zeigen eine ansteigende oder eine abfallende Tendenz.

Man kann diese beiden Kriterien als 2 verschiedene Tests betrachten. Formuliere jeweils die zulässige Hypothese und die Nullhypothese. Gib den kritischen Bereich an und berechne die Wahrscheinlichkeit für einen Fehler 1. Art. Zeichne jeweils die OC-Kurve. – Hinweis zu **2)**: Wähle als Testgröße die Maximalzahl der monoton liegenden Werte.

b) Bestimme unabhängig vom Testproblem die Wahrscheinlichkeitsverteilung der im *Hinweis* angesprochenen Zufallsgröße in Abhängigkeit von $p := P(»\text{Meßwert ist größer als der vorhergehende}«)$; dabei wird angenommen, daß die Wahrscheinlichkeit für einen mit dem vorhergehenden Meßwert übereinstimmenden Meßwert Null ist. Bestimme ihren Erwartungswert und ihre Varianz und deren Werte für $p = \frac{1}{2}$.

***57.** »Bomber« Huber, der Fußballstar, schießt in einem Spiel Z Tore. Die Zufallsgröße Z sei *Poisson*-verteilt mit dem Erwartungswert μ. Der FC. X. will Huber anwerben, wenn er auf lange Sicht pro Spiel im Mittel mehr als 1,5 Tore schießt. Das nächste Spiel soll die Entscheidung bringen. Schießt Huber mindestens 3 Tore, dann wird man ihm eine passende Geldsumme anbieten. Wie groß ist die Wahrscheinlichkeit, daß der FC. X. Huber irrtümlich einkauft? Zeichne die OC-Kurve.

Zu 17.6.

58. Zeige: Der Test der Nullhypothese »$p = \frac{1}{2}$« über den Parameter p einer *Bernoulli*-Kette der Länge 10 ist verfälscht, wenn man als kritischen Bereich $K := [0; 3] \cup [8; 10]$ wählt. – Zeichne auch die OC-Kurve und bestimme ihren Hochpunkt.

59. **a)** Eine Urne enthält 10 Kugeln, darunter womöglich rote. Man testet die Nullhypothese »Die Urne enthält genau 3 rote Kugeln«, indem man 6 Kugeln mit Zurücklegen entnimmt und die Anzahl Z der roten Kugeln in der Stichprobe bestimmt.
Gib einen kritischen Bereich zum Signifikanzniveau 25% an und zeichne die OC-Kurve des Tests. Ist er verfälscht?

b) Löse **a)** durch Ziehen ohne Zurücklegen.

c) Löse **a)** für die Nullhypothese »Die Urne enthält mindestens 3 rote Kugeln«.

d) Löse **a)** für die Nullhypothese »Die Urne enthält mindestens 3 rote Kugeln« durch Ziehen ohne Zurücklegen.

60. **a)** Z sei nach $B(12; p)$ verteilt. Zeige, daß der Annahmebereich »$1 \leqq Z \leqq 3$« zur Nullhypothese »$p = \frac{1}{6}$« bezüglich der zulässigen Hypothese $[0; 1]$ einen verfälschten Test liefert. Für welche Nullhypothese ist der Test unverfälscht?

b) Z sei nach $B(20; p)$ verteilt. Zur Nullhypothese »$p = \frac{1}{5}$« werde der Annahmebereich »$1 \leqq Z \leqq 7$« festgesetzt. Zeige, daß dieser Test bezüglich $H = [0; 1]$ verfälscht ist. Zeichne die OC-Kurve des Annahmebereichs.

61. Untersuche, welche der Tests von Aufgabe **52** verfälscht sind.

62. Bei einem Blutalkoholgehalt von mehr als 0,8 Promille ist Autofahren strafbar. Das Gesetz zieht rigoros diese Grenze. In einer Klinik kann der Blutalkohol praktisch zweifelsfrei gemessen werden; der Schnelltest auf der Straße ist nicht so zuverlässig. Das Testergebnis – es lautet »Alkoholgehalt größer bzw. kleiner als 0,8‰« – kann in zweifacher Weise falsch sein. Erläutere die beiden Fehlermöglichkeiten und ihre Folgen! Welche Wahl der Fehlerwahrscheinlichkeiten entspricht unserem Rechtsgrundsatz »in dubio pro reo«? Welche Konsequenzen ergäben sich, wenn der Test verfälscht wäre? Welche besondere Problematik ergibt sich daraus, daß der Blutalkoholgehalt eines Fahrers auch beliebig genau bei 0,8‰ liegen kann?

Zu 17.7.

63. Eine Abfüllanlage soll Zuckerpakete zu je 1000 g abfüllen. Die Zufallsgröße X gebe den wirklichen Inhalt in g an. Aus Erfahrung weiß man, daß $\operatorname{Var} X = 25$ gilt. Eine Messung von 50 Paketen soll darüber entscheiden, ob die Anlage neu eingestellt werden muß. Man wählt als Testgröße das arithmetische Mittel $\bar{X}$ der 50 Messungen und nimmt an, daß es normalverteilt ist. Bestimme für das Signifikanzniveau 5% den kritischen Bereich
a) für einen einseitigen Test, wo man sich nur für zuviel Zucker im Paket interessiert,
b) für einen zweiseitigen Test.

64. Eine Lehrmittelfirma liefert Widerstände und behauptet, ihr Nennwert 50 Ω sei bei einer Varianz von 25 Ω^2 gesichert. Bestimme zu einem Signifikanzniveau von 5% den kritischen Bereich für einen zweiseitigen Test, wenn 10 Widerstände gemessen werden und als Testgröße das arithmetische Mittel der gemessenen Widerstände genommen wird. Wie wird man sich entscheiden, wenn die Messung der 10 Widerstände folgende Werte in Ω ergab: 49,0 46,9 50,0 46,8 53,1 50,6 50,2 47,7 49,0 48,5.

65. Bei Werkzeugmaschinen kennt man oft die Streuung für die Maße der Produkte aus Erfahrung, während der Mittelwert von der jeweiligen Einstellung der Maschine abhängt. Eine Maschine produziert Bolzen der Länge L mm. Die Zufallsgröße L sei normalverteilt mit der Standardabweichung $\sigma = 0{,}5$. Wenn der Erwartungswert $\mathcal{E} L = \mu$ außerhalb des Intervalls $[97; 103]$ liegt, muß die Maschine neu eingestellt werden. Ein solcher Fall soll mit mindestens 98% Sicherheit erkannt werden. Es wird 1 beliebig herausgegriffener Bolzen genau gemessen. In welchem Intervall muß seine Länge liegen, wenn die Maschine weiterlaufen darf? Wie groß ist die Mindestwahrscheinlichkeit dafür, daß die Maschine auf Grund des Tests unnötigerweise neu eingestellt wird? Zeichne die OC-Kurve.