

Wahrscheinlichkeitsrechnung und Statistik

Eine Kurzeinführung im Rahmen der Vorlesung “Mathematik und Statistik für Molekularbiologen”

STEFAN BORESCH

stefan@mdy.univie.ac.at, <http://www.mdy.univie.ac.at/en/sbhome.html>

Molecular Dynamics and Biomolecular Simulation Group,
Institut für Theoretische Chemie und Molekulare Strukturbiologie,
Universität Wien, Währingerstraße 17, 1090 Wien, Austria

17. Januar 2003

Copyright (c) 2002 Stefan Boesch

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

Although every reasonable effort has been made to incorporate accurate and useful information into this booklet, the copyright holder makes no representation about the suitability of this book or the information therein for any purpose. It is provided “as is” without expressed or implied warranty. In particular, the copyright holder declines to be liable in any way should errors result from the use of the examples and the information given here in practical work.

1 Grundlegendes

1.1 Der Ereignisraum

Der Begriff Ereignisraum definiert sich aus den möglichen Ergebnissen (idealisierter) Experimente bzw. Beobachtungen. Einfachste Beispiele sind der Ereignisraum für das Werfen einer Münze, welcher aus zwei Elementen, K(opf) und Z(ahl), $\{K, Z\}$, besteht, oder der Ereignisraum für das Rollen eines Würfels, welcher sich aus den sechs möglichen Augenziffern $\{1, 2, \dots, 6\}$ zusammensetzt.

Bei der Beschreibung von Ereignissen muß man zwischen nicht zusammengesetzten, sogenannten *Elementarereignissen* und *zusammengesetzten* Ereignissen unterscheiden. Das Ereignis “Augenziffer 6” ist ein Elementarereignis, das Ereignis “gerade Augenziffer” hingegen kann durch die drei Elementarereignisse “2”, “4”, “6” realisiert werden und ist daher ein zusammengesetztes Ereignis. Als weiteres Beispiel betrachten wir das Werfen zweier Würfel. Ein Elementarereignis stellt z.B. das Ergebnis “(1, 5), erster Würfel zeigt 1, zweiter 5” dar. Hingegen handelt es sich bei “Augensumme von 6” um ein zusammengesetztes Ereignis, denn dieses Ergebnis kann durch die fünf einfachen Ereignisse (1, 5), (2, 4), (3, 3), (4, 2), (5, 1) realisiert werden kann. Als zweites Beispiel möge das Alter einer Person dienen: Jeder mögliche Wert x (z.B. 50 Jahre) ist ein einfaches Ereignis. Die Aussage, jemand sei zwischen fünfzig und sechzig Jahre alt, stellt hingegen das zusammengesetzte Ereignis $50 \leq x \leq 60$ (Jahre) dar.

Wir können jetzt die Definition des Begriffs “Ereignisraum” präzisieren: Die Gesamtheit aller Elementarereignisse (Punkte) (des interessierenden Zufallsexperiments) wird als *Ereignisraum* \mathcal{S} bezeichnet. Jedes denkbare Ergebnis eines (idealisierten) Experiments (oder einer Beobachtung) wird vollständig durch ein (und genau ein) Element von \mathcal{S} beschrieben.

Unsere einfachsten Beispiele (Werfen einer Münze, Rollen eines Würfels) sind Beispiele von endlichen, diskreten Ereignisräumen. In der Praxis kommt man oft nicht mit diskreten und/oder endlichen Ereignisräumen aus: Das Gewicht eines Hühnereis möge zum Beispiel typischerweise zwischen 70 und 75 g liegen, es kann aber beliebige Zwischenwerte (72.5 g, 71.315 g, ...) annehmen. Der zugehörige Ereignisraum kann daher nicht diskret sein, und darüberhinaus enthält er unendlich viele Punkte. Dazu kommt, daß es in der Praxis vorteilhaft ist, auch total unsinnig erscheinende Werte nicht auszuschließen (d.h. in unserem Beispiel (formal) als Gewicht eines Hühnereis auch 500 g zuzulassen). Die Wahrscheinlichkeit eines solchen Ereignis ist so gering, daß wir so ein “Ei” in der Realität nicht erwarten, die mathematische Behandlung wird jedoch vereinfacht. *NB*: Ein diskreter Ereignisraum muß nicht endlich sein. Der Ereignisraum für das zweimalige Werfen einer Münze ist KK, KZ, ZK, ZZ, für das dreimalige Werfen KKK, KKZ, KZK, ZKK, KZZ, ZKZ, ZZK, ZZZ usw. Die Extrapolation auf unendlich wiederholtes Werfen einer Münze führt auf einen diskreten, jedoch unendlichen Ereignisraum.

Wir werden uns zunächst vornehmlich (Abschnitte 1–4) auf diskrete Ereignisräume und die Wahrscheinlichkeiten diskreter Ereignisse konzentrieren. In Abschnitt 5 werden Verteilungsfunktionen für diskrete und kontinuierliche Zufallsvariablen eingeführt, damit wird automatisch der Umstieg zu kontinuierlichen Ereignisräumen vorgenommen. Diese sind für uns primär bei der statistischen Behandlung von Meßwerten von Interesse (Abschnitt 6.)

1.2 Definitionen und Regeln

1.2.1 Ereignisse

Gegeben sei ein Ereignisraum \mathcal{S} , dessen Elemente (Elementarereignisse) mit x bezeichnet werden, d.h. $x \in \mathcal{S}$ für alle x . Wir verwenden Großbuchstaben A, B usw. um ganz allgemein ein Ereignis (einfach oder zusammengesetzt) zu bezeichnen. Die Notation $x \in A$ heißt, daß ein Punkt x im Ereignis A inbegriffen ist. (Wenn z.B. A das Ereignis "Würfelsumme 6 beim Rollen zweier Würfel" bezeichnet, dann gilt $(5, 1) \in A$.) Die Gleichheit zweier Ereignisse wird durch

$$A = B$$

verdeutlicht.

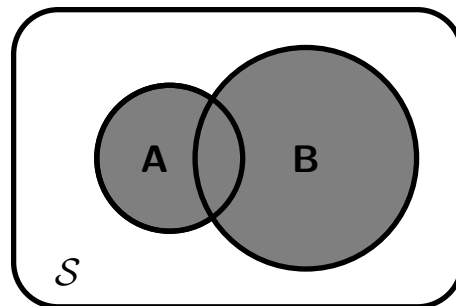
Ereignisse werden durch bestimmte Bedingungen für ihre Elemente definiert. Wir brauchen daher ein Symbol, das den Fall beschreibt, daß kein Element einem bestimmten Satz von Bedingungen genügt, dafür schreiben wir

$$A = 0.$$

Gilt $A = 0$, so enthält A keinen Punkt von \mathcal{S} , wir sagen, daß Ereignis A ist *unmöglich*.

Weiters gibt es zu jedem Ereignis A ein weiteres Ereignis, daß sich durch die Bedingung "A geschieht nicht" definiert. Es enthält alle Elemente von \mathcal{S} , die nicht in A enthalten sind. Wir sprechen vom komplementären Ereignis A' (oft auch als \bar{A} bezeichnet).

$$A \cup B$$



$$AB (= A \cap B)$$

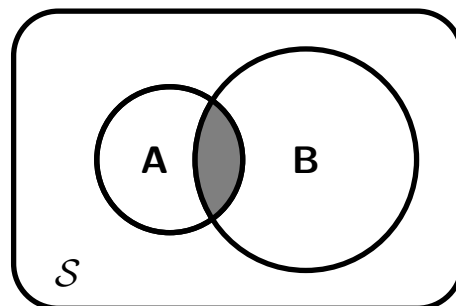


Abbildung 1: Vereinigung und Durchschnitt zweier Ereignisse.

Aus zwei Ereignissen A und B lassen sich neue Ereignisse durch die Forderungen "sowohl A als auch B treffen zu" (*Durchschnitt*) und "entweder A oder B oder beide treffen zu" (*Vereinigung*)

definieren. Die beiden Fälle sind in Abb. 1 illustriert. Wir schreiben für den Durchschnitt AB ¹ und für die Vereinigung $A \cup B$. Sei A das Ereignis “Augenziffer 1” und B das Ereignis “ungerade Augenziffer”. Dann ist $AB = \{1\}$ und $A \cup B = \{1, 3, 5\}$. Schließen sich A und B aus, dann haben sie keine gemeinsamen Elemente, und AB ist unmöglich. In solchen Fällen gilt

$$AB = 0,$$

dies ist in Abb. 2 illustriert. Ein Beispiel ist das gleichzeitige Auftreten der Augenziffer 1 und einer geraden Augenziffer beim Rollen eines Würfels, diese beiden Ereignisse schließen einander aus. Das Ereignis AB' bedeutet, daß sowohl A als auch B' zutreffen, d.h., A aber nicht B treten ein. $A'B'$ bedeutet, daß weder A noch B zutreffen.

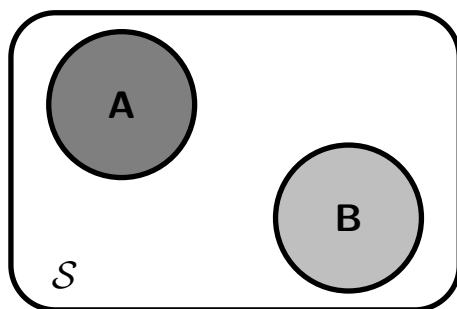


Abbildung 2: Disjunkte Ereignisse

Klarerweise lassen sich alle eben eingeführten Konzepte auf mehr als zwei Ereignisse verallgemeinern ($ABC \dots, A \cup B \cup C \cup \dots$). Insbesondere schließen sich die Ereignisse A, B, C usw. einander aus, wenn kein einziges Ereignispaar ein gemeinsames Element hat, d.h. $AB = 0, AC = 0, \dots, BC = 0$, usw.

$$A \subset B$$

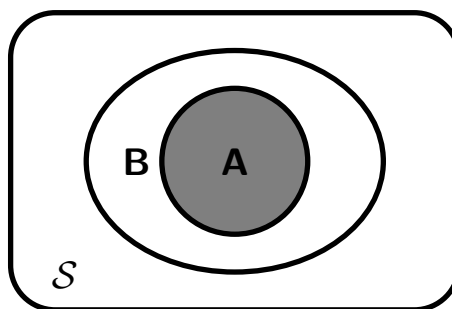


Abbildung 3: Obermenge

Das letzte fehlende Konzept betrifft den Fall, daß A nicht eintreten kann, ohne daß auch B eintritt (wobei die Umkehrung im Allgemeinen nicht gilt! — “Alle Bienen sind Insekten, aber nicht alle Insekten sind Bienen.”). Man sagt auch, B sei die *Obermenge* von A (s. Abb. 3). Wir führen die äquivalenten Symbole $A \subset B$ und $B \supset A$ ein, die bedeuten, daß jedes Element von A in B enthalten ist (A impliziert B , B ist von A impliziert). In diesem Fall führen wir weiters die alternative Schreibweise $B - A$ für BA' ein. Die Notation $B - A$ gestattet uns auch zu schreiben $A' = S - A$ und $A - A = 0$.

¹Die Standardnotation für den Durchschnitt zweier Mengen ist $A \cap B$, AB ist jedoch einfacher zu schreiben.

► **Beispiele:** (a) Wenn sich A und B ausschließen, dann bedingt A das Nichteintreten von B und umgekehrt. Somit bedeuten $AB = 0$, $A \subset B'$ und $B \supset A'$ dasselbe.

(b) Das Ereignis $A - AB$ bedeutet A aber nicht sowohl A als auch B . Somit gilt $A - AB = AB'$

(c) Bridge (jeder der vier Spieler bekommt 13 der insg. 52 Karten): Die Ereignisse A , B , C und D bedeuten, daß der jeweilige Spieler mindestens ein As hat. (Da in Bridge alle Karten ausgeteilt werden, muß mindestens eines der vier Ereignisse eintreten, d.h., mindestens ein Spieler hat mindestens ein As.) Ein wenig Überlegung zeigt, daß (i) $A \cup B \cup C \cup D = \mathcal{S}$, d.h. den gesamten Ereignisraum des Beispiels ergibt. (ii) Der Fall $ABCD \neq 0$ kann nur eintreten, wenn jeder Spieler genau ein As hat. (iii) Ein Spieler (assoziiert mit Ereignis D) habe alle vier Asse bekommen. In diesem Fall können weder A , B , C zutreffen, mit anderen Worten A' , B' und C' treten gleichzeitig ein oder, mit noch anderen Worten, das Ereignis $A'B'C'$ trifft zu.



1.2.2 Wahrscheinlichkeiten

In einfachen Fällen ist der Begriff der Wahrscheinlichkeit unmittelbar einsichtig. So bereitet es intuitiv keine Schwierigkeiten dem Ereignis Kopf oder Zahl beim Werfen einer (fairen) Münze jeweils die Wahrscheinlichkeit $\mathbf{P}(K) = \mathbf{P}(Z) = 1/2$ zuzuordnen, oder die Wahrscheinlichkeit beim Würfeln (mit einem fairen Würfel) eine Sechs zu erhalten mit $\mathbf{P}(6) = 1/6$ zu präzisieren. Bei beiden Beispielen handelt es sich um Elementarereignisse des jeweiligen Ereignisraums, und in beiden Fällen tritt der Sonderfall ein, daß jedes der Elementarereignisse gleichwahrscheinlich ist. Betrachten wir unter diesen Voraussetzungen ein zusammengesetztes Ereignis A , so ist $\mathbf{P}(A)$ die Summe der Wahrscheinlichkeiten der in A enthaltenen Elementarereignisse. Sei A das Ereignis "gerade Würfelzahl", so ist A durch die Elementarereignisse "2", "4" und "6" realisiert. Somit ist $\mathbf{P}(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$. Allgemein gilt für die Wahrscheinlichkeit eines zusammengesetzten Ereignis A , das sich aus r Elementarereignissen zusammensetzt, und einem Ereignisraum \mathcal{S} , der aus n Elementarereignissen besteht

$$\mathbf{P}(A) = \frac{\text{Für } A \text{ günstige Fälle}}{\text{Insgesamt mögliche Fälle}} = \frac{r}{n} \quad (1)$$

Gleichung 1 ist die s.g. Laplace'sche Definition der Wahrscheinlichkeit. Sie ist nur für endliche, diskrete Ereignisräume anwendbar und gilt *nur wenn die Elementarereignisse gleichwahrscheinlich sind*.²

Nehmen wir jetzt an, es interessiert Sie, ob eine Münze oder ein Würfel wirklich fair sind. Eine Möglichkeit dies herauszufinden besteht darin, das Zufallsexperiment (Werfen der Münze, Würfeln) n -mal zu wiederholen. Ist die Münze fair, dann erwarten wir das K(opf) bzw. Z(ahl) ungefähr $n/2$ -mal auftreten werden (bzw. für den Würfel erwarten wir, daß jede Augenzahl gleich oft, d.h. in einem Sechstel der Fälle auftreten wird. Wir führen die relative Häufigkeit des Ereignis A $h(A) = n_A/n$ ein, wobei $n_A \leq n$ die Anzahl des Auftretens von A in n Experimenten (also die absolute Häufigkeit) bezeichnet. Ist die Münze fair, so erwarten wir daß $h(K) = h(Z) = 1/2$ um so besser erfüllt ist, je

²Die Wichtigkeit der zweiten Bedingung sieht man an folgendem Beispiel: Der Ereignisraum für das Experiment "Augensumme zweier Würfel" ist $\{2, 3, 4 \dots 12\}$ Ein Versuch auf diesem Ereignisraum mit Definition (1) zu operieren würde jedoch käglich scheitern, da die Wahrscheinlichkeiten der Elementarereignisse nicht gleich sind. Die Wahrscheinlichkeit die "Augensumme 2" zu bekommen ist z.B. $\frac{1}{36}$, hingegen kommt die "Augensumme 7" mit der Wahrscheinlichkeit $\frac{1}{6}$ vor.

$$\mathbf{P}(A \cup B) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 A_2)$$

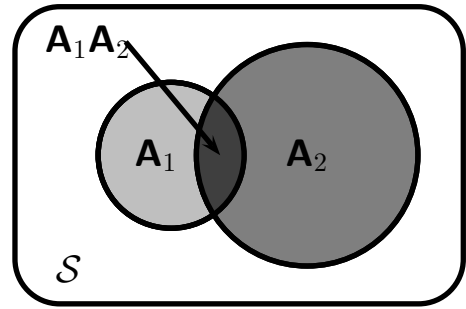


Abbildung 4: Zum Additionssatz

größer die Anzahl der Experimente ist. Mathematisch suggeriert das zum Grenzwert $\lim_{n \rightarrow \infty} h(A)$ überzugehen. Dieser führt zur *statistische Definition* der Wahrscheinlichkeit, die besagt, daß dieser Grenzwert (wenn er existiert) gleich der Wahrscheinlichkeit des Ereignis A ist, d.h.,

$$\lim_{n \rightarrow \infty} h(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} = \mathbf{P}(A). \quad (2)$$

Im Gegensatz zu Gl. 1 gilt Gl. 2 auch für ungleich Wahrscheinlichkeiten der Elementarereignisse. Für eine gezinkte Münze könnte man z.B. $\lim_{n \rightarrow \infty} h(Z) = \mathbf{P}(Z) = 0.55$ anstatt $\mathbf{P}(Z) = 1/2$ für die faire Münze finden.

Die moderne Wahrscheinlichkeitstheorie hat gezeigt, daß es vorteilhaft ist, Wahrscheinlichkeiten durch axiomatische Definitionen einzuführen.³ *In jedem diskreten Ereignisraum \mathcal{S} mit Elementarereignissen $E_1, E_2 \dots$ ist mit jedem dieser Ereignisse eine Zahl assoziiert, die wir als Wahrscheinlichkeit von E_i bezeichnen und durch $\mathbf{P}(E_i)$ symbolisieren. Die Wahrscheinlichkeit ist nicht-negativ und gehorcht der Bedingung*

$$\mathbf{P}(E_1) + \mathbf{P}(E_2) + \dots = 1. \quad (3)$$

Weiters gilt: *Die Wahrscheinlichkeit $\mathbf{P}(A)$ eines beliebigen Ereignisses A ist die Summe aller in A enthaltenen Elementarereignisse. Da Gl. 3 aber auch besagt, daß $\mathbf{P}(\mathcal{S}) = 1$, folgt daraus für jedes Ereignis A*

$$0 \leq \mathbf{P}(A) \leq 1 \quad (4)$$

Wir betrachten jetzt zwei Ereignisse A_1 und A_2 . Um die Wahrscheinlichkeit $\mathbf{P}(A_1 \cup A_2)$ zu erhalten, müssen wir die Wahrscheinlichkeiten aller Elementarereignisse von A_1 und A_2 aufaddieren, wobei wir zu berücksichtigen haben, daß kein Elementarereignis doppelt gezählt wird (Punkte die sowohl in A_1 als auch in A_2 liegen!), s. Abb. 4. Es gilt daher

$$\mathbf{P}(A_1 \cup A_2) \leq \mathbf{P}(A_1) + \mathbf{P}(A_2) \quad (5)$$

Jedes Ereignis, daß sowohl zu A_1 als auch A_2 gehört, wird auf der rechten Seite doppelt gezählt, während es auf der linken Seite nur einmal vorkommt. Folglich gilt

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 A_2) \quad (6)$$

Schließen sich A_1 und A_2 aus, d.h. $A_1 A_2 = 0$, dann vereinfacht sich Gl. 6 zu

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) \quad (7)$$

³Derartige Definitionen gehen meist auf den russischen Mathematiker A. N. Kolmogoroff zurück.

► **Beispiel:** Eine Münze wird zweimal geworfen. Unser Ereignisraum besteht also aus KK, KZ, ZK, ZZ, wir gehen von einer fairen Münze aus, somit hat jedes Ereignis die Wahrscheinlichkeit $\frac{1}{4}$. Die Ereignisse A_1 und A_2 seien als “K(opf) im ersten bzw. zweiten Wurf” definiert. Somit enthält A_1 KK und KZ, A_2 KK und ZK. Die Vereinigung $A = A_1 \cup A_2$ besteht aus KK, ZK, KZ. Es gilt nach Gleichung 6

$$\mathbf{P}(A_1 \cup A_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$



Gleichung 5 hat eine Verallgemeinerung für beliebige Ereignisse $A_1, A_2, A_3 \dots$. Es gilt

$$\mathbf{P}(A_1 \cup A_2 \cup \dots) \leq \mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3) + \dots \quad (8)$$

Das Gleichheitszeichen gilt nur dann wenn sich die Ereignisse $A_1, A_2, A_3 \dots$ ausschließen.

2 Einschub: Elemente der Kombinatorik

Es bietet sich an dieser Stelle an, ganz kurz ein paar Elemente der Kombinatorik zu wiederholen. Im Rahmen der Wahrscheinlichkeitstheorie ermöglicht Kombinatorik die Anzahl der Elementarereignisse N eines endlichen, diskreten Ereignisraums zu bestimmen. Kombinatorische Fragestellungen tauchen aber auch im Alltag auf — denken Sie z.B. an die Frage nach der Anzahl der Möglichkeiten im Lotto 6 aus 45. Gilt weiters der Spezialfall, daß jedes Elementarereignis gleichwahrscheinlich ist, dann kann man natürlich aus der Kenntnis der Größe des Ereignisraums unmittelbar Wahrscheinlichkeiten berechnen (aus Gl. 3 folgt in diesem Fall sofort $\mathbf{P}(E_i) = 1/N$ und es gilt Gl. 1).

2.1 Vorbemerkung(en)

Sie haben m Elemente a_1, a_2, \dots, a_m und n Elemente b_1, b_2, \dots, b_n . Daraus lassen sich $m \times n$ Paare (a_i, b_j) bilden, die jeweils ein Element jeder Gruppe enthalten. Um sich davon zu überzeugen, ordnen Sie die Paare in einem rechteckigen Raster in Form einer Multiplikationstabelle mit m Reihen und n Spalten an, sodaß (a_i, b_j) der Schnittpunkt der i -ten Reihe mit der j -Spalte ist. In einer derartigen Tabelle scheint jedes Paar einmal auf, und die Richtigkeit der Behauptung ist damit offensichtlich.

Dies lässt sich wie folgt verallgemeinern: Für n_1 Elemente a_1, a_2, \dots, a_{n_1} , und n_2 Elemente b_1, b_2, \dots, b_{n_2} , und n_3 Elemente c_1, c_2, \dots, c_{n_3} usw. gibt es $n_1 \times n_2 \times n_3 \dots$ Anordnungen der Form (a_i, b_j, c_k, \dots) . Der Fall eines Paares wurde bereits behandelt. Nun ist aber jedes Triplet (a_i, b_j, c_k) auch ein Paar der Form $((a_i, b_j), c_k)$, und somit ist der Tripletfall auf den bewiesenen Paarfall reduziert. Durch Induktion folgt der allgemeine Beweis der Behauptung. Viele Anwendungen beruhen auf der folgenden alternativen Formulierung diese Theorems: *r -maliges hintereinanderfolgendes Auswählen (r hintereinanderfolgende Entscheidungen) mit genau n_k Möglichkeiten im k -ten Schritt können $n_1 \cdot n_2 \cdot \dots \cdot n_r$ unterschiedliche Resultate haben.*

► **Beispiele:** (i) Personen werden nach Geschlecht, Zivilstatus (verheiratet/ledig) und Beruf klassifiziert. Unterscheidet man z.B. zwischen 17 Berufen, dann gibt es $2 \times 2 \times 17 = 68$ unterschiedliche Klassen (Kategorien).

(ii) Das Legen von i Bällen in n Schachteln,⁴ wobei Mehrfachbelegungen möglich sind, läuft auf das Auswählen einer Schachtel (n Möglichkeiten) für jeden Ball (die Auswahl wird i -mal wiederholt) hinaus. Somit können die i Bälle auf n^i verschiedene Weisen in die n Schachteln plaziert werden. ◀

2.2 Permutationen

Gegeben sind n verschiedene Elemente (z.B. unterschiedlich gefärbte Bälle). Auf wieviele Arten kann man diese Elemente in einer Reihe anordnen?

Für den ersten Ball hat man n Plätze zur Verfügung. Für den zweiten Ball bleiben dementsprechend $n - 1$, für den dritten Ball $n - 2$ Plätze usw. Für den n -ten Ball bleibt immer genau ein Platz übrig. Somit ist die Zahl der *Permutationen* P_n von n Elementen durch

$$P_n = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1 = n! \quad (9)$$

gegeben.

Die bisherige Ableitung hat vorausgesetzt, daß die Elemente unterschiedlich (bzw. unterscheidbar) sind (z.B. daß jeder Ball eine andere Farbe hat). Nehmen wir jetzt an, daß wir wie zuvor n Elemente reihen sollen, n_1 der n Elemente aber gleich (ununterscheidbar) seien. (5 Bälle von denen 3 die gleiche Farbe haben sollen auf 5 Plätze angeordnet werden). Wir nehmen an, wir kennen die gesuchte Zahl P_{n,n_1} und fragen uns, um welchen Faktor sich diese erhöhen würde, wenn die n_1 gleichen Elemente unterscheidbar wären. Dieser gesuchte Faktor ist aber nach Gl. 9 $n_1!$. Somit gilt

$$P_n = n_1! \cdot P_{n,n_1}$$

womit wir sofort für P_{n,n_1}

$$P_{n,n_1} = \frac{n!}{n_1!} \quad (10)$$

finden.

Wenn nun von den n Elementen jeweils n_1, n_2, n_3 usw. gleich sind, so ergibt sich sofort als Verallgemeinerung von Gl. 10

$$P_{n,n_1,n_2,n_3,\dots} = \frac{n!}{n_1!n_2!n_3!\dots} \quad (11)$$

► **Beispiel:** Was ist die Anzahl N aller fünfziffrigen Zahlen, die aus den Ziffern 4 und 7 bestehen. Die möglichen Zahlen haben die Form 44444 ($5 \times 4, 0 \times 7$), 74444 ($4 \times 4, 1 \times 7$) und alle Permutationen, 77444 ($3 \times 4, 2 \times 7$) und alle Permutationen usw. Das Symbol n_4 benenne wie oft die Ziffer 4, n_7 wie oft die Ziffer 7 in der Zahl vorkommt. Es muß gelten $n_4 + n_7 = 5$. Für jedes Paar (n_4, n_7) gibt es

$$P_{5,n_4,n_7} = \frac{5!}{n_4!n_7!}$$

Möglichkeiten. Da dieser Ausdruck bezüglich Vertauschung von n_4 und n_7 spiegelbildlich ist, können wir uns ein wenig Rechenaufwand ersparen und N in folgender Weise berechnen:

$$N = 2 \times (P_{5,5,0} + P_{5,4,1} + P_{5,3,2}) = 2 \times \left(\frac{5!}{5!0!} + \frac{5!}{4!1!} + \frac{5!}{3!2!} \right) = 32$$

⁴Fast jedes kombinatorische Problem lässt sich mit Hilfe von Bällen und ggf. Schachteln illustrieren

(NB: $0! = 1$). ◀

2.3 Variationen

Gegeben sind n voneinander verschiedene Elemente. Wieviele Möglichkeiten gibt es, aus diesen Elementen i Elemente herauszugreifen und in verschiedener Weise anzuordnen (d.h. die Reihenfolge der Anordnung ist relevant)? Die Frage kann auch lauten, wieviele Stichproben der Größe i aus n Elementen gezogen werden können.

Zur Beantwortung derartiger Fragen muß man unterscheiden, ob die gezogenen Elemente/Bälle zurückgelegt werden dürfen oder nicht, bzw. ob ein Element in der Stichprobe mehrfach vorkommen darf oder nicht. Man spricht von *Variationen von n Elementen zur i -ten Klasse mit $(\bar{V}_{n,i})$ und ohne $(V_{n,i})$ Wiederholung*. Im Fall mit Wiederholung gibt es für jede Stichprobe (jedes Ziehen) n Möglichkeiten, da wir i -mal ziehen, ergibt sich

$$\bar{V}_{n,i} = n^i \quad (12)$$

Das klassische Beispiel zur Illustration dieses Falles ist die Anzahl der möglichen Kodierungen für die 20 Aminosäuren durch den Dreibuchstabencode aus den vier DNA (RNA). Dies entspricht dem Fall $n = 4$ und $i = 3$, somit gibt es die wohlbekannten $4^3 = 64$ Möglichkeiten.

Bei Fragestellungen dieser Art stellt sich manchmal die Schwierigkeit was in Gl. 12 n und was i ist. n ist die Anzahl der Möglichkeiten pro Auswahlschritt, i ist die Anzahl der Wiederholungen des Auswahltritts. Im Falle des genetischen Codes sind 3 Positionen zu besetzen (es wird dreimal gewählt, $i = 3$), für jede Auswahl gibt es $n = 4$ Möglichkeiten. Man vergleiche hierzu das schon erwähnte Problem des Plazierens von z.B. 4 (unterscheidbaren) Bällen in 3 Schachteln, wobei Mehrfachbelegungen (= Wiederholungen) möglich sind: Hier wird $i = 4$ mal gewählt, pro Auswahl gibt es $n = 3$ Möglichkeiten. Daher gibt es $3^4 = 81$ Möglichkeiten die Bälle zu plazieren.

Wenn wir keine Wiederholung zulassen, so gibt es beim ersten Mal ziehen n Möglichkeiten, beim zweiten Mal $n - 1$ Möglichkeiten, und schließlich $n - i + 1$ Möglichkeiten beim letzten (= i -ten) Ziehen. Somit ergibt sich für die Anzahl von Variationen von n Elementen zur i -ten Klasse ohne Wiederholung

$$V_{n,i} = n \cdot (n - 1) \cdot \dots \cdot (n - i + 1) = \frac{n!}{(n - i)!} = (n)_i \quad (13)$$

► **Beispiel:** Die Anzahl dreistelliger Zahlen, die sich aus den Ziffern eins bis neun schreiben läßt (ohne daß eine Ziffer mehr als einmal vorkommt) ist

$$V_{9,3} = \frac{9!}{(9 - 3)!} = \frac{9!}{6!} = 9 \cdot 8 \cdot 7 = 504. \quad \blacktriangleleft$$

Wir wissen also, daß aus einer Population mit n Elementen n^i bzw. $n!/(n - i)!$ Stichproben der Größe i gezogen werden können, je nachdem ob die Stichproben mit oder ohne Wiederholung gezogen werden. In Wahrscheinlichkeitstheorie und Statistik hört man oft den Ausdruck "Zufallsstichproben".

Diesen Ausdruck können wir jetzt präzisieren: Im beschriebenen Auswahlprozeß handelt es sich genau dann um Zufallsstichproben, wenn die Wahrscheinlichkeit jeder Stichprobe E genau gleich ist, d.h. die Wahrscheinlichkeit $\mathbf{P}(E) = 1/n^i$ bzw. $\mathbf{P}(E) = (n-i)!/n!$ (mit bzw. ohne Wiederholung).

2.4 Kombinationen

Wir stellen uns dasselbe Problem wie im vorigen Unterabschnitt, nur daß es diesmal nicht auf die Reihenfolge der Elemente in der Stichprobe ankommt. Gegeben sind n verschiedene Elemente. Auf wieviele Arten lassen sich aus ihnen i Elemente herausgreifen, wenn es auf die Reihenfolge der herausgegriffenen Elemente nicht ankommt, und jedes Element nur einmal vorkommen darf. Dies entspricht der Variation von n Elementen zur i -ten Klasse ohne Wiederholung, jedoch ignorieren wir die $i!$ Möglichkeiten die Stichprobe anzuordnen. Man findet daher für die Anzahl der *Kombinationen von n Elementen zur i -ten Klasse*

$$C_{n,i} = \frac{V_{n,i}}{i!} = \frac{(n)_i}{i!} = \frac{n!}{(n-i)!i!} = \binom{n}{i} \quad (14)$$

► **Beispiel:** Die Gemeindevertretung eines Ortes setzt sich aus 18 Männern und 5 Frauen zusammen. Es soll eine dreigliedrige Abordnung an die Bezirkshauptmannschaft delegiert werden. Berechne die Anzahl der Möglichkeiten, eine solche Abordnung zu wählen, die a) nur aus Männern, b) aus zwei Männern und einer Frau besteht:

$$\text{a) } \binom{18}{3} = 816 \quad \text{b) } \binom{18}{2} \cdot \binom{5}{1} = 765.$$

(Im Fall b) wird das in Abschnitt 2.1 besprochene Grundtheorem der Kombinatorik verwendet, d.h. aus der Möglichkeit unter Männern und Frauen zu wählen, wird das Produkt gebildet.) ◀

Der eben besprochene Fall schließt Wiederholungen aus. Ähnlich wie bei Variationen gibt es auch bei Kombinationen den Fall der *Kombination von n Elementen zur i -ten Klasse mit Wiederholung*. Dies ist der einzige Fall, bei dem das Ergebnis nicht unmittelbar einsichtig ist, und man betrachtet am besten das folgende äquivalente Problem: Wie können i ununterscheidbare Objekte (z.B. Bälle gleicher Farbe) auf n Zellen (Schachteln) verteilt werden? Wir stellen die i Bälle durch Sternchen und die n Schachteln durch n Abstände zwischen $n+1$ Strichen dar. Z.B. stellt $|***|*|||****|$ eine mögliche Verteilung von $i=8$ Bällen auf $n=6$ Schachteln dar. Diese Darstellung muß immer einen Strich am Anfang und am Ende enthalten, aber die verbleibenden $n-1$ Striche (Zelltrennungen) können in beliebiger Ordnung auftreten. Das heißt aber nichts anderes als daß die Anzahl unterscheidbarer Anordnungen gleich der Anzahl von Möglichkeiten i Positionen aus $n+i-1$ entspricht. Diese Anzahl ist aber durch Gl. 14 gegeben, somit finden wir

$$\bar{C}_{n,i} = \binom{n-1+i}{i} = \binom{n-1+i}{n-1} \quad (15)$$

Eine interessante Nebenfrage ist noch, wieviele unterscheidbare Anordnungen es gibt, in der keine Zelle leer bleibt. In unserer symbolischen Schreibweise bedeutet dies, daß keine zwei Striche nebeneinander stehen dürfen. Die 8 Bälle (Sterne) lassen $i-1$ Stellen, die von den $n-1$ frei verschiebbaren Strichen eingenommen werden müssen. Dies ist genau auf $\binom{i-1}{n-1}$ Möglichkeiten realisierbar.

► **Beispiele:** a) Wie viele verschiedene Augenzahlen kann man beim Würfeln mit drei Würfeln erhalten? Jeder der drei Würfel kann eine Augenzahl zwischen eins und sechs aufweisen (diese 6 Möglichkeiten sind die "Zellen" oder "Schachteln"). Mögliche Ergebnisse sind z.B. (1,5,2) oder (2,2,6). (Wir unterscheiden nicht zwischen (1,5,2) und (1,2,5)!) Somit läßt sich die Frage umformulieren, auf wieviele Arten man 3 Werte aus insgesamt 6 Augenzahlen herausgreifen kann, wobei Wiederholungen zugelassen sind. Einsetzen in Gl. 15 mit $i = 3$ und $n = 6$ gibt

$$\bar{C}_{6,3} = \binom{6+3-1}{3} = \binom{8}{3} = 56.$$

b) 100 Personen wurden zufällig ausgewählt, und in Raucher (r) und Nichtraucher (n) unterteilt. Weiters wird zwischen Männern (M) und Frauen (F) unterschieden. Unsere Stichprobe von 100 Personen ist also durch das Quadrupel (F_r, F_n, M_r, M_n) charakterisiert. Wieviele solche Quadrupel gibt es? (NB: Diese Anzahl ist der Ereignisraum dieses Problems) Dieses Problem ist äquivalent zur Aufgabe $i = 100$ Bälle auf $n = 4$ Zellen aufzuteilen. Einsetzen in Gl. 15 ergibt

$$\bar{C}_{100,4} = \binom{100+4-1}{100} = \binom{103}{100} = 176851.$$



2.5 Binomialkoeffizienten

Es wurden jetzt bereits einige Male Binomialkoeffizienten verwendet, die hoffentlich aus der Mittelschule bekannt sind. Der Vollständigkeit halber einige Beziehungen und Definitionen (siehe auch Ihre Formelsammlungen). Die Definition des Binomialkoeffizienten lautet:

$$\binom{n}{i} = \frac{(n)_i}{i!} = \frac{n \cdot (n-1) \dots (n-i+1)}{i!} = \frac{n!}{(n-i)! i!}$$

Aus der Definition sieht man unmittelbar, daß

$$\binom{n}{i} = \binom{n}{n-i}.$$

Weiters führt man folgende Definitionen ein

$$\binom{n}{0} = 1 \quad \binom{n}{i} = 0 \text{ wenn } i < 0 \quad \binom{n}{i} = 0 \text{ wenn } i > n$$

Es gilt folgender Summensatz

$$\binom{n}{i-1} + \binom{n}{i} = \binom{n+1}{i}$$

Der Binomialkoeffizient ist eng mit dem binomischen Lehrsatz verknüpft

$$(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i},$$

welcher für $a = b = 1$ auf den interessanten Zusammenhang

$$(1+1)^n = \sum_{i=0}^n \binom{n}{i} = 2^n$$

führt.

Der Rest dieses Abschnitt ist optional und als Vertiefung für Interessierte gedacht und daher nicht Prüfungsstoff

Wir nützen die Gelegenheit zu einem Exkurs im Exkurs und führen eine verallgemeinerte Form des Binomialkoeffizienten ein:

$$\binom{a}{k} = \frac{a(a-1)\dots(a-k+1)}{k!} \quad (16)$$

mit $a \in \mathbb{R}$ und k eine ganze Zahl.⁵ Sie können sich unschwer davon überzeugen, daß sich Gl. 16 für $a = n \in \mathbb{N}$ auf den üblichen Binomialkoeffizienten $\frac{n!}{(n-k)!k!}$ reduziert.

Dieser verallgemeinerte Binomialkoeffizient gestattet es, die Taylorreihe der Funktion $(1+x)^a$ (vgl. Abschnitt über Taylorreihen!) in folgender Form

$$(1+x)^a = \underbrace{\binom{a}{0}}_1 x^0 + \binom{a}{1} x^1 + \binom{a}{2} x^2 + \binom{a}{3} x^3 + \dots \quad (17)$$

zu schreiben.

Weiters weisen wir auf folgende kombinatorische Interpretation der schon erwähnten Beziehung

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n$$

hin: Die linke Seite repräsentiert die Anzahl der Möglichkeiten, auf die eine Population von n Elementen in zwei Subpopulationen aufgeteilt werden kann, wenn die Anzahl der ersten Gruppe jede Größe $k = 0, 1, \dots, n$ annehmen kann. Umgekehrt kann man diese Aufteilung auch erreichen, in dem man für jedes Element der Population entscheidet, ob es in die eine oder die andere Untergruppe fällt — dies ergibt 2^n Möglichkeiten (rechte Seite).

2.6 Die Sterling'sche Näherung für $n!$

Dieser Abschnitt ist optional und als Vertiefung für Interessierte gedacht und daher nicht Prüfungsstoff

Es ist Ihnen vermutlich schon aufgefallen, daß die Fakultät $n!$ mit steigendem n sehr rasch anwächst, und z.B. rasch die Kapazität des Taschenrechners übersteigt. Dazu kommt, daß rein prinzipiell die Berechnung von $n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$ eine aufwendige Rechenoperation ist ($n-2$ Multiplikationen). Es ist daher von Interesse, Näherungsformeln für die Berechnung von $n!$ (bzw. $\ln n!$) zur Verfügung zu haben.

Wir starten mit der Definition der Fakultät

$$n! = 1 \times 2 \times 3 \times \dots \times (n-1) \times n$$

und ziehen den Logarithmus

$$\ln n! = \ln 1 + \ln 2 + \ln 3 + \dots + \ln(n-1) + \ln n = \sum_{m=1}^n \ln m. \quad (18)$$

⁵Es gelten weiterhin alle Definitionen, wie für den normalen Binomialkoeffizienten, s. oben

Wir nähern nun die Summe durch das Integral

$$\ln n! = \sum_{m=1}^n \ln m \approx \int_1^n dx \ln x = [x \ln x - x]_1^n = n \ln n - n + 1 \approx n \ln n - n. \tag{19}$$

Die Näherung der Summe durch ein Integral wird um so besser, je größer n , da der Logarithmus für große n eine sehr langsam ansteigende, sehr langsam variierende Funktion ist. Die letzte Approximation in Gl. 19 besteht darin, den von der unteren Grenze stammenden Faktor 1, der im Vergleich zu großen n unbedeutend ist, zu vernachlässigen. (*Testen Sie die Güte der Näherung für einige n . Wichtig ist der relative Fehler $(n \ln n - n) / \ln n!$*)

Der relative Fehler von Gl. 19 sinkt für $n = 100$ schon unter 1%. Eine noch bessere Näherung für $n!$ (bzw. $\ln n!$) erhält man durch die Gleichung

$$\ln n! \approx n \ln n - n + \frac{1}{2} \ln(2\pi n), \tag{20}$$

in diesem Fall ist der relative Fehler für $n = 10(!)$ bereits kleiner als 1 Promille!

Wir zeigen die Ableitung von 20, da sie fast alle Ihre bisher erlernten mathematischen Kenntnisse fordert. Startpunkt ist das (bestimmte) Integral

$$\int_0^\infty dx x^n e^{-x}.$$

Wir suchen in einer Integraltafel nach dem Integral und finden (leicht nachrechenbar durch partielle Integration!!)

$$\int dx x^n e^{-x} = -x^n e^{-x} + n \int dx x^{n-1} e^{-x} \tag{A}$$

Als nächstes setzen wir in (A) die Grenzen 0 und ∞ ein. Im Term $[-x^n e^{-x}]_0^\infty$ gibt die untere Grenze 0 sicher Null. Die obere Grenze muß als Grenzwert berechnet werden, d.h. $\lim_{x \rightarrow \infty} -x^n e^{-x}$. Da die Exponentialfunktion für positive Argumente stärker steigt als jede Potenz (von x), und für negative Argumente schneller gegen Null geht als jede Potenz (von x),⁶ ist dieser Grenzwert aber für jedes n immer Null. Somit erhält man aus (A) nach Einsetzen der Grenzen folgende Rekursionsformel für das bestimmte Integral

$$\int_0^\infty dx x^n e^{-x} = n \int_0^\infty dx x^{n-1} e^{-x} \tag{B}.$$

Schritt (B) muß $n - 1$ mal wiederholt werden. Schließlich erhält man

$$\int_0^\infty dx x^n e^{-x} = n(n-1) \dots 2 \underbrace{\int_0^\infty dx e^{-x}}_1 = n!, \tag{21}$$

weil $\int_0^\infty dx e^{-x} = -[e^{-x}]_0^\infty = -[0 - 1] = 1$, wobei Einsetzen der oberen Grenze (∞) eigentlich wieder als Grenzwert durch Grenzwertbildung berechnet werden müßte.

Gl. 21 hilft uns zunächst nicht weiter. Wir müssen uns mit den Eigenschaften des Integranden $f(x) = x^n e^{-x}$ beschäftigen. Für positive x (und große n , daran sind wir ja interessiert) ist x^n eine rasch anwachsende Funktion von x und e^{-x} eine (noch) rasch(er) abfallende Funktion von x . Die Funktion hat daher vermutlich ein Maximum (wo sich x^n und e^{-x} die Waage halten), bevor e^{-x} "gewinnt". Weiters wird das Maximum umso schärfer sein, je größer der Exponent n . Als nächstes suchen wir die Position des Maximums x_0 . Anstatt $f(x)$ direkt zu diskutieren, betrachten wir $\ln f(x) = \ln(x^n e^{-x})$ und finden

$$(\ln f(x))' = \frac{d}{dx}(n \ln x - x) = \frac{x}{n} - 1 = 0$$

und somit

$$x_0 = n. \tag{22}$$

Wie schon gesagt ist $f(x)$ nur in der unmittelbaren Nähe dieses Maximums merklich von 0 verschieden, d.h., dieser Bereich trägt das Meiste zum gesuchten Integral bei. Wir gehen jetzt wie folgt vor: Zuerst entwickeln wir $\ln f(x)$ in eine Taylorreihe um $x_0 = n$ und behalten nur den ersten nichttrivialen Term. Danach wenden wir die Exponentialfunktion auf unsere Näherung für $\ln f(x)$ an, und erhalten eine Näherung für $f(x)$. Diese Näherung wird schließlich integriert, und damit haben wir nach Gl. 21 die gesuchte Näherung für $n!$. Wir schreiben

$$x = n + \xi, \quad \text{mit } \xi \ll n$$

und berechnen wie angekündigt die Taylorreihe von

$$\ln f(x) = n \ln x - x = n \ln(n + \xi) - (n + \xi) \tag{C}$$

⁶Um sich das überlegen, genügt es die MacLaurin'sche Reihe der Exponentialfunktion mit einer beliebigen Potenzfunktion zu vergleichen!

um den Entwicklungspunkt $x_0 = n$. Genauer gesagt brauchen wir uns nur um den Term $\ln(n + \xi)$ kümmern,

$$\ln(n + \xi) = \ln n + \underbrace{\ln\left(1 + \frac{\xi}{n}\right)}_{\text{Formelsammlung!!}} = \ln n + \frac{\xi}{n} - \frac{1}{2} \frac{\xi^2}{n^2} + \dots \quad (\text{D})$$

Einsetzen von (D) in (C) führt auf

$$\ln f(x) = n \ln n - n - \frac{1}{2} \frac{\xi^2}{n}. \quad (\text{E})$$

(Der lineare Term in ξ hebt sich weg, weil wir um das Maximum der Funktion entwickelt haben.) Aus (E) finden wir nun sofort

$$f(x) = n^n e^{-n} e^{-1/2(\xi^2/n)}.$$

Damit können wir integrieren (vgl. Gl. 21), wobei wir bei den Grenzen berücksichtigen müssen, daß wir eine Variablentransformation von x auf ξ vorgenommen haben!

$$n! \approx \int_{-n}^{\infty} d\xi n^n e^{-n} e^{-1/2(\xi^2/n)} = n^n e^{-n} \int_{-\infty}^{\infty} d\xi e^{-1/2(\xi^2/n)} = n^n e^{-n} \sqrt{2\pi n}.$$

Zur Berechnung des Integrals mußten wir ein wenig "Zaubern." Zunächst einmal wird die untere Grenze $-n$ durch $-\infty$ ersetzt. Sie können sich jedoch überlegen, daß der Integrand für $\xi < -n$ bereits vernachlässigbar klein ist, somit die Ausweitung der Grenze keinen Einfluß hat. Das verbleibende Integral $\int_{-\infty}^{\infty} d\xi \exp(-\xi^2/2n) = \sqrt{2\pi n}$ können wir mit unseren Mitteln nicht beweisen.⁷

Es gilt also für große n

$$n! \approx \sqrt{2\pi n} n^n e^{-n} \quad (23)$$

Gleichungen 23 bzw. 20 (und manchmal auch Gl. 19) sind unter dem Namen *Stirling'sche Näherung* der Fakultät bekannt.

3 Bedingte Wahrscheinlichkeit — Statistische Unabhängigkeit

3.1 Bedingte Wahrscheinlichkeit

Nach diesem Einschub kehren wir zur Berechnung von Wahrscheinlichkeiten (in diskreten Ereignisräumen) zurück. Mit Gl. 6 bzw. dem Spezialfall Gl. 7 können bereits viele wichtige Fragestellungen untersucht werden. In der Praxis stellt sich jedoch auch häufig folgendes Problem: Man betrachtet zwei Ereignisse A und H . H ist eingetroffen. Wie groß ist die Wahrscheinlichkeit dafür, daß mit H gleichzeitig auch A eintritt. Man nennt diese Wahrscheinlichkeit die *bedingte Wahrscheinlichkeit* von A (bezüglich der Hypothese H) und bezeichnet sie mit $\mathbf{P}(A|H)$.

Wir betrachten eine Population von N Personen, die N_H Frauen und N_A Raucher umfasse. Die Ereignisse A und H bedeuten "es handelt sich um einen Raucher (eine Raucherin)" und "es handelt sich um eine Frau". Sowohl aus unserem intuitiven Verständnis von Wahrscheinlichkeiten, als auch aus der am Ende von Abschnitt 2.3 gegebenen Definition von Zufallstichproben folgt, daß die Wahrscheinlichkeit, daß eine zufällig ausgewählte Person ein(e) Raucher(in) ist, $\mathbf{P}(A) = N_A/N$ beträgt, ebenso ist die Wahrscheinlichkeit, daß eine zufällig ausgewählte Person weiblich ist, $\mathbf{P}(H) = N_H/N$. Beschränken wir uns auf die Subpopulation aller Frauen, so beträgt die Wahrscheinlichkeit eine Raucherin zu wählen N_{AH}/N_H . Diese Wahrscheinlichkeit ist aber genau die bedingte Wahrscheinlichkeit,

$$\mathbf{P}(A|H) = \frac{\mathbf{P}(AH)}{\mathbf{P}(H)}, \quad (24)$$

⁷Es gilt die Identität $\int_{-\infty}^{\infty} dx \exp(-x^2) = \sqrt{\pi}$, aus der man sich obiges Integral durch die Substitution $\xi/\sqrt{2n} = x$ ableiten kann; die Beziehung ist wichtig genug, daß Sie sie in Ihre Formelsammlungen hinzufügen sollten.

daß eine Person raucht, unter der Voraussetzung, daß die gezogene Person weiblich ist.

Gl. 24 ist die Definition der bedingten Wahrscheinlichkeit. Durch triviale Umformung gilt weiters die Beziehung

$$\mathbf{P}(AH) = \mathbf{P}(H) \cdot \mathbf{P}(A|H) = \mathbf{P}(HA) = \mathbf{P}(A) \cdot \mathbf{P}(H|A), \quad (25)$$

Die weiteren Identitäten in Gl. 25 folgen sofort aus der Kommutativität des Durchschnitts zweier Mengen ($\mathbf{P}(AH) = \mathbf{P}(HA)$) und der Möglichkeit, die Rolle von A und H zu vertauschen ($\mathbf{P}(H|A)$ ist die bedingte Wahrscheinlichkeit, daß H eintritt unter der Voraussetzung, daß A eingetreten ist.) (NB: $\mathbf{P}(A|H) \neq \mathbf{P}(H|A)$!!)

Bevor wir zu einem Beispiel kommen, verallgemeinern wir noch Gl. 25. Nehmen wir an, wir haben drei Ereignisse A, B, C . Wir betrachten zunächst $BC = H$ als die Hypothese und wenden Gl. 25 zweimal an. Damit erhalten wir sofort

$$\mathbf{P}(ABC) = \mathbf{P}(A|BC) \cdot \mathbf{P}(BC) = \mathbf{P}(A|BC) \cdot \mathbf{P}(B|C) \cdot \mathbf{P}(C), \quad (26)$$

und die weitere Verallgemeinerung zu vier und mehr Ereignissen sollte offensichtlich sein.

► **Beispiel:** (a) Aus einem normalen Kartenspiel (52 Karten) wird eine Karte gezogen. Man weiß, daß die gezogene Karte rot ist. Wie groß ist die Wahrscheinlichkeit, daß die gezogene Karte ein König ist? In unserer Notation entspricht “rot” der Hypothese H , und “König” dem Ereignis A . Es gibt 26 rote Karten, darunter 2 Könige. Somit ist $\mathbf{P}(AH) = 2/52$ und $\mathbf{P}(H) = 26/52$ und gemäß Gl. 24 $\mathbf{P}(A|H) = \frac{2/52}{26/52} = 1/13$.

(b) Für einen Betriebsrat soll eine Person nachgewählt werden. Es kandidieren 5 Frauen und 8 Männer. 3 der Frauen und 3 der Männer sind Angestellte, die restlichen Kandidat(inn)en sind Arbeiterinnen bzw. Arbeiter. Uns interessieren nicht nur die Wahrscheinlichkeiten der Wahlausgänge M : “der gewählte Kandidat ist männlich” und A : “der/die gewählte Kandidat(in) ist ein(e) Arbeiter(in)”, sondern auch die bedingte Wahrscheinlichkeit $\mathbf{P}(A|M)$, d.h., unter der Voraussetzung, daß der gewählte Kandidat männlich ist, wie hoch ist die Wahrscheinlichkeit, daß es ein Arbeiter ist. Wir haben $\mathbf{P}(M) = 8/13$ und $\mathbf{P}(AM) = 3/13$. Somit finden wir gemäß Gl. 24 $\mathbf{P}(A|M) = \frac{3/13}{8/13} = \frac{3}{8}$. ◀

3.2 Statistische (stochastische) Unabhangigkeit

Ausgehend von Gl. 24 sind zwei Grenzfalle von Interesse. Tritt A immer nur zusammen mit H auf, so ist $\mathbf{P}(AH) = \mathbf{P}(A)$ und wir erhalten

$$\mathbf{P}(A|H) = \frac{\mathbf{P}(A)}{\mathbf{P}(H)} \quad (27)$$

Dieser Spezialfall tritt z.B. in folgendem Beispiel ein. Es wurde eine gerade Augenzahl gewurfelt. Was ist die Wahrscheinlichkeit, daß es eine “2” ist? Im Prinzip ist das mit $A = “2”$ und $H = “gerade Augenzahl”$ die bedingte Wahrscheinlichkeit $\mathbf{P}(A|H)$, die sich aber, weil das Ereignis “2” immer eine gerade Augenzahl (die Hypothese) voraussetzt, direkt aus den Einzelwahrscheinlichkeiten $\mathbf{P}(A|H) = \mathbf{P}(A)/\mathbf{P}(H) = \frac{1/6}{1/2} = \frac{1}{3}$ berechnen laßt.

Der andere Spezialfall tritt dann ein, wenn $\mathbf{P}(A)$ und $\mathbf{P}(H)$ voneinander unabhangig sind. Das bedeutet $\mathbf{P}(A|H) = \mathbf{P}(A)$. (Sie werfen zwei Wurfel hintereinander. Der erste zeigt “2”. Die Wahrscheinlichkeit, mit dem zweiten eine “5” zu wurfeln, ist davon vollig unberuhrt, und daher $\frac{1}{6}$.)

Eingesetzt in Gl. 25 bedeutet dies aber

$$\mathbf{P}(AH) = \mathbf{P}(HA) = \mathbf{P}(A) \cdot \mathbf{P}(H) \quad (28)$$

Wenn Gl. 27 zutrifft, so heißen die Ereignisse A und H *statistisch (stochastisch) unabhängig*. Die Wahrscheinlichkeit aus einem Kartenspiel mit 52 Karten einen Herz König zu ziehen, ist einerseits $\frac{1}{52}$ — es gibt 52 Karten, davon einen Herz König, und jede Karte ist gleichwahrscheinlich. Man kann sich die Wahrscheinlichkeit aber auch als das Produkt der Wahrscheinlichkeiten, gleichzeitig eine Herz ($\frac{13}{52} = \frac{1}{4}$) und einen König ($\frac{4}{52} = \frac{1}{13}$) zu ziehen, denken, d.h. $\frac{1}{13} \cdot \frac{1}{4} = \frac{1}{52}$.

Sehr oft ist die statistische Unabhängigkeit zweier Ereignisse intuitiv klar. Als ein Beispiel, wo dies nicht offensichtlich ist, betrachten wir Familien mit drei Kindern und interessieren uns für das Geschlecht der Kinder (b = Bub, m = Mädchen), wobei es auf das Alter der Kinder ankommt. Die Notation bbg bedeutet: “erstes (ältestes) Kind ein Bub, zweites Kind ein Bub, drittes (jüngstes) Kind ein Mädchen”. Wir nehmen an, daß jede der acht Möglichkeiten (bbb, bbm, bmb, mbb, bmm, mbm, mmb, mmm) gleich wahrscheinlich ist.⁸ Die Hypothese H sei “Die Familie hat Kinder beiderlei Geschlechts,” ($\mathbf{P}(H) = 6/8$) das Ereignis A sei “. . . aber maximal ein Mädchen” ($\mathbf{P}(A) = 4/8$). Das gleichzeitige Eintreffen AH bedeutet eine der Möglichkeiten bbm, bmb, mbb. Somit ist $\mathbf{P}(AH) = 3/8 = \mathbf{P}(A) \cdot \mathbf{P}(H) = 3/4 \cdot 1/2$. In Familien mit drei Kindern sind die beiden Ereignisse A und H also unabhängig. Sie können sich aber leicht überzeugen, daß dies in Familien mit zwei oder vier Kindern *nicht* der Fall ist!

So wie sich Gl. 25 zu Gl. 26 verallgemeinern läßt, gilt im Falle statistischer Unabhängigkeit für mehrere Ereignisse $A, B, C \dots$

$$\mathbf{P}(ABC\dots) = \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(C) \dots \quad (29)$$

Weiters muß jedes Paar AB, AC, BC usw. statistisch unabhängig sein (d.h. Gl. 28 muß für jedes Paar gelten). (Umgekehrt reicht die paarweise statistische Unabhängigkeit dreier Ereignisse allein nicht aus, um Gl. 29 zu garantieren!)

3.3 Der Satz von Bayes

Wir kehren jetzt zum Rechnen mit bedingten Wahrscheinlichkeiten zurück. Es seien n unvereinbare Ereignisse A_1, A_2, \dots, A_n gegeben, von denen eines immer eintreffen muß, und ein weiteres Ereignis B , das stets mit genau einem jener Ereignisse auftritt. Es gilt also

$$B = \sum_{i=1}^n BA_i$$

bzw. unter Berücksichtigung von Gl. 8

$$\mathbf{P}(B) = \sum_{i=1}^n \mathbf{P}(BA_i).$$

Mit Hilfe von Gl. 25 gilt aber weiters

$$\mathbf{P}(B) = \sum_{i=1}^n \mathbf{P}(A_i) \mathbf{P}(B|A_i) \quad (30)$$

⁸Das ist eine Vereinfachung, weil die Geburtenrate von Buben und Mädchen nicht völlig gleich ist!

Gl. 30 wird auch als der *Satz über die totale Wahrscheinlichkeit* bezeichnet und ist nützlich, weil es oft einfacher ist, die auftretenden bedingten Wahrscheinlichkeiten zu berechnen, als $\mathbf{P}(B)$ direkt zu ermitteln.

- **Beispiel:** (a) Gegeben sind 3 Urnen, und zwar:
 2 Urnen mit je 4 schwarzen und 1 gelben Kugel,
 1 Urne mit 3 schwarzen und 5 gelben Kugeln.

Aus einer dieser Urnen, wobei nicht festgestellt wurde, aus welcher, wird eine Kugel herausgezogen. Wie groß ist die Wahrscheinlichkeit dafür, daß es eine gelbe ist? Wir bezeichnen das Ereignis, daß man eine der beiden oben zuerst angeführten Urnen getroffen hat, mit A_1 , dasjenige, daß man die oben als dritte Urne angeführte getroffen hat mit A_2 . B bezeichnet "Ziehen einer gelben Kugel". BA_1 ist dann das Ereignis eine Kugel aus einer der ersten beiden Urnen zu ziehen, BA_2 das Ereignis eine Kugel aus der dritten Urne zu ziehen. Da die gelbe Kugel aus einer der drei Urnen stammen muß, gilt

$$BA_1 + BA_2 = B$$

woraus sich Gl. 30 zufolge ergibt

$$\mathbf{P}(B) = \mathbf{P}(A_1)\mathbf{P}(B|A_1) + \mathbf{P}(A_2)\mathbf{P}(B|A_2).$$

Nun ist $\mathbf{P}(A_1) = 2/3$, das es drei Urnen gibt, von denen zwei zum Ereignis führen. Entsprechend ist $\mathbf{P}(A_2) = 1/3$. Weiters ist $\mathbf{P}(B|A_1) = 1/5$, da im ersten Typ von Urne 5 Kugeln liegen, von denen nur eine gelb ist. $\mathbf{P}(B|A_2) = 5/8$. Wir erhalten somit

$$\mathbf{P}(B) = \frac{2}{3} \cdot \frac{1}{5} + \frac{1}{3} \cdot \frac{5}{8} = \frac{41}{120}.$$

(b) Sie spielen in einem nicht ganz ehrlichen Kasino.⁹ 99% aller Würfel sind fair (F), aber 1% sind so gezinkt (U), daß in 50% aller Würfe eine Sechs (6) kommt. Sie nehmen einen Würfel aus einem großen Gefäß voller Würfel. Was sind $\mathbf{P}(6|U)$, $\mathbf{P}(6|F)$? Was sind $\mathbf{P}(6U)$ ($=\mathbf{P}(6 \cap U)$) und $\mathbf{P}(6F)$? Was ist die Wahrscheinlichkeit mit dem zufällig herausgegriffenen Würfel eine Sechs zu würfeln? Diese Aufgabe ist völlig analog zu (a): Die Wahrscheinlichkeit mit einem fairen Würfel eine Sechs zu würfeln ist $\mathbf{P}(6|F) = 1/6$, mit einem gezinkten Würfel beträgt die Wahrscheinlichkeit (lt. Angabe) $\mathbf{P}(6|U) = 1/2$. Weiters gilt $\mathbf{P}(6U) = \mathbf{P}(U)\mathbf{P}(6|U) = 1/100 \cdot 1/2 = 1/200$ und analog $\mathbf{P}(6F) = \mathbf{P}(F)\mathbf{P}(6|F) = 99/100 \cdot 1/6 = 33/200$. Die Wahrscheinlichkeit mit dem zufällig gezogenen Würfel eine Sechs zu würfeln ist die totale Wahrscheinlichkeit $\mathbf{P}(6) = \mathbf{P}(U)\mathbf{P}(6|U) + \mathbf{P}(F)\mathbf{P}(6|F) = 1/200 + 33/200 = 17/100$. ◀

Unter den oben genannten Voraussetzungen (n unvereinbare Ereignisse A_1, A_2, \dots, A_n und ein weiteres Ereignis B , das stets mit genau einem dieser Ereignisse zusammen auftreten muß), lassen sich zwei weitere interessante Beziehungen ableiten. Aus der Definition der bedingten Wahrscheinlichkeit (Gl. 24) folgt

$$\mathbf{P}(A_i B) = \mathbf{P}(B)\mathbf{P}(A_i|B) = \mathbf{P}(A_i)\mathbf{P}(B|A_i),$$

was nichts anderes als Gl. 25 ist. In dieser Form geschrieben läßt sich aber weiter auf

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B|A_i)}{\mathbf{P}(B)} \quad (31)$$

⁹NB: Ich empfehle nicht einmal Spielen in einem ehrlichen Kasino!

umformen, woraus man durch Einsetzen von Gl. 30 schließlich

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B|A_i)}{\sum_{i=1}^n \mathbf{P}(A_i)\mathbf{P}(B|A_i)} \quad (32)$$

erhält. Man bezeichnet Gl. 31 und Gl. 32 als die *Formeln von Bayes* oder auch als Formeln über die Wahrscheinlichkeit von Hypothesen (Ursachen). Die Ursachen in diesem Zusammenhang sind die Ereignisse A_j , deren Wahrscheinlichkeiten $\mathbf{P}(A_j)$ als bekannt vorausgesetzt werden (*a priori* Wahrscheinlichkeiten). Gl. 32 ermöglicht nun die Berechnung der bedingten Wahrscheinlichkeiten $\mathbf{P}(A_j|B)$, daß die Hypothesen (Ursachen) A_j zutreffen, wenn (“weil”) das Ereignis B eingetreten ist. Diese Fragestellung vertauscht die Bedeutung von Hypothese (Ursache) und Ereignis (“Resultat”, “Wirkung”). Die Standardfragestellung in diesem Zusammenhang wäre ja $\mathbf{P}(B|A_j)$, d.h. die Wahrscheinlichkeit, daß ein Ereignis B eintritt, vorausgesetzt A_j trifft zu (ist eingetreten). Man spricht in diesem Zusammenhang deshalb auch von den *a posteriori* Wahrscheinlichkeiten $\mathbf{P}(A_j|B)$ der Hypothesen (Ursachen).

► **Beispiel:** (a) Gegeben sind die gleichen Urnen wie im Beispiel zur totalen Wahrscheinlichkeit. Aus einer dieser Urnen (die nichtunterscheidbar sind) wird eine Kugel gezogen. Die Kugel ist gelb. Wie groß ist die Wahrscheinlichkeit, daß die Kugel aus einer der beiden ersten Urnen stammt. Wir verwenden die gleichen Bezeichnungen wie im vorherigen Beispiel. Damit gilt nach dem Satz von Bayes

$$\mathbf{P}(A_1|B) = \frac{\mathbf{P}(A_1)\mathbf{P}(B|A_1)}{\mathbf{P}(A_1)\mathbf{P}(B|A_1) + \mathbf{P}(A_2)\mathbf{P}(B|A_2)} = \frac{\frac{2}{3} \cdot \frac{1}{5}}{\frac{2}{3} \cdot \frac{1}{5} + \frac{1}{3} \cdot \frac{5}{8}} = \frac{16}{41}.$$

(b) Das folgende Beispiel ist die Fortsetzung des “unehrlichen Kasino” Beispiels: Sie haben nach wie vor den in der “letzten Folge” gezogenen Würfel und würfeln drei Sechsen (3×6) in Folge. Was ist die Wahrscheinlichkeit $\mathbf{P}(U|3 \times 6)$, daß der von ihnen gezogene Würfel gezinkt ist. Der Satz von Bayes auf dieses Beispiel angewandt besagt

$$\mathbf{P}(F|3 \times 6) = \frac{\mathbf{P}(3 \times 6|F)\mathbf{P}(F)}{\mathbf{P}(3 \times 6)},$$

wobei $\mathbf{P}(3 \times 6) = \mathbf{P}(3 \times 6|F)\mathbf{P}(F) + \mathbf{P}(3 \times 6|U)\mathbf{P}(U)$ die totale Wahrscheinlichkeit 3 Sechsen zu würfeln ist. Weil es jetzt um 3 Sechsen in Folge geht, können wir nicht direkt die Ergebnisse von (a) verwenden. Jedes Werfen eines Würfel ist ein statistisch unabhängiger Vorgang zu allen vorhergehenden Würfeln, d.h., $\mathbf{P}(3 \times 6) = \mathbf{P}(6)^3$. Daher ist die Wahrscheinlichkeit $\mathbf{P}(3 \times 6|U) = 0.5^3 = 0.125$, und wir erhalten somit

$$\mathbf{P}(F|3 \times 6) = \frac{0.5^3 \cdot 0.01}{\left(\frac{1}{6}\right)^3 \cdot 0.99 + 0.5^3 \cdot 0.01} = 0.21$$

Somit ist es trotz dreimaligem Würfeln einer Sechs noch immer wahrscheinlicher, daß wir einen fairen Würfel gezogen haben. ◀

► **Ein biologisches Beispiel:** Bedingte Wahrscheinlichkeiten und der Satz von Bayes spielen eine wichtige Rolle in vielen statistischen Anwendungen, u.a. auch in der Bioinformatik. Nehmen wir an, es interessiert Sie die Frage, ob extrazelluläre Protein möglicherweise eine leicht unterschiedliche Aminosäurezusammensetzung haben als intrazelluläre Proteine. Sie gehen (zum Beispiel) davon aus, daß Cystein in extrazellulären Proteinen häufiger vorkommt als im intrazellulären Proteinen. Versuchen wir diese Information dazu zu benutzen, um zu entscheiden ob eine neue Proteinsequenz $x = x_1, x_2, \dots, x_n$ (eher) auf ein

extra- oder intrazelluläres Protein hindeutet. Wir nehmen die Trainingssets der SWISS-PROT Datenbank¹⁰ und klassifizieren sie in extra- und intrazelluläre Proteine.

Daraus können wir eine Satz von Häufigkeiten q_a^{int} für intra- und einen entsprechenden Satz q_a^{ext} für extrazelluläre Proteine ableiten. Um den Satz von Bayes anwenden zu können, brauchen wir weiters einen Schätzwert für die Wahrscheinlichkeit, daß eine neubestimmte Sequenz extra- (\mathbf{P}^{ext}) bzw. intrazellulär (\mathbf{P}^{int}) ist. Wir nehmen weiters an, daß jede Sequenz entweder extra- oder intrazellulär ist, d.h. $\mathbf{P}^{int} = 1 - \mathbf{P}^{ext}$. (Dies ist die Hauptschwäche dieses Ansatzes: Die Möglichkeit von Transmembranproteinen kann nicht berücksichtigt werden!) Die beiden Wahrscheinlichkeiten \mathbf{P}^{int} und \mathbf{P}^{ext} sind *a priori* Wahrscheinlichkeiten, denn Sie stellen unsere Vermutung bezüglich der Rolle der Sequenz dar *bevor* wir die Sequenz selbst gesehen haben.

Weiters haben wir die bedingten Wahrscheinlichkeiten $\mathbf{P}(x|ext) = \prod_i q_{x_i}^{ext}$ und $\mathbf{P}(x|int) = \prod_i q_{x_i}^{int}$. Da wir annehmen, daß die Sequenz entweder extra- oder intrazellulär sein muß, ist die totale Wahrscheinlichkeit $\mathbf{P}(x) = \mathbf{P}^{ext}\mathbf{P}(x|ext) + \mathbf{P}^{int}\mathbf{P}(x|int)$. Aus dem Satz von Bayes folgt jetzt

$$\mathbf{P}(ext|x) = \frac{\mathbf{P}^{ext} \prod_i q_{x_i}^{ext}}{\mathbf{P}^{ext} \prod_i q_{x_i}^{ext} + \mathbf{P}^{int} \prod_i q_{x_i}^{int}}$$

$\mathbf{P}(ext|x)$ ist die gesuchte *a posteriori* Wahrscheinlichkeit, daß die Sequenz extrazellulär ist und ist unsere Vermutung *nachdem* wir die Sequenz gesehen haben. ◀

4 Die Binomialverteilung und davon abgeleitete Grenzfälle

4.1 Binomialverteilung

Wir interessieren uns jetzt für die Berechnung von Wahrscheinlichkeiten von wiederholten Versuchen, die voneinander (statistisch) unabhängig sind. Jeder Einzelversuch einer derartigen Kette habe zwei mögliche Ergebnisse S ("success") und F ("failure"), deren Wahrscheinlichkeiten $\mathbf{P}(S) = p$ und $\mathbf{P}(F) = q$ über die Dauer des Experiments (der gesamten Versuchskette) konstant bleiben. Selbstverständlich gilt $p + q = 1$. Eine derartige Versuchsanordnung nennt man *Bernoulli-Schema*.

Wir fragen jetzt nach der Wahrscheinlichkeit $\mathbf{P}(\mathbf{S}_n = m) = b(m; n, p)$, daß S bei n Versuchen m -mal eintritt (und dementsprechen F ($n - m$)-mal). Wir nehmen zunächst an, daß eine bestimmte Reihenfolge verlangt wird. Wir betrachten die Ereigniskette $SSFFSF \dots S$. Da die Einzelereignisse laut Voraussetzung voneinander unabhängig sind, gilt für die Wahrscheinlichkeit dieser Kette $p \times p \times q \times q \times p \times q \times \dots \times p$ (Gl. 28). Insbesondere gilt für den Fall, daß S in den ersten m , und F in den weiteren $(n - m)$ Versuchen auftritt

$$p^m q^{n-m} = p^m (1 - p)^{n-m}.$$

Dieses Produkt ist wegen der Unabhängigkeit der Einzelversuche aber auch die Wahrscheinlichkeit jeder Versuchskette, in der S m -mal und F ($n - m$)-mal auftritt.

Unsere eigentliche Frage war nach der Wahrscheinlichkeit $b(m; n, p)$, daß S bei n Versuchen m -mal eintritt, die Reihenfolge des Auftretens der S und F ist uns dabei egal. Wir überlegen uns daher, auf

¹⁰<http://www.ebi.ac.uk/swissprot/>

wieviel verschiedene Arten und Weisen man m Ereignisse S und $(n - m)$ Ereignisse F anordnen kann. Diese Zahl bekommt man entweder durch die Anzahl der Permutation $P_{n,m,n-m} = n!/m!(n - m)!$ (Gl. 11), oder durch Stellen der (äquivalenten) Frage, wieviele Stichproben der Größe m man aus n Elementen (den Versuchen) ziehen kann — dies ist auf $C_{n,m} = \binom{n}{m}$ Weisen möglich (vgl. Gl. 14). Die gesuchte Wahrscheinlichkeit ist jetzt die Wahrscheinlichkeit eines Einzelereignisses $p^m q^{n-m}$ multipliziert mit der Anzahl der Möglichkeiten, daß dieses realisiert wird, d.h.

$$\mathbf{P}(S_n = m) = b(m; n, p) = \frac{n!}{m!(n - m)!} p^m q^{n-m} = \frac{n!}{m!(n - m)!} p^m (1 - p)^{n-m} = \binom{n}{m} p^m (1 - p)^{n-m}. \tag{33}$$

Wegen des Auftretens des Binomialkoeffizienten $\binom{n}{m}$ wird Gl. 33 *Binomialverteilung* genannt. Die Bedeutung des Begriffs "Verteilung" wird in Abschnitt 5 näher analysiert.

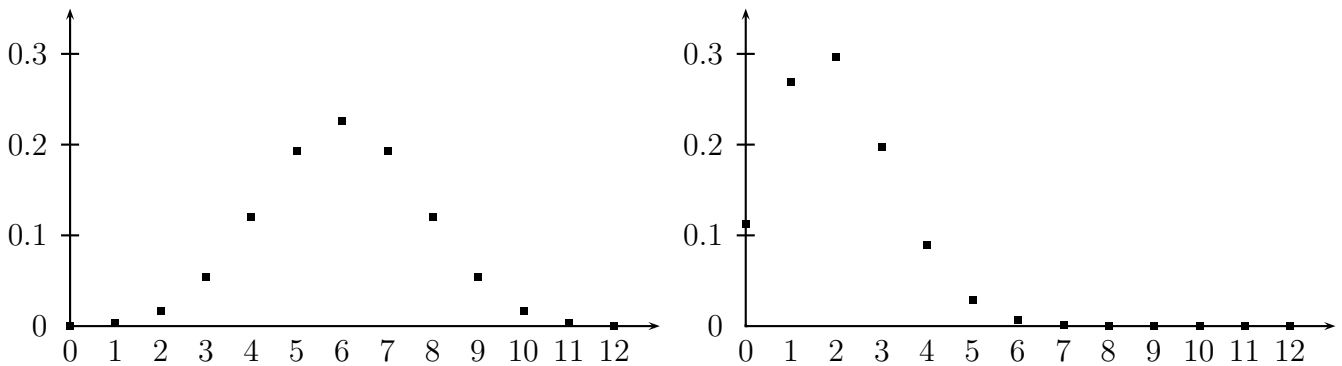


Abbildung 5: Binomialverteilung $b(m; 12, 1/2)$ (links) und $b(m; 12, 1/6)$ (rechts).

Gl. 33 ist in Abb. 5 illustriert. Ist die Wahrscheinlichkeit des Einzelereignisses genau $1/2$, dann ist $b(m; n, p)$ symmetrisch (linker Plot in Abb. 5, entspricht dem Auftreten von m -mal K(opf) beim zwölfmaligem Werfen mit einer fairen Münze), ansonsten ist $b(m; n, p)$ asymmetrisch (rechter Plot in Abb. 5, entspricht der Wahrscheinlichkeit m Sechsen beim zwölfmaligen Würfeln mit einem fairen Würfel zu bekommen).

► **Beispiel:** Eine faire Münze ($\mathbf{P}(K) = p = \mathbf{P}(Z) = 1 - p = q = 1/2$) wird dreimal geworfen. Wie hoch ist die Wahrscheinlichkeit zweimal K(opf) zu werfen? Wir bezeichnen das Ereignis K mit S und Z(ahl) mit F , somit läßt sich die Frage durch Gl. 33 beantworten. Wir haben

$$b(2; 3, \frac{1}{2}) = \binom{3}{2} \underbrace{\left(\frac{1}{2}\right)^2}_{p^2} \underbrace{\left(\frac{1}{2}\right)}_q = \frac{3!}{2! (3 - 2)!} \times \frac{1}{8} = \frac{3}{8}.$$

Anmerkungen: (i) In diesem Beispiel kann man sich die gefragten Wahrscheinlichkeiten natürlich auch direkt überlegen (*machen Sie das!!*), und sich somit die Richtigkeit der allgemeinen Gl. 33 verdeutlichen. Die Nützlichkeit von Gl. 33 wird dann klar, wenn z.B. $n = 20$ und $m = 11$. (ii) Obwohl es durch die Ableitung klar sein sollte, betonen wir zur Sicherheit, daß Gl. 33 keineswegs die Gleichwahrscheinlichkeit der Einzelereignisse voraussetzt. Die gleichen Rechenschritte sind auch auf den Fall einer unfairen Münze anwendbar (z.B. $p = 0.55$, $q = 1 - p = 0.45$).

Gl. 33 stellt weiters die Basis für kombinierte Fragestellungen dar. Wir fragen jetzt wie hoch bei drei Würfeln die Wahrscheinlichkeit $\mathbf{P}(\mathbf{S}_3 \geq 2)$ ist, mindestens zweimal K(opf) zu erhalten. Dieses Ereignis tritt ein, wenn man zweimal K ($b(2; 3, 1/2)$) oder dreimal K ($b(3; 3, 1/2)$) erhält. Da $b(2; 3, 1/2)$ und $b(3; 3, 1/2)$ voneinander unabhängig sind, ist die gesuchte Größe durch $\mathbf{P}(\mathbf{S}_3 \geq 2) = b(2; 3, 1/2) + b(3; 3, 1/2)$ gegeben. Mit $b(3; 3, 1/2) = \frac{3!}{3!(3-3)!}(1/2)^3(1/2)^0 = 1/8$ und dem bereits bekannten $b(2; 3, 1/2) = 3/8$ findet man $\mathbf{P}(\mathbf{S}_3 \geq 2) = 3/8 + 1/8 = 1/2$. ◀

Bernoulli-Schemata, deren Wahrscheinlichkeiten durch Gl. 33 gegeben sind, treten in vielerlei Anwendungen auf, z.B.: (i) Die Wahrscheinlichkeit eines fehlerhaften Produkts ist p . Wie hoch ist die Wahrscheinlichkeit unter n Produkten m fehlerhafte zu finden? (ii) Wirksamkeit eines Medikaments: Ein neuer Wirkstoff versagt nur in 10% der Fälle (korrekte Diagnose und Indikation vorausgesetzt). Was ist die Wahrscheinlichkeit, daß von n Patienten genau m , weniger als m usw. kuriert werden?

Nach Gl. 33 berechnete Wahrscheinlichkeiten sollten aber immer mit der nötigen Sorgfalt interpretiert werden, wie das folgende Beispiel zeigt:

▶ Die Ansteckungsrate einer Rinderkrankheit beträgt 25%. Eine neuentwickelte Impfung soll getestet werden, und n (gesunden) Rindern wird der Impfstoff verabreicht, danach kommen die Tiere in eine Umgebung, in der Infektion möglich ist. Wie sind die Ergebnisse eines derartigen Versuchs zu interpretieren? Nun, für einen nichtwirksamen Impfstoff ist die Wahrscheinlichkeit, daß genau k der n Tiere gesund bleiben durch $b(k; n, 0.75)$ gegeben. Für $k = n = 10$ ist diese Wahrscheinlichkeit 0.056, für $k = n = 12$ 0.032. Bleiben also von 10 (12) Tieren alle gesund, bleibt noch immer ein “Restzweifel” von 5.6 bzw. 3.2%. Klarerweise ist der Fall $k = n = 12$ aussagekräftiger. Vergrößert man n ein wenig, so kann man ein paar interessante “Effekte” sehen: Im Falle des nichtwirksamen Impfstoffs, ist die Wahrscheinlichkeit, daß von 17 Rindern maximal eines erkrankt 0.0501 und die Wahrscheinlichkeit, daß von 23 Rindern maximal 2 erkranken 0.0492. *Das heißt aber, daß der Fall “ein Rind von 17” erkrankt ein stärkeres Indiz für die Wirksamkeit des Impfstoffs ist (kleinerer “Restzweifel”) als der Fall “10 von 10 Rindern bleiben gesund”.* Größen von Stichproben bzw. Populationen sollten daher niemals zu klein gewählt werden. ◀

4.2 Multinomialverteilung

Die Binomialverteilung ist auch auf Versuchsketten anwendbar, bei denen jeder Einzelschritt mehr als zwei mögliche Ergebnisse haben kann, und zwar dann, wenn man primär an einem bestimmten Ereignis interessiert ist (“Erfolg”), während alle anderen Ergebnisse als “Mißerfolg” gewertet werden. Das klassische Beispiel ist Würfeln, wenn z.B. die “Sechs” als Erfolg (Ereignis S), jede andere Augenzahl jedoch als Mißerfolg gilt (Ereignis F). In solchen Fällen kann Gl. 33 unmittelbar verwendet werden (vgl. Abb. 5, rechts).

Es ist aber auch nicht schwer, die Betrachtungen des letzten Abschnitts auf eine Folge von unabhängigen Versuchen zu verallgemeinern, bei der in jedem Versuch eines von k unvereinbaren Versuchsergebnissen A_1, A_2, \dots, A_k eintreten kann. Durch Überlegungen, die der Ableitung von Gl. 33 analog sind kommt man auf die Wahrscheinlichkeit dafür, daß bei einer derartigen Versuchskette von n Einzelversuchen m_1 mal A_1, m_2 mal A_2 , usw. ... bis m_k mal A_k auftritt — diese ist gegeben durch

$$\frac{n!}{m_1! m_2! \dots m_k!} p_1^{m_1} p_2^{m_2} \dots p_k^{m_k}, \quad (34)$$

wobei gelten muß, daß $m_1 + m_2 + \dots + m_k = n$ und $p_1 + p_2 + \dots + p_k = 1$. (Der Vorfaktor in Gl. 34

folgt aus Gl. 11.)

► **Beispiel** Was ist die Wahrscheinlichkeit, bei 5 Würfeln mit einem Würfel zweimal eine Sechs, und dreimal eine Zwei zu erhalten? Es gibt sechs Ereignisse A_1 bis A_6 , nämlich die sechs Augenzahlen. Es ist $m_1 = m_3 = m_4 = m_5 = 0$, $m_2 = 3$ und $m_6 = 6$. Ferner ist $p_1 = p_2 = \dots = p_6 = 1/6$. Wir erhalten somit

$$\frac{5!}{0!3!0!0!0!2!} \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^3 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^2 = 10 \cdot \left(\frac{1}{6}\right)^5 = 0.00129$$



4.3 Poissonverteilung

In vielen Anwendungen liegen Bernoulli-Schemata vor, bei denen n (die Anzahl der Gesamtversuche) vergleichsweise groß, und p (die Erfolgswahrscheinlichkeit eines Einzelexperiments) vergleichsweise klein ist. Das Produkt

$$\lambda = np \tag{35}$$

sei eine endliche Zahl von eher bescheidener Größe. In solchen Fällen kann $b(m; n, p)$ in einer von Poisson zum ersten Mal abgeleiteten Form genähert werden (daß so eine Näherung wünschenswert ist, leuchtet jedem ein, der versucht hat Gl. 33 für, sagen wir, $n = 100$ auszuwerten).

► **Ableitung der Poissonverteilung** Für den Fall $m = 0$ erhält man aus Gl. 33 und 35

$$b(0; n, p) = (1 - p)^n = \left(1 - \frac{\lambda}{n}\right)^n$$

Ziehen des Logarithmus und Anwendung der Taylorreihe $\ln(1 - x) = -[x + x^2/2 + x^3/3 \dots]$ ergibt

$$\ln b(0; n, p) = n \ln \left(1 - \frac{\lambda}{n}\right) = -\lambda - \underbrace{\frac{\lambda^2}{2n} - \dots}_{\rightarrow 0 \text{ wenn } n \text{ groß}} \tag{A}$$

Erinnern Sie sich an unsere Voraussetzungen: großes n , jedoch moderates λ , und somit kann man die Taylorreihe wie in (A) angedeutet nach dem linearen Term abbrechen. Somit folgt aber aus (A)

$$b(0; n, p) \approx e^{-\lambda}. \tag{B}$$

Aus Gl. 33 kann man weiters folgende Abschätzung für den Quotienten

$$\frac{b(m; n, p)}{b(m - 1; n, p)} = \frac{(n - 1 + m)p}{mq} = \frac{\lambda - (1 - m)p}{m(1 - p)} = \frac{\lambda - (1 - m)\frac{\lambda}{n}}{m(1 - \frac{\lambda}{n})}$$

machen. Für große n wird λ/n verschwindend klein und man kann schreiben

$$\frac{b(m; n, p)}{b(m - 1; n, p)} \approx \frac{\lambda}{m}. \tag{C}$$

Aus Gl. (B) und (C) erhält man aber jetzt sofort

$$\begin{aligned}
 b(1; n, p) &\approx \lambda \cdot b(0; n, p) \approx \lambda e^{-\lambda} \\
 b(2; n, p) &\approx \frac{1}{2} \lambda \cdot b(1; n, p) \approx \frac{1}{2} \lambda^2 e^{-\lambda} \\
 b(3; n, p) &\approx \frac{1}{3} \lambda \cdot b(2; n, p) \approx \frac{1}{3 \cdot 2 \cdot 1} \lambda^3 e^{-\lambda} \\
 &\dots \\
 b(m; n, p) &\approx \frac{1}{m} \lambda \cdot b(m-1; n, p) \approx \frac{1}{m!} \lambda^m e^{-\lambda}
 \end{aligned}$$

Die letzte Zeile ist die gesuchte Näherung. ◀

Wegen der großen Wichtigkeit der sogenannten *Poissonverteilung* führen wir die Notation

$$p(m; \lambda) = e^{-\lambda} \frac{\lambda^m}{m!} \tag{36}$$

ein, die, genügend großes n vorausgesetzt, eine gute Näherung für $b(m; n, \lambda/n)$ ist.

► **Beispiel:** Das folgende Beispiel illustriert die gute Übereinstimmung von Binomial- und Poissonverteilung. Was ist die Wahrscheinlichkeit daß in einer Firma mit 500 Angestellten genau k ihren Geburtstag am 1. Jänner haben. Unter der Annahme, daß diese 500 Leute zufällig ausgewählt sind, handelt es sich um ein Bernoulli-Schema mit Erfolgswahrscheinlichkeit $p = 1/365$ und $n = 500$. Aus Gl. 35 berechnet sich $\lambda = 500/365 = 1.3699$. Die korrekten Wahrscheinlichkeiten (Gl. 33) und deren Poissonnäherungen (Gl. 36) sind

k	0	1	2	3	4	5	6
Binomial	0.2537	0.3484	0.2388	0.1089	0.0372	0.0101	0.0023
Poisson	0.2541	0.3481	0.2385	0.1089	0.0373	0.0102	0.0023

Die Übereinstimmung ist aber bereits für viel kleinere n als in eben gebrachten Beispiel recht akzeptabel. Überprüfen Sie dies selbst für $n = 6, p = 1/6$ und $m = 1, 2, \dots$ — dies entspricht (z.B.) der Wahrscheinlichkeit in sechs Würfeln mit einem fairen Würfel 1, 2, usw. Sechsen zu erhalten. ◀

Wir haben bis jetzt die Poissonverteilung als Näherung der Binomialverteilung betrachtet. Es gibt aber durchaus Prozesse, die unmittelbar der Poissonverteilung gehorchen, dazu gehören z.B. radioaktiver Zerfall, Chromosomenaustausch in Zellen in Folge von Röntgenbestrahlung, die Anzahl falsch verbundener Telefonanrufe(!), die räumliche Verteilung von Bakterien in einer Petrischale oder von roten Blutkörperchen in einer Zählkammer usw.

Zwischen der Binomial- und der Poissonverteilung sind folgende Unterschiede anzumerken: In der Poissonverteilung kommt die Anzahl der Versuche n nicht mehr vor, somit kann die Anzahl der Erfolge m beliebig große Werte annehmen, in der Binomialverteilung gilt immer $m \leq n$. Aus diesem Grund kann man von der Poissonverteilung auch nicht mehr auf relative Häufigkeiten m/n zurückschließen. Wir werden im Abschnitt 5 sehen, daß $\lambda = np$ (Gl. 35) dem wahrscheinlichsten Fall/Ergebnis eines Bernoulli-Schemas, das durch die Binomialverteilung beschrieben ist, entspricht (ganz genau handelt es sich um den Erwartungswert der Zufallsvariable $\mathbf{S}_n = k$). (Illustriert ist das z.B. durch Abb. 5. Im linken Fall ist $n = 12, p = 1/2$, der häufigste Ausgang des Experiments ist S

in $12 \cdot 1/2 = 6$ Fällen. Im rechten Fall ($n = 12, p = 1/6$) ist $np = 2$, und tatsächlich ist der Fall “2 mal S ” am wahrscheinlichsten.) Man spricht daher bei Problemen, die durch die Poissonverteilung (zumindestens näherungsweise) beschrieben werden, oft davon, daß S im Mittel λ -mal vorkommt. Die Bedeutung von $\lambda = np$ muß immer klar von der Wahrscheinlichkeit des Einzelereignis p eines Bernoulli-Schemas unterschieden werden.

4.4 Gaussverteilung

Wir haben die Poissonverteilung als Näherung der Binomialverteilung für große n und kleine p abgeleitet. Eine weitere wichtige Näherung erhält man für große n unter der Bedingung, daß p ungefähr gleich $1/2$ ist (wobei im Grenzwert $n \rightarrow \infty$ beliebige p möglich sind, numerisch ist für kleine (große) p allerdings dann die Poissonverteilung vorzuziehen). Es gilt dann¹¹

$$b(m; n, p) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(m-np)^2}{2np(1-p)}} \quad (37)$$

Gl. 37 ist der sogenannte Grenzwertsatz von Moivre-Laplace. Gl. 37 ist eine Gaußsche Glockenkurve, deren Maximum bei $m = np$. Mit den Abkürzungen

$$\sigma^2 = np(1-p) \quad (38)$$

und

$$\mu = np \quad (39)$$

läßt sich dies auch kompakter als

$$b(m; n, p) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(m-\mu)^2}{2\sigma^2}} \quad (40)$$

schreiben.

► **Beispiel:** Was ist die Wahrscheinlichkeit mit einer fairen Münze in 100 Würfeln 40 Mal K(opf) zu bekommen. Da in diesem Fall $p = 1/2$ gilt, haben wir einen idealen Testfall für Gl. 37 und finden

$$b(m; n, p) \approx \frac{1}{\sqrt{2\pi \cdot 100 \cdot 0.5 \cdot (1-0.5)}} e^{-\frac{(40-100 \cdot 0.5)^2}{2 \cdot 100 \cdot 0.5 \cdot (1-0.5)}} = 0.010798$$

Zum Vergleich: Der exakte Wert nach Gl. 33 beträgt 0.010844. ◀

Näherung (37) bzw. 40 ist aber nicht nur für $p \approx 1/2$ verwendbar. Abbildung 6 zeigt den Vergleich zwischen einer Binomialverteilung $b(k; 10, 1/5)$ (die Stufenfunktion) und der Näherung durch Gl. 37 (stetige Funktion) — obwohl $p \neq 1/2$ und n klein, ist die Näherung überraschend gut.

Wir interessieren uns jetzt für die Wahrscheinlichkeit, daß m zwischen zwei Werten a und b liegt. Für ein Bernoulli-Schema müßten wir exakterweise die Summe $\sum_{\nu=a}^b b(\nu; n, p)$ bilden (vgl. Gl. 33).

¹¹Beweis wird bei Gelegenheit ins Skriptum aufgenommen

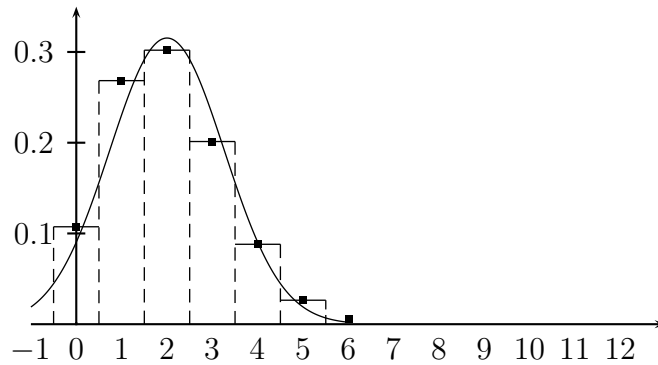


Abbildung 6: Vergleich zwischen $b(k; 10, 1/5)$ und der Näherung durch Gl. 37

Wenn wir Gl. 40 (bzw. Gl. 37) ausnützen wollen, so müssen wir die Summe durch das bestimmte Integral über die rechte Seite von Gl. 40 ersetzen. Wir erhalten also die Näherung

$$\sum_{\nu=a}^b b(\nu; n, p) \approx \frac{1}{\sigma\sqrt{2\pi}} \int_a^b dm e^{-\frac{(m-\mu)^2}{2\sigma^2}}. \tag{41}$$

Der Vollständigkeit sollte hinzugefügt werden, daß die mathematische Literatur eigentlich Gl. 41 bzw. Grenzwertdarstellungen, die zeigen daß der Fehler für $n \rightarrow \infty$ gegen Null geht als Grenzwertsatz von Moivre-Laplace bezeichnet.

Wir betrachten zunächst prinzipielle Eigenschaften dieses Integrals. Durch die Variablensubstitution $t = m - \mu/\sigma$ erhält das zu berechnende Integral die Form (*vollziehen Sie das nach, inklusive Änderung der Grenzen!!*)

$$\frac{1}{\sqrt{2\pi}} \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} dt e^{-t^2/2}$$

Wir haben schon im Abschnitt über Integration darauf hingewiesen, daß es für Integranden des Typs $\exp(ax^2)$ keine (in geschlossener Form schreibbare) Stammfunktion gibt. Integrale der Form

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z dt \exp(-t^2/2) \tag{42}$$

sind aber (wie wir noch sehen werden) in Wahrscheinlichkeitstheorie und Statistik von extremer Wichtigkeit und daher in detaillierten Tabellen verfügbar, bzw. heutzutage am Computer mit beliebiger Genauigkeit numerisch berechenbar. Man bezeichnet Gl. 42 als *Gaußsche Wahrscheinlichkeitsfunktion* oder auch als *Standard-Normalverteilung* $N(0, 1)$. Unter Ausnutzung von Gl. 42 läßt sich jetzt die Lösung unserer Frage in kompakter Form als

$$\sum_{\nu=a}^b b(\nu; n, p) \approx [\Phi((b - \mu)/\sigma) - \Phi((a - \mu)/\sigma)]. \tag{43}$$

schreiben. Die Werte von Φ bekommt man wie schon gesagt aus Tabellen (z.B. "kleiner Bartsch") oder durch numerische Integration. Für die Benutzung der Tabellen muß man bloß wissen, daß wegen

der Symmetrie des Integranden $\exp(-t^2/2)$ die Beziehung

$$\Phi(-z) = 1 - \Phi(z) \quad (44)$$

gilt, weswegen $\Phi(x)$ nur für positive x tabelliert ist.

► **Beispiel:** In Fortsetzung des obigen Beispiels fragen wir jetzt, was die Wahrscheinlichkeit ist, daß mit der fairen Münze zwischen 45 und 55 mal Kopf geworfen wird. Wir finden mit Gl. 43 und $\mu = np = 100 \cdot 0.5 = 50$ und $\sigma^2 = 100 \cdot 0.5 \cdot 0.5 = 25$

$$\sum_{\nu=45}^{55} b(\nu; 100, 1/2) \approx \Phi((55 - 50)/\sqrt{25}) - \Phi((45 - 50)/\sqrt{25}) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.68.$$

Der nach Gl. 33 berechnete exakte Wert ist allerdings 0.73, der Fehler der Näherung ist also nicht gerade klein. ◀

Wegen des im eben gerechneten Beispiel erhaltenen Fehler ist es notwendig, Näherung (43) nochmals zu diskutieren. Leitet man alle Schritte, die zu (41) bzw. (42) führen sorgfältig ab, so sieht man, daß im Grenzwert $n \rightarrow \infty$ der Fehler Null wird. Wie das Beispiel drastisch verdeutlicht, ist aber "100 \neq ∞ ". Es wäre daher wünschenswert, wenn wir eine Version von Gln. 41/43 hätten, die eine bessere Näherung für endliche n darstellt. Zu diesem Zweck betrachten wir nochmals Abb. 6. Das bestimmte Integral ist der Flächeninhalt unter der Glockenkurve, die die Binomialverteilung annähert. Interessieren tut uns aber eigentlich der Flächeninhalt unter der Stufenfunktion. Betrachtet man die Stufenfunktion aber genauer, so sieht man, daß an jedem Punkt ν der Funktionswert $b(\nu; n, p)$ für das Intervall $(\nu - 1/2, \nu + 1/2]$ aufgetragen ist. Das bedeutet aber, daß wir für die Summe $a \leq \nu \leq b$ (linke Seite in (41) eigentlich als Integrationsgrenzen $a - 1/2$ und $b + 1/2$ nehmen sollten, d.h.

$$\sum_{\nu=a}^b b(\nu; n, p) \approx \frac{1}{\sigma\sqrt{2\pi}} \int_{a-1/2}^{b+1/2} dm e^{-\frac{(m-\mu)^2}{2\sigma^2}} = \Phi((b + 1/2 - \mu)/\sigma) - \Phi((a - 1/2 - \mu)/\sigma). \quad (45)$$

► **Beispiel:** Als Test, ob Gl. 45 tatsächlich besser als (43) ist, berechnen wir nochmals die Wahrscheinlichkeit, daß mit der fairen Münze zwischen 45 und 55 mal Kopf geworfen wird. Wir finden mit $\mu = np = 100 \cdot 0.5 = 50$ und $\sigma^2 = 100 \cdot 0.5 \cdot 0.5 = 25$

$$\sum_{\nu=45}^{55} b(\nu; 100, .5) \approx \Phi((55 - 50 + 1/2)/\sqrt{25}) - \Phi((45 - 50 - 1/2)/\sqrt{25}) = \Phi(1.1) - \Phi(-1.1) = 2\Phi(1.1) - 1 = 0.73,$$

in ausgezeichneter Übereinstimmung mit dem exakten Wert. ◀

Für endliche n und höchste Genauigkeit sollte also Gl. 45 verwendet werden, um die Binomialverteilung zu nähern. Die Stärke der Gaußschen Näherung der Binomialverteilung liegt genau in Fragestellungen vom Typ Gl. 41. Nicht nur, daß für große n die Berechnung eines Werts $b(k; n, p)$ mühsam wird, die Berechnung von $\sum_{k=\alpha}^{\beta} b(k; n, p)$ mit der Binomialverteilung Gl. 33 würde $\beta - \alpha$ derartige Berechnungen erfordern. Gl. 43 (bzw. (45)) reduziert diese Rechnung auf das Nachschauen zweier tabellierter Werte.

5 Verteilungen, Erwartungswert und Varianz

5.1 Zufallsvariablen

Bis jetzt haben wir versucht, Wahrscheinlichkeiten von Elementar- oder zusammengesetzten Ereignissen eines Ereignisraums zu berechnen. Eine Hürde dabei war, daß Ereignisse oft mühsam mit Worten beschrieben werden mußten, was andererseits sicherlich der Vorstellung zuträglich war. Als nächstes führen wir einen Abstraktionsschritt durch, der es in Folge gestattet, den vollen mathematischen Apparat auf Probleme der Wahrscheinlichkeitsrechnung (und Statistik) anzuwenden (allerdings um den Preis, daß Vorstellbarkeit verlorengeht, und man sich mit etwas arkanen Bezeichnungen anfreunden muß).

Die zentrale Idee besteht darin, Funktionen (im mathematischen Sinn) einzuführen, die als Definitionsmenge den Ereignisraum eines interessierenden Zufallsexperiments haben. Eine Funktion \mathbf{X} , die jedem Elementarereignis E_i des Ereignisraums \mathcal{S} eines Zufallsexperiments eine reelle Zahl zuordnet, heißt Zufallsvariable.¹² Die Definitionsmenge von \mathbf{X} ist der Ereignisraum \mathcal{S} , die Wertemenge W ist eine Teilmenge von \mathbb{R} :

$$\mathbf{X} : \mathcal{S} \rightarrow W, E_i \mapsto \mathbf{X}(E_i)$$

Man unterscheidet zwischen *diskreten*, *stetigen* und *allgemeinen* Zufallsvariablen. Wir beschränken uns auf diskrete und stetige Zufallsvariable, diese reichen für die von Ihnen gebrauchten Anwendungen von Wahrscheinlichkeitsrechnung und Statistik. Die Behandlung allgemeiner Zufallsvariablen würde viel mehr mathematisches Rüstzeug (Maßtheorie, die ihrerseits wieder einen verallgemeinerten Integralbegriff voraussetzt) erfordern (und somit überlassen wir dieses Gebiet gerne den Mathematikern).

Diskrete Zufallsvariablen¹³ sind Funktionen, deren Wertemenge abzählbar oder abzählbar unendlich¹⁴ sind. Beispiele von diskreten Zufallsvariablen sind z.B. n-maliges Werfen einer Münze oder eines Würfels, die dabei entstehenden Ereignisse sind diskret, wengleich der Ereignisraum (und somit die Wertemengen von darauf definierten Zufallsvariablen) mit steigendem n rasch sehr groß werden können. Durch den Grenzübergang $n \rightarrow \infty$ gelangt man zu Beispielen von diskreten Zufallsvariablen mit abzählbar unendlichen Ereignisräumen bzw. Wertemengen. Eine diskrete Zufallsvariable mit abzählbar unendlicher Wertemenge ist z.B. die Anzahl von Würfelversuchen ("Mensch ärgere Dich nicht!" Spiel!), die benötigt wird, um zum ersten Mal eine Sechs zu würfeln ($\mathbf{X} = 1, 2, \dots, k$). Es gibt hierbei keine obere Grenze für die Anzahl k der Würfe, die dazu benötigt wird, und somit für den Wert den \mathbf{X} annehmen kann, wengleich die Wahrscheinlichkeit mit steigendem k sehr stark abnimmt $\mathbf{P}(\mathbf{X} = k) = (5/6)^{k-1}(1/6)$

Um sich *stetige* Zufallsvariable vorzustellen denken Sie an ein Roulettespiel, dessen Ergebnis (Er-

¹²In vielen Lehrbüchern heißt es, daß man besser von Zufallsfunktion sprechen sollte, aber die Gewohnheit ...

¹³Achtung: Manche Mittelschulbücher, z.B. Laub, "Lehrbuch der Mathematik", 4. Band, Seite 114 (Ausgabe von 1981) verwendet einen eingeschränktere Definition von diskreten Zufallsvariablen. Die dort getroffene Unterscheidung erscheint ein wenig spitzfindig und macht die Behandlung stetiger Zufallsvariablen komplizierter. Die hier verwendete Unterscheidung folgt [Bosch99]

¹⁴Das klassische Beispiel abzählbar unendlicher Mengen ist \mathbb{N}_0 . Eine abzählbar unendliche Menge muß auf \mathbb{N}_0 zurückführbar sein, dies ist z.B. für \mathbb{Z} und \mathbb{Q} der Fall, nicht jedoch für \mathbb{R} , bei der es sich um eine überabzählbar unendliche Menge handelt

eignisraum) nicht die Zahlen von 0 bis 36 sind, sondern der mit arbiträrer Genauigkeit¹⁵ gemessene Winkel θ zwischen 0 und 360° ($0 < \theta \leq 2\pi$). Klarerweise ist der zu diesem Zufallsexperiment gehörende Ereignisraum nicht diskret, und die zur beschriebenen Zufallsvariable gehörende Wertemenge ist überabzählbar unendlich.

5.2 Diskrete Zufallsvariablen

Wir verweilen noch kurz bei diskreten Ereignisräumen und Zufallsvariablen, denn diese eignen sich sehr gut, die nächsten wichtigen Konzepte, Wahrscheinlichkeitsverteilung (Abschnitt 5.2.1), Verteilungsfunktion (Abschnitt 5.2.2) und Erwartungswerte (Abschnitt 5.2.3), einzuführen. In Abschnitt 5.3 werden diese Konzepte dann für stetige Zufallsvariablen besprochen.

5.2.1 Wahrscheinlichkeitsverteilung — Wahrscheinlichkeitsfunktion

Es sei \mathbf{X} eine Zufallsvariable (oder besser Zufallsfunktion), die die Werte x_1, x_2, x_3, \dots annehmen kann. Die Anhäufung aller Ereignisse für die \mathbf{X} einen bestimmten Wert x_i annimmt bildet das Ereignis $\mathbf{X} = x_i$, wir bezeichnen seine Wahrscheinlichkeit mit $\mathbf{P}(\mathbf{X} = x_i) = f(x_i)$. Die Funktion

$$\mathbf{P}(\mathbf{X} = x_i) = f(x_i) \quad (i = 1, 2, \dots) \quad (46)$$

wird als *Wahrscheinlichkeitsverteilung* oder auch als *Wahrscheinlichkeitsfunktion* der Zufallsvariablen \mathbf{X} bezeichnet. Aus den Eigenschaften von Wahrscheinlichkeiten (Abschnitt 1.2.2) folgen sofort die folgenden Eigenschaften von $f(x_i)$

$$f(x_i) \geq 0 \quad (47)$$

und

$$\sum f(x_i) = 1. \quad (48)$$

Als Beispiel einer Wahrscheinlichkeitsfunktion betrachten wir das Würfeln mit zwei Würfeln unterschiedlicher Farbe. Die Zufallsvariable \mathbf{X} ist die Augenzahl, die Werte zwischen 2 und 12 annehmen kann. Die für eine Augenzahl günstigen Ereignisse, die Augenzahl (d.h. die Zufallsvariable) und die Wahrscheinlichkeitsverteilung (Wahrscheinlichkeitsfunktion) sind in Tabelle 1 (S. 29) dargestellt. Die Summe $\sum f(x_i)$ ist wie lt. Gleichung 48 gefordert gleich 1.

5.2.2 Verteilungsfunktion

Die eben eingeführte Wahrscheinlichkeitsverteilung $f(x_i)$ gibt die Wahrscheinlichkeit $\mathbf{P}(\mathbf{X} = x_i)$ an, mit der die Zufallsvariable \mathbf{X} den Wert x_i aus ihrer Wertemenge annimmt. Oft interessiert man sich für die Wahrscheinlichkeit dafür, daß eine Zufallsvariable \mathbf{X} Werte annimmt, die nicht größer als ein fest vorgegebener Wert x sind, d.h. für $\mathbf{P}(\mathbf{X} \leq x)$. Läßt man x die Zahlengerade \mathbb{R} durchlaufen, so wird durch

$$F(x) = \mathbf{P}(\mathbf{X} \leq x), \quad x \in \mathbb{R} \quad (49)$$

¹⁵Wir halten uns hier nicht mit Problemen auf, die uns hierbei irgendwann aus der Unschärferelation der Quantenmechanik erwachsen würden.

günstige Ereignisse	\mathbf{X}	$\mathbf{P}(x_i) = f(x_i)$
(1,1)	2	1/36
(1,2);(2,1)	3	2/36
(1,3);(2,2);(3,1)	4	3/36
(1,4);(2,3);(3,2);(4,1)	5	4/36
(1,5);(2,4);(3,3);(4,2);(5,1)	6	5/36
(1,6);(2,5);(3,4);(4,3);(5,2);(6,1)	7	6/36
(2,6);(3,5);(4,4);(5,3);(6,2)	8	5/36
(3,6);(4,5);(5,4);(6,3)	9	4/36
(4,6);(5,5);(6,4)	10	3/36
(5,6);(6,5)	11	2/36
(6,6)	12	1/36
		$\Sigma f(x_i) = 1$

Tabelle 1: Illustration von Zufallsvariable $\mathbf{X} = 2, 3, \dots, 12$ und Wahrscheinlichkeitsverteilung $f(x_i)$ am Beispiel der Augensumme beim Würfeln mit zwei (unterscheidbaren) Würfeln.

eine reellwertige Funktion F erklärt. Diese Funktion F , die durch die Zufallsvariable \mathbf{X} bestimmt ist, heißt *Verteilungsfunktion* von \mathbf{X} .¹⁶ Für diskrete Zufallsvariablen berechnet sich $F(x)$ gemäß

$$F(x) = \mathbf{P}(\mathbf{X} \leq x) = \sum_{x_i \leq x} \mathbf{P}(\mathbf{X} = x_i) = \sum_{x_i \leq x} f(x_i) \tag{50}$$

Der Zusammenhang zwischen $f(x_i)$ und $F(x)$ ist in Abb. 7 für den bereits in Tabelle 1 diskutierten Fall der Augensumme zweier (unterscheidbarer) Würfel illustriert.

Jede Verteilungsfunktion $F(x)$ einer diskreten Zufallsvariablen \mathbf{X} besitzt folgende Eigenschaften, die unmittelbar aus der Definition (50) bzw. den Eigenschaften der Wahrscheinlichkeitsverteilung $f(x_i)$, Gln. 47 und 48, folgen: $F(x)$ ist eine Treppenfunktion, die nur an den Stellen x_i aus der Wertemenge W von \mathbf{X} Sprünge der Höhe $f(x_i) = \mathbf{P}(\mathbf{X} = x_i)$ besitzt. Weiters folgt aus

$$x < \hat{x} \quad \Rightarrow \quad F(x) \leq F(\hat{x}). \tag{51}$$

$F(x)$ kann also nie kleiner werden (wenn sie auch zwischen den x_i konstant bleibt), man bezeichnet dieses Verhalten als *monoton nichtfallend*. Dieses Verhalten ist sehr schön in Abb. 7 für unser Beispiel der Augensumme zweier Würfel illustriert. Weiters werden die Funktionswerte $F(x)$ beliebig klein, wenn nur x klein genug gewählt wird (in Abb. 7 z.B. ist $F(x)=0$ für alle $x < 2$. Andererseits geht $F(x)$ für große x gegen 1 (in Abb. 7 z.B. ist dies für $x \geq 12$ der Fall). Formal läßt sich dieses Verhalten durch die beiden Grenzwerte

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1 \tag{52}$$

ausdrücken.

¹⁶Bitte verwechseln Sie nicht eine *Wahrscheinlichkeitsverteilung* $f(x_i)$ mit einer *Verteilungsfunktion* $F(x)$. Wegen dieser Verwechslungsgefahr verwenden manche Lehrbücher den Ausdruck “Wahrscheinlichkeitsfunktion” für $f(x_i)$.

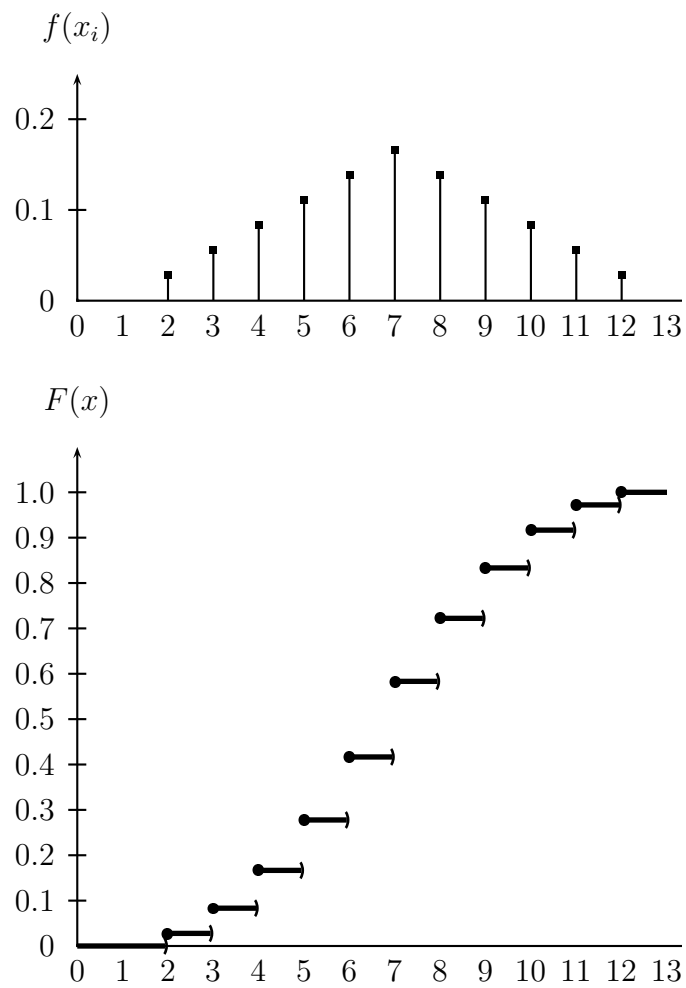


Abbildung 7: $f(x_i)$ und $F(x)$ für die Zufallsvariable “Augensumme zweier unterscheidbarer Würfel” (siehe auch Tabelle 1)

Verteilungsfunktionen spielen in der Wahrscheinlichkeitstheorie eine zentrale Rolle. Einer der Gründe dafür ist, daß aus der Kenntnis der Verteilungsfunktion $F(x)$ (im Falle einer diskreter Zufallsvariablen \mathbf{X}) sowohl die Wertemenge W von \mathbf{X} als auch die Wahrscheinlichkeitsverteilung $f(x_i)$ abgeleitet werden kann. Erfüllt nämlich eine Treppenfunktion die im vorigen Absatz beschriebenen Eigenschaften, insb. Gl. 51 und 52, so besteht W aus den Sprungstellen dieser Funktion (d.h. der Verteilungsfunktion) und $f(x_i) = \mathbf{P}(\mathbf{X} = x_i)$ ist gleich der Sprunghöhe an der Stelle x_i . Weiters lassen sich aus $F(x)$ sehr einfach die Wahrscheinlichkeit dafür berechnen, daß \mathbf{X} Werte aus einem Intervall annimmt. Es gilt

$$\mathbf{P}(a < \mathbf{X} \leq b) = F(b) - F(a) \quad (53)$$

$$\mathbf{P}(a \leq \mathbf{X} \leq b) = F(b) - F(a - o) \quad (54)$$

$$\mathbf{P}(a < \mathbf{X}) = 1 - F(a). \quad (55)$$

Die Notation $F(a - o)$ bezeichnet den linksseitigen Grenzwert von $F(x)$ an der Stelle a : Dieser ist gleich $F(a)$, wenn a keine Sprungstelle ist, und sonst gleich dem Wert der Treppenstufe, die unmittelbar *links* neben dem Punkt a liegt.

5.2.3 Erwartungswerte

Die letzten beiden Abschnitte haben hoffentlich verdeutlicht, daß Kenntnis der Wahrscheinlichkeitsverteilung bzw. der Verteilungsfunktion die vollständige Information über die einem Zufallsexperiment zugeordneten Werte der Zufallsvariable und deren Wahrscheinlichkeiten beinhaltet. Ausgewählte, sogenannte *Erwartungswerte* der Zufallsvariablen \mathbf{X} bzw. von Funktionen von \mathbf{X} , $\phi(\mathbf{X})$, dienen nun dazu die Wahrscheinlichkeitsverteilung bzw. Verteilungsfunktion in möglichst kompakter Form zu charakterisieren.

Vorbemerkung bzw. “Disclaimer”: Die folgenden Definitionen beinhalten *Reihen* der Form $E(\phi(\mathbf{X})) = \sum \phi(x_i)f(x_i)$. Im Rahmen dieser Vorlesung war es nicht möglich, Reihen in systematischer Form zu behandeln und uns mit deren Konvergenz zu beschäftigen. Wir setzen daher von jetzt an stillschweigend voraus, daß die auftretenden Reihen *absolut konvergent* sind,¹⁷ nur in diesem Fall existiert der entsprechende Erwartungswert $E(\phi(\mathbf{X}))$.

Wir führen zunächst den Erwartungswert einer Zufallsvariablen \mathbf{X} ein. Es sei \mathbf{X} eine Zufallsvariable mit Werten x_1, x_2, \dots und korrespondierenden Wahrscheinlichkeiten $f(x_1), f(x_2), \dots$. Dann bezeichnet man

$$E(\mathbf{X}) = \sum x_i f(x_i) \quad (56)$$

als den Erwartungswert von \mathbf{X} . Überlegen wir uns einmal, was Gl. 56 heißt. Jeder Wert der Zufallsvariable wird mit seiner Wahrscheinlichkeit (NB: $f(x_i) = \mathbf{P}(\mathbf{X} = x_i)$) multipliziert (gewichtet), und alle Terme werden summiert. Die entspricht genau der Berechnungsvorschrift eines gewichteten Mittelwerts, d.h., wenn jeder Wert nicht gleich häufig auftritt bzw. nicht gleichwahrscheinlich ist. (Ein praktisches Beispiel eines gewichteten Mittelwerts ist die Berechnung des Notendurchschnitts:

$$\frac{1 \times n_1 + 2 \times n_2 + 3 \times n_3 + 4 \times n_4 + 5 \times n_5}{n} = 1 \times \frac{n_1}{n} + 2 \times \frac{n_2}{n} + 3 \times \frac{n_3}{n} + 4 \times \frac{n_4}{n} + 5 \times \frac{n_5}{n},$$

¹⁷d.h. daß nicht nur die Reihe, sondern auch $\sum |\phi(x_i)|f(x_i)$ konvergiert.

d.h. eins mal der relativen Häufigkeit der Eins (n_1/n) plus zwei mal der relativen Häufigkeit der Zwei usw.) Die Berechnung dieses empirischen Notenmittelwerts ist völlig analog der Berechnung des Erwartungswerts von \mathbf{X} in unserem Modellbeispiel Augensumme zweier Würfel (s. Tabelle 1): Jede mögliche Augensumme (= Werte der Zufallsvariablen) wird mit ihrer entsprechenden Wahrscheinlichkeit multipliziert und diese Produkte werden summiert. Aufgrund der Ähnlichkeit zwischen einem gewichteten (empirischen) Mittelwert und dem Erwartungswert einer Zufallsvariablen bezeichnet man

$$E(\mathbf{X}) = \mu \tag{57}$$

auch als Mittel- oder Durchschnittswert und auch als Mittelwert der Verteilung. Weitere, in Anwendungen gebräuchliche Abkürzungen sind $\langle \mathbf{X} \rangle$ und $\bar{\mathbf{X}}$.

Wendet man eine beliebige Funktion ϕ auf eine Zufallsvariable an, so erhält man die neue Zufallsvariable $\phi(\mathbf{X})$. In Verallgemeinerung von Gl. 56 definiert man den Erwartungswert dieser neuen Zufallsvariablen als

$$E(\phi(\mathbf{X})) = \sum \phi(x_i)f(x_i) \tag{58}$$

In der mathematischen Wahrscheinlichkeitstheorie sind die sogenannten r -ten Momente der Zufallsvariablen \mathbf{X} um den Ursprung, $E(\mathbf{X}^r) = \sum x_i^r f(x_i)$ von großem Interesse. Wir begnügen uns mit dem zweiten Moment $E(\mathbf{X}^2)$, ersetzen jedoch \mathbf{X} durch die Abweichung vom Mittelwert $\mathbf{X} - \mu$. Wir berechnen also

$$E((\mathbf{X} - \mu)^2) = \sum_i (x_i^2 - 2\mu x_i + \mu^2)f(x_i) = \underbrace{\sum_i x_i^2 f(x_i)}_{=E(\mathbf{X}^2)} - 2\mu \underbrace{\sum_i x_i f(x_i)}_{=\mu} + \mu^2 \underbrace{\sum_i f(x_i)}_{=1},$$

wobei wir im zweiten Schritt ausgenützt haben, daß wir Terme, die nicht von der Zählvariablen abhängen, aus der Summation herausheben können, und Gln. 58, 56 sowie 48 angewandt haben. Nach dem Zusammenfassen der Terme erhält man

$$E((\mathbf{X} - \mu)^2) = E(\mathbf{X}^2) - \mu^2 = \text{Var}(\mathbf{X}) = \sigma^2. \tag{59}$$

Man bezeichnet diese Größe als *Varianz* von \mathbf{X} , $\text{Var}(\mathbf{X}) = \sigma^2$. Die positive Wurzel der Varianz σ bezeichnet man als Standardabweichung.

► **Beispiel:** Mit Hilfe von Tabelle 1 berechnen wir μ und σ^2 unseres Beispiels Augenzahl zweier Würfel:

$$\mu = \sum_{j=2}^{12} j \mathbf{P}(\mathbf{X} = j) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + \dots = 7$$

$$\sigma^2 = \sum_{j=2}^{12} j^2 \mathbf{P}(\mathbf{X} = j) - \mu^2 = 2^2 \times \frac{1}{36} + 3^2 \times \frac{2}{36} + \dots - 7^2 \approx 5.83$$



Um dieses Kapitel abzuschließen betrachten wir noch die Eigenschaften von μ und σ unter der linearen Transformation $\mathbf{X} \rightarrow a\mathbf{X} + b$, wobei a und b Konstanten sind. Es gilt

$$E(a\mathbf{X} + b) = aE(\mathbf{X}) + b \tag{60}$$

und

$$\text{Var}(a\mathbf{X} + b) = a^2\text{Var}(\mathbf{X}). \quad (61)$$

Um Gleichungen 60 und 61 zu beweisen, berechnen wir zunächst den Erwartungswert einer Konstanten $E(c)$,

$$E(c) = \sum cf(x_i) = c \sum f(x_i) = c.$$

(Hinweis: Gl. 48!) Somit erhalten wir für

$$E(a\mathbf{X} + b) = \sum (ax_i + b)f(x_i) = \sum ax_if(x_i) + \sum bf(x_i) = a \sum x_if(x_i) + b \sum f(x_i) = aE(\mathbf{X}) + b$$

womit Gl. 60 bewiesen ist. Um Gl. 61 zu verifizieren, nützen wir gleich Gl. 60:

$$\begin{aligned} \text{Var}(a\mathbf{X} + b) &= E((a\mathbf{X} + b - E(a\mathbf{X} + b))^2) = E((a\mathbf{X} + b - (aE(\mathbf{X}) + b))^2) = \\ &= E((a(\mathbf{X} - E(\mathbf{X})))^2) = E(a^2(\mathbf{X} - \mu)^2) = a^2E((\mathbf{X} - \mu)^2) = a^2\text{Var}(\mathbf{X}). \end{aligned}$$

Intuitiv ist es logisch, daß sich der Mittelwert durch Multiplikation bzw. Verschiebung des Ursprungs verändert, gleichzeitig ist es verständlich, daß eine Verschiebung des Ursprungs keinen Einfluß auf die Varianz hat.

Gleichungen 60 und 61 sind einerseits für Beweise in Zusammenhang mit Mittelwerten und Varianzen wichtig, andererseits verwendet man sie, um die sogenannte *Standardisierte* \mathbf{X}^* der Zufallsvariablen \mathbf{X} einzuführen. Hat eine Zufallsvariable \mathbf{X} Mittelwert μ und Varianz σ^2 , so folgt aus Gl. 60, daß $\mathbf{X} - \mu$ den Mittelwert 0 hat ($E(\mathbf{X} - \mu) = E(\mathbf{X}) - \mu = 0$). Die Zufallsvariable $\frac{\mathbf{X} - \mu}{\sigma}$ hat weiters nach Gl. 61 die Varianz $\frac{1}{\sigma^2}\sigma^2 = 1$, wegen dieser Eigenschaften bezeichnet man

$$\frac{\mathbf{X} - \mu}{\sigma} = \mathbf{X}^* \quad (62)$$

als Standardisierte von \mathbf{X} .

5.2.4 Binomialverteilung revisited

In diesem und dem nächsten Abschnitt betrachten wir einige Eigenschaften der in Abschnitt 4 eingeführten Binomial- bzw. Poissonverteilung (Absch. 5.2.5) unter den in diesem Kapitel eingeführten Gesichtspunkten (Zufallsvariable, Wahrscheinlichkeitsverteilung, Erwartungswerte).

Die Zufallsvariable \mathbf{X} sei \mathbf{S}_n , die Anzahl der Erfolge in einem n -mal durchgeführten Bernoulliexperiment. Die Wertemenge von \mathbf{S}_n ist demzufolge $\mathbf{S}_n = 0, 1, 2, \dots, n$ und

$$\mathbf{P}(\mathbf{S}_n = m) = b(m; n, p) = f(m) \quad (63)$$

wobei p die Wahrscheinlichkeit $\mathbf{P}(S) = p$ des Erfolgs des Einzelexperiments ist (vgl. 4.1). Nachdem durch Gl. 63 $b(m; n, p)$ als Wahrscheinlichkeitsfunktion identifiziert wurde, erklärt sich auch warum man von *Binomialverteilung* spricht. Die zugehörige Verteilungsfunktion $F(x)$ ist durch

$$F(x) = \sum_{\nu \leq x} b(\nu; n, p)$$

gegeben. (Für $x < 0$ gilt klarerweise $F(x) = 0$, für $x > n$ $F(x) = 1$, dies braucht nicht extra definiert werden, sondern folgt aus den Eigenschaften des Binomialkoeffizienten (s. Abschnitt 2.5, S. 11).

Als nächstes beweisen wir die Gültigkeit von Gl. 48 für die Binomialverteilung:

$$\sum_{\nu=0}^n b(\nu; n, p) = \sum_{\nu=0}^n \binom{n}{\nu} p^\nu q^{n-\nu} = (p + q)^n = (p + (1 - p))^n = 1^n = 1;$$

diese folgt unmittelbar aus dem binomischen Lehrsatz $(a + b)^n = \sum_{j=0}^n \binom{n}{j} a^j b^{n-j}$.

Weiters interessieren wir uns für Mittelwert und Varianz. (*Achtung:* Die technischen Details der folgenden Ableitung sind ziemlich kompliziert. Dasselbe Ergebnis wird in Abschnitt 5.4 nochmals abgeleitet, die dortige Rechnung erfordert zwar mehr theoretischen Background, ist aber technisch um vieles einfacher.)

$$\mu = \sum_{\nu=0}^n \nu \binom{n}{\nu} p^\nu q^{n-\nu} = \sum_{\nu=1}^n \nu \binom{n}{\nu} p^\nu q^{n-\nu},$$

da der erste Term wegfällt. Wir heben jetzt np aus der Summe und kürzen durch ν :

$$\mu = \sum_{\nu=1}^n \nu \frac{n!}{\nu!(n-\nu)!} p^\nu q^{n-\nu} = np \sum_{\nu=1}^n \frac{(n-1)!}{(\nu-1)!(n-\nu)!} p^{\nu-1} q^{n-\nu}$$

Als nächstes ändern wir jetzt den Summationsindex und setzen $\nu - 1 = j$, diese Summationsvariable beginnt daher bei $j = 0$ und endet bei $j = n - 1$. An allen Stellen wo noch ν vorkommt, ersetzen wir ν durch $j + 1$.

$$\mu = np \sum_{\nu=1}^n \frac{(n-1)!}{(\nu-1)!(n-\nu)!} p^{\nu-1} q^{n-\nu} = np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-(j+1))!} p^j q^{n-(j+1)} = np \underbrace{\sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j}}_{(p+q)^{n-1}=1}$$

In den letzten Schritten haben wir $n - (j + 1)$ zu $(n - 1) - j$ umgruppiert.

Die Berechnung der Varianz erfolgt analog, wir zeigen die Rechnung mit Zwischenschritte, aber ohne weitere Erläuterungen:

$$\begin{aligned} \sigma^2 &= \sum_{\nu=0}^n \nu^2 \binom{n}{\nu} p^\nu q^{n-\nu} - \mu^2 = \sum_{\nu=1}^n \nu^2 \binom{n}{\nu} p^\nu q^{n-\nu} - \mu^2 = np \sum_{\nu=1}^n \nu \frac{(n-1)!}{(\nu-1)!(n-\nu)!} p^{\nu-1} q^{n-\nu} - \mu^2 = np \sum_{\nu=1}^n (\nu-1+1) \frac{(n-1)!}{(\nu-1)!(n-\nu)!} p^{\nu-1} q^{n-\nu} - \mu^2 = \\ & np \left(\sum_{\nu=2}^n (\nu-1) \frac{(n-1)!}{(\nu-1)!(n-\nu)!} p^{\nu-1} q^{n-\nu} + \underbrace{\sum_{\nu=1}^n \frac{(n-1)!}{(\nu-1)!(n-\nu)!} p^{\nu-1} q^{n-\nu}}_{=1} \right) - \mu^2 = np \left((n-1)p \sum_{\nu=2}^n \frac{(n-2)!}{(\nu-2)!(n-\nu)!} p^{\nu-2} q^{n-\nu} + 1 \right) - \mu^2 = \\ & np \left((n-1)p \underbrace{\sum_{j=0}^{n-2} \frac{(n-2)!}{j!(n-2-j)!} p^j q^{n-2-j}}_{=(p+q)^{n-2}=1} + 1 \right) - \mu^2 = np((n-1)p + 1) - (np)^2 = np(1-p) = npq \end{aligned}$$

Wir fassen zusammen: Mittelwert und Varianz einer binomialverteilten Zufallsvariablen sind durch

$$\mu = np \quad \sigma^2 = npq = np(1-p) \tag{64}$$

gegeben.

5.2.5 Poissonverteilung revisited

In Abschnitt 4.3 wurde die Poissonverteilung einerseits als Grenzfall der Binomialverteilung für große n und kleine p abgeleitet, weiters wurde darauf hingewiesen, daß viele Prozesse der Poissonverteilung selbst gehorchen (z.B. radioaktiver Zerfall). Wir zeigen zunächst, daß die Poissonverteilung $p(m; \lambda) = e^{-\lambda} \frac{\lambda^m}{m!}$ der Bedingung (48) genügt, d.h., daß es sich um eine Wahrscheinlichkeitsverteilung handelt.

$$\sum_{m=0}^{\infty} e^{-\lambda} \frac{\lambda^m}{m!} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda} e^{\lambda} = 1$$

Ich hoffe, Sie haben die Reihe $\sum_{m=0}^{\infty} x^m/m!$ als die Taylorreihe der Exponentialfunktion e^x um den Entwicklungspunkt $x_0 = 0$ erkannt! Es handelt sich also bei der Poissonverteilung tatsächlich um eine Wahrscheinlichkeitsverteilung $p(m; \lambda) = f(m)$. Die Zufallsvariable ist wie im Falle der Binomialverteilung das m -malige Eintreten des interessierenden Ereignisses (z.B. m radioaktive Zerfälle). \mathbf{X} kann die Werte 1, 2, 3, ... annehmen. Im Gegensatz zu binomialverteilten Ereignissen, sind bei Prozessen, die der Poissonverteilung gehorchen sowohl der Ereignisraum, als auch die Wertemenge von \mathbf{X} (abzählbar) unendlich.

Da die Berechnung von Mittelwert und Varianz der Poissonverteilung im Vergleich zur Binomialverteilung einfacher ist, skizzieren wir kurz die notwendigen Schritte:

$$\begin{aligned} \mu &= \sum_{m=0}^{\infty} m e^{-\lambda} \frac{\lambda^m}{m!} = \sum_{m=1}^{\infty} m e^{-\lambda} \frac{\lambda^m}{m!} = e^{-\lambda} \lambda \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} = e^{-\lambda} \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda \\ \sigma^2 &= \left(\sum_{m=0}^{\infty} m^2 e^{-\lambda} \frac{\lambda^m}{m!} \right) - \lambda^2 = \lambda \left(\underbrace{\sum_{m=1}^{\infty} (m-1) e^{-\lambda} \frac{\lambda^{(m-1)}}{(m-1)!}}_{= \lambda} + \underbrace{\sum_{m=1}^{\infty} e^{-\lambda} \frac{\lambda^{(m-1)}}{(m-1)!}}_{= 1} \right) - \lambda^2 = \lambda \end{aligned}$$

Wir fassen zusammen: Mittelwert und Varianz einer Poisson-verteilten Zufallsvariablen sind durch

$$\mu = \lambda \quad \sigma^2 = \lambda \quad (65)$$

gegeben. Dieses Ergebnis für den Mittelwert $\mu = \lambda = np$ (vgl. Gl. 35) erklärt im übrigen auch warum man bei (annähernd) Poisson-verteilten Prozessen davon spricht, ein Ereignis trete *im Mittel* $\lambda = np$ mal auf (vgl. Seite 24).

5.3 Stetige Zufallsvariablen

5.3.1 Verteilungsfunktion — Dichte

Genauso wie wir uns die sehr allgemeinen Konzepte Wahrscheinlichkeitsverteilung, Verteilungsfunktion und Erwartungswerte für diskrete Zufallsvariablen zunächst am einfachen Beispiel der Augenzahl zweier Würfel verdeutlicht haben, verwenden wir zunächst das Beispiel des fiktiven, "kontinuierlichen" Roulettespiels von Seite 27, um den Fall stetiger Zufallsvariablen zu behandeln. Die Wertemenge der Zufallsvariablen $\mathbf{X} = \theta$ dieses modifizierten Roulettespiels umfasst wie gesagt das halboffene

Intervall $I : 0 < \theta \leq 2\pi$, diese Wertemenge ist überabzählbar unendlich und somit kann es sich bei θ um keine diskrete Zufallsvariable handeln.

Wir überlegen uns nun, wie für dieses Beispiel eine in Analogie zu Gl. 49 eingeführte *Verteilungsfunktion* $F(x)$ aussieht, $F(x) = \mathbf{P}(\theta \leq x)$, $x \in \mathbb{R}$. Da nur Winkel aus dem Intervall $0 < \theta \leq 2\pi$ möglich sind, müssen folgende Wahrscheinlichkeiten gelten: $\mathbf{P}(0 < \theta \leq 2\pi) = 1$, $\mathbf{P}(\theta \leq 0) = 0$ und $\mathbf{P}(\theta > 2\pi) = 0$. Weiters gehen wir davon aus, daß unser Roulette fair ist, somit muß für jedes $x \in I$ die Verhältnisgleichung

$$\mathbf{P}(\theta \leq x) : \mathbf{P}(\theta \leq 2\pi) = x : 2\pi$$

gelten. Zu diesen Überlegungen paßt folgende Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ \frac{x}{2\pi} & \text{für } 0 < x \leq 2\pi \\ 1 & \text{für } x > 2\pi \end{cases}$$

Wir verknüpfen jetzt diese Überlegung bezüglich der Verteilungsfunktion in unserem Beispiel mit folgender Definition: *Eine Zufallsvariable \mathbf{X} heißt stetig, wenn eine nichtnegative, integrierbare Funktion f existiert, so daß für ihre Verteilungsfunktion $F(x) = \mathbf{P}(\mathbf{X} \leq x)$ die Integraldarstellung*

$$F(x) = \mathbf{P}(\mathbf{X} \leq x) = \int_{-\infty}^x du f(u) \quad (66)$$

gilt. Eine derartige Funktion f heißt *Dichte* der stetigen Zufallsvariablen \mathbf{X} .

Für unser Beispiel findet man (z.B. durch Differentiation nach x) sofort die zu $F(x)$ gehörige Dichte

$$f(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ \frac{1}{2\pi} & \text{für } 0 < x \leq 2\pi \\ 0 & \text{für } x > 2\pi \end{cases}$$

Der Zusammenhang zwischen $f(x)$ und $F(x)$ für das Beispiel des "kontinuierlichen Roulettespiels ist in Abb. 8 dargestellt.

Welche Eigenschaften haben nun Verteilungsfunktion und Dichte stetiger Zufallsvariablen, welche Unterschiede (wenn überhaupt) gibt es zu diskreten Zufallsvariablen? $F(x)$ ist die Wahrscheinlichkeit $\mathbf{P}(\mathbf{X} \leq x)$, daß \mathbf{X} kleiner oder gleich x ist. Gemäß der Definition Gl. 66 fängt die Wertemenge von \mathbf{X} bei $-\infty$ an (ist die Wertemenge von \mathbf{X} wie in unserem Beispiel auf ein Intervall $a < x \leq b$ beschränkt, so gilt automatisch für alle $x \leq a$, daß $F(x) = \mathbf{P}(\mathbf{X} \leq x | x \leq a) = 0$. Wenn wir jetzt das Argument von F gegen $+\infty$ gehen lassen, so überdecken wir ganz sicher die komplette Wertemenge von \mathbf{X} , d.h., es handelt sich um das sichere Ereignis, und somit gilt $\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} \mathbf{P}(\mathbf{X} \leq x) = 1$. Zusammen mit Gl. 66 heißt das aber

$$\int_{-\infty}^{+\infty} dx f(x) = 1. \quad (67)$$

Gleichung 67 ist das Analogon zu Gleichung 48 für stetige Zufallsvariable.

Für $F(x)$ gelten weiters folgende Eigenschaften:

$$\mathbf{P}(a < \mathbf{X} \leq b) = F(b) - F(a) = \int_a^b du f(u) \quad (68)$$

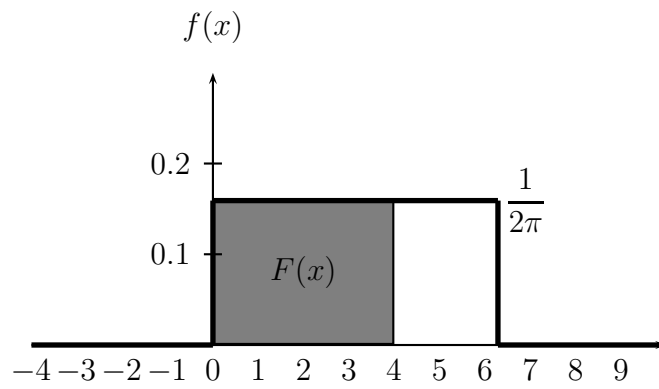
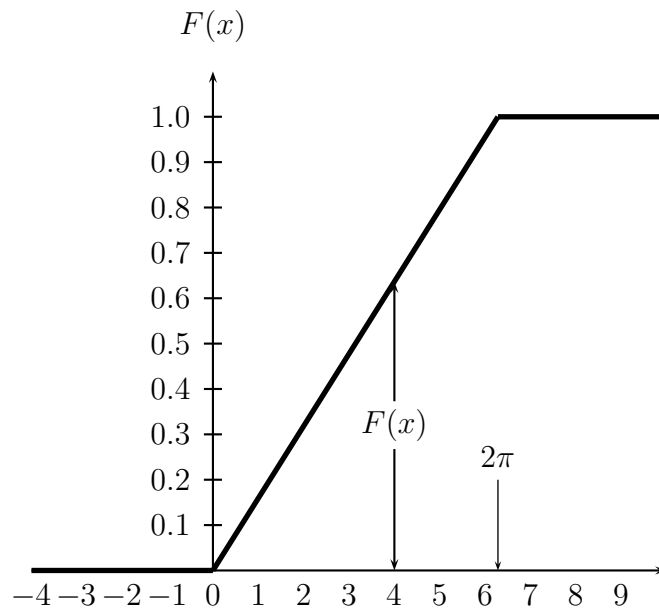


Abbildung 8: Dichte und Verteilung des “kontinuierlichen Roulettespiels”

und

$$\mathbf{P}(\mathbf{X} > c) = \int_c^{+\infty} du f(u) = \underbrace{\int_{-\infty}^c du f(u) + \int_c^{+\infty} du f(u)}_{=1} - \int_{-\infty}^c du f(u) = 1 - F(c) \quad (69)$$

Der Beweis der beiden Beziehungen (wenn nicht ohnedies angedeutet) folgt unmittelbar aus den Eigenschaften des bestimmten Integrals.

Die Verteilungsfunktion stetiger Zufallsvariablen ist *stetig* (einer der großen Unterschiede zu allgemeinen Zufallsvariablen). Die Dichte hingegen muß nicht stetig sein, wie man schon an unserem Beispiel sieht (Sprungstellen bei 0 und 2π). Um die Dichte $f(x)$ besser zu verstehen, fragen wir jetzt nach der Wahrscheinlichkeit $\mathbf{P}(\mathbf{X} = x_0)$. Wir beginnen mit dem Hilfskonstrukt

$$\mathbf{P}(x_0 - h < \mathbf{X} \leq x_0) = F(x_0) - F(x_0 - h) = \int_{x_0-h}^{x_0} du f(u)$$

welches sicher größer als die gesuchte Wahrscheinlichkeit $\mathbf{P}(\mathbf{X} = x_0)$ ist. Machen wir h beliebig klein ($\lim_{h \rightarrow 0}$), so erhalten wir zunächst das verblüffende Ergebnis

$$\mathbf{P}(\mathbf{X} = x_0) = \lim_{h \rightarrow 0} \int_{x_0-h}^{x_0} du f(u) = \int_{x_0}^{x_0} du f(u) = 0$$

für jedes(!) $x_0 \in \mathbb{R}$. Die Betrachtung sollte klar machen, daß $f(x_0)$ *nicht* die Wahrscheinlichkeit ist, mit der \mathbf{X} den Wert x_0 annimmt, dies ist ein großer Unterschied zwischen einer Dichte $f(x)$ und einer (diskreten) Wahrscheinlichkeitsverteilung $f(x_i)$. Da es sich nicht um eine Wahrscheinlichkeit handelt, kann die Dichte auch Werte größer als 1 annehmen! Für stetige Zufallsvariablen ist der Begriff der Wahrscheinlichkeit nur mehr für Intervalle der Form $\delta x = x_0 - (x_0 - h)$ (auch wenn diese arbiträr klein werden können), nicht jedoch für den isolierten Punkt x_0 sinnvoll. Das Produkt $\delta x f(x_0) = [x_0 - (x_0 - h)]f(x_0) = h f(x_0)$ ist die Rechtecksnäherung des Integrals $\int_{x_0-h}^{x_0} du f(u)$. Im Grenzübergang von Summation zu Integration wird das $\delta x = h$ zum dx des Integrals.

5.3.2 Erwartungswerte

Wir ignorieren in diesem Abschnitt mathematische Feinheiten (insbesondere Unstetigkeiten der Dichte) und erarbeiten die für uns wichtigen Ergebnisse durch Analogieschluß vom Fall diskreter Zufallsvariablen. Im letzten Abschnitt hatten wir gesehen, daß die Summation, die von einer (diskreten) Wahrscheinlichkeitsverteilung $f(x_i)$ auf eine (diskrete) Verteilungsfunktion $F(x)$ führte, bei stetigen Zufallsvariablen durch ein Integral über eine entsprechende Dichte ersetzt wurde. Wenngleich eine Dichte nicht eins zu eins einer Wahrscheinlichkeitsverteilung entspricht, so übt sie doch die gleiche Funktion aus. Dies legt nahe, einen Erwartungswert $E(\phi(\mathbf{X}))$ einer stetigen Zufallsvariablen durch

$$E(\phi(\mathbf{X})) = \int_{-\infty}^{+\infty} dx \phi(x) f(x) \quad (70)$$

zu definieren (vorausgesetzt, das Integral existiert). Der Übergang von Summation zu Integration bietet sich ja auch aus der Ableitung des (einzigen uns bekannten Riemannsches) Integralbegriffs an. Wir belassen die Grenzen des Integrals bei $\pm\infty$. Ist, wie bei unserem Roulettebeispiel, die Dichte nur

auf einem Intervall ungleich Null, so tragen alle Bereich außerhalb dieses Intervalls nicht zu $E(\phi(\mathbf{X}))$ bei; alternativ können dann natürlich als Integrationssschranken auch die Intervallgrenzen verwendet werden.

Insbesondere definieren wir wieder Mittelwert und Varianz durch

$$E(\mathbf{X}) = \mu = \int_{-\infty}^{+\infty} dx x f(x) \quad (71)$$

und

$$E((\mathbf{X} - \mu)^2) = \sigma^2 = \int_{-\infty}^{+\infty} dx (x - \mu)^2 f(x) = E(\mathbf{X}^2) - \mu^2. \quad (72)$$

Die letzte Identität in Gl. 72 folgt unmittelbar aus den Eigenschaften des Integrals, Definition (71) und Gl. 67. Die positive Wurzel der Varianz heißt wie im diskreten Fall Standardabweichung. Für den Erwartungswert $E(\mathbf{X})$ gilt weiters für beliebige reelle Zahlen a, b die Identität

$$E(a\mathbf{X} + b) = aE(\mathbf{X}) + b; \quad (73)$$

auch hier folgt der Beweis sofort aus den Eigenschaften des Integrals und Gl. 67.

5.3.3 Normalverteilung revisited

Die wichtigste stetige Verteilung ist die Gaußsche Normalverteilung. Wir haben bereits gesehen, daß sie als Grenzwert der Binomialverteilung für große n auftritt.¹⁸ Die wirkliche Relevanz und Allgemeinheit der Normalverteilung folgt aus dem zentralen Grenzwertsatz (s. Kap. 5.5.3). Da wir bei den in Zusammenhang mit der Normalverteilung auftretenden Integralen an die Grenzen unserer Kenntnisse stoßen, ist dieser Unterabschnitt nicht viel mehr als eine Rekapitulation von Kapitel 4.4 mit Betonung der Konzepte Verteilung und Erwartungswerte (Mittelwert, Varianz).

Die sogenannte $N(0, 1)$ Gaußsche Normalverteilung wurde bereits in Kapitel 4.4 durch

$$N(0, 1)(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x du e^{-\frac{1}{2}u^2} \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (74)$$

eingeführt, wobei $\varphi(x)$ die zu $\Phi(x)$ gehörige Dichte ist. $\Phi(x)$ kann nicht analytisch berechnet werden, wegen seiner großen Wichtigkeit ist es extensiv tabelliert (z.B. "kleiner Bartsch S. 232) bzw. kann heute einfach durch numerische Integration berechnet werden (sogar von vielen Taschenrechnern). Eine wichtige Eigenschaft von $\varphi(x)$ ist die Symmetrie bezüglich des Ursprungs, d.h.

$$\varphi(-x) = \varphi(x)$$

Daraus folgt die für die Benutzung von Tabellen wichtige Beziehung

$$\Phi(-x) = \int_{-\infty}^{-x} du \varphi(u) = \int_x^{+\infty} du \varphi(u) = 1 - \int_{-\infty}^x du \varphi(u) = 1 - \Phi(x), \quad (75)$$

wobei sich die zweite Identität aus der eben erwähnten Symmetrie von $\varphi(x)$ ergibt. Der letzte Schritt folgt aus Gl. 69.

¹⁸Sogar die Poissonverteilung kann durch die Normalverteilung angenähert werden, wenn $\lambda = np$ groß ist.

Da uns der mathematische Apparat fehlt, um Gl. 67 für $N(0, 1)$ explizit zu beweisen, verwenden wir Gl. 67 um folgende Beziehung zu beweisen

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} du e^{-\frac{1}{2}u^2} = 1 \quad \Rightarrow \quad \int_{-\infty}^{+\infty} du e^{-\frac{1}{2}u^2} = \sqrt{2\pi} \quad \Rightarrow \quad \int_{-\infty}^{+\infty} dt e^{-at^2} = \sqrt{\frac{\pi}{a}}, \quad (76)$$

wobei man die letzte Identität durch die Variablensubstitution $u/\sqrt{2} = \sqrt{a}t$ bekommt (Gl. 76 wird z.B. für den Beweis der vollständigen Stirling'schen Näherung für $n!$ benötigt (Abschn. 2.6). Wir geben jetzt ohne weiteren Beweis die Ausdrücke für Mittelwert und Varianz der $N(0, 1)$ Verteilung an.

$$\mu = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dx x e^{-\frac{1}{2}x^2} = 0 \quad (77)$$

$$\sigma^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dx (x - \mu)^2 e^{-\frac{1}{2}x^2} = 1 \quad (78)$$

(Gl. 77 erhält man ohne jegliche Integration aus der Eigenschaft des bestimmten Integrals einer ungeraden Funktion $\int_{-\gamma}^{+\gamma} dx f(x) = 0$ wenn $f(-a) = -f(a)$. Gl. 78 kann man sich selbst durch partielle Integration und Gln. 76, 77 herleiten.)

Beim Übergang von der Binomial- auf die Gaußverteilung sind wir zunächst auf folgende Dichte der allgemeinen Gaußverteilung gekommen

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\mu)^2}{2\sigma^2}}. \quad (79)$$

Die zugehörige Verteilung

$$N(\mu, \sigma^2)(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x du e^{-\frac{(u-\mu)^2}{2\sigma^2}} \quad (80)$$

läßt sich aber durch die Substitution $t = \frac{u-\mu}{\sigma}$ immer auf die Standardnormalverteilung $\Phi(\frac{x-\mu}{\sigma})$ zurückführen. Durch diese Vorschrift genügt die Tabellierung der Standardnormalverteilung $N(0, 1)$ zur Berechnung beliebiger $N(\mu, \sigma^2)$ Verteilungen. Die Variablentransformation entspricht der für den diskreten Fall (S. 33) besprochenen Einführung der Standardisierten der zu (80) gehörigen Zufallsvariablen. Der Vollständigkeit geben wir Mittelwert und Varianz der $N(\mu, \sigma^2)$ -Verteilung an:

$$\mu = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} dx x e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \mu \quad (81)$$

$$\sigma^2 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} dx (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \sigma^2 \quad (82)$$

Diese Eigenschaften zusammen mit Gln. 77, 78 erklären die weitverbreitete Notation $N(\mu, \sigma^2)$ für die allgemeine, und $N(0, 1)$ für die Standardnormalverteilung.

5.4 Zu Wahrscheinlichkeitsverteilungen und Erwartungswerten mehrerer Zufallsvariablen

Wir beschränken uns auf diskrete Zufallsvariablen. Die Verallgemeinerung auf stetige Zufallsvariablen durch Analogieschluß ist unproblematisch (Summation \rightarrow Integral), bringt aber inhaltlich wenig

Neues. Darüberhinaus wurden die mathematischen Grundlagen von Doppel- und Mehrfachintegralen nie besprochen. Des weiteren bringen wir alle Ableitungen für den Spezialfall zweier Zufallsvariablen \mathbf{X} und \mathbf{Y} , die Verallgemeinerung auf n Zufallsvariablen $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ ist immer offensichtlich.

Eine Zufallsvariable ist eine Funktion, die als Definitionsmenge den Ereignisraum eines Zufallsexperiments hat. Es gibt keinen Grund, nicht mehr als eine Zufallsvariable auf demselben Ereignisraum zuzulassen. Die Zufallsvariablen \mathbf{X} und \mathbf{Y} seien zwei derartige Funktionen, die auf einem Ereignisraum definiert sind. (Als Beispiel denken Sie an unsere zwei Würfel: \mathbf{X} sei die Augensumme, \mathbf{Y} das Produkt der Augenziffern.) Die zu \mathbf{X} und \mathbf{Y} gehörigen Wahrscheinlichkeitsverteilungen (Wahrscheinlichkeitsfunktionen) bezeichnen wir als $f(x_i)$ und $g(y_i)$. Die Funktion

$$\mathbf{P}(\mathbf{X} = x_j, \mathbf{Y} = y_k) = p(x_j, y_k) \quad (j, k = 1, 2, \dots) \quad (83)$$

wird dann als gemeinsame (Wahrscheinlichkeits)Verteilung von \mathbf{X} und \mathbf{Y} bezeichnet, sie besitzt folgende Eigenschaften

$$p(x_j, y_k) \geq 0 \quad \sum_{j,k} p(x_j, y_k) = 1. \quad (84)$$

Darüberhinaus findet man für festgehaltenes j bzw. k

$$\sum_k p(x_j, y_k) = \mathbf{P}(\mathbf{X} = x_j) = f(x_j) \quad \sum_j p(x_j, y_k) = \mathbf{P}(\mathbf{Y} = y_k) = g(y_k) \quad (85)$$

wieder $f(x_j)$ und $g(y_k)$, die man auch als Randverteilungen bezeichnet.

Wir kommen jetzt auf unser Konzept der bedingten Wahrscheinlichkeit zurück. Unter Verwendung der Notation aus Gl. 83 ist die bedingte Wahrscheinlichkeit des Ereignis $\mathbf{Y} = y_k$ unter der Voraussetzung $\mathbf{X} = x_i$ durch

$$\mathbf{P}(\mathbf{Y} = y_k | \mathbf{X} = x_i) = \frac{p(x_i, y_k)}{f(x_i)} \quad (86)$$

gegeben (vorausgesetzt $f(x_i) > 0!$). Gilt jedoch $p(x_j, y_k) = f(x_j)g(y_k)$ für alle x_j, y_k , dann sind die Ereignisse $\mathbf{X} = x_j$ und $\mathbf{Y} = y_k$ stochastisch (statistisch) unabhängig, \mathbf{X} und \mathbf{Y} werden dann als unabhängige Zufallsvariablen bezeichnet. Wir reformulieren also das Konzept der statistischen Unabhängigkeit in der Sprache von Zufallsvariablen

$$\mathbf{P}(\mathbf{X} = x_i, \mathbf{Y} = y_j, \dots) = \mathbf{P}(\mathbf{X} = x_i) \cdot \mathbf{P}(\mathbf{Y} = y_j) \cdot \dots \quad (87)$$

Was können wir nun über die Erwartungswerte für zwei (oder mehr) auf demselben Ereignisraum definierten Zufallsvariablen sagen? Uns interessiert zunächst $E(\mathbf{X}) + E(\mathbf{Y})$, wobei \mathbf{X} und \mathbf{Y} die gemeinsame Verteilung $p(x_i, y_j)$ haben. Es gilt

$$E(\mathbf{X}) + E(\mathbf{Y}) = \sum_{i,j} x_i p(x_i, y_j) + \sum_{i,j} y_j p(x_i, y_j) = \sum_{i,j} (x_i + y_j) p(x_i, y_j) = E(\mathbf{X} + \mathbf{Y})$$

Um den ersten Schritt zu rechtfertigen bzw. mit der gewöhnlichen Definition von $E(\mathbf{X})$ in Einklang zu bringen, betrachten wir den Terme $\sum_{i,j} x_i p(x_i, y_j)$ genauer. Nach Gl. 85 gilt nämlich

$$\sum_{i,j} x_i p(x_i, y_j) = \sum_i x_i \sum_j p(x_i, y_j) = \sum_i x_i f(x_i),$$

und analoges gilt für die Summe $\sum_{i,j} y_j p(x_i, y_j) = \sum_j y_j g(y_j)$. Für n Zufallsvariablen gilt ganz allgemein

$$E(\mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_n) = E(\mathbf{X}_1) + E(\mathbf{X}_2) + \dots + E(\mathbf{X}_n) \quad (88)$$

Für das Produkt $E(\mathbf{X}\mathbf{Y})$ gibt es keine allgemeine Entsprechung zu Gl. 88, d.h., im allgemeinen gilt $E(\mathbf{X}\mathbf{Y}) \neq E(\mathbf{X})E(\mathbf{Y})$. Sind \mathbf{X} und \mathbf{Y} jedoch stochastisch unabhängig, so findet man

$$E(\mathbf{X}\mathbf{Y}) = \sum_{i,j} x_i y_j p(x_i, y_j) = \sum_{i,j} x_i y_j f(x_i) g(y_j) = \sum_i x_i f(x_i) \sum_j y_j g(y_j) = E(\mathbf{X})E(\mathbf{Y}) \quad (89)$$

mit entsprechender Verallgemeinerung für n Zufallsvariablen.

Es bleibt uns noch Gesetzmäßigkeiten für die Varianz $\text{Var}(\mathbf{X} + \mathbf{Y})$ zweier auf demselben Ereignisraum definierter Zufallsvariablen zu untersuchen. Gesucht ist also $\text{Var}(\mathbf{X} + \mathbf{Y}) = E((\mathbf{X} + \mathbf{Y} - (\mu_x + \mu_y))^2)$. Wir berechnen zunächst

$$((\mathbf{X} + \mathbf{Y}) - (\mu_x + \mu_y))^2 = ((\mathbf{X} - \mu_x) + (\mathbf{Y} - \mu_y))^2 = (\mathbf{X} - \mu_x)^2 + (\mathbf{Y} - \mu_y)^2 + 2(\mathbf{X} - \mu_x)(\mathbf{Y} - \mu_y).$$

Nimmt man von dieser Größe den Erwartungswert, so bekommt man (unter impliziter Verwendung von Gl. 88)

$$E(((\mathbf{X} + \mathbf{Y}) - (\mu_x + \mu_y))^2) = E((\mathbf{X} - \mu_x)^2) + E((\mathbf{Y} - \mu_y)^2) + 2E((\mathbf{X} - \mu_x)(\mathbf{Y} - \mu_y))$$

Die ersten beiden Erwartungswerte auf der rechten Seite sind aber $\text{Var}(\mathbf{X})$ und $\text{Var}(\mathbf{Y})$. Für den letzten Erwartungswert führt man die Bezeichnung *Kovarianz* $\text{Cov}(\mathbf{X}, \mathbf{Y})$ ein. Es gilt

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = E((\mathbf{X} - \mu_x)(\mathbf{Y} - \mu_y)) = E(\mathbf{X}\mathbf{Y}) - \mu_x E(\mathbf{Y}) - \mu_y E(\mathbf{X}) + \mu_x \mu_y = E(\mathbf{X}\mathbf{Y}) - \mu_x \mu_y \quad (90)$$

Sind \mathbf{X} und \mathbf{Y} stochastisch unabhängig, d.h., gilt $E(\mathbf{X}\mathbf{Y}) = E(\mathbf{X})E(\mathbf{Y})$, dann verschwindet die Kovarianz und es gilt

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = 0 \quad (91)$$

Achtung: Die Umkehrung gilt nicht, nur weil die Kovarianz verschwindet müssen die entsprechenden Zufallsvariablen *nicht* unabhängig sein. Dasselbe Problem existiert für den mittels der Kovarianz definierten *Korrelationskoeffizienten*

$$\rho(\mathbf{X}, \mathbf{Y}) = \frac{\text{Cov}(\mathbf{X}, \mathbf{Y})}{\sigma_x \sigma_y}. \quad (92)$$

Sind \mathbf{X} und \mathbf{Y} stochastisch unabhängig, so verschwindet ρ , hingegen bedeutet $\rho = 0$ nicht zwingend die stochastische Unabhängigkeit von \mathbf{X} und \mathbf{Y} .

Nach diesem Exkurs über die Kovarianz können wir jetzt für $\text{Var}(\mathbf{X} + \mathbf{Y})$ kompakt schreiben

$$\text{Var}(\mathbf{X} + \mathbf{Y}) = \sigma_x^2 + \sigma_y^2 + 2\text{Cov}(\mathbf{X}, \mathbf{Y}) \quad (93)$$

bzw. im Falle der stochastischen Unabhängigkeit von \mathbf{X} und \mathbf{Y}

$$\text{Var}(\mathbf{X} + \mathbf{Y}) = \sigma_x^2 + \sigma_y^2 \quad (94)$$

Sind n Zufallsvariablen $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ stochastisch unabhängig, so gilt für die Varianz ihrer Summe $\mathbf{S}_n = \sum_i \mathbf{X}_i$

$$\text{Var}(\mathbf{S}_n) = \sigma_{x_1}^2 + \sigma_{x_2}^2 + \dots = \sum_i \sigma_{x_i}^2. \quad (95)$$

Eigenschaften 88 und 95 werden noch von großer Wichtigkeit sein, wenn die Meßsituation in der Sprache der Wahrscheinlichkeitsrechnung ausgedrückt werden wird. (Jede Messung unterliegt dem sogenannten statistischen Fehler, ist also in dieser Hinsicht die Realisierung eines Zufallsexperiment, und somit ist das Meßergebnis eine Zufallsvariable. Weiters nimmt man an, daß eine Messung von denen davor und danach stochastisch unabhängig ist.)

► **Beispiel:** Die Berechnung von μ und σ^2 der Binomialverteilung (S. 34) war einigermaßen mühsam. Unter Verwendung von (88) und (95) können wir diese Größen nun viel einfacher ableiten. Die Zufallsvariable der Binomialverteilung \mathbf{S}_n ist die Summe der Zufallsvariablen \mathbf{X}_i der Einzelexperimente des Bernoullischen. Kennen wir aber $E(\mathbf{X}_i)$ und $\text{Var}(\mathbf{X}_i)$, so erhält man gemäß (88) und (95) $E(\mathbf{S}_n) = n E(\mathbf{X}_i)$ und $\text{Var}(\mathbf{S}_n) = n \text{Var}(\mathbf{X}_i)$.

Die Zufallsvariable eines Einzelexperiments \mathbf{X}_i kann nur die Werte 0 und 1 annehmen, und $f(0) = q = 1 - p$, $f(1) = p$. Somit ist

$$E(\mathbf{X}_i) = 0 \times (1 - p) + 1 \times p = p,$$

$$\text{Var}(\mathbf{X}_i) = E(\mathbf{X}_i^2) - E(\mathbf{X}_i)^2 = [0^2 \times (1 - p) + 1^2 \times p] - p^2 = p(1 - p)$$

woraus unmittelbar das schon auf S. 34 erhaltene Ergebnis

$$E(\mathbf{S}_n) = np, \quad \text{Var}(\mathbf{S}_n) = np(1 - p)$$

folgt. ◀

5.5 Gesetzmäßigkeiten für große Zahlen

In diesem Abschlußkapitel der Wahrscheinlichkeitsrechnung beschäftigen wir uns mit Eigenschaften der Verteilung von Zufallsvariablen \mathbf{S}_n , die die *Summe* von n (stochastisch unabhängigen (auf dem gleichen Ereignisraum definierten)) Zufallsvariablen \mathbf{X}_i sind,

$$\mathbf{S}_n = \sum_{i=1}^n \mathbf{X}_i,$$

wenn die Anzahl der Variablen groß wird, oder, mathematisch formuliert, wenn $n \rightarrow \infty$. Manchmal ist an Stelle von \mathbf{S}_n auch die davon abgeleitete Größe \mathbf{S}_n/n von Interesse. Laut Voraussetzung gelten für den Erwartungswert und die Varianz von \mathbf{S}_n Gln. 88 und 95.

Qualitativ lassen sich die Aussagen folgendermaßen formulieren: (i) Aus je mehr Summanden \mathbf{X}_i sich die Zufallsvariable \mathbf{S}_n zusammensetzt, desto unwichtiger werden die Abweichungen vom Erwartungswert $E(\mathbf{S}_n/n)$. Zumindestens indirekt kann man aus diesem Ergebnis erkennen, warum statistische Betrachtungen für das Studium makroskopischer Größen als Funktion mikroskopischer Eigenschaften so wichtig und zielführend sind (dies ist das Arbeitsgebiet der Statistischen Mechanik). Makroskopische Größen können (oft) als Erwartungswerte formuliert werden, wobei über alle Teilchen

(Atome, Moleküle) gemittelt (summiert) wird. Da die Anzahl der Teilchen in der Größenordnung von N_L (der Loschmidtzahl) liegt, ist die Voraussetzung großer n "hinreichend" erfüllt.

(ii) Der sogenannte *zentrale Grenzwertsatz* steigert die erste Aussage insofern, daß er nicht nur den Erwartungswert von \mathbf{S}_n , sondern die Verteilungsfunktion von \mathbf{S}_n selbst beschreibt. *Fast alle Stichprobenverteilungen streben beim Vergrößern ihres Umfangs einer Normalverteilung zu.* Dieser Satz ist unter anderem die wahrscheinlichkeitstheoretische Grundlage von vielen statistischen Anwendungen. Etwas exakter formuliert lautet der zentrale Grenzwertsatz: Eine Summe von sehr vielen unabhängigen Zufallsvariablen ($\mathbf{S}_n = \sum_{i=1}^n \mathbf{X}_i$, n groß) hat eine nahezu normale Verteilung $N(\mu, \sigma^2)$, wenn die Varianzen der einzelnen Zufallsvariablen gegenüber der Varianz ihrer Summe klein sind (die einzelnen \mathbf{X}_i müssen keineswegs normalverteilt sein, noch die gleiche Verteilung besitzen!).

Der Rest dieses Abschnitts versucht eine zumindest exakte mathematische Formulierung dieser Aussagen (wenngleich oft ohne Beweis), er ist optional und als Vertiefung für Interessierte gedacht und daher nicht Prüfungstoff

5.5.1 Tschebyscheffsche Ungleichung

Ist die Verteilung bzw. die Verteilungsfunktion einer Zufallsvariablen \mathbf{X} bekannt, so läßt sich die Wahrscheinlichkeit $\mathbf{P}(|\mathbf{X} - \mu| \geq a)$ exakt berechnen. Häufig kennt man jedoch die Verteilungsfunktion von \mathbf{X} nicht, wohl aber μ und σ^2 . Die Varianz ist wie immer (und unabhängig von der Wahrscheinlichkeitsverteilung) durch $\sigma^2 = \sum (x_i - \mu)^2 f(x_i)$ gegeben. Summieren wir jetzt nur über x_i für die $|x_i - \mu| \geq a$, so gilt sicher folgende Ungleichung

$$\sigma^2 \geq \sum_{|x_i - \mu| \geq a} (x_i - \mu)^2 f(x_i)$$

Für jeden einzelnen Summanden auf der rechten Seite gilt weiters

$$(x_i - \mu)^2 f(x_i) \geq a^2 f(x_i)$$

wodurch umso mehr gelten muß

$$\sigma^2 \geq a^2 \sum_{|x_i - \mu| \geq a} f(x_i).$$

Division durch a^2 führt auf

$$\sum_{|x_i - \mu| \geq a} f(x_i) = \mathbf{P}(|\mathbf{X} - \mu| \geq a) \leq \frac{\sigma^2}{a^2}. \quad (96)$$

Gleichung 96 ist als Tschebyscheffsche Ungleichung bekannt. Für $a \leq \sigma$ ist (96) inhaltslos, da die rechte Seite dann ≥ 1 wird. Interessant wird (96) für $a = k\sigma$, $k > 1$, denn dann erhält man

$$\mathbf{P}(|\mathbf{X} - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

5.5.2 Schwaches Gesetz der großen Zahlen

Ein Zufallsexperiment werde n -mal unter denselben Bedingungen durchgeführt. Bei jeder Wiederholung nimmt die Zufallsvariable \mathbf{X} einen Wert aus ihrer Wertemenge an. Wir bezeichnen die so erhaltenen Werte als $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ (wobei durchaus auch gleiche Werte auftreten können). Der Wert \hat{x}_i ist also die Realisierung von

\mathbf{X} bei der i -ten Versuchsdurchführung. Wir fassen jetzt die n -maliger Durchführung der Einzelexperimente zu einem neuen Zufallsexperiment mit Zufallsvariable $\mathbf{Z}_n = 1/n \sum_i \mathbf{X}_i$ zusammen (es handelt sich also um ein verallgemeinertes Bernoulli-Schema, bei dem wir die Bedingungen des Einzelexperiments offenlassen). Die \hat{x}_i sind daher Realisierungen von stochastisch unabhängigen Zufallsvariablen \mathbf{X}_i . Aufgrund der Konstruktion stimmen Verteilungsfunktion, Erwartungswerte und Varianzen der Zufallsvariablen \mathbf{X}_i und \mathbf{X} überein, es sei $E(\mathbf{X}_i) = \mu$ und $\text{Var}(\mathbf{X}_i) = \sigma^2$.

Das arithmetische Mittel

$$\bar{\hat{x}} = \frac{\sum_i \hat{x}_i}{n}$$

ist dann weiters die Realisierung der Zufallsvariablen \mathbf{Z}_n . Es gilt

$$E(\mathbf{Z}_n) = \frac{1}{n} \sum_i E(\mathbf{X}_i) = \frac{n\mu}{n} = \mu$$

und weiters wegen der stochastischen Unabhängigkeit der \mathbf{X}_i auch

$$\text{Var}(\mathbf{Z}_n) = \frac{1}{n^2} \sum_i \text{Var}(\mathbf{X}_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

(Wegen des ersten Schritts s. Gl. 61.)

Nach Gl. 96 gilt aber jetzt für jedes $\epsilon > 0$

$$\mathbf{P}(|\mathbf{Z}_n - \mu| \geq \epsilon) \leq \frac{\text{Var}(\mathbf{Z}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}. \quad (97)$$

Die rechte Seite wird aber jetzt für ein beliebiges (fixes!) $\epsilon > 0$ beliebig klein, wenn nur n groß genug ist. Die Wahrscheinlichkeit, daß die Zufallsvariable $\mathbf{Z}_n = \frac{1}{n} \sum_i \mathbf{X}_i$ Werte annimmt, die von μ mehr als ϵ abweichen, ist somit für große n sehr klein. Diese Tatsache (die man noch als $\lim_{n \rightarrow \infty}$ formulieren könnte), wird als *schwaches Gesetz der großen Zahlen* bezeichnet.

Gleichung 97 impliziert weiters, daß $\bar{\hat{x}}$ meistens in der Nähe des Erwartungswerts μ liegt. Diese Eigenschaft ermöglicht es, Näherungswerte für μ mit Hilfe von Zufallsexperimenten (darunter fällt auch Stichproben ziehen, Messungen durchführen usw.) zu gewinnen.

5.5.3 Zentraler Grenzwertsatz

Einziges Ziel dieses Unterabschnitts, ist den zentralen Grenzwertsatz mathematisch zu formulieren, ein Beweis liegt außerhalb unserer Möglichkeiten. Wir betrachten wieder n Zufallsvariablen $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ die stochastisch unabhängig sind, und jedes \mathbf{X}_i besitze Mittelwert μ_i und Varianz σ_i (die nicht gleich zu sein haben!). Gemäß Abschnitt 5.4 gilt dann für Summenvariable $\mathbf{S}_n = \sum_i \mathbf{X}_i$

$$E(\mathbf{S}_n) = \sum_i \mu_i$$

$$\text{Var}(\mathbf{S}_n) = \sum_i \sigma_i^2.$$

Wir erinnern uns an die Standardisierte einer Zufallsvariablen (s. S. 33) $\mathbf{X}^* = (\mathbf{X} - E(\mathbf{X}))/\sqrt{\text{Var}(\mathbf{X})}$. Der zentrale Grenzwertsatz besagt jetzt, daß die Standardisierte \mathbf{S}_n^* der Zufallsvariablen \mathbf{S}_n ,

$$\mathbf{S}_n^* = \frac{\sum_i (\mathbf{X}_i - \mu_i)}{\sqrt{\sum_i \sigma_i^2}}$$

für große n unter extrem schwachen Voraussetzungen (die in der Praxis typischerweise keinerlei Einschränkung bedeuten) ungefähr $N(0, 1)$ -verteilt ist. Mathematisch läßt sich das in der Form

$$\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{S}_n^* \leq x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x du e^{-u^2/2} \quad (98)$$

ausdrücken.

5.6 Testverteilungen

In der mathematischen Statistik spielen drei weitere stetige Wahrscheinlichkeitsverteilungen eine große Rolle, die für bestimmte Funktionen von Zufallsvariablen \mathbf{X} (bzw. Funktionen mehrerer Zufallsvariablen $\mathbf{X}_1, \mathbf{X}_2, \dots$) auftreten, wobei die \mathbf{X}_i selbst $N(0, 1)$ bzw. Chi-Quadrat-verteilt (s.u.) sind. Diese Verteilungen bzw. ihre Dichten sind mathematisch ziemlich kompliziert, sodaß wir hier nur der Vollständigkeit halber die Namen erwähnen: Es handelt sich um die *Chi-Quadrat-Verteilung*, die *Studentsche t-Verteilung*, und die *F-Verteilung* von Fisher. Alle drei sind ähnlich wie die $N(0, 1)$ Verteilung ausführlich tabelliert. Ein wichtiger Unterschied zu $N(0, 1)$ Tabellen besteht darin, daß in diese Verteilungen noch die Anzahl der sogenannten *Freiheitsgrade* eingeht. Für eine typische Anwendung der t-Verteilung ist die Anzahl der Freiheitsgrade z.B. die Anzahl der Stichproben oder Meßergebnisse *minus* 1 (s. Abschnitt 6.3.2).

6 Die statistische Behandlung experimenteller Daten

6.1 Einleitung

Wir verlassen jetzt endgültig die Wahrscheinlichkeitsrechnung und wenden uns der Statistik, oder genauer der statistischen Behandlung von Meßdaten, illustriert an ein paar Standardsituationen, zu. Dieser letzte Abschnitt ist folgendermaßen organisiert. In diesem Unterabschnitt (Abschn. 6.1) folgen ein paar Betrachtungen über Arten von Fehlern, und eine Präzisierung des Begriffs "Genauigkeit". Abschnitt 6.2 beginnt mit der Einführung empirischen Kenngrößen zur Beschreibung eines Datensatzes, wie Mittelwert und Ausmaß der Streuung der Daten um diesen Mittelwert (empirische Varianz). Danach folgen zwei wichtige Verfahren, die nicht unmittelbar mit der Wahrscheinlichkeitsrechnung verknüpft sind, Fehlerfortpflanzung und lineare Regression. Abschnitt 6.3 nimmt schließlich den Faden der Wahrscheinlichkeitstheorie wieder auf, und versucht die Prinzipien zu erklären, auf denen einige statistischen Methoden beruht, wie z.B. die Berechnung von Vertrauensbereichen oder statistische Tests.

Sobald man Messungen durchführt, ist man vor Fehlern nicht gefeit. Für die Behandlung der aus Messungen erhaltenen Daten, die möglicherweise nicht fehlerfrei sind, ist es wichtig, sich über die Arten von Fehlern im klaren zu sein. Für die erste Art von Fehlern gibt es keine einheitliche Bezeichnung. Manche Lehrbücher sprechen von "verbotenen" Fehlern, vielleicht wären auch "trivialer Fehler" oder "dummer Fehler" zutreffende Bezeichnungen. Gemeint ist folgendes: Wenn man eine neue Methode zum erstenmal verwendet, hat man oft die Arbeitsanleitung nicht völlig verstanden

und produziert "Hausnummern". Manchmal mißt man zwar richtig, macht aber beim Einsetzen in Gleichungen, mit denen man aus den Daten die gewünschte Größe extrahiert, Unsinn (z.B. Eintippen von Grad in den Taschenrechner und Berechnen der Winkelfunktion, wenn der Taschenrechner auf Radiant eingestellt ist, oder z.B. Vermengen von Einheiten). Im Normalfall merkt man derartige Fehler rasch und sie sind leicht zu korrigieren.

Systematische Fehler Etwas schwieriger wird es bei den sogenannten *systematischen* Fehlern. Diese treten auf, wenn man im Prinzip alles richtig macht, nur z.B. vergißt, einen Parameter, der auf die Meßdaten Einfluß hat zu berücksichtigen. Beispiel: Sie messen die Abhängigkeit der Reaktionsgeschwindigkeit von der Konzentration eines Reaktanden. Bei Vergleich mit Literaturwerten für ähnliche Systeme stellen Sie fest, daß Ihre Ergebnisse zwar qualitativ gut übereinstimmen, die quantitative Übereinstimmung jedoch schlecht ist. Schließlich lesen Sie den "Kleindruck": Die Literaturdaten wurden bei 25° C gemessen, Sie mußten aber (aus welchen Gründen auch immer) bei 15° C arbeiten. Wenn Sie wissen, wie sich die betreffende Reaktionsgeschwindigkeit als Funktion der Temperatur ändert, können Sie ggf. Ihre Daten korrigieren, und somit mit den Referenzdaten vergleichen. Wichtig ist (daher die Bezeichnung), daß ein derartiger Fehler unter Ihrer Kontrolle ist: Sie können ihn entweder vermeiden oder Sie können die Auswirkung dieses Fehlers berechnen und dafür korrigieren. Systematische Fehler verfälschen ein Resultat in eine bestimmte Richtung ("zu hoch", "zu niedrig"), das unterscheidet Sie von den oben erwähnten "dummen Fehlern"¹⁹ und den sogenannten zufälligen Fehlern.

Zufällige (statistische) Fehler Sie machen "alles richtig." Die systematischen Fehlerquellen haben Sie ausgeschaltet bzw. Sie wissen, wie Sie dafür korrigieren können. Sie wiederholen Ihr Experiment, und jede Wiederholung gibt Ihnen ein leicht anderes Ergebnis (wenn Sie z.B. jemals versucht haben, so genau als möglich zu titrieren, dann wird Ihnen das beschriebene Phänomen nicht unbekannt sein). Vorausgesetzt, Sie sitzen wirklich keinem "dummen" oder versteckten systematischen Fehler auf, dann haben Sie soeben Bekanntschaft mit dem *zufälligen* (oder auch *statistischen*) Fehler gemacht. In jeden Meßablauf schleichen sich Fluktuationen ein, die *nicht* kontrollierbar sind. Wenn Sie eine Konzentration naßchemisch bestimmen, müssen Sie in der Regel ein- oder mehrmals pipettieren und titrieren. Da Sie kein Roboter sind, und Konzentrationen immer nur bis zu einer gewissen Genauigkeit ablesen können, werden Sie beim Wiederholen der Prozedur auf leicht andere Ergebnisse kommen. Messen Sie die Konzentration photometrisch, dann sollte im Idealfall die Probe thermostatisiert sein. Was die Güte und Effizienz des Thermostaten betrifft, müssen Sie sich aber auf das Gerät verlassen, und Schwankungen in der Thermostatisierung können Ihr Ergebnis beeinflussen. Diese zufälligen Fehler gilt es durch Statistik zu charakterisieren und zu bestimmen. Wie Sie sehen werden können Sie in letzter Konsequenz den zufälligen Fehler nur durch mehr Messungen verkleinern.

"Genauigkeit" Was heißt jetzt in Hinblick auf das eben Gesagte, so "genau als möglich" zu messen? Wann ist eine Messung "genauer" als eine andere. Das Wort Genauigkeit ist in Hinblick auf die eben erfolgte Diskussion von systematischem und zufälligem Fehler doppeldeutig. Das Englische

¹⁹Ein wirklich "dummer" Fehler in Zusammenhang mit dem Reaktionskinetikbeispiel wäre, überhaupt nicht auf die Temperatur zu achten, dann können Sie auch nicht mehr nachträglich korrigieren und Ihre Messung ist tatsächlich wertlos.

kennt zwei Worte für Genauigkeit: *accuracy* und *precision*. Im normalen (englischen) Sprachgebrauch sind die beiden Wörter völlig synonym, womit dasselbe Problem wie im Deutschen existiert. Die Statistiker haben jedoch die Gelegenheit ausgenutzt, und verwenden *accuracy* in Zusammenhang mit dem systematischen Fehler, und *precision* in Zusammenhang mit dem zufälligen Fehler. *Accuracy* bezeichnet den Grad der Abweichung vom “wahren Wert” (wobei man natürlich in den seltensten Fällen den wahren Wert kennt). Etwas holprig könnte man also im Deutschen das Wort Korrektheit oder Richtigkeit verwenden. *Precision* hat mit der Größe des statistischen Fehlers zu tun, also wie sehr die einzelnen Meßergebnisse streuen. Wenn zwei Studenten die Konzentration der gleichen Probe bestimmen, und einer bekommt aus fünf Meßungen Konzentrationen, die zwischen 0,13 und 0,14 mol/l liegen, der zweite aus fünf Meßungen Konzentrationen, die zwischen 0,12 und 0,15 mol/l liegen, so hat der erste “more precisely” gemessen. Nichtsdestotrotz könnten beide falsch (“inaccurately”) gemessen haben, dann nämlich wenn die Konzentration der vom Saalassistenten ausgegebenen Probe bei 0,11 mol/l liegt! Eine Messung sollte natürlich immer *accurate* und *precise* sein. Eine *precise* Messung kann (durch das Auftreten eines systematischen Fehlers) falsch sein und ist damit wertlos. Zu niedriger *precision* wiederum kann die Aussagekraft einer *accurate* Messung sehr erniedrigen.

Statistik kann nichts gegen systematische Fehler machen, die Vermeidung (bzw. Korrektur) von systematischen Fehlern fordert das Fachwissen und Geschick des jeweiligen Experten auf seinem Gebiet. Hingegen ist die Charakterisierung des zufälligen Fehlers Aufgabe der Statistik, und wir kommen darauf in Abschnitt 6.3 zurück.

6.2 Empirische Behandlung und Methoden, die nicht direkt mit Wahrscheinlichkeitstheorie verknüpft sind

Wie schon in der Einleitung angesprochen, ist dieses Kapitel in gewisser Weise ein Exkurs. Zuerst wenden wir unseren Hausverstand an, um Meßergebnisse kompakt zusammenzufassen und zu beschreiben (die tiefere Bedeutung dieser Größen und ihre Verknüpfung mit der Wahrscheinlichkeitsrechnung wird in Kapitel 6.3.1 klar werden). Danach stellen wir zwei Methoden zur Datenbehandlung vor, deren Ableitung zwar einige Mathematikkenntnisse, jedoch keine wahrscheinlichkeitstheoretischen Ergebnisse voraussetzt.

6.2.1 Empirischer Mittelwert, empirische Varianz und Kovarianz

Sie haben eine Messung fünfmal durchgeführt. Welcher der Werte x_i ($i = 1, \dots, 5$) ist der wahre Wert, dem wahren Wert am nächsten? Wie bringen Sie Ihr Ergebnis in eine kompakte Form (besonders interessant, wenn Sie statt fünf 50 Meßergebnisse haben)?

Da es keinen Grund gibt, ein Meßergebnis zu bevorzugen,²⁰ ist ganz sicher das arithmetische Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (99)$$

eine gute Wahl, um Ihre Messungen in einem Wert zusammenzufassen. Man bezeichnet Gl. 99 als

²⁰Wenn es einen solchen gibt, dann sollten Sie Ihren Meßvorgang *sehr* kritisch überdenken!

(empirischen) Mittelwert.²¹

Bei der Behandlung von Meßdaten ist \bar{x} die ideale Wahl, um die einzelnen Meßergebnisse zu repräsentieren. In anderen Anwendungen können aber alternative, sogenannte *Lageparameter* die Stichprobe besser beschreiben: Für den *Median* (*Zentralwert*) einer Stichprobe ordnet man die x_i nach Größe. Für ungerade n ist der Median \tilde{x} genau der Wert x_i , für den gilt, daß genausoviele Werte kleiner (gleich) und grösser (gleich) sind als \tilde{x} , d.h. $\tilde{x} = x_i$ mit $i = \frac{n+1}{2}$. (Wenn Ihre nach Größe geordneten Meßwerte 1.1, 1.3, 1.4, 1.8, 1.9 mmol/l seien, dann ist der Median $\tilde{x} = 1.4$ mmol/l, zum Vergleich der Mittelwert ist $\bar{x} = 1.5$ mmol/l.) Für gerade n gilt $\tilde{x} = \frac{x_{n/2} + x_{n/2+1}}{2}$.

Hauptsächlich für diskrete Stichproben, die durch ganze Zahlen charakterisiert sind, verwendet man auch den *Modalwert* (*Modus*, *Mode*) der Stichprobe. Der Modalwert ist derjenige Wert der Stichprobe, der am häufigsten vorkommt. (Für eine Stichprobe mit Werten 1, 2, 2, 2, 3, 3, 4 ist 2 der Modalwert.)

Als nächste Kenngröße ist die Abweichung der einzelnen Meßwerte x_i vom Mittelwert \bar{x} von Interesse, dieses (noch festzulegende Maß der Streuung innerhalb der Meßwerte gibt Aufschluß über die *precision* der Daten.) Da immer gilt

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) = \frac{1}{n} \left(\sum_{i=1}^n x_i - n\bar{x} \right) = \bar{x} - \frac{1}{n} n\bar{x} = 0$$

brauchen wir einen besseren Ansatz. Eine Möglichkeit ist an Stelle der Abweichungen $(x_i - \bar{x})$ über den Betrag der Abweichungen $|x_i - \bar{x}|$ zu summieren. Ebenso möglich ist Summation über die Abweichungsquadrate $(x_i - \bar{x})^2$, und dies ist die in der Praxis (der Meßdatenanalyse) bevorzugte Vorgehensweise. Konkret verwendet man die *empirische Varianz* s^2

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right] = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]. \quad (100)$$

Der erste Ausdruck ist die Definition, die Ausdrücke weiter rechts (Umformungen bitte selbst durchführen) dienen der rechnerischen Effizienz (sofern man heute noch in die unglückliche Lage kommt, derartige Größen mit der Hand oder dem Taschenrechner ausrechnen zu müssen). Die positive Quadratwurzel s heißt *empirische Standardabweichung*. Die Verwendung des Faktors $1/(n-1)$ an Stelle des vielleicht erwarteten $1/n$ erklärt sich daraus, daß die n x_i und \bar{x} wegen (99) ja nicht unabhängig sind, und daher hat man nur $n-1$ sogenannte Freiheitsgrade. Weitere theoretische Vorzüge des Faktors $1/(n-1)$ werden in Abschn. 6.3.1 besprochen.

Die Analogie zwischen s^2 und $\text{Var}(\mathbf{X})$ ist nicht zu übersehen (auf den tatsächlichen Zusammenhang wird nochmals auf Abschn. 6.3.1 verwiesen. In Abschnitt 5.4 wurde für zwei Zufallsvariablen die Kovarianz eingeführt. Hat man es in einer Messung mit Wertepaaren zu tun (z.B. Konzentration und Temperatur als Funktion der Zeit), so ist es naheliegend für Meßdaten der Form (x_i, y_i) mit Mittelwerten \bar{x} und \bar{y} die *empirische Kovarianz*

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (101)$$

²¹Das Adjektiv empirisch dient dazu, um \bar{x} formal von $\mu = E(\mathbf{X})$, dem Erwartungswert einer Zufallsvariablen, abzugrenzen, kann aber entfallen, wenn keine Verwechslungsgefahr besteht. Zum Zusammenhang zwischen \bar{x} und μ siehe Abschnitt 6.3.1.

einzuführen. Achtung, wir schreiben im Gegensatz zu den Varianzen s_x^2, s_y^2 nur s_{xy} (nicht s_{xy}^2 , kein Tippfehler!).²² Weiters sei noch der *empirische Korrelationskoeffizient*

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (102)$$

erwähnt.

6.2.2 Propagierung von Unsicherheiten / Standardabweichungen (Fehlerfortplanzung)

Folgende Situation tritt im Laboralltag vermutlich täglich auf. Die durch Mittelwert \bar{x} und Standardabweichung s_x beschriebenen Meßergebnisse, werden in eine Formel eingesetzt, aus der die eigentlich interessierende Größe $w = w(x)$ berechnet wird. Was sind \bar{w} und s_w . Knifflig wird das Problem, wenn w eine Funktion von mehr als einer Meßgröße ist, d.h. $w = w(x, y, \dots)$. Ein ganz simples Beispiel ist das Volumen V eines Quaders, wenn wir einmal annehmen, daß es für jede der drei Kantenlängen a, b, c eine Standardabweichung s_a, s_b, s_c gibt.²³ Eine Möglichkeit, die Standardabweichungen (Fehler) zu propagieren, besteht darin, die Extremfälle zu betrachten, d.h. $w_{min} = w(x_{min}, y_{min}, \dots)$ und $w_{max} = w(x_{max}, y_{max}, \dots)$ zu berechnen, wobei die Subskripte min und max die jeweils kleinsten bzw. größten Werte von $x, y \dots$ und in Folge w bezeichnen. Die Differenz $w_{max} - w_{min}$ ist ein Maß für die Unsicherheit. Diese Vorgehensweise ist legitim, aber in vielen Fällen zu pessimistisch, denn es kann zu Fehlerkompensation kommen.

In vielen Fällen kann man mit Taylorentwicklungen (die man beim linearen Glied (Glieder 1. Ordnung) abbricht) zu einer vernünftigen Abschätzung kommen. Wir illustrieren die Methode zunächst im eindimensionalen Fall und betrachten $\bar{w} = w(\bar{x})$ sowie $w_i = w(x_i)$, die Werte der interessierenden Größe w als Funktion des Mittelwerts x bzw. als Funktion eines Einzelmeßwerts x_i . Wir entwickeln jetzt w_i in eine Taylorreihe erster Ordnung um den Entwicklungspunkt \bar{x} :

$$w_i = w(x_i) = w(\bar{x}) + \left(\frac{dw}{dx} \right)_{\bar{x}} (x_i - \bar{x}) + \dots \quad (A)$$

und brechen (wie gezeigt) die Reihe nach dem linearen Term (wie gezeigt) ab. Uns interessiert die Abschätzung für $s_w^2 = 1/(n-1) \sum (w_i - \bar{w})^2$. Aus (A) folgt

$$w_i - \bar{w} \approx \left(\frac{dw}{dx} \right)_{\bar{x}} (x_i - \bar{x}) \quad (B)$$

und somit ist

$$s_w^2 = \frac{1}{n-1} \sum_{i=1}^n (w_i - \bar{w})^2 \approx \frac{1}{n-1} \sum_{i=1}^n \left[\left(\frac{dw}{dx} \right)_{\bar{x}} (x_i - \bar{x}) \right]^2 =$$

²²Manchmal ist allerdings auch die Schreibweise s_{xy}^2 in Verwendung...

²³Für so etwas simples wie eine Länge mag eine Standardabweichung ungewöhnlich erscheinen. Jedoch erinnern Sie sich, daß die Standardabweichung ein Maß für die *precision* ist. Sie können z.B. mit einem Maßband nicht viel genauer als $\pm 0,5$ mm ablesen, d.h. diese 0,5 mm sind eine realistische Abschätzung der *precision* ihrer Längenmessung. Alternativ denken Sie an das Volumen von quaderförmigen Schachteln, die fabriksmäßig erzeugt werden. Durch Fluktuationen im Fabrikationsprozeß wird nicht jede Schachtel exakt gleich sein, und somit kommen Sie ebenfalls auf Standardabweichungen in den Schachtellängen.

$$= \left(\frac{dw}{dx} \right)_{\bar{x}}^2 \underbrace{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}_{s_x^2} = s_x^2 \left(\frac{dw}{dx} \right)_{\bar{x}}^2. \quad (\text{C})$$

Als nächstes wiederholen wir die Schritte (A) bis (C) für den Fall, daß w eine Funktion von zwei Variablen x und y ist, die aus Messungen ermittelt werden (die weitere Verallgemeinerung ist dann hoffentlich trivial). In diesem Fall entwickeln wir $w_i = w(x_i, y_i)$ in eine Taylorreihe um den Punkt (\bar{x}, \bar{y}) .

$$w_i = w(x_i, y_i) = w(\bar{x}, \bar{y}) + \left(\frac{\partial w}{\partial x} \right)_{\bar{x}, \bar{y}} (x_i - \bar{x}) + \left(\frac{\partial w}{\partial y} \right)_{\bar{x}, \bar{y}} (y_i - \bar{y}) + \dots \quad (\text{D})$$

woraus

$$w_i - \bar{w} \approx \left(\frac{\partial w}{\partial x} \right)_{\bar{x}, \bar{y}} (x_i - \bar{x}) + \left(\frac{\partial w}{\partial y} \right)_{\bar{x}, \bar{y}} (y_i - \bar{y}) \dots \quad (\text{E})$$

folgt. Für s_w^2 folgt daraus

$$\begin{aligned} s_w^2 &= \frac{1}{n-1} \sum_{i=1}^n (w_i - \bar{w})^2 \approx \frac{1}{n-1} \sum_{i=1}^n \left[\left(\frac{\partial w}{\partial x} \right)_{\bar{x}, \bar{y}} (x_i - \bar{x}) + \left(\frac{\partial w}{\partial y} \right)_{\bar{x}, \bar{y}} (y_i - \bar{y}) \right]^2 = \\ &= \frac{1}{n-1} \sum_{i=1}^n \left[(x_i - \bar{x})^2 \left(\frac{\partial w}{\partial x} \right)_{\bar{x}, \bar{y}}^2 + (y_i - \bar{y})^2 \left(\frac{\partial w}{\partial y} \right)_{\bar{x}, \bar{y}}^2 + 2(x_i - \bar{x})(y_i - \bar{y}) \left(\frac{\partial w}{\partial x} \right)_{\bar{x}, \bar{y}} \left(\frac{\partial w}{\partial y} \right)_{\bar{x}, \bar{y}} \right] \\ s_w^2 &\approx s_x^2 \left(\frac{\partial w}{\partial x} \right)_{\bar{x}, \bar{y}}^2 + s_y^2 \left(\frac{\partial w}{\partial y} \right)_{\bar{x}, \bar{y}}^2 + 2s_{xy} \left(\frac{\partial w}{\partial x} \right)_{\bar{x}, \bar{y}} \left(\frac{\partial w}{\partial y} \right)_{\bar{x}, \bar{y}} \end{aligned} \quad (103)$$

Die gewöhnlichen Ableitungen des eindimensionalen Falls wurden durch partielle Ableitungen ersetzt. Zum Schluß wurde im gemischten Term Gebrauch von in Kapitel 6.2.1 eingeführten (empirischen) Kovarianz Gl. 101 gemacht. Ist die Kovarianz Null bzw. ist keine Information darüber vorhanden, so fällt der Term weg bzw. wird notgedrungenerweise vernachlässigt.

Die Verwendung nur des linearen Terms kann natürlich eine zu grobe Näherung sein, dennoch ist Gl. 103 die Standardmethode der Fehlerfortpflanzung.

► **Beispiele:** (i) Die Fläche eines Kreises berechnet sich nach $A = \pi r^2$. Was ist die Standardabweichung der Fläche, wenn $\bar{r} = 10$ cm und $s_r = 3$ mm? Für s_A gilt

$$s_A^2 = \left(\frac{dA}{dr} \right)^2 s_r^2 = s_r^2 (2\pi \bar{r})^2 = 0.3^2 (2\pi 10)^2 = 0.09 \cdot 400\pi^2 = 36\pi^2$$

Somit ist $s_A = 6\pi$ cm².

(ii) Was ist nun die oben angesprochene Standardabweichung des Volumens eines Quaders. Es gilt $V = a \times b \times c$, und somit für s_V

$$s_V^2 = s_a^2 \cdot (\bar{b} \bar{c})^2 + s_b^2 \cdot (\bar{a} \bar{c})^2 + s_c^2 \cdot (\bar{a} \bar{b})^2 + 2s_{ab}(bc)(ac) + 2s_{bc}(ac)(ab) + 2s_{ac}(bc)(ab)$$

wobei die letzten drei Terme nur von Interesse sind, wenn die drei Kantenlängen (bzw. die Fehler, die in ihrer Bestimmung auftreten) korreliert sind.

(iii) Eine Größe ξ berechne sich als Funktion von u und v gemäß $\xi = u^2 + v^2$. Wenn s_u, s_v, s_{uv} bekannt sind, was ist s_ξ ? Es ist $\partial\xi/\partial u = 2u, \partial\xi/\partial v = 2v$. Somit ist

$$s_\xi^2 = 4\bar{u}^2 s_u^2 + 4\bar{v}^2 s_v^2 + 8\bar{u}\bar{v} s_{uv}.$$

◀

Hinweis: Viele Formelsammlungen behandeln nur den Fall unkorrelierter Größen, d.h. Verschwinden aller Kovarianzen. Überprüfen Sie Ihre Formelsammlung und fügen Sie ggf. die allgemeine Gl. 103 hinzu!

6.2.3 Lineare Regression

Die in diesem Abschnitt behandelte Technik gehört eigentlich in die statistische Behandlung zweidimensionaler Datensätze. Da im Rahmen dieser Vorlesung allerdings die Zeit fehlt, die Behandlung mehrdimensionaler Datensätze systematisch auf wahrscheinlichkeitstheoretischer Grundlage zu behandeln, begnügen wir uns mit einem wichtigen Spezialfall, den wir mit unserem Wissen über mehrdimensionale Differentialrechnung behandeln können.

Folgende Problemstellung: Sie messen eine Größe $y = y(x)$ als Funktion von x (Verdünnungsreihe für photometrische Messungen (also Absorption als Funktion der Konzentration), Konzentration als Funktion der Zeit (Reaktionskinetik) usw.) In vielen Fällen gilt nun zwischen y und x eine lineare Beziehung

$$y = ax + b \quad (\text{A})$$

oder Sie können durch eine geeignete Umformung auf eine lineare Beziehung kommen (Häufigstes Beispiel ist vermutlich Logarithmieren y , d.h., $Y = \ln y = ax + b$. Was ist die "ideale Gerade," die Sie durch Ihren Datensatz $\{(x_i, y_i)\}$ legen können. Wenn Ihre Messungen in Ordnung sind, und die gemessenen Größen theoretisch der Beziehung (A) gehorchen, dann ist zwar optisch (durch Plotten der Datenpunkte) klar, daß die Daten idealerweise auf einer Gerade liegen sollten, aber wie diese Gerade konstruiert werden soll, verlangt mehr als Intuition.^{24,25}

Die Idee, die hinter den meisten *Regressionsanalysen* steht, ist die Minimierung der Summe der Abweichungsquadrate

$$\chi^2 = \sum_{i=1}^n (y_i - y_i^{id})^2. \quad (104)$$

Die y_i sind die tatsächlich gemessenen Werte von $y = y(x)$, während die y_i^{id} die Werte sind, die aus der zu suchenden Fitfunktion berechnet werden. Für den Fall einer linearen Abhängigkeit (A) soll

²⁴Der Autor zieht es vor, die vor allem in Praktika der analytischen Chemie vorkommende Meinung nicht weiter zu kommentieren, daß jeder, der versucht durch mehr als zwei Datenpunkte eine Gerade zu legen, selber schuld sei.

²⁵Der Autor empfiehlt unbedingt, in der Praxis die Daten zu plotten: Wenn diese nicht wie eine Gerade aussehen, dann ist die im Folgenden beschriebene Prozedur vermutlich Zeitverschwendung!!

also

$$\chi^2 = \sum_{i=1}^n (y_i - (a x_i + b))^2 = \min \quad (\text{B})$$

minimiert werden.

Anders ausgedrückt sind also die Werte \hat{a} und \hat{b} gesucht, für die (B) ein Minimum hat. Nach dem in den Kapiteln über Funktionen mehrerer Variablen gesagtem, heißt das aber, daß wir das Gleichungssystem

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= 0 = \frac{\partial \sum_{i=1}^n (y_i - (a x_i + b))^2}{\partial a} \\ \frac{\partial \chi^2}{\partial b} &= 0 = \frac{\partial \sum_{i=1}^n (y_i - (a x_i + b))^2}{\partial b} \end{aligned}$$

lösen müssen (wobei wir nicht überprüfen werden, ob die gefundene Kandidatenstelle tatsächlich ein Minimum ist). Die partiellen Ableitungen ergeben²⁶

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= 2 \left(\sum_{i=1}^n (y_i - (a x_i + b))(-x_i) \right) = 0 \\ \frac{\partial \chi^2}{\partial b} &= 2 \left(\sum_{i=1}^n (y_i - (a x_i + b))(-1) \right) = 0 \end{aligned}$$

oder

$$\begin{aligned} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i &= 0 \\ \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - n b &= 0. \end{aligned}$$

Aus der zweiten Gleichung folgt aber

$$\hat{b} = \frac{1}{n} \sum_{i=1}^n y_i - a \frac{1}{n} \sum_{i=1}^n x_i = \bar{y} - a \bar{x} \quad (105)$$

womit man in die erste Gleichung einsetzen kann:

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i^2 + (\bar{y} - a \bar{x}) \sum_{i=1}^n x_i$$

Auflösen nach a führt auf

$$\hat{a} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{s_{xy}}{s_x^2} \quad (106)$$

²⁶Lassen Sie sich beim Ableiten nicht von den Summationszeichen irritieren. Da immer gilt $\partial(f+g)/\partial a = (\partial f/\partial a) + (\partial g/\partial a)$, genügt es für einen Term abzuleiten, z.B. $\partial(y_i - (a x_i + b))^2/\partial a = [\text{Kettenregel!!!}] 2(y_i - (a x_i + b))(\partial(y_i - (a x_i + b))/\partial a) = [\text{innere Ableitung ausrechnen!!!}] 2(y_i - (a x_i + b))(-x_i) = 2(-x_i y_i + a x_i^2 + b x_i)$. Danach stellen Sie einfach das Summationszeichen voran

Hinweis: Die Ausdrücke für a und b können auf vielerlei Art geschrieben werden, machen Sie sich mit der Darstellung Ihrer Formelsammlung vertraut. Die Ausarbeitung der letzten Darstellung von \hat{a} als Funktion von Varianz und Kovarianz bleibt dem Leser überlassen.

Die eben gezeigte Methode läßt sich übrigens auf allgemeine lineare Zusammenhänge erweitern. Für Funktionen der Form $y = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n$ kann man wie für den eben behandelten Spezialfall $y = a_0x + a_1 = ax + b$ analytische Ausdrücke für die Koeffizienten a_0, a_1, \dots, a_n finden.

► **Beispiel:** Die Reaktionsgeschwindigkeit einer Reaktion 1. Ordnung ($A \rightarrow$ Produkte) gehorcht der Differentialgleichung

$$-\frac{dc_A(t)}{dt} = k c_A(t)$$

mit der Anfangsbedingung $c_A(t=0) = c_{A,0}$. Die Symbole bedeuten die Konzentration der Ausgangssubstanz A, c_A , zur Zeit t . Die Substanzmenge zu Beginn des Experiments ($t=0$) ist $c_{A,0}$. Die für die Art der Reaktion spezifische Größe ist die Geschwindigkeitskonstante k .

(a) Lösen Sie die Differentialgleichung!

(b) Die nachstehende Tabelle enthält "Meß"ergebnisse²⁷ für eine solche Reaktion 1. Ordnung. Gegeben ist c_A als Funktion der Zeit (in Minuten) (die Anfangskonzentration ($c_{A,0}$) beträgt 1 mol/l). Ermitteln Sie unter Verwendung von linearer (!) Regression die Geschwindigkeitskonstante.

t/min	$c_A(t)$
0	1,000
1	0,599
2	0,358
3	0,215
4	0,128
5	0,077
6	0,046
7	0,028
8	0,017
9	0,010
10	0,006

Die Differentialgleichung

$$-\frac{dc_A(t)}{dt} = k c_A(t)$$

wird durch Trennung der Variablen auf

$$\frac{dc_A}{c_A} = -k dt$$

umgeformt, woraus man nach Integration

$$\ln c_A = -kt + \alpha'$$

bzw.

$$c_A = \alpha \exp(-kt)$$

erhält. Aus der Anfangsbedingung $c_A(t=0) = c_{A,0}$ folgt $\alpha = c_{A,0}$ und somit

$$c_A(t) = c_{A,0} \exp(-kt).$$

²⁷Ergebnisse wurden am Computer generiert, daher die Anführungszeichen!

Der entscheidende Schritt zur Verwendung *linearer* Regression besteht in der Verwendung der logarithmierten Form der Lösung von (a)

$$\ln c_A(t) = \ln c_{A,0} - kt = d - kt.$$

Eine Auftragung des Logarithmus der Konzentration als Funktion der Zeit ist also (von Abweichungen durch Meßfehler bzw. -ungenauigkeiten abgesehen) eine Gerade. Aus der Angabe $c_{A,0} = 1 \text{ mol/l}$ folgt weiters $d = 0$. Lineare Regression ergibt die Koeffizienten der Geradengleichung

$$y = a + bt,$$

und somit ergibt sich aus dem Vergleich mit der vorigen Gleichung, daß $d = a$ und $k = -b$.

Im folgenden eine Tabellierung der wichtigsten Zwischenergebnisse, wie sie für die Berechnung der Regressionsgeraden mit einem einfachen Taschenrechner notwendig sind.

$\sum t_i$	55,0
$\sum t_i^2$	385,0
$\sum y_i$	-28,2
$\sum y_i^2$	101,3
$\sum y_i t_i$	-197,5
$\Delta = N \sum t_i^2 - (\sum t_i)^2$	1210,0
$a = \frac{1}{\Delta} (\sum t_i^2 \sum y_i - \sum t_i \sum t_i y_i)$	0,0
$b = \frac{1}{\Delta} (N \sum y_i t_i - \sum t_i \sum y_i)$	-0,513

Somit findet man eine Reaktionskonstante von $0,513 \text{ min}^{-1}$. ◀

Die Berechnung von linearer Regression ist heute natürlich trivial und besteht aus dem Drücken einer Maustaste am Computer. Das vielleicht interessanteste Element des eben vorgerechneten Beispiels ist daher die Rückführung eines nichtlinearen Problems auf den Fall linearer Regression. Numerisch sind auch beliebige nichtlineare Fitprobleme $y = f(x)$, also z.B. $y = b \exp(ax)$ lösbar, oft ebenfalls per Mausdruck. Dennoch ist in Fällen wie diesen Linearisierung vorzuziehen. Die analytische Lösung des linearen Problems $\ln y = ax + b$ ist exakt, während der Versuch $y = b \exp(ax)$ zu fitten, u.U. numerisch instabil sein kann. Daher sollte man Probleme, die auf eine lineare Fragestellung reduzierbar sind, auch auf eine solche umformen, anstatt blindlings eine Computerfunktion aufzurufen.

6.3 Aus der Wahrscheinlichkeitsrechnung folgende Methoden

Dank der rapiden Entwicklung der Computertechnologie (und Open Source Programmpaketen wie z.B. <http://www.r-project.org>) sind heute statistische Methoden, die wegen ihres Rechenaufwands (oder Preises) früher nur Experten zugänglich waren, heute für jedermann verfügb- und anwendbar. Die größte Schwierigkeit besteht heute daher darin, die für ein Problem korrekte Methode auszuwählen.

Im folgenden beschränken wir uns auf eindimensionale Problemstellungen. Zunächst wollen wir die in Abschn. 6.2 eingeführten empirischen Größen mathematisch begründen, um ein Gefühl für die Zusammenhänge zwischen Wahrscheinlichkeitsrechnung und Statistik zu schaffen.. Danach stellen wir uns die (empirisch nicht lösbare) Frage, wie wir z.B. die Unsicherheit einer durch Mittelwertbildung gewonnenen Größe (also z.B. des Mittelwerts von fünf Messungen der Konzentration eines Schadstoffs) präzisieren können. Schließlich illustrieren wir anhand einfachster Beispiele die Theorie, die statistischen Tests zu Grunde liegt.

6.3.1 Parameterschätzung

Es ist zunächst einmal notwendig, die Sprache der Wahrscheinlichkeitsrechnung (“Zufallsvariable, Erwartungswert, Verteilung”) und die reale Situation (“Tabelle mit Meßdaten”) zur Deckung zu bringen. Gegen Ende von Abschnitt 5, insbesondere im Unterkapitel 5.5 wurde einige Male von einer Zufallsvariablen \mathbf{S}_n gesprochen, welche die Summe von n (stochastisch) unabhängigen Zufallsvariablen $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ ist. Kapitel 5.5 beschäftigte sich mit einigen allgemeinen Grenzwerten für Erwartungswerte von \mathbf{S}_n bzw. $\mathbf{Z}_n = \mathbf{S}_n/n$. Die Statistik faßt jede Einzelmessung als Zufallsexperiment auf, das von den Messungen davor und danach (stochastisch) unabhängig ist.²⁸ Ein konkret erhaltener Meßwert \hat{x}_i ist die Realisierung der Zufallsvariablen \mathbf{X}_i . Damit ist aber die Brücke zwischen Wahrscheinlichkeitsrechnung und konkreter Meßsituation geschlagen: Es gilt z.B. der zentrale Grenzwertsatz, d.h., würde ich unendlich viele Messungen durchführen, so wäre \mathbf{S}_n $N(\mu, \sigma^2)$ -verteilt. In der Praxis trifft das schon ab $n > 30$ in ausreichend guter Näherung zu, und aus diesem Grund ist der zentrale Grenzwertsatz so wichtig für die statistischen Theorie. Durch die Korrespondenz

$$\text{Meßwert } \hat{x}_i \Leftrightarrow \text{Realisation der Zufallsvariablen } \mathbf{X}_i$$

können weiters viele theoretische Erkenntnisse der Wahrscheinlichkeitsrechnung auf die angewandte Statistik übertragen werden (hierbei spielen die in Abschnitt 5.6 zumindestens erwähnten Testverteilungen eine große Rolle).

Exkurs: Bernoulli-Experimente: Aus Gründen der Anschaulichkeit machen wir jetzt gewissermaßen einen weiteren Exkurs und betrachten ein im Vergleich zur Behandlung von Meßdaten einfacheres Problem: Stichproben werden auf ein Qualitätskriterium hin getestet. Dieser Vorgang (der auch auf viele andere Situationen zutrifft) ist ein Bernoulli-Schema (vgl. Abschn. 4.1) Die Zufallsvariable, die einer Stichprobe entspricht, kann nur zwei Werte annehmen, 0 (Qualitätskriterium nicht erfüllt) und 1 (Qualitätskriterium erfüllt). Was kann ich aus der Häufigkeit, mit der in diesen Stichproben das Kriterium erfüllt bzw. nicht erfüllt ist, über die Qualität der Gesamtheit (z.B. Wahrscheinlichkeit von Ausschuß in einer Produktion, Wahrscheinlichkeit der Wirksamkeit eines neuen Arzneistoffs) sagen. Die folgenden Betrachtungen verwenden in intensiver Weise Material der Abschnitte 4.1, 5.2.4 und 5.5.

Sie erinnern sich, daß der Begriff der Wahrscheinlichkeit entweder über die relative Häufigkeit

$$h(A) = \frac{\text{Anzahl der Versuche, in denen A eintritt}}{\text{Gesamtanzahl der Versuche}}$$

eingeführt werden kann, oder daß die axiomatisch eingeführte Wahrscheinlichkeit Eigenschaften besitzt, die denen von $h(A)$ analog sind. Es liegt also nahe, die in dem Bernoulli-Experiment gefundene relative Häufigkeit, daß das Kriterium A erfüllt ist, als Schätzwert für die Wahrscheinlichkeit von A zu verwenden

$$p = \mathbf{P}(A) \approx h_A. \quad (\text{A})$$

Zieht man ein paar mal n Stichproben, so werden die daraus gewonnenen relativen Häufigkeiten im allgemeinen variieren. Wie berechtigt ist die Abschätzung (A)?

²⁸Soweit die Prämisse der mathematischen Statistik, die in der Praxis natürlich hinterfragt bzw. realisiert werden muß

Zu dem beschriebenen Bernoulliexperiment gehört zum i -ten Versuch die Zufallsvariable

$$\mathbf{X}_i = \begin{cases} 1, & \text{Kriterium erfüllt} \\ 0, & \text{sonst} \end{cases}$$

der die Realisierung

$$x_i = \begin{cases} 1, & \text{Kriterium erfüllt} \\ 0, & \text{sonst} \end{cases}$$

zugeordnet ist. Die zu \mathbf{X}_i gehörige Verteilungsfunktion f hat die Werte $f(0) = 1-p$, $f(1) = p$. Damit erhält man die Erwartungswerte (s. Abschn. 5.2.3)

$$E(\mathbf{X}_i) = 0 \cdot (1-p) + 1 \cdot p = p = \mu$$

$$\text{Var}(\mathbf{X}_i) = 0^2 \cdot (1-p) + 1^2 \cdot p - \mu^2 = p - p^2 = p(1-p) = \sigma^2$$

Die Summe $\mathbf{S}_n = \sum \mathbf{X}_i$ ist nach Abschnitt 5.2.4 binomialverteilt mit $E(\mathbf{S}_n) = np$ und $\text{Var}(\mathbf{S}_n) = np(1-p)$. Diese neue Zufallsvariable \mathbf{S}_n hat in einem Bernoulliexperiment vom Umfang n die absolute Häufigkeit, mit der das Kriterium erfüllt wird, als Realisierung. Die relative Häufigkeit

$$h(A) = \frac{1}{n} \sum_{i=1}^n x_i$$

ist somit die Realisierung der Zufallsvariablen $\bar{\mathbf{X}} = 1/n \mathbf{S}_n$. Was sind jetzt $E(h(A))$ und $\text{Var}(h(A))$?

$$E(h(A)) = E(\bar{\mathbf{X}}) = E\left(\frac{1}{n} \mathbf{S}_n\right) = \frac{1}{n} E(\mathbf{S}_n) = \frac{1}{n} np = p \quad (\text{B})$$

und

$$\text{Var}(h(A)) = \text{Var}(\bar{\mathbf{X}}) = \text{Var}\left(\frac{1}{n} \mathbf{S}_n\right) = \frac{1}{n^2} \text{Var}(\mathbf{S}_n) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n} \quad (\text{C})$$

Die Umformungen bedienen sich der Eigenschaften von E und Var (s. S. 32). Gl. (B) zeigt aber, daß Ansatz (A) begründet war: Der Erwartungswert von $h(A)$ ist die gesuchte (unbekannte) Wahrscheinlichkeit p , mit der das fragliche Kriterium in der Gesamtheit erfüllt ist. Man bezeichnet $h(A) = \bar{\mathbf{X}}$ als *erwartungstreue Schätzfunktion* für den Parameter p , der Zahlenwert $h(A)$ heißt Schätzwert.

Ganz fertig sind wir jedoch noch nicht. Gl. (B) ist zwar schön und gut, aber nichtsdestotrotz werden konkrete Realisierungen von $\bar{\mathbf{X}}$, d.h. Zahlenwerte $h(A)$, mehr oder weniger stark fluktuieren. Gefragt ist daher eine *konsistente* Schätzfunktionen für p , für die für beliebiges $\epsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \mathbf{P}(|h(A) - p| > \epsilon) = 0$$

Wenn Sie sich Abschnitt 5.5 im Detail durchgelesen haben, dann sollten Sie in dieser Bedingung Gl. 97 wiedererkennen, das schwache Gesetz der großen Zahlen. Gemäß diesem allgemeinen Gesetz ist $h(A)$ also sicher auch eine konsistente Schätzfunktion für p . Man kann sich das aber auch leicht aus Gl. (C) ableiten, die besagt, daß $\text{Var}(h(A)) = \frac{p(1-p)}{n}$. Die rechte Seite wird aber für $p = 1/2$ maximal, und somit gilt die Ungleichung

$$\text{Var}(h(A)) = \frac{p(1-p)}{n} \leq \frac{1}{4n}$$

Kombiniert man diese Ungleichung mit der Tschebyscheffschen Ungleichung (Gl. 96, S. 44)

$$\mathbf{P}(|h(A) - E(h(A))| > \epsilon) < \frac{\text{Var}(h(A))}{\epsilon^2}$$

so erhält man

$$\mathbf{P}(|h(A) - \underbrace{E(h(A))}_p| > \epsilon) < \frac{1}{4n\epsilon^2}$$

wobei die rechte Seite für $n \rightarrow \infty$ natürlich verschwindet.

Begründung der empirisch eingeführten Größen Mittelwert \bar{x} und Varianz s^2 : Wir betrachten jetzt die in Abschnitt 6.2.1 eingeführten Größen \bar{x} und s^2 als Realisierung der Zufallsvariablen $\bar{\mathbf{X}} = 1/n \sum \mathbf{X}_i$ und $\mathbf{S}^2 = 1/(n-1) \sum (\mathbf{X}_i - \bar{\mathbf{X}})^2$, wobei die \mathbf{X}_i paarweise (stochastisch) unabhängige Zufallsvariablen mit Realisierungen x_i sind. Die \mathbf{X}_i haben alle denselben Erwartungswert $\mu = E(\mathbf{X}_i)$ und dieselbe Varianz $\sigma^2 = \text{Var}(\mathbf{X}_i)$. (Jetzt sind wir bei der konkreten Situation von Meßdaten: jede Messung ist ein Zufallsexperiment mit Zufallsvariable \mathbf{X}_i ; der in einer Messung erhaltene Meßwert x_i ist eine Realisierung von \mathbf{X}_i .) Wir zeigen nun die Richtigkeit der folgenden drei Gleichungen:

$$E(\bar{\mathbf{X}}) = \mu \quad (107)$$

$$\text{Var}(\bar{\mathbf{X}}) = \frac{\sigma^2}{n} \quad (108)$$

$$E(\mathbf{S}^2) = \sigma^2 \quad (109)$$

► **Beweis:** Zur Erinnerung die Voraussetzungen:

$$E(\mathbf{X}_i) = \mu \quad \text{Var}(\mathbf{X}_i) = \sigma^2 \quad \text{für alle } i$$

Somit gilt

$$E(\bar{\mathbf{X}}) = E\left(\frac{1}{n} \sum_i \mathbf{X}_i\right) = \frac{1}{n} \sum_i E(\mathbf{X}_i) = \frac{1}{n} n \mu = \mu$$

womit (107) bewiesen ist. Der Beweis von (108) stützt sich auf die Additivität der Varianz von stochastisch unabhängigen Zufallsvariablen:

$$\text{Var}(\bar{\mathbf{X}}) = \text{Var}\left(\frac{1}{n} \sum_i \mathbf{X}_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_i \mathbf{X}_i\right) = \frac{1}{n^2} \sum_i \text{Var}(\mathbf{X}_i) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$

Um (109) zu beweisen, formen wir zunächst $\mathbf{S}^2 = 1/(n-1) \sum (\mathbf{X}_i - \bar{\mathbf{X}})^2$ um

$$(n-1) \mathbf{S}^2 = \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})^2 = \sum_i \mathbf{X}_i^2 - 2 \sum_i \mathbf{X}_i \bar{\mathbf{X}} + n \bar{\mathbf{X}}^2 = \sum_i \mathbf{X}_i^2 - n \bar{\mathbf{X}}^2$$

Bei der Einführung der Varianz (Gl. 59, S. 32) haben wir ganz allgemein gefunden, daß für eine beliebige Zufallsvariable \mathbf{Z} gilt $\text{Var}(\mathbf{Z}) = E(\mathbf{Z}^2) - (E(\mathbf{Z}))^2 = E(\mathbf{Z}^2) - \mu_{\mathbf{Z}}^2$, woraus

$$E(\mathbf{Z}^2) = \text{Var}(\mathbf{Z}) + (E(\mathbf{Z}))^2$$

folgt. Insbesondere gilt dann aber

$$E(\mathbf{X}_i^2) = \text{Var}(\mathbf{X}_i) + (E(\mathbf{X}_i))^2 = \sigma^2 + \mu^2$$

$$E(\bar{\mathbf{X}}^2) = \text{Var}(\bar{\mathbf{X}}) + (E(\bar{\mathbf{X}}))^2 = \frac{\sigma^2}{n} + \mu^2$$

Mit diesen Vorergebnissen ausgestattet berechnen wir jetzt

$$E((n-1)\mathbf{S}^2) = E\left(\sum_i \mathbf{X}_i^2 - n\bar{\mathbf{X}}^2\right) = \sum_i E(\mathbf{X}_i^2) - nE(\bar{\mathbf{X}}^2) = \sum_i (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) =$$

$$= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = (n-1)\sigma^2 = (n-1)E(\mathbf{S}^2).$$

Division durch $(n-1)$ beweist Gl. 109. Die Rechnung zeigt weiters die Notwendigkeit des Faktors $1/(n-1)$ in der Definition von s^2 . Hätten wir nur $1/n$ genommen, hätten wir anstelle von $E(\mathbf{S}^2) = \sigma^2$ den nicht erwartungstreuen Schätzwert $\sigma^2 - \sigma^2/n$ bekommen. ◀

Der Schätzwert \bar{x} für μ ist übrigens nicht nur erwartungstreu, sondern auch konsistent. Dies folgt entweder direkt aus dem schwachen Gesetz der großen Zahlen (Abschn. 5.5.2) oder durch Kombination von Gl. 108 mit der Tschebyscheffschen Ungleichung (96),

$$\mathbf{P}(|\bar{\mathbf{X}} - \mu| > \epsilon) \leq \frac{\text{Var}(\bar{\mathbf{X}})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Der obige Ausdruck geht für beliebige $\epsilon > 0$ gegen Null, wenn nur n groß genug ist.

6.3.2 Fehlerbalken, Konfidenzintervalle für den Mittelwert

Gleichungen 107–109 zusammen mit dem schwachen Gesetz der großen Zahlen zeigen, daß die Verwendung der zunächst empirisch eingeführten Größen \bar{x} und s^2 zu konsistenten Schätzwerten für $E(\bar{\mathbf{X}}) = \mu$ und $\text{Var}(\bar{\mathbf{X}})$ führt. Dennoch wird der empirische Mittelwert \bar{x} im allgemeinen nicht exakt mit μ übereinstimmen. In vielen Anwendungen ist es aber wünschenswert sagen zu können, in welchem Intervall $I : \bar{x} - z \leq \bar{x} \leq \bar{x} + z$ sich μ befindet. Man kann z zwar nicht exakt bestimmen, aber es ist möglich, Werte für z zu bestimmen, sodaß μ mit einer gewissen Wahrscheinlichkeit (typischerweise 95% oder 99%) innerhalb des Intervalls I liegt.

“Vorübung”: Wie schon im letzten Abschnitt über Parameterschätzung beginnen wir mit einem vereinfachten Beispiel, das einerseits die Verknüpfung mit der Wahrscheinlichkeitsrechnung zeigt, und andererseits an einem verständlichen Beispiel gestattet, die in dieser Beziehung etwas esoterische Sprache der Statistik einzuführen. Wir betrachten eine Zufallsvariable \mathbf{X} , die $N(\mu, \sigma^2)$ -verteilt sei, d.h. die Verteilung von \mathbf{X} ist durch Gl. 80 gegeben (die korrespondierende Dichte ist der Integrand (samt Vorfaktor) in (80)). Wir wissen, daß die Dichte bei μ ein Maximum hat, andererseits wissen wir auch, daß wie bei jeder stetigen Verteilung die Wahrscheinlichkeit an einem Punkt gleich Null ist, d.h., es gilt auch $\mathbf{P}(\mathbf{X} = \mu) = 0$. Es ist daher nur sinnvoll, sich für die Wahrscheinlichkeit zu interessieren, daß \mathbf{X} Werte auf einem um μ zentrierten Intervall einnimmt. Die Antwort kennen wir bereits seit Abschnitt 5.3.3, es gilt

$$\mathbf{P}(\mu - z \leq \mathbf{X} \leq \mu + z) = \mathbf{P}(|\mathbf{X} - \mu| \leq z) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mu-z}^{\mu+z} du e^{-\frac{(u-\mu)^2}{2\sigma^2}}$$

In Abschnitt 5.3.3 wurde weiters gezeigt, daß jede Aufgabenstellung, die eine $N(\mu, \sigma)$ Verteilung involviert, durch die Substitution $t = (x - \mu)/\sigma$ mit Hilfe der $N(0, 1)$ Standardnormalverteilung ausgedrückt und gelöst werden kann. Es gilt

$$N(\mu, \sigma^2)(x) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

wobei $\Phi(x)$ durch Gl. 74 gegeben ist, und somit ist die gesuchte Wahrscheinlichkeit

$$\mathbf{P}(|\mathbf{X} - \mu| \leq z) = \Phi(z/\sigma) - \Phi(-z/\sigma) = 2\Phi(z/\sigma) - 1.$$

(Die letzte Umformung folgt aus Gl. 76.) Die Wahrscheinlichkeit hängt klarerweise von z ab (σ ist ja vorgegeben). Man kann die Fragestellung aber auch umdrehen, und denjenigen Wert von z suchen, für den die Wahrscheinlichkeit einen bestimmten Wert annimmt, d.h.

$$\Phi(z/\sigma) = \frac{1 + \mathbf{P}(|\mathbf{X} - \mu| \leq z)}{2}. \quad (110)$$

► **Beispiel:** Am besten setzen wir jetzt einmal mit konkreten Zahlen ein. Für den Mittelwert $\mu = 10.0$ und $\sigma^2 = 0.25$ sei der Bereich $\mu \pm z$ zu bestimmen, in dem sich 95% aller möglichen Werte der $N(10, 0.25)$ -verteilten Zufallsvariable befinden. Es muß gelten

$$\Phi(z/0.5) = \Phi(c) = \frac{1.95}{2} = 0.975.$$

Mit einer Tabelle der Standardnormalverteilung findet man $c = 1.96$, oder $z = c \cdot \sigma = 1.96 \cdot 0.5 = 0.98$. Mit anderen Worten, auf dem Intervall $9.02 \leq \mathbf{X} \leq 10.98$ (oder 10 ± 0.98) befinden sich 95% aller Werte, die die Zufallsvariable \mathbf{X} einnehmen kann. ◀

Statistiker bezeichnen Zielwahrscheinlichkeiten der Form $\mathbf{P}(|\mathbf{X} - \mu| \leq z)$ als *Konfidenzniveau* γ , und die Größe $c (= z/\sigma)$ wird auch das $\frac{1 + \gamma}{2}$ -Quantil der $N(0, 1)$ Standardnormalverteilung genannt.

Konfidenzintervall für den Erwartungswert μ einer normalverteilten Stichprobe bei bekannter Varianz σ_0^2 : Was hat das Ganze mit unserer ursprünglichen Fragestellung zu tun? Zur Beantwortung dieser (berechtigten) Frage betrachten wir jetzt folgende Situation. Sie führen Messungen einer Größe durch, deren Gesamtheit normalverteilt sei,²⁹ die Meßwerte bezeichnen wir wie üblich mit x_1, x_2, \dots, x_n . Ihr bester Schätzwert für den Erwartungswert μ der Probe ist $\bar{x} = 1/n \sum x_i$. Wir behandeln insofern einen Spezialfall, als die Varianz der Meßwerte $\text{Var}(\mathbf{X}_i) = \sigma_0^2$ bekannt sei. Dieser Fall kann in der Praxis auftreten, z.B. bei (gutkalibrierten) Automatpipetten: Sie geben die auszugebende Menge vor. Die Varianz der ausgegebenen Menge hängt nicht von der Einstellung ab, sondern ist eine Charakteristik des Produkts, die Ihnen entweder der Hersteller garantiert, oder die Sie aus Erfahrung kennen. Die Zufallsvariable $\bar{\mathbf{X}}$ ist also normalverteilt, $E(\bar{\mathbf{X}}) = \mu$ können wir schätzen, und $\text{Var}(\bar{\mathbf{X}}) = \sigma_0^2/n$ ist bekannt. Somit können wir aber den Vertrauensbereich zu einem bestimmten Konfidenzniveau γ völlig analog zur "Vorübung" durchführen. Ganz in Analogie zu Gl. 110 bestimmen wir ein z gemäß

$$\Phi(z\sqrt{n}/\sigma_0) = \Phi(c) = \frac{1 + \mathbf{P}(|\bar{\mathbf{X}} - \mu| \leq z)}{2} \approx \frac{1 + \mathbf{P}(|\bar{\mathbf{X}} - \bar{x}| \leq z)}{2}. \quad (111)$$

²⁹Erinnern Sie sich an den zentralen Grenzwertsatz. Er garantiert, daß die Annahme einer Normalverteilung der stochastisch unabhängigen Zufallsvariablen \mathbf{X}_i mit Realisierung x_i (=Meßwerte!) in vielen Fällen (zumindest in guter Näherung) gerechtfertigt ist!

Achtung: Der Faktor \sqrt{n} in $c = z\sqrt{n}/\sigma_0$ kommt daher, weil die interessierende Zufallsvariable $\bar{\mathbf{X}}$ $N(\mu, \sigma_0^2/n)$ -verteilt ist! Für ein vorgegebenes Konfidenzniveau $\gamma = \mathbf{P}(|\bar{\mathbf{X}} - \mu| \leq z) = \mathbf{P}(|\bar{\mathbf{X}} - \mu| \leq \sigma_0 c / \sqrt{n})$ bestimmen Sie wieder $c = \sqrt{n}z/\sigma_0$ mit Hilfe einer Tabelle der $N(0,1)$ Verteilung und berechnen sich dann z um das gesuchte Intervall $\bar{x} - z \leq \bar{x} \leq \bar{x} + z$ zu bestimmen. Die Zufallsvariable \mathbf{T} , die der $N(0,1)$ Verteilung genügt ist

$$\mathbf{T} = \sqrt{n} \frac{\bar{\mathbf{X}} - \mu}{\sigma_0}, \quad (112)$$

wobei für $\mu \approx \bar{x}$ verwendet wird.

► **Beispiel:** Eine Automatpipette wird auf 0.1 ml Ausgabemenge eingestellt. Die bekannte (von der eingestellten Menge unabhängige) Varianz betrage $\sigma_0^2 = 0.001 \text{ ml}^2$. Was ist der Vertrauensbereich zum Konfidenzniveau 0.99 für den Mittelwert von 100 mit dieser Pipette entnommenen Proben? Es muß gelten

$$\Phi(\sqrt{100}z/\sqrt{0.001}) = \Phi(c) = \frac{1.99}{2} = 0.995.$$

Mit einer Tabelle der Standardnormalverteilung findet man $c = 2.575$, oder $z = c \cdot \sigma_0/\sqrt{n} = 2.575 \cdot \sqrt{0.001}/10 = 0.008$. Das empirische Konfidenzintervall für den Mittelwert der 100 Proben ist daher $[0.092, 0.108]$ ml. ◀

Konfidenzintervall für den Erwartungswert μ einer normalverteilten Stichprobe bei unbekannter Varianz: Der in der Praxis viel realistischere Fall ist, daß man weder Mittelwert noch Varianz kennt, und nur die Schätzwerte \bar{x} , s^2 zur Verfügung hat. Es bleibt einem nicht viel anderes übrig, als σ durch s zu ersetzen. Dann gehorcht aber die Gl. 112 analoge Zufallsvariable

$$\mathbf{T} = \sqrt{n} \frac{\bar{\mathbf{X}} - \mu}{\mathbf{S}}, \quad (113)$$

nicht mehr einer $N(0,1)$ Verteilung, sondern der sogenannten t-Verteilung. (Die t-Verteilung ist eine der Test-Verteilungen, die im Abschnitt 5.6 erwähnt wurden.) Sie ist ebenso wie die $N(0,1)$ Verteilung tabelliert, wir bezeichnen sie im folgenden als $F_t^m(x)$. Gl. 111 ist somit durch

$$F_t^m(z\sqrt{n}/s) = F_t^m(c) = \frac{1 + \mathbf{P}(|\bar{\mathbf{X}} - \mu| \leq z)}{2}. \quad (114)$$

zu ersetzen. Der prinzipielle Ablauf der Berechnung eines Vertrauensbereichs bleibt also gleich, nur ist eine Tabelle von $F_t^m(x)$ zu verwenden. Diese sind meistens anders organisiert als Tabellen der $N(0,1)$ Verteilung. Erstens hängt der Wert von $F_t^m(x)$ von der *Anzahl der Freiheitsgrade* m ab, diese sind durch

$$m = n - 1,$$

also Anzahl der Datenpunkte minus 1, gegeben. Zweitens, da die Hauptaufgabe dieser Tabellen die Berechnung von Quantilen ist, sind sie nach den gängigsten Konfidenzniveaus γ einerseits, und nach Freiheitsgraden andererseits, geordnet (im Gegensatz zu $\Phi(x)$ Tabellen, die einfach nach x geordnet sind). Die t-Verteilung geht übrigens im Limes $n \rightarrow \infty$ in die Standardnormalverteilung über (was man wegen des Zusammenhangs zwischen σ^2 und s^2 auch erwarten würde). Weiters, sind viele Tabellen nicht nach γ , sondern nach $\alpha = (1 - \gamma)/2$ (oder $\tilde{\alpha} = 1 - \gamma$) geordnet, α ($\tilde{\alpha}$) wird manchmal als Irrtumswahrscheinlichkeit bezeichnet. Um zu illustrieren, was gemeint ist, ein kurzer Ausschnitt aus einer Tabelle von $F_t^m(x)$

m	Irrtumswahrscheinlichkeit					
	0.10	0.05	0.025	0.01	0.005	0.001
1.	3.078	6.314	12.706	31.821	63.657	318.313
2.	1.886	2.920	4.303	6.965	9.925	22.327
3.	1.638	2.353	3.182	4.541	5.841	10.215
4.	1.533	2.132	2.776	3.747	4.604	7.173
5.	1.476	2.015	2.571	3.365	4.032	5.893
6.	1.440	1.943	2.447	3.143	3.707	5.208
7.	1.415	1.895	2.365	2.998	3.499	4.782
8.	1.397	1.860	2.306	2.896	3.355	4.499
9.	1.383	1.833	2.262	2.821	3.250	4.296
10.	1.372	1.812	2.228	2.764	3.169	4.143
usw. . .						

► **Beispiel:** Betrachten wir ein einfaches Beispiel: Sie haben 6 Messungen durchgeführt ($n = 6$) und daraus $\bar{x} = 0.8405$, $s = 0.00394$ berechnet. Was ist der Vertrauensbereich für \bar{x} zum Konfidenzniveau $\gamma = 0.95$ und $\gamma = 0.99$. Die gezeigte Tabelle ist nach Irrtumswahrscheinlichkeiten $\alpha = (1 - \gamma)/2$ geordnet, wir müssen also in den Spalten $(1 - 0.95)/2 = 0.025$ und $(1 - 0.99)/2 = 0.005$ suchen, $m = n - 1 = 5$.

Wir finden für $\gamma = 0.95$ $c = 2.571$, somit ist $z = c \cdot s/\sqrt{n} = 2.571 \cdot 0.00394/\sqrt{6} = 0.0041$ und der Vertrauensbereich ist $[0.8364, 0.8446]$.

Für $\gamma = 0.99$ findet man $c = 4.032$, somit ist $z = c \cdot s/\sqrt{n} = 4.032 \cdot 0.00394/\sqrt{6} = 0.0065$ und der Vertrauensbereich ist $[0.8340, 0.8470]$. ◀

6.3.3 Beispiel eines statistischen Tests

Zum Abschluß “kratzen” wir an der Oberfläche des riesigen Untergebiets *Statistische Tests*. Das im folgenden besprochene Beispiel illustriert die Problemstellung und die Denkweise, die statistischen Tests zugrunde liegt. Eine allgemeine Einführung ist aus Zeitgründen nicht möglich. Die heute vermutlich größte Schwierigkeit bei der Auswahl eines statistischen Tests ist die Entscheidung, welcher Test auf das aktuelle Problem paßt. Das kann ein nichttriviales Problem sein, das durch die einfache Verfügbarkeit von (auch kostenloser) statistischer Software oft übersehen wird. Meine Empfehlung lautet: Handelt es sich um ein klar umrissenes Problem, auf daß gemäß der Literatur Test *abc* zutrifft, so verwenden Sie ihn. Weiterführende Literatur und “Kochrezepte” sind im Internet ohne Problem zu finden. Ausgezeichnete Onlinereferenzen sind

<http://www.statsoftinc.com/textbook/stathome.html> sowie
<http://www.itl.nist.gov/div898/handbook/index.htm>

Im Zweifelsfall oder wenn die zu beantwortende Frage sich nicht mit den Ihnen bekannten Standardsituationen deckt, *suchen Sie Rat bei einem Experten*, bevor Sie ein “Kochrezept” blind verwenden!!³⁰

³⁰Für die beste Statistiksoftware gilt leider noch immer das Gesetz von “Garbage in, garbage out”.

Wir betrachten die folgende Problemstellung: Ein Falschspieler besitzt zwei äußerlich nicht unterscheidbare Würfel. Einer ist fair, der andere würfelt Sechsen mit $p = 0.3$ (anstatt der fairen $p = 1/6$). Mit einem der Würfel wird n -mal gewürfelt, sie notieren die Häufigkeit von Sechsen und wollen entscheiden, ob mit dem fairen oder dem gezinkten Würfel gewürfelt wurde.

Wir nehmen an, daß als mögliche Werte der unbekanntenen Wahrscheinlichkeit $p = \mathbf{P}(A)$ nur die beiden (bekannten) Werte $p_0 = 1/6$ (fairer Würfel) und $p_1 = 0.3$ (gezinkter Würfel) in Frage kommen (man ordnet die Wahrscheinlichkeiten immer so, daß $p_0 < p_1$. Man stellt nun die sogenannte

$$\text{Nullhypothese } H_0: p = p_0$$

auf, die richtig oder falsch sein kann. Offensichtlich gibt es auch die

$$\text{Alternativhypothese } H_1: p = p_1.$$

Der Test läuft prinzipiell wie folgt ab: Es wird eine kritische Zahl c bestimmt (Details in Kürze), die in Kombination mit der aus einem Bernoulli-Experiment vom Umfang n (= n -mal Würfeln) bestimmten relativen Häufigkeit $h(6)$ für das Ereignis "Würfeln einer Sechs", zur Testentscheidung gemäß

$$h(6) > c \Rightarrow \text{Entscheidung für } H_1 \text{ und}$$

$$h(6) \leq c \Rightarrow \text{Entscheidung für } H_0$$

verwendet wird. Dabei können zwei Arten von Fehlern passieren: Eine Entscheidung für H_1 , obwohl H_0 richtig ist, heißt *Fehler 1. Art*, eine Entscheidung für H_0 , obwohl H_1 richtig ist, dagegen *Fehler 2. Art*. Die Wahrscheinlichkeit, daß ein Fehler 1. Art unterläuft, sei α , die eines Fehlers 2. Art β .

Man gibt sich zunächst ein α (z.B. $\alpha = 0.05$) für die Fehlerwahrscheinlichkeit 1. Art vor, "danach wird gewürfelt", und zwar n -mal, d.h., es wird ein Bernoulli-Experiment vom Umfang n durchgeführt. Die Zufallsvariable $\bar{\mathbf{X}}$ beschreibe die relative Häufigkeit $h(6)$, wobei die Wahrscheinlichkeit des Ereignis "Würfeln einer Sechs" entweder $p_0 = 1/6$ oder $p_1 = 0.3$ beträgt. Die Zufallsvariable $n\bar{\mathbf{X}}$ kennzeichnet dann die absolute Häufigkeit, also die Anzahl der Sechsen bei n -mal Würfeln. $n\bar{\mathbf{X}}$ ist binomialverteilt, und zwar mit $p = p_0$ falls H_0 richtig ist, ansonsten mit $p = p_1$. Klarerweise wird n relativ groß sein müssen, und somit nähern wir die Binomialverteilung durch eine Normalverteilung. In diesem Fall ist $n\bar{\mathbf{X}}$ näherungsweise $N(np_0, np_0(1-p_0))$ -verteilt, falls H_0 zutrifft. Somit ist weiters $\bar{\mathbf{X}} = \frac{1}{n}n\bar{\mathbf{X}}$ näherungsweise $N(p_0, \frac{1}{n}p_0(1-p_0))$ -verteilt. In Analogie zu der im letzten Abschnitt verwendeten Vorgehensweise findet man nun die kritische Zahl c aus

$$\mathbf{P}(\bar{\mathbf{X}} > c | p = p_0) = 1 - \mathbf{P}(\bar{\mathbf{X}} \leq c | p = p_0) \approx 1 - \Phi\left(\frac{(c-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}\right) = \alpha$$

Mit $\Phi\left(\frac{(c-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}\right) = 1 - \alpha$ findet man mit Hilfe von Tabellen der Standardnormalverteilung den $(1 - \alpha)$ -Quantil $z_{1-\alpha}$, und erhält damit für die kritische Zahl

$$c = p_0 + z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \quad \text{mit} \quad \Phi(z_{1-\alpha}) = 1 - \alpha. \quad (115)$$

Durch c ist nun aber auch die Wahrscheinlichkeit β für den Fehler 2. Art bestimmt. Falls nämlich H_1 richtig ist, so ist $\bar{\mathbf{X}}$ näherungsweise $N(p_1, \frac{p_1(1-p_1)}{n})$ -verteilt. Daraus folgt

$$\beta = \mathbf{P}(\bar{\mathbf{X}} \leq c | p = p_1) \approx \Phi\left(\frac{(c-p_1)\sqrt{n}}{\sqrt{p_1(1-p_1)}}\right) \quad (116)$$

Aus einem Vergleich von Gln. 115 und 116 sieht man, daß bei konstantem n eine Verkleinerung von α eine Vergrößerung von c und damit eine Vergrößerung von β bedingt. Würde man umgekehrt ein kleineres β erzwingen wollen, würde sich α vergrößern.

Wählt man hingegen n hinreichend groß, so können die Varianzen $p_0(1-p_0)/n$ und $p_1(1-p_1)/n$ beliebig klein gemacht werden, somit werden mit wachsendem n beide Fehlerwahrscheinlichkeiten gleichzeitig kleiner. Bei Alternativtests wie in diesem Beispiel, kann man daher auch α und β vorgeben, und daraus den notwendigen Stichprobenumfang n und die kritische Zahl c bestimmen. Aus Gln. 115 und 116 folgt wegen $z_\beta = -z_{1-\beta}$ weiters

$$\begin{aligned} c &= p_0 + z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \\ c &= p_1 + z_{1-\beta} \sqrt{\frac{p_1(1-p_1)}{n}} \end{aligned} \quad (117)$$

Aus Gl. 117 lassen sich einige interessante Zusammenhänge ableiten. Subtraktion der beiden Gleichungen ergibt

$$p_1 - p_0 = \frac{1}{\sqrt{n}} \left(z_{1-\alpha} \sqrt{p_0(1-p_0)} + z_{1-\beta} \sqrt{p_1(1-p_1)} \right)$$

oder

$$n = \frac{\left(z_{1-\alpha} \sqrt{p_0(1-p_0)} + z_{1-\beta} \sqrt{p_1(1-p_1)} \right)^2}{(p_1 - p_0)^2} \quad (118)$$

Man sieht, daß die Stichprobenanzahl n für gegebenes α, β stark von der Differenz $p_1 - p_0$ abhängt. Aus (117) und (118) folgt weiters

$$c = p_0 + \frac{z_{1-\alpha} \sqrt{p_0(1-p_0)}(p_1 - p_0)}{\left(z_{1-\alpha} \sqrt{p_0(1-p_0)} + z_{1-\beta} \sqrt{p_1(1-p_1)} \right)}$$

und hieraus

$$c = \frac{p_0 z_{1-\beta} \sqrt{p_1(1-p_1)} + p_1 z_{1-\alpha} \sqrt{p_0(1-p_0)}}{\left(z_{1-\alpha} \sqrt{p_0(1-p_0)} + z_{1-\beta} \sqrt{p_1(1-p_1)} \right)} \quad (119)$$

► **Beispiel:** Mit einem der beschriebenen Würfel ($p_0 = 1/6, p_1 = 0.3$) werde $n = 300$ Mal gewürfelt. Aus (115) und (116) folgt

$$c = \frac{1}{6} + z_{1-\alpha} \sqrt{\frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{300}} = \frac{1}{6} + \frac{z_{1-\alpha}}{\sqrt{2160}}$$

und

$$\beta = \Phi \left((c - 0.3) \sqrt{\frac{300}{0.3 \cdot 0.7}} \right) = \Phi(37.796(c - 0.3))$$

Für $\alpha = 0.01$ erhalten wir $z_{0.99} = 2.33, c = 0.217$ und $\beta = \Phi(-3.1479) = 1 - \Phi(3.1479) = 0.001$. Das heißt, daß für $n = 300$ die Entscheidung für $p = 0.3$ mit $\alpha = 0.01$ angenommen wird, wenn $h(6) > c = 0.217$.

Alternativ geben wir $\alpha = \beta = 0.0001$ vor, und erhalten wegen $z_{1-\alpha} = z_{1-\beta} = 3.719$ aus (118) den minimalen Stichprobenumfang

$$n \geq \frac{3.719^2 (\sqrt{5/36} + \sqrt{0.21})^2}{(0.3 - 1/6)^2} = 537.17,$$

also $n = 538$. ◀

Verwendete Literatur

- AHS Mathematikschulbücher (7. und 8. Klasse)
- H. G. Zachmann “Mathematik für Chemiker”, 4. Auflage, Verlag Chemie, Weinheim, Deerfield Beach, Basel 1981.
- W. Feller “An Introduction to Probability Theory and Its Applications”, 3. Auflage, John Wiley & Sons, New York, 1968.
- K. Bosch “Elementare Einführung in die Wahrscheinlichkeitsrechnung”, 7. Auflage, Vieweg, Braunschweig, Wiesbaden 1999.
- K. Bosch “Elementare Einführung in die angewandte Statistik”, 6. Auflage, Vieweg, Braunschweig, Wiesbaden 1997.
- R. Durbin, S. R. Eddy, A. Krogh, G. Mitchinson “Biological Sequence Analysis”, Cambridge University Press, Cambridge, New York, 1998.

Auflistung von Änderungen — “revision history”

Jänner 2003 Version 0.9. Erste vollständige Version, sicher noch mit vielen Fehlern behaftet.

Dezember 2002 Version \approx 0.6. Die Kapitel “Verteilungen, Erwartungswert und Varianz” sowie “Die statistische Behandlung experimenteller Daten” fehlen noch völlig, auch in den anderen Teilen sind noch Änderungen zu erwarten.

GNU Free Documentation License

GNU Free Documentation License
Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.
59 Temple Place, Suite 330, Boston, MA 02111-1307 USA
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “you”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “Modified Version” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document’s license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document’s license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled “History”, Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled “History” in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the “History” section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled “Acknowledgements” or “Dedications”, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting

on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (c) YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.