

Statistik II
Induktive Statistik
Skript zur Vorlesung WS 2005/2006

Jörg-Peter Schräpler
email: joerg-peter.schraepler@ruhr-uni-bochum.de

14. Dezember 2005

Inhaltsverzeichnis

1	Vorwort	3
2	Grundlagen der Wahrscheinlichkeitsrechnung	3
2.1	Zufallsexperiment und Ereignisse	3
2.1.1	Das Rechnen mit Ereignissen	4
2.2	Wahrscheinlichkeitsbegriffe	4
2.2.1	Das Gleichmöglichkeitsmodell	5
2.2.2	Statistische Wahrscheinlichkeit	7
2.3	Axiome von Kolmogoroff	9
2.3.1	Addition von Wahrscheinlichkeiten	10
2.3.2	Multiplikation von Wahrscheinlichkeiten	11
2.3.3	Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit	12
2.3.4	Satz der totalen Wahrscheinlichkeit	14
2.3.5	Theorem von Bayes	15
3	Zufallsvariable und Wahrscheinlichkeitsverteilung	17
3.1	Zufallsvariable	18
3.2	Wahrscheinlichkeitsverteilung	18
3.2.1	Wahrscheinlichkeitsfunktion bei diskreten Zufallsvariablen	18
3.2.2	Wahrscheinlichkeitsfunktion bei stetigen Zufallsvariablen	20
3.3	Mehrdimensionale Zufallsvariablen	23
3.3.1	Gemeinsame Verteilung bei zwei diskreten Zufallsvariablen	23
3.3.2	Gemeinsame Verteilung bei zwei stetigen Zufallsvariablen	25
3.4	Unabhängigkeit zweier Zufallsvariablen	25
3.5	Erwartungswert und Varianz	26
3.5.1	Erwartungswert	26
3.5.2	Varianz einer Zufallsvariablen	27
4	Ausgewählte Wahrscheinlichkeitsverteilungen	29
4.1	Binomialverteilung	29
4.2	Multinomialverteilung	32
4.3	Normalverteilung	33
4.4	Approximation von Verteilungen	37
4.4.1	Approximation der Binomialverteilung	38
4.5	Chi-Quadrat-Verteilung	39

5 Stichprobe, Stichprobenvariable und Stichprobenverteilung	40
5.1 Grundgesamtheit und Stichprobe	40
5.2 Repräsentativität	40
5.3 Zufällige und bewußte Auswahl	41
5.4 Einfache Zufallsstichprobe	41
5.5 Stichprobenvariable	42
5.6 Stichprobenverteilungen	42
5.6.1 Spezielle Stichprobenverteilungen	43
6 Das Testen von statistischen Hypothesen	45
6.1 Richtige und falsche Testentscheidungen	48
7 Das Schätzen von Parametern	50
7.1 Punktschätzung	51
7.1.1 Schätzfunktionen und ihre Eigenschaften	51
7.1.2 Eigenschaften von Schätzfunktionen	51
7.2 Intervallschätzung	52
A Tabellierte Werte der Binomialverteilung	54
B Tabellierte Werte der Standardnormalverteilung	57
Literatur	58

1 Vorwort

Das vorliegende Skript ist für Studenten der Veranstaltung „Statistik II“ im Grundstudium des WS 2005/2006 konzipiert und liefert eine Zusammenfassung des Vorlesungsstoffs. Im weitesten Sinne ist es eine Art Stichwortliste und Formelsammlung. Weitergehende Ausführungen zum Thema „Induktive Statistik“ finden sich in der im Anhang erwähnten Literatur.

Die schließende Statistik befaßt sich im wesentlichen mit der Beurteilung und Verallgemeinerung von statistischen Befunden, die aus sogenannten *Zufallsstichproben* stammen. Ziel ist dabei, Rückschlüsse von einer derartigen Stichprobe auf die Grundgesamtheit zu ziehen. Man geht davon aus, daß dies bei Zufallsstichproben gewährleistet ist. Das Phänomen „Zufall“ ist aber Gegenstand der Wahrscheinlichkeitsrechnung.

2 Grundlagen der Wahrscheinlichkeitsrechnung

Um überhaupt von „verallgemeinerbaren“ Ergebnisse von Stichproben sprechen zu können, muß die Gewinnung der Daten prinzipiell unter den gleichen Randbedingungen wiederholt durchgeführt werden können. Man verwendet in diesem Zusammenhang häufig den Begriff „Zufallsexperiment“.

2.1 Zufallsexperiment und Ereignisse

Ein **Zufallsexperiment** ist durch folgende Eigenschaften gekennzeichnet:

- Es gibt mehrere mögliche Ergebnisse des Vorgangs
- die Ergebnisse sind nicht mit Sicherheit vorhersagbar
- der Vorgang ist beliebig oft wiederholbar
- der Vorgang läuft nach festen Regeln ab, d.h. unter konstanten Randbedingungen

Ergebnis, Ergebnisraum: Der Ausgang eines Zufallsexperimentes wird als Versuchsergebnis oder auch nur als **Ergebnis** ω bezeichnet. Alle tatsächlich möglichen Ergebnisse des Zufallsexperiments bilden dann den **Ergebnisraum**¹ $\Omega := \{\omega_1, \omega_2, \dots, \omega_n\}$.

Ereignis: Jede Teilmenge A des Ergebnisraumes Ω , $A \subset \Omega$, wird dann **Ereignis** genannt. Das Ereignis A tritt dann ein, wenn Ergebnis ω beobachtet wird, das zu A gehört, für das mit anderen Worten $\omega \in A$ gilt. Ein **Elementarereignis** liegt dann vor, wenn A lediglich ein Versuchsergebnis enthält, also $A = \{\omega\}$.

Stichprobe als Zufallsexperiment: Eine zufällige Auswahl von Datenträgern, z.B. eine Zufallsstichprobe von 100 Personen wird ebenfalls als Zufallsexperiment angesehen. Das „Versuchsergebnis“ besteht jeweils aus der ausgewählten Person und die tatsächlich möglichen Auswahlsergebnisse bilden den Ergebnisraum bzw. hier den Stichprobenraum.

Beispiel Würfel: Anhand eines Würfelwurfes lassen sich die Begriffe und die Merkmale eines Zufallsexperimentes sehr gut demonstrieren. Der Ergebnisraum Ω besteht bei einem einfachen Würfelwurf aus sechs Versuchsergebnissen ω_i , den Augenzahlen $1, 2, \dots, 6$. Beim gleichzeitigen Wurf zweier Würfel oder einem zweimaligen Würfelwurf können insgesamt 36 Versuchsergebnisse

¹Schlittgen (1998, S.58) verwendet hierfür die Bezeichnung *Ergebnismenge* oder *Stichprobenraum*.

unterschieden werden. Jedes Ergebnis im ersten Wurf kann mit sechs möglichen Ergebnissen des zweiten Wurfes kombiniert werden. Also:

$$\Omega = \{1, 1; 1, 2; 1, 3; 1, 4; 1, 5; 1, 6; 2, 1; 2, 2; 2, 3; 2, 4; 2, 5; 2, 6; \dots; 6, 6\}$$

Ein Ereignis A könnte dann z.B. der „Wurf von zwei geraden Zahlen“ sein. Dieses Ereignis enthält die folgenden Ergebnisse:

$$A = \{2, 2; 2, 4; 2, 6; 4, 2; 4, 4; 4, 6; 6, 2; 6, 4; 6, 6\}$$

Oder das Ereignis B „höchstens drei Augen“ wäre dann:

$$B = \{1, 1; 1, 2; 2, 1\}$$

2.1.1 Das Rechnen mit Ereignissen

Die üblichen mengentheoretischen Operationen und Sprechweisen erhalten im Rahmen der Wahrscheinlichkeitsrechnung eine Umdeutung und Umformulierung. Die Tabelle 1 übersetzt die verwendete Notation im Kontext der Mengenlehre in den Kontext des Rechnens mit Ereignissen.

Die Darstellung der Begriffsbildungen erfolgt häufig auch über sogenannte **Venn-Diagramme**. Bei einem Venn-Diagramm wird die Ergebnismenge Ω durch die Fläche eines Rechteck charakterisiert. Ereignisse werden durch Teilflächen repräsentiert.

Die **Gleichheit zweier Ereignisse** A und B liegt dann vor, wenn $A = B$, wenn also jedes Ergebnis, das zu A gehört auch zu B gehört und umgekehrt.

$$\omega \in A \Leftrightarrow \omega \in B$$

B ist ein **Teilergebnis** von A , wenn nur die eine Richtung gilt: Jedes Ergebnis aus B gehört auch zu A .

$$\omega \in B \Rightarrow \omega \in A$$

Siehe hierzu Abbildung 2. Man sagt auch, daß das Ereignis B das Ereignis A nach sich zieht, tritt B ein, so tritt auch A ein.

Das **Komplementäreignis** \bar{A} des Ereignisses A besteht aus allen Ergebnissen von Ω , die nicht zu A gehören. \bar{A} tritt genau dann ein, wenn A nicht eintritt. Siehe hierzu Abbildung 5.

Die **Vereinigung** $A \cup B$ zweier Ereignisse A und B besteht aus allen Ergebnissen die zu A und B gehören. Man sagt auch A **oder** B tritt ein, wenn $A \cup B$ eintritt. Siehe hierzu Abbildung 4. Als **Differenz** $A \setminus B$ zweier Ereignisse wird das Ereignis $A \cap \bar{B}$ bezeichnet. Siehe hierzu Abbildung 3.

Sind A und B zwei Ereignisse, so ist $A \cap B$ als Teilmenge von Ω wieder ein Ereignis, der sogenannte **Durchschnitt** von A und B . das Ereignis $A \cap B$ besteht aus den Ergebnissen, die in A **und** B vorkommen. Das Eintreten von $A \cap B$ wird auch mit „ A und B treten zugleich ein“ bezeichnet.

2.2 Wahrscheinlichkeitsbegriffe

Das erste größere Werk zur Wahrscheinlichkeitsrechnung stammt von dem Mathematiker Jakob Bernoulli, es trägt den Titel *Ars conjectandi* („Kunst der Vermutungen“) und wurde posthum im Jahr 1713 veröffentlicht. Er beschäftigt sich in den ersten Kapiteln seines Buches vor allem mit Glücksspielen, wie Münzwurf, Würfel- und Kartenspiele. Zum Schluß geht er allerdings darüber hinaus und versucht eine allgemeine Theorie vernünftiger Urteilsbildung für Situationen der Unsicherheit zu entwickeln. Ausgangspunkt seiner Betrachtungen bilden allerdings Glücksspiele oder allgemeiner auch sogenannte „Zufallsgeneratoren“². Wahrscheinlichkeitsaussagen,

²Eine ausführliche Darstellung und Kritik der Wahrscheinlichkeitskonzeptionen findet sich bei Rohwer (2000). Ein Kapitel beschäftigt sich dabei ausführlich mit Zufallsgeneratoren.

TABELLE 1: Übersetzung mengentheoretischer Sprechweisen und Operationen in Sprechweisen und Operationen mit Ereignissen

Im Kontext der Mengenlehre		Im Kontext von Ereignissen	
Notation	Begriff	Begriff	Beschreibung
Ω	Grundmenge	sicheres Ereignis	Ω tritt mit Sicherheit ein
$\{\}$ bzw. \emptyset	leere Menge	unmögliches Ereignis	\emptyset tritt nie ein
$A \subset B$	A ist Teilmenge von B	A ist Teilereignis von B	wenn A eintritt, tritt auch B ein
$A = B$	identische Mengen	äquivalente Ereignisse	A tritt genau dann ein, wenn B eintritt
$A \cap B$	Schnittmenge (Durchschnittsmenge)	Durchschnitt der Ereignisse A und B	sowohl A als auch B tritt ein (A und B)
$A \cap B = \emptyset$	disjunkte Teilmengen	disjunkte Ereignisse (unvereinbare Ereignisse)	A und B schließen sich aus
$A \cup B$	Vereinigungsmenge	Vereinigung der Ereignisse A und B	mindestens eines der Ereignisse A oder B tritt ein (A oder B)
\bar{A}	Komplementärmenge	zu A komplementäres Ereignis	A tritt nicht ein
$A - B$	Differenzmenge	Differenz der Ereignisse A und B	A aber nicht B tritt ein
$\omega \in A$	ω ist Element von A	Versuchsergebnis ω ist in A enthalten	wenn ω beobachtet wird, tritt A ein
$\omega \notin A$	ω ist kein Element von A	Versuchsergebnis ω ist nicht in A enthalten	wenn ω beobachtet wird, tritt A nicht ein

die sich auf Zufallsgeneratoren stützen können, werden *aleatorische Wahrscheinlichkeitsaussagen* genannt. Sie beziehen sich auf Eigenschaften eines Zufallsgenerators wie z.B. Glücksspiele. Einige Autoren (z.B. Rohwer 2000) vertreten die Auffassung, daß diese Aussagen von *epistemischen Wahrscheinlichkeitsaussagen*, die sich auf Hypothesen beziehen, streng unterschieden werden müssen.

2.2.1 Das Gleichmöglichkeitsmodell

Das Gleichmöglichkeitsmodell ist eine Wahrscheinlichkeitskonzeption, die auf Glücksspiele zurückgeht und ist in diesem Sinne auch von Bernoulli verwendet worden. Zur Formulierung des Modells kann man z.B. von Urnen, die mit N vollständig symmetrischen Objekten, wie z.B. Kugeln gefüllt sind, ausgehen. Die Kugeln unterscheiden sich nur durch ihre Farbe und Nummer. Die Objekte werden gut gemischt und eines davon wird blind herausgegriffen. Gegebenenfalls wird dieses Herausgreifen wiederholt. Falls die gezogene Kugel jeweils wieder zurückgelegt wird, soll durch erneutes Mischen die Ausgangssituation wieder hergestellt werden.

Das Wesentliche beim Urnenmodell besteht darin, daß bei diesem Zufallsexperiment bei jedem Zug die *Chance* jeder noch in der Urne befindlichen Kugel gleich groß ist. Man kann auch sagen, daß das Verfahren keine Erinnerung an frühere Anwendungen haben darf.

Wenn diese Bedingungen erfüllt sind, läßt sich ein Wahrscheinlichkeitsmaß durch folgende Definition festlegen:

$$P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Ergebnisse}}{\text{Anzahl der gleich möglichen Ergebnisse}}$$

Diese Definition der Wahrscheinlichkeit stammt von Laplace (1749 - 1827). Etwas allgemeiner ausgedrückt:

Im endlichen Gleichmöglichkeitsmodell mit N Versuchsergebnissen, von denen m im Sinne der Fragestellung günstig sind und das Ereignis A bilden, wird die Wahrscheinlichkeit $P(A)$ für die Realisation von A bei einem Zufallsexperiment mit m/N bemessen.

$$P(A) = \frac{m}{N}$$

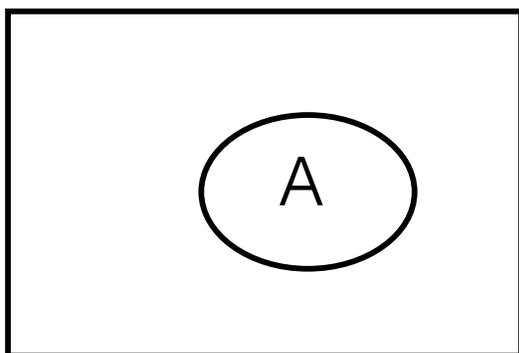
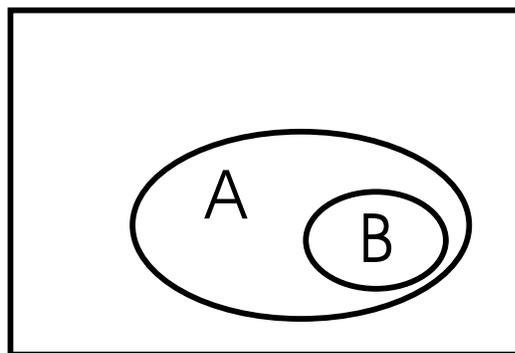
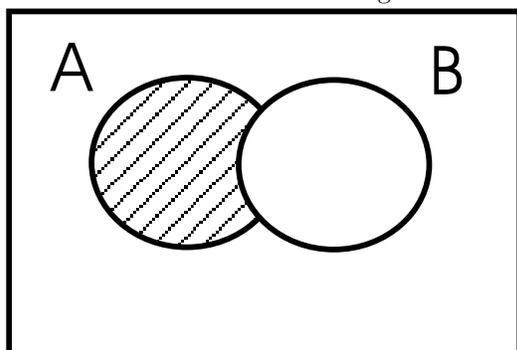
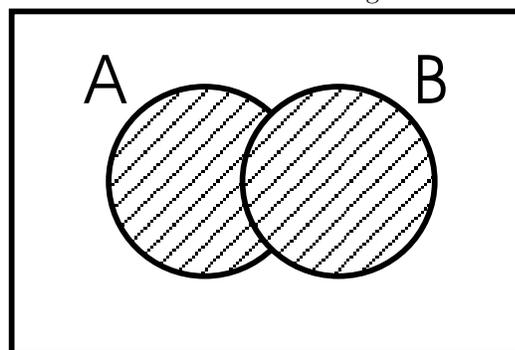
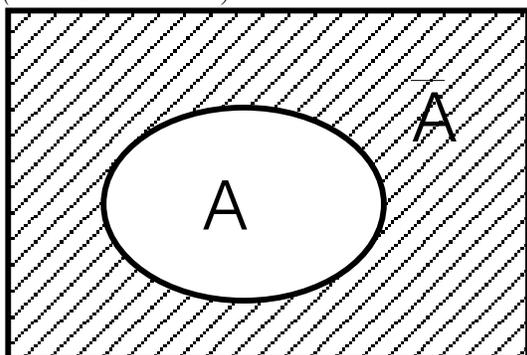
ABBILDUNG 1: A als Teilmenge von Ω ABBILDUNG 2: B als Teilereignis von A ABBILDUNG 3: Differenz $A \setminus B$ von A, B (schraffierte Fläche)ABBILDUNG 4: Vereinigung $A \cup B$ von A, B (graue Fläche)

ABBILDUNG 5: Komplementäreignis (schraffierte Fläche)

Die Wahrscheinlichkeit $P(A)$ ist nach diesem Konzept also auf der Grundlage theoretischer Überlegungen **vor** der Durchführung eines Zufallsexperimentes bestimmbar. Deshalb wird diese Wahrscheinlichkeit auch **A-priori-Wahrscheinlichkeit** genannt.

Während sich ursprünglich dieses (aleatorischen) Wahrscheinlichkeitskonzepts von Bernoulli ausschließlich auf die Verwendung von Zufallsgeneratoren bezog, versuchten A. de Moivre (1756) sowie Laplace (1814) diesen Wahrscheinlichkeitsbegriff allgemein anwendbar zu machen.

Die Begründung der Gleichmöglichkeit der Elementarereignisse erfolgte durch das etwas dubiose „Prinzip des mangelnden (unzureichenden) Grundes“, alle Ergebnisse bzw. Elementarereignisse des Zufallsexperiments können **a priori** als gleich möglich bewertet werden, wenn es keinen Grund gibt, der dagegen spricht, bzw. wenn man über die Ereignisse „in gleicher Weise un-schlüssig“ ist.

Beispiele: Das Urnenmodell läßt sich auf einige Glücksspiele anwenden bzw. diese in das Urnenmodell übersetzen:

- Würfeln: In der Urne sind 6 Kugeln. Die Zahl auf der gezogenen Kugel wird als Augenzahl des Würfels interpretiert. Bei mehrmaligem Würfeln wird die gezogene Kugel vor jedem neuen Zug wieder zurückgelegt.

Bei dem Spiel „Mensch ärgere Dich nicht“ ist es am Anfang wichtig, eine „Sechs“ zu würfeln, um seine Spielfigur auf das Spielfeld setzen zu können. Die Wahrscheinlichkeit, bei einem Versuch eine Sechs zu würfeln läßt sich dann mit $P(A) = \frac{1}{6}$ angeben, da ein Elementarereignis im Sinne der Fragestellung „günstig“ und insgesamt sechs Elementarereignisse gleich möglich sind.

Die Wahrscheinlichkeit für ein Ereignis B , eine gerade Zahl zu würfeln ist bei $B = \{2, 4, 6\}$ analog $P(B) = \frac{3}{6}$.

- Münzwurf: Die Urne enthält 2 Kugeln. Die Zahl 1 entspricht der Kopfseite (K), die Zahl 2 der Wappenseite (W).

Die Wahrscheinlichkeit, bei einem zweimaligen Münzwurf mindestens einmal Kopf zu werfen läßt sich dann auch a-priori angeben. Insgesamt mögliche Ereignisse: KK, KW, WK, WW . Davon sind drei im Sinne der Fragestellung „günstig“. Also ist die Wahrscheinlichkeit $P(A) = \frac{3}{4}$.

- Roulette: Die 37 Kugeln in der Urne werden mit den Zahlen von 0 bis 36 des Roulette-Spiels gleichgesetzt.

2.2.2 Statistische Wahrscheinlichkeit

Für die Entwicklung von statistischen oder auch frequentistischen Wahrscheinlichkeitskonzepten lassen sich vor allem drei Motive angeben (vgl. Rohwer 2000). 1. Die Erfahrungstatsache, daß man sich in vielen Fällen von einem Zusammenhang zwischen aleatorischen Wahrscheinlichkeiten und relativen Häufigkeiten überzeugen kann. 2. Die Einschränkung des ursprünglichen klassischen Ansatzes auf Glücksspiele. Frequentistische Ansätze scheinen es dagegen zu erlauben, die Anwendungsmöglichkeiten auf alle Gebiete auszuweiten, in denen sich sinnvoll von Massenerscheinungen und relativen Häufigkeiten sprechen läßt. 3. Man versuchte wegen des etwas dubiosen „Prinzips des unzureichenden Grunde“ einen empirischen Ansatz zum Verständnis von Wahrscheinlichkeiten zu finden.

Es ist eine Erfahrungstatsache, daß sich bei einer wachsenden Zahl von Wiederholungen eines Zufallsexperimentes die relativen Häufigkeiten von Ereignissen jeweils um einen festen Wert stabilisieren. Dieses *Prinzip der großen Zahl* wurde erstmalig von Cardano (1501 - 1576) formuliert. Diese Beobachtung war dann Anlaß zur Formulierung eines frequentistischen Konzepts von Wahrscheinlichkeit.

Ausgangspunkt sind hierbei Beobachtungen, die unter gleichen Bedingungen durchgeführt werden. Das Konzept besagt, daß bei einem hinreichend großen Beobachtungsumfang N eines Merkmals die Merkmalsausprägung A in der Häufigkeit n_A gezählt werden kann. Die Wahrscheinlichkeit $P(A)$, daß bei einer einmaligen Beobachtung die Ausprägung A auftritt, kann dann durch die relative Häufigkeit $h_A = \frac{n_A}{N}$ angegeben werden:

$$P(A) = \frac{n_A}{N}$$

Im Gegensatz zu der a-priori Wahrscheinlichkeit läßt sich die Wahrscheinlichkeit nach diesem frequentistische Konzept erst nach Durchführung von Beobachtungen oder Experimenten bestimmen, weshalb sie auch **A-posteriori-Wahrscheinlichkeit** genannt wird.

Folgende konzeptionellen Schwierigkeiten stellen sich:

- Die Beobachtungen müssen stets unter den gleichen Bedingungen erfolgen. Dies ist unrealistisch und nur im Falle von künstlichen Vorgängen, wie z.B. Situationen, die durch Zufallsgeneratoren erzeugt werden möglich.
- Der Hinweis auf einen 'hinreichend großen Beobachtungsumfang' ist nur vage, es gibt keine allgemein gültigen Regeln hierfür. Klar ist anscheinend nur, daß der Stichprobenumfang nach Möglichkeit groß sein sollte.

Der wichtigste Vertreter einer frequentistischen Auffassung, R. v. Mises (1883 - 1953) verweist auf die Existenz einer theoretischen Konstanten, eines **Grenzwertes**, dem sich die relativen Häufigkeiten bei unendlich großer Stichprobengröße annähern.

Definition : Die Wahrscheinlichkeit $P(A)$ ist als Grenzwert der relativen Häufigkeit $\frac{n_A}{n}$ anzusehen, wenn n gegen unendlich strebt.

$$P(A) = \lim_{n \rightarrow \infty} \left(\frac{n_A}{n} \right)$$

Im Prinzip lassen sich nach dieser Auffassung keine Wahrscheinlichkeiten bestimmen, relative Häufigkeiten bilden nur mehr oder weniger genaue Näherungswerte hierfür.

Beispiel Münzwurf: Das „Prinzip der großen Zahl“ wird in der Literatur fast ausschließlich mit Beispielen auf Basis von Zufallsgeneratoren erläutert. Die Ergebnisse des Münzwurfexperiments in Tabelle 2 stammen aus Schlittgen (1996, S.67). Die Tabelle zeigt, daß sich bei wachsender Zahl von Würfeln die relativen Häufigkeiten beim Wert 0.5 stabilisieren.

Im Gegensatz zu dem klassischen aleatorischen Wahrscheinlichkeitskonzept, welches die Gleichmöglichkeit aus den Eigenschaften eines Verfahrens, hier eines Zufallsgenerators, wie etwa einem Glücksspiel, ableitet, bezieht sich das frequentistische Wahrscheinlichkeitskonzept auf relative Häufigkeiten, d.h. auf Situationen. Die **metaphysische Grundannahme** ist, „daß

TABELLE 2: *Häufigkeiten des Ereignisses Kopf*

n	n(Kopf)	h(Kopf)
10	7	0.700
20	11	0.550
40	17	0.425
60	24	0.400
80	34	0.425
100	47	0.470
200	92	0.460
400	204	0.510
600	348	0.580
800	404	0.505
1000	492	0.492
2000	1010	0.505
3000	1530	0.510
4000	2032	0.508
5000	2515	0.503

man empirisch ermittelbare relative Häufigkeiten so auffassen könne, daß sich in ihnen näherungsweise gewisse 'Konstanten' manifestieren; und daß diese 'Konstanten' als Eigenschaften irgendwelcher, aber nicht explizit gemachter Prozesse gedeutet werden können, die die Situationen hervorbringen, auf die sich die relativen Häufigkeiten beziehen" (Rohwer 2000, S.135).

Hier liegt der eigentliche **wunde Punkt** des frequentistischen Ansatzes: Für die Annahme, daß den empirischen relativen Häufigkeiten sogenannte 'Konstanten' oder 'Grenzwerte' zugrundeliegen, gibt es keine Begründung. Derartige Konstanten lassen sich nur in Bezug auf ein Verfahren, wie einem Zufallsgenerator, ableiten, durch die erst die Situationen entstehen, aus denen dann die relativen Häufigkeiten ermittelt werden. Dies wäre dann aber wiederum der Verweis auf das klassische Wahrscheinlichkeitskonzept.

2.3 Axiome von Kolmogoroff

Die konzeptionell festgelegten Wahrscheinlichkeiten müssen einige zentrale Eigenschaften erfüllen, damit sie sich für Berechnungen eignen. Die Eigenschaften entsprechen weitestgehend denen, die auch für relative Häufigkeiten gelten. Für die Wahrscheinlichkeitsrechnung wurden 1933 von Kolmogoroff folgende Eigenschaften definiert:

Definition: Ein Wahrscheinlichkeitsmaß P ist eine Abbildung, die allen Ereignissen $A \subset \Omega$ eines Zufallsexperiments eine Zahl $P(A)$ zuordnet, und die folgende Bedingungen erfüllt:

- 1) $0 \leq P(A) \leq 1$
- 2) $P(\Omega) = 1$
- 3) $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ falls $A_1 \cap A_2 = \emptyset$

$P(A)$ heißt die Wahrscheinlichkeit des Ereignisses A .

Die Wahrscheinlichkeit eines Ereignisses ist nicht negativ und liegt zwischen 0 und 1. Das sichere Ereignis hat die Wahrscheinlichkeit 1 und die Wahrscheinlichkeit für die Vereinigung von A und B ist so groß wie die Summe der Wahrscheinlichkeiten der einzelnen Ereignisse, falls diese Ereignisse disjunkt sind.

Aus diesen Axiomen läßt sich weiter ableiten:

$$P(\bar{A}) = 1 - P(A)$$

Wenn nicht nur zwei sondern mehrere Ereignisse A_1, A_2, \dots, A_n mit $A_i \cap A_j = \emptyset$ für $i \neq j$, sich paarweise ausschließen, ist die Wahrscheinlichkeit für die Vereinigung dieser Ereignisse so groß wie die Summe der Einzelwahrscheinlichkeiten:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$$

Beispiel der einfache Münzwurf: Definition der Ereignisse: Z für Zahlseite oben, W für Wappenseite oben. Beide erhalten aus Symmetriegründen die gleiche Wahrscheinlichkeit. Das Ereignis K für 'Kante' ist extrem selten, ihm wird deshalb die Wahrscheinlichkeit 0 zugeordnet: $P(K) = 0$. Die Wahrscheinlichkeiten für die zwei anderen Ereignisse sind dann:

$$P(Z) = P(W) = 0.5$$

des weiteren muß gelten:

$$\begin{aligned} P(\Omega) &= P(Z \cup W \cup K) = 1 \\ P(\emptyset) &= 0 \\ P(Z \cup K) &= 0.5 \\ P(Z \cup W) &= 1 \\ P(W \cup K) &= 0.5 \end{aligned}$$

2.3.1 Addition von Wahrscheinlichkeiten

Das dritte Axiom von Kolmogoroff besagt, daß Wahrscheinlichkeiten addiert werden dürfen, wenn Sie sich gegenseitig ausschließen bzw. disjunkt sind. Wie verhält es sich dann mit nicht disjunkten Ereignissen? Hier kann folgende Überlegung weiter führen:

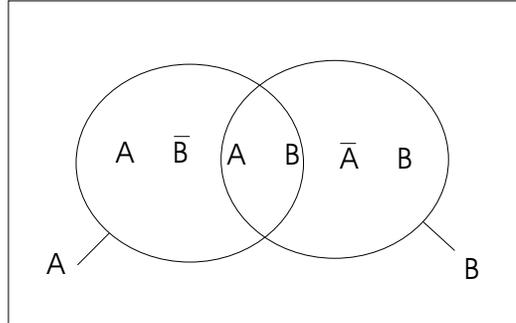


ABBILDUNG 6: Vereinigung der Ereignisse A und B über drei sich paarweise ausschließende Ereignisse

Die Vereinigungsmenge der Ereignisse A und B kann in drei sich paarweise ausschließende Ereignisse zerlegt werden (siehe hierzu Abbildung 6):

$$A \cup B = (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B)$$

Die Wahrscheinlichkeit für $A \cup B$ ist somit

$$P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B)$$

Das Ereignis A läßt sich in zwei sich ausschließende Ereignisse zerlegen, $A = (A \cap \bar{B}) \cup (A \cap B)$, so daß

$$P(A) = P(A \cap \bar{B}) + P(A \cap B)$$

Analog dazu gilt für Ereignis $B = (A \cap B) \cup (\bar{A} \cap B)$ so daß,

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

Man erhält dann für $P(A \cup B)$ durch Einsetzen die Gleichung des allgemeinen Additionssatzes für beliebige Ereignisse:

$$\begin{aligned} P(A \cup B) &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

Falls mehr als drei Ereignisse vorliegen, erhält man:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Beispiel: Roulette Bei einem Roulette-Spiel setzt ein Spieler auf drei Ereignisse. A : „Schwarz“, B : „gerade Zahl“, C : „erstes Dutzend“ (Zahlen von 1 - 12). Bei 37 Feldern inklusive der „Null“ ergeben sich folgende Wahrscheinlichkeiten: $P(A) = 18/37$; $P(B) = 18/37$; $P(C) = 12/37$; $P(A \cap B) = 9/37$; $P(B \cap C) = 6/37$; $P(A \cap C) = 6/37$ und $P(A \cap B \cap C) = 3/37$. Gesucht ist die Wahrscheinlichkeit dafür, daß ein Ergebnis eintritt, auf das gesetzt wurde, also $P(A \cup B \cup C)$.

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= \frac{18}{37} + \frac{18}{37} + \frac{12}{37} - \frac{9}{37} - \frac{6}{37} - \frac{6}{37} + \frac{3}{37} \\ &= \frac{30}{37} = 0.81 \end{aligned}$$

2.3.2 Multiplikation von Wahrscheinlichkeiten

Neben dem allgemeinen Additionssatz für beliebige Ereignisse, der sich aus den Axiomen von Kolmogoroff ableiten läßt, existieren auch Rechenregeln für eine sinnvolle multiplikative Verknüpfung von Wahrscheinlichkeiten. Hierzu wird auf das Konzept der bedingten relativen Häufigkeit zurückgegriffen.

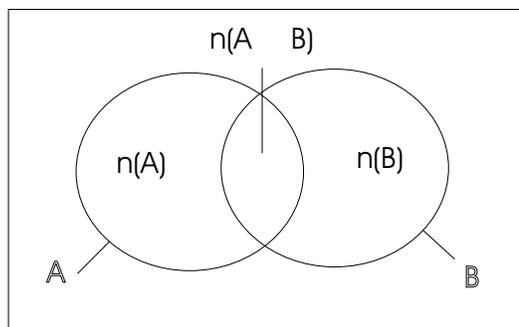


ABBILDUNG 7: Ergebnisraum von zwei Ereignissen

Ausgangspunkt bildet ein Zufallsexperiment, dessen Ergebnisraum aus n gleich möglichen Versuchsergebnissen besteht. Das Ereignis A enthält $n(A)$ Ergebnisse. Der Durchschnitt mit dem Ereignis B (mit $n(B)$ Ergebnisse) enthält $n(A \cap B)$ Ergebnisse. Somit ist $P(A) = n(A)/n$ und $P(A \cap B) = n(A \cap B)/n$. Durch Erweiterung mit $n(A)$ ergibt sich:

$$P(A \cap B) = \frac{n(A \cap B)}{n} = \frac{n(A)}{n} \cdot \frac{n(A \cap B)}{n(A)} = P(A) \cdot \frac{n(A \cap B)}{n(A)}$$

Die Wahrscheinlichkeit $P(A \cap B)$ läßt sich somit als Multiplikation der Wahrscheinlichkeit $P(A)$ mit $\frac{n(A \cap B)}{n(A)}$ ausdrücken. In dem Bruch $n(A \cap B)/n(A)$ wird die Anzahl der gemeinsamen Ergebnisse von A und B auf die Ergebnisse von A bezogen. Dieser Ausdruck läßt sich als eine Wahrscheinlichkeit konzipieren, so daß $n(A \cap B)/n(A)$ die Wahrscheinlichkeit für den Eintritt des Ereignisses B angibt, unter der Bedingung, daß das Ereignis A eintritt, bzw. schon eingetreten ist.

Man spricht in diesem Zusammenhang von einer **bedingten Wahrscheinlichkeit** $P(B|A)$, die „Wahrscheinlichkeit für B unter der Bedingung A “. Daraus folgt:

$$P(A \cap B) = P(A)P(B|A)$$

Die Darstellung erfolgt hier aus der Sicht des Ereignisses A , bzw. die gemeinsamen Ergebnisse werden auf die Ergebnisse von A bezogen. Analog hierzu läßt sich aber auch das Ereignis B mit seinen $n(B)$ Ergebnissen vollständig in die Betrachtung einbeziehen:

$$P(A \cap B) = \frac{n(A \cap B)}{n} = \frac{n(B)}{n} \cdot \frac{n(A \cap B)}{n(B)} = P(B) \cdot \frac{n(A \cap B)}{n(B)}$$

Die bedingte Wahrscheinlichkeit $n(A \cap B)/n(B)$ ist die Wahrscheinlichkeit für A unter der Bedingung B , also $P(A|B)$. Der **Multiplikationssatz für zwei Ereignisse** ist dann:

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

Beispiel (aus Schlittgen 1996, S.78): Eine Frau bewahrt Batterien in einer Schachtel auf. Umweltbewußt werden verbrauchte Batterien nicht weggeworfen, sondern ebenfalls in die Schachtel gelegt. Nun werden von ihr zwei Batterien für eine Taschenlampe benötigt. Sie weiß, daß von den sechs Batterien in der Schachtel zwei leer sind.

Frage: Wie groß ist die Wahrscheinlichkeit, daß gleich die ersten Batterien, die sie heraus nimmt, noch unverbraucht sind?

Zuerst erfolgt die Definition der Ereignisse: A: „Die erste Batterie ist gut“; B: „Die zweite Batterie ist gut“. Die Wahrscheinlichkeit, daß die erste Batterie unverbraucht ist, beträgt:

$$P(A) = \frac{4}{6}$$

Die Wahrscheinlichkeit, daß die zweite Batterie noch gut ist, wenn beim ersten Zug eine unverbrauchte herausgenommen wurde, beträgt:

$$P(B|A) = \frac{3}{5}$$

Die gesuchte Wahrscheinlichkeit, daß beide geladen sind, ist die Wahrscheinlichkeit der Ereignisse $P(A \cap B)$:

$$P(A \cap B) = P(A)P(B|A) = \frac{4}{6} \cdot \frac{3}{5} = 0.4$$

2.3.3 Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit

Ein wichtiger Sonderfall in der Beziehung zwischen zwei Ereignissen liegt dann vor, wenn die Ereignisse voneinander **stochastisch unabhängig** sind. Als Beispiel kann man sich z.B. eine Studie vorstellen, die prüfen soll, ob ein Medikament wirksam ist oder nicht. Folgende Ereignisse können definiert werden:

A: „Therapieerfolg“;

B: „Das echte Medikament wird verabreicht“ (kein Plazebo);

Offensichtlich hat das Medikament keine spezifische Wirkung, wenn die Chancen für einen Therapieerfolg in den Gruppen mit und ohne echtem Medikament gleich groß sind:

$$P(A|B) = P(A|\bar{B})$$

Man sagt dann, daß A von B stochastisch unabhängig ist. Folgende Beziehung läßt sich formulieren:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap \bar{B})}{P(\bar{B})} = P(A|\bar{B}) \\ P(A \cap B) \cdot P(\bar{B}) &= P(A \cap \bar{B}) \cdot P(B) \\ P(A \cap B)(1 - P(B)) &= P(A \cap \bar{B})P(B) \\ P(A \cap B) - P(A \cap B)P(B) &= P(A \cap \bar{B})P(B) \\ P(A \cap B) &= P(B)[P(A \cap B) + P(A \cap \bar{B})] \\ &= P(A) \cdot P(B) \end{aligned}$$

stochastische Unabhängigkeit: Zwei Ereignisse A und B heißen stochastisch unabhängig, wenn gilt: $P(A \cap B) = P(A) \cdot P(B)$.

Daraus folgt, daß zwei Ereignisse immer dann unabhängig voneinander sind, wenn $P(A|B) = P(A)$ und $P(B|A) = P(B)$, also wenn die bedingten Wahrscheinlichkeiten gleich den nicht-bedingten Wahrscheinlichkeiten sind.

Bei mehr als zwei Ereignissen liegt es nahe, von Unabhängigkeit zu sprechen, wenn jeweils das Eintreten eines Teils der Ereignisse die Chance für die restlichen Ereignisse nicht verändert. Also die Wahrscheinlichkeiten von Durchschnitten gleich den Produkten der Einzelwahrscheinlichkeiten sind.

Die Ereignisse A_1, A_2, \dots, A_k heißen stochastisch unabhängig, wenn für jede Auswahl $A_{i_1}, A_{i_2}, \dots, A_{i_m}$ mit $m \leq k$ gilt:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_m})$$

Wenn Ereignisse sich gegenseitig ausschließen sind sie stets voneinander abhängig! Die Wahrscheinlichkeit, daß sie gemeinsam auftreten ist Null:

$$P(A \cap B) = 0 = P(A)P(B|A)$$

Da $P(A)$ und $P(B)$ ungleich Null sind, muß $P(B|A) = 0$ sein. Dann ist aber auch $P(B|A) \neq P(B)$.

Beispiel 1: Zwei Studenten lösen in einer Klausur unabhängig voneinander eine Aufgabe. Die Wahrscheinlichkeit, daß Student A die Aufgabe richtig löst, beträgt 0.8. Die für Student B ist 0.7, so daß $P(A) = 0.8$ und $P(B) = 0.7$. Mit welcher Wahrscheinlichkeit wird die Aufgabe von beiden Schülern nicht richtig gelöst?

Gesucht ist die Wahrscheinlichkeit $P(\bar{A} \cap \bar{B})$, die wegen der versprochenen Unabhängigkeit beim richtigen oder unzutreffenden Lösen der Aufgabe leicht zu berechnen ist:

$$P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B}) = 0.2 \cdot 0.3 = 0.06$$

Beispiel 2: Von 100 Männern wird erfragt, ob sie Angeln (A) bzw. nicht angeln (\bar{A}) und/oder kegeln (K) oder nicht kegeln (\bar{K}). Es ergeben sich folgende Häufigkeiten:

$A \cap K = 25$ Angler und Kegler zugleich

$A \cap \bar{K} = 5$ nicht kegelnde Angler

$\bar{A} \cap K = 15$ nicht angelnde Kegler

$\bar{A} \cap \bar{K} = 55$ weder Angler noch Kegler

Man kann dies in Form einer Vierfeldertafeln schreiben:

	K	\bar{K}	Σ
A	25	5	30
\bar{A}	15	55	70
Σ	40	60	100

Angenommen, dies seien Prozente einer bestimmten Grundgesamtheit von Männern. Dann lassen sich jetzt folgende Wahrscheinlichkeiten angeben:

$P(A) = 0.30$ Wahrscheinlichkeit zu angeln

$P(\bar{A}) = 0.70$ Wahrscheinlichkeit nicht zu angeln

$P(K) = 0.40$ Wahrscheinlichkeit zu kegeln

$P(\bar{K}) = 0.60$ Wahrscheinlichkeit nicht zu kegeln

$$P(K|A) = P(K \cap A)/P(A) = 0.25/0.30 = 0.8333$$

$$P(K|\bar{A}) = P(K \cap \bar{A})/P(\bar{A}) = 0.15/0.70 = 0.2143$$

$$P(A|K) = P(A \cap K)/P(K) = 0.25/0.40 = 0.6250$$

$$P(A|\bar{K}) = P(A \cap \bar{K})/P(\bar{K}) = 0.05/0.60 = 0.0833$$

Ein zufällig ausgewählter Mann sei z.B. Angler. Dann ist die bedingte Wahrscheinlichkeit, daß er auch Kegler ist $P(K|A) = 83\%$. Aus $P(K|A) = 0.8333 \neq 0.2143 = P(K|\bar{A})$ bzw. aus $P(K|A) = 0.8333 \neq 0.40 = P(K)$ folgt, daß Angeln und Kegeln voneinander stochastisch abhängig sind.

2.3.4 Satz der totalen Wahrscheinlichkeit

Betrachtet wird ein durch sich paarweise ausschließende Ereignisse A_1, A_2, \dots, A_n ausgeschöpfter Ergebnisraum Ω :

$$A_i \cap A_j = \emptyset \quad \text{für } i \neq j \quad \text{und} \quad A_1 \cup A_2 \cup \dots \cup A_n = \Omega$$

Zudem wird das Ereignis E gebildet, welches ebenfalls durch sich paarweise ausschließende Ereignisse, die Durchschnitte der Ereignisse E mit A_i , zerlegt ist (siehe Abbildung 8).

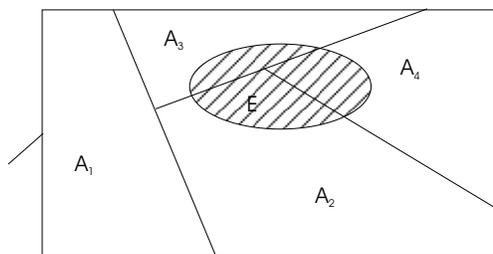


ABBILDUNG 8: Zum Satz der totalen Wahrscheinlichkeit

Die Wahrscheinlichkeit für E ergibt sich mit

$$P(E) = P[(A_1 \cap E) \cup (A_2 \cap E) \cup \dots \cup (A_n \cap E)] = \sum_{i=1}^n P(A_i \cap E)$$

Die Anwendung des Multiplikationssatzes auf die einzelnen Wahrscheinlichkeiten $P(A_i \cap E)$ führt zu dem *Satz der totalen Wahrscheinlichkeit* :

$$P(E) = \sum_{i=1}^n P(E|A_i)P(A_i), \quad \text{wobei} \quad \sum_{i=1}^n P(A_i) = 1 \quad (1)$$

Die Wahrscheinlichkeit für das Ereignis E entspricht der Summe der gewogenen bedingten Wahrscheinlichkeiten $P(E|A_i)$, wobei die Gewichte die Einzelwahrscheinlichkeiten $P(A_i)$ sind.

Beispiel 1: Drei Personen bewerben sich um ein öffentliches Amt. Eine Meinungsumfrage ergebe die individuellen Wahlchancen 0.25, 0.35 und 0.40. Die Chancen, daß die drei nach ihrer Wahl einen Brückenbau durchsetzen, betragen 0.60, 0.90 und 0.80.

Wie groß ist die Wahrscheinlichkeit, daß die Bevölkerung nach der Wahl die Brücke erhält?

Die Wahlchancen sind:

$$P(A_1) = 0.25$$

$$P(A_2) = 0.35$$

$$P(A_3) = 0.40$$

Die Chancen für den Brückenbau nach der Wahl sind:

$$P(E|A_1) = 0.60$$

$$P(E|A_2) = 0.90$$

$$P(E|A_3) = 0.80$$

Dann ist die totale Wahrscheinlichkeit für den Brückenbau:

$$P(E) = \sum_1^3 P(A_i)P(E|A_i) = 0.25 \cdot 0.60 + 0.35 \cdot 0.90 + 0.40 \cdot 0.80 = 0.785$$

Beispiel 2: Ein Gemüsehändler erhalte Karotten aus drei Gärtnereien: 50% stamme aus A_1 , 30% aus A_2 und 20% aus A_3 . Der Händler weiß, daß A_1 1% Ausschuß liefert, A_2 3% und A_3 4%. Wieviel Prozent Ausschuß sind zu erwarten?

Die Wahrscheinlichkeiten für die Herkunft der Karotten sind:

$$P(A_1) = 0.50$$

$$P(A_2) = 0.30$$

$$P(A_3) = 0.20$$

Die Wahrscheinlichkeiten für einen Ausschuß bezogen auf die Gärtnereien sind:

$$P(E|A_1) = 0.01$$

$$P(E|A_2) = 0.03$$

$$P(E|A_3) = 0.04$$

Dann ist die totale Wahrscheinlichkeit für einen Ausschuß:

$$P(E) = \sum_1^3 P(A_i)P(E|A_i) = 0.5 \cdot 0.01 + 0.3 \cdot 0.03 + 0.2 \cdot 0.04 = 0.022$$

2.3.5 Theorem von Bayes

Für die algebraische Herleitung des *Theorem von Bayes* wird neben dem Satz der totalen Wahrscheinlichkeit auch der Multiplikationssatz benötigt. Wir verwenden wieder die Ereignisse A_i und E .

$$P(A_i \cap E) = P(A_i)P(E|A_i) = P(E)P(A_i|E) \quad (2)$$

$$P(A_i|E) = \frac{P(A_i \cap E)}{P(E)} \quad (3)$$

Im Zähler der Gleichung 3 kann der Ausdruck $P(A_i)P(E|A_i)$ eingesetzt werden, und im Nenner wird für E die Gleichung 1 für die totale Wahrscheinlichkeit eingesetzt. Man erhält dann:

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_{i=1}^n P(E|A_i)P(A_i)}, \quad P(E) > 0 \quad (4)$$

Die Bayes'sche Formel stellt einen interessanten Zusammenhang zwischen den bedingten Wahrscheinlichkeiten $P(E|A_i)$ und $P(A_i|E)$ her. Man findet häufig folgende Konzeption:

A_i - sind Tatsachen, Gründe, wahre Ursachen (z.B. A_1 Person hat 'Krebs')
 E - sind Erfahrungsdaten (z.B. Diagnose 'Krebs')

$P(A_i)$ - Wahrscheinlichkeit für eine Tatsache (z.B. $P(A_1)$ Wahrscheinlichkeit für das Vorliegen von Krebs)

$P(E|A_i)$ - Wahrscheinlichkeit für das Eintreten von E unter der Voraussetzung von A_i . Diese Wahrscheinlichkeit wird häufig auch **A-priori-Wahrscheinlichkeit** genannt, weil sie im Vorfeld geschätzt werden muß (z.B. $P(E|A_1)$ Eigenschaft eines Diagnoseverfahrens: Wahrscheinlichkeit, daß bei krebskranken Personen die Diagnose auf 'Krebs' lautet).

$P(A_i|E)$ - Wahrscheinlichkeit, für das Vorliegen der Tatsache A_i unter der Voraussetzung von E . Oder: daß für das Erfahrungsdatum E die Tatsache A_i der Grund war. Sie wird häufig auch **A-posteriori-Wahrscheinlichkeit** genannt, weil sie im nachhinein - also nach dem empirischen Befund - die Wahrscheinlichkeit beschreibt, daß eine entsprechende Tatsache auch wirklich vorliegt (z.B. $P(A_1|E)$ Wahrscheinlichkeit, daß eine Person mit der Diagnose 'Krebs' tatsächlich Krebs hat). In Abbildung 9 ist zur Veranschaulichung ein Pfaddiagramm

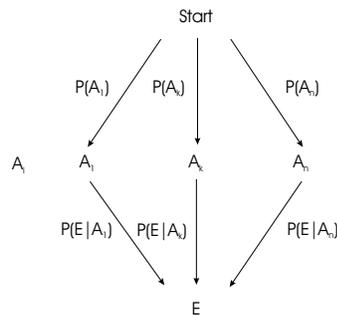


ABBILDUNG 9: Pfaddiagramm

dargestellt. Mit Hilfe der Pfeile und der folgenden Pfadregeln kann die totale Wahrscheinlichkeit als auch die A-posteriori-Wahrscheinlichkeit relativ leicht bestimmt werden:

1. Die Wahrscheinlichkeit eines Pfades ergibt sich als Produkt der Einzelwahrscheinlichkeiten längs des Pfades.
2. Die Wahrscheinlichkeit eines Ereignisses ergibt sich als Summe der Wahrscheinlichkeiten aller Pfade, die zu dem Ereignis führen.

Die Wahrscheinlichkeit, vom **Start** nach E zu gelangen, ist $P(E) = \sum_{i=1}^n P(A_i) \cdot P(E|A_i)$. Wenn man nun über A_k nach E gelangt, ist die Wahrscheinlichkeit $P(A_k|E)$ der Anteilswert:

$$P(A_k|E) = \frac{P(A_k) \cdot P(E|A_k)}{\sum_{i=1}^n P(A_i) \cdot P(E|A_i)}$$

Zur Erläuterung des Bayes'schen Theorems folgen ein paar Beispiele.

Beispiel 1: Das Bayes'sche Theorem wird des öfteren für die Prüfung der Zuverlässigkeit von medizinischen Diagnosen verwendet. Ein typisches Beispiel wäre folgendes Problem: Es lassen sich drei Ereignisse angeben:

E - „Diagnose 'Krebs'“
 A_1 - „Person hat Krebs“

A_2 - „Person hat keinen Krebs“

Die Wahrscheinlichkeiten für die Tatsache 'Krebs' bzw. 'kein Krebs' wird über die relativen Häufigkeiten angegeben:

$$P(A_1) = 0.002$$

$$P(A_2) = 0.998$$

Die Eigenschaften des Diagnoseverfahrens führt zu folgenden bedingten Wahrscheinlichkeiten:

$P(E|A_1) = 0.88$ Wahrscheinlichkeit dafür, daß bei krebskranken Personen die Diagnose auf 'Krebs' lautet.

$P(E|A_2) = 0.05$ Wahrscheinlichkeit dafür, daß bei nicht-krebskranken Personen die Diagnose auf 'Krebs' lautet.

Gefragt wird nun nach $P(A_1|E)$, der Wahrscheinlichkeit, daß eine Person mit der Diagnose 'Krebs' tatsächlich auch Krebs hat. Die algebraische Lösung hierfür liefert dann das Bayes'sche Theorem:

$$P(A_1|E) = \frac{P(E|A_1)P(A_1)}{P(E|A_1)P(A_1) + P(E|A_2)P(A_2)} = \frac{0.88 \cdot 0.002}{0.88 \cdot 0.002 + 0.05 \cdot 0.998} = 0.034$$

Die mit diesem Verfahren als krank identifizierten Personen sind mit einer Wahrscheinlichkeit von nur 3.4% wirklich krank! Hier macht sich die geringe Prävalenz, also die geringe Zahl der Krebserkrankten in der Bevölkerung bemerkbar ($P(A_1) = 0.002$).

Beispiel 2: Auf dem Flughafen werden alle Passagiere vorsorglich kontrolliert. Ein Terrorist werde mit einer Wahrscheinlichkeit von 0.98 festgenommen, ein Nicht-Terrorist mit 0.001. Jeder hunderttausendste Flugpassagier sei ein Terrorist ($P(A_1) = 0.00001$). Wie groß ist die Wahrscheinlichkeit, daß eine Festnahme tatsächlich einen Terroristen erbringt?

$P(A_1) = 0.00001$ Wahrscheinlichkeit für Terrorist

$P(A_2) = 1 - P(A_1) = 0.99999$ Wahrscheinlichkeit für Kein-Terrorist

$P(E|A_1) = 0.98$ Wahrscheinlichkeit für eine Festnahme unter der Bedingung 'Terrorist'

$P(E|A_2) = 0.001$ Wahrscheinlichkeit für eine Festnahme unter der Bedingung 'Nicht-Terrorist'

gesucht ist $P(A_1|E) = ?$

$$P(A_1|E) = \frac{P(E|A_1)P(A_1)}{P(E|A_1)P(A_1) + P(E|A_2)P(A_2)} = \frac{0.98 \cdot 0.00001}{0.98 \cdot 0.00001 + 0.001 \cdot 0.99999} = 0.0097$$

Trotz der Zuverlässigkeit der Kontrollen erfolgen somit über 99% aller Festnahmen zu Unrecht.

3 Zufallsvariable und Wahrscheinlichkeitsverteilung

Die Ergebnisse von Experimenten, Befragungen und Beobachtungen werden i.d.R. durch Zahlen repräsentiert. Die Zuordnung der Zahlen zu den möglichen Ergebnissen geschieht dann mittels geeigneter Variablen. Im Falle von Zufallsexperimenten spricht man unter Berücksichtigung des Generierungsprozesses der Ergebnisse von Zufallsvariablen.

3.1 Zufallsvariable

Ganz allgemein bezeichnet man eine Variable X , die jedem möglichen Ergebnis ω eines Zufallsexperiments eine (reelle) Zahl zuordnet, als Zufallsvariable. Die möglichen Ergebnisse ω_i bilden dabei den endlichen Ergebnisraum Ω .

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$$

Formal läßt sich dann eine Zufallsvariable als Funktion bzw. Abbildung definieren:

$$X : \Omega \rightarrow \tilde{X} \subseteq \mathfrak{R}$$

Jedem Ergebnis ω_i in dem Ergebnisraum Ω wird ein Zahlenwert $X(\omega) \in \tilde{X}$ zugeordnet. \tilde{X} ist also der Merkmalsraum, innerhalb dessen die Zufallsvariable X ihre Werte annehmen kann. Zur numerischen Repräsentation ist der Merkmalsraum in die Menge der reellen Zahlen eingebettet.

Formal besteht damit eine Analogie zu der Konzeption der „statistischen Variablen“. Es ist allerdings wichtig, daß diese beiden Begriffe unterschieden werden (vgl. z.B. Rohwer 2000, S.83). Statistische Variablen beziehen sich auf Objekte und Situationen in unserer Erfahrungswelt, während sich Zufallsvariablen allein auf eine Ergebnismenge von Zufallsexperimenten bezieht. Allerdings kann diese Ergebnismenge dazu genutzt werden, um Hypothesen über Ereignisse oder Sachverhalte zu formulieren.

Beispiel: Eine Münze wird dreimal geworfen. Die möglichen Ergebnisse der Münze sind 'Wappen' oder 'Kopf'. Insgesamt sind acht Ergebnisse möglich. Diesen Ergebnissen können nun Zahlen zugeordnet werden, die jeweils die Anzahlen für das Ergebnis 'Kopf' anzeigen. Die zugeordneten Zahlen sollen Werte einer Zufallsvariablen X sein. Die Zuordnung kann dann tabellarisch dargestellt werden:

Ergebnisraum Ω	$X(\omega)$
$\omega_1 = WWW$	$x_1 = 0$
$\omega_2 = WWK$	$x_2 = 1$
$\omega_3 = WKW$	$x_2 = 1$
$\omega_4 = KWW$	$x_2 = 1$
$\omega_5 = WKK$	$x_3 = 2$
$\omega_6 = KWK$	$x_3 = 2$
$\omega_7 = KKW$	$x_3 = 2$
$\omega_8 = KKK$	$x_4 = 3$

3.2 Wahrscheinlichkeitsverteilung

Bei der Betrachtung von Zufallsvariablen sind die Wahrscheinlichkeiten der durch sie festgelegten Ereignisse von zentraler Bedeutung.

Wahrscheinlichkeitsverteilung: Die Wahrscheinlichkeitsverteilung einer Zufallsvariablen X ist die Zuordnung der Wahrscheinlichkeiten zu allen durch X festgelegten Ereignissen. Analog dem Vorgehen in der deskriptiven Statistik wird zwischen diskreten und stetigen Variablen unterschieden.

3.2.1 Wahrscheinlichkeitsfunktion bei diskreten Zufallsvariablen

Diskrete Variablen sind dadurch gekennzeichnet, daß die Realisierungsmöglichkeit eine 'diskrete' oder abzählbare Menge bildet. Die einzelnen Werte kommen bei wiederholter Beobachtung in der Regel mehrmals vor.

Definition: Die Funktion $f(x)$, die jeder reellen Zahl x die Wahrscheinlichkeit $P(X = x)$ zuordnet, heißt **Wahrscheinlichkeitsfunktion** der Zufallsvariablen X :

$$f(x) = P(X = x)$$

Sind $x_1, x_2, \dots, x_i, \dots$ die Realisierungsmöglichkeiten der diskreten Zufallsvariablen X , so wird die Wahrscheinlichkeitsfunktion i.d.R. in der Form

$$f(x_i) = P(X = x_i) = p_i \quad i = 1, 2, \dots$$

angegeben.

Die Wahrscheinlichkeitsfunktion ist für alle reellen Zahlen definiert. Ist aber x keine Realisierungsmöglichkeit von X so ist $\{X = x\} = \emptyset$ und es gilt

$$f(x) = P(X = x) = 0$$

Die Festlegung der Wahrscheinlichkeitsverteilung kann bei diskreten Zufallsvariablen durch die Angabe aller einzelnen Wahrscheinlichkeiten $P(X = x)$ erfolgen. Für zusammengesetzte Ereignisse können entsprechende Summen gebildet werden. Wenn x_1, \dots, x_k die Realisierungsmöglichkeiten mit $x_i \leq x$ sind:

$$\{X \leq x\} = \{X = x_1\} \cup \{X = x_2\} \cup \dots \cup \{X = x_k\}$$

Für das Wahrscheinlichkeitsmaß gilt dann:

$$P(X \leq x) = P(X = x_1) + P(X = x_2) + \dots + P(X = x_k)$$

Wenn die Zufallsvariable X die Realisierungsmöglichkeit x_i , $i = 1, 2, 3, \dots$, so gilt

$$\begin{aligned} \sum_i p_i &= \sum_i P(X = x_i) = P(\{X = x_1\} \cup \{X = x_2\} \cup \dots) \\ &= 1 \end{aligned}$$

Die Summe aller p_i ist stets 1.

Verteilungsfunktion: Die Verteilungsfunktion einer Zufallsvariablen X heißt die Funktion $F(x)$, die jedem x die Wahrscheinlichkeit $P(X \leq x)$ zuordnet:

$$F(x) = P(X \leq x)$$

Die Verteilungsfunktion $F(x)$ einer Zufallsvariablen X hat dann folgende *Eigenschaften*:

1. $F(x)$ nimmt nur Werte zwischen 0 und 1 an, $0 \leq F(x) \leq 1$ für alle $x \in \mathfrak{R}$
2. $F(x)$ steigt für wachsendes x monoton an (oder bleibt zumindest auf gleicher Höhe):
 $x_1 < x_2 \rightarrow F(x_1) \leq F(x_2)$
3. $F(x) \rightarrow 1$ für $x \rightarrow \infty$
4. $F(x) \rightarrow 0$ für $x \rightarrow -\infty$

Die Verteilungsfunktion $F(x)$ einer diskreten Zufallsvariablen X ist eine Treppenfunktion: Sie weist bei den Realisierungsmöglichkeiten x_i Sprünge der Höhe $P(X = x_i)$ auf. Zwischen der Verteilungsfunktion $F(x)$ und der Wahrscheinlichkeitsfunktion $f(x)$ der diskreten Zufallsvariablen X mit den Realisierungsmöglichkeiten $x_1, x_2, \dots, x_i, \dots$ bestehen folgende Zusammenhänge:

$$\begin{aligned}
 (1) \quad F(x) &= \sum_{x_i \leq x} f(x_i) \\
 (2) \quad F(x_i) - F(x_{i-1}) &= f(x_i) \\
 (3) \quad F(x_i) &= P(X \leq x_i) = P(X < x_i) + P(X = x_i) \\
 &F(x_i) = F(x_{i-1}) + P(X = x_i)
 \end{aligned}$$

Beispiel: Dreifacher Münzwurf In einem Zufallsexperiment wird wieder eine Münze dreimal geworfen. Die möglichen Ergebnisse eines Wurfs sind 'Wappen' oder 'Kopf'. Es wird dann eine Zufallsvariable definiert, die jedem Ergebnis ω_i eine Zahl mit der Anzahl der 'Kopf' - Würfe zuweist. In der nachstehenden Tabelle sind jedem Ergebnis die Werte der Zufallsvariablen zugeordnet worden. Zudem wird über eine Wahrscheinlichkeitsfunktion jedem Wert der Zufallsvariablen die Wahrscheinlichkeit $P(X = x)$ zugewiesen. In der letzten Spalte ist die Verteilungsfunktion $F(x) = \sum_{x_i \leq x} f(x_i)$ angegeben.

Ergebnisraum Ω	Werte $X(\omega)$ der Zufallsvariablen X	Wahrscheinlichkeitsfunktion $f(x)$	Verteilungsfunktion $F(x)$
$\omega_1 = WWW$	$x_1 = 0$	$P(X = 0) = P(0) = 1/8$	$F(0) = P(X \leq 0) = 1/8$
$\omega_2 = WWK$	$x_2 = 1$	$P(X = 1) = P(1) = 3/8$	$F(1) = P(X \leq 1) = 4/8$
$\omega_3 = WKW$	$x_2 = 1$	$P(X = 2) = P(2) = 3/8$	$F(2) = P(X \leq 2) = 7/8$
$\omega_4 = KWW$	$x_2 = 1$	$P(X = 3) = P(3) = 1/8$	$F(3) = P(X \leq 3) = 1$
$\omega_5 = WKK$	$x_3 = 2$		
$\omega_6 = KWK$	$x_3 = 2$		
$\omega_7 = KKW$	$x_3 = 2$		
$\omega_8 = KKK$	$x_4 = 3$		

Während die Wahrscheinlichkeitsverteilung im Falle einer diskreten Variablen in der Regel durch ein Stabdiagramm dargestellt wird, lässt sich die Verteilungsfunktion in Form einer Treppenfunktion abbilden (siehe hierzu Abbildung 10).

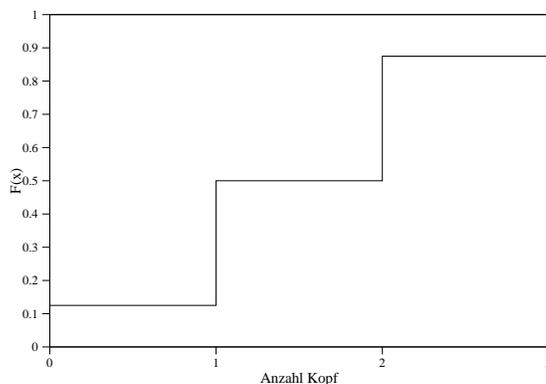


ABBILDUNG 10: Verteilungsfunktion

3.2.2 Wahrscheinlichkeitsfunktion bei stetigen Zufallsvariablen

Bei stetigen Variablen ist jede Zahl aus einem Intervall eine Realisierungsmöglichkeit, d.h. sie kann quasi überabzählbar viele Werte annehmen. In der deskriptiven Statistik erfolgt die Darstellung der Verteilung einer Variablen im Falle von empirischen Datensätzen in der Regel durch

ein Histogramm. Entsprechend dem Prinzip der *Flächentreue* ist der Flächeninhalt der Blöcke dabei proportional zur relativen Häufigkeit der jeweiligen Klassen. Bei einem sehr großen Stichprobenumfang kann das Histogramm durch eine glatte Kurve approximiert werden. Eine durch zwei Punkte a und b begrenzte Fläche unter dieser Kurve kann dann als Wahrscheinlichkeit angesehen werden, einen Wert aus diesem Intervall zu erhalten.

Die Wahrscheinlichkeit des Intervalls $(a, b]$ wird demnach formal durch ein Integral beschrieben: $P(a < X \leq b) = \int_a^b f(x)dx$.

Definition: Wahrscheinlichkeitsdichte: X sei eine Zufallsvariable. Es gebe eine Funktion $f(x)$, so daß

$$P(a < X \leq b) = \int_a^b f(x)dx \quad \text{für alle } a, b \text{ mit } a \leq b$$

Dann heißt die Funktion $f(x)$ Wahrscheinlichkeitsdichte der Verteilung von X .

Bei der Interpretation ist zu beachten, daß im Gegensatz zur Wahrscheinlichkeitsfunktion bei diskreten Zufallsvariablen, die Dichtefunktion bei einer stetigen Zufallsvariablen an einer Stelle x nicht direkt einen Wahrscheinlichkeitswert angibt. Nur die über Intervalle $(a, b]$ abgegrenzten Flächen unterhalb der Dichtefunktion entsprechen den Wahrscheinlichkeiten $P(a < X \leq b)$. Um die Fläche unterhalb einer Dichtefunktion als Wahrscheinlichkeit interpretieren zu können, müssen zudem folgende Eigenschaften erfüllt sein:

$$(1) \quad f(x) \geq 0$$

$$(2) \quad \int_{-\infty}^{+\infty} f(x)dx = 1$$

Der Flächeninhalt darf nie negativ sein und die Gesamtfläche unterhalb der Dichtefunktion muß den Inhalt 1 haben, denn durch $\{-\infty < X < +\infty\}$ wird das sichere Ereignis beschrieben.

Beispiel (aus Schlittgen 1998, S.98): Herr Jedermann geht ohne auf die Uhr zu sehen zur S-Bahn. Die S-Bahn fährt im 20-Minuten-Takt und Verspätungen sowie zu frühes Kommen werden ausgeschlossen. Die Zufallsvariable „ X = Wartezeit in Minuten“ kann dann jeden Wert aus dem Intervall $[0; 20]$ annehmen. Über diese Annahmen gelangt man dann zu:

$$P(0 \leq X \leq 20) = 1$$

Aufgrund des unkoordinierten Vorgehens erscheint es zudem plausibel, gleich langen Teilintervallen gleiche Wahrscheinlichkeiten zuzuordnen. Also ist die Wahrscheinlichkeit eine Wartezeit zwischen a und b , $0 < a < b < 20$ zu erhalten, proportional zu der Länge des Intervalls $[a; b]$:

$$P(a \leq X \leq b) = k \cdot (b - a)$$

Dann gilt:

$$k \cdot (b - a) = \int_a^b k dx$$

Die Dichtefunktion von X bleibt daher über dem Intervall $[0; 20]$ konstant, bzw. $f(x) = k$. Mit

$$P(0 \leq X \leq 20) = k \cdot 20 = 1 \quad \text{folgt } k = \frac{1}{20}$$

Vollständig beschrieben lautet dann die Dichtefunktion:

$$f(x) = \begin{cases} 1/20 & 0 < x \leq 20 \\ 0 & \text{sonst.} \end{cases}$$

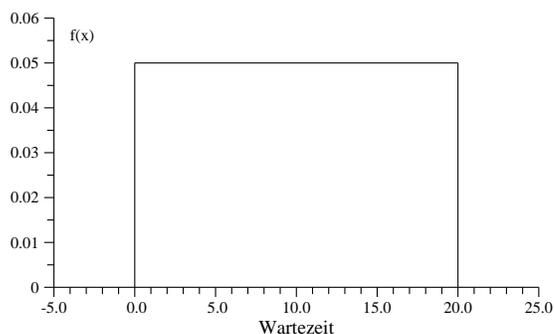


ABBILDUNG 11: Dichtefunktion von $X = \text{'Wartezeit'}$

Verteilungsfunktion: Im Falle einer stetigen Zufallsvariablen X , welche eine Dichte $f(x)$ besitzt, werden Wahrscheinlichkeiten für Ereignisse der Form $\{a < X \leq b\}$ durch Flächeninhalte repräsentiert. Somit gilt:

$$F(b) = P(X \leq b) = P(-\infty < X \leq b)$$

Der Wert der Verteilungsfunktion an der Stelle b , also $F(b)$ ist dann der Inhalt der links von b durch $f(x)$ begrenzten Fläche. Somit ist dies das Integral der Funktion $f(x)$ von $-\infty$ bis b .

Zwischen der Dichtefunktion $f(x)$ und der Verteilungsfunktion $F(x)$ einer stetigen Zufallsvariablen X besteht dann folgender Zusammenhang:

$$(1) \quad F(b) = \int_{-\infty}^b f(x) dx$$

$$(2) \quad \frac{dF(x)}{dx} = f(x)$$

Der Wert der Verteilungsfunktion $F(x)$ an der Stelle $x = b$ entspricht dem Integral der Dichtefunktion bis b . Und die erste Ableitung der Verteilungsfunktion $F(x)$ ist die Dichtefunktion $f(x)$.

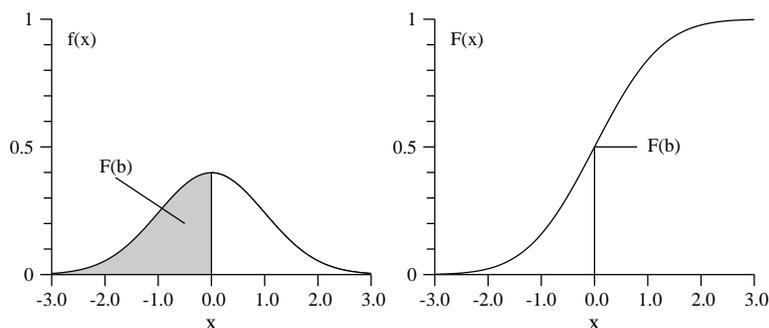


ABBILDUNG 12: Zusammenhang von Dichte- und Verteilungsfunktion

Bezogen auf das obige Beispiel erhält man durch Integration die Verteilungsfunktion:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{20}x & 0 \leq x \leq 20 \\ 1 & \text{sonst.} \end{cases}$$

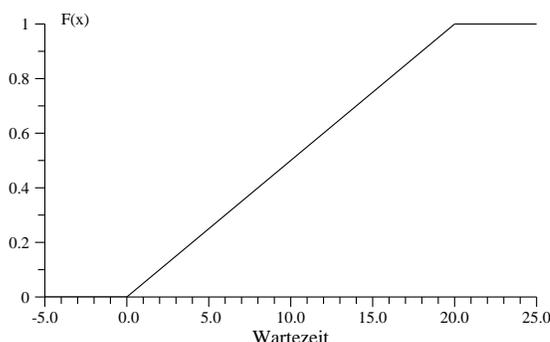


ABBILDUNG 13: Verteilungsfunktion von $X = \text{'Wartezeit'}$

In dem beschriebenen Vorgehen wurde auf Basis von Annahmen ein entsprechendes Verteilungsmodell entwickelt. Sehr häufig findet man aber auch Anwendungen, in denen eine geeignete Dichtefunktion für eine schon vorliegende empirische Verteilung gesucht wird, die diese approximiert.

3.3 Mehrdimensionale Zufallsvariablen

Werden in einem Zufallsexperiment nicht nur eine sondern gleichzeitig mehrere Variablen betrachtet, die den gleichen Ergebnisraum betreffen, wird konzeptionell analog der deskriptiven Statistik im Falle der simultanen Betrachtung von empirischen Variablen vorgegangen.

3.3.1 Gemeinsame Verteilung bei zwei diskreten Zufallsvariablen

Bei zwei **diskreten Zufallsvariablen** X und Y läßt sich die gemeinsame Verteilung in Anlehnung an relative Häufigkeiten z.B. in einer Vier-Felder-Tabelle, durch die Wahrscheinlichkeiten

$$f(x_i, y_j) = p_{ij} = P(X = x_i, Y = y_j)$$

angeben, wobei die Ereignisse wie folgt definiert sind:

$$\{X = x_i, Y = y_j\} = \{\omega \mid \omega \in \Omega, X(\omega) = x_i \text{ und } Y(\omega) = y_j\}$$

Die Funktion, die dann bei zwei Zufallsvariablen X und Y jedem Paar von Realisierungsmöglichkeiten (x_i, y_j) die Wahrscheinlichkeit $P(X = x_i, Y = y_j) = p_{ij}$ zuordnet, wird als **gemeinsame Wahrscheinlichkeitsfunktion** bezeichnet.

Beispiel: Zweifacher Würfelwurf Es werden zwei Würfel gleichzeitig geworfen, der Ergebnisraum ist dann:

$$\Omega = \{(i, j) \mid i, j = 1, 2, 3, 4, 5, 6\}$$

Mit i wird die Augenzahl des ersten und mit j die Augenzahl des zweiten Würfels bezeichnet. Somit sind zwei Zufallsvariablen X und Y auf Ω definiert:

$$\begin{aligned} X : \Omega &\rightarrow \tilde{X} \subseteq \mathfrak{R} \\ \omega = (i, j) &\rightarrow X(\omega) = i \end{aligned}$$

sowie:

$$Y : \Omega \rightarrow \tilde{Y} \subseteq \mathfrak{R}$$

$$\omega = (i, j) \rightarrow Y(\omega) = j$$

Wenn man das Gleichmöglichkeitsmodell voraussetzt, gilt:

$$P(X = i, Y = j) = P(\{(i, j)\}) = \frac{1}{36}$$

Analog den relativen Häufigkeiten läßt sich aus der tabellarisch angegebenen gemeinsamen Wahrscheinlichkeitsfunktion zweier Zufallsvariablen die Randverteilungen durch eine zeilen- und spaltenweise Summation der gemeinsamen Wahrscheinlichkeiten bestimmen (siehe hierzu Tabelle 3).

TABELLE 3: Kreuztabelle für gemeinsame Wahrscheinlichkeitsverteilung

Zufalls- variable X	Zufallsvariable Y								$\sum_{j=1}^m p_{ij}$
	y_1	y_2	y_3	y_4	\dots	y_j	\dots	y_m	
x_1	p_{11}	p_{12}	p_{13}	p_{14}	\dots	p_{1j}	\dots	p_{1m}	$p_{1.}$
x_2	p_{21}	p_{22}	p_{23}	p_{24}	\dots	p_{2j}	\dots	p_{2m}	$p_{2.}$
x_3	p_{31}	p_{32}	p_{33}	p_{34}	\dots	p_{3j}	\dots	p_{3m}	$p_{3.}$
x_4	p_{41}	p_{42}	p_{43}	p_{44}	\dots	p_{4j}	\dots	p_{4m}	$p_{4.}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_i	p_{i1}	p_{i2}	p_{i3}	p_{i4}	\dots	p_{ij}	\dots	p_{im}	$p_{i.}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_k	p_{k1}	p_{k2}	p_{k3}	p_{k4}	\dots	p_{kj}	\dots	p_{km}	$p_{k.}$
$\sum_{i=1}^k p_{ij}$	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{.4}$	\dots	$p_{.j}$	\dots	$p_{.m}$	1

Es gilt also:

$$\sum_{j=1}^m p_{ij} = p_{i.} \quad \text{sowie} \quad \sum_{i=1}^k p_{ij} = p_{.j} \quad \text{und} \quad \sum_i p_{i.} = \sum_j p_{.j} = 1$$

Für das Beispiel „Zweifacher Würfelwurf“ erhält man dann folgende Tabelle:

TABELLE 4: Gemeinsame Wahrscheinlichkeitsverteilung beim zweifachen Würfelwurf

Zufalls- variable X	Zufallsvariable Y						$\sum_{j=1}^6 p_{ij}$
	1	2	3	4	5	6	
1	1/36	1/36	1/36	1/36	1/36	1/36	6/36
2	1/36	1/36	1/36	1/36	1/36	1/36	6/36
3	1/36	1/36	1/36	1/36	1/36	1/36	6/36
4	1/36	1/36	1/36	1/36	1/36	1/36	6/36
5	1/36	1/36	1/36	1/36	1/36	1/36	6/36
6	1/36	1/36	1/36	1/36	1/36	1/36	6/36
$\sum_{i=1}^6 p_{ij}$	6/36	6/36	6/36	6/36	6/36	6/36	1

Allgemein kann die Wahrscheinlichkeitsfunktion zweier diskreter Zufallsvariablen X und Y mit $f(x, y) = P(X = x, Y = y)$ angegeben werden. Die Wahrscheinlichkeitsfunktion der **Randverteilung** von X ist dann die Summe über die Realisierungsmöglichkeiten von Y :

$$P(X = x) = f(x) = \sum_y f(x, y)$$

Die Wahrscheinlichkeitsfunktion der **bedingten Verteilung** von X bei gegebenem $\{Y = y\}$, wobei $f(y) = P(Y = y) > 0$, ist:

$$f(x|y) = P(X = x|Y = y) = \frac{f(x, y)}{f(y)}$$

Für die Randverteilung und die bedingte Verteilung von Y bei gegebenem $\{X = x\}$ gilt entsprechendes.

3.3.2 Gemeinsame Verteilung bei zwei stetigen Zufallsvariablen

Im Falle von zwei stetigen Zufallsvariablen X und Y ist die gemeinsame Dichte $f(x, y)$ wie folgt definiert (vgl. z.B. Schlittgen 1998, S.110):

gemeinsame Dichte: Das durch die Funktion $f(x, y)$ begrenzte Volumen über einem Rechteck ist gleich der Wahrscheinlichkeit, daß X und Y Werte annehmen, so daß (x, y) in diesem Rechteck liegt.

Entsprechend dem diskreten Fall sind dann die Randdichten und bedingten Dichten definiert:

Randdichte: X und Y seien zwei Zufallsvariablen mit der gemeinsamen Dichtefunktion $f(x, y)$. Die Randdichten von X und Y sind dann

$$f(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad \text{sowie} \quad f(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

anstelle der Summen im diskreten Fall stehen nun Integrale.

bedingte Dichte: Die bedingte Dichte von X bei gegebenen $\{Y = y\}$ ist für den Fall $f(y) > 0$ wie folgt

$$f(x|y) = \frac{f(x, y)}{f(y)}$$

Analog dazu läßt sich auch $f(y|x)$ definieren:

$$f(y|x) = \frac{f(x, y)}{f(x)}$$

für den Fall $f(x) > 0$.

3.4 Unabhängigkeit zweier Zufallsvariablen

Das Konzept der Unabhängigkeit bei Ereignissen läßt sich auch auf Zufallsvariablen übertragen. Damit zwei Variablen X und Y unabhängig voneinander sind, ist erforderlich, daß alle durch X festgelegten Ereignisse unabhängig sind von denen, die durch Y beschrieben werden. Insofern läßt sich folgende Definition angeben:

Unabhängigkeit: Zwei Zufallsvariablen X und Y mit der gemeinsamen Dichte oder Wahrscheinlichkeitsfunktion $f(x, y)$ sind unabhängig, wenn für alle x und y gilt:

$$f(x, y) = f(x) \cdot f(y)$$

wobei $f(x)$ die Dichte bzw. Wahrscheinlichkeitsfunktion von X ist und $f(y)$ die von Y .

3.5 Erwartungswert und Varianz

In der deskriptiven Statistik wurde gezeigt, daß eine statistische Variable durch ihre Verteilung sowie deren Mittelwert und Streuung beschrieben werden kann. Analog hierzu lassen sich auch für Zufallsvariablen Lageparameter und Varianz berechnen.

3.5.1 Erwartungswert

Der Begriff des *Erwartungswertes* stammt aus der Frühphase der Wahrscheinlichkeitsrechnung, in der man sich in erster Linie mit Glücksspielen beschäftigt hat. So kann man den Erwartungswert als denjenigen Betrag auffassen, von dem man erwarten kann, daß man ihn bei einer häufigen Wiederholung eines Spiels im Durchschnitt gewinnen oder verlieren wird. In diesem Sinne kann man den Erwartungswert einer Zufallsvariablen als den zu erwartenden Durchschnittswert einer größeren Anzahl von Realisierungen der Zufallsvariablen auffassen (vgl. Rohwer 2000, S.101).

Erwartungswert bei einer diskreten Zufallsvariablen: Analog einer wiederholt beobachteten diskreten statistischen Variablen, bei der sich der Mittelwert auf Basis der empirischen Häufigkeitsverteilung durch $\bar{x} = \sum_i^k x_i \cdot h_i$ ergibt, läßt sich der *Erwartungswert* einer Zufallsvariablen X mit der Wahrscheinlichkeitsfunktion $p_i = P(P = x_i); i = 1, 2, 3, \dots$ ermitteln.

$$E(X) = \mu = \sum_i x_i p_i$$

Beispiel: Dreifacher Münzwurf und Häufigkeit für „Kopf“ (diskrete Zufallsvariable)

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1 \frac{1}{2}$$

Erwartungswert bei einer stetigen Zufallsvariablen: Im Fall einer stetigen Zufallsvariablen X mit einer Dichtefunktion $f(x)$ ist der Erwartungswert gegeben mit:

$$E(X) = \mu = \int_{-\infty}^{+\infty} x f(x) dx$$

Beispiel: Betrachtet wird wieder das vorhergehende Beispiel der Zufallsvariable X „Wartezeit auf den nächsten S-Bahn Zug“ mit der Gleichverteilung über dem Intervall $[0; 20]$ bzw. mit der Dichte:

$$f(x) = \begin{cases} 1/20 & 0 \leq x \leq 20 \\ 0 & \text{sonst.} \end{cases}$$

Der Erwartungswert ist dann:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{20} x \frac{1}{20} dx = \frac{1}{20} \left[\frac{1}{2} x^2 \right]_0^{20} = \frac{1}{20} \left[\frac{1}{2} 20^2 - \frac{1}{2} 0^2 \right] = \frac{20}{2} = 10$$

Herr Jedermann muß also im Durchschnitt mit einer Wartezeit von 10 Minuten rechnen.

3.5.2 Varianz einer Zufallsvariablen

Neben dem Mittelwert kann auch die Varianz einer Zufallsvariablen in Analogie zum Begriff der Varianz einer statistischen Variablen definiert werden. Anstelle der Häufigkeitsfunktion wird bei Zufallsvariablen die Wahrscheinlichkeitsfunktion bzw. Dichtefunktion verwendet.

Varianz einer Zufallsvariablen X sei eine Zufallsvariable mit der Wahrscheinlichkeitsfunktion $p_i = P(X = x_i)$, wobei $i = 1, 2, \dots$ bzw. der Dichte $f(x)$. Der Erwartungswert von X sei $E(X) = \mu$. Die Varianz von X bzw. von der Verteilung von X ist dann:

$$\text{Var}(X) = \sigma^2 = \sum_i (x_i - E(X))^2 p_i = \sum_i (x_i - \mu)^2 p_i \quad \text{falls } X \text{ diskret ist;}$$

$$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \quad \text{falls } X \text{ stetig ist;}$$

Die Standardabweichung von X ist $\sigma = \sqrt{\sigma^2}$.

Analog der empirischen Varianz einer statistischen Variablen ist die Varianz einer Zufallsvariablen die erwartete quadratische Abweichung der Variablen X von ihrem Erwartungswert $E(X)$:

$$\text{Var}(X) = E((X - \mu)^2)$$

und läßt sich auch wie folgt berechnen:

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Beispiel 1: Bei einem dreimaligen Münzwurf werden wieder die „Kopfwürfe“ als Werte einer Zufallsvariablen angesehen. Mit $E(X) = 1.5$ ist die Varianz:

$$\text{Var}(X) = \sum_i (x_i - E(X))^2 p_i = (0 - 1.5)^2 \cdot \frac{1}{8} + (1 - 1.5)^2 \cdot \frac{3}{8} + (2 - 1.5)^2 \cdot \frac{3}{8} + (3 - 1.5)^2 \cdot \frac{1}{8} = 0.75$$

Beispiel 2: Wir verwenden wieder das S-Bahn Beispiel und als Zufallsvariable die Wartezeit X mit dem Intervall $[0; 20]$ und den schon weiter oben berechneten Erwartungswert von X :

$$\mu = E(X) = 10$$

Die Varianz berechnet sich dann aus:

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = \int_0^{20} (x - 10)^2 \cdot \frac{1}{20} dx \\ &= \frac{1}{20} \int_0^{20} (x^2 - 20x + 100) dx \\ &= \frac{1}{20} \left[\frac{1}{3} x^3 - \frac{1}{2} 20x^2 + 100x \right]_0^{20} \\ &= \frac{1}{20} \left[\frac{1}{3} 20^3 - \frac{1}{2} \cdot 20 \cdot 20^2 + 100 \cdot 20 \right] \\ &= \frac{1}{3} 400 - \frac{1}{2} \cdot 400 + 100 \\ &= 33.33 \end{aligned}$$

Als Standardabweichung erhält man $\sigma = \sqrt{\text{Var}(X)} = 5.77$.

Eigenschaften des Erwartungswertes und der Varianz einer Zufallsvariablen Die Eigenschaften des arithmetischen Mittels und der empirischen Varianz lassen sich übertragen:

X und Y seien zwei Zufallsvariablen mit den Erwartungswerten $E(X)$ und $E(Y)$ und den Varianzen $Var(X)$ und $Var(Y)$. Es gilt:

TABELLE 5: *Eigenschaften von Erwartungswerten und Varianzen von Zufallsvariablen*

Erwartungswert	Varianz
$E(a) = a; a \in \mathfrak{R}$	$Var(a) = 0$
$E(X + a) = E(X) + a$	$Var(X + a) = Var(X)$
$E(a \cdot X) = aE(X)$	$Var(a \cdot X) = a^2 \cdot Var(X)$
$E(X + Y) = E(X) + E(Y)$	
$E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i)$	

wenn die Zufallsvariablen X u. Y unabhängig sind:

$$E(X \cdot Y) = E(X) \cdot E(Y) \quad | \quad Var(X + Y) = Var(X) + Var(Y)$$

wenn die Zufallsvariablen X_i paarweise unabhängig sind:

$$| \quad Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$$

Standardisierung einer Zufallsvariablen Die Varianz und der Erwartungswert einer Zufallsvariablen kann dazu verwendet werden, die Zufallsvariable zu **standardisieren**:

$$K = \frac{X - E(X)}{+\sqrt{Var(X)}}$$

Hier wird die Differenz zwischen jedem Wert der Zufallsvariablen X und ihrem Erwartungswert in Einheiten ihrer Standardabweichung ausgedrückt. Der Erwartungswert $E(K)$ ist dann 0 und die Varianz 1. Dies ist vorteilhaft, da die Werte der Verteilungsfunktion einer standardisierten Zufallsvariablen in Tabellenform vorliegen, und damit die Durchführung von Tests und Schätzungen erleichtert wird.

$$E(K) = E\left[\frac{X - E(X)}{+\sqrt{Var(X)}}\right] = \frac{E[X - E(X)]}{+\sqrt{Var(X)}} = \frac{E(X) - E(X)}{+\sqrt{Var(X)}} = 0$$

$$Var(K) = Var\left[\frac{X - E(X)}{+\sqrt{Var(X)}}\right] = \frac{Var[X - E(X)]}{Var(X)} = \frac{Var(X)}{Var(X)} = 1$$

Beispiel: Es wird wiederum eine Münze dreimal geworfen und eine Zufallsvariable X soll dabei die Werte für die Anzahl der „Kopf“-Würfe annehmen. Für den Erwartungswert ergab sich dabei $E(X) = 1\frac{1}{2}$ und die Varianz ist $Var(X) = 0.75$. Die standardisierten Werte von X werden dann mit $k = (x - 1.5)/\sqrt{0.75}$ berechnet und sind in der nachstehenden Tabelle aufgeführt.

x	0	1	2	3
k	-1.732	-0.577	0.577	1.732
p_i	1/8	3/8	3/8	1/8

4 Ausgewählte Wahrscheinlichkeitsverteilungen

Sogenannte theoretische Verteilungen werden häufig an empirische Verteilungen angepaßt. Ein Grund dafür ist, daß in vielen Fällen theoretische Verteilungen gut geeignet sind, eine große Anzahl an empirischen Verteilungen hinreichend gut zu approximieren. Einige häufig verwendete Verteilungsmodelle werden kurz skizziert.

4.1 Binomialverteilung

Die Binomialverteilung gehört zu den bedeutendsten diskreten Wahrscheinlichkeitsverteilungen. Da sie auf eine von Jakob Bernoulli (1655 - 1705) entwickelte Versuchsanordnung zurückgeht, wird sie des öfteren auch Bernoulli-Verteilung genannt.

Als Bernoulli-Versuchsanordnung wird dann eine wiederholte Durchführung eines Zufallsvorgangs (Bernoulli-Experiment) bezeichnet, wobei gilt:

- Bei der Durchführung eines Experiments sind lediglich zwei Ergebnissalternativen möglich, das Ergebnis A bzw. das Ergebnis \bar{A} . Die Wahrscheinlichkeiten für A bzw. \bar{A} sind $P(A) = p$ bzw. $P(\bar{A}) = 1 - p$.
- Das Experiment wird n -mal durchgeführt, bei jeder einzelnen Durchführung muß p gleich groß sein, bzw. die Realisation von A und \bar{A} bleibt konstant.
- Die Wiederholungen sind unabhängig voneinander.

Die Fragestellung ist dann: Wie groß ist die Wahrscheinlichkeit, daß das Ergebnis A bei n Versuchen x -mal realisiert wird, (wobei $x = 0, 1, \dots, n$) ?

Beispiel (aus Tiede 2000): Eine Urne enthält 50 schwarze und 30 weiße Kugeln. Die Wahrscheinlichkeit bei einem Zufallszug eine weiße Kugel zu ziehen beträgt dann $p = 30/(50 + 30) = 3/8$. Die Ziehung wird dann $n = 10$ mal durchgeführt, wobei nach jedem Zug die ausgewählte Kugel wieder in die Urne zurückgelegt wird, so daß sich die Wahrscheinlichkeit $p = 3/8$ nicht ändert. Gefragt ist dann nach der Wahrscheinlichkeit, daß sich bei den 10 Ziehungen z.B. $x = 0$ -mal (oder $x = 1$ -mal etc.) eine weiße Kugel gezogen wird.

Einen Lösungsansatz findet man durch folgende Überlegung: Die Wahrscheinlichkeit, daß bei n Experimenten x -mal das Ergebnis A auftritt und $(n - x)$ -mal das Ergebnis \bar{A} , beträgt nach dem Multiplikationssatz für unabhängige Ereignisse:

$$\underbrace{p \cdot p \cdot \dots \cdot p}_x \cdot \underbrace{(1 - p) \cdot (1 - p) \cdot \dots \cdot (1 - p)}_{n - x} = p^x (1 - p)^{n - x}$$

Nun sind zudem noch verschiedene Reihenfolgen für die Ergebnisse A und \bar{A} möglich, für die gilt, daß A insgesamt x -mal und \bar{A} insgesamt $(n - x)$ -mal enthalten ist. Jede dieser möglichen anderen Reihenfolgen tritt dann ebenfalls mit der angegebenen Wahrscheinlichkeit auf. Die Anzahl der verschiedenen Reihenfolgen muß bei der Berechnung der Wahrscheinlichkeit mit berücksichtigt werden.

Aus der Kombinatorik ergibt sich, daß sich die Ergebnisse A und \bar{A} , die bei n Experimenten insgesamt x - bzw. $(n - x)$ -mal auftreten, in

$$\frac{n!}{x!(n - x)!} = \binom{n}{x}$$

unterschiedlichen Reihenfolgen anordnen lassen. Die Wahrscheinlichkeit für jede dieser Reihenfolge beträgt $p^x (1 - p)^{n - x}$. Also ist die Wahrscheinlichkeit, bei einem aus n Wiederholungen

bestehenden Bernoulli-Experiment genau x mal das Ereignis A zu erhalten $\binom{n}{x} p^x (1-p)^{n-x}$.

Allgemein ausgedrückt: Eine Zufallsvariable X heißt **binomialverteilt** mit den Parametern n und p , bzw. $X \sim B(n, p)$, wenn die **Wahrscheinlichkeitsfunktion** von X gegeben ist durch

$$f(x_i) = \frac{n!}{x_i! \cdot (n - x_i)!} \cdot p^{x_i} (1 - p)^{n - x_i} = \binom{n}{x_i} p^{x_i} (1 - p)^{n - x_i} \quad x_i = 0, 1, \dots, n$$

Die **Verteilungsfunktion** $F(x)$ ergibt sich dann mit:

$$F(x) = \sum_{x_i \leq x} f(x_i) = \sum_{x_i \leq x} \binom{n}{x_i} p^{x_i} (1 - p)^{n - x_i} \quad x_i = 0, 1, \dots, n$$

Wie die meisten theoretischen Verteilungen ist die Binomialverteilung auch tabelliert. Die im Anhang beigefügte Tabelle enthält für verschiedene p und n die Wahrscheinlichkeitsfunktion $f(x; n, p)$ der zugehörigen $B(n, p)$ -Verteilung. Die Tabelle gibt zwar nur für Wahrscheinlichkeiten an, welche kleiner oder gleich 0.5 sind. Aber wegen der Beziehung zwischen den Zufallsvariablen X und $Y = n - X$ und $X \sim B(n, p)$ gilt $Y \sim B(n, q)$, wobei $q = 1 - p$. Wenn $y = n - x$ dann gilt nämlich (vgl. Schlittgen 1998):

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x (1 - p)^{n - x} = \binom{n}{n - x} q^{n - x} (1 - q)^{n - (n - x)} \\ &= \binom{n}{y} q^y (1 - q)^{n - y} = P(Y = y) \end{aligned}$$

Für $p > 0.5$ nutzt man daher die Beziehung:

$$P(x|n, p) = P(n - x|n, 1 - p)$$

Die gesuchte Wahrscheinlichkeit für x der Verteilung $B(n, p)$ ist dann so groß wie die in der Tabelle angegebene Wahrscheinlichkeit für $n - x$ der Verteilung $B(n, 1 - p)$. Falls z.B. $P(4)$ der Verteilung $B(10, 0.6)$ gesucht wird, nutzt man die Tabelle der Verteilung $B(10, 1 - 0.6) = B(10, 0.4)$ für $P(10 - 4) = P(6)$ und liest den Wahrscheinlichkeitswert $P(6) = 0.1115$ ab.

Eigenschaften der Binomialverteilung: Der Erwartungswert $E(X)$ und die Varianz $Var(X)$ der Binomialverteilung mit den Parametern n und p sind wie folgt:

$$\begin{aligned} E(X) &= n \cdot p, \\ Var(X) &= n \cdot p \cdot (1 - p) = E(X) \cdot (1 - p) \end{aligned}$$

Fall der Wert für p nahe null oder eins liegt, verläuft die binomiale Verteilung linkssteil bzw. rechtssteil. Ist $p = 0.5$ so liegt eine symmetrische Verteilung vor.

Zudem gilt: Sind X und Y unabhängige, binomialverteilte Zufallsvariablen mit dem gleichen p , also $X \sim B(n, p)$ und $Y \sim B(m, p)$, so ist die Summe ebenfalls binomialverteilt:

$$Z = X + Y \sim B(n + m, p)$$

Beispiel 1: Wir gehen wieder von dem dreimaligen Münzwurf aus. Geprüft wird zunächst, ob es sich um eine Bernoulli-Versuchsanordnung handelt: 1. Es gibt nur zwei Möglichkeiten, Kopf oder Zahl. Die Wahrscheinlichkeit für „Kopf“ ist bei einem Wurf $p = 0.5$. 2. Die Wahrscheinlichkeit für „Kopf“ ändert sich nicht, wenn mehrere Münzwürfe fair ausgeführt werden. 3. Die Wiederholungen sind unabhängig voneinander.

Gefragt ist dann beispielsweise nach der Wahrscheinlichkeit für „0“ Kopfwürfe, also daß bei einem dreimaligen Münzwurf, „keinmal“ Kopf auftritt. Die Häufigkeiten für das Ergebnis „Kopf“ können durch eine nach $B(3, 0.5)$ verteilte Zufallsvariable X angegeben werden, für deren Wert $x = 0$ sich dann die Wahrscheinlichkeit

$$P(0) = \binom{n}{x} p^x (1-p)^{(n-x)} = \binom{3}{0} 0.5^0 (1-0.5)^{(3-0)} = \frac{3!}{0!(3-0)!} 0.5^3 = \frac{1}{8} = 0.125$$

ergibt.

Die Wahrscheinlichkeit für den Wert $x = 1$ wäre dann:

$$P(1) = \binom{3}{1} 0.5^1 (1-0.5)^{(3-1)} = 0.375$$

für $x = 2$:

$$P(2) = \binom{3}{2} 0.5^2 (1-0.5)^{(3-2)} = 0.375$$

und für $x = 3$:

$$P(3) = \binom{3}{3} 0.5^3 (1-0.5)^{(3-3)} = 0.125$$

Diese Werte lassen sich aber auch ohne Berechnung bequem aus der Tabelle ablesen.

Beispiel 2: (aus Tiede 2000) Nach einer Bypass-Operation liegt die fünfjährige Überlebenschance für Frauen einer bestimmten Altersgruppe bei 0.8 und die für Männer bei 0.7. An einem bestimmten Tag werden zwei Frauen und drei Männer dieser Altersgruppe operiert. Mit welcher Wahrscheinlichkeit sind von diesen Patienten nach fünf Jahren noch genau zwei Personen am Leben?

Wir betrachten diese Situation als ein Bernoulli-Experiment mit jeweils konstanter Erfolgswahrscheinlichkeit (Überleben) und n -Wiederholungen (2 Frauen und 3 Männer). Die Anzahl der überlebenden Frauen kann dann durch eine nach $B(2, 0.8)$ und die der Männer durch eine $B(3, 0.7)$ verteilten Zufallsvariablen erfaßt werden. Gesucht ist die Wahrscheinlichkeit für das Überleben von genau zwei Personen. Folgende Möglichkeiten bestehen:

- zwei Frauen und kein Mann
- zwei Männer und keine Frau
- eine Frau und ein Mann

Die Wahrscheinlichkeit, daß $x = 0, 1, 2$ Frauen überleben, kann unter den gemachten Voraussetzung der Binomialverteilung $B(2, 0.8)$ entnommen werden. Da nur bis $p = 0.5$ tabelliert wird, läßt sich zum praktischen Vorgehen folgende Beziehung nutzen: $P(x|n, p) = P(n-x|n, 1-p)$.

$$\begin{aligned} P_F(0|2, 0.8) &= P(2-0|2, 1-0.8) = P(2|2, 0.2) = 0.04 \\ P_F(1|2, 0.8) &= P(2-1|2, 1-0.8) = P(1|2, 0.2) = 0.32 \\ P_F(2|2, 0.8) &= P(2-2|2, 1-0.8) = P(0|2, 0.2) = 0.64 \end{aligned}$$

Analog gilt für die Wahrscheinlichkeit, daß $x = 0, 1, 2, 3$ Männer überleben, unter Verwendung von $B(3, 0.7)$:

$$\begin{aligned}
P_M(0|3, 0.7) &= P(3-0|3, 1-0.7) = P(3|3, 0.3) = 0.027 \\
P_M(1|3, 0.7) &= P(3-1|3, 1-0.7) = P(2|3, 0.3) = 0.189 \\
P_M(2|3, 0.7) &= P(3-2|3, 1-0.7) = P(1|3, 0.3) = 0.441 \\
P_M(3|3, 0.7) &= P(3-3|3, 1-0.7) = P(0|3, 0.3) = 0.343
\end{aligned}$$

Genau zwei Personen überleben dann mit der Wahrscheinlichkeit:

$$P_F(0|2, 0.8) \cdot P_M(2|3, 0.7) + P_F(1|2, 0.8) \cdot P_M(1|3, 0.7) + P_F(2|2, 0.8) \cdot P_M(0|3, 0.7) = 0.0954$$

4.2 Multinomialverteilung

Aus der Binomialverteilung läßt sich nur die Wahrscheinlichkeit für die Häufigkeit des Eintrittes oder Nicht-Eintrittes eines einzelnen Ereignisses A bei n Versuchen ableiten. Liegen dagegen mehr als zwei disjunkte Ereignisse vor, die den gesamten Ergebnisraum aufspannen, wird die **Multinomialverteilung** verwendet. Die Konzeption ist dann wie folgt:

Es werden k disjunkte Ereignisse A_1, A_2, \dots, A_k betrachtet, die den gesamten Ergebnisraum Ω aufspannen, so daß gilt: $A_1 \cup A_2 \cup \dots \cup A_k = \Omega$. Gesucht ist dann die gemeinsame Verteilung der Anzahlen des Eintretens aller Ereignisse bei n Versuchswiederholungen.

Multinomialverteilung: Die Zufallsvariablen X_1, X_2, \dots, X_k heißen **multinomialverteilt** mit den Parametern n, p_1, p_2, \dots, p_k , bzw. $(X_1, X_2, \dots, X_k) \sim M(n, p_1, p_2, \dots, p_k)$, wenn die zugehörige Wahrscheinlichkeitsfunktion die folgende Gestalt hat:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

wobei $x_1, x_2, \dots, x_k \geq 0$ und $x_1 + x_2 + \dots + x_k = n$.

Als **Erwartungswert** ergibt sich:

$$E(X_i) = n \cdot p_i$$

Und die **Varianz** ist:

$$Var(X_i) = n \cdot p_i(1 - p_i)$$

Beispiel: Vierseitiger Steinzeitwürfel In der Steinzeit kannte man schon das Würfelspiel, man verwendete dabei für die Herstellung eines vierseitigen Würfels spezielle Tierknochen. Aus Experimenten in Museen kennt man die statistische Wahrscheinlichkeit für die vier Seiten, wenn man sie durchnummeriert sind dies $p_1 = p_2 = 0.4$ und $p_3 = p_4 = 0.1$. Gefragt ist nun, wie die Anzahl der Würfe mit den verschiedenen Seiten verteilt ist, wenn ein derartiger Würfel insgesamt n mal geworfen wird.

Es lassen sich nun Zufallsvariablen X_1, X_2, X_3, X_4 definieren, die die Anzahlen für die vier Seiten angeben. Analog der Binomialverteilung erhält man eine spezielle Serie, in der x_1, x_2, x_3 und x_4 mal die entsprechenden Seiten nach oben zu liegen kommen, die Wahrscheinlichkeit:

$$p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$

Diese Wahrscheinlichkeit muß mit der Anzahl der möglichen Anordnungen multipliziert werden:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) = \frac{n!}{x_1! x_2! x_3! x_4!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$

Wenn man z.B. nach der Wahrscheinlichkeit fragt, daß bei 12 Würfeln alle Seiten gleich häufig nach oben zu liegen kommen (also jede 3 mal), ist dies:

$$P(X_1 = 3, X_2 = 3, X_3 = 3, X_4 = 3) = \frac{12!}{3!3!3!3!} 0.4^3 0.4^3 0.1^3 0.1^3 = 0.0015$$

4.3 Normalverteilung

Eine der populärsten und wichtigsten stetigen Wahrscheinlichkeitsverteilungen ist die **Normalverteilung**. Zu Ehren ihres Erfinders wird sie häufig auch Gauß-Verteilung oder Gaußsche Fehlerkurve genannt. Aus zwei Gründen hat diese Verteilung in der Statistik ein besonderes Gewicht bekommen: Zum einen kann die Normalverteilung bei einer Vielzahl von statistischen Maßzahlen als Verteilungsmodell unterstellt werden, wenn nur die Stichproben genügend groß sind. Zum zweiten weist die Normalverteilung einige positive formale Eigenschaften auf, die das empirische Arbeiten erleichtern (vgl. z.B. Schlittgen 1998, S.230). Die Dichtefunktion läßt sich wie folgt beschreiben:

Dichtefunktion: Die Zufallsvariable X heißt normalverteilt mit den Parametern μ und σ , bzw. $X \sim N(\mu, \sigma)$, wenn ihre Dichte gegeben ist durch

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Wie Abbildung 14 zeigt, beeinflussen die beiden Parameter μ und σ die Lage der Verteilung.

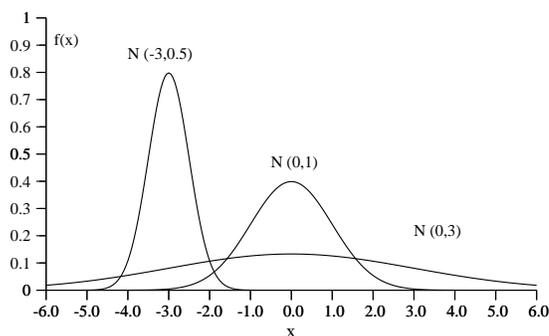


ABBILDUNG 14: Dichte der Normalverteilung für verschiedene Parameter

Die beiden expliziten Parameter sind zugleich der Erwartungswert als auch die Standardabweichung von X , eine normalverteilte Zufallsvariable wird daher auch mit $N(\mu, \sigma)$ gekennzeichnet. Auf Basis dieser beiden Parameter lassen sich prinzipiell unendlich viele Normalverteilungen konstruieren.

Erwartungswert, Varianz: Die Parameter μ und σ einer normalverteilten Zufallsvariablen X , bzw. $X \sim N(\mu, \sigma)$, entsprechen dem Erwartungswert und der Standardabweichung:

$$\begin{aligned} E(X) &= \mu \\ \text{Var}(X) &= \sigma^2 \end{aligned}$$

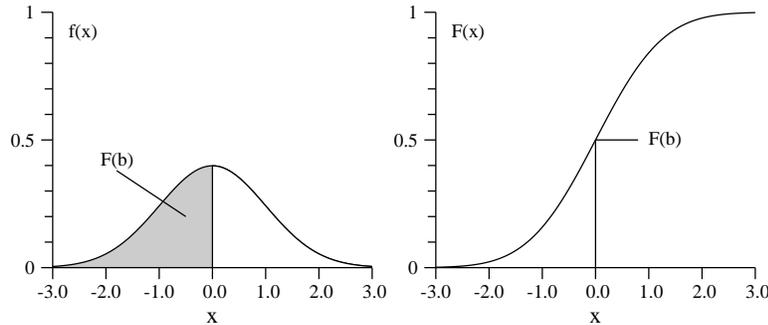


ABBILDUNG 15: Dichtefunktion und Verteilungsfunktion einer Normalverteilung

Verteilungsfunktion der Normalverteilung Die Verteilungsfunktion der Normalverteilung gibt den Inhalt der Fläche unter der Dichtefunktion mit den Grenzen $-\infty$ und x an, wobei die Kurve die Form eines liegenden „S“ annimmt.

Die Verteilungsfunktion ist somit das Integral der Dichtefunktion:

$$F(x) = \int_{-\infty}^x f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(x-\mu)^2/2\sigma^2}$$

Eigenschaften der Normalverteilung: Die Normalverteilung hat folgende Kennzeichen:

- Die Dichtefunktion hat stets im Wert $x = \mu$ ihr Maximum
- Die Wendepunkte liegen in $x = \mu - \sigma$ und $x = \mu + \sigma$
- Die Dichtefunktion ist symmetrisch zum Lot im Wert $x = \mu$
- Die Dichtefunktion nähert sich asymptotisch den Werten $-\infty$ und $+\infty$ auf der Abzisse

Eine Normalverteilung mit dem Mittelwert $\mu = 0$ und der Standardabweichung von $\sigma = 1$ wird als Standardnormalverteilung bezeichnet. Eine nach $N(0,1)$ verteilte Zufallsvariable wird häufig auch mit K abgekürzt und ihre Werte mit k . Jede beliebige Normalverteilung $N(\mu, \sigma)$ kann mit Hilfe der folgenden Transformation in eine entsprechende Normalverteilung transformiert werden:

$$K = \frac{X - \mu}{\sigma}$$

Setzt man in die Dichtefunktion der Normalverteilung für x die Werte für k ein, erhält man die **Dichtefunktion der Standardnormalverteilung:**

$$f(k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k^2}, \quad -\infty < k < +\infty$$

Zwischen den beiden Dichtefunktionen besteht folgende Beziehung:

$$f(x)\sigma = f(k)$$

Die Beziehung der Verteilungsfunktionen von X und K ist dann:

$$P(X \leq x) = P(K \leq k), \quad \text{mit } K = \frac{X - \mu}{\sigma}$$

Die Wahrscheinlichkeit, daß eine nach $N(\mu, \sigma)$ verteilte Zufallsvariable X den Wert realisiert, der kleiner oder gleich x ist, entspricht der Wahrscheinlichkeit, daß die Standardnormalverteilung einen Wert realisiert, der kleiner oder gleich dem Wert $k = \frac{x-\mu}{\sigma}$ ist. Da die Verteilungsfunktion der Standardnormalverteilung in fast allen einschlägigen Lehrbüchern im Anhang tabelliert vorliegt, kann jede Wahrscheinlichkeit für X unter Verwendung der Tabelle der Verteilungsfunktion für K ermittelt werden.

Die tabellierten Werte weisen zwar nur die Wahrscheinlichkeit für $P(K \geq k)$ wobei $k > 0$ auf, da die Normalverteilung jedoch vollkommen symmetrisch zu ihrem Mittelwert verläuft, sind die Flächen unterhalb der Verteilung im negativen Bereich also für $P(K \leq k)$, wobei $k < 0$, genauso groß.

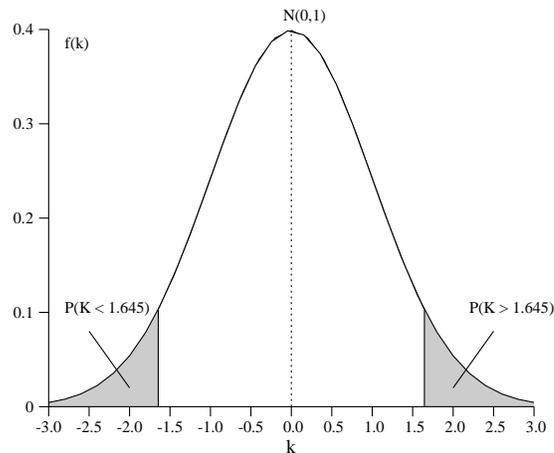


ABBILDUNG 16: Symmetrie der Flächen unterhalb der Normalverteilung

Beispiel: Es wird angenommen, daß die Körpergrößenverteilung von Autofahrern und Autofahrerinnen den folgenden Normalverteilungen folgen:

- Männer: $N(176.8 \text{ cm}, 5.2 \text{ cm})$
- Frauen: $N(170.2 \text{ cm}, 5.0 \text{ cm})$

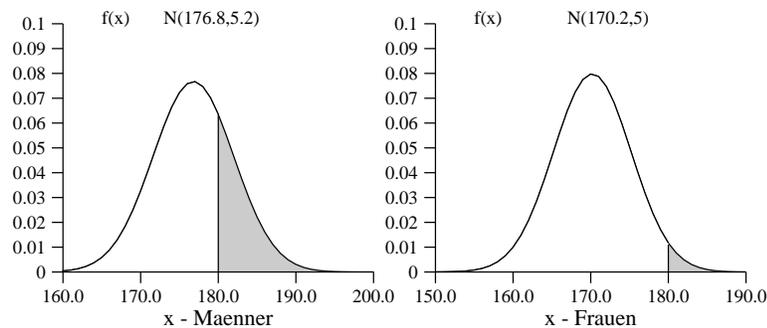


ABBILDUNG 17: Körpergröße von Männern u. Frauen - Dichtefunktion

Gesucht ist nun die Wahrscheinlichkeit, daß die beiden Zufallsvariablen X_M und X_F einen Wert realisieren, der größer ist als 180, bzw. die Wahrscheinlichkeit, daß männliche und weibliche

Autofahrer eine Körpergröße > 180 cm aufweisen. Für eine tabellarische Ermittlung ist daher eine Transformation notwendig:

$$k_M = \frac{x - \mu}{\sigma} = \frac{180 - 176.8}{5.2} = 0.615 \quad (5)$$

$$k_F = \frac{x - \mu}{\sigma} = \frac{180 - 170.2}{5.0} = 1.96 \quad (6)$$

Aus der Tabelle können dann folgende Werte für $P(K \geq k)$ abgelesen werden³: $P(K \geq 0.615) = 0.269$ und $P(K \geq 1.96) = 0.025$. Die Wahrscheinlichkeit, daß ein männlicher Autofahrer größer oder gleich 180 cm ist beträgt also ca. 27% und bei weiblichen Autofahrern nur 2.5%.

Beispiel 2: Eine Zufallsvariable X (Körpergewicht von Männern) sei gemäß $N(75, 5)$ verteilt. Gesucht ist die Wahrscheinlichkeit für $P(70 \leq X \leq 85)$. Im ersten Schritt werden die Grenzwerte standardisiert:

$$k_1 = \frac{70 - 75}{5} = -1$$

$$k_2 = \frac{85 - 75}{5} = 2$$

dann wird die Beziehung $P(70 \leq X \leq 85) = P(-1 \leq K \leq 2)$ genutzt. Die gesuchte Wahrscheinlichkeit ergibt sich dann aus der Addition von $P(0 \leq K \leq 2)$ und $P(-1 \leq K \leq 0)$, wobei der zweite Summand genauso groß ist wie $P(0 \leq K \leq 1)$. Somit erhält man:

$$\begin{aligned} P(0 \leq K \leq 2) &= 0.5 - P(K \geq 2) = 0.5 - 0.02275 = 0.4773 \\ P(0 \leq K \leq 1) &= 0.5 - P(K \geq 1) = 0.5 - 0.15866 = 0.3413 \\ P(-1 \leq K \leq 2) &= 0.4773 + 0.3413 = 0.8186 \end{aligned}$$

Die gesuchte Wahrscheinlichkeit beträgt also ca. 82%.

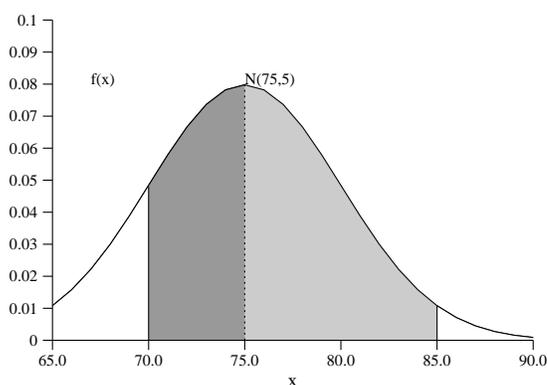


ABBILDUNG 18: *Ermittelte Flächen*

³Zwischenwerte lassen sich interpolieren.

4.4 Approximation von Verteilungen

Die Normalverteilung wurde von Gauß ursprünglich als Modell für die Verteilung von Meßfehlern verwendet. Dabei kann häufig angenommen werden, daß positive und negative Meßfehler gleich wahrscheinlich sind. In der Regel muß man allerdings mehrere unabhängige Fehlerquellen bei seiner Untersuchung berücksichtigen, die sich jeweils als Zufallsvariable konzipieren lassen. In dem vorliegenden Beispiel soll von fast gleichverteilten unabhängigen Zufallsvariablen X_1, X_2, X_3 und X_4 ausgegangen werden. Abbildung 19 zeigt die Dichtefunktion von X_1 und X_2 , welche durch eine Simulation auf Basis eines Zufallsgenerator (Werte zwischen 0 und 1, $n = 9000$) erzeugt wurden.

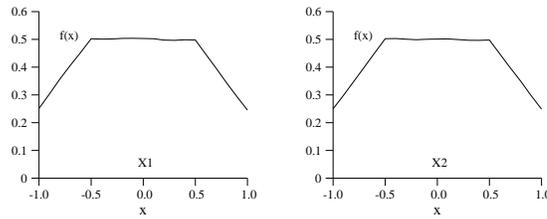


ABBILDUNG 19: Dichtefunktion zweier Zufallsvariablen

Der Gesamtfehler ergibt sich aus der Addition der einzelnen Fehlerquellen, bzw. der Zufallsvariablen. In Abbildung 20 sind die resultierenden Dichtefunktionen dargestellt.

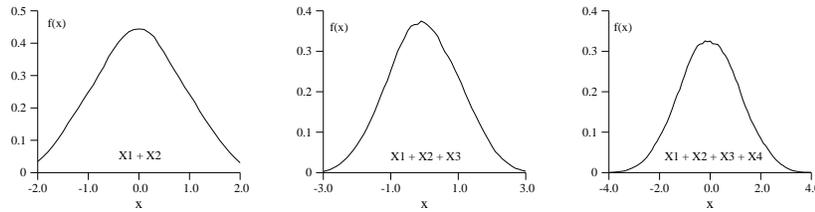


ABBILDUNG 20: Addition von Zufallsvariablen

Werden zunächst nur zwei Fehlerquellen berücksichtigt, also $X_1 + X_2$, ist das Ergebnis eine dreieckförmige (Simpson-) Dichtefunktion, bei Addition von mehreren Zufallsvariablen, tendiert die Verteilung von $\sum X_i$ mit wachsender Zahl der Summanden gegen die Normalverteilung. Dieser Sachverhalt wird als **zentraler Grenzwertsatz** bezeichnet und gilt nicht nur für rechteckverteilte und symmetrische Zufallsvariablen sondern prinzipiell für alle Verteilungsformen einer Zufallsvariablen. Die Konvergenz gegen die Normalverteilung ist allerdings im Falle von symmetrischen Ausgangsverteilungen besonders schnell.

zentraler Grenzwertsatz: X_1, X_2, \dots, X_n seien identisch verteilte, unabhängige Zufallsvariablen mit

$$E(X_i) = \mu \quad \text{und} \quad \text{Var}(X_i) = \sigma^2 > 0$$

Dann konvergiert die Verteilung der standardisierten Summe dieser Zufallsvariablen⁴

$$Z_n = \frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$$

⁴Die Standardisierung ist notwendig, weil bei $\mu \neq 0$ die Dichte verschoben und wegen $\sigma^2 > 0$ immer flacher würde.

mit steigender Summandenzahl gegen eine standardisierte Normalverteilung:

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \phi(z)$$

Dies kann auch wie folgt geschrieben werden: $Z_n \sim N(0, 1)$.

4.4.1 Approximation der Binomialverteilung

Ein wichtiger Spezialfall des zentralen Grenzwertsatzes ist die Approximation der Binomialverteilung für mittlere und große n durch die Normalverteilung. Für diese Fälle sind die Wahrscheinlichkeiten nicht mehr tabelliert. Man verwendet stattdessen den sogenannten **lokalen Grenzwertsatz** von *Laplace* und *de Moivre*. Sie konnten unabhängig voneinander zeigen, daß die Binomialverteilung mit wachsendem n und konstantem p in Richtung Normalverteilung $N(np, \sqrt{np(1-p)})$ tendiert und für $n \rightarrow \infty$ in sie übergeht. Dieser Grenzwertsatz wird „lokal“ genannt, weil er sich auf die diskreten Stellen von X bezieht und die Dichtefunktion der Normalverteilung betrifft.

lokaler Grenzwertsatz: Falls X binomialverteilt ist mit den Parametern n und p , bzw. $X \sim B(n, p)$, so entspricht die Wahrscheinlichkeit $P(x)$ bei $n \rightarrow \infty$ der Wahrscheinlichkeitsdichte $f(x)$ einer nach $N(np, \sqrt{np(1-p)})$ verteilten Variablen. Es gilt also:

$$\frac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1)$$

Werden keine Wahrscheinlichkeiten für einzelne x -Werte gesucht sondern die Wahrscheinlichkeit für die Realisierung eines x -Wertes in einem bestimmten Bereich $P(a \leq X \leq b)$, wird anstelle des lokalen Grenzwertsatzes der **allgemeine Grenzwertsatz** verwendet. Er bezieht sich auf die Verteilungsfunktion der Normalverteilung.

allgemeiner Grenzwertsatz: Falls X binomialverteilt ist mit den Parametern n und p , bzw. $X \sim B(n, p)$, so entspricht die Wahrscheinlichkeit $P(a \leq X \leq b)$ wobei $a, b = 0, 1, 2, \dots, n$ bei $n \rightarrow \infty$ dem Wert der Verteilungsfunktion $F(x)$ einer nach $N(\mu, \sigma) = N(np, \sqrt{np(1-p)})$ verteilten Variablen zwischen den Grenzen $\frac{a-\mu-0.5}{\sigma}$ und $\frac{b-\mu+0.5}{\sigma}$. Die Approximation enthält hierbei einen Korrekturfaktor von 0.5, da die Treppenstufen der Binomialverteilung von der stetigen Verteilungsfunktion der Normalverteilung ungefähr in der Mitte getroffen werden⁵.

Die Qualität der Approximation wird um so besser, je größer n ist und je näher p an dem Wert 0.5 liegt. Als Faustregel gilt, daß die Approximation hinreichend genau ist, wenn (vgl. Tiede 2000, S.74):

$$n \geq \frac{9}{p(1-p)}$$

Beispiel 1 (aus Schlittgen 1998, S.242): Ein Marktforschungsinstitut, das in Foto-Fachgeschäften eine Erhebung durchführen will, stützt sich bei der zufälligen Auswahl von $n = 80$ Geschäften auf eine erworbene Adressenliste. Von den über 1000 Adressen auf der Liste sind 15% nicht gültig. Aufgrund der großen Anzahl von Adressen wird dieser Wert als konstante Wahrscheinlichkeit p aufgefaßt. Gesucht ist die Wahrscheinlichkeit, daß bei den $n = 80$ gezogenen Geschäften genau $x = 8$ ungültig sind, also $P(X = 8) = ?$.

⁵Schlittgen (1998) verwendet die Verteilungsfunktion der Normalverteilung auch für den Fall $P(X = x)$ anstelle der Dichtefunktion:

$$P(X = x) = \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Für $n = 80$ liegen keine tabellierten Werte der Binomialverteilung vor, es sollte daher geprüft werden, ob eine Approximation durch eine Normalverteilung möglich ist:

$$n \geq \frac{9}{p(1-p)} = \frac{9}{0.15 \cdot (1-0.15)} = 70.59$$

Da $n = 80 > 70.6$ kann, gemäß einigen Autoren, die Approximation gewählt, bzw. es kann $B(n, p) \sim N(np, \sqrt{np(1-p)})$ angenommen werden. Man erhält $B(80, 0.15) \sim N(80 \cdot 0.15, \sqrt{80 \cdot 0.15(1-0.15)}) = N(12, 3.194)$. Die Beziehung zwischen den Dichtefunktionen ist:

$$\begin{aligned} k &= \frac{8-12}{3.194} = 1,252 \\ f(k) &= 0.1816 \\ f(x) &= \frac{f(k)}{\sigma} = \frac{0.1816}{3.194} = 0.057 \end{aligned}$$

Beispiel 2: Es soll im Anschluß an Beispiel 1 nun die Wahrscheinlichkeit ermittelt werden, daß von den $n = 80$ ausgewählten Adressen der Fehlerbereich bei 9 bis 15 Adressen liegt, bzw. gefragt ist nach $P(9 \leq X \leq 15)$. Wie schon weiter oben festgestellt wurde, läßt sich unter den gegebenen Umständen auf die Approximation durch die Normalverteilung zurückgreifen. Da sich die gesuchte Wahrscheinlichkeit nicht auf einen einzelnen Wert sondern auf ein Intervall bezieht, muß nun die Verteilungsfunktion verwendet werden:

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a - \mu - 0.5}{\sigma} \leq K \leq \frac{b - \mu + 0.5}{\sigma}\right) \\ P(9 \leq X \leq 15) &= P\left(\frac{9 - 12 - 0.5}{3.194} \leq K \leq \frac{15 - 12 + 0.5}{3.194}\right) \\ &= P(-1.096 \leq K \leq 1.096) \\ &= (0.5 - 0.136) + (0.5 - 0.136) = 0.728 \end{aligned}$$

4.5 Chi-Quadrat-Verteilung

Neben normalverteilten Zufallsvariablen können natürlich auch andere - in bestimmten Kontexten sehr nützliche - Verteilungsformen für Zufallsvariablen definiert werden. Eine besondere Bedeutung hat dabei die χ^2 -Verteilung. Sie ist definiert mit:

$$\chi^2 = \sum_{i=1}^{\nu} K_i^2, \quad \text{wobei } K_i \text{ nach } N(0, 1) \text{ verteilt ist}$$

Man erhält eine χ^2 -verteilte Zufallsvariable, wenn man die Summe der Quadrate von ν untereinander unabhängigen standardnormalverteilten Zufallsvariablen bildet. Die Anzahl ν der frei variierbaren und unabhängigen Zufallsvariablen wird häufig als *Freiheitsgrad* bezeichnet, weil diese bezüglich ihrer Realisationen keinen Nebenbedingungen bzw. Einschränkungen unterworfen sind.

Erwartungswert und Varianz: Ist eine Zufallsvariable Y χ^2 -verteilt, mit ν Freiheitsgraden, so gilt $Y \sim \chi^2_{\nu}$. Der Erwartungswert einer χ^2 -verteilten Zufallsvariablen ist gleich dem Freiheitsgrad, die Varianz dem zweifachen Freiheitsgrad:

$$E(Y) = \nu, \quad \text{Var}(Y) = 2 \cdot \nu$$

Eine χ^2 -verteilte Zufallsvariable kann nur positive Werte realisieren und für den üblicherweise verwendeten Bereich sind die Quantile der Verteilung tabelliert.

5 Stichprobe, Stichprobenvariable und Stichprobenverteilung

5.1 Grundgesamtheit und Stichprobe

Datenträger bzw. statistische Einheiten mit gleichen Identifikationsmerkmalen bilden eine statistische Masse bzw. die *Grundgesamtheit*. Die Grundgesamtheit ist jene Gesamtheit, für die eine Untersuchung gelten soll.

Eine *Totalerhebung* liegt dann vor, wenn von allen Datenträgern der Grundgesamtheit Informationen erhoben werden. Beschränkt man sich dagegen nur auf einen Teil der Grundgesamtheit, wird diese Erhebung Teilerhebung oder *Stichprobe* genannt.

Beispiel: Eine Untersuchung soll sich auf die Wohnbevölkerung der Bundesrepublik Deutschland im Jahre 01.01.2000 beziehen. Dies ist also die angestrebte Grundgesamtheit. Die *statistischen Einheiten* wären dann die Personen, die zu diesem Zeitpunkt ihren ersten Wohnsitz in der BRD bei einem Einwohnermeldeamt gemeldet haben. Eine Totalerhebung liegt dann vor, wenn die gesamte Wohnbevölkerung in der Untersuchung befragt wird, von einer Stichprobe spricht man, wenn nur ein Teil davon zur Befragung ausgewählt wird.

Eine Stichprobe hat gegenüber einer Totalerhebung den Vorteil, daß sie weniger aufwendig und damit mit geringeren Kosten verbunden ist. Zudem läßt sich eine Befragung auf Basis einer Stichprobe in einem kürzeren Zeitraum realisieren, was die Vergleichbarkeit der Daten in Bezug auf den Erhebungszeitpunkt erhöht. Ein Problem stellt allerdings der Punkt dar, daß die erhobenen Informationen auf Basis der Stichprobe generalisiert werden sollen, bzw. sich nicht nur auf die Teilerhebung sondern auch auf die Grundgesamtheit beziehen soll. Ansätze zur Lösung dieses Problems werden im Rahmen des Testens und Schätzens behandelt.

5.2 Repräsentativität

Im Rahmen von Stichprobenkonzepten wird häufig der Begriff „Repräsentativität“ verwendet. So findet man kaum eine Umfragestudie, die ohne einen Exkurs über deren „Repräsentativität“ auskommt. Trotz dieser Beliebtheit fehlt diesem Begriff allerdings weitestgehend eine theoretische Fundierung und in mathematisch-statistischen Lehrbüchern wird man vergeblich nach solch einem Schlagwort suchen (vgl. Pötter/Rendtel 1993).

In einem Einführungsbuch für Sozialwissenschaftler (Böltken 1976, S.128) wird Repräsentativität wie folgt definiert: „Unter Repräsentativität verstehen wir, daß in einer Auswahl alle für die Grundgesamtheit typischen und charakteristischen Merkmale und Merkmalskombinationen getreu ihrer relativen Häufigkeit vertreten sein müssen und somit die Auswahl ein verkleinertes Abbild der Grundgesamtheit selbst für solche Merkmale ist, von deren Vorhandensein wir (noch) gar nichts wissen.“

Diese Definition kann vor allem in zwei Punkten kritisiert werden:

1. Wenn die Stichprobe in allen Punkten mit der Grundgesamtheit übereinstimmen würde, müßte man aus ihr ebenfalls wiederum eine Sub-Stichprobe ziehen können und die repräsentativ wäre und aus der folgenden ebenfalls, bis man letztlich bei einem Umfang von eins bei einem repräsentativen Menschen endet.

2. Die Idee der Übereinstimmung multivariater Verteilungen in Grundgesamtheit und Stichprobe ist allein schon wegen der hohen Anzahl an möglichen Merkmalskombinationen unrealistisch. So ergeben sich z.B. bei fünf Variablen mit vier Merkmalskombinationen $4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 1024$ Zellen. Eine Verteilung von 1024 Beobachtungen in 1024 Zellen wird aber mit Sicherheit nicht mehr mit der Verteilung in der Grundgesamtheit übereinstimmen.

Und dies wird sich selbst auch dann nicht ändern, wenn man hier den dreifachen Stichprobenumfang wählen würde. Andererseits könnte man argumentieren, daß mit zunehmendem Stichprobenumfang, die „Repräsentativität“ zunimmt.

Eine andere Vorstellung bezieht den Begriff „Repräsentativität“ auf die Möglichkeit, statistische Befunde, die aus Zufallsstichproben stammen, beurteilen und verallgemeinern zu können, wobei die hiermit einhergehende Ungenauigkeit in der Aussage wahrscheinlichkeitstheoretisch berücksichtigt wird.

5.3 Zufällige und bewußte Auswahl

Die Stichprobenziehung kann zunächst danach beurteilt werden, ob sie eine zufällige Auswahl oder eine bewußte Auswahl darstellt.

Bewußte Auswahl: Wenn die Entscheidung, ob eine Einheit einer Grundgesamtheit in die Stichprobe kommt, mehr oder weniger vollständig im Ermessen des Auszuwählenden liegt, spricht man von einer bewußten Auswahl. Bei solch einem Auswahlmodus läßt sich nicht die Ziehungswahrscheinlichkeit für eine Einheit der Grundgesamtheit angeben.

Zu den bewußten Auswahlen zählt vor allem das Quotenauswahlverfahren, welches häufig in der Markt- und Meinungsforschung verwendet wird, aber auch die Erhebung von typischen Einzelfällen.

Zufällige Auswahl: Hier wird die Entscheidung, ob eine Einheit der Grundgesamtheit in die Stichprobe gelangt dem Ermessen des Auszuwählenden entzogen und von einem kontrollierten Zufallsprozeß gefällt. Diese zufällige Auswahl läßt sich dann als Zufallsexperiment auffassen, für die sich die Wahrscheinlichkeitstheorie anwenden läßt.

Zu den zufälligen Auswahlen zählt die einfache Zufallsstichprobe, sowie die geschichtete Auswahl und die Klumpenauswahl, die - einzeln oder kombiniert - auch mehrstufig verwendet werden können.

5.4 Einfache Zufallsstichprobe

Allgemein formuliert stellt eine Stichprobe eine endliche Teilmenge von Elementen der Grundgesamtheit dar. Durch diese Definition wird zum Ausdruck gebracht, daß

- Jedes Element nur einmal in der Stichprobe enthalten sein kann.
- Die Reihenfolge, in der die Elemente gezogen werden bedeutungslos ist.

Eine **Zufallsstichprobe** liegt dann vor, wenn jedes Element der Grundgesamtheit eine gleiche Chance hat, in die Stichprobe zu gelangen. Eine Möglichkeit dies zu realisieren besteht z.B. darin, die N Elemente der Grundgesamtheit durchnummerieren und zu mischen und dann zufällig mehrere Stichproben in Form von kompletten Nummernsets zu ziehen. Jeder Nummernset hat somit die gleiche Chance gezogen zu werden.

Von einer **einfachen Zufallsstichprobe** spricht man, wenn sich zudem die Auswahlwahrscheinlichkeit der Elemente von Zug zu Zug nicht ändert. Dies läßt sich allerdings nur bei Stichproben „mit Zurücklegen“ realisieren, was aber der Stichprobendefinition widerspricht. Bei Stichproben mit einem relativ kleinen Auswahlsatz z.B. in der Größenordnung von 1% ist die Änderung in der Auswahlwahrscheinlichkeit allerdings so gering, daß annähernd von einer „einfachen Zufallsstichprobe“ ausgegangen werden kann.

Neben dem erwähnten „Orginalverfahren“, wie Numerierung der Elemente der Grundgesamtheit, Mischen, zufällige Auswahl, gibt es eine ganze Reihe von anderen Verfahren, die zur Ziehung von Zufallsstichproben verwendet werden. So werden anstelle der Durchnummerierung

und dem Mischen den Elementen der Grundgesamtheit teilweise direkt Zufallsziffern zugeordnet. In Abhängigkeit vom gewünschten Umfang der Stichprobe werden dann mehrere Zahlen ausgelost, die jene Elemente identifizieren, die in die Stichprobe gelangen. Weitere Methoden sind das Schlußziffernverfahren, Auswahlen nach Namensanfang, Buchstabenverfahren und Geburtstagsverfahren.

5.5 Stichprobenvariable

Die einfache Zufallsstichprobe läßt sich konzeptionell auf zwei unterschiedliche Weisen darstellen:

- Die Zufallsstichprobe erfaßt n Werte von Realisierungen x_i einer Zufallsvariablen X , die die Stichprobe bilden.
- Die n Stichprobenwerte x_i stellen einzelne Realisierungen von n unabhängigen identisch verteilten Zufallsvariablen X_i dar. Jede der Zufallsvariablen X_i realisiert damit in der Stichprobe einen Wert x_i , so daß die Stichprobe aus n unabhängigen Stichprobenwerten x_i besteht.

Man kann sich also im Kontext der Zufallsstichprobe einerseits auf Stichprobenwerte beziehen (und von einer einzigen Zufallsvariablen ausgehen), oder aber von Realisierungen von n Stichprobenvariablen ausgehen.

In den folgenden Abschnitten sollen als Stichprobe die Stichprobenvariablen X_1, \dots, X_n bezeichnet werden.

Stichprobenvariable: Eine Stichprobe aus einer Grundgesamtheit, in der die Zufallsvariable X die Verteilungsfunktion $F(x)$ hat, besteht aus n Zufallsvariablen X_i , $i = 1, \dots, n$, den Stichprobenvariablen. Für diese gilt:

- Die Stichprobenvariablen X_1, \dots, X_n haben dieselbe Wahrscheinlichkeitsverteilung wie X .
- Die Stichprobenvariablen X_1, \dots, X_n sind unabhängig voneinander.

5.6 Stichprobenverteilungen

Aus den Werten x_1, \dots, x_n der Stichprobenvariablen X_1, \dots, X_n werden in der Regel statistische Maßzahlen bzw. empirische Parameter berechnet, etwa um Näherungswerte- oder Schätzwerte für unbekannte theoretische Parameter (Parameter in der Grundgesamtheit) zu erhalten, wie z.B. das arithmetische Mittel, ein Anteilswert oder die Varianz. Formal lassen sich diese Maßzahlen als Funktion der Werte x_i betrachten, z.B. $g(x_1, \dots, x_n)$. So ist in diesem Kontext das arithmetische Mittel eine Funktion von Zufallsvariablen und wird damit selbst zu einer Zufallsvariablen:

$$\bar{X} = g(X_1, \dots, X_n)$$

Eine statistische Maßzahl, die funktional abhängig von den Stichprobenvariablen X_i ist, wird als *Stichprobenfunktion* bezeichnet. Der arithmetische Mittelwert \bar{X} stellt die mögliche Realisation für das arithmetische Mittel $\bar{x} = 1/n \sum_{i=1}^n x_i$ dar, wenn n Stichprobenvariablen X_i jeweils einen Wert x_i annehmen. Anders ausgedrückt: Der aus einer konkreten Zufallsstichprobe stammende Wert x_i ist einer von vielen Mittelwerten, die hätten realisiert werden können. Prinzipiell können sehr viele Stichproben unter den gleichen Randbedingungen gezogen und jeweils der Mittelwert bestimmt werden. Die Verteilung dieser Mittelwerte ist die Verteilung der Variablen \bar{X} und die voneinander unterscheidbaren Werte \bar{x} geben die möglichen Werte und den Wertebereich von \bar{X} an.

Die Wahrscheinlichkeitsverteilung einer Stichprobenfunktion wie z.B. die Verteilung der Zufallsvariablen \bar{X} , wird *Stichprobenverteilung* dieser Funktion genannt.

5.6.1 Spezielle Stichprobenverteilungen

Arithmetische Mittel Eine spezielle Stichprobenfunktion ist das arithmetische Mittel $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Ihre Verteilung ist die Verteilung jener Mittelwerte, die sich aus den vielen möglichen verschiedenen Stichproben gleichen Umfangs berechnen lassen, die aus einer Grundgesamtheit gezogen werden können. Für diese Stichprobenverteilung läßt sich dann der Erwartungswert $E(\bar{X})$ ermitteln, wobei für $E(X_i)$ (der Mittelwert in der Grundgesamtheit) der Buchstabe μ verwendet wird:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

Das arithmetische Mittel all jener Mittelwerte, die in den vielen verschiedenen Stichproben gleichen Umfangs bestimmt werden können, ist genau so groß wie der Mittelwert in der Grundgesamtheit.

Die durchschnittliche quadratische Abweichung der Mittelwerte in den vielen Stichproben ist die Varianz um den Mittelwert der Grundgesamtheit. Wenn für $Var(X_i) = \sigma^2$ (der Varianz in der Grundgesamtheit) verwendet wird, ergibt sich:

$$Var(\bar{X}) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Die Wurzel aus der Varianz ist dann bekanntlich die Standardabweichung s , mit $s = \frac{\sigma}{\sqrt{n}}$. Dieser Ausdruck wird bisweilen auch als \sqrt{n} -Gesetz bezeichnet, andere verwendete Ausdrücke sind „Standardfehler“, „durchschnittlicher Stichprobenfehler“ oder „Stichprobenfehler“.

Die Stichprobenverteilung des arithmetischen Mittels besitzt also eine Standardabweichung $Var(\bar{X})$, die proportional zur Standardabweichung der Grundgesamtheit σ und umgekehrt proportional zur Wurzel aus dem Umfang der Stichprobe n ist.

Eine Vergrößerung der Stichprobe bewirkt dann zwar erwartungsgemäß eine Verkleinerung der Standardabweichung von \bar{X} , dies aber nicht linear: eine Verdopplung des Stichprobenumfangs verringert die Standardabweichung um den Faktor $1/\sqrt{2} = 0.707$, eine 4-facher Stichprobenumfang ist dann nötig, um die Streuung von \bar{X} zu halbieren.

Neben dem Erwartungswert und der Varianz läßt sich die **Verteilung der Stichprobenfunktion** \bar{X} über den schon weiter oben erläuterten **zentralen Grenzwertsatz** angeben. Er besagt, daß die Summe von n unabhängigen, gleichverteilten Zufallsvariablen X_i normalverteilt ist, wenn n gegen unendlich geht, bzw. $\sum_{i=1}^n X_i$ folgt der Normalverteilung für $n \rightarrow \infty$. Dies gilt natürlich auch für den Durchschnitt der Summe von Stichprobenvariablen X_i bzw. damit auch für $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Daraus folgt unter den Bedingungen der einfachen Zufallsstichprobe:

Die Stichprobenverteilung des arithmetischen Mittels ist für große Stichprobenumfänge approximativ normalverteilt, \bar{X} folgt also approximativ der Verteilung $N(\mu, \frac{\sigma}{\sqrt{n}})$. Sind bereits die Stichprobenvariablen X_i normalverteilt, bzw. $X_i \sim N(\mu, \sigma)$, so ist \bar{X} exakt normalverteilt. Entsprechend den Aussagen des zentralen Grenzwertsatzes gilt dies unabhängig davon, welche Verteilung in der Grundgesamtheit vorliegt.

Beispiel 1: In einem großen Unternehmen betragen die durchschnittlichen Fehlzeiten pro Jahr $\mu = 6$ Tage, wobei die Standardabweichung den Wert $\sigma = 3$ Tage hat. Nun werden 40 Beschäftigte zufällig ausgewählt und bezüglich ihrer Fehlzeiten befragt. Wie groß ist die Wahrscheinlichkeit, daß die durchschnittliche Fehlzeit \bar{X} dieser Personen im Bereich zwischen 5 und 6 Tagen liegt?

Gesucht ist $P(5 \leq \bar{X} \leq 6)$. \bar{X} ist verteilt nach $N(\mu, \sigma/\sqrt{n}) = N(6, 3/\sqrt{40}) = N(6, 0.474)$.

Zur Bestimmung der Wahrscheinlichkeit wird wiederum standardisiert:

$$\begin{aligned} P(5 \leq \bar{X} \leq 6) &= P\left(\frac{5-6}{0.474} \leq K \leq \frac{6-6}{0.474}\right) \\ &= P(-2.11 \leq K \leq 0) \\ &= (0.5 - 0.0174) + (0.5 - 0.5) = 0.4826 \end{aligned}$$

Die Wahrscheinlichkeit beträgt damit ca. 48%, daß die durchschnittliche Fehlzeit bei den befragten 40 Personen im Intervall 5 bis 6 Tagen liegt.

Beispiel 2: Bezug nehmend auf das vorherige Beispiel wird nun der Stichprobenumfang n gesucht, der gewährleistet, daß die Sicherheit der Aussage, die durchschnittliche Fehlzeit liege im Intervall 5 bis 6 Tage, von 48.26% auf eine Wahrscheinlichkeit von 49.9% ansteigt.

Wir können also von $P(k_1 \leq K \leq k_2) = 0.499$ ausgehen. Die rechte Intervallgrenze soll mit $x = 6$ Tage und damit $k_2 = 0$ festgelegt sein. Zwischen n und k besteht folgende Beziehung:

$$\begin{aligned} k &= \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{(\bar{x} - \mu)\sqrt{n}}{\sigma} \\ \sqrt{n} &= \frac{k \cdot \sigma}{\bar{x} - \mu} \\ n &= \left(\frac{k \cdot \sigma}{\bar{x} - \mu}\right)^2 \end{aligned}$$

Nun muß noch der Wert für k_1 aus der Tabelle ermittelt werden. Tabelliert sind Werte für $P(K \geq k)$, wir müssen also in der Tabelle für die Verteilungsfunktion den Wahrscheinlichkeitswert $0.5 - 0.499 = 0.001$ suchen und k_1 ermitteln. Der gesuchte Wert beträgt dann $k_1 = 3.09$. Die notwendige Stichprobengröße ist dann :

$$n = \left(\frac{k \cdot \sigma}{\bar{x} - \mu}\right)^2 = \left(\frac{-3.09 \cdot 3}{5 - 6}\right)^2 = 85.93$$

Notwendig ist dann eine Stichprobengröße von $n = 86$.

Stichprobenverteilung der relativen Häufigkeit (des Anteilswertes) Die relative Häufigkeit eines Ereignisses A ist ebenfalls eine Stichprobenfunktion: Von Stichprobe zu Stichprobe variiert der Anteil $h(A)$ des Eintretens von A . Jede der voneinander unabhängigen Versuchsdurchführungen bzw. Stichprobenziehungen läßt sich durch eine Stichprobenvariable X_i beschreiben, mit den Werten

$$x_i = \begin{cases} 1 & \text{falls } A \text{ eintritt} \\ 0 & \text{andernfalls.} \end{cases}$$

Dann ist $Y = \sum_{i=1}^n X_i$ die Anzahl derjenigen Versuche, bei denen das Ereignis A eintritt, also $Y = n(A)$ (absolute Häufigkeit). Die Stichprobenfunktion der relativen Häufigkeit $h(A)$ ist dann

$$h(A) = \frac{Y}{n} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

Die relative Häufigkeit (der Anteilswert) der mit 1 kodierten beobachteten Ergebnisse an allen Ergebnissen entspricht dem arithmetischen Mittel der 0-1-kodierten Ergebnisse.

Die Verteilung von Y folgt einer Binomialverteilung: die Zufallsvariable X realisiert bei n Bernoulli-Experimenten das Ereignis A (mit 1 kodiert) bei jedem Versuch mit einer Wahrscheinlichkeit $P(A) = p$. Somit kann die **Verteilung der relativen Häufigkeit** $f\left(\frac{y}{n}\right)$ mit

$$f\left(\frac{y}{n}\right) = P\left(\frac{Y}{n} = \frac{y}{n}\right) = \frac{1}{n} \cdot B(n, p)$$

angegeben werden. Der **Erwartungswert** und die **Varianz** sind dann:

$$\begin{aligned} E\left(\frac{Y}{n}\right) &= p \\ \text{Var}\left(\frac{Y}{n}\right) &= \frac{p(1-p)}{n} \end{aligned}$$

Da die relative Häufigkeit dem arithmetischen Mittel der 0-1-kodierten Ergebnisse entspricht, läßt sich die Stichprobenverteilung des arithmetischen Mittels auch auf die der relativen Häufigkeit übertragen, es gilt bei großem n :

$\frac{Y}{n}$ folgt approximativ der Verteilung $N(p, \sqrt{p(1-p)/n})$
und die absolute Häufigkeit Y folgt approximativ der Verteilung $N(np, \sqrt{np(1-p)})$.

6 Das Testen von statistischen Hypothesen

Unter statistischen Hypothesen versteht man Aussagen bzw. Vermutungen über die Charakteristik von Zufallsvariablen bzw. Untersuchungsvariablen. Je nach dem welche Eigenschaft von Interesse ist, unterscheidet man zwischen Parameterhypothesen, Verteilungshypothesen und Unabhängigkeitshypothesen. Diese statistischen Hypothesen werden dann anhand eines hierfür erhobenen empirischen Befundes überprüft. Ein statistischer Test ist also ein Verfahren, welches darüber entscheidet, ob eine statistische Hypothese auf Grund von Stichprobenergebnissen akzeptiert werden kann oder verworfen werden sollte.

- Gegenstand von *Parametertests* sind Parameterhypothesen, die Vermutungen über einen Zustand einer statistischen Maßzahl in einer Grundgesamtheit anstellen. Ein Beispiel hierfür wäre die Hypothese, daß der mittlere Wert einer Variablen Y in der Grundgesamtheit nicht größer als ein Wert a ist, oder z.B. der Anteil von Studienabbrechern bei den Studenten der Sozialwissenschaften genauso groß ist, wie bei den Studenten der Wirtschaftswissenschaften.
- *Anpassungstests* beziehen sich auf Verteilungshypothesen, wie z.B. auf die Vermutung, daß eine Fehlervariable in einem Modell normalverteilt ist.
- *Unabhängigkeitstests* betreffen Hypothesen, die Aussagen über die Unabhängigkeit von Variablen formulieren.

Ausgangspunkt eines Hypothesentests bildet stets die Formulierung einer *Nullhypothese*.

Nullhypothese: Die Angabe eines hypothetischen Wertes θ_0 eines Parameters θ bezeichnet man als **Nullhypothese**:

$$H_0 : \theta = \theta_0$$

Liegt eine weitere Hypothese in Bezug auf den Parameter θ vor, welche der Nullhypothese entgegen steht, spricht man in der Regel von einer **Alternativhypothese** H_a . Wenn außer der Nullhypothese keine weitere Hypothese formuliert wird, wird das Vorgehen, das angibt, wie auf

Stichprobenbasis über die Beibehaltung der Hypothese H_0 oder ihre Ablehnung zu entscheiden ist, **Signifikanztest** genannt.

Das formale Vorgehen eines Hypothesentests läßt sich wie folgt beschreiben (vgl. Tiede 2000):

1. **Hypothesenformulierung:** Die Nullhypothese H_0 und gegebenenfalls auch eine Alternativhypothese H_a muß formuliert werden.

Es wird beispielsweise als Nullhypothese für den arithmetischen Mittelwert einer Grundgesamtheit μ_0 der Wert 10 abgenommen, und als Alternativhypothese H_a Werte für den Mittelwert, die größer als 10 sind $\mu_a > 10$. Man schreibt dafür:

$$\begin{aligned} H_0 : \mu_0 &= 10 \\ H_a : \mu_a &> 10 \end{aligned}$$

2. **Aufstellung einer geeigneten Prüfvariablen:** Als Prüfvariable muß ein Konstrukt in Form einer Zufallsvariable verwendet werden, welche für die Prüfung der Nullhypothese geeignet ist. Dabei müssen zwei Bedingungen erfüllt werden:

1. Die Werte einer Prüfvariablen müssen mit der Fragestellung bzw. den Aussagen der Hypothesen korrespondieren. So eignet sich die Zufallsvariable „arithmetische Mittel“ \bar{X} zur Prüfung des arithmetischen Mittels μ in der Grundgesamtheit, aber nicht für das Überprüfen einer u-förmigen Verteilung.
2. Die Verteilungsfunktion der Prüfvariablen muß „unter der Nullhypothese“ bekannt sein. So ist es notwendig, den Wertebereich der Prüfvariablen in einen Annahmebereich und einen Rückweisungsbereich für die Nullhypothese aufzuteilen. Der *Annahmebereich* umfaßt den Wertebereich der Prüfvariable, die zur Annahme der Nullhypothese, der Rückweisungsbereich jenen Bereich der Werte die zur Rückweisung der Nullhypothese führen.

3. **Festlegung des Annahme- und Rückweisungsbereich** Vor Ziehung einer oder mehrerer Stichproben sollte klar sein, bei welcher Stichprobenrealisation der Prüfvariablen die Nullhypothese angenommen oder verworfen wird.
4. **Ziehung der Stichprobe(n)** Nach der Hypothesenformulierung, Festlegung von Prüfvariable sowie Annahme- und Rückweisungsbereich wird eine oder mehrere Stichproben gezogen.
5. **Annahme oder Ablehnung der Nullhypothese** Je nach realisiertem Stichprobenwert, wird die Nullhypothese auf Basis des vorher konstruierten Annahme- und Rückweisungsbereich angenommen und verworfen.

Beispiel: Vor zehn Jahren zeigten mehrmalige Erhebungen an der Ruhr-Universität Bochum, daß der durchschnittliche tägliche Zigarettenverbrauch von Studenten bei konstant 10 Zigaretten/pro Kopf lag, die Standardabweichung betrug vier Zigaretten. Es wird nun vermutet, daß sich innerhalb der zehn Jahre der durchschnittliche Zigarettenverbrauch nicht verändert hat. Um dies zu prüfen, wird eine aktuelle Erhebung in Form einer Stichprobe mit einem Umfang von 150 Studenten gezogen.

Im ersten Schritt erfolgt nun die Formulierung der Nullhypothese. Für die Grundgesamtheit galt vor zehn Jahren:

$$\mu = 10 \text{ und } \sigma = 4$$

Die Nullhypothese behauptet, daß der Mittelwert unverändert bei 10 Zigaretten liegt:

$$H_0 : \mu_0 = 10 \text{ und } n = 150$$

Als Prüfvariable eignet sich die Stichprobenvariable \bar{X} , da eine relativ starke Abweichung eines empirischen Mittelwerts \bar{x} von dem in der Nullhypothese formulierten Durchschnittswert μ eher gegen die Beibehaltung der Hypothese sprechen. Des weiteren ist die Wahrscheinlichkeitsverteilung der Stichprobenverteilung unter der Nullhypothese, also $\mu = \mu_0$ bekannt, \bar{X} folgt approximativ $N(\mu, \frac{\sigma}{\sqrt{n}})$. In dem Beispiel folgt die Prüfvariable \bar{X} approximativ $N(10, \frac{4}{\sqrt{150}}) = N(10, 0.327)$.

Nach der Bestimmung einer geeigneten Prüfvariablen erfolgt die Festlegung des Annahme- und Rückweisungsbereiches durch einen Rückweisungspunkt. Hierbei wird ein sogenanntes **Signifikanzniveau** festgelegt, welches bei der Testentscheidung die Wahrscheinlichkeit für den Fehler beschreibt, die in Wahrheit zutreffende Nullhypothese zu verwerfen. Dieser Fehler wird als **α -Fehler** oder als **Fehler 1. Art** bezeichnet. Man unterscheidet in diesem Zusammenhang zudem zwischen einem **einseitigen** und einem **zweiseitigen Test**. Besteht das Rückweiskriterium in einer Abweichung sowohl unterhalb als auch oberhalb von μ_0 , wird man einen zweiseitigen Test durchführen. Formuliert man dagegen die Alternativhypothese so, daß nur Abweichungen nach einer bestimmten Seite hin zu einer Zurückweisung der Nullhypothese führen, verwendet man einen einseitigen Test.

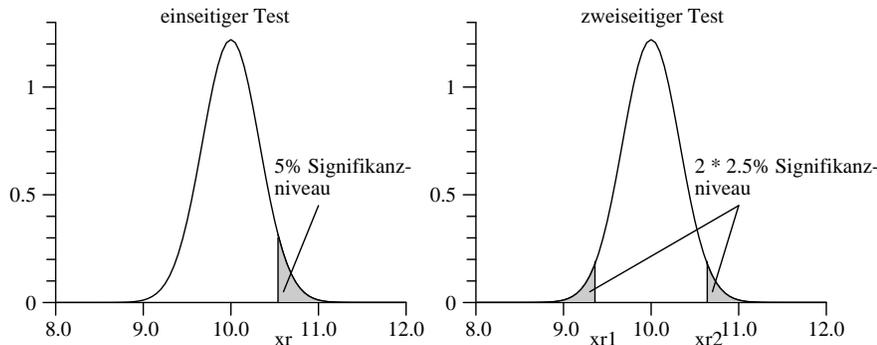


ABBILDUNG 21: Beispiel für einen einseitigen und zweiseitigen Hypothesentest

In unserem Beispiel gehen wir von einem zweiseitigen Signifikanzniveau von $\alpha = 5\%$ aus und legen die Rückweisungspunkte \bar{x}_{r1} und \bar{x}_{r2} durch folgende Ungleichung fest⁶:

$$P(\bar{x}_{r1} \geq \bar{X} \geq \bar{x}_{r2}) = 0.05 = P(k_{r1} \geq K \geq k_{r2}) = P(-1.96 \geq K \geq 1.96)$$

Die Rückweisungspunkte \bar{x}_{r1} und \bar{x}_{r2} ergeben sich dann aus der Beziehung:

$$k_r = \frac{\bar{x}_r - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$\bar{x}_r = \mu_0 + k_r \cdot \frac{\sigma}{\sqrt{n}}$$

Man erhält in unserem Beispiel die Rückweisungspunkte

$$\bar{x}_{r1} = 10 - 1.96 \cdot 0.327 = 9.359$$

$$\bar{x}_{r2} = 10 + 1.96 \cdot 0.327 = 10.641$$

⁶Die Werte für k_{r1} und k_{r2} lassen sich aus der Tabelle 11 im Anhang bei $P = 0.025$ ablesen.

Im nächsten Schritt wird eine Stichprobe gezogen und der durchschnittliche Zigarettenverbrauch berechnet. Sollte dann der empirische Wert größer als 10.641 oder kleiner als 9.359 sein, wird nach Maßgabe dieses Tests die Nullhypothese bei einem Signifikanzniveau von $\alpha = 5\%$ verworfen.

Neben der Methode, die Testentscheidung auf Basis eines vorher festzulegenden Annahme- und Rückweisungsbereichs abhängig zu machen, wird häufig auch ein anderes Verfahren verwendet. So besteht die Möglichkeit, daß man die Stichprobenrealisation der Prüfvariablen, also in diesem Fall den in der Stichprobe ermittelten durchschnittliche Zigarettenverbrauch, als Rückweisungspunkt festlegt und dann das **empirische Signifikanzniveau** erhält. Dieses Verfahren wird in den gängigen statistischen Softwarepaketen verwendet. Wird z.B. bei einem Stichprobenbefund mit einem sehr kleinen empirischen Signifikanzniveau die Nullhypothese zurückgewiesen, ist die Wahrscheinlichkeit für den Fehler 1. Art ebenfalls nur so groß wie das Signifikanzniveau.

Ergibt sich in unserem Beispiel ein Stichprobenbefund von durchschnittlich 11 Zigaretten, so beträgt das empirische Signifikanzniveau 0.11%:

$$P(\bar{X} \geq 11) = P\left(K \geq \frac{11 - 10}{0.327}\right) = P(K \geq 3.058) = 0.0011$$

Die Rückweisung der Nullhypothese wäre in diesem Fall mit einer sehr geringen Irrtumswahrscheinlichkeit und damit Fehler 1. Art verbunden.

6.1 Richtige und falsche Testentscheidungen

Bei Testentscheidungen können prinzipiell zwei Fehlentscheidungen vorliegen:

- Eine zutreffende Nullhypothese wird abgelehnt. Dieser Fall wurde weiter oben schon erwähnt und als **α -Fehler** bzw. **Fehler 1. Art** bezeichnet. Die Wahrscheinlichkeit für diese Fehlentscheidung läßt sich als bedingte Wahrscheinlichkeit formulieren: Es ist die Wahrscheinlichkeit, die Nullhypothese unter der Bedingung abzulehnen, daß die Nullhypothese in Wahrheit doch zutrifft, also $\alpha = P(R_0|H_0)$.
- Eine unzutreffende Nullhypothese wird angenommen. Diese Fehlentscheidung wird in der Regel als **β -Fehler** oder **Fehler 2. Art** bezeichnet. Auch die Wahrscheinlichkeit für diese Fehlentscheidung läßt sich als bedingte Wahrscheinlichkeit formulieren: Es ist die Wahrscheinlichkeit, die Nullhypothese unter der Bedingung anzunehmen, daß die Alternativhypothese in Wahrheit zutrifft, also $\beta = P(A_0|H_a)$.

Die Wahrscheinlichkeit für den β -Fehler läßt sich offensichtlich nur dann bestimmen, wenn die Verteilungsfunktion der Prüfvariablen unter der Alternativhypothese bekannt ist. Wird, wie bei einem Signifikanztest, die Alternativhypothese nicht oder nur als Bereichshypothese formuliert, ist eine Angabe des β -Fehlers nicht möglich.

Die vier Entscheidungsmöglichkeiten sind in Tabelle 6 dargestellt.

TABELLE 6: Fehlerarten bei Testentscheidungen

Testentscheidung	Realität	
	H_0 trifft zu	H_0 trifft nicht zu
Annahme von H_0	richtige Entscheidung, $P(A_0 H_0) = 1 - \alpha$	Ent- Fehler 2. Art, $P(A_0 H_a) = \beta$
Ablehnung von H_0	Fehler 1. Art, $P(R_0 H_0) = \alpha$	richtige Entscheidung, $P(R_0 H_a) = 1 - \beta$

Beispiel: Der Zusammenhang zwischen Fehler 1. Art und Fehler 2. Art läßt sich am anschaulichsten anhand von zwei Verteilungen demonstrieren. Die Abbildung 22 zeigt beispielhaft die Wahrscheinlichkeitsverteilungen unter der Nullhypothese $N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$ und der Alternativhypothese $N\left(\mu_a, \frac{\sigma}{\sqrt{n}}\right)$.

In der Nullhypothese wird davon ausgegangen, daß z.B. der durchschnittliche Zigarettenverbrauch bei $\mu_0 = 10$ Zigaretten pro Tag liegt. Die formulierte Alternativhypothese geht nur von $\mu_a = 9$ Zigaretten aus. Die beiden Verteilungen sollen sich nur in Bezug auf ihren Erwartungswert unterscheiden.

Der **Fehler 1. Art**, also α , wird frei festgelegt. In diesem Fall wird ein zweiseitiges Signifikanzniveau angenommen, so daß sowohl auf der linken als auch der rechten Seite der Verteilung unter H_0 der Bereich $\alpha/2$ abgetragen wird. Der Annahmehbereich A_0 liegt dazwischen und die Wahrscheinlichkeit, eine zutreffende Nullhypothese anzunehmen beträgt $P(A_0|H_0) = 1 - \alpha$. In dem obigen Zigarettenbeispiel haben wir ein $\alpha = 5\%$ gewählt, es ergibt sich also für $P(A_0|H_0)$ ein Wert von $1 - 0.05 = 0.95$.

Der **Fehler 2. Art** beschreibt die Wahrscheinlichkeit für die Fehlentscheidung, die Nullhypothese anzunehmen, obwohl die Alternativhypothese korrekt wäre, also $P(A_0|H_a) = \beta$. Da wir eine Alternativhypothese spezifiziert haben, läßt sich β bestimmen. Wir gehen wieder von dem Zigarettenbeispiel aus und nehmen an, daß \bar{X} nach $N(\mu_0, \frac{\sigma}{\sqrt{n}}) = N(10, 0.327)$ und alternativ nach $N(\mu_a, \frac{\sigma}{\sqrt{n}}) = N(9, 0.327)$ verteilt ist. Die Rückweisungspunkte erhielten wir durch

$$P(\bar{x}_{r1} \geq \bar{X} \geq \bar{x}_{r2}) = 0.05 = P(k_{r1} \geq K \geq k_{r2}) = P(-1.96 \geq K \geq 1.96)$$

und

$$\begin{aligned}\bar{x}_r &= \mu_0 + k_r \cdot \frac{\sigma}{\sqrt{n}} \\ \bar{x}_{r1} &= 10 - 1.96 \cdot 0.327 = 9.359 \\ \bar{x}_{r2} &= 10 + 1.96 \cdot 0.327 = 10.641\end{aligned}$$

Gesucht ist jetzt allerdings β , die Wahrscheinlichkeit $P(\bar{x}_{r1} \leq \bar{X} \leq \bar{x}_{r2})$ für die Verteilung $N(\mu_a, \frac{\sigma}{\sqrt{n}}) = N(9, 0.327)$. Es ergibt sich dann

$$P(9.359 \leq \bar{X} \leq 10.641) = P\left(\frac{9.359 - 9}{0.327} \leq K \leq \frac{10.641 - 9}{0.327}\right)$$

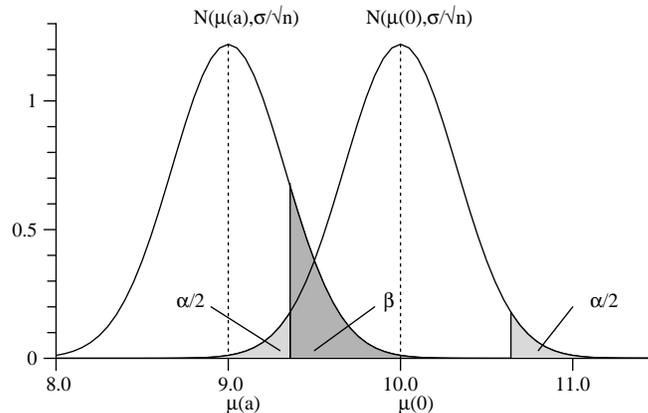


ABBILDUNG 22: Wahrscheinlichkeiten für den Fehler 1. und 2. Art

$$= P(1.098 \leq K \leq 5.018)$$

Die Wahrscheinlichkeit für β ist somit $P(K \geq 1.098) - P(K \geq 5.018) = 0.1357 - 0 = 0.1357$. Wenn in Wirklichkeit der mittlere Zigarettenverbrauch $\mu_a = 9$ ist, beträgt unter diesen Umständen die Wahrscheinlichkeit, die Nullhypothese dennoch anzunehmen 13.57%.

Aus der Abbildung 22 geht zudem hervor, daß die Wahrscheinlichkeit für den β -Fehler ansteigt, wenn unter den gleichen Umständen der Alternativmittelwert μ_a näher an den Mittelwert der Nullhypothese μ_0 heranrückt. Dies ist auch nicht weiter verwunderlich, da eine Diskriminierung mit zunehmender Annäherung immer schwieriger wird.

Des Weiteren besteht ein Zusammenhang zwischen Stichprobenumfang n und der Höhe des β -Fehlers. Je größer der Stichprobenumfang ist, desto kleiner wird σ/\sqrt{n} und umso schmäler und steiler werden die Verteilungen. Hierdurch reduziert sich auch der β -Fehler. Die Wahrscheinlichkeit $P(R_0|H_a) = 1 - \beta$ ist dann die Wahrscheinlichkeit, daß der β -Fehler vermieden wird, in unserem Beispiel ist dies $1 - 0.1357 = 0.8643$, also ca. 87%.

Für einen geeigneten Test sollte der β -Fehler bei einem vorgegeben Signifikanzniveau α möglichst klein sein. Zur Einschätzung eines Testes wird üblicherweise die Wahrscheinlichkeit, sich richtigerweise für H_a zu entscheiden verwendet, also $1 - \beta$. Diese Wahrscheinlichkeit wird oft als Funktion der durch die Nullhypothese und Alternativhypothese festgelegten Parameter dargestellt, und mit **Gütefunktion des Tests** bezeichnet.

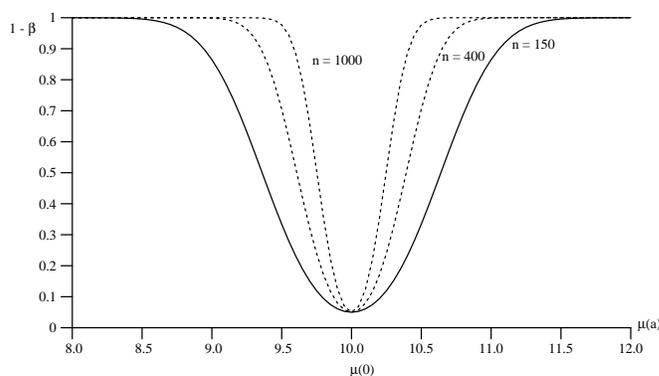


ABBILDUNG 23: Gütefunktion des Tests für $H_0 : \mu = \mu_0$ gegen $H_a : \mu \neq \mu_0$ bei drei verschiedenen Stichprobengrößen und $\sigma = 4$

Die Abbildung 23 zeigt die Gütefunktion des Tests aus unserem Beispiel für verschiedene Stichprobengrößen. Es läßt sich erkennen, daß mit der Annäherung von μ_a an μ_0 die Wahrscheinlichkeit $1 - \beta$ absinkt und in dem Punkt μ_0 den Wert α hat. Die Wahrscheinlichkeit, eine vorhandene Differenz zwischen wahren und hypothetischen Parameterwert aufzudecken, wird umso kleiner, je kleiner die Differenz zwischen beiden ist. Zudem wird mit zunehmendem Stichprobenumfang die Standardabweichung kleiner und die Gütefunktion immer steiler. Die Wahrscheinlichkeit, vorhandene Unterschiede aufzudecken, wächst mit dem Stichprobenumfang und die Wahrscheinlichkeit für den Fehler 2. Art nimmt ab.

7 Das Schätzen von Parametern

Die Parameterschätzung ist ein Verfahren, bei dem für unbekannte Parameter der Grundgesamtheit unter Verwendung des Befundes einer Stichprobe geeignete Schätzwerte bestimmt werden. Wird ein Grundgesamtheitsparameter lediglich durch einen einzelnen Wert geschätzt, handelt es sich um eine Punktschätzung, wird der Parameter durch einen Bereich geschätzt, spricht man von einer Intervallschätzung.

7.1 Punktschätzung

7.1.1 Schätzfunktionen und ihre Eigenschaften

Die Güte einer Punktschätzung bezieht sich auf die **Schätzfunktion**, welche eine Stichprobenfunktion ist.

Schätzfunktion: Eine Stichprobenfunktion $\hat{\theta}(X_1, \dots, X_n)$, deren Wert $\hat{\theta}(x_1, \dots, x_n)$ als Schätzwert für einen theoretischen Parameter oder Grundgesamtheitsparameter θ dienen soll, heißt **Schätzfunktion** oder kurz **Schätzer** (für θ). Häufig wird auch $\hat{\theta}$ für $\hat{\theta}(X_1, \dots, X_n)$ bzw. $\hat{\theta}(x_1, \dots, x_n)$ geschrieben. $\hat{\theta}$ besitzt daher die Eigenschaften einer Zufallsvariablen.

7.1.2 Eigenschaften von Schätzfunktionen

Bei Schätzfunktionen interessiert man sich dafür, wie eng sich die Realisationen um den zu schätzenden Parameter konzentrieren. In Anlehnung an die Varianz mißt man dies mit dem quadratischen Abstand, mit dem im Mittel zwischen zu schätzenden Parameter und Schätzwert zu rechnen ist.

Der durchschnittliche quadratische Abstand zwischen Schätzfunktion $\hat{\theta}$ und dem Parameter θ oder der mittlere quadratische Fehler von $\hat{\theta}$ bezüglich θ ist

$$MQF(\hat{\theta}, \theta) = E((\hat{\theta}(X_1, \dots, X_n) - \theta)^2)$$

Effizienz: Sind $\hat{\theta}_1$ und $\hat{\theta}_2$ zwei Schätzfunktionen für denselben Parameter θ , so heißt $\hat{\theta}_1$ **effizienter** als $\hat{\theta}_2$, wenn für alle Werte von θ

$$MQF(\hat{\theta}_1, \theta) \leq MQF(\hat{\theta}_2, \theta)$$

Erwartungstreue: Eine Schätzfunktion $\hat{\theta}(X_1, \dots, X_n)$ für den Parameter θ heißt **unverzerrt** oder auch **erwartungstreu**, wenn ihr **Bias** $b(\hat{\theta}, \theta) = E(\hat{\theta}) - \theta$ verschwindet:

$$b(\hat{\theta}, \theta) = 0$$

bzw.

$$E(\hat{\theta}) = E(\hat{\theta}(X_1, \dots, X_n)) = \theta$$

Andernfalls heißt $\hat{\theta}$ verzerrt. Asymptotisch erwartungstreu ist eine Schätzfunktion, wenn sie die angegebene Bedingung nur bei $n \rightarrow \infty$ erfüllt.

Die Schätzung des arithmetischen Mittels der Grundgesamtheit μ durch das arithmetische Mittel der Stichprobe \bar{x} also $\mu \approx \hat{\mu} = \bar{x}$ ist erwartungstreu:

$$E(\hat{\mu}) = E(\bar{X}) = \mu$$

Da der Anteilswert bzw. die relative Häufigkeit ebenso als spezielles arithmetisches Mittel angesehen werden kann gilt dies auch für diese Schätzung.

Die Schätzung der Varianz der Grundgesamtheit durch die Varianz aus der Stichprobe $\sigma^2 \approx \hat{\sigma}^2 = s^2$ ist hingegen nicht erwartungstreu, es ergibt sich, wie sich zeigen läßt, für den Erwartungswert der Schätzfunktion nicht σ^2 , sondern ein etwas kleinerer Wert:

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

so daß die unbekannte Varianz der Grundgesamtheit etwas unterschätzt wird. Man erhält eine erwartungstreue Schätzung hingegen durch die Schätzfunktion:

$$\hat{\sigma}_1^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

Daher wird die Stichprobenvarianz bei kleineren Stichprobenumfängen häufig so bestimmt, daß die Summe der Quadrate durch $n - 1$ und nicht durch n dividiert wird.

Konsistenz: $\hat{\theta}(X_1, \dots, X_n)$ heißt eine konsistente Schätzfunktion für den Parameter θ , falls

$$\lim_{n \rightarrow \infty} MQF(\hat{\theta}, \theta) = 0$$

oder wenn gilt

$$\lim_{n \rightarrow \infty} b(\hat{\theta}, \theta) = 0 \quad \text{und} \quad \lim_{n \rightarrow \infty} Var(\hat{\theta}) = 0$$

Die Konsistenz ist quasi eine Versicherung, daß ein größerer Aufwand bzw. eine größere Stichprobe bessere Schätzergebnisse liefert. Sie ist eine minimale Anforderung an eine Schätzfunktion.

7.2 Intervallschätzung

Neben der Angabe eines einzelnen Schätzwertes ist es in der Regel auch sinnvoll, anzugeben, in welchen Grenzen der Grundgesamtheitsparameter bei vorgegebener Sicherheit liegt. Man spricht auch von einem **Vertrauensbereich**. Das konkrete Vorgehen wird hierbei am Beispiel des arithmetischen Mittels demonstriert. So soll bei einem vorgegebenen Vertrauensniveau (Konfidenzniveau) der unbekannte Parameter μ in einem Bereich $[\hat{\mu}_u, \hat{\mu}_o]$ liegen.

Die Stichprobenverteilung des arithmetischen Mittels folgt bei großem n approximativ der Verteilung $N(\mu, \frac{\sigma}{\sqrt{n}})$. Die Standardabweichung in der Grundgesamtheit σ wird dabei durch die Standardabweichung in der Stichprobe s geschätzt. Wir geben nun ein Wahrscheinlichkeitsintervall $1 - \alpha$ vor, in welchem der Stichprobenmittelwert \bar{x} liegen soll.

$$P(\bar{x}_1 \leq \bar{X} \leq \bar{x}_2) = 1 - \alpha$$

Mit $\bar{x} = \mu + k \cdot \frac{\sigma}{\sqrt{n}}$ ergibt sich:

$$P(\mu + k_1 \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + k_2 \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Durch geeignetes Umformen der Gleichung⁷ erhält man

$$P(\bar{X} - k_2 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - k_1 \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Das durch die zwei Zufallsvariablen $\bar{X} - k_2 \frac{\sigma}{\sqrt{n}}$ sowie $\bar{X} - k_1 \frac{\sigma}{\sqrt{n}}$ gebildete Intervall

$$\left[\bar{X} - k_2 \frac{\sigma}{\sqrt{n}}, \bar{X} - k_1 \frac{\sigma}{\sqrt{n}} \right]$$

⁷In der Gleichung wird der Term $-\mu - \bar{X}$ addiert:

$$P(-\bar{X} + k_1 \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + k_2 \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Nach Multiplikation mit (-1) erhält man:

$$P(\bar{X} - k_1 \frac{\sigma}{\sqrt{n}} \geq -\mu \geq \bar{X} - k_2 \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

bzw.

$$P(\bar{X} - k_2 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - k_1 \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

wird Zufallsintervall genannt. Vor Ziehung einer Stichprobe kann das Zufallsintervall als Wahrscheinlichkeit $1 - \alpha$ interpretiert werden, daß \bar{X} in der Stichprobe einen Wert realisiert, der zu einem realisierten Intervall

$$\left[\bar{x} - k_2 \frac{\sigma}{\sqrt{n}}, \bar{x} - k_1 \frac{\sigma}{\sqrt{n}} \right]$$

führt, das μ einschließt. Liegt eine Stichprobenrealisation vor, kann der Wert $1 - \alpha$ des Zufallsintervall als Vertrauens- oder Konfidenzintervall interpretiert werden, daß das realisierte Intervall Grenzen hat, die den Grundgesamtheitsparameter μ einschließen.

Konfidenzintervall: \bar{X} sei eine Zufallsvariable mit Erwartungswert μ . Dann ist für große n die Realisierung des Zufallsintervalls

$$\left[\bar{X} - k_2 \frac{\sigma}{\sqrt{n}}, \bar{X} - k_1 \frac{\sigma}{\sqrt{n}} \right] = [\hat{\mu}_u, \hat{\mu}_o]$$

ein approximatives $(1 - \alpha)$ -Konfidenzintervall für μ . n gilt nach einer Faustregel als groß, wenn $n \geq 30$. Als gesamte Ungleichung wird häufig auch geschrieben:

$$\text{Konf} (\hat{\mu}_u \leq \mu \leq \hat{\mu}_o)$$

Mit einem Vertrauen von $1 - \alpha$ liegt μ im Bereich $[\hat{\mu}_u, \hat{\mu}_o]$.

Aus der Gleichung für das Konfidenzintervall läßt sich folgendes ableiten:

1. Bei einer Erhöhung des Stichprobenumfangs n wird der Vertrauensbereich enger und damit die Aussage über den Grundgesamtheitsparameter schärfer.
2. Je größer die Streuung σ ist desto breiter wird das Konfidenzintervall.
3. Bei Verringerung des Vertrauensniveaus $(1 - \alpha)$ wird der Vertrauensbereich enger. Unter Lasten des Vertrauensniveaus können also für den Parameter der Grundgesamtheit schärfere Aussagen gemacht werden.

Beispiel: (aus Schlittgen 1998, S.310) Der Geschäftsführer einer Reparaturwerkstatt interessiert sich für die durchschnittliche Reparaturdauer der bei ihm in Auftrag gegebenen Pkw. Eine Stichprobe von 99 zufällig gezogenen Karteikarten, auf denen alte Aufträge vermerkt sind, liefert folgende Maßzahlen: das arithmetische Mittel beträgt $\bar{x} = 157.6$ min, in der Stichprobe beträgt die Standardabweichung $s = 143.65$. In welchem Bereich liegt bei einem Vertrauen von 0.95 die durchschnittliche Reparaturdauer? Für das Konfidenzniveau $1 - \alpha = 0.95$ betragen die Werte für die Quantile der Standardnormalverteilung $k_1 = -1.96$ und $k_2 = 1.96$.⁸ Man erhält dann:

$$\begin{aligned} \hat{\mu}_u &= 157.6 - 1.96 \cdot \frac{143.65}{\sqrt{99}} = 129.3 \\ \hat{\mu}_o &= 157.6 + 1.96 \cdot \frac{143.65}{\sqrt{99}} = 185.9 \\ &= [129.3, 185.9] \end{aligned}$$

Bei einem Vertrauensniveau von 0.95 schließt also das Intervall $[129.3, 185.9]$ den Wert für die unbekannte durchschnittliche Reparaturdauer, μ , mit ein.

⁸Man erhält die Werte aus der tabellierten Verteilungsfunktion im Anhang. Die Werte sind bei $P = 0.025$ abzulesen.

B Tabellierte Werte der Standardnormalverteilung

TABELLE 10: Standardnormalverteilung $N(0, 1)$ - Dichtefunktion

K	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.39894	.39892	.39886	.39876	.39862	.39844	.39822	.39797	.39767	.39733
0.1	.39695	.39654	.39608	.39559	.39505	.39448	.39387	.39322	.39253	.39181
0.2	.39104	.39024	.38940	.38853	.38762	.38667	.38568	.38466	.38361	.38251
0.3	.38139	.38023	.37903	.37780	.37654	.37524	.37391	.37255	.37115	.36973
0.4	.36827	.36678	.36526	.36371	.36213	.36053	.35889	.35723	.35553	.35381
0.5	.35207	.35029	.34849	.34667	.34482	.34294	.34105	.33912	.33718	.33521
0.6	.33322	.33121	.32918	.32713	.32506	.32297	.32086	.31874	.31659	.31443
0.7	.31225	.31006	.30785	.30563	.30339	.30114	.29887	.29659	.29431	.29200
0.8	.28969	.28737	.28504	.28269	.28034	.27798	.27562	.27324	.27086	.26848
0.9	.26609	.26369	.26129	.25888	.25647	.25406	.25164	.24923	.24681	.24439
1.0	.24197	.23955	.23713	.23471	.23230	.22988	.22747	.22506	.22265	.22025
1.1	.21785	.21546	.21307	.21069	.20831	.20594	.20357	.20121	.19886	.19652
1.2	.19419	.19186	.18954	.18724	.18494	.18265	.18037	.17810	.17585	.17360
1.3	.17137	.16915	.16694	.16474	.16256	.16038	.15822	.15608	.15395	.15183
1.4	.14973	.14764	.14556	.14350	.14146	.13943	.13742	.13542	.13344	.13147
1.5	.12952	.12758	.12566	.12376	.12188	.12001	.11816	.11632	.11450	.11270
1.6	.11092	.10915	.10741	.10567	.10396	.10226	.10059	.09893	.09728	.09566
1.7	.09405	.09246	.09089	.08933	.08780	.08628	.08478	.08329	.08183	.08038
1.8	.07895	.07754	.07614	.07477	.07341	.07206	.07074	.06943	.06814	.06687
1.9	.06562	.06438	.06316	.06195	.06077	.05959	.05844	.05730	.05618	.05508
2.0	.05399	.05292	.05186	.05082	.04980	.04879	.04780	.04682	.04586	.04491
2.1	.04398	.04307	.04217	.04128	.04041	.03955	.03871	.03788	.03706	.03626
2.2	.03547	.03470	.03394	.03319	.03246	.03174	.03103	.03034	.02965	.02898
2.3	.02833	.02768	.02705	.02643	.02582	.02522	.02463	.02406	.02349	.02294
2.4	.02239	.02186	.02134	.02083	.02033	.01984	.01936	.01888	.01842	.01797
2.5	.01753	.01709	.01667	.01625	.01585	.01545	.01506	.01468	.01431	.01394
2.6	.01358	.01323	.01289	.01256	.01223	.01191	.01160	.01130	.01100	.01071
2.7	.01042	.01014	.00987	.00961	.00935	.00909	.00885	.00861	.00837	.00814
2.8	.00792	.00770	.00748	.00727	.00707	.00687	.00668	.00649	.00631	.00613
2.9	.00595	.00578	.00562	.00545	.00530	.00514	.00499	.00485	.00470	.00457
3.0	.00443	.00430	.00417	.00405	.00393	.00381	.00370	.00358	.00348	.00337
3.1	.00327	.00317	.00307	.00298	.00288	.00279	.00271	.00262	.00254	.00246
3.2	.00238	.00231	.00224	.00216	.00210	.00203	.00196	.00190	.00184	.00178
3.3	.00172	.00167	.00161	.00156	.00151	.00146	.00141	.00136	.00132	.00127
3.4	.00123	.00119	.00115	.00111	.00107	.00104	.00100	.00097	.00094	.00090
3.5	.00087	.00084	.00081	.00079	.00076	.00073	.00071	.00068	.00066	.00063
3.6	.00061	.00059	.00057	.00055	.00053	.00051	.00049	.00047	.00046	.00044
3.7	.00042	.00041	.00039	.00038	.00037	.00035	.00034	.00033	.00031	.00030
3.8	.00029	.00028	.00027	.00026	.00025	.00024	.00023	.00022	.00021	.00021
3.9	.00020	.00019	.00018	.00018	.00017	.00016	.00016	.00015	.00014	.00014

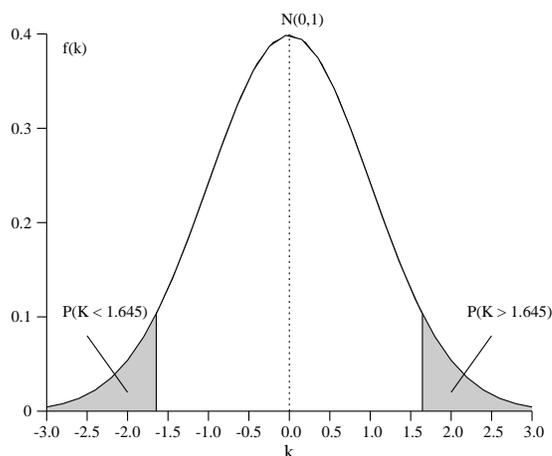


ABBILDUNG 24: Flächen unterhalb der Normalverteilung

TABELLE 11: Standardnormalverteilung $N(0, 1)$ - Verteilungsfunktion: Werte für $P(K \geq k)$

k	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003

Literatur

Boeltken, F. 1976. *Auswahlverfahren*. Teubner, Stuttgart.

Pötter, U. and Rendtel, U. 1993. Über Sinn und Unsinn von Repräsentativitätsstudien. *Allgemeines Statistisches Archiv*, 77:260–280.

Rohwer, G. 2000. Epistemische und aleatorische Wahrscheinlichkeit. Skript zur Veranstaltung 'Wissenschaftstheorie II' Februar 2000. Version 1. Ruhr-Universität Bochum. Fakultät für Sozialwissenschaften.

Schlittgen, R. 1998. *Einführung in die Statistik*. Oldenbourg, München - Wien.

Tiede, M. and Voss, W. 2000. *Schließen mit Statistik - Verstehen*. Oldenbourg, München - Wien.