

UNTERLAGEN
zum Lehrveranstaltung
MATHEMATIK für
NATURWISSENSCHAFTEN (STATISTIK)

Ferdinand Österreichler FB Mathematik
der Universität Salzburg

Salzburg, März 2008

Contents

1	BESCHREIBENDE STATISTIK	5
1.1	Darstellungsformen eindimensionaler Daten	5
1.1.1	Einführende Beispiele	5
1.1.2	Tabellarische Darstellung	6
1.1.3	Graphische Darstellung	6
1.2	Kenngrößen eindimensinaler Daten: Zentral- und Streumaße .	13
1.2.1	Stichprobenmittel und Standardabweichung	13
1.2.2	Stichprobenmedian und mittlere absolute Abwei-chung	17
1.3	Weitere Mittelwerte	20
1.3.1	Das geometrische Mittel	20
1.3.2	Das harmonische Mittel	23
1.3.3	Der Spezialfall $n=2$	25
1.4	Lineare Regression und Korrelation	26
1.4.1	Aufgabenstellung und einführendes Beispiel	26
1.4.2	Herleitung der Gleichung der homogenen Regressions- geraden	27
1.4.3	Anwendungsbeispiele	29
1.4.4	Herleitung der Gleichung der allgemeinen Regression- sgeraden	31
2	WAHRSCHEINLICKEITSRECHNUNG	37
2.1	Die Binomialverteilung	37
2.2	Die Poissonverteilung	40
2.3	Die Exponentialverteilung	46
2.4	Die Normalverteilung	50
2.4.1	Allgemeines	50
2.4.2	Das Wahrscheinlichkeitsnetz	54

3	BEURTEILENDE STATISTIK	59
3.1	DER χ^2 -TEST	59
3.1.1	Einleitung	59
3.1.2	Anpassung von Modellen	64
3.1.3	Test zweier Wahrscheinlichkeitsverteilungen auf Gleichheit (Test auf Homogenität)	66
4	ÜBUNGSAUFGABEN	69

Chapter 1

BESCHREIBENDE STATISTIK

1.1 Darstellungsformen eindimensionaler Daten

1.1.1 Einführende Beispiele

Beispiel 1: Chuck-a-luck

In den Lehrbüchern der Wahrscheinlichkeitsrechnung aus dem angelsächsischen Raum findet sich häufig das folgende Spielchen, "Chuck-a-luck" genannt, das in geringfügig modifizierter Form auch im Casinospiel "Sic-Bo" integriert ist.

Man setzt einen Einsatz e auf eine Zahl $z \in \{1, \dots, 6\}$. Dann werden drei Würfel geworfen. Kommt z unter den drei Augenzahlen nicht vor, ist der Einsatz verloren. Kommt z genau ein-, zwei- oder dreimal vor, gewinnt man den ein-, zwei- bzw. dreifachen Einsatz.

Man spiele 36 Spiele und trage den Gewinn jedes Spieles gegen dessen Nummer auf.

Der Wertebereich des Gewinns X ist, bezogen auf den Einsatz $e = 1$, gleich $W_X = \{-1, 1, 2, 3\}$.

Beispiel 2: Warten bis zum ersten Erfolg

Ein Versuch bestehe darin, eine Münze so oft zu werfen, bis zum ersten Mal das Ereignis "Kopf" eintritt. X bezeichne die Anzahl der nötigen Münzwürfe.

Man führe $2^6 = 64$ Versuche durch und erstelle ein Histogramm für die Häufigkeiten der beobachteten "Wartezeiten" $X_i, i \in \{1, \dots, 64\}$, und zwar für die Klassen $K_1 = \{1\}, \dots, K_6 = \{6\}$ und $K_7 = \{7, 8, \dots\}$.

Der Wertebereich einer "Wartezeit" X ist die abzählbar unendliche Menge $W_X = \mathbb{N} \cup \{\infty\}$.

1.1.2 Tabellarische Darstellung

Wir gehen auch weiterhin davon aus, dass der Wertebereich W_X der zugrundeliegenden Variablen X eine endliche oder abzählbar unendliche Teilmenge der reellen Zahlen ist, dass also gilt $W_X = \{\omega_1, \omega_2, \dots\} \subset \mathbb{R}$. Wenn möglich - und das ist vielfach der Fall - nehmen wir an, dass gilt $\omega_1 < \omega_2 < \dots$.

Ausgehend von einer Urliste, das ist eine Folge $x_1, x_2, \dots, x_n \in W_X$ von Beobachtungswerten der Variablen X erstellen wir eine Häufigkeitstabelle.

Dann sind

$$\begin{aligned} H_j &= |\{i \in \{1, \dots, n\} : x_i = \omega_j\}| \\ &= \text{die Anzahl der Versuche mit dem Ausfall } \omega_j \end{aligned}$$

bzw.

$$h_j = \frac{H_j}{n} = \frac{\text{Anzahl der Versuche mit dem Ausfall } \omega_j}{\text{Gesamtanzahl der Versuche}}$$

die *absolute Häufigkeit* bzw. die *relative Häufigkeit* des Ausfalls $\omega_j, j \geq 1$.

1.1.3 Graphische Darstellung

a) Die Strichliste

Ist der Wertebereich der Variablen im Vergleich mit dem gewählten Stichprobenumfang klein und ist man von vornherein nicht an der Reihenfolge interessiert, in welcher die Daten erhoben werden, so ist zu erwägen, eine Strichliste anzufertigen¹. Dabei sind die einzelnen Beobachtungswerte unmittelbar bei deren Erhebung durch je einen waagrechten (oder senkrechten) Strich über (bzw. neben) dem zugehörigen Wert der Variablen in einer Strichliste

¹Es gibt Anwendungen, bei welchen man sich speziell für die Reihenfolgen von Messwerten interessiert. Typisch dafür sind Folgen von Tagen mit und ohne Niederschlag. Hinsichtlich der Untersuchung sogenannter *Läufe von binären Zufallsfolgen* sei auf die Diplomarbeit von Frau *Radauer* [?] verwiesen.

einzutragen. Zur Wahrung der Übersicht - und um das spätere Abzählen zu erleichtern - formt man in der hergebrachten Weise Blöcke zu je 5 Strichen. Auf diese Weise kann bereits während der Datenerhebung eine graphische Darstellung der Verteilung der Beobachtungswerte entstehen.²

b) Das Histogramm³

Bei einem Histogramm für die relativen Häufigkeiten der Werte einer metrischen Variablen trägt man über jeder Klasse die relative Häufigkeit pro Einheit der Variablen auf, also

$$\text{Höhe} = \frac{\text{relative Häufigkeit}}{\text{Klassenbreite}}$$

Ein Histogramm beruht auf einer geeigneten Klasseneinteilung. Eine solche ist die Partition eines Intervalls I von \mathbb{R} , welches den Wertebereich der Variablen umfasst, in endlich viele Intervalle (Klassen).

Für Variable mit abzählbar unendlichem oder überabzählbarem Wertebereich ist eine Klassenzusammenfassung unumgänglich. (Beispiele: Wartezeit bis zum ersten Erfolg, Anzahl der radioaktiven Zerfälle in einem Zeitintervall vorgegebener Länge. Messergebnisse mit prinzipiell beliebig genauer Messgenauigkeit.)

Bezeichnung: Ist der Wertebereich W_X einer Variablen X eine Teilmenge von $\{\mu + c \cdot z : z \in \mathbb{N}_0\}$ oder $\{\mu + c \cdot z : z \in \mathbb{Z}\}$ mit $c > 0$ und $\mu \in \mathbb{R}$, so sagt man, W_X habe eine *Gitterstruktur mit der Gitterkonstante*⁴ c . Seien $h_{\mu+c \cdot z}$ die relativen Häufigkeiten der Ausfälle $\mu + c \cdot z$ eines Versuches. Dann heißt die Funktion

$$f(x) = \sum_{z \in \mathbb{N}_0(\mathbb{Z})} \frac{h_{\mu+c \cdot z}}{c} \cdot 1_{[\mu+c(z-1/2), \mu+c(z+1/2))}(x), \quad x \in \mathbb{R}$$

das zugehörige *Histogramm*. (Man stellt das Histogramm zumeist so dar, dass man zu jedem Wert $\mu + c \cdot z$ ein Rechteck mit der Basis $[\mu + c(z -$

²Manche Beobachter verzichten bewusst darauf, während des Beobachtungsverlaufs Strichlisten anzufertigen, um jegliche subjektive Erwartungshaltung auszuschließen.

³Die Bezeichnung "Histogramm" (histogram) wurde vermutlich vom englischen Statistiker *Karl Pearson* (1857 – 1936) im Jahre 1895 erstmals verwendet.

⁴Da man stets davon ausgehen kann, dass die Messgenauigkeit jedes noch so guten Messinstruments beschränkt ist, ist zumeist auch die Annahme einer Gitterstruktur angebracht.

$1/2), \mu + c(z + 1/2))$ und dem Flächeninhalt $h_{\mu+c \cdot z}$ - und somit der Höhe $h_{\mu+c \cdot z}/c$ - zeichnet.)

Beispiel 3: Radioaktiver Zerfall

Ernest Rutherford (1871 – 1937) und *Hans W. Geiger* (1882 – 1945) verwendeten bei ihrem klassischen Experiment im Jahre 1910 eine Polonium-Quelle und registrierten für 2608 disjunkte Zeitintervalle von je 7.5 Sekunden Dauer die Anzahl der Szintillationen. Im Folgenden sind die beobachteten absoluten Häufigkeiten H_j der Intervalle mit $j, j \in \{0, \dots, 14\}$ Szintillationen dargestellt (vgl. beispielsweise [13], S. 36).

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	≥ 15
H_j	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1	0

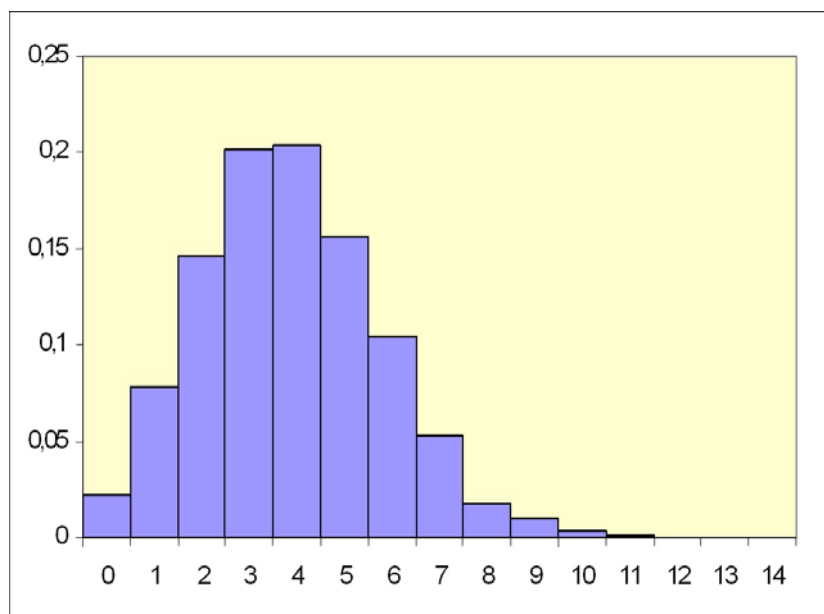


Abbildung: Zugehöriges Histogramm

Allgemeine Hinweise zur Klassenbildung

Generell lässt sich sagen, dass die Wahl der Klassen einige Erfahrung voraussetzt. Verwenden Sie nicht das erstbeste Histogramm. Experimentieren Sie, um eine möglichst charakteristische Form zu erzielen!

- Wahl der Anzahl der Klassen

Diese sollte man nicht zu klein und nicht zu groß wählen. Zu viele Klassen erzeugen ein unerwünscht erratisches Muster der Häufigkeiten, zu wenige ein unangebracht grobes Muster (Siehe Aufgabe 1). Als Orientierungshilfe kann im Fall *unimodaler* (*eingipfliger*) Verteilungen der gerundete Funktionswert $\frac{1}{5} \cdot (1 + \log_2(n))$ des Stichprobenumfangs dienen⁵.

- Wahlen von μ und der Klassenbreite c

Es ist vielfach bequem, $\mu = 0$ zu wählen. Für die Wahl der Klassenbreite c kann ein geeignetes Stängel-Blatt-Diagramm der Daten als Orientierungshilfe dienen. Weitere Orientierungshilfen hinsichtlich der Wahl von Klassenbreite und Klassengrenzen finden Sie in Abschnitt 1.3.3 von [17].

Verwendungszweck von Histogrammen

Histogramme werden verwendet,

- (i) um eine Datenmenge zwecks guter visueller Wahrnehmbarkeit von allgemeinen Charakteristika der Verteilung der Beobachtungswerte, wie repräsentativem Wert, Streuungsverhalten und charakteristischer Gestalt zu verdichten,
- (ii) um Hinweise zur Wahl eines geeigneten Wahrscheinlichkeitsmodells und nötigenfalls einer geeigneten Variablentransformation zu erhalten - welcher später eine genauere statistische Analyse folgen kann - und
- (iii) um ein unerwartetes Verhalten der zugrundeliegenden Variablen festzustellen und/oder ungewöhnliche Beobachtungswerte zu entdecken.

c) Das Stängel-Blatt-Diagramm⁶

Ist der Wertebereich einer Variablen vergleichsweise groß und ist man nicht an der Reihenfolge interessiert, in welcher die Daten erhoben werden, so kann die Anfertigung eines sogenannten *Stängel-Blatt-Diagramms* -

⁵Für Beispiel 2 stimmt $\frac{1}{5} \cdot (1 + \log_2(64)) = 7$ mit der vorgeschlagenen Klassenanzahl überein.

⁶Das *Stängel-Blatt-Diagramm* (*stem-and-leaf-display*) ist - wie das *Kastenbild* (*box-plot*) in Abschnitt 1.2.2 eine Darstellungsform der *Explorativen Datenanalyse* (*Exploratory Data Analysis*). Dieser Zweig der modernen beschreibenden Statistik ist eine Schöpfung des US-amerikanischen Statistikers *John W. Tukey* (1915 – 2000), vom dem übrigens auch die Kurzbezeichnung *bit* für *binary digit* stammt.

wie die einer Strichliste - bereits während der Datenerhebung erfolgen. Im wesentlichen entspricht dies nämlich dem Erstellen einer Strichliste mit zusätzlicher Klassenbildung. Im folgenden sei diese, vom amerikanischen Statistiker *John W. Tukey* (1977), dem Begründer der sogenannten Explorativen Datenanalyse, vorgeschlagene visuelle Darstellung von Daten anhand eines Beispiels erläutert.

Anmerkungen zur Anfertigung eines Stängel-Blatt-Diagramms

für eine Liste von Rohdaten mit Werten beispielsweise aus $\{70, \dots, 115\}$

Zunächst werden alle möglichen Zehnerzahlen (Zehnerziffer bzw. Paar, bestehend aus Hunderter- und Zehnerziffer), also $7, \dots, 11$ im "Stängel" - d.h. vertikal - eingetragen,

anschließend wird die Liste der Rohdaten der Reihe nach durchgegangen und für jeden Wert der Liste wird hinter der zugehörigen Zehnerzahl im Stängel die Einerziffer als Blatt - d.h. horizontal - eingetragen.

Modifikationsmöglichkeit: Sollte sich die Klassenbildung nach Zehnerzahlen als zu grob erweisen, ist zu erwägen, die Zehnerzahlen im Stängel jeweils zweimal anzuschreiben und hinter der jeweils ersten Zehnerzahl die Einerziffern von 0 bis 4 und hinter der jeweils zweiten Zehnerzahl die Einerziffern von 5 bis 9 als Blatt anzufügen.

d) Die empirische Verteilungsfunktion

Seien x_1, \dots, x_n reelle Beobachtungswerte, sei x eine weitere reelle Zahl und sei

$$|\{i \in \{1, \dots, n\} : x_i \leq x\}|$$

die absolute Häufigkeit der Beobachtungen, deren Beobachtungswerte kleiner oder gleich x sind. Dann heißt die Abbildung

$$x \longmapsto F_n(x) = \frac{|\{i \in \{1, \dots, n\} : x_i \leq x\}|}{n}$$

die *empirische Verteilungsfunktion* der Beobachtungswerte x_1, \dots, x_n .

Anmerkung: Im folgenden wird die empirische Verteilungsfunktion mit Hilfe des Wertebereichs ihrer Funktionswerte definiert. Zu diesem Zweck ordnet man die Beobachtungswerte ihrer Größe nach: Es bezeichne

$x_{1:n}$ den kleinsten von n Beobachtungswerten,
 $x_{2:n}$ den zweit-kleinsten Beobachtungswert,
 \dots \dots
 $x_{n:n}$ den größten Beobachtungswert.

Dann gelten $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$, $\{x_{1:n}, \dots, x_{n:n}\} = \{x_1, \dots, x_n\}$ und

$$F_n(x) = \begin{cases} 0 & \text{für } x < x_{1:n} \\ \frac{i}{n} & \text{für } x_{i:n} \leq x < x_{i+1:n}, \quad i \in \{1, \dots, n-1\} \\ 1 & \text{für } x_{n:n} \leq x \end{cases}$$

Anmerkung zur "Kurve von Quetelet"⁷

Der nachstehende Bericht des Mathematikers *B.L. van der Waerden* gibt einerseits einen interessanten Einblick in die Ideengeschichte der Statistik und andererseits eine Anregung, eine empirische Verteilungsfunktion plastisch darzustellen.

"Lebhaft erinnere ich mich noch, wie mein Vater mich als Knaben eines Tages an den Rand der Stadt führte, wo am Ufer die Weiden standen, und mich hundert Weidenblätter willkürlich pflücken ließ. Nach Aussonderung der beschädigten Spitzen blieben noch 89 unversehrte Blätter übrig, die wir dann zu Hause, nach abnehmender Größe geordnet, wie Soldaten in Reih und Glied stellten. Dann zog mein Vater durch die Spitzen eine gebogene Linie und sagte: "Dies ist die Kurve von QUETELET. Aus ihr siehst du, wie die Mittelmäßigen immer die große Mehrheit bilden und nur wenige nach oben hervorragen oder nach unten zurückbleiben." "

Beispiel 4: Zur "Kurve von Quetelet"

W. Rohm (HTL Saalfelden) und F. Österreicher haben 1982 gemäß *B.L. van der Waerden*s Beschreibung einer Föhre 100 Nadeln aufs Geratewohl entnommen und deren Längen (in mm) gemessen. Die geordneten Messergebnisse sind im nachstehenden Stängel-Blatt-Diagramm dargestellt.

7		044778
8		000134455677777888889999
9		000111112222333444445555566666777778899
10		001122222233333444556667777
11		0115

⁷Der belgische Statistiker *Lambert-Adolphe-Jacques Quetelet* (1796 – 1874) erschloss der Normalverteilung in der Anthropometrie ein gänzlich neues und unvermutetes Anwendungsgebiet und übte damit einen prägenden Einfluss auf *Francis Galton* (1822 – 1911) aus. Auf seinen Einfluss geht auch die Gründung vieler statistischer Behörden in Europa zurück.

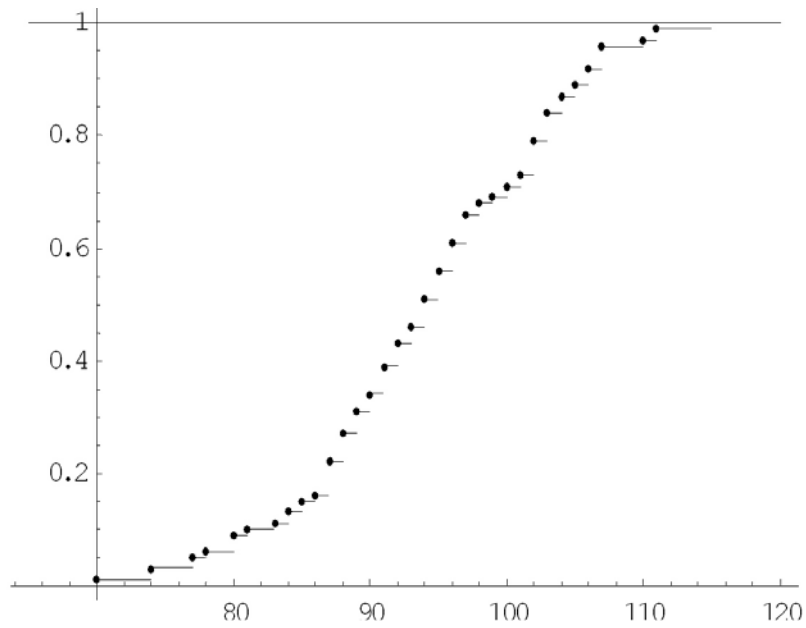


Abbildung: Zugehörige empirische Verteilungsfunktion

1.2 Kenngrößen eindimensionaler Daten: Zentral- und Streumaße

Wir gehen im folgenden - sofern nicht anders angenommen - von Daten x_1, \dots, x_n einer eindimensionalen Variablen (d.h. mit Werten aus \mathbb{R}), versehen mit der natürlichen Ordnung " $<$ " und dem üblichen Abstand, aus.

1.2.1 Stichprobenmittel und Standardabweichung

a) Das Stichprobenmittel

Das *arithmetische Mittel*

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

der Stichprobenwerte heißt *Stichprobenmittel* oder *Mittelwert*.

Anmerkung 1: Der britische Wissenschaftler *Thomas Simpson* (1710 – 1761) schlug für die Handhabung von mehrfachen, erfahrungsgemäß verschiedenen Messwerten einer Messgröße die Verwendung des Stichprobenmittels vor. Mit ein Motiv dafür war wohl, die Subjektivität der Beobachter beim Umgang mit solchen Daten hintanzuhalten und durch eine standardisierte Vorgangsweise den Austausch von Messergebnissen zu erleichtern. Er schreibt:

"Zusammenfassend scheint es, dass das Bestimmen des arithmetischen Mittels einer Anzahl von Messwerten die Chance kleiner Fehler beträchtlich verringert und nahezu jede Möglichkeit für große ausschließt. Diese Erwägung allein scheint ausreichend, um die Verwendung dieser Methode nicht nur Astronomen zu empfehlen, sondern allen, die Präzisionsmessungen durchführen. Je mehr Beobachtungen oder Experimente gemacht werden, desto weniger werden die Resultate fehleranfällig sein, vorausgesetzt, eine Wiederholung der Messungen ist unter gleichen Bedingungen möglich."

Anmerkung 2: Liegen bereits die relativen Häufigkeiten h_j der einzelnen Ausfälle ω_j , $j \geq 1$, der Variablen X vor, so ist es zweckmäßiger, das Stichprobenmittel gemäß

$$\bar{x}_n = \sum_{j \geq 1} \omega_j \cdot h_j$$

mit deren Hilfe auszudrücken. Falls möglich denken wir uns dabei die ω_j wieder der Größe nach geordnet: $\omega_1 < \omega_2 < \dots$.

Zu den beiden Darstellungen des Stichprobenmittels äußerte sich *Henri Lebesgue*⁸, der Schöpfer des nach ihm benannten Lebesgue-Integrals, welches in der höheren Analysis das *Riemann*-Integral ersetzt, wie folgt.

”Man kann auch sagen, dass man sich bei der Verwendung der ersten Berechnungsart wie ein Kaufmann ohne System verhält, der Geldstücke und Banknoten in der zufälligen Reihenfolge zählt, wie er sie in die Hand bekommt. Während wir bei der zweiten Rechenart vorgehen wie ein umsichtiger Kaufmann, der sagt:

Ich habe $H(E_1)$ Münzen zu einer Krone, macht $1 \times H(E_1)$,
 ich habe $H(E_2)$ Münzen zu zwei Kronen, macht $2 \times H(E_2)$,
 ich habe $H(E_5)$ Münzen zu fünf Kronen, macht $5 \times H(E_5)$, usw.;

ich habe also insgesamt

$$S = 1 \times H(E_1) + 2 \times H(E_2) + 5 \times H(E_5) + \dots,$$

weil er - wie reich er auch sein mag - nur eine endliche Anzahl von Banknoten zu zählen hat.”

Die **Minimaleigenschaft des Stichprobenmittels** bezüglich der Summe der Abweichungsquadrate

Behauptung: Die Funktion

$$f(x) = \frac{1}{n} \sum_{i=1}^n (x_i - x)^2$$

ist genau dann minimal, wenn $x = \bar{x}_n$ ist.

Beweisvariante 1: Die Ableitung von $f(x)$ ist

$$f'(x) = -\frac{2}{n} \sum_{i=1}^n (x_i - x) = 2(x - \bar{x}_n) \begin{cases} < 0 & \text{für } x < \bar{x}_n \\ = 0 & \text{für } x = \bar{x}_n \\ > 0 & \text{für } x > \bar{x}_n. \end{cases}$$

⁸ *Henri Lebesgue* (1875-1941), französischer Mathematiker

1.2. KENNGRÖSSEN EINDIMENSIONALER DATEN: ZENTRAL- UND STREUMASSE 15

Somit wird das Minimum im Punkt $x = \bar{x}_n$ angenommen.

Beweisvariante 2 bedient sich einer Aussage aus der Mechanik, welche trotz ihrer Einfachheit äußerst nützlich ist. Diese Aussage ist

Der Steinersche Verschiebungssatz⁹: Es gilt

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

Wegen $n(x - \bar{x}_n)^2 \geq 0$ ergibt sich daraus als **Folgerung** die gewünschte Aussage, nämlich

$$\sum_{i=1}^n (x_i - x)^2 \geq \sum_{i=1}^n (x_i - \bar{x}_n)^2,$$

wobei Gleichheit offensichtlich genau dann gilt, wenn $x = \bar{x}_n$ ist.

Anmerkung 3: Interpretiert man die ω_j , $j \geq 1$, als Punkte der x -Achse, in denen die Massen H_j sitzen, und lässt man diese Massen um eine Achse rotieren, die durch den Punkt x der x -Achse geht und normal auf dieser steht, dann ist die Summe

$$\sum_{i=1}^n (x_i - x)^2 = \sum_{j \geq 1} (\omega_j - x)^2 H_j$$

das zugehörige Trägheitsmoment der Massenverteilung¹⁰. Die Folgerung des Steinerschen Verschiebungssatzes besagt also, dass das Trägheitsmoment genau dann kleinstmöglich ist, wenn die Drehachse durch den Massenmittelpunkt \bar{x}_n der Massenverteilung geht.

Beweis des Satzes: Die Anwendung der unmittelbar einsichtigen (und auch geometrisch interpretierbaren) Beziehung

$$a^2 = b^2 + (a - b)^2 + 2(a - b)b$$

für $a = x_i - x$ und $b = x_i - \bar{x}_n$ ergibt

$$(x_i - x)^2 = (x_i - \bar{x}_n)^2 + (\bar{x}_n - x)^2 + 2(\bar{x}_n - x)(x_i - \bar{x}_n).$$

⁹parallel-axis theorem, *Jakob Steiner* (1796-1863), Schweizer Geometer

¹⁰Der Begriff des Trägheitsmoments wurde - wie der des Erwartungswerts in der Wahrscheinlichkeitsrechnung - vom holländischen Wissenschaftler *Christiaan Huygens* (1629 - 1695) geprägt. Er ist überdies der Erfinder der Pendeluhr, der Entdecker der Saturnringe und der Urheber des *Huygensschen Prinzips* in der Optik.

Daraus ergibt sich die Aussage des Verschiebungssatzes durch Summation über $i \in \{1, \dots, n\}$ und Berücksichtigen von $\sum_{i=1}^n (x_i - \bar{x}_n) = \sum_{i=1}^n x_i - n\bar{x}_n = 0$.

b) **Stichprobenvarianz und Standardabweichung**¹¹

Die Größe

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

heißt Stichprobenvarianz. Deren Quadratwurzel

$$s_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}$$

heißt *Standardabweichung der Stichprobe*. Nicht die Stichprobenvarianz, sondern die Standardabweichung ist die eigentliche Maßzahl für die Abweichung der Daten vom Mittelwert. Dies kann man sich etwa mit Hilfe einer Dimensionsbetrachtung überlegen, wie sie in der Physik üblich ist. Sind beispielsweise die x_i die Ergebnisse der Messungen einer Länge, so sind auch die Differenzen $x_i - \bar{x}_n$ Längen. Deren Quadrate $(x_i - \bar{x}_n)^2$ sind demnach die Flächen der zugehörigen Quadrate, sodass auch s_n^2 ein Maß für eine Fläche ist. Erst deren Quadratwurzel, die Standardabweichung s_n , hat wieder die richtige "Dimension", nämlich die einer Länge.

Warum dividiert man durch $n-1$ und nicht durch n ?

Vordergründige Antwort: Wegen der Bedingung $\sum_{i=1}^n (x_i - \bar{x}_n) = 0$ ist eine der Abweichungen $x_i - \bar{x}_n$ durch die restlichen $n-1$ Abweichungen festgelegt. Da also nur $n-1$ Summanden frei variieren können, dividiert man durch $n-1$. Einer physikalischen Tradition folgend nennt man diese Zahl auch die *Anzahl der Freiheitsgrade*.

Hintergründige Antwort: Dazu bedarf es eines stochastischen Modells. Sei X eine Zufallsvariable mit Erwartungswert μ und unbekannter Varianz σ^2 . Wäre nun der Erwartungswert μ bekannt, so würde man σ^2

¹¹Die Bezeichnung "standard deviation" (Standardabweichung) wurde Karl Pearson im Jahre 1893 geprägt. Der Begriff selbst wurde jedoch bereits vom deutschen Mathematiker Carl Friedrich Gauß (1777 – 1855) im Rahmen der Fehlerrechnung verwendet. Bei Gauß hieß diese Größe "mittlerer zu befürchtender Fehler".

1.2. KENNGRÖSSEN EINDIMENSIONALER DATEN: ZENTRAL- UND STREUMASSE¹⁷

naturgemäß durch

$$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

schätzen. Ist jedoch der Erwartungswert unbekannt, so hat man diesen durch seinen Schätzwert \bar{x}_n zu ersetzen. Gemäß der Minimaleigenschaft von \bar{x}_n gilt

$$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \geq \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 .$$

Damit schätzt man die Varianz "im Durchschnitt etwas zu kurz". Durch die Multiplikation mit dem Faktor $\frac{n}{n-1} > 1$ wird dieses Defizit behoben und man erhält den angegebenen Schätzwert

$$\frac{n}{n-1} \times \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 .$$

Es lässt sich zeigen, dass der zugehörigen Schätzer S_n^2 *erwartungstreu* (unverfälscht, oder - englisch - *unbiased*) ist¹². Dazu bedarf es freilich eines stochastischen Modells.

1.2.2 Stichprobenmedian und mittlere absolute Abweichung

a) Der Stichprobenmedian¹³

Zur Bestimmung des Stichprobenmedians ist es notwendig, die ursprünglichen Daten x_1, \dots, x_n der Größe nach zu ordnen. Seien also $x_{1:n}, \dots, x_{n:n}$ die der Größe nach geordneten Daten (somit ist $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ und es gilt $\{x_{1:n}, \dots, x_{n:n}\} = \{x_1, \dots, x_n\}$). Dann ist

$$\tilde{x}_n = \begin{cases} x_{(n+1)/2:n} & \text{für } n \text{ ungerade} \\ (x_{n/2:n} + x_{n/2+1:n})/2 & \text{für } n \text{ gerade} \end{cases}$$

¹²Der englische Statistiker *William Searly Gosset* (1876–1937), der unter dem Pseudonym "*Student*" für die Brauerei Guinness arbeitete, verwendete aus diesem Grund anstelle des von *Karl Pearson* benützten Nenners n den Nenner $n-1$. Dies veranlasste Karl Pearson zu der Äußerung: "Only naughty brewers take n so small that the difference is not of the order of the probable error!"

¹³Vom lateinischen Wort *medius* 3: *der, die, das Mittlere* abgeleitet

der *Median* der Daten. Dabei sind

$x_{(n+1)/2:n}$ der Wert in der Mitte der geordneten Datenliste,
 $(x_{n/2:n} + x_{n/2+1:n})/2$ das arithmetische Mittel der beiden Werte in der Mitte.

Anmerkung 1: Sofern n ungerade ist, verwendet man für die Definition des Stichprobenmedians nur die Ordnungsrelation " $<$ ". Demnach ist die Definition des Stichprobenmedians in diesem Fall auch für ordinale Variable - wie z.B. für Schulnoten - möglich.

Anmerkung 2: Sei $\lceil x \rceil$ die kleinste ganze Zahl $\geq x$, d.h. sei $\lceil x \rceil = \min \{ i \in \mathbb{Z} : i \geq x \}$. Mit Hilfe dieser Bezeichnung kann der Stichprobenmedian ohne Fallunterscheidung definiert werden:

$$\tilde{x}_n = (x_{\lceil n/2 \rceil:n} + x_{\lceil (n+1)/2 \rceil:n}) / 2.$$

Anmerkung 3: Wird im Fall $n \geq 3$ ein Beobachtungswert $x_i = x_{n:n}$ durch einen extrem großen Wert ersetzt, so hat dies keinen Einfluss auf den Stichprobenmedian, hingegen einen beträchtlichen auf das Stichprobenmittel. Man sagt: Der Stichprobenmedian ist *robust* (*resistent*) gegen *Ausreißer*. Das Stichprobenmittel hingegen ist *sensitiv* gegen Ausreißer.

Die **Minimaleigenschaft des Stichprobenmedians** hinsichtlich der Summe der Absolutbeträge der Abweichungen

Behauptung: Die Funktion

$$g(x) = \frac{1}{n} \sum_{i=1}^n |x_i - x|$$

ist genau dann minimal, wenn $x \in [x_{\lceil n/2 \rceil:n}, x_{\lceil (n+1)/2 \rceil:n}]$ ist.

Beweis: Dieser erfolgt durch Fallunterscheidung. Eine ausführliche Darstellung findet sich in [17], Abschnitt 1.4.2.

b) **Die mittlere absolute Abweichung**¹⁴ vom Stichprobenmedian ist

$$\tilde{s}_n = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}_n|.$$

¹⁴Der französische Mathematiker *Pierre Simon Laplace* (1749–1827) bevorzugte die mittlere absolute Abweichung, wogegen *Carl Friedrich Gauß* die Standardabweichung (unter der Bezeichnung *mittlerer zu befürchtenden Fehler*) vorzog.

1.2. KENNGRÖSSEN EINDIMENSIONALER DATEN: ZENTRAL- UND STREUMASSE 19

Diese ist ein mit der Standardabweichung vergleichbares Maß für die Abweichung der Beobachtungswerte vom zugehörigen Zentralmaß.

c) Das einfachste Streumaß ist die **Spannweite der Stichprobe**, nämlich die Differenz

$$x_{n:n} - x_{1:n}$$

zwischen Stichprobenmaximum und Stichprobenminimum.

In der nachstehenden Definition betrachten wir eine Verallgemeinerung des Medians.

Definition: Sei $\alpha \in (0, 1)$. Dann heißt

$$q_{\alpha,n} = \begin{cases} x_{[\alpha \cdot n]:n} & \text{für } \alpha \cdot n \notin \mathbb{N} \\ (1 - \alpha)x_{\alpha \cdot n:n} + \alpha x_{\alpha \cdot n + 1:n} & \text{für } \alpha \cdot n \in \mathbb{N} \end{cases}$$

das α -Quantil der Stichprobe.

Die Spezialfälle für $\alpha \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, nämlich das 1. Quartil, der Median und das 3. Quartil, lassen sich folgendermaßen ohne Fallunterscheidung definieren

$$\begin{aligned} \text{das 1. Quartil } q_{1/4,n} &= (3x_{[n/4]:n} + x_{[(n+1)/4]:n}) / 4 \\ \text{der Median } \tilde{x}_n = q_{1/2,n} &= (x_{[n/2]:n} + x_{[(n+1)/2]:n}) / 2 \\ \text{das 3. Quartil } q_{3/4,n} &= (x_{[3n/4]:n} + 3x_{[(3n+1)/4]:n}) / 4. \end{aligned}$$

Zusammen mit dem Stichprobenminimum $x_{1:n}$ und dem Stichprobenmaximum $x_{n:n}$ benötigt man diese zum Anfertigen eines Kasten-Bilds¹⁵.

Ein Streumaß, welches - im Unterschied zur Spannweite - robust gegen Ausreißer ist, ist der **Interquartilabstand**

$$q_{3/4,n} - q_{1/4,n}.$$

¹⁵Das *Kastenbild* (*box plot*) ist eine knappe visuelle Darstellung der Verteilung der Stichprobenwerte. Es stammt, wie das *Stängel-Blatt-Diagramm* aus Abschnitt 1.2.2, von John W. Tukey.

1.3 Weitere Mittelwerte

1.3.1 Das geometrische Mittel

Beispiel 1: Man bestimme die Wachstumsrate für das im Schaubild in *Küttings* Artikel [15] dargestellte Wirtschaftswachstum. Offensichtlich wächst die Wirtschaft im angegebenen Zeitraum von 6 Jahren mit dem Faktor

$$(1 + 0.044) \cdot (1 + 0.004) \cdot (1 - 0.02) \cdot (1 + 0.057) \cdot (1 + 0.026) \cdot (1 + 0.034) = 1.1519.$$

Um dieses Resultat bei einer für jedes Jahr gleichen Wachstumsrate r zu erreichen, muss gelten

$$(1 + r) \cdot (1 + r) \cdot (1 + r) \cdot (1 + r) \cdot (1 + r) \cdot (1 + r) = (1 + r)^6 = 1.1519$$

und somit

$$1 + r = \sqrt[6]{1.1519} = 1 + 0.0238.$$

Anmerkung 1: Es gilt

$$1 + \frac{0.044 + 0.004 - 0.02 + 0.057 + 0.026 + 0.034}{6} = 1 + 0.143 > 1 + 0.0238.$$

Anders ausgedrückt: Bei einem durchschnittlichen Prozentsatz von 14.3 % würde die Wirtschaft gemäß $1.143^6 = 2.2299$ mit dem Faktor 2.2299 wachsen!

Definition: Seien $x_1, \dots, x_n \geq 0$. Dann nennt man

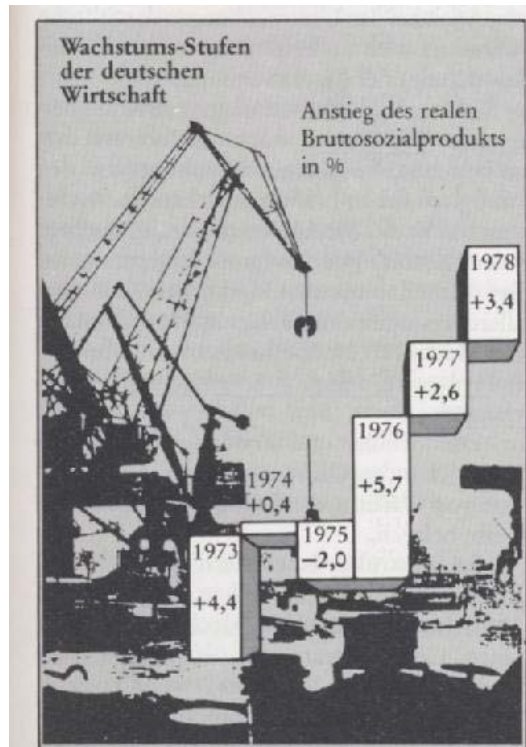
$$\hat{x}_n = \sqrt[n]{x_1 \cdot \dots \cdot x_n}$$

das *geometrische Mittel* der Zahlen x_1, \dots, x_n .

Bezeichnung: Für $x_i = 1 + r_i$, $i \in \{1, \dots, n\}$ nennt man die Größe

$$\hat{x}_n - 1 = \sqrt[n]{(1 + r_1) \cdot \dots \cdot (1 + r_n)} - 1$$

die *Wachstumsrate* (*growth rate*).



Anmerkung 2: Das geometrische Mittel eignet sich für Größen, die multiplikativ verknüpft werden. Typisch dafür ist der durchschnittliche *Aufzinsungsfaktor* in der Zinseszinsrechnung. Demgemäß wird es häufig als Durchschnittswert bei zeitlich aufeinanderfolgenden Veränderungsraten verwendet, also etwa bei jährlich erhobenen Kosten- und Preisindizes, den Quotienten der Börsenkurse von Ende zu Beginn eines Tages und den Wachstumsraten von Individuen und Populationen.

Behauptung 1: Es gilt

$$\sqrt[n]{\prod_{i=1}^n x_i} \leq \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$$

mit Gleichheit genau dann, wenn gilt $x_1 = \dots = x_n$.

Beweis: Siehe [17], Abschnitt 1.5.1.

Anmerkung 3: Wegen

$$(\sqrt[3]{a \cdot b \cdot c})^3 = a \cdot b \cdot c$$

ist das geometrische Mittel von a , b und c die Seitenlänge eines Würfels, welcher dasselbe Volumen besitzt, wie ein Quader mit den Seitenlängen a , b und c . Da das Volumen eines dreiaxigen Ellipsoids proportional zu $a \cdot b \cdot c$ ist, ist das geometrische Mittel $\sqrt[3]{a \cdot b \cdot c}$ somit ein besseres Maß für den mittleren Durchmesser eines Gerölls als das arithmetische Mittel $(a + b + c)/3$.

Wegen

$$\ln\left(\sqrt[n]{\prod_{i=1}^n x_i}\right) = \frac{1}{n} \sum_{i=1}^n \ln(x_i)$$

sind logarithmische Skalen mit dem geometrischen Mittel eng verknüpft.

Logarithmische Skalen sind in verschiedenen Wissenschaftsdiziplinen gebräuchlich, wie zum Beispiel die beiden physiologischen Größen Schallpegel (in Dezibel) und die scheinbare Helligkeit von Sternen (Magnitudo).¹⁶

Der *pH-Wert*

$$pH = -\log_{10} [H^+]$$

von wasserhaltigen Flüssigkeiten misst die Konzentration der H^+ -Ionen und gibt an, wie sauer oder basisch eine Lösung ist. So bedeuten

$$pH = 0 : \text{ extrem sauer} \quad pH = 7 : \text{ neutral} \quad pH = 14 : \text{ extrem basisch.}$$

Das Symbol "p" steht übrigens für "negativer dekadischer Logarithmus von" und ist somit ein Synonym für die logarithmische Skalierung.

Die in der Petrographie verwendete *Udden-Wentworth-Skala* wird zur Klassifikation von Korngrößen herangezogen. Ist d der Korndurchmesser (in mm), dann ist der zugehörige *Udden-Wentworth-Skalenwert* $\Phi(d)$ gegeben durch

$$\Phi(d) = -\log_2(d).$$

¹⁶Die physiologischen Größen Schallpegel und Helligkeit messen die durch die physikalischen Größen Schallintensität bzw. Strahlungsstrom beim Menschen hervorgerufene Empfindung. Dabei kommt das *Weber-Fechnersche Gesetz* der Psychophysik zur Anwendung, nach dem die Empfindungsstärke proportional zum Logarithmus der Reizstärke ist.

Millimeters	μm	Phi (ϕ)	Wentworth size class	
4096		-20	Boulder (-8 to -12 ϕ)	Gravel
1024		-12		
256		-10		
64		-8	Pebble (-6 to -8 ϕ)	
16		-6		
4		-4	Pebble (-2 to -6 ϕ)	
		-2		
3.36		-1.75	Gravel	
2.83		-1.50		
2.38		-1.25		
2.00		-1.00		
1.68		-0.75	Very coarse sand	
1.41		-0.50		
1.19		-0.25		
1.00		-0.00	Coarse sand	
0.84		0.25		
0.71		0.50		
0.59		0.75	Medium sand	
1/2 - 0.50	500	1.00		
0.42	420	1.25		
0.35	350	1.50		
0.30	300	1.75	Fine sand	
1/4 - 0.25	250	2.00		
0.210	210	2.25		
0.177	177	2.50	Very fine sand	
0.149	149	2.75		
1/8 - 0.125	125	3.00		
0.105	105	3.25	Coarse silt	
0.088	88	3.50		
0.074	74	3.75		
1/16 - 0.0625	63	4.00	Medium silt	
0.0530	53	4.25		
0.0440	44	4.50		
0.0370	37	4.75	Fine silt	
1/32 - 0.0310	31	5		
0.020	20	6		
1/64 - 0.0156	15.6	7	Very fine silt	
1/128 - 0.0078	7.8	8		
1/256 - 0.0039	3.9	9		
0.0020	2.0	10	Clay	
0.00098	0.98	11		
0.00049	0.49	12		
0.00024	0.24	13		
0.00012	0.12	14		
0.00006	0.06	14		

Udden-Wentworth-Scale grain-size scale

1.3.2 Das harmonische Mittel

Beispiel 2: Die ersten 100 km legt ein Zug mit einer konstanten Geschwindigkeit von 70 km/h zurück, die zweiten 100 km mit einer konstanten Geschwindigkeit von 110 km/h. Mit welcher Durchschnittsgeschwindigkeit fährt der Zug?

Bezeichnen $s_1, s_2; v_1, v_2; t_1, t_2$ Weg, Geschwindigkeit und Zeit der jeweiligen Teilstrecken, und \tilde{v} die gesuchte Durchschnittsgeschwindigkeit, dann gilt wegen der Beziehungen $s = \tilde{v} \cdot t$ (Weg = Geschwindigkeit \times Zeit), $s = s_1 + s_2$ und $t = t_1 + t_2$

$$\begin{aligned} \tilde{v} &= \frac{s}{t} = \frac{s_1 + s_2}{t_1 + t_2} = \frac{s_1 + s_2}{\frac{s_1}{v_1} + \frac{s_2}{v_2}} \\ &= \frac{1}{\frac{s_1}{s_1 + s_2} \cdot \frac{1}{v_1} + \frac{s_2}{s_1 + s_2} \cdot \frac{1}{v_2}}. \end{aligned}$$

Die numerische Lösung ist daher $\tilde{v} = \frac{1}{\frac{1}{2} \cdot \frac{1}{70} + \frac{1}{2} \cdot \frac{1}{110}} = 85,56 \text{ km/h}$.

Anmerkung 1: Die durchschnittliche Geschwindigkeit wäre 90 km/h, das geometrische Mittel der beiden Geschwindigkeiten wäre 87,75 km/h!

Definition: Seien $x_1, \dots, x_n > 0$. Dann nennt man

$$\frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}}$$

das *harmonische Mittel* der Zahlen x_1, \dots, x_n .

Anmerkung 2: Das harmonische Mittel ist der Kehrwert des arithmetischen Mittels der Kehrwerte der x_i . Es eignet sich als Mittelwert von Größen, die indirekt proportional zu anderen Größen sind. Anwendungsmöglichkeiten sind z.B.: Absolutbetrag der Krümmung (indirekt proportional zum Krümmungsradius, $|\kappa(\rho)| = 1/\rho$), Frequenz (indirekt proportional zur Wellenlänge, $\nu(\lambda) = 1/\lambda$), Geschwindigkeit (indirekt proportional zur Zeit, $v(t) = \frac{s}{t}$), Dichte (indirekt proportional zum Volumen, $\rho(V) = \frac{m}{V}$), Stromstärke (indirekt proportional zum Widerstand, $I(R) = \frac{U}{R}$), Druck eines idealen Gases (indirekt proportional zum Volumen, $p(V) = \frac{RT}{V}$).

Behauptung 2: Es gilt

$$\frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} \leq \sqrt[n]{\prod_{i=1}^n x_i}$$

mit Gleichheit genau dann, wenn gilt $x_1 = \dots = x_n$.

Beweis: Diese Ungleichung ist gleichbedeutend mit

$$\sqrt[n]{\prod_{i=1}^n \frac{1}{x_i}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i},$$

was identisch mit der Ungleichung zwischen dem geometrischen und dem arithmetischen Mittel der Kehrwerte $y_i = \frac{1}{x_i}$, $i \in \{1, \dots, n\}$ ist. \square

Aus den Behauptungen 2 und 3 ergibt sich zusammenfassend

Behauptung 3: Für $x_i \in (0, \infty)$, $i \in \{1, \dots, n\}$ gilt

$$\min(x_1, \dots, x_n) \leq \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} \leq \hat{x}_n \leq \bar{x}_n \leq \max(x_1, \dots, x_n)$$

mit Gleichheit jeweils genau dann, wenn gilt $x_1 = \dots = x_n$.

1.3.3 Der Spezialfall $n=2$

Für $0 < x < y < \infty$ bezeichne $a = a(x, y)$, $g = g(x, y)$ und $h = h(x, y)$ das arithmetische, geometrische bzw. harmonische Mittel von x und y . Der nachstehenden Abbildung entnimmt man - mittels der Ähnlichkeit jeweils zweier Dreiecke -

$$\frac{x}{g} \quad (= \tan \alpha) \quad = \quad \frac{g}{y} \quad \text{und daher} \quad g^2 = x \cdot y \quad \dots \quad (1_\alpha)$$

und

$$\frac{h}{g} \quad (= \cos \beta) \quad = \quad \frac{g}{a} \quad \text{und daher} \quad g^2 = h \cdot a \quad \dots \quad (1_\beta)$$

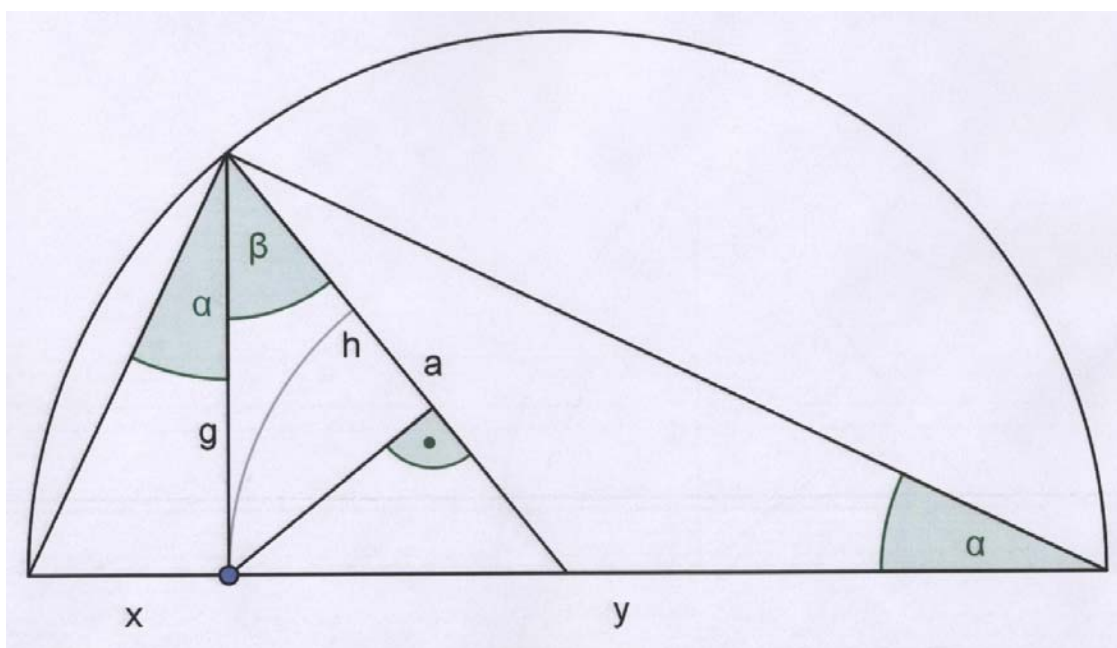


Abbildung: arithmetisches, geometrisches und harmonisches Mittel von x und y

Also gilt

$$h(x, y) \cdot a(x, y) = xy \quad (2)$$

und somit im Hinblick auf (1_α) $g(x, y) = \sqrt{xy}$ und wegen $a(x, y) = \frac{x+y}{2}$ im Hinblick auf (2)

$$h(x, y) = \frac{1}{\frac{1}{2}(\frac{1}{x} + \frac{1}{y})}.$$

Ferner entnimmt man der Abbildung die Gültigkeit der Ungleichungskette

$$x < a(x, y) < g(x, y) < a(x, y) < y. \quad (3)$$

Für den Fall $x = y$ gilt offensichtlich überall das Gleichheitszeichen.

1.4 Lineare Regression und Korrelation

1.4.1 Aufgabenstellung und einführendes Beispiel

Gegeben seien Wertepaare $\{(x_1, y_1), \dots, (x_n, y_n)\} \subseteq \mathbb{R}^2$, derart, dass weder alle x_i noch alle y_i gleich sind. Gesucht ist eine Ausgleichsgerade, d.h. die Gleichung

$$y = kx + d \quad \text{bzw.} \quad x = k'y + d'$$

einer Geraden, die die gegebene "Punktwolke" möglichst gut beschreibt.

Beispiel 1: Case Study 10.2.1 aus [?]

Präzisierung der Zielfunktion: Eine formale Behandlung der Aufgabenstellung setzt voraus, dass wir zunächst zwei Fragen klären.

Frage 1: Welche Abstände von Punkten und Geraden werden gemessen?

Variante 1: Vertikale Abstände $y_i - (k x_i + d)$, $i \in \{1, \dots, n\}$. Durch diese Wahl zeichnet man die Variable X als *unabhängige Variable* und die Variable Y als *abhängige Variable* aus.

Variante 1': Horizontale Abstände $x_i - (k' y_i + d')$, $i \in \{1, \dots, n\}$. Durch diese Wahl zeichnet man die Variable Y als *unabhängige Variable* und die Variable X als *abhängige Variable* aus.

Variante 2: Normalabstände

Frage 2: Wie werden die Abstände von Punkten und Geraden gemessen?

Variante 1: Durch die Abstandsquadrate, etwa $(y_i - (k x_i + d))^2$, $i \in \{1, \dots, n\}$.

Variante 2: Durch die Absolutbeträge der Abweichungen.

Als Zugeständnis an die mathematische Bequemlichkeit schließen wir die beiden Varianten 2 aus. Darüber hinaus beschränken wir uns auf die Behandlung der Vertikalabstände, da die der Horizontalabstände lediglich auf eine Vertauschung der beiden Variablen X und Y hinausläuft.

Demnach haben wir die Absicht, der Liste von Paaren $\{(x_1, y_1), \dots,$

$(x_n, y_n)\} \subseteq \mathbb{R}^2$ eine Gerade $y = kx + d$ durch geeignete Wahl der Parameter k und d so anzupassen, dass die Zielfunktion

$$\sum_{i=1}^n (y_i - (kx_i + d))^2$$

minimal wird. Die so bestimmte Gerade heißt **Regressionsgerade**. Die verwendete Methode nennt man *die Methode der kleinsten Quadrate*^{17, 18}.

1.4.2 Herleitung der Gleichung der homogenen Regressionsgeraden

Gegeben seien Wertepaare $\{(x_1, y_1), \dots, (x_n, y_n)\} \subseteq \mathbb{R}^2$ derart, dass weder alle x_i noch alle y_i gleich 0 sind.

Gesucht ist die sogenannte *homogene Regressionsgerade*, das ist jene Gerade $y = kx$ durch den Koordinatenursprung, die die gegebene "Punktwolke" im obigen Sinn möglichst gut beschreibt. Demgemäß werden wir den Anstieg k der Geraden so wählen, dass die Summe der Quadrate der sogenannten *Residuen*

$$y_i - kx_i, \quad i \in \{1, \dots, n\},$$

möglichst klein ist.

1. Lösungsvariante: Durch Differenzieren

¹⁷Die Methode der kleinsten Quadrate wurde vom französischen Mathematiker *Adrien Marie Legendre* (1752 – 1833) in seiner Arbeit "*Methode de la Moindre Quaree*" 1805 publiziert. Unabhängig von diesem wurde sie auch von *Carl Friedrich Gauß* entwickelt und in der Astronomie und der Geodesie angewandt.

¹⁸Die beiden Begriffe "*Regression*" und "*Korrelation*" wurden vom englischen Statistiker *Sir Francis Galton* (1822-1911) im Rahmen seiner Studien zur Vererbung geprägt. Das Wort "*Regression*" wurde von ihm im Zusammenhang mit "regression towards mediocrity", also dem Rückschritt zum Mittelmaß verwendet. Das Wort "*Korrelation*" könnte man mit "wechselseitiger Beziehung" übersetzen.

Galton war äußerst vielseitig: Er ist Schöpfer des nach ihm benannten *Galtonbretts*, der Urheber des Einsatzes von Fingerabdrücken zur Identifikation von Personen in der Kriminologie und der Entdecker der Antizyklogen in der Meteorologie.

Die Hälfte der Ableitung unserer Zielfunktion

$$f(k) = \sum_{i=1}^n (kx_i - y_i)^2$$

ist

$$\frac{1}{2}f'(k) = \sum_{i=1}^n x_i \cdot (kx_i - y_i) = k \cdot \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \cdot y_i = \sum_{i=1}^n x_i^2 \cdot (k - \hat{k})$$

mit $\hat{k} = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sum_{i=1}^n x_i^2}$. Wegen

$$f'(k) \begin{cases} < 0 & \text{für } k < \hat{k} \\ = 0 & \text{für } k = \hat{k} \\ > 0 & \text{für } k > \hat{k} \end{cases}$$

nimmt $f(k)$ für $k = \hat{k}$ sein Minimum an.

2. Lösungsvariante: Durch quadratisches Ergänzen

Ausquadrieren der Summanden der Funktion $f(k)$ und quadratisches Ergänzen in der folgenden Form

$$A \cdot k^2 + B \cdot k + C = A(k + \frac{B/2}{A})^2 + C(1 - \frac{(B/2)^2}{AC})$$

ergibt

$$\begin{aligned} f(k) &= \sum_{i=1}^n (k \cdot x_i - y_i)^2 \\ &= \sum_{i=1}^n (k^2 x_i^2 - 2k x_i y_i + y_i^2) \\ &= k^2 \sum_{i=1}^n x_i^2 - 2k \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 \\ &= \sum_{i=1}^n x_i^2 (k - \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2})^2 + \sum_{i=1}^n y_i^2 (1 - \frac{(\sum_{i=1}^n x_i y_i)^2}{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i^2}). \end{aligned}$$

Anmerkung 1: Sei nun \hat{k} wie oben definiert und

$$\hat{r} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i^2}}$$

der sogenannte *homogene Korrelationskoeffizient* (der Datenmenge $\{(x_1, y_1), \dots, (x_n, y_n)\}$). Dann nimmt die Funktion $f(k)$ ihr Minimum

$$f(\hat{k}) = \sum_{i=1}^n (y_i - \hat{k} \cdot x_i)^2 = \sum_{i=1}^n y_i^2 (1 - \hat{r}^2)$$

offensichtlich genau dann an, wenn $k = \hat{k}$ ist. Die Größe \hat{r}^2 heißt *homogenes Bestimmtheitsmaß* und ist eine Maß für die Güte der Anpassung, denn $f(\hat{k})$ ist als Summe von Quadraten nichtnegativ und genau dann gleich Null, wenn $y_i = \hat{k} \cdot x_i \quad \forall i \in \{1, \dots, n\}$ gilt. Dies ist gleichbedeutend damit, dass

$$\hat{r}^2 \leq 1$$

ist, wobei $\hat{r}^2 = 1$ genau dann gilt, wenn alle Punkte der Punktwolke auf einer Geraden durch den Ursprung liegen.

Für den homogenen Korrelationskoeffizienten \hat{r} gilt demnach

$$-1 \leq \hat{r} \leq 1,$$

wobei Gleichheit genau im eben angesprochenen Fall besteht. Sofern $\hat{r} > 0$ ist, ist der Anstieg \hat{k} der Regressionsgeraden positiv; sofern $\hat{r} < 0$ ist, ist dieser negativ.

1.4.3 Anwendungsbeispiele

Beispiel 2¹⁹: Zur Schätzung von π

Gegeben seien Messungen (d_i, u_i) , $i \in \{1, \dots, n\}$, von Durchmesser und Umfang zylindrischer Gefäße mit kreisförmiger Grundfläche. Seien $\frac{u_i}{d_i}$, $i \in \{1, \dots, n\}$, die Verhältnisse von Kreisumfang und Kreisdurchmesser. Dann

¹⁹Die Zahl π wird in den Anwendungen gelegentlich nach dem deutschen Mathematiker *Ludolf van Ceulen* (1540 – 1610) Ludolfsche Zahl genannt.

ist - wegen des bekannten linearen Zusammenhangs $u = d \cdot \pi$ zwischen Kreisumfang und Kreisdurchmesser - das arithmetische Mittel

$$\hat{\pi}_1 = \frac{1}{n} \sum_{i=1}^n \frac{u_i}{d_i}$$

ein naheliegender Schätzwert für π .

Wendet man hingegen die Methode der kleinsten Quadrate auf die Punktwolke (d_i, u_i) , $i \in \{1, \dots, n\}$, an, so erhält man nicht bloß den alternativen Schätzwert

$$\begin{aligned} \hat{\pi}_2 &= \frac{\sum_{i=1}^n d_i u_i}{\sum_{i=1}^n d_i^2} \\ &= \sum_{i=1}^n \frac{d_i^2}{\sum_{j=1}^n d_j^2} \frac{u_i}{d_i} \end{aligned}$$

für π , sondern zudem eine empirische Bestätigung des obigen linearen Zusammenhangs zwischen Kreisumfang und Kreisdurchmesser. Aus der zweiten Darstellung erkennt man, dass dieser Schätzwert ein mittels der Gewichte $\frac{d_i^2}{\sum_{j=1}^n d_j^2}$ gewichtetes Mittel der Quotienten $\frac{u_i}{d_i}$, $i \in \{1, \dots, n\}$, ist. Durch diese Gewichte werden Quotienten mit großem Durchmesser sehr viel stärker bewertet werden als solche mit kleinem. Dieser Sachverhalt ist verträglich mit der Vorstellung, dass Messergebnisse von größeren Gefäßen genauer sind als solche von kleinen.

Beispiel 3: Keplers drittes Gesetz²⁰

Dieses lautet

Das Quadrat der Umlaufzeit eines Planeten
ist proportional zum Kubus der großen Halbachse seiner Bahn.

In Zeichen

$$U^2 = k \cdot a^3$$

²⁰Keplers erstes Gesetz lautet bekanntlich: Die Planeten bewegen sich auf Ellipsen, in deren gemeinsamen Brennpunkt die Sonne steht.

In seinem 1604 veröffentlichten Werk *Paralipomena*, welches er auch einfach *Optica* nannte, beschäftigte sich *Johannes Kepler* (1571 – 1630) auch mit Linsen, deren Krümmungskurven die Form von allgemeinen Kegelschnitten besitzen. In diesem Zusammenhang prägte er den Begriff *Brennpunkt*.

Die Daten der jeweiligen Planeten sind wie folgt

Planet	große Halbachse	siderische Umlaufzeit
Merkur	0.3870938	0.2408
Venus	0.7233276	0.6152
Erde	1.0009071	1
Mars	1.5237020	1.8808
Jupiter	5.2026255	11.8618
Saturn	9.5402393	29.4566
Uranus	19.2685300	84.0120
Neptun	30.2080026	164.7819
Pluto	39.8397880	247.6867

In moderner Darstellung ist die Beziehung zwischen großer Halbachse a und Umlaufzeit U

$$U = c \cdot a^{\frac{3}{2}}.$$

In dieser Form ist es jedoch für die Anwendung der homogenen linearen Regression nicht brauchbar. Um letztere anwenden zu können, ist es zweckmäßig, auf die ursprüngliche Formulierung des Gesetzes zurückzugreifen und die Größen a^3 und U^2 gegeneinander aufzutragen.

Der Schätzwert für die Konstante k ist

$$\hat{k} = \frac{\sum_{i=1}^9 a_i^3 u_i^2}{\sum_{i=1}^9 a_i^6} = .97271.$$

Das Bestimmtheitsmaß ist

$$\hat{r}^2 = \frac{(\sum_{i=1}^9 a_i^3 u_i^2)^2}{\sum_{i=1}^9 a_i^6 \cdot \sum_{i=1}^9 u_i^4} = .99997.$$

Damit ist die Gültigkeit des Gesetzes denkbar gut bestätigt.

1.4.4 Herleitung der Gleichung der allgemeinen Regressionsgeraden

Gegeben seien Wertepaare $\{(x_1, y_1), \dots, (x_n, y_n)\} \subseteq \mathbb{R}^2$, derart, dass weder alle x_i noch alle y_i gleich sind.

Gesucht ist die Regressionsgerade, d.h. die Gleichung $y = kx + d$ jener Geraden, die diese "Punktwolke" im obigen Sinn möglichst gut beschreibt.

Demgemäß haben wir die Parameter k und d so zu wählen, dass die Summe der Quadrate der Residuen

$$y_i - (kx_i + d), \quad i \in \{1, \dots, n\}$$

möglichst klein ist. Der erste Schritt, die Zielfunktion

$$f(d, k) = \sum_{i=1}^n (y_i - kx_i - d)^2$$

zu minimieren, besteht darin, den Spezialfall

$$\sum_{i=1}^n z_i^2 = \sum_{i=1}^n (z_i - \bar{z}_n)^2 + n \cdot \bar{z}_n^2$$

des Steinerschen Verschiebungssatzes auf die Größen

$$z_i = y_i - kx_i - d$$

anzuwenden und $\bar{z}_n = \bar{y}_n - k\bar{x}_n - d$ zu berücksichtigen. Auf diese Weise erhält man

$$\sum_{i=1}^n (y_i - kx_i - d)^2 = \sum_{i=1}^n (y_i - \bar{y}_n - k(x_i - \bar{x}_n))^2 + n(\bar{y}_n - k\bar{x}_n - d)^2.$$

Für den - zunächst noch vom Anstieg k abhängigen - Ordinatenabschnitt $\tilde{d} = \tilde{d}(k) = \bar{y}_n - k\bar{x}_n$ gilt demnach

$$f(d, k) = f(\tilde{d}, k) + n(d - \tilde{d})^2 \geq f(\tilde{d}, k)$$

mit Gleichheit genau dann, wenn $d = \tilde{d}$ ist.

Die neue Zielfunktion $g(k) = f(\tilde{d}(k), k) = \sum_{i=1}^n (y_i - \bar{y}_n - k(x_i - \bar{x}_n))^2$ zu minimieren, bedeutet jedoch, der Punktwolke $\{(x'_i, y'_i) = (x_i - \bar{x}_n, y_i - \bar{y}_n), i \in \{1, \dots, n\}\}$ eine Gerade durch den Koordinatenursprung anzupassen. Somit können wir das Resultat aus dem vorangehenden Abschnitt auf die transformierte Punktwolke anwenden.

Mit Hilfe der Stichprobenvarianzen der x - und y -Werte und der Kovarianz der Stichprobe

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2, \\ s_y^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2 \quad \text{und} \\ s_{xy} &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n) \cdot (y_i - \bar{y}_n) \end{aligned}$$

lassen sich der entsprechende Anstieg der Regressionsgeraden und der entsprechende Korrelationskoeffizient folgendermaßen ausdrücken

$$\tilde{k} = \frac{s_{xy}}{s_x^2} \quad \text{und} \quad \tilde{r} = \frac{s_{xy}}{s_x \cdot s_y}.$$

Für diese Größen gilt

$$\begin{aligned} g(k) &= \sum_{i=1}^n (y_i - \bar{y}_n - k(x_i - \bar{x}_n))^2 \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 (k - \tilde{k})^2 + \sum_{i=1}^n (y_i - \bar{y}_n)^2 (1 - \tilde{r}^2) \\ &\geq \sum_{i=1}^n (y_i - \bar{y}_n)^2 (1 - \tilde{r}^2) \end{aligned}$$

mit Gleichheit genau dann, wenn $k = \tilde{k}$ ist.

Die Gleichung der *Regressionsgeraden der Variablen Y bezüglich der unabhängigen Variablen X* ist daher

$$y = \bar{y}_n + \frac{s_{xy}}{s_x^2} (x - \bar{x}_n).$$

Der zugehörige Ordinatenabschnitt ist somit $\tilde{d}(\tilde{k}) = \bar{y}_n - \frac{s_{xy}}{s_x^2} \bar{x}_n$.

Anmerkung: Die Größe \tilde{r}^2 lässt sich mit Hilfe der durch das lineare Modell für die gegebenen Werte x_i der unabhängigen Variablen X für die abhängige Variable Y vorhergesagten Werte

$$\tilde{y}_i = \bar{y}_n + \frac{s_{xy}}{s_x^2} (x_i - \bar{x}_n), \quad i \in \{1, \dots, n\},$$

gemäß

$$\frac{1}{n-1} \sum_{i=1}^n (\tilde{y}_i - \bar{y}_n)^2 = \frac{s_{xy}^2}{(s_x^2)^2} \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{s_{xy}^2}{s_x^2}$$

folgendermaßen ausdrücken

$$\tilde{r}^2 = \frac{\sum_{i=1}^n (\tilde{y}_i - \bar{y}_n)^2}{\sum_{i=1}^n (y_i - \bar{y}_n)^2}.$$

Dies ermöglicht die

Interpretation: \tilde{r}^2 ist der Anteil der Variation in der abhängigen Variablen, der durch das lineare Modell erklärt wird.

Variante 1': Indem man die Variablen X und Y vertauscht, erhält man die Gleichung der *Regressionsgeraden von X bezüglich der unabhängigen Variablen Y* , nämlich

$$x = \bar{x}_n + \frac{s_{yx}}{s_y^2} (y - \bar{y}_n) = \bar{x}_n + \frac{s_{xy}}{s_y^2} (y - \bar{y}_n) .$$

In ihrer üblichen Form

$$y_{X(Y)}(x) = \bar{y}_n + \frac{s_y^2}{s_{xy}} (x - \bar{x}_n)$$

besitzt diese den Anstieg $\tilde{k}_{X(Y)} = \frac{s_y^2}{s_{xy}}$, während die *Regressionsgerade von Y bezüglich der unabhängigen Variablen X* den Anstieg $\tilde{k}_{Y(X)} = \frac{s_{xy}}{s_x^2}$ besitzt.

Für das *Bestimmtheitsmaß* $\tilde{r}^2 = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2}$ der Korrelation, welches - wie im homogenen Fall - ein Maß für die Stärke des linearen Zusammenhanges von X und Y ist, gilt im Hinblick auf die Anstiege $\tilde{k}_{Y(X)}$ und $\tilde{k}_{X(Y)}$ folgende Aussage

$$\tilde{r}^2 \leq 1 \iff \begin{cases} \frac{s_{xy}}{s_x^2} \leq \frac{s_y^2}{s_{xy}} & \text{für } s_{xy} > 0 \\ y = \bar{y}_n \text{ und } x = \bar{x}_n & \text{für } s_{xy} = 0 \\ \frac{s_{xy}}{s_x^2} \geq \frac{s_y^2}{s_{xy}} & \text{für } s_{xy} < 0 . \end{cases}$$

Der Fall $\tilde{r}^2 = 1$ ist mit jeder der beiden folgenden Aussagen gleichbedeutend.

- Alle Punkte (x_i, y_i) liegen auf einer Geraden.
- Die beiden Regressionsgeraden fallen zusammen.

Der *Pearsonsche Korrelations-Koeffizient* $\tilde{r} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$ ist naturgemäß ein Wert aus $[-1, 1]$. Er bleibt selbstverständlich unverändert, wenn die Maßeinheiten in den beiden Variablen geändert werden. Positive Korrelation liegt vor, wenn zu großen Werten der einen Variablen auch große Werte der

anderen Variablen gehören; negative Korrelation, wenn zu großen Werten der einen Variablen kleine Werte der anderen Variablen gehören. Im Fall $\tilde{r} = 0$ nennt man die beiden Variablen *unkorreliert*. Das bedeutet, dass zwischen den beiden Variablen kein linearer Zusammenhang existiert. Dabei liegt jedoch die Betonung auf dem Wort **linear**, denn $\tilde{r} = 0$ schließt einen anderen funktionellen Zusammenhang nicht aus! Ist nämlich ein Zusammenhang durch eine gerade Funktion²¹ gegeben, so kann ohne Weiters $\tilde{r} = 0$ gelten.

²¹Eine solche ist durch die Eigenschaft $f(-x) = f(x)$ charakterisiert. Typische Beispiele sind die Polynome $f(x) = x^{2n}$, $n \in \mathbb{N}$.

Chapter 2

WAHRSCHEINLICKEITSRECHNUNG

2.1 Die Binomialverteilung

Eine Urne enthalte N Kugeln, davon r rote und $w = N - r$ weiße, welche - von ihrer Farbe abgesehen - ununterscheidbar sind. Es werden n Kugeln zufällig und mit Zurücklegen gezogen. S_n sei die Anzahl der roten Kugeln in der *Stichprobe*.

Die möglichen Werte k der Zufallsgröße S_n sind $k \in \{0, \dots, n\}$. ($S_n = 0$ bedeutet, dass die Stichprobe keine einzige rote Kugel enthält, $S_n = n$ bedeutet, dass alle n gezogenen Kugeln rot sind.)

Bezeichnet $p = \frac{r}{N}$ den Anteil der roten Kugeln in der Urne und $q = 1 - p = \frac{w}{N}$ den Anteil der weißen Kugeln in der Urne, so ist die Wahrscheinlichkeit des Ereignisses $S_n = k$ (d.i. das Ereignis, dass die Stichprobe genau k rote Kugeln enthält)

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}, \quad k \in \{0, \dots, n\}.$$

Die durch diese Formel gegebene Wahrscheinlichkeitsverteilung heißt *Binomialverteilung* mit den Parametern n und p .

Bezeichnung: Das mathematische Symbol $\binom{n}{k}$ - gelesen "n über k" - heißt *Binomialkoeffizient*. Mithilfe der Größe

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

- n -Faktorielle oder n -Fakultät genannt - ist der Binomialkoeffizient $\binom{n}{k}$ folgendermaßen definiert

$$\binom{n}{k} := \frac{n!}{k! \cdot (n-k)!}.$$

Beispielsweise ist

$$\binom{5}{3} = \frac{5!}{3! \times 2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \times 2 \cdot 1} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10.$$

Aus diesem Beispiel erkennt man, dass man $\binom{n}{k}$ gemäß

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-(k-1))}{k!}$$

mit etwas weniger Rechenaufwand berechnen kann.

Anmerkung 1: Man erinnere sich an die aus der Schule geläufigen Spezialfälle des *Binomischen Lehrsatzes*

$$(a+b)^2 = a^2 + 2ab + b^2$$

und

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

Die Koeffizienten sind die Binomialkoeffizienten. So ergibt sich beispielsweise für die vierte Potenz

$$\begin{aligned} (b+a)^4 &= \binom{4}{0} b^0 a^{4-0} + \binom{4}{1} b^1 a^{4-1} + \binom{4}{2} b^2 a^{4-2} + \binom{4}{3} b^3 a^{4-3} + \binom{4}{4} b^4 a^{4-4} \\ &= 1 \cdot a^4 + 4 \cdot a^3 b + 6 \cdot a^2 b^2 + 4 \cdot a b^3 + 1 \cdot b^4 \\ &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4. \end{aligned}$$

Allgemein ergibt sich für $a = p$ und $b = q$ wegen $p+q=1$

$$1 = (p+q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}$$

und das ist ein durchaus verständlicher Sachverhalt: Die Summe aller Wahrscheinlichkeiten ist 1.

Prominente Anwendungen der Binomialverteilungen sind in der Genetik zu finden.

Vererbungsmechanismus nach Gregor Mendel (1822–1884): Mendel führte bekanntlich Kreuzungsversuche mit Erbsenpflanzen durch. Diese haben

Genotyp A ... violett-rote Blüten und graubraune Samenschalen
 Genotyp a ... weiße Blüten und weiße Samenschalen,

wobei, wie Mendel sagte, der Genotyp A dominant und der Genotyp a rezessiv ist. Jede Pflanze besitzt je ein Allel von jedem der beiden Elternpflanzen. Die möglichen Kombinationen sind somit (A, A) , (A, a) , (a, A) und (a, a) :

Allelpaar des Genoms	Erbigkeit	Phänotyp
(A, A)	... reinerbig dominant	A
$(A, a), (a, A)$... mischerbig	A
(a, a)	... reinerbig rezessiv	a

Wir gehen im Folgenden davon aus, dass wir zwei mischerbige Erbsenpflanzen kreuzen, nämlich

$$(A, a) \times (A, a)$$

Mendel hatte nachweislich Kenntnisse der Wahrscheinlichkeitsrechnung und schloss - vermutlich aufgrund der Binomischen Formel -

$$\begin{aligned} (A + a)^2 &= A^2 + 2Aa + a^2 \\ &= A \times A + 2 \cdot A \times a + a \times a \end{aligned}$$

auf folgenden Vererbungsmechanismus: Die Tochterpflanze erhält von jeder der beiden Elternpflanzen ein Allel, wobei dessen Auswahl rein zufällig erfolgt: Sie erhält mit jeweils Wahrscheinlichkeit $1/2$ die Allele A und a , so als würde aus zwei Urnen, welche je eine rote Kugel und eine weiße Kugel enthalten, jeweils eine Kugel gezogen. Ist S_2 die Anzahl der gezogenen roten Kugeln, dann sind

$$\begin{aligned} P(S_2 = 0) &= \binom{2}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^2 = \binom{2}{0} \frac{1}{2^2} = \frac{1}{4}, \\ P(S_2 = 1) &= \binom{2}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^1 = \binom{2}{1} \frac{1}{2^2} = \frac{1}{2}, \\ P(S_2 = 2) &= \binom{2}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = \binom{2}{2} \frac{1}{2^2} = \frac{1}{4}. \end{aligned}$$

Da in den Fällen $S_2 = 1$ und $S_2 = 2$ die Tochterpflanze den Phänotyp A besitzt, ist die W-keit, dass sie violett-rote Blüten und graubraune Samenschalen besitzt, gleich

$$P(S_2 \geq 1) = P(S_2 = 1) + P(S_2 = 2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

Beispiel 2: Erzeugt man durch Kreuzung zweier mischerbiger Erbsenpflanzen n Tochterpflanzen, so ist die Anzahl X_n der Tochterpflanzen mit violett-roten Blüten und graubraunen Samenschalen Binomialverteilt mit den Parametern n und $p = 3/4$, d.h. es gilt

$$P(X_n = k) = \binom{n}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{n-k} = \binom{n}{k} \frac{3^k}{4^n}, \quad k \in \{0, \dots, n\}.$$

Anmerkung 2 (zur Geschichte der Genetik): Der britische Arzt *John Arbuthnott* verglich bereits im Jahr 1710 das Zustandekommen des Geschlechts eines neugeborenen Kindes mit dem Wurf einer Münze mit dem Seiten M und F . In seiner Arbeit "*An Argument for Devine Providence, taken from the constant Regularity observ'd in the Birth of both Sexes*" verwendete die Binomischen Formeln

$$M^2 + 2MF + F^2$$

und

$$M^4 + 4M^3F + 6M^2F^2 + 4MF^3 + F^4.$$

Er hatte jedoch festgestellt, dass in den Taufregistern von London der Jahre 1629 bis 1710 in allen 82 Jahren mehr männliche als weibliche Neugeborene aufscheinen, sodass von einer zugunsten M gefälschten Münze auszugehen ist.

Der *Erwartungswert* und die *Varianz* der Binomialverteilung mit den Parametern n und p sind

$$\begin{aligned} E(S_n) &= n \cdot p & \text{und} \\ V(S_n) &= n \cdot p \cdot q. \end{aligned}$$

Da die Varianz die Dimension einer Fläche besitzt, ist es zweckmäßig, die Quadratwurzel aus der Varianz zu ziehen und die sogenannte *Standardabweichung*

$$\sigma_n = \sqrt{n \cdot p \cdot q}$$

als Maß für die Variabilität der Zufallsgröße S_n zu verwenden.

2.2 Die Poissonverteilung

Wir versuchen mithilfe des folgenden Beispiels das wichtige Rüstzeug der Exponentialfunktion bereitzustellen.

Beispiel 1¹: Über den Ausschuss bei der Produktion von Glasflaschen.

”Bei der Herstellung von Glasflaschen tritt folgendes Problem auf. Im geschmolzenen Glas, aus dem die Flaschen hergestellt werden, bleiben kleine feste Körper zurück, die wir einfach ‘Steine’ nennen wollen. Kommt ein solcher Stein in das Material einer Flasche, so ist sie unbrauchbar. Die Steine verteilen sich im flüssigen Glas unregelmäßig, bei konstanten Bedingungen der Fabrikation befinden sich aber in derselben Menge flüssigen Glases durchschnittlich etwa gleich viele Steine. Wir können also annehmen, dass in der Menge flüssigen Glases für 100 Flaschen durchschnittlich x Steine enthalten sind. Wir fragen, welcher Prozentsatz der hergestellten Flaschen dadurch unbrauchbar ist, dass die Flaschen Steine enthalten. Auf den ersten Blick könnte man etwa folgendes vermuten: Das Material für 100 Flaschen enthält x Steine, von den hergestellten Flaschen sind also x unbrauchbar. Diese Überlegung ist jedoch falsch. Da nicht beachtet wurde, dass in das Material einer Flasche auch mehrere Steine kommen können und der Anteil an Ausschussware dadurch vermindert wird.”

Die Wahrscheinlichkeitsrechnung ist das geeignete Instrument, diesen Sachverhalt aufzuklären:

Die Tatsache, dass im Durchschnitt x (zum Beispiel $x = 100$) Steine im Material für 100 Flaschen enthalten sind, könnte man grob so interpretieren, dass im Material jeder Flasche unabhängig voneinander ein oder kein Stein enthalten ist; und zwar mit den Wahrscheinlichkeiten

$$p = \frac{x}{100} \quad (\text{im Beispiel } = 1) \quad \text{bzw.} \quad 1 - \frac{x}{100} \quad (= 0)$$



Die Interpretation wäre schon deutlich besser, wenn man annimmt, dass in jedem zehnten Teil des Materials einer Flaschen unabhängig voneinander ein oder kein Stein enthalten ist; und zwar mit den Wahrscheinlichkeiten

¹aus Rényi, A.: Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie, VEB Deutscher Verlag der Wissenschaften, Berlin 1962, Seite 101 ff

$p/10$ (im Beispiel $= 1/10$) bzw. $1 - p/10$ ($= 9/10$). Demgemäß wäre die Wahrscheinlichkeit, dass sich im Material einer Flaschen kein Stein befindet

$$(1 - \frac{p}{10})^{10} \quad (\cong 0.3483).$$

Die Interpretation wäre noch besser, wenn man anstelle des zehnten Teils des Materials einer Flasche den 100-sten, 1000-sten, ... betrachtet. Die Wahrscheinlichkeit, dass sich im Material einer Flasche kein Stein befindet wäre sodann

$$\begin{array}{ll} (1 - \frac{p}{100})^{100} & (\cong 0.3660) \\ (1 - \frac{p}{1000})^{1000} & (\cong 0.3667) \\ \dots & \dots \end{array}$$

Mit einer ∞ feinen Unterteilung würde man allerdings gewiss über das Ziel hinausschießen, da eine solche ja der unrealistischen Annahme gleichkommt, dass Steine unendlich klein sind. Dennoch gehören unrealistische Grenzübergänge - wie dieser - zur beliebten, weil bequemen Praxis des Mathematikers.

Der Mathematiker zögert in diesem Fall keinen Augenblick, als Wahrscheinlichkeit dafür, dass sich im Material einer Flasche kein Stein befindet,

$$\lim_{n \rightarrow \infty} (1 - \frac{p}{n})^n \quad (\cong 0.3679)$$

anzusehen, wofür er - wegen der enormen Bedeutung dieser Funktion - kurz e^{-p} schreibt ²⁾. Die Zahl $e \cong 2.71828$ heißt zu Ehren des Schweizer Mathematikers *Leonhard Euler* (1707 – 1783) *Euler'sche Zahl*.

Da somit die Wahrscheinlichkeit, dass sich ein oder mehrere Steine im Material für eine Flaschen befinden,

$$1 - e^{-p} \quad (\cong 0.6321)$$

ist, werden von 100 Flaschen etwa

$$100 \cdot (1 - e^{-p}) \quad (\cong 63)$$

unbrauchbar sein.

Für unser Beispiel von $x = 100$ sind das beträchtlich weniger als die eingangs erwähnte erste Vermutung liefert.

Für $x = 30$ - und das kann praktisch noch vorkommen - erhält man anstelle der vermuteten 30 nur etwa 26 unbrauchbare Flaschen. Für sehr kleines x ist die Anzahl der unbrauchbaren Flaschen tatsächlich ungefähr gleich x .

Rényi schreibt abschließend:

”Es ist bemerkenswert, dass es bei einem hohen Gehalt an Steinen wirtschaftlicher ist, kleinere Flächen herzustellen (natürlich, wenn kein Verfahren zur Verfügung steht, das flüssige Glas zu reinigen). Macht man z.B. bei $x = 30$ mit dem Material für 100 Flaschen ursprünglicher Größe 400 kleinere Flaschen, so sinkt der Prozentsatz an Ausschussware von 25.92% auf 7.22%. Dies ist ein Beispiel dafür, dass man mit Hilfe der Wahrscheinlichkeitsrechnung Auskunft über konkrete Fragen der Produktion erhalten kann.”

Zwecks Motivation der Poissonverteilung betrachten wir einen Versuch, der darin besteht, $74 = 2 \times 37$ einzelne Versuche beim Roulette durchzuführen und zu beobachten, wie oft die Kugel dabei auf ”0” zu liegen kommt.

Aufgabe 1: Entnehmen Sie die Ergebnisse von 100 solchen Versuchen aus den monatlich veröffentlichten Permanenzen 3) des Casino Salzburg und tragen Sie die Versuchsergebnisse in die folgende Tabelle ein.

Tabelle 1: Absolute und relative Häufigkeiten für die Anzahl X der ”0”-en bei 100 Versuchen von jeweils 74 Einzelversuchen beim Roulette.

Anzahl k der ”0”-en	Strichliste	$H_{100}(k)$	$h_{100}(k)$
0			
1			
2			
3			
4			
5			
> 5			

Nun zur Theorie:

Gemäß Beispiel 1 wird die Wahrscheinlichkeit $P(X = 0)$, dass die Kugel kein einziges mal auf ”0” zu liegen kommt, durch

$$\begin{aligned} P(X = 0) &= \left(1 - \frac{1}{37}\right)^{74} = \left[\left(1 - \frac{1}{37}\right)^{37}\right]^2 \\ &\cong [e^{-1}]^2 = e^{-2} \end{aligned}$$

approximiert.

Die Wahrscheinlichkeiten $P(X = 0)$, $P(X = 1)$, ... sind durch die Binomialverteilung $B_{74, 1/37}$ bestimmt. So ist z.B. die Wahrscheinlichkeit $P(X = 3)$,

dass die Kugel 3-mal auf "0" zu liegen kommt,

$$\begin{aligned} P(X=3) &= \binom{74}{3} \left(\frac{1}{37}\right)^3 \left(1 - \frac{1}{37}\right)^{74-3} \\ &= \frac{74 \cdot 73 \cdot 72}{3!} \cdot \frac{\left(\frac{1}{37}\right)^3}{1} \cdot \frac{\left(1 - \frac{1}{37}\right)^{74}}{\left(1 - \frac{1}{37}\right)^3} . \end{aligned}$$

Wegen

$$74 \cdot 73 \cdot 72 \cdot \left(\frac{1}{37}\right)^3 = \frac{74}{37} \cdot \frac{73}{37} \cdot \frac{72}{37} \cong 2^3, \quad \left(1 - \frac{1}{37}\right)^{74} \cong e^{-2}$$

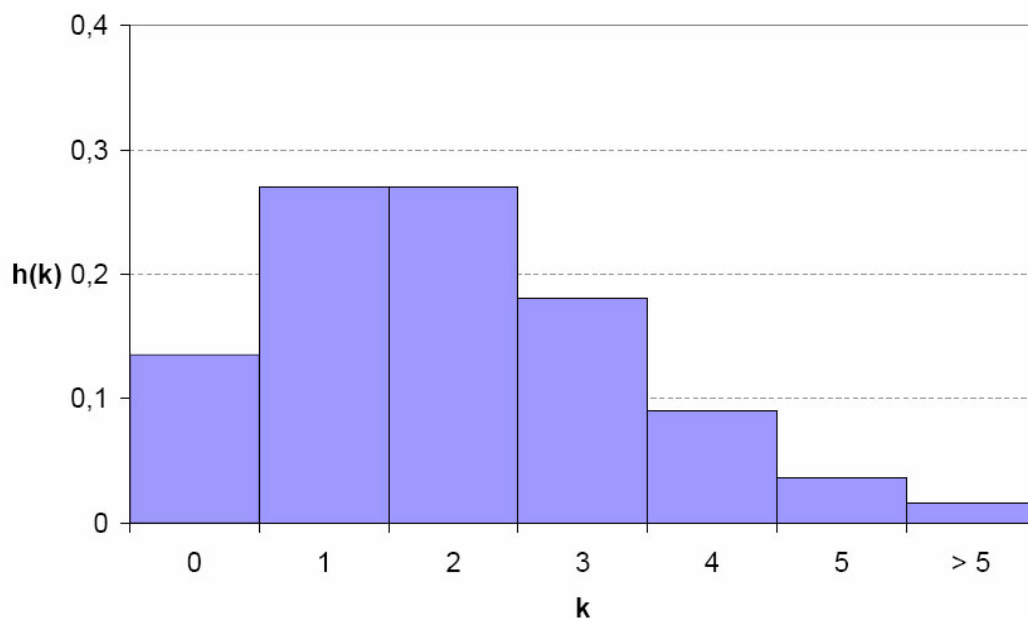
und - in sehr grober Näherung -

$$\left(1 - \frac{1}{37}\right)^3 \cong 1$$

ist

$$P(X=3) \cong \frac{2^3 \cdot e^{-2}}{3!} .$$

Abbildung 1: Theoretische (—) und empirische (—) Verteilung zu Aufgabe 1.



Poissonverteilung^{2,3}: Die durch

$$P_\lambda(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}, \quad k \in \{0, 1, \dots\} \quad \text{für } \lambda > 0$$

definierte Verteilung heißt Poissonverteilung. Der Parameter λ hat mehrfache Bedeutung: λ ist zugleich Erwartungswert und Varianz.

Radioaktiver Zerfall

*Ernest Rutherford*⁴ (1871 – 1937) geht bei seiner Beschreibung der Emission radioaktiver Strahlung von der Differentialgleichung

$$\frac{dn}{dt} = -\lambda n$$

aus. Der Anfangswert zum Zeitpunkt $t = 0$ ist N . Die äquivalente Darstellung

$$\frac{d \ln n}{dt} = \frac{\frac{dn}{dt}}{n} = -\lambda$$

liefert durch Integration

$$\ln n = \ln c - \lambda t$$

bzw.

$$n(t) = N \cdot e^{-\lambda t},$$

wobei wegen des Anfangswerts zum Zeitpunkt $t = 0$ $c = N$ ist.

Anmerkung 1: Die entscheidende Kenngröße für eine radioaktive Substanz ist bekanntlich deren Halbwertszeit τ . Das ist die Zeit, die vergeht, bis genau

² *Siméon D. Poisson* (1781 – 1840) hat auf Seite 206 seines Buches "Recherches sur la probabilité de jugements en matière criminelle et en matière civile", Paris 1837, erstmals den Grenzübergang

$$\lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

durchgeführt. Deswegen trägt die Grenzverteilung der Binomialverteilung B_{n,p_n} für $p_n = \frac{\lambda}{n}$ seinen Namen.

³ *Simon Newcomb* (1835 – 1909) dürfte die Poissonverteilung erstmals in statistischem Zusammenhang verwendet haben, indem er sie bei einem Treffen der American Academy of Arts and Sciences zu einer Untersuchung der jahrhundertealten Frage heranzog, ob die Sterne des Himmels zufällig verteilt sind.

⁴ *Rutherford, E.*: (1900): A Radio-active Substance emitted from Thorium Compounds. Philosophical Magazine **5**, p. 1 – 14

die Hälfte der radioaktiven Atome zerfallen sind, also die Lösung der Gleichung

$$n(\tau) = \frac{N}{2},$$

nämlich

$$\tau = \frac{1}{\lambda} \ln\left(\frac{N}{N/2}\right) = \frac{\ln 2}{\lambda}.$$

In einer Notiz zur Arbeit von *Rutherford* und *Geiger* "The Probability Variations in the Distribution of α Particles" ⁵ im Jahre 1910 beschreibt der englische Mathematiker *Harry Bateman* (1882 – 1946) die Anzahl X der Szintillationen in einem Zeitintervall der Länge t erstmals mithilfe der Poissonverteilung

$$P_{\lambda t}(X = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n \in \mathbb{N}_0$$

mit dem Parameter λt .

2.3 Die Exponentialverteilung

Ein Versuch besteht darin, eine Münze so lange zu werfen, bis "Kopf" fällt. Die Anzahl A der Würfe ist von Interesse.

Die Verteilung, die einen solchen Versuch beschreibt, ist die **Geometrische Verteilung**:

$$P(A = k) = p \cdot q^{k-1} \quad (P(A > k) = q^k), \quad k \in \mathbb{N},$$

wobei $P(\text{"Kopf"}) = p$ und $P(\text{"Adler"}) = q = 1 - p$ ist.

Experiment 1 ⁶: Das Experiment bestehe aus 64 Versuchen im obigen Sinn, wobei mit einer gewöhnlichen ("fairen") Münze geworfen wird.

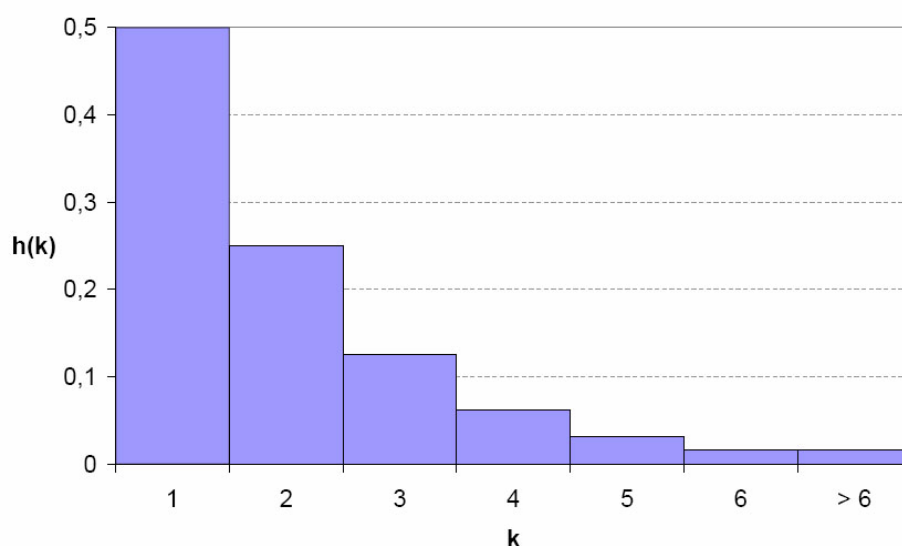
Gemäß der geometrischen Verteilung (für $p = q = \frac{1}{2}$) ist zu erwarten, dass von 64 Versuchen 32 Versuche einen Wurf, 16 Versuche 2 Würfe, 8 Versuche 3 Würfe, 4 Versuche 4 Würfe, 2 Versuche 5 Würfe, 1 Versuch 6 Würfe und 1 Versuch mehr als 6 Würfe benötigen.

⁵ *Rutherford, E., H. Geiger, H. and Bateman, H.* (1910): Philosophical Magazine, 6th Series, **20**, 689 – 707

⁶ Vgl. *Pearson, K.*: Science and Monte Carlo. Fortnightly Review, February 1892, wo die empirische Verteilung der Längen von "runs" einer Farbe beim Roulette mit der theoretischen Verteilung verglichen wird.

Tabelle1: Absolute und relative Häufigkeiten für die verschiedenen der Münzwürfe bei 64 Versuchen.

Anzahl k der Münzwürfe	Strichliste	$H_{64}(k)$	$h_{64}(k)$
1			
2			
3			
4			
5			
6			
> 6			

Abbildung 1: Theoretische (————) und empirische Verteilung (— — —) zu Experiment 1.

Wenn man sich vorstellt, dass zu jeder vollen Sekunde ein "Münzwurf" stattfindet, so stimmt A mit der Versuchsdauer T überein. Die zu erwartende Versuchsdauer ist somit $E(T) = E(A) = \frac{1}{p}$ (für den Spezialfall $p = \frac{1}{2}$ also, wie zu vermuten, 2 Sekunden).

Will man aber je einen "Münzwurf" zu Ende jeder $\frac{1}{10}$ Sekunde durchführen, die zu erwartende Versuchsdauer $E(T) = \frac{1}{p}$ jedoch beibehalten, so muss für

die zu erwartende Anzahl $E(A)$ der Würfe pro Versuch wegen $A = 10 \cdot T$

$$E(A) = 10 \cdot E(T) = \frac{10}{p} = \frac{1}{p/10}$$

gelten. Das bedeutet, dass die Wahrscheinlichkeit für "Kopf" nunmehr $\frac{p}{10}$ betragen muss. Die Wahrscheinlichkeit, dass ein Versuch länger als t Sekunden dauert, ist somit

$$p_{10}(T > t) = \left(1 - \frac{p}{10}\right)^{10 \cdot t},$$

bzw.

$$p_n(T > t) = \left(1 - \frac{p}{n}\right)^{n \cdot t},$$

wenn man je einen "Münzwurf"⁷ pro $\frac{1}{n}$ Sekunde durchführen will.

Der Grenzübergang für $n \rightarrow \infty$ ist zwar - wie immer - etwas wirklichkeitsfremd aber bequem. Er verhilft hier zur Darstellung der fraglichen Wahrscheinlichkeit mit Hilfe der, bei Mathematikern sehr beliebten Exponentialfunktion, welche - unter anderem - durch Grenzübergang

$$P(T > t) = \lim_{n \rightarrow \infty} \left(1 - \frac{p}{n}\right)^{n \cdot t} = e^{-p \cdot t}$$

definiert ist. Die Zahl $e \cong 2.71828$ heißt zu Ehren des Schweizer Mathematikers *Leonhard Euler* (1707 – 1783) *Euler'sche Zahl*.

Exponentialverteilung: Die durch

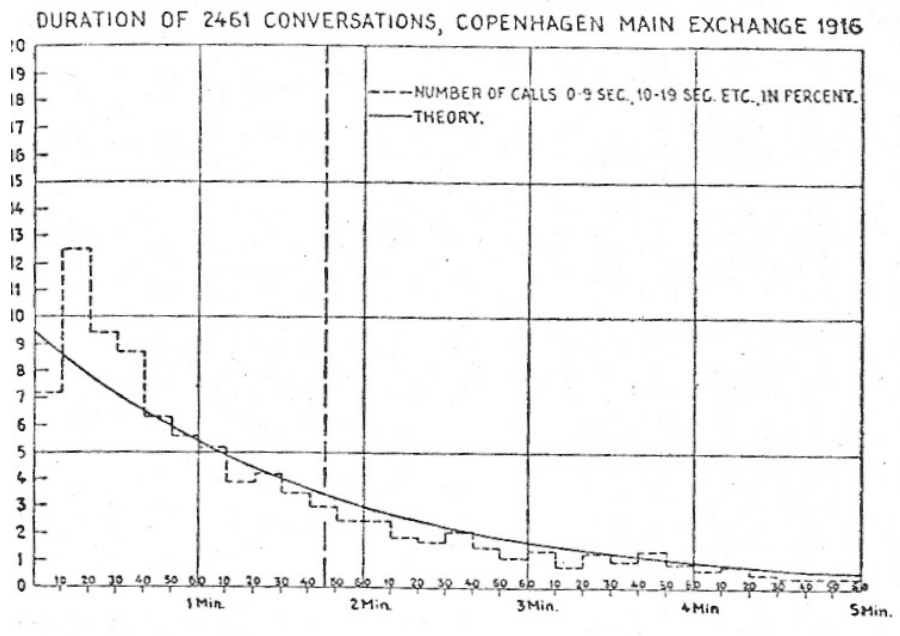
$$P(T > t) = e^{-\lambda \cdot t}, \quad \lambda > 0^8$$

definierte Verteilung heißt Exponentialverteilung. Sie hat den Erwartungswert $E(T) = \frac{1}{\lambda}$ und approximiert die Lebensdauer- (Versuchsdauer-) Verteilung bzw. die Wartezeitverteilung "seltener Ereignisse" namentlich dann, wenn (analog zum Münzwurf) die bereits absolvierte Wartezeit keinen Einfluss auf die noch zu absolvierende hat. Denn die Eigenschaft

$$\frac{P(T > t + h)}{P(T > t)} = P(T > h)$$

charakterisiert sowohl die geometrische Verteilung wie die Exponentialverteilung.

⁷Die Bezeichnung "Münzwurf" wurde lediglich aus methodischen Gründen beibehalten. Was wir mit unserer Überlegung zu beschreiben beabsichtigen, ist hingegen recht allgemein.

Beispiel 1⁹: *Über die Dauer von Telefongesprächen*

Die Dauer eines Telefongesprächs ist früher in der Regel als fest angesehen worden. Dies ist ziemlich richtig für den Fall von Ferngesprächen, welche nach dem Ablauf einer gewissen Zeit unterbrochen werden. Für die wichtigste Klasse von Gesprächen, nämlich die der gewöhnlichen Ortsgespräche, gilt mit guter Näherung das Verteilungsgesetz

$$S(t) = P(T > t) = e^{-t} \quad \text{und daher} \quad S'(t) = -e^{-t},$$

wobei Zeiteinheit und mittleren Gesprächsdauer als identisch angenommen wurden. Die beiden Gleichungen drücken aus, dass die Wahrscheinlichkeit, dass ein laufendes Gespräch im Begriff ist, zu Ende zu gehen, unabhängig von der Zeit ist, die es bereits gedauert hat - was vernünftigermaßen von vornherein hätte erwartet werden können. Der entscheidende Punkt ist jedoch, dass Praxis und Theorie recht gut übereinstimmen, wie die obige graphische Darstellung der Erfahrungen der Hauptfernsprechzentrale in Kopenhagen zeigen.

⁹ von A.K. Erlang: The application of the theory of probabilities in telephon administration. Kopenhagen 1920 (frei übersetzt)

Beispiel 2: Radioaktiver Zerfall

Marsden und Barratt (1911)¹⁰ stellen in ihrer Arbeit "The α Particles emitted by the Active Deposits of Thorium and Actinium" für den radioaktiven Zerfall den Zusammenhang zwischen der Poissonverteilung und Exponentialverteilung her. Gemäß Anmerkung 1 in Abschnitt 2.2 ist

$$P_{\lambda t}(X = 0) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

die Wahrscheinlichkeit, dass im Zeitintervall $[0, t]$ kein Zerfall stattfindet. Anders ausgedrückt heißt das, dass für die Wartezeit T bis zum Zerfall des ersten Atoms gilt

$$P(T > t) = e^{-\lambda t} \quad \text{oder, gleichbedeutend,} \quad P(T \leq t) = 1 - e^{-\lambda t}.$$

Die Funktion

$$F_{\lambda}(t) = 1 - e^{-\lambda t}, \quad t \in [0, \infty)$$

heißt *Verteilungsfunktion* der Exponentialverteilung mit dem Parameter λ . Deren Ableitung

$$\frac{d}{dt} F_{\lambda}(t) = \lambda e^{-\lambda t}, \quad t \in [0, \infty)$$

heißt *Dichtefunktion* der Exponentialverteilung. Die durchschnittliche Zeitdauer δ zwischen zwei aufeinanderfolgenden Zerfällen ist - wie wir bereits wissen -

$$\delta = \frac{1}{\lambda}.$$

2.4 Die Normalverteilung

2.4.1 Allgemeines

Wir gehen von einer Zufallsgröße S_n aus, welche gemäß einer Binomialverteilung mit den Parametern n und $p = \frac{1}{2}$ verteilt ist. Gemäß Abschnitt 2.1 sind Erwartungswert und Standardabweichung von S_n

$$E(S_n) = \frac{n}{2} \quad \text{und} \quad \sigma_n = \frac{\sqrt{n}}{2}.$$

¹⁰Marsden, E. and T. Barratt: In: Proceedings of the Physical Society of London 24 (1911), S. 5061

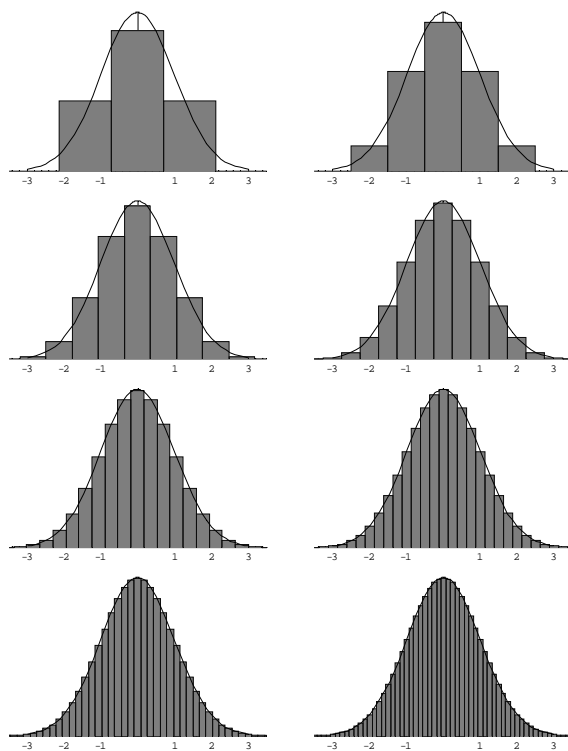
Im Folgenden betrachten wir, wie sich die Verteilung der zentrierten und normierten Zufallsgröße

$$S_n^* = \frac{S_n - \frac{n}{2}}{\sqrt{n}/2}$$

für wachsendes n verhält. Die entsprechenden Histogramme legen nahe, dass die Verteilungen von S_n^* einer Grenzverteilung zustreben.

Die Kurve, der die Histogramme mit wachsendem n zustreben, ist die berühmte *Gaußsche Glockenkurve*. Sie hat die Form

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$



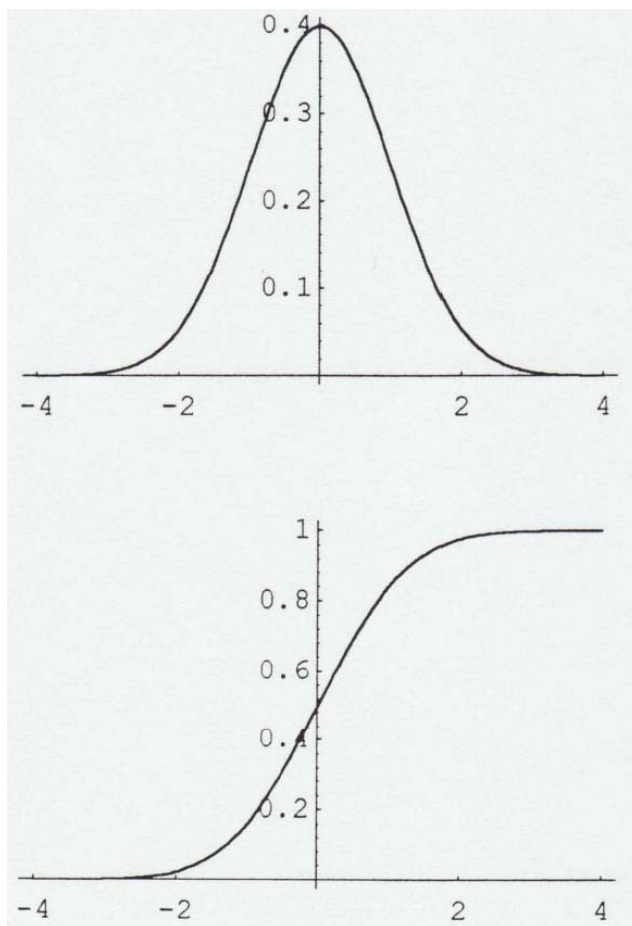
Das Integral

$$\Phi(z) = \int_{-\infty}^z \varphi(x) dx$$

ist die Fläche zwischen der Funktion φ und der Abszisse bis zum Wert z . Die Integralfunktion

$$\Phi(z), \quad z \in \mathbb{R},$$

heißt *Gaußsche Summenfunktion*.



Die Dichtefunktion der Normalverteilung $N(\mu, \sigma^2)$ mit Erwartungswert μ und Varianz σ^2 ist

$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}.$$

Die Beziehung

$$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X - \mu}{\sigma} \sim N(0, 1)$$

gewährleistet, dass es reicht, bloß Dichte- und Verteilungsfunktion

$$\Phi(z) = \int_{-\infty}^z \varphi(x) dx$$

der Standard-Normalverteilung zu tabellieren.

Die Normalverteilung verdankt ihre herausragende Bedeutung einer weitreichenden Verallgemeinerung des ursprünglichen *Satzes von De Moivre-Laplace*¹¹, welcher die eingangs dargestellte Konvergenz der Folge der Histogramme gegen die Normalverteilung mathematisch beschreibt. Dieser Grenzübergang erfolgt auch für die zentrierte und normierte Zufallsgrößen

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}$$

für allgemeine $p \in (0, 1)$, allerdings weit weniger rasch als für den Fall $p = \frac{1}{2}$. Eine für praktische Zwecke brauchbare Annäherung von

$$P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq z\right) \cong \Phi(z) \quad \text{für } z \in [-3, 3]$$

ist dann gegeben, wenn die Faustregel

$$np(1-p) \geq 9$$

erfüllt ist.

Zur Geschichte: Quetelet und Galton

Der belgische Statistiker *Lambert-Adolphe-Jacques Quetelet* (1796 – 1874), auf dessen Einfluss die Gründung vieler statistischer Behörden in Europa zurückgeht, hat der Normalverteilung in der Anthropometrie ein gänzlich neues und unvermutetes Anwendungsgebiet erschlossen. Bei seinen Überlegungen geht er von jenem Gebiet aus, für welches das Fehlerverteilungsgesetz ursprünglich modelliert wurde: den Präzisionsmessungen in Astronomie und Geodäsie, und bedient sich der Fiktion des *homme moyen* (des mittleren Menschen). Zur Illustration seiner Ideen verwendet er Daten aus statistischen Erhebungen.

¹¹ *Abraham de Moivre* (1667 – 1754), *Pierre-Simon de Laplace* (1749 – 1827)

In den vorhergehenden Beispielen wussten wir trotz der Schwankungen der Messwerte, dass sehr wohl eine Messgröße in Wirklichkeit existiert, deren Wert wir zu bestimmen versuchten: das war entweder die Körpergröße eines Menschen oder die Rektaszension des Polarsterns.

Es stellt sich hier eine Frage von größter Wichtigkeit. Man kann sich fragen, ob in einem Volk ein Modellmensch existiert: ein Mensch, der dieses Volk bezüglich der Körpergröße repräsentiert. Im Vergleich zu diesem müssten alle Menschen seines Volkes so betrachtet werden, wie wenn sie mehr oder weniger große Abweichungen aufweisen. Wenn man diese misst, dann hätte man Werte, die sich um den Mittelwert gruppieren, wie jene Zahlen, die man erhielte, wie wenn mit mehr oder weniger groben Mitteln gemessen worden wäre.

Quetelet, 1834

Man vergleiche dazu Beispiel 5: Zur "Kurve von Quetelet" in Abschnitt 1.3.3 von Teil 1 der Seminarunterlagen und für weitere Hintergrundinformationen [17].

Der Engländer *Francis Galton* (1822–1911) hat die Arbeiten *Quetelets* weitergeführt. U.a. hat er sein sogenanntes *Galton-Brett* dazu ersonnen, um verständlich zu machen, warum die Daten vieler Beobachtungen und Experimente gut durch eine Normalverteilung beschrieben werden können. Diese nüchterne Aussage klingt in *Galtons* Worten ¹² ganz anders:

Order in Apparent Chaos. - I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they would have known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, the more perfect is its sway. It is the supreme law of Unreason.

2.4.2 Das Wahrscheinlichkeitsnetz

Aus der empirischen Verteilungsfunktion einer im Rahmen einer Beobachtungsstudie oder im Rahmen eines Experiments erhobenen Datenmenge lässt

¹²aus: *Galton, F.:* "Natural Inheritance", Macmillan & Co., London - New York 1889

sich üblicherweise visuell schwer erkennen, ob diese aus einer normalverteilten Grundgesamtheit stammt oder nicht. Wir können jedoch leichter erkennen, ob eine Punktwolke gut durch eine geeignete Gerade beschrieben werden kann. Diese Eigenschaft unseres Wahrnehmungsapparats nützt das sogenannte *Wahrscheinlichkeitsnetz* (*normal quantile plot*).

Sei Φ die Verteilungsfunktion der $N(0, 1)$ -Verteilung und Φ^{-1} die zugehörige Inverse. Dann gilt offensichtlich $\Phi^{-1}(\Phi(x)) = id(x)$. Darauf beruht das *Wahrscheinlichkeitsnetz*, in dem man die Werte $\Phi^{-1}(i/n)$ gegen die Werte $x_{i:n}$ aufträgt. Die zugehörigen Punktepaare

$$(x_{i:n}, \Phi^{-1}(i/n)), \quad i \in \{1, \dots, n\},$$

liegen erwartungsgemäß annähernd auf einer Geraden, wenn die Daten aus einer normalverteilten Grundgesamtheit stammen. Da Ordinatenabschnitt und Anstieg der Geraden vom Stichprobenmittel \bar{x}_n und der Stichprobenvarianz s_n abhängen, kann es zweckmäßig sein, die Beobachtungswerte zuvor gemäß

$$\tilde{x}_{i:n} = \frac{x_{i:n} - \bar{x}_n}{s_n}$$

zu zentrieren und zu standardisieren und dann die "Punktwolke"

$$(\tilde{x}_{i:n}, \Phi^{-1}(i/n)), \quad i \in \{1, \dots, n\},$$

mit der Geraden $y = x$ zu vergleichen.

Wir wenden das Wahrscheinlichkeitsnetz im Rahmen folgenden Anwendungsaufgabe an.

Sind Schwangerschaftsdauern normalverteilt? (siehe [?], 1986, Case Study 4.3.3, S. 407. Man vergleiche dazu auch die Aufgaben 3 und 5 aus Teil 1 der Seminarunterlagen.)

Die Fallstudie beginnt mit folgendem Zitat: In einer allseits bekannten Kolumne für Verliebte des *Tennessean* (Nashville, January 27, 1973), ist folgender Brief erschienen.

Dear Abby: You wrote in your column that a woman is pregnant for 266 days. Who said so? I carried my baby for ten month and five days (310 days), and there is no doubt about it because I know the exact date my baby was conceived. My husband is in the

Navy and it couldn't have been possible any other time because I saw him only once for an hour, and I didn't see him again until the day before the baby was born.

I don't drink or run around, and there is no way this baby isn't his, so please print a retraction about the 266-carrying time because otherwise I am in a lot of trouble.

San Diego Reader

Entsprechend gut dokumentierter Aufzeichnungen sind Erwartungswert und Standardabweichung der Dauer T einer Schwangerschaft $\mu = 266$ und $\sigma = 16$ Tage. Wenn angenommen werden könnte, dass T überdies normalverteilt ist, wäre die Wahrscheinlichkeit, dass die Schwangerschaftsdauer 310 Tage oder länger dauert,

$$P(T \geq 310) = P\left(\frac{T - 266}{16} \geq \frac{310 - 266}{16}\right) = 1 - \Phi(2.75) \cong 0.003.$$

Die Interpretation dieses Ergebnisses sei der Leserin bzw. dem Leser überlassen. (Mehr dazu könnte in einem Abschnitt über das "Testen von Hypothesen" gesagt werden. Zur Abhandlung dieses Themas für Wahrscheinlichkeiten und Anteile sei auf Abschnitt 2.2 des Skriptums [11] verwiesen.)

Ähnlich wie in dieser Fallstudie stellt sich bei vielen statistischen Untersuchungen die entscheidende Frage, ob die zugrundeliegende Grundgesamtheit (mit gutem Grund) als normalverteilt angenommen werden kann, denn erst dann lassen sich spezifische, auf der Normalverteilungsannahme beruhende statistische Verfahren anwenden und schließlich wohlbegründete Aussagen treffen. In der beurteilenden Statistik wird zur Überprüfung der Gültigkeit der Normalverteilungsannahme üblicherweise der sogenannte *Lilliefors Test* verwendet.

Die Untersuchung der Daten mit Hilfe des Wahrscheinlichkeitsnetzes dient Anwendern gelegentlich als wichtige Orientierungshilfe.

Die nachstehende Tabelle zeigt die in einem der *County General Hospitals* in

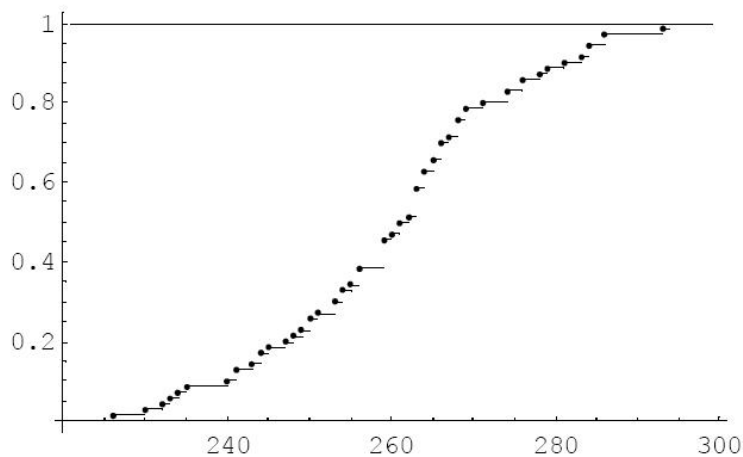
den U.S.A. im Jahre 1978 erhobenen Dauern von 70 Schwangerschaften.

251	264	234	283	226	244	269	241	276	274
263	243	254	276	241	232	260	248	284	253
265	235	259	279	256	256	254	256	250	269
240	261	263	262	259	230	268	284	259	261
268	268	264	271	263	259	294	259	263	278
267	293	247	244	250	266	286	263	274	253
281	286	266	249	255	233	245	266	265	264

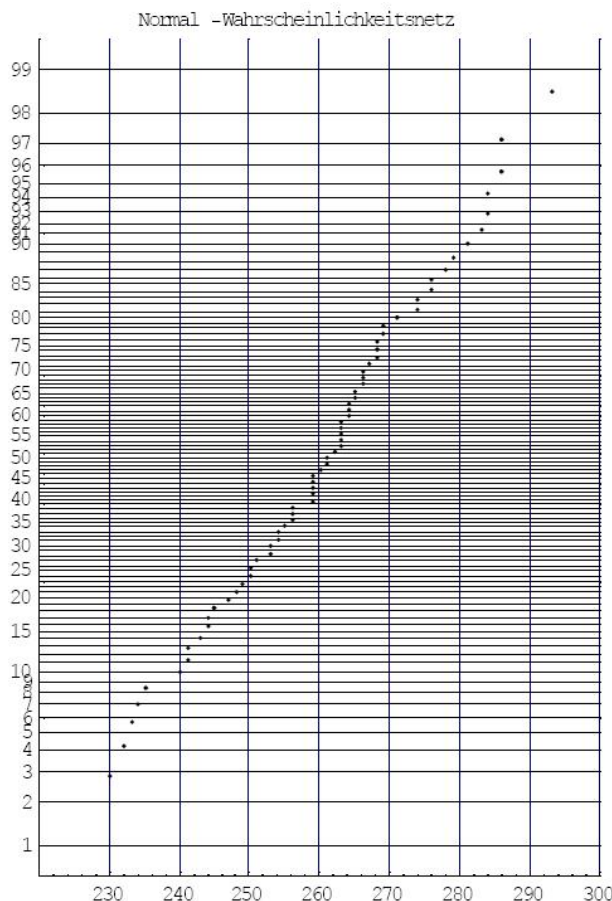
Die beiden nachstehenden Graphiken zeigen das zugehörige Stängel-Blatt-Diagramm der geordneten Daten

22		6
23		02345
24		0113445789
25		0013344566699999
26		011233333344455666788899
27		1446689
28		14466
29		34

und die empirische Verteilungsfunktion.



Die folgende Graphik zeigt schließlich die in ein *Wahrscheinlichkeitsnetz* (*normal quantile plot*)¹³ eingetragene "Punktwolke" für die zugehörigen geordneten, jedoch nicht zentrierten und standardisierten Daten. In einem üblichen Koordinatensystem ist dies die Punktwolke $(x_{i:70}, \Phi^{-1}(i/70))$, $i \in \{1, \dots, n\}$.



¹³Dieses ist im Handel erhältlich; beispielsweise bei Fa. Schleicher & Schüll GmbH, Grimsehlstraße 23, D-37574 Einbeck.

Weiters ist es unter <http://www.statistik.tuwien.ac.at/lv-guide/> auf der Homepage des Instituts für Statistik & Wahrscheinlichkeitstheorie der TU Wien verfügbar (16.1.2007).

Chapter 3

BEURTEILENDE STATISTIK

3.1 DER χ^2 -TEST

3.1.1 Einleitung

Wir beginnen mit einer einfachen Anwendung des χ^2 -Tests¹, der die nötigen allgemeinen Betrachtungen folgen.

Beispiel: Es sollte untersucht werden, ob ein bestimmter Würfel fair ist. Zu diesem Zweck wurde der Würfel $n = 600$ mal geworfen. Die Anzahlen b_j der Würfe mit der Augenzahl $j \in \{1, \dots, 6\}$ waren wie folgt:

b_1	b_2	b_3	b_4	b_5	b_6
142	79	92	102	110	75

Kann auf Grund dieses Ergebnisses die Annahme (Hypothese H_0) aufrecht erhalten werden, dass der Würfel fair ist?

Seien X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariable mit Wertebereich $W_{X_1} = \{\omega_0, \omega_1, \dots, \omega_m\}$ und Verteilung $P = (p_0, p_1, \dots, p_m)$ mit positiven Wahrscheinlichkeiten. Weiters sei $\mathbb{B}_m^{(n)} = (B_0, B_1, \dots, B_m)$ der Vektor der Ausfallshäufigkeiten

$$B_j = |\{i \in \{1, \dots, n\} : X_i = \omega_j\}|, \quad j \in \{0, 1, \dots, m\}.$$

¹Der χ^2 -Test ist eine Erfindung des englischen Statistikers *Karl Pearson*. Sein bahnbrechender Artikel erschien im Jahr 1900. Der χ^2 -Test ist der bekannteste Vertreter der sogenannten *Anpassungstests* (*Goodness-of-fit tests*).

(Somit ist $\sum_{j=0}^m B_j = n$). Dann sind die Zufallsvariablen B_j Binomialverteilt mit den Parametern n und p_j und es gelten daher

$$E(B_j) = n p_j \quad \text{und} \quad V(B_j) = n p_j (1 - p_j), \quad j \in \{0, \dots, m\}.$$

Wir definieren nun ein "Abstandsmaß" des Zufallsvektors $\mathbb{B}_m^{(n)} = (B_0, B_1, \dots, B_m)$ der Beobachtungen vom Vektor $\mathbb{E}_m^{(n)} = nP = (np_0, np_1, \dots, np_m)$ der zugehörigen Erwartungswerte, und zwar durch das n -fache des Erwartungswerts der Quadrate der relativen Abweichungen

$$\chi_m^2 = \chi^2(\mathbb{B}_m^{(n)}, nP) = \sum_{j=0}^m \frac{(B_j - n p_j)^2}{n p_j}.$$

Für den Spezialfall $m = 1$ können wir uns mit den uns zur Verfügung stehenden Mitteln davon überzeugen, dass der dadurch definierte " χ^2 -Abstand", der übrigens keineswegs die Eigenschaften einer Metrik besitzt, statistisch brauchbar ist. Wegen $B_0 + B_1 = n$ und $n p_0 + n p_1 = n$ gilt nämlich

$$\chi_1^2 = \sum_{j=0}^1 \frac{(B_j - n p_j)^2}{n p_j} = \left(\frac{B_1 - n p_1}{\sqrt{n p_1 (1 - p_1)}} \right)^2.$$

Nun besagt der Satz von DeMoivre-Laplace

$$\lim_{n \rightarrow \infty} P \left(\frac{B_1 - n p_1}{\sqrt{n p_1 (1 - p_1)}} \leq z \right) = \Phi(z)$$

mit $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, $z \in \mathbb{R}$. Somit gilt für $z > 0$

$$\lim_{n \rightarrow \infty} P \left(\left(\frac{B_1 - n p_1}{\sqrt{n p_1 (1 - p_1)}} \right)^2 \leq z \right) = 2 \Phi(\sqrt{z}) - 1$$

und daher wegen $\frac{d}{dz} (2 \Phi(\sqrt{z}) - 1) = \frac{1}{\sqrt{2\pi z}} e^{-z/2}$

$$\lim_{n \rightarrow \infty} P \left(\left(\frac{B_1 - n p_1}{\sqrt{n p_1 (1 - p_1)}} \right)^2 \leq z \right) = \int_0^z \frac{1}{\sqrt{2\pi x}} e^{-x/2} dx.$$

Die Funktion

$$f_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2}, \quad x \in (0, \infty),$$

ist der Spezialfall der Klasse der Dichtefunktionen der χ^2 -Verteilungen mit m Freiheitsgraden für den Spezialfall $m = 1$.

Definition: Es sei $m \in \mathbb{N}$. Dann heißt die stetige Verteilung mit dem Träger $(0, \infty)$ und der Dichtefunktion

$$f_m(x) = \frac{x^{m/2-1} e^{-x/2}}{2^{m/2} \Gamma(m/2)}$$

χ^2 -Verteilung mit m Freiheitsgraden. Dabei ist $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, $\alpha > 0$, die sogenannte *Gamma-Funktion*.

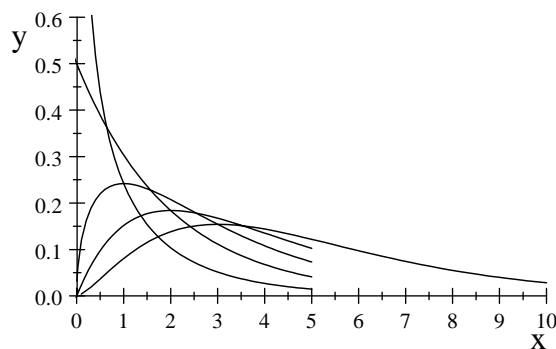


Abbildung der Dichtefunktionen $f_m(x)$ für $m = 1, 2, 3, 4$ und 5

Anmerkung 1: Erwartungswert und Varianz der χ^2 -Verteilung mit m Freiheitsgraden sind m bzw. $2m$. Nun ist der Erwartungswert der χ^2 -Statistik

$$E(\chi^2(\mathbb{B}_m^{(n)}, nP)) = m,$$

also tatsächlich gleich dem Erwartungswert der χ^2 -Verteilung mit m Freiheitsgraden. Die zugehörige Varianz ist

$$V(\chi^2(\mathbb{B}_m^{(n)}, nP)) = 2m + \frac{1}{n} \left[\sum_{j=0}^m \frac{1}{p_j} - (m+2)^2 + 3 \right],$$

also nur asymptotisch gleich der Varianz $2m$ dieser Verteilung.

Der χ^2 -Test beruht auf dem von *Karl Pearson* im Jahr 1900 veröffentlichten Satz, dessen Gültigkeit man aus dem behandelten Spezialfall für $m = 1$ und aus Anmerkung 1 vermuten kann.

Satz: Seien $m \in \mathbb{N}$ fest und $P = (p_0, p_1, \dots, p_m)$ eine beliebige Wahrscheinlichkeitsverteilung mit positiven Werten p_j . Dann gilt

$$\lim_{n \rightarrow \infty} P(\chi^2(\mathbb{B}_m^{(n)}, nP) \leq z) = \int_0^z f_m(x) dx.$$

Statistische Folgerung des Satzes von Pearson: Seien X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariable mit Wertebereich $W_{X_1} = \{\omega_0, \omega_1, \dots, \omega_m\}$ und bezeichne $V_m = \{P' = (p'_0, p'_1, \dots, p'_m) : p'_j \geq 0 \forall j \in \{0, 1, \dots, m\}, \sum_{j=0}^m p'_j = 1\}$ die Menge der Wahrscheinlichkeitsverteilungen auf $\{\omega_0, \omega_1, \dots, \omega_m\}$. Ferner sei $P = (p_0, p_1, \dots, p_m) \in V_m$ derart, dass alle p_j positiv sind. Um die

- Hypothese $H_0 : X_1 \sim P = (p_0, p_1, \dots, p_m)$
gegen die
- Hypothese $H_1 : \text{Es gibt eine Verteilung } Q \in V_m \setminus \{P\} \text{ derart, dass } X_1 \sim Q$

auf einem Niveau $\cong \alpha$ mit $0 < \alpha \ll 1$ zu testen, kann man sich des folgenden (verteilungsfreien) Verfahrens bedienen. Man lehne H_0 zugunsten H_1 ab, wenn

$$\chi_m^2 = \sum_{j=0}^m \frac{(b_j - n p_j)^2}{n p_j} \geq \chi_{m, 1-\alpha}^2$$

gilt. Dabei ist der kritische Wert $\chi_{m, 1-\alpha}^2$ das $(1 - \alpha)$ -Quantil der Verteilungsfunktion der χ^2 -Verteilung mit m Freiheitsgraden.

Im Folgenden ist die Tabelle der kritischen Werte $\chi_{m, 1-\alpha}^2$ für $m \in$

$\{1, \dots, 10\}$ und $\alpha \in \{0.5, 0.1, 0.05, 0.01\}$ angegeben.

$m \setminus \alpha$	0.5	0.2	0.1	0.05	0.01
1	0.455	1.642	2.706	3.841	6.635
2	1.386	3.219	4.605	5.991	9.210
3	2.366	4.642	6.251	7.815	11.345
4	3.357	5.989	7.779	9.488	13.277
5	4.351	7.289	9.236	11.070	15.086
6	5.348	8.558	10.645	12.592	16.812
7	6.346	9.803	12.017	14.067	18.475
8	7.344	11.030	13.362	15.507	20.090
9	8.343	12.242	14.684	16.919	21.666
10	9.342	13.442	15.987	18.307	23.209

Anmerkung 2: Da der *Satz von Pearson* eine asymptotische Aussage ist, ist ein Niveau $\cong \alpha$ nur dann gesichert, wenn n "hinreichend groß" ist. Eine gebräuchliche Faustregel, dies zu überprüfen, ist $n \times \min \{p_0, p_1, \dots, p_m\} \geq 5$.

Fortsetzung des einführenden Beispiels:

Da die Hypothese H_0 , dass der Würfel fair ist, durch die Wahrscheinlichkeitsverteilung $P = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ beschrieben wird und der Stichprobenumfang $n = 600$ ist, ist $n \times \min \{p_0, p_1, \dots, p_m\} = 600 \times \frac{1}{6} = 100$. Also ist die Faustregel bestens erfüllt, sodass wir den χ^2 -Test bedenkenlos anwenden können. Der beobachtete χ^2 -Wert ist gleich

$$\begin{aligned}
 \chi_5^2 &= \frac{\frac{(142-100)^2}{100}}{\frac{(102-100)^2}{100}} + \frac{\frac{(79-100)^2}{100}}{\frac{(110-100)^2}{100}} + \frac{\frac{(92-100)^2}{100}}{\frac{(75-100)^2}{100}} + \\
 &= \frac{1}{100} [42^2 + 21^2 + 8^2 + 2^2 + 10^2 + 25^2] \\
 &= 29.98.
 \end{aligned}$$

Die Anzahl der Freiheitsgrade ist $6 - 1 = 5$. Wegen $29.98 \geq \chi_{5,1-0.01}^2 = 15.086 > \chi_{5,1-0.05}^2 = 11.07$ ist die Hypothese H_0 , dass der Würfel fair ist, nicht bloß auf dem Niveau $\alpha = 0.05$, sondern auch auf dem Niveau $\alpha = 0.01$ zu verwerfen. Der P -Wert ist übrigens $P(\chi_5^2 \geq 29.98) \cong 1.5 \times 10^{-5}$. Demgemäß ist die durch die Daten gegebene Evidenz gegen H_0 sehr stark.²

² Ronald A. Fisher hat maßgeblich dazu beigetragen, die Niveaus $\alpha = 0.05$ und

3.1.2 Anpassung von Modellen

Im Folgenden behandeln wir die Anpassung von Wahrscheinlichkeitsmodellen. In diesem Fall ist eine Familie von Wahrscheinlichkeitsverteilungen gegeben, deren Parameter aus den Daten geschätzt wird. Anschließend wird mit Hilfe des χ^2 -Tests untersucht, ob die durch den Schätzwert bestimmte Verteilung dem zugehörigen Histogramm gut angepasst ist.

Seien X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariable mit endlichem oder abzählbar unendlichem Wertebereich W_{X_1} . Ferner sei I ein offenes, nicht notwendigerweise endliches Intervall der Menge \mathbb{R} der reellen Zahlen und

$$P_\theta, \theta \in I \subseteq \mathbb{R}$$

eine Familie von (diskreten) Wahrscheinlichkeitsverteilungen auf W_{X_1} .

Um die Nullhypothese

$$H_0 : \exists \theta \in I : X_1 \sim P_\theta$$

gegen die Alternativhypothese

$$H_1 : \exists \text{ kein } \theta \in I : X_1 \sim P_\theta$$

auf der Basis der Stichprobe (x_1, \dots, x_n) zu testen,

- ermittelt man zunächst den Schätzwert $\hat{\theta}_n = \hat{\theta}(x_1, \dots, x_n)$ des Parameters θ mit Hilfe eines geeigneten Schätzers $\hat{\theta}_n$,
- macht nötigenfalls eine geeignete Klasseneinteilung $\{K_0, K_1, \dots, K_m\}$ des Wertebereichs W_{X_1} der betrachteten Zufallsvariablen, sodass die Faustregel $n \times \min \{P_{\hat{\theta}_n}(K_0), \dots, P_{\hat{\theta}_n}(K_m)\} \geq 5$ erfüllt ist,
- berechnet den Wert χ^2 der zugehörigen χ^2 -Statistik

$$\chi_{m-d}^2 = \chi^2(\mathbb{B}_m^{(n)}, nP_{\hat{\theta}_n}) = \sum_{j=0}^m \frac{(B(K_j) - nP_{\hat{\theta}_n}(K_j))^2}{nP_{\hat{\theta}_n}(K_j)}$$

- und verwirft H_0 zugunsten H_1 auf dem Niveau $\cong \alpha$, sofern

$$\chi_{m-d}^2 \geq \chi_{m-d, 1-\alpha}^2$$

ist.

$\alpha = 0.01$ populär zu machen. In einer Zeit, in welcher die Berechnung des P -Werts aufgrund mangelnder Rechenhilfen sehr mühsam war, war es auch naheliegend, die kritischen Werte für einige wenige Niveaus tabellarisch festzuhalten. Dies umso mehr, wenn die Prüfverteilung - wie z.B. die χ^2 -Verteilung - noch dazu von einem Parameter abhängt.

Die Anzahl der Freiheitsgrade ist dabei gleich der Anzahl der Klassen vermindert um 1 und überdies vermindert um die Dimension d des zu schätzenden Parameters³. Das ist $(m + 1) - 1 - d = m - d$.

Anwendungsbeispiel: Der radioaktive Zerfall

In diesem Anwendungsbeispiel wollen wir nachprüfen, dass - wie behauptet wird - die Verteilung der Anzahl der mit einem *Geiger-Müller*-Zähler in einem Zeitintervall vorgegebener Länge registrierten Szintillationen tatsächlich gut durch eine Poissonverteilung beschrieben wird.

E. Rutherford und *H. Geiger* haben bei ihrem klassischen, im Jahre 1910 durchgeführten Experiment eine Poloniumquelle benützt und die Szintillationen von 2608 disjunkten Zeitintervallen von je 7.5 Sekunden Dauer registriert.

Es bezeichnen

$b(j)$... die beobachteten Häufigkeiten der Zeitintervalle mit j Szintillationen und

$e_{\hat{\lambda}}(j) = 2608 \cdot \frac{\hat{\lambda}^j e^{-\hat{\lambda}}}{j!}$... die unter der Annahme der Poissonverteilung zu erwartenden Häufigkeiten, $j \in \{0, 1, \dots, 14\}$,

wobei $\hat{\lambda}$ der durch das Stichprobenmittel $\hat{\lambda} = \frac{1}{2608} \sum_{j=0}^{14} j \cdot b(j) = 3.872$ geschätzte Parameter ist. Dies deswegen, weil eine gemäß einer Poissonverteilung mit dem Parameter λ verteilte Zufallsgröße X den Erwartungswert $E(X) = \lambda$ hat und das Stichprobenmittel der erwartungstreue Schätzer von λ mit kleinster Varianz ist.

j	0	1	2	3	4	5	6	7	8	9
$b(j)$	57	203	383	525	532	408	273	139	45	27
$e_{\hat{\lambda}}(j)$	54.3	210.3	407.1	525.3	508.4	393.7	254.0	140.5	68.0	29.2

j	10	11	12	13	14	≥ 15
$b(j)$	10	4	0	1	1	0
$e_{\hat{\lambda}}(j)$	11.3	4.0	1.3	0.4	0.1	0.1

Im vorliegenden Fall ist es daher naheliegend, die folgenden Klassen zu wählen, welche offensichtlich alle die genannte Faustregel erfüllen

$$K_0 = \{0\}, K_1 = \{1\}, \dots, K_{10} = \{10\} \quad \text{und} \quad K_{11} = \{11, 12, \dots\}.$$

Der beobachtete χ^2 -Wert ist 12.961.

³salopp ausgedrückt: überdies vermindert um die "Anzahl der zu schätzenden Parameter"

Da im vorliegenden Fall die Anzahl der Klassen 12 und die Dimension des zu schätzenden Parameters $d = 1$ ist, ist die Anzahl der Freiheitsgrade gleich 10.

Wegen $P(\chi_{10}^2 \geq 12.961) > 0.2$ gibt es im vorliegenden Fall daher tatsächlich keinen Grund, die Nullhypothese zu verwerfen, dass die zugrundeliegende Verteilung eine Poissonverteilung ist.

Anmerkung 3: Bei genauerer Analyse der Funktionsweise eines Geiger-Müller-Zählers müsste man dessen sogenannte Totzeit berücksichtigen. Zur Beschreibung dieses Phänomens sei auf die Abschnitte 12 und 16 in [13] verwiesen.

3.1.3 Test zweier Wahrscheinlichkeitsverteilungen auf Gleichheit (Test auf Homogenität)

Wir gehen im Folgenden von zwei Stichproben mit den Stichprobenumfängen n_0 und n_1 aus. $N = n_0 + n_1$ bezeichne den Gesamtumfang beider Stichproben.

Die erste Stichprobe sei durch n_0 unabhängige, identisch verteilte Zufallsvariable X_{01}, \dots, X_{0n_0} gegeben, deren Wertebereich $W = \{\omega_0, \omega_1, \dots, \omega_m\}$ und deren Wahrscheinlichkeitsverteilung $P_0 = (p_{00}, p_{01}, \dots, p_{0m})$ ist. $\mathbb{B}_m^{(n_0)} = (B_{00}, B_{01}, \dots, B_{0m})$ bezeichne den zugehörigen Vektor der Ausfallshäufigkeiten

$$B_{0j} = |\{i \in \{1, \dots, n_0\} : X_{0i} = \omega_j\}|, \quad j \in \{0, 1, \dots, m\}.$$

Die zweite Stichprobe sei durch n_1 unabhängige, identisch verteilte Zufallsvariable X_{11}, \dots, X_{1n_1} mit demselben Wertebereich $W = \{\omega_0, \omega_1, \dots, \omega_m\}$ und der Wahrscheinlichkeitsverteilung $P_1 = (p_{10}, p_{11}, \dots, p_{1m})$ gegeben. $\mathbb{B}_m^{(n_1)} = (B_{10}, B_{11}, \dots, B_{1m})$ bezeichne den zugehörigen Vektor der Ausfallshäufigkeiten

$$B_{1j} = |\{i \in \{1, \dots, n_1\} : X_{1i} = \omega_j\}|, \quad j \in \{0, 1, \dots, m\}.$$

Die beiden Vektoren sind in der folgenden $2 \times (m+1)$ -Tafel zusammen mit den Zeilen- und Spaltensummen dargestellt.

Ausfall	ω_0	ω_1	...	ω_m	
erste Stichprobe	B_{00}	B_{01}	...	B_{0m}	n_0
zweite Stichprobe	B_{10}	B_{11}	...	B_{1m}	n_1
	$B_{00} + B_{10}$	$B_{01} + B_{11}$...	$B_{0m} + B_{1m}$	N

Unsere Aufgabe sei es,

die Nullhypothese $H_0 : P_1 = P_0$ gegen die Alternativhypothese $H_1 : P_1 \neq P_0$ zu testen.

Da die zugehörigen Vektoren der Erwartungswerte der Beobachtungshäufigkeiten unter der Annahme der Nullhypothese offensichtlich die entsprechenden Vielfachen des Vektors $\hat{P} = (\frac{B_{00}+B_{10}}{N}, \frac{B_{01}+B_{11}}{N}, \dots, \frac{B_{0m}+B_{1m}}{N})$, nämlich $\mathbb{E}_m^{(n_0)} = n_0 \hat{P}$ und $\mathbb{E}_m^{(n_1)} = n_1 \hat{P}$ sind, ist die zugehörige χ^2 -Statistik

$$\chi_m^2 = \sum_{j=0}^m \sum_{i=0}^1 \frac{(B_{ij} - E_{ij})^2}{E_{ij}} = \frac{1}{n_0(1 - \frac{n_0}{N})} \sum_{j=0}^m \frac{(B_{0j} - n_0 \frac{B_{0j}+B_{1j}}{N})^2}{\frac{B_{0j}+B_{1j}}{N}}.$$

Da unter der Nullhypothese die Wahrscheinlichkeitsverteilung \hat{P} als gegeben angenommen werden kann und die Stichprobenumfänge zudem fest vorgegeben sind (d.h. dass alle Spalten- und Zeilensummen unserer $2 \times (m+1)$ -Tafel fest sind), können von den $(m+1)^2$ Eintragungen nur m frei gewählt werden. Daher ist die Anzahl der Freiheitsgrade der χ^2 -Statistik gleich m .

Bei gegebenem Signifikanzniveau α ist die Nullhypothese H_0 zugunsten der Alternativhypothese H_1 demnach dann zu verwerfen, wenn der beobachtete Wert von $\chi_m^2 \geq \chi_{m,1-\alpha}^2$ ist.

Chapter 4

ÜBUNGSAUFGABEN

1. Augensumme beim Wurf mit drei Würfeln

Man werfe $6^3/2 = 108$ mal drei symmetrische Würfel, stelle die jeweilige Augensumme X fest und fertige eine Strichliste an. Danach erstelle man je ein Histogramm für die Häufigkeit der Ausfälle von X

a) für die einzelnen Werten $j \in \{3, \dots, 18\}$ des Wertebereichs W_X

b) für die Klassen $K_1 = \{3, 4\}$, $K_2 = \{5, 6\}$, $K_3 = \{7, 8\}$, $K_4 = \{9, 10\}$, $K_5 = \{11, 12\}$, $K_6 = \{13, 14\}$, $K_7 = \{15, 16\}$, $K_8 = \{17, 18\}$

c) für die Klassen $K_1 = \{3, 4, 5, 6\}$, $K_2 = \{7, 8, 9, 10\}$, $K_3 = \{11, 12, 13, 14\}$, $K_4 = \{15, 16, 17, 18\}$

Der Wertebereich der Augensumme X ist die Menge $W_X = \{3, \dots, 18\}$.

2. Beispiel 2 aus Abschnitt 1.1.1

3. Schwangerschaftsdauern (siehe [6], S. 407)

Die nachstehende Tabelle zeigt die in einem der *County General Hospitals* in den U.S.A. im Jahre 1978 erhobenen Dauern von 70 Schwangerschaften.

(a) Erstellen Sie ein *Stängel-Blatt-Diagramm* für die ihrer Größe nach geordneten Daten, (b) zeichnen Sie die zugehörige empirische Verteilungs-

funktion und (c) fertigen Sie ein Kastenbild an.

251	264	234	283	226	244	269	241	276	274
263	243	254	276	241	232	260	248	284	253
265	235	259	279	256	256	254	256	250	269
240	261	263	262	259	230	268	284	259	261
268	268	264	271	263	259	294	259	263	278
267	293	247	244	250	266	286	263	274	253
281	286	266	249	255	233	245	266	265	264

4. Am National Bureau of Standards in Washington, U.S.A., wurden in den Jahren 1962 – 1963 Messungen des dortigen Standardgewichts NB 10 (Nominalwert von 10 Pond) mit einem der besten Meßgeräte und unter Gewährleistung von möglichst gleichbleibenden Bedingungen durchgeführt. Da alle Messwerte etwa 400 Mikropond unter dem Nominalwert lagen, war es vorteilhaft, die Differenzen der Messwerte vom Nominalwert in Mikropond anzugeben und als Stichprobenwerte zu verwenden. (So ist z.B. für einen Messwert von 9.999591 Pond diese Differenz $10 - 9.999591 = 0.000409$ Pond oder 409 Mikropond.) Bei 100 aufeinanderfolgenden Messungen ergaben sich folgende Stichprobenwerte (vgl. [1], § 6 Measurement Error).

409	400	406	299	402	406	401	403	401	403
398	403	407	402	401	399	400	401	405	402
408	399	399	402	399	397	407	401	399	401
403	400	410	401	407	423	406	406	402	405
405	409	399	402	407	406	413	409	404	402
404	406	407	405	411	410	410	410	401	402
404	405	392	407	406	404	403	408	404	407
412	406	409	400	408	404	401	404	408	406
408	406	401	412	393	437	418	515	404	401
401	407	412	375	409	406	398	406	403	404

(a) Erstellen Sie ein *Stängel-Blatt-Diagramm* für die ihrer Größe nach geordneten Daten, (b) zeichnen Sie die zugehörige empirische Verteilungsfunktion, (c) fertigen Sie ein Kastenbild an und (d) tragen Sie die Daten in das Wahrscheinlichkeitsnetz ein.

5. Changing the choice of classes can change the appearance of a histogram. Here is an example in which a small shift in the classes, with

no change in the number of classes, has an important effect on the histogram. The data are acidity levels (measured by pH) in 105 samples of rainwater. Distilled water has pH 7.00. As the water becomes more acid, the pH goes down. The pH of rainwater is important to environmentalists because of the problem of acid rain (aus [10], S. 35).

4.33	4.38	4.48	4.48	4.50	4.55	4.59	4.59	4.61	4.61
4.75	4.76	4.78	4.82	4.82	4.83	4.86	4.93	4.94	4.94
4.94	4.96	4.97	5.00	5.01	5.02	5.05	5.06	5.08	5.09
5.10	5.12	5.13	5.15	5.15	5.15	5.16	5.16	5.16	5.18
5.19	5.23	5.24	5.29	5.32	5.33	5.35	5.37	5.37	5.39
5.41	5.43	5.44	5.46	5.46	5.47	5.50	5.51	5.53	5.55
5.55	5.56	5.61	5.62	5.64	5.65	5.65	5.66	5.67	5.67
5.68	5.69	5.70	5.75	5.75	5.75	5.76	5.76	5.79	5.80
5.81	5.81	5.81	5.81	5.85	5.85	5.90	5.90	6.00	6.03
6.03	6.04	6.04	6.05	6.06	6.07	6.09	6.13	6.21	6.34
6.43	6.61	6.62	6.65	6.81					

(a) Make a histogram of pH with 14 classes, using class boundaries 4.2, 4.4, ..., 7.0. How many modes does your histogram show? More than one mode suggests that the data contain groups that have different distributions.

(b) Make a second histogram, also with 14 classes, using class boundaries 4.14, 4.34, ..., 6.94. The classes are those from (a) moved 0.06 to the left. How many modes does the new histogram show?

(c) Use your software's histogram function to make a histogram without specifying the number of classes and their boundaries. How does the software's default histogram compare with those in (a) and (b)?

(d) Make a normal quantile plot of these data.

6. In 1798 the English scientist *Henry Cavendish* measured the density of the earth by careful work with a torsion balance. The variable recorded was the density of the earth as a multiple of the density of water. Here are the Cavendish's 29 measurements (aus [10], S. 37):

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.58	

- (a) Present the measurements graphically by either a stemplot or a histogram and explain the reason for your choice.
- (b) Then briefly discuss the main features of the distribution. In particular, what is your estimate of the density of the earth based on these measurements?
- (c) Find \bar{x}_n and s_n for these data.
7. How much oil the wells in a given field will ultimately produce is key information in deciding whether to drill more wells. Here are the estimated total amounts of oil recovered from 64 wells in the Devonian Richmond Dolomite area of the Michigan basin, in thousand of barrels (aus [10], S. 37):

21.7	53.2	46.4	42.7	50.4	97.7	103.1	51.9	43.4	69.5
156.5	34.6	37.9	12.9	2.5	31.4	79.5	26.9	18.5	14.7
32.9	196.0	24.9	118.2	82.2	35.1	47.6	54.2	63.1	69.8
57.4	65.6	56.4	49.4	44.9	34.6	92.2	37.0	58.8	21.3
36.6	64.9	14.8	17.6	29.1	61.4	38.6	32.5	12.0	28.3
204.9	44.5	10.3	37.7	33.7	81.1	12.1	20.1	30.5	7.1
10.1	18.0	3.0	2.0						

Graph the distribution and describe the main features.

8. Many businesses operate call centers to serve customers who want to place an order or make an inquiry. Customers want their requests handed thoroughly. Businesses want to treat customers well, but they also want to avoid wasted time on the phone. They therefore monitor the length of calls and encourage their representatives to keep calls short. Here is an example of the difficulties this policy can cause.

We have data on the length of all 31 492 calls made to the customer service center of a small bank in a month. The following table displays the length of the first 80 calls. The complete data set is *eg01-004.dat*, which you can find on the text CD and web site (aus [10], S. 9 f.).

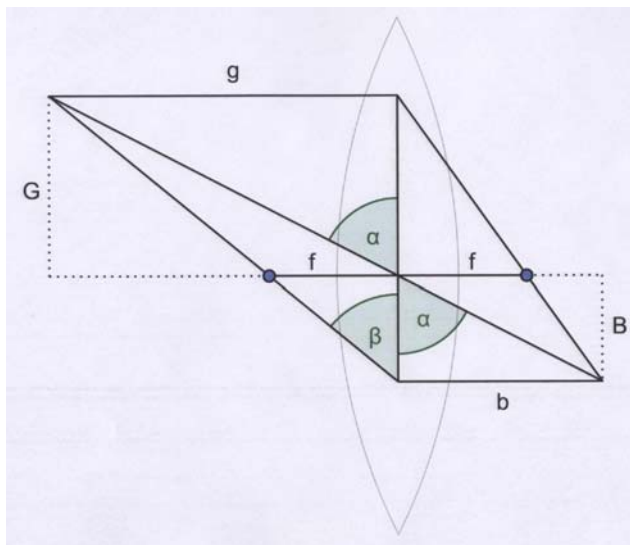
Take a look at the data in the following table. The numbers are meaningless without some background information. The *individuals* are calls made to the bank's call center. The *variable* recorded is the length of each call. The *units* are seconds. We see that the call lengths vary a great deal. The longest call lasted 2631 seconds, almost 44 minutes.

More striking is that 8 of these 80 calls lasted less than 10 seconds. What's going on?

77	289	128	59	19	148	157	203
126	118	104	141	290	48	3	2
372	140	438	56	44	274	479	211
179	1	68	386	2631	90	30	57
89	116	225	700	40	73	75	51
148	9	115	19	76	138	178	76
67	102	35	80	143	951	106	55
4	54	137	367	277	201	52	9
700	182	73	199	325	75	103	64
121	11	9	88	1148	2	465	25

(a) Order the data from smallest to largest and display them in form of a stemplot, (b) make a histogram with a suitable number of classes and suitable class lengths, (c) draw the empirical distribution function and (d) a normal quantile plot.

9. Eine Bakterienkultur wächst während der ca. 16 Tagstunden um 20 % pro Stunde und während der Nacht um nur 12 % pro Stunde. Berechne das durchschnittliche Wachstum pro Stunde! (Beispiel E 1418 aus [4])
10. Bei einem Autorennen sind 5 Runden zu fahren. Die mittleren Geschwindigkeiten in den einzelnen Runden betragen für einen Fahrer: 183, 210, 201, 180, 182 (km/h). Wie groß ist die mittlere Geschwindigkeit für alle 5 Runden? (Beispiel E 1419 aus [4]).
11. Von zwei Körpern mit einem Gewicht von jeweils 12 kg verdrängt der eine 2 l und der andere 1 l Wasser. Die Dichten der beiden Körper sind daher 6 kg/dm^3 bzw. 12 kg/dm^3 . Berechnen Sie die durchschnittliche Dichte der beiden Körper (d.i. die Dichte jenes Körpers, welcher durch Zusammenfügen der beiden einzelnen Körper entsteht.)
12. Für eine *bikonvexe Linse* bezeichne
 G ... die Gegenstandsgröße, g ... die Gegenstandsweite,
 B ... die Bildgröße, b ... die Bildweite und



f ... die Brennweite.

(a) Entnehmen Sie der obigen Abbildung - mittels Ähnlichkeit zweier Dreiecke - zunächst die Beziehung (1_α) und leiten Sie mit deren Hilfe aus einer weiteren analog entnommenen Beziehung (1_β) die Abbildungsgleichung (2) her:

$$\text{Abbildungsmaßstab} \quad \frac{G}{g} = \frac{B}{b} \quad (1_\alpha)$$

$$\text{Abbildungsgleichung} \quad \frac{1}{f} = \frac{1}{g} + \frac{1}{b} \quad (2)$$

(b) Überzeugen Sie sich davon, dass $2f$ das harmonische Mittel von g und b ist.

13. Führen Sie eigene Versuche durch, um die Zahl π zu schätzen und die Beziehung $u = \pi \cdot d$ zwischen Kreisumfang u und Kreisdurchmesser d experimentell zu prüfen.

Orientieren Sie sich dabei an Beispiel 2 von Abschnitt 1.4.

14. Der Franziskanermönch, Mathematiker und Musiktheoretiker *Marin Mersenne* (1588 – 1648), ein Zeitgenosse von *Galilei*, hat Fallversuche durchgeführt, um die Erdbeschleunigung g zu bestimmen. Die in seinem Werk *Harmonie Universelle* (Paris 1636) veröffentlichten Messergeb-

nisse sind in der folgenden Tabelle dargestellt,

Falldauer	Fallhöhe
1/2	3
1	12
2	48
3	110
3 1/2	146 1/2

wobei die Falldauern t_i in Sekunden und die Fallhöhen s_i in "königlichen" Fuß (32.87 cm) angegeben sind.

Ermitteln Sie aus diesen Daten und der Beziehung $s = \frac{g}{2}t^2$ einen Schätzwert für g . Berücksichtigen Sie bei der Interpretation des Ergebnisses freilich, dass die Zeitmessung für *Mersenne* noch ein beträchtliches Problem war.

15. **Hubble's law** (siehe [6], Case Study 10.4.1)

One of the most startling and profound scientific revelations of the twentieth century was the discovery in 1929 by the American astronomer *Edwin Hubble* that the universe is expanding. Hubble's announcement shattered forever the ancient belief that the heavens are basically in a state of cosmic equilibrium: quite the contrary, galaxies are receding from each other at fantastic velocities. If v is a galaxy's recession velocity (relative to any other galaxy) and d , its distance (from that other galaxy), Hubble's law states that

$$v = H \cdot d,$$

where H is known as Hubble's constant. The following table lists distance (in millions of light years) and velocity (thousands of miles/sec) measurements made on 11 galactic clusters. Estimate Hubble's con-

stant and from that the age $1/H$ of the universe.

Cluster	Distance	Velocity
Virgo	22	0.75
Pegasus	68	2.4
Perseus	108	3.2
Coma Berenices	137	4.7
Ursa Major No. 1	255	9.3
Leo	315	12.0
Corona Borealis	390	13.4
Gemini	405	14.4
Bootes	685	24.5
Ursa Major No. 2	700	26.0
Hydra	1100	38.0

16. Das Hooksche Gesetz

Die Feder einer Federwaage habe die Länge von d cm, wenn sie ohne Gewicht ist. Wenn ein Gewicht von x Kilogramm an den Haken am Ende der Feder gehängt wird, wird die Feder zur neuen Länge von y cm gestreckt. Aufgrund des Hookschen Gesetzes gilt (selbstverständlich für nicht allzu große Gewichte), dass der Längenzuwachs proportional zum Gewicht ist. Für die Länge y gilt somit die lineare Beziehung

$$y = k \cdot x + d.$$

Folgende Wertepaare (x_i, y_i) wurden ermittelt:

$(0, 439.00)$, $(2, 439.12)$, $(4, 439.21)$, $(6, 439.31)$, $(8, 439.40)$, $(10, 439.50)$.

Ermitteln Sie die Regressionsgerade und das Bestimmtheitsmaß.

17. Aufgabe 1 aus Abschnitt 2.2: Vergleichen Sie das empirische Histogramm mit dem Histogramm der zu erwartenden Häufigkeiten für eine (a) $B_{74,1/37}$ -verteilte Zufallsvariable und (b) eine P_2 -verteilte Zufallsvariable.
18. In der nachstehenden Tabelle sind die absoluten Häufigkeiten H_j der Tage dargestellt, an den in Oxford (England) in den Jahren 1925–1930 ($N = 2191$ Tage) $j \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ Erdbeben (ohne Nachbeben) registriert wurden (aus [12], S. 70 f.).

j	0	1	2	3	4	5	6	7
H_j	685	792	467	160	68	13	5	1

(a) Erstellen Sie ein zugehöriges Histogramm und (b) vergleichen Sie dieses mit dem Histogramm einer Poissonverteilung P_λ , deren Parameter λ durch das Stichprobenmittel

$$\hat{\lambda} = \frac{1}{N} \sum_{j=0}^7 j \cdot H_j$$

gegeben ist. (c) Führen Sie einen χ^2 -Test auf Anpassung der Daten durch diese Poissonverteilung durch, wobei Sie die Klassen gemäß der Faustregel $n \times \min \{P_{\hat{\lambda}_n}(K_0), \dots, P_{\hat{\lambda}_n}(K_m)\} \geq 5$ festlegen.

19. Experiment 1 in Abschnitt 2.3: Führen Sie dieses Experiment durch und vergleichen Sie das empirische Histogramm mit dem Histogramm der zu erwartenden Häufigkeiten für eine $G_{1/2}$ -verteilte Zufallsvariable.
20. Over "short" geological periods, a volcano's eruptions are believed to be Poisson events - that is, they are thought to occur independently and at a constant rate. If so, the pdf (probability density function) describing the intervals between eruptions should have the form

$$f_\lambda(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

Collected for the purpose of testing that presumption are the data in the following table, showing the intervals (in month) that elapsed between thirty-seven consecutive eruptions of Mauna Loa, a fourteen thousand-foot volcano in Hawaii. During the period covered - 1832 to 1950 - eruptions were occurring at the rate of $\lambda = 0.027$ per month (or once every 3.1 years). Is the variability in these thirty-six consecutive t_i s consistent with the exponential law? (siehe [7], Case Study 4.2.4)

126	73	3	6	37	23
73	23	2	65	94	51
26	21	6	68	16	20
6	18	6	41	40	18
41	11	12	38	77	61
26	3	38	50	91	12

Bibliography

- [1] *Freedman, D., Pisani, R. and R. Purves*: Statistics. Norton & Co., New York 1978
- [2] *Barth·Haller*: Stochastik Leistungskurs. Ehrenwirth Verlag, München 1983
- [3] *Hartung, J., Elpelt, B. und K-H. Klösener*: Statistik: Lehr- und Handbuch der angewandten Statistik. Oldenbourg Verlag, München - Wien 1991
- [4] *Kronfellner, M.Peschek·Blasonig·Fischer·Kronfellner,J.:* Angewandte Mathematik 4. Verlag Hölder Pichler Tempsky, Wien 1997
- [5] *Krämer, W.:* Statistik verstehen: Eine Gebrauchsanweisung. Campus Verlag, Frankfurt - New York, 1999
- [6] *Larsen, R.J. and M.L. Marx*: An Introduction to Mathematical Statistics and its Applications. Prentice Hall, Englewood Cliffs, New Jersey 1986
- [7] *Larsen, R.J. and M.L. Marx*: An Introduction to Mathematical Statistics and its Applications. Pearson Prentice Hall, Upper Saddle River, New Jersey 2006
- [8] *Marsal, D.:* Statistische Methoden für Erdwissenschaftler. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart 1967
- [9] *Moore, D.S.:* Statistics: Concepts and Controversies. W.H. Freeman & Co., New York 2001

- [10] *Moore, D.S. and G.P. McCabe*: Introduction to the Practice of Statistics. W.H. Freeman & Co., New York 2004
- [11] *Österreicher, F.*: Die Normalverteilung in Wort und Bild - Einführung in die stochastische Modellbildung am Beispiel der Meßfehler. In: Ausflüge in die Mathematik - 21 Jahre Institut für Mathematik, Abakus Verlag, Salzburg 1988
- [12] *Schönwiese, Ch.-D.*: Praktische Statistik für Meteorologen und Geowissenschaftler. Gebrüder Borntraeger, Berlin-Stuttgart 2006
- [13] *Topsoe, F.*: Spontane Phänomene - Stochastische Modelle und ihre Anwendungen. Vieweg, Braunschweig - Wiesbaden 1983

Zeitschriften

- [14] *Engel, A.*: Statistik in der Schule: Ideen und Beispiele aus neuerer Zeit. Der Mathematikunterricht, Jahrgang 28, Heft 1 (1982), S. 57 – 85
- [15] *Kütting, H.*: Stochastisches Denken in der Schule - Grundlegende Ideen und Methoden. Der Mathematikunterricht, Heft 4 (1985), S. 87 – 106

Skripten

- [16] *Österreicher, F.*: Ausgewählte Kapitel der Statistik. LV-Unterlagen, Salzburg 1986
- [17] *Österreicher, F.*: Skriptum zur Lehrveranstaltung Statistik für Lehramt. Salzburg 2007

Diplomarbeiten

- [18] *Weiß, M.*: Binomialverteilung und Normalapproximation: Grundlegendes und Hintergrundinformation für den Stochastikunterricht. Diplomarbeit, Salzburg 1995
- [19] *Jandl, M.*: Computereinsatz im Stochastikunterricht. Diplomarbeit, Salzburg 1997
- [20] *Huber, Ch.*: Die Radioaktivität als Hilfsmittel der Geochronologie - Von Röntgen bis DAREosDAT. Diplomarbeit, Salzburg 2008