Wahrscheinlichkeitsrechnung und Statistik

Dr. Sven Feldmann, svendrfeldmann@yahoo.de

Inhaltsverzeichnis

1 Einführung

Man unterscheidet zwischen der schließenden und der beschreibenden Statistik. Aufgabe der beschreibenden Statistik ist die Kompression unübersichtlicher Datenmengen in einige wenige aussagekräftige Kenngrößen. Solche Kenngrößen können der Mittelwert, die mittlere quadratische Streuung, die Spannweite und der Zentralwert sein, falls es sich bei den Daten um numerische Werte handelt, wie die Größe eines Werkstücks, der Abfüllmenge in einer Packung oder die Besucherzahlen eines Kinofilms.

Unter Umständen können die Datenmengen so groß sein, daß eine vollständige Erhebung aus Zeit- und Kostengründen nicht möglich ist. Dann muß man sich mit dem Ziehen von sogenannten Stichproben begnügen und versuchen, aus diesem Stichprobenmaterial zuverlässige Aussagen über die Kenngrößen des unzugänglichen gesamten Datenmaterials zu bekommen. Den Grad der Zuverlässigkeit einer so gewonnenen Aussage zu messen, ist Aufgabe der schließenden Statistik. Beispielsweise kann man am Mittelwert der Länge von 1 Mio Werkstücken (Grundgesamtheit) interessiert sein, man aber nur über eine Kapazität verfügen, die das Vermessen von 100 zufällig ausgewählten Werkstücken gestattet. Welchen Aussagewert besitzt der Mittelwert dieser Stichprobe im Umfang von 100 in Bezug auf den tatsächlichen Mittelwert der Grundgesamtheit? Eine Bemessungsgrundlage dafür liefert die Wahrscheinlichkeitsrechnung. Mit ihrer Hilfe ist man in der Lage, aus der Stichprobe ein Vertrauensintervall zu berechnen, daß mit einer gewissen Wahrscheinlichkeit den Mittelwert der Grundgesamtheit überdeckt. In natürlicher Weise schließt sich die Frage an nach dem Umfang, den eine Stichprobe wenigstens haben muß, um ein Vertrauensintervall mit einer vorgegebenen Überdeckungswahrscheinlichkeit konstruieren zu können. Deswegen besteht der erste Teil der Vorlesung aus für uns nützlichen Elementen dieser Theorie, die sich auf die Begriffe

- Ereignis
- Wahrscheinlichkeit
- Unabhängigkeit von Ereignissen,
- Zufallsgröße (diskret und stetig),
- Erwartungswert
- Varianz und Standardabweichung
- Verteilungsfunktion

stützen. Im Rahmen des Verteilungsbegriffs werden die für uns wichtigen Beispiele der

- Binomialverteilung $\mathscr{B}(n,p)$ (diskret)
- Poisson-Verteilung $\mathscr{P}(\lambda)$ (diskret)

- Geometrischen Verteilung $\mathscr{G}(p)$ (diskret)
- Hypergeometrischen Verteilung $\mathscr{H}(M, N, n)$ (diskret)
- Gleichverteilung (diskret, stetig)
- Dreiecksverteilung (diskret, stetig)
- Exponentialverteilung $\mathscr{E}(\lambda)$ (stetig)
- Normalverteilung $\mathscr{N}(\mu, \sigma^2)$ (stetig)
- χ^2-Verteilung (stetige Testverteilung)
- STUDENTsche oder t-Verteilung (stetige Testverteilung)
- WEIBULL-Verteilung (stetig)
- F-Verteilung (stetige Testverteilung)

behandelt und in Beziehung zueinander gesetzt.

Es schließt sich ein Teil an, der der **schließenden** Statistik gewidmet ist. In diesem Rahmen wird eine Stichprobentheorie entwickelt, die unter anderem folgendes beinhaltet.

- Es wird die landläufige Vorstellung von einer Stichprobe in die Wahrscheinlichkeitstheorie eingebettet.
- Es wird der Begriff der Stichprobenfunktion beleuchtet.
- Es werden spezielle Stichprobenfunktionen, die man auch als Schätzfunktionen oder kurz als Schätzer bezeichnet, eingeführt, um die Parameter der Verteilungsfunktion für eine Grundgesamtheit durch Erhebung von Stichproben abschätzen zu können. Diese Parameter sind bei uns der Mittelwert, die Varianz und der Anteilswert.
- Es werden sogenannte Vertrauensintervalle für oben genannte Parameter konstruiert.
- Der Wert solcher Vertrauensintervalle wird mit einer sog. Irrtumswahrscheinlichkeit α bemessen. Naturgemäß ergibt sich die Frage nach dem Umfang einer zu erhebenden Stichprobe, um daraus ein Vertrauensintervall berechnen zu können, das eine vorgegebene Vertrauenswahrscheinlichkeit besitzt.

Die Vorlesung behandelt einige Testverfahren, die sich auf die Begriffe

- Irrtumswahrscheinlichkeit, Vertrauensniveau
- Ablehnungs- und Annahmebereich

stützen. Innerhalb der Testverfahren unterscheidet man parametrische und nichtparametrische Tests. Hinsichtlich der ersten Sorte behandeln wir den Test

- für den Anteilswert p einer binomialverteilten Zufallsgröße (Zg)
- für den Mittelwert μ einer normalverteilten Zg
- für die Varianz σ^2 einer normalverteilten Zg
- auf gleiche Mittelwerte bei bekannten Varianzen zweier normalverteilte Zg
- auf gleiche Mittelwerte bei unbekannten aber gleichen Varianzen zweier normalverteilte Zg
- auf gleiche Varianzen zweier normalverteilter Zg (F-Test)

Hinsichtlich der zweiten Sorte gehen wir auf

- den χ^2-Anpassungstest und
- den Kolmogorow-Smirnow-Test ein.

Es schließt sich ein Teil an, der der **beschreibenden** Statistik gewidmet ist. Dort geht es um einige Feststellungen, die man anhand von Stichproben über die Eigenschaften (Schiefe, Wölbung, Konzentration) der Verteilungsfunktion der zugrundeliegenden Grundgesamtheit machen kann.

Das Skript endet mit einigen Übungsaufgaben einschließlich ihrer Lösungen.

2 Wahrscheinlichkeitsrechnung

Zuerst stellen wir einige Tatsachen aus dem Bereich der Kombinatorik zusammen.

2.1 Permutationen und Binomialkoeffizient

Bei einer **Permutation** handelt es sich um eine spezielle Hintereinanderanordnung der Elemente aus einer Menge $\{a_1, \ldots, a_n\}$. Für $n = 3$ stellen die geordneten Tripel

$$(a_1, a_2, a_3), \quad (a_1, a_3, a_2), \quad (a_2, a_1, a_3), \quad (a_2, a_3, a_1), \quad (a_3, a_1, a_2), \quad (a_3, a_2, a_1)$$

alle Permutation der Elemente a_1, a_2, a_3 dar. Offenbar sind das sechs Stück, was mit

$$3! := 1 \times 2 \times 3$$

übereinstimmt. Im allgemeinen lassen sich n Elemente auf

$$n! := 1 \times 2 \times 3 \times \cdots \times n$$

verschiedene Arten hintereinander anordnen. Sprechweise “n Fakultät”.

Beispiel 1 Buchproblem

Gegeben sind n nicht notwendig verschiedene Bücher, die in

$$m_1, \ldots, m_s$$

Exemplaren vorhanden sind. Dann gibt es

$$\frac{n!}{m_1! \ldots m_s!}$$

verschiedene Möglichkeiten, diese Bücher auf einem Bücherbord hintereinander anzuordnen, wobei gleiche Exemplare nicht unterschieden werden. $\diamond$

Die Zahl $\frac{n!}{m_1! \ldots m_s!}$ wird als Polynomial- oder als Multinomialkoeffizient bezeichnet. Für $s = 2$ und $k := m_1$ folgt zwangsläufig $m_2 = n - k$ resultierend in dem wichtigen Spezialfall

$$\frac{n!}{k!(n-k)!}.$$

Diesen Quotienten bezeichnet man **Binomialkoeffizienten**:

$$\boxed{\binom{n}{k} := \frac{n!}{(n-k)!k!}, \quad 0 \leq k \leq n, \quad k, n \in \mathbb{N}_0 := \{0, 1, 2 \ldots\}, \quad 0! := 1}$$

Sprechweise 'n über k'. Der Name rührt von der Beziehung

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \quad a, b \in \mathbb{R} \tag{1}$$

her, wobei $(a+b)^n$ als ein Binom bezeichnet wird und $\mathbb{R}$ die Menge aller reellen Zahlen darstellt. Offenbar ist $(a+b)^n$ die Summe aller Produkte der Form

$$\underbrace{aaabbbaab \ldots abbaaab}_{n \text{ Faktoren}}. \tag{2}$$

Die Reihenfolge der Faktoren ist aufgrund der Kommutativität des Produktes zweier reeller Zahlen für den Wert des Produkts gleichgültig. Die Anzahl aller Produkte mit k-mal a und $(n-k)$-mal b ergibt sich wie beim Buchproblem für n Bücher, wobei nur zwischen zwei verschiedenen Exemplaren in den Anzahlen k und $n-k$ unterschieden wird.

Geeignete Interpretation der Zeichenkette (2) führt zur Beantwortung folgender Frage. In einer Urne befinden sich n Kugeln

$$k_1, \ldots, k_n.$$

Wieviele Möglichkeiten gibt es, k Kugeln aus dieser Urne zu ziehen?

Interpretiert man ein a in der Kette (2) an der Stelle i so, daß die Kugel k_i gezogen wurde, so stimmt die gesuchte Anzahl mit der Anzahl aller Ketten überein, die genau k-mal ein a enthalten. Demnach stimmt die gesuchte Anzahl mit dem Koeffizienten des Produkts $a^k b^{n-k}$ in (1) überein, d.h. es gibt

$$\binom{n}{k}$$

Möglichkeiten k Kugeln aus n auszuwählen.

Der Binomialkoeffizient besitzt folgende wichtige Eigenschaften:

$$\binom{n}{0} = 1, \quad \binom{n}{1} = n, \quad \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

$$\underbrace{\binom{n}{k} = \binom{n}{n-k},}_{\text{Symmetrie}} \qquad \underbrace{\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}}_{\text{Additionstheorem, Pascalsches Dreieck}} \qquad . \tag{3}$$

Falls die natürlichen Zahlen M, N, n die Ungleichungen $0 \leq M \leq n \leq N - M$ erfüllen, so gilt

$$\sum_{k=0}^{M} \binom{M}{k} \binom{N-M}{n-k} = \binom{N}{n}. \tag{4}$$

Diese Gleichung gestattet die Einführung der hypergeometrischen Verteilungsfunktion, die bei der Kontrolle von Handelswaren auf Fehlerhaftigkeit eine gewisse Rolle spielt.

In **mathcad** liefert die Auswertung von

$$\mathrm{combin}(n, k)$$

den Wert von $\binom{n}{k}$.

Beispiel 2 Binomialkoeffizient

An 10 Erzeugnissen wird eine Gut-Schlecht-Prüfung vorgenommen.

a) Wieviele verschiedene Prüfprotokolle sind insgesamt möglich?

Modelliert man das Testergebnis ‘gut’ durch Vergabe einer 1 und das Testergebnis ’schlecht’ durch Vergabe einer 0, so kann jedes Prüfprotokoll mit einer 10-stelligen Binärzahl identifiziert werden. Insgesamt gibt es 2^{10} verschiedene solche Binärzahlen, d.h. es existieren 1024 verschiedene Prüfprotokolle.

b) Wieviele Prüfprotokolle enthalten genau 7 mal ’gut’?

Diese Frage ist äquivalent zur Frage auf wieviel verschiedene Weisen man 7 Plätze aus 10 auswählen kann, d.h. man hat

$$\binom{10}{7} = \frac{10!}{3!7!} = \frac{8 \times 9 \times 10}{1 \times 2 \times 3} = 4 \times 3 \times 10 = 120$$

zu berechnen. Demnach existieren 120 Prüfprotokolle, in denen genau 7 mal mit 'gut' bewertet wird.

c) Wieviel Prüfprotokolle enthalten mindestens 7 mal 'gut'?

In Hinblick auf b) hat man die Summe

$$\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}$$

zu berechnen. Bei Ausnutzung des Additionstheorems (3) bekommt man

$$\underbrace{\binom{10}{7} + \binom{10}{8}}_{\binom{11}{8}} + \underbrace{\binom{10}{9} + \binom{10}{10}}_{\binom{11}{10} = 11} = \frac{9 \times 10 \times 11}{1 \times 2 \times 3} + 11 = 176$$

d.h. es existieren 176 Prüfprotokolle, in denen mindestens 7 Erzeugnisse mit 'gut' bewertet werden.

d) Wieviele Prüfprotokolle enthalten höchstens 7 mal gut?

Zu Beantwortung hat man

$$S := \binom{10}{0} + \ldots + \binom{10}{7}$$

zu berechnen. Entsprechend (3) gilt

$$\binom{10}{0} + \ldots + \binom{10}{10} = (1+1)^{10} = 2^{10} = 1024.$$

Zusammen mit b) und c) gilt daher

$$S = 1024 - \left(\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right) + \binom{10}{7} = 1024 - 176 + 120 = 968.$$

$\diamond$

2.2 Klassisches Wahrscheinlichkeitsmodell

Wir betrachten ein Zufallsexperiment und bezeichnen mit

$$\omega_1, \ldots, \omega_n$$

die Ergebnisse des Experimentierens. Wir nennen jede einelementige Menge $\{\omega_i\}$ ein **Elementarereignis** des Experiments, bezeichnen mit Ω die Menge aller ω_i und nennen jede Teilmenge

$$A \subseteq \Omega$$

ein **Ereignis**. Die Menge Ω nennen wir **Ereignisraum** oder **Gesamtereignis** und setzen vorläufig ihre Endlichkeit voraus. Zur Vereinfachung schreiben wir anstatt von $\{\omega_i\}$ oft nur ω_i.

Nun wird jedem Ereignis A durch die Vorschrift

$$\boxed{P(A) := \frac{\#A}{\#\Omega}} \tag{5}$$

eine nichtnegative Zahl zugeordnet, die **Wahrscheinlichkeit** von A heißt. Wie üblich bezeichnet

$$\#A$$

die Anzahl der Elemente von A. Da A eine Teilmenge von Ω ist, ist die Anzahl der Elemente von A kleiner oder gleich der Anzahl der Elemente von Ω. Demnach ist $P(A)$ eine Zahl zwischen 0 und 1, wobei diese beiden Werte auch angenommen werden können. Offenbar wird jedem Elementarereignis dieselbe Wahrscheinlichkeit

$$\frac{1}{\#\Omega}$$

zugeordnet. Die Annahme, daß alle Elementarereignisse gleiche Wahrscheinlichkeit besitzen, wird auch als Laplace-Annahme bezeichnet. Später werden wir sehen, daß andere Zuweisungen ebenfalls sinnvoll sind, wie das Bernoulliexperiment zeigen wird.

Die Bezeichnung $P(A)$ geht auf das französische Wort *probabilité* für Wahrscheinlichkeit zurück und die Definition von $P(A)$ auf Pierre Simon Laplace (1749-1827).

2.3 Rechnen mit Ereignissen

Für zwei disjunkte Teilmengen A und B von Ω, d.h. für zwei Ereignisse, die sich gegenseitig ausschließen, gilt offenbar

$$\#(A \cup B) = \#A + \#B.$$

Dem entspricht für die zugehörigen Wahrscheinlichkeiten die Gleichung

$$P(A \cup B) = P(A) + P(B).$$

Für zwei sich nicht notwendig ausschließende Ereignisse gilt dagegen

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

d.h. man hat die Wahrscheinlichkeit für das gleichzeitige Eintreten von A und B abzuziehen. Insbesondere gelten die Gleichungen

$$P(\emptyset) = 0, \quad P(\Omega) = 1.$$

Hier bezeichnet

$$\emptyset$$

die leere Menge.

Entsprechend [20], S. 1036, Tabelle 6.4 gilt folgende **Algebra der Ereignisse**.

	Ereignis	Interpretation	Wahrscheinlichkeit
1	Ω	Gesamtereignis	$P(\Omega) = 1$
2	$\emptyset$	unmögliches Ereignis	$P(\emptyset) = 0$
3	A	beliebiges Ereignis	$0 \leq P(A) \leq 1$
4	$A \cup B$	Es treten die Ereignisse A oder B ein	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5	$A \cap B$	Es treten die Ereignisse A und B ein	
6	$A \cap B = \emptyset$	Die Ereignisse A und B können nicht simultan eintreten	$P(A \cup B) = P(A) + P(B)$
7	$A \subseteq B$	Tritt das Ereignis A ein, so tritt auch das Ereignis B ein	$P(A) \leq P(B)$
8	$A \setminus B$	Es tritt das Ereignis A ein und das Ereignis B nicht	$P(A \setminus B) = P(A) - P(A \cap B)$

Wie üblich bezeichnet

$$A \setminus B := \{\omega \in A : \omega \notin B\}$$

die **Differenzmenge** von A und B, d.h. man nimmt aus A alle Elemente heraus, die zu B gehören.

Die Menge

$$\overline{A} := \Omega \setminus A$$

wird als das **Komplementärereignis** von A bezeichnet.

Beispiel 3 Wahrscheinlichkeitsbegriff, Geburtstagsproblem

Das folgende Geburtstagsproblem soll den Wahrscheinlichkeitsbegriff illustrieren. Auf einer Party befinden sich n Gäste. Wie groß ist die Wahrscheinlichkeit, daß mindestens zwei Gäste am gleichen Tag Geburtstag haben? Es wird vorausgesetzt, daß das Jahr 365 Tage besitzt und n kleiner oder gleich 365 ist.

Zur Lösung betrachten wir Listen ω_i der Länge n deren Einträge aus der Menge

$$\{1, 2, 3, \ldots, 365\} \subset \mathbb{N} := \{1, 2, 3, \ldots\}$$

stammen. Jede Liste codiert ein Elementarereignis: z.B. bedeutet

$$(345, 100, 2, 2, \ldots, 50) \in \{1, \ldots, 365\}^n$$

daß

Gast 1 am Tag 345
Gast 2 am Tag 100
Gast 3 am Tag 2
Gast 4 am Tag 2
$\vdots$
Gast n am Tag 50

Geburtstag hat. Das Experiment besteht darin, sich einfach n Personen einzuladen und diese nach ihren Geburtstagen zu fragen. Der Ausgang des Experiments wird in einer Liste ω_i gespeichert. Offenbar gibt es 365^n Elementarereignisse, d.h.

$$\#\Omega = 365^n.$$

Wir führen zwei Ereignisse ein:

- A ergibt sich aus allen Listen, die wenigstens zwei gleiche Einträge besitzen.
- B ergibt sich aus allen Listen, bei denen alle Einträge paarweise verschieden sind. Offenbar ist B das Komplementärereignis zu A.

Demnach gilt

$$\Omega = A \cup B, \quad A \cap B = \emptyset.$$

Laut Aufgabenstellung sind wir an $P(A)$ interessiert. Aufgrund der letzten Gleichungen gilt

$$P(A) = P(\Omega \setminus B) = \frac{\#(\Omega \setminus B)}{\#\Omega} = \frac{365^n - N}{365^n} = 1 - \frac{N}{365}, \quad N := \#B.$$

Die Zahl N stimmt offenbar mit der Anzahl aller Möglichkeiten überein, 365 verschiedene Zeichen auf n Plätze zu verteilen: d.h. man hat 365 Möglichkeiten Platz 1 zu belegen. Nachdem man sich für die Belegung von Platz 1 entschieden hat, verbleiben noch 364 Möglichkeiten, Platz 2 zu belegen, sodaß man für die Belegung der Plätze 1 und 2 mit verschiedenen Zeichen insgesamt 365×364 Möglichkeiten besitzt. So fortfahrend bekommt man

$$N = \underbrace{365 \times 364 \times \ldots (365 - n + 1)}_{n \text{ Faktoren}}.$$

Die Tabelle

n	20	23	30	40	50
$P(A)$	0.411	0.507	0.706	0.891	0.970

zeigt

$$P(A) = 1 - \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{365-n+1}{365}$$

für verschiedene n, d.h. bei einer Feier mit 40 Gästen ist es zu fast 90 % sicher, daß wenigstens zwei Gäste am selben Tag Geburtstag haben. ◇

Beispiel 4 Wahrscheinlichkeitsbegriff und Binomialkoeffizient, 6 aus 49

Das Beispiel 'Lotto am Samstag: 6 aus 49' kombiniert den Wahrscheinlichkeitsbegriff mit dem Binomialkoeffizienten.

a) Wie groß ist die Wahrscheinlichkeit, 6 Richtige zu tippen? Zur Beantwortung bilden wir den Ereignisraum Ω aus allen 6-Tupeln mit paarweise verschieden Einträgen aus der Menge $\{1, \ldots, 49\}$. Die Anzahl der Elemente von Ω stimmt mit dem Binomialkoeffizienten

$$\binom{49}{6}$$

überein. Das Ereignis A ist durch irgend ein beliebiges Element von Ω definiert. Demzufolge gilt $\#A = 1$, d.h.

$$P(A) = \frac{\#A}{\#\Omega} = \frac{1}{\binom{49}{6}} = \frac{1}{13\,983\,816} =: a.$$

b) Wie groß ist die Wahrscheinlichkeit bei einem Tipp von 6 Zahlen 5 Richtige und eine Falsche zu haben, d.h. einen Fünfer zu haben? Antwort:

$$\binom{6}{5}\binom{49-6}{1}a.$$

Zur Begründung sei

$$\omega := \{s_1, \ldots, s_6\} \in \Omega$$

ein Elementarereignis, d.h. eine konkrete Ziehung von sechs Elementen aus der Menge $\{1, \ldots, 49\}$. Dann können $\binom{6}{5}(= 6)$ Fünfer

$$f_i := \omega \setminus \{s_i\}, \quad i = 1, \ldots, 6$$

aus ω gebildet werden. Da die sechste Zahl falsch sein soll, ergibt sich das Ereignis A, einen "Fünfer" zu machen, mengenmäßig entsprechend

$$A := \{(f, s) : f \in \{f_1, \ldots, f_6\}, s \in \{1, \ldots, 49\} \setminus \omega\}$$

d.h. man bildet A aus allen Sechsern, die sich durch Auffüllung eines beliebigen f_i mit einem Element aus

$$\{1, \ldots, 49\} \setminus \omega$$

ergeben. Offenbar gilt

$$\#A = \binom{6}{5}\binom{49-6}{1} = 6 \times 43$$

d.h. es gibt 258 Möglichkeiten, einen Sechsertipp abzugeben, der bei Ziehung von ω fünf Richtige und eine Falsche ergibt. Division durch die Anzahl aller Elementarereignisse liefert

$$P(A) = \frac{\#A}{\#\Omega} = \frac{258}{13\,983\,816} = 0,000.02.$$

In der gleichen Weise schließt man auf die Wahrscheinlichkeit

$$\frac{\binom{6}{k}\binom{49-6}{6-k}}{\binom{49}{6}}. \tag{6}$$

mit der man k Richtige und $6-k$ Falsche tippt. Mit der Bezeichnung

$$H(k,M,N) := \frac{\binom{M}{k}\binom{N-M}{M-k}}{\binom{N}{M}}.$$

bekommt man

k	$H(k,6,49)$
0	0.435 965
1	0.413 019
2	0.132 378
3	0.017 650 4
4	0.000 968 619 7
5	0.000 018 449 9
6	0.000 000 071 511 2

Die Funktion $H(k,M,N)$ führt auf die hypergeometrische Verteilung. Man beachte, daß

$$\sum_{k=0}^{6} H(k,6,49) = 1$$

gilt.

c) Zur Ergänzung betrachten wir ein überschaubares Beispiel. Die Wahrscheinlichkeit für einen Dreier bei 4 aus 6 läßt sich durch Aufzählung aller Elementarereignisse direkt ablesen: Es ist zu erwarten, daß der gesamte Ereignisraum Ω aus $\binom{6}{4}$, d.h. aus 15 Elementen besteht. Tatsächlich gilt

$$\Omega = \left\{ \begin{array}{ccccc} 1234 & 1235 & 1236 & 1245 & 1246 \\ 1256 & 1345 & 1346 & 1356 & 1456 \\ 2345 & 2346 & 2356 & 2456 & 3456 \end{array} \right\}$$

wobei alle vierstelligen Zahlen mit den Ziffern 1,2,3,4,5,6, bei denen jede Ziffer nur einmal vorkommt, aufsteigend sortiert sind. Wir betrachten das willkürlich ausgewählte Elementarereignis

$$\omega := 2346$$

und fragen nach der Anzahl N aller Elementarereignisse (Elemente von Ω), für die dieser Tipp einen Dreier liefert. Offenbar lassen sich aus ω die vier Dreier

$$234, \quad 236, \quad 246, \quad 346$$

bilden. Jeder dieser Dreier läßt sich mittels Ergänzung durch 1 oder 5 zu einem Element von Ω vervollständigen, sodaß bei dessen Ziehung der Tipp ω einen Dreier liefert. Demnach gilt

$$N = \binom{4}{3}\binom{6-4}{1} = 8$$

in Übereinstimmung mit der Plausibilitätsbetrachtung von oben. Die Wahrscheinlichkeit, daß ein bestimmtes Elementarereignis eintritt, liegt bei $a = 1/15$. Daher beträgt die Wahrscheinlichkeit bei 4 aus 6 einen Dreier zu machen bei $N/a = 8/15$. $\diamond$

2.4 Bedingte Wahrscheinlichkeiten, Unabhängigkeit von Ereignissen

Es ist naheliegend, verschiedene Ereignisse miteinander zu kombinieren und so zu neuen Ereignissen zu gelangen.

Die **bedingte** Wahrscheinlichkeit für das Auftreten des Ereignisses A unter der Voraussetzung, daß das Ereignis B schon eingetreten ist, wird durch die Zahl

$$\boxed{P(A|B) := \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0} \tag{7}$$

erklärt. Nach Definition gilt

$$P(A|B) = \frac{\#(A \cap B)/\#\Omega}{\#B/\#\Omega} = \frac{\#(A \cap B)}{\#B}$$

d.h. falls jedes Elementarereignis die gleiche Eintrittswahrscheinlichkeit von $\frac{1}{\#\Omega}$ besitzt, so ergibt sich $P(A|B)$ aus der Anzahl aller Elementarereignisse, die sowohl zu A als auch zu B gehören dividiert durch die Anzahl aller Elementarereignisse, die zu B gehören. Die Annahme gleicher Eintrittswahrscheinlichkeit für jedes Elementarereignis hatten wir weiter oben Laplacesche Annahme genannt.

Zwei Ereignisse

$$A, B \subseteq \Omega$$

heißen **unabhängig**, falls die Gleichung

$$\boxed{P(A \cap B) = P(A)P(B)} \tag{8}$$

erfüllt ist, d.h. die Wahrscheinlichkeit für das Eintreten des Ereignisses

$$A \cap B$$

stimmt mit dem Produkt aus der Wahrscheinlichkeit für das Eintreten der Ereignisses A und der Wahrscheinlichkeit für das Eintreten der Ereignisses B überein. Genauer müßte man eigentlich von der Unabhängigkeit von A und B bezüglich P sprechen, da für eine andere Definition von P die Ereignisse A und B nicht mehr unabhängig sein müssen. Bisher hatten wir nur endliche Ω und das sogenannte 'Zählmaß' (Anzahl der Elemente einer endlichen Mengen) betrachtet. Im Zusammenhang mit dem Bernoulli-Experiment, welches in Abschnitt 2.7 beschrieben wird, werden wir in natürlicher Weise auf eine andere Definition von P stoßen.

Offenbar gilt für zwei unabhängige Ereignisse

$$P(A|B) = P(A), \quad P(B|A) = P(B).$$

Eine Familie

$$A_i \subseteq \Omega, \quad i = 1, \ldots, n$$

von Ereignissen heißt **unabhängig**, falls für jede Auswahl

$$\{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}, \quad i_1 < i_2 < \cdots < i_m$$

von m Elementen und für beliebiges $m \in \{2, \ldots, n\}$ die Gleichung

$$\boxed{P(A_{i_1} \cap \cdots \cap A_{i_m}) = \prod_{k=1}^{m} P(A_{i_k})}$$

gilt ([20], S.1038). (Der Fall $m = 1$ ist trivialerweise erfüllt.)

Offenbar folgt aus der Unabhängigkeit einer Ereignisfamilie die paarweise Unabhängigkeit aller ihrer Mitglieder. Das folgende Beispiel zeigt, daß der umgekehrte Schluß falsch ist, d.h. aus der paarweisen Unabhängigkeit aller Familienmitglieder kann man **nicht** auf die Unabhängigkeit der Familie schließen.

Beispiel 5 Paarweise Unabhängigkeit impliziert nicht Unabhängigkeit

Es wird mit zwei Würfeln gewürfelt. Der Ereignisraum Ω kann durch das Kartesische Produkt der Menge $\{1, \ldots, 6\}$ mit sich selbst, d.h. entsprechend

$$\Omega = \{(i,j) : i,j \in \{1, \ldots, 6\}\} = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$$
$$= \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}$$

modelliert werden. Wir definieren drei Ereignisse.

- A_1 tritt ein, falls die Anzahl der Augen des ersten Würfels ungerade ist:

$$A_1 := \left\{ \begin{array}{c} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6) \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6) \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6) \end{array} \right\}.$$

- A_2 tritt ein, falls die Anzahl der Augen des zweiten Würfels ungerade ist:

$$A_2 := \left\{ \begin{array}{c} (1,1),(2,1),(3,1),(4,1),(5,1),(6,1) \\ (1,3),(2,3),(3,3),(4,3),(5,3),(6,3) \\ (1,5),(2,5),(3,5),(4,5),(5,5),(6,5) \end{array} \right\}.$$

- A_3 tritt ein, falls die Summe der Anzahl der Augen ungerade ist:

$$A_3 := \left\{ \begin{array}{c} (2,1),(1,2) \\ (4,1),(3,2),(2,3),(1,4) \\ (6,1),(5,2),(4,3),(3,4),(2,5),(1,6) \\ (6,3),(5,4),(4,5),(3,6) \\ (6,5),(5,6) \end{array} \right\}.$$

Offenbar bedeutet das Ereignis

$$A_1 \cap A_2$$

daß sowohl der erste als auch der zweite Würfel eine ungerade Augenanzahl zeigt. Da die Summe zweier ungerader Zahlen immer gerade ist, gilt

$$A_1 \cap A_2 \cap A_3 = \emptyset$$

woraus

$$P(A_1 \cap A_2 \cap A_3) = 0$$

folgt. Offenbar gelten die Gleichungen

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$$

so daß

$$P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3)$$

d.h. die Familie $\{A_1, A_2, A_3\}$ ist abhängig. Andererseits gilt

$$A_1 \cap A_2 = \left\{ \begin{array}{c} (1,1),(1,3),(1,5) \\ (3,1),(3,3),(3,5) \\ (5,1),(5,3),(5,5) \end{array} \right\}, \quad P(A_1 \cap A_2) = \frac{1}{4} = P(A_1)P(A_2)$$

$$A_1 \cap A_3 = \left\{ \begin{array}{c} (1,2),(1,4),(1,6) \\ (3,2),(3,4),(3,6) \\ (5,2),(5,4),(5,6) \end{array} \right\}, \quad P(A_1 \cap A_3) = \frac{1}{4} = P(A_1)P(A_3)$$

$$A_2 \cap A_3 = \left\{ \begin{array}{c} (2,1),(4,1),(6,1) \\ (2,3),(4,3),(6,3) \\ (2,5),(4,5),(6,5) \end{array} \right\}, \quad P(A_2 \cap A_3) = \frac{1}{4} = P(A_2)P(A_3).$$

Demnach ist die Familie $\{A_1, A_2, A_3\}$ paarweise unabhängig, aber insgesamt abhängig. $\diamond$

Beispiel 6 Bedingte Wahrscheinlichkeit, Würfeln mit zwei Würfeln

Es wird mit zwei Würfeln gewürfelt. Der Ereignisraum Ω kann durch das Kartesische Produkt der Menge $\{1, \ldots, 6\}$ mit sich selbst gebildet werden:

$$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\} := \{(i, j) : i, j \in \{1, \ldots, 6\}\}.$$

- Das Ereignis A ist durch alle Elementarereignisse definiert, für die die Summe der Augenanzahl gleich 4 ist, und
- das Ereignis B durch alle Elementarereignisse, bei denen beide Würfel die gleiche Augenanzahl zeigen.

Offenbar gilt

$$A = \{(1, 3), (2, 2), (3, 1)\}, \quad B = \{(1, 1), (2, 2), \ldots, (6, 6)\} \tag{9}$$

sodaß sich das simultane Eintreten von A und B mengenmäßig durch

$$A \cap B = \{(2, 2)\}$$

äußert. Offenbar gilt

$$P(A|B) = \frac{P(\{(2, 2)\})}{P(B)} = \frac{1/36}{6/36} = \frac{1}{6}.$$

Dieses Ergebnis läßt sich dahingehend interpretieren, daß man als neuen Ereignisraum die Menge B verwendet und nach der Anzahl der Elemente von B fragt, für die die Summe der Augenanzahlen gleich 4 ist. Offenbar besitzt B genau ein solches Element und besteht selber aus sechs Elementen, daher besitzt die Wahrscheinlichkeit, daß A eintritt unter der Bedingung, daß B eingetreten ist, den Wert 1/6.

Offenbar sind wegen

$$P(A) = 3/36, \quad P(B) = 6/36, \quad P(A \cap B) = 1/36$$

die Ereignisse (9) nicht unabhängig.

Wählt man dagegen

- als Ereignis A alle Elementarereignisse, bei denen die Augenanzahl, die der zweite Würfel zeigt, kleiner als 4 ist, und
- als Ereignis B alle Elementarereignisse, bei denen die Augenanzahl, die der erste Würfel zeigt, kleiner als 4 ist,

so sind A und B unabhängig: Offenbar besteht das Ereignis $A \cap B$ aus allen Elementarereignissen, bei denen die Augenanzahlen beider Würfel kleiner als 4 sind, d.h. $P(A \cap B) = 1/4$. Außerdem gilt $P(A) = P(B) = 1/2$. $\diamond$

2.5 Satz von der totalen Wahrscheinlichkeit und der Satz von Bayes

Falls die Ereignisse $B_1, \ldots, B_n$ eine **disjunkte Zerlegung** von Ω bilden, d.h.

$$\Omega = B_1 \cup B_2 \cup \cdots \cup B_n, \quad i \neq j \Rightarrow B_i \cap B_j = \emptyset$$

so gelten für jedes Ereignis $A \subseteq \Omega$ die Gleichungen

$$\begin{aligned}
\sum_{k=1}^{n} P(A|B_k)P(B_k) &= \sum_{k=1}^{n} \frac{P(A \cap B_k)}{P(B_k)} P(B_k) \\
&= \sum_{k=1}^{n} P(A \cap B_k) = P(\cup_{k=1}^{n}(A \cap B_k)) \\
&= P(A \cap (\cup_{k=1}^{n} B_k)) = P(A \cap \Omega) = P(A).
\end{aligned}$$

Die Gleichung

$$\boxed{P(A) = \sum_{k=1}^{n} P(A|B_k)P(B_k)}$$

heißt der **Satz von der totalen Wahrscheinlichkeit**.

Nach Definition gelten die Gleichungen

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B|A) = \frac{P(B \cap A)}{P(A)}.$$

Ihre Kombination liefert den **Satz von Bayes** (1702-1761):

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

Beispiel 7 Anwendung des Satzes von Bayes

Wir ziehen eine Kugel aus einer von zwei gleichberechtigten Urnen, wobei

- die erste Urne eine weiße und vier schwarze (**w s s s s**) und
- die zweite Urne eine weiße und zwei schwarze (**w s s**)

Kugeln enthält.

Es sei eine schwarze Kugel gezogen worden. Wie groß ist die Wahrscheinlichkeit, daß sie aus der ersten Urne stammt?

Die Intuition sagt uns, daß die gezogene Kugel wahrscheinlich aus der ersten Urne stammt, da das Verhältnis aus schwarzen und weißen Kugeln in der ersten Urne mehr zu Gunsten der schwarzen Kugeln ist als in der zweiten. Quantitativ Genaueres kann wiefolgt festgestellt werden.

- Das Ereignis A bedeutet, daß eine schwarze Kugel gezogen wurde,
- das Ereignis B_1, daß die gezogene Kugel unabhängig davon, ob sie weiß oder schwarz ist, aus Urne 1 stammt,
- das Ereignis B_2, daß die gezogene Kugel unabhängig davon, ob sie weiß oder schwarz ist, aus Urne 2 stammt.

Demzufolge interessieren wir uns laut Aufgabenstellung für die Zahl

$$P(B_1|A)$$

d.h. für die Wahrscheinlichkeit, in Urne 1 gegriffen zu haben, wenn man eine schwarze Kugel gezogen hat.

Mittels des Satzes von der totalen Wahrscheinlichkeit bekommen wir

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2).$$

Da beide Urnen gleichberechtigt sind, gilt

$$P(B_1) = P(B_2) = 1/2.$$

Entsprechend der Anteile an schwarzen und weißen Kugeln bekommen wir

$$P(A|B_1) = \frac{4}{5}, \quad P(A|B_2) = \frac{2}{3}.$$

Daher gilt

$$P(A) = \left(\frac{4}{5} \times \frac{1}{2}\right) + \left(\frac{2}{3} \times \frac{1}{2}\right) = \frac{11}{15}$$

sodaß wir für die uns interessierende Wahrscheinlichkeit

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)} = \frac{\frac{1}{2} \times \frac{4}{5}}{\frac{11}{15}} = \frac{4}{10} \times \frac{15}{11} = \frac{6}{11}$$

bekommen.

Zur Ergänzung stellen wir fest, daß die Wahrscheinlichkeit beim Ziehen einer schwarzen Kugel in Urne 2 gegriffen zu haben

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{P(A)} = \frac{\frac{1}{2} \times \frac{2}{3}}{\frac{11}{15}} = \frac{1}{3} \times \frac{15}{11} = \frac{5}{11}$$

beträgt. ◇

2.6 Wichtige Begriffe aus der Wahrscheinlichkeitsrechnung

2.6.1 Zufallsgröße

Um über die Begriffe Erwartung und Varianz reden zu können, benötigen wir zuerst den Begriff der **Zufallsgröße** (Zg). Dabei handelt es sich um eine eindeutige Abbildung des Ereignisraums Ω in den Bereich der reellen Zahlen:

$$\boxed{X : \Omega \mapsto \mathbb{R}}$$

d.h. jedem Elementarereignis ω wird durch X genau eine reelle Zahl zugeordnet. Eine Zg kann auch als Zufallsvariable bezeichnet werden

Eine Zg, deren Wertevorrat abzählbar ist, d.h. durchnummeriert werden kann, heißt **diskret**. Eine Zg, deren Wertevorrat ein gesamtes Intervall $[a, b] \subseteq \mathbb{R}$ umfaßt, heißt **stetig**.

Wir benutzen folgende Schreibweisen. Der Wertevorrat von X wird mit

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\}$$

bezeichnet.

Oft trifft man auf die Schreibweise

$$P(X = x_k)$$

womit $P(A_k)$ gemeint ist und die Menge A_k sich durch das **Urbild** von x_k vermöge X ergibt:

$$\boxed{A_k := \{X = x_k\} := \{\omega \in \Omega : X(\omega) = x_k\}}$$

Demnach liefert $P(X = x_k)$ die Wahrscheinlichkeit dafür, daß die Zg X den Wert x_k annimmt. Falls x_k nicht im Wertevorrat von X liegt, wird $A_k = \emptyset$ und demzufolge $P(A_k) = 0$.

In gleicher Weise setzt man für $x, y \in \mathbb{R}$ mit $x \leq y$ die Ereignisse

$$\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\}, \quad \{x < X \leq y\} := \{\omega \in \Omega : x < X(\omega) \leq y\}.$$

Beispiel 8 Zufallsgröße, Würfeln mit zwei Würfeln

Ein Beispiel dafür liefert das Würfeln mit zwei Würfeln. Der Ereignisraum Ω kann durch das Kartesische Produkt der Menge $\{1, \ldots, 6\}$ mit sich selbst, d.h. entsprechend

$$\Omega = \{(i, j) : i, j \in \{1, \ldots, 6\}\} = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$$

$$= \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}$$

modelliert werden.

Eine Möglichkeit, eine Zufallsgröße zu bilden, besteht in der Summierung der Augenanzahlen:

$$X((i,j)) := i + j.$$

Offenbar kann X nur die Werte $2, \ldots, 12$ annehmen, d.h.

$$X : \Omega \mapsto X(\Omega) := \{2, \ldots, 12\}$$

und insbesondere ist X diskret. Die Urbildmengen $A_k := \{X = k\}$ lauten

$$\begin{array}{ll}
A_2 & = \{(1,1)\} \\
A_3 & = \{(1,2),(2,1)\} \\
A_4 & = \{(1,3),(2,2),(3,1)\} \\
A_5 & = \{(1,4),(2,3),(3,2),(4,1)\} \\
A_6 & = \{(1,5),(2,4),(3,3),(4,2),(5,1)\} \\
A_7 & = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \\
A_8 & = \{(2,6),(3,5),(4,4),(5,3),(6,2)\} \\
A_9 & = \{(3,6),(4,5),(5,4),(6,3)\} \\
A_{10} & = \{(4,6),(5,5),(6,4)\} \\
A_{11} & = \{(5,6),(6,5)\} \\
A_{12} & = \{(6,6)\}
\end{array}$$

Man prüft leicht nach, daß die Ereignisfamilie $\{A_2, \ldots, A_{12}\}$ eine disjunkte Zerlegung von Ω etabliert. $\diamond$

2.6.2 Unabhängige Zufallsgrößen, Axiome von Kolmogorow

Es seien X und Y zwei Zufallsgrößen über dem gleichen Ereignisraum Ω. Beide Zg heißen **unabhängig**, falls für jede Wahl von $x, y \in \mathbb{R}$ die Ereignisse

$$\{X \leq x\}, \quad \{Y \leq y\}$$

unabhängig sind. In Hinblick auf Gleichung (8) sind X und Y demnach unabhängig genau dann, wenn für alle $x, y \in \mathbb{R}$ die Gleichung

$$\boxed{P(\{X \leq x\} \cap \{Y \leq y\}) = P(X \leq x) P(Y \leq y)}$$

gilt. Falls

$$X : \Omega \mapsto \{x_0, x_1 \ldots\}, \quad Y : \Omega \mapsto \{y_0, y_1 \ldots\}$$

beide diskret sind, so ist für ihre Unabhängigkeit hinreichend und notwendig die Gültigkeit von

$$\boxed{P(\{X = x_i\} \cap \{Y = y_j\}) = P(X = x_i) P(Y = y_j)}$$

für alle Paare $(i,j) \in \mathbb{N}_0^2$.

Offenbar hängt die Definition (5) der Unabhängigkeit zweier Ereignisse A und B vom zugrundeliegenden Maß P ab, sodaß dieses auch in die Definition der Unabhängigkeit zweier Zg eingeht. In (5) hatten wir nur endliches Ω betrachtet und P durch das Zählmaß definiert. Etwas abstrakter kann man P als eine Abbildung der von Ω erzeugten Potenzmenge in das Intervall $[0, 1]$ auffassen. Die Potenzmenge 2^Ω von Ω ergibt sich aus dem Mengensystem, das aus allen Teilmengen von Ω besteht [20], S.933. Wie nachfolgende Beispiele zeigen werden, gibt es viele andere Möglichkeiten der Definition von P, die fast alle in folgenden Rahmen passen.

- Aus der Menge Ω erzeugt man eine sogenannte σ-Algebra $\mathfrak{M}$. Darunter versteht man ein aus Teilmengen von Ω gebildetes Mengensystem $\mathfrak{M}$ mit den Eigenschaften

  1. $\Omega \in \mathfrak{M}$
  2. Falls $A \in \mathfrak{M}$, so folgt $\Omega \setminus A \in \mathfrak{M}$
  3. Falls $A_1, A_2, \ldots, A_n, \ldots \in \mathfrak{M}$, so gehört auch $\cup_{k=1}^{\infty} A_k$ zu $\mathfrak{M}$.

  [14], S.9. Die Elemente von $\mathfrak{M}$ heißen Ereignisse. Man prüft leicht nach, daß 2^Ω diese drei Eigenschaften besitzt und daß es echte Teilmengen von 2^Ω gibt, wie z.B. $\{\emptyset, \Omega\}$, die ebenfalls diesen Anforderungen genügen. Demnach kann man aus einer Menge Ω verschiedene σ-Algebren erzeugen. Das Mengensystem

  $$\mathscr{E} := \{\{\omega_1\}, \ldots, \{\omega_n\}\}$$

  bestehend aus den einelementigen Mengen $\{\omega_k\}$, erzeugt die σ-Algebra 2^Ω und wird daher ihr Erzeuger genannt. Wählt man $\mathscr{E} := \{A, \Omega\}$, wobei $A \subseteq \Omega$ eine beliebige Teilmenge darstellt, so wird das vierelementige Mengensystem

  $$\mathfrak{M} := \{\emptyset, A, \Omega \setminus A, \Omega\}$$

  als σ-Algebra erzeugt.

- Auf der σ-Algebra $\mathfrak{M}$ betrachtet man eine Mengenfunktion P mit den Eigenschaften

  1. $P : \mathfrak{M} \mapsto [0, 1]$
  2. $P(\Omega) = 1$
  3. $A_i \cap A_j = \emptyset \Rightarrow P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$.

  Eine solche Funktion heißt ein Wahrscheinlichkeitsmaß auf $\mathfrak{M}$. Diese drei Anforderungen an an eine Mengenfunktion werden die Axiome von Kolmogorow (A.N. Kolmogorow, 1903-1987) genannt. Sie wurden von ihm im Jahre 1933 in seinem Buch *Grundbegriffe der Wahrscheinlichkeitsrechnung* vorgeschlagen.

Benutzt man solch ein Wahrscheinlichkeitsmaß P, muß man für eine Zg X ihre **Meßbarkeit** bezüglich P verlangen. Da entsprechend [14], S.58, die σ-Algebra für das Lebesgue-Maß auf $\mathbb{R}$ durch die halboffenen Intervalle $(x, y]$ erzeugt wird, ist X meßbar bezüglich P, falls für alle $x, y \in \mathbb{R}$ die Enthaltenseinsrelation

$$\{x < X \leq y\} := \{\omega \in \Omega : x < X(\omega) \leq y\} \in \mathfrak{M} \tag{10}$$

gilt. Zusammenfassend sagen wir daher, daß $X : \Omega \mapsto \mathbb{R}$ auf dem Wahrscheinlichkeitsraum

$$(\Omega, \mathfrak{M}, P)$$

[1], S.4, als eine Zg aufgefaßt werden kann, falls für alle $x, y \in \mathbb{R}$ die Relation (10) gilt.

Beispiel 9 Wir betrachten auf der Potenzmenge

$$2^\Omega = \{\emptyset, \underbrace{\{\omega_1\}}_{=:\,A_1}, \underbrace{\{\omega_2\}}_{=:\,A_2}, \underbrace{\{\omega_3\}}_{=:\,A_3}, \underbrace{\{\omega_1,\omega_2\}}_{=:\,A_4}, \underbrace{\{\omega_1,\omega_3\}}_{=:\,A_5}, \underbrace{\{\omega_2,\omega_3\}}_{=:\,A_6}, \Omega\}$$

von $\Omega := \{\omega_1, \omega_2, \omega_3\}$ zwei verschiedene Mengenfunktionen. Setzen wir für $k = 1, 2, 3$

$$P_1(A_k) := \frac{1}{3}$$

so folgt mit dem dritten Axiom von Kolmogorow für die Mengen $\emptyset, A_4, A_5, A_6, \Omega$ zwangsläufig

$$P_1(\emptyset) =: 0, \quad P_1(A_4) = P_1(A_5) = P_2(A_6) := \frac{2}{3}, \quad P_1(\Omega) = 1.$$

Setzen wir für $k = 1, 2, 3$

$$P_2(A_k) := \frac{k}{6}$$

so folgt mit dem dritten Axiomen von Kolmogorow für die Mengen $\emptyset, A_4, A_5, A_6, \Omega$ zwangsläufig

$$\begin{aligned}
&P_2(\emptyset) := 0 \\
&P_2(A_4) := P_2(A_1 \cup A_2) = \frac{1+2}{6} = \frac{1}{2} \\
&P_2(A_5) := P_2(A_1 \cup A_3) = \frac{1+3}{6} = \frac{2}{3} \\
&P_2(A_6) := P_2(A_2 \cup A_3) = \frac{2+3}{6} = \frac{5}{6} \\
&P_2(\Omega) := P(A_1 \cup A_2 \cup A_3) = \frac{1+2+3}{6} = 1.
\end{aligned}$$

Es verbleibt zu überprüfen, ob das dritte Axiom für alle Folgen paarweiser durchschnittsfremder Ereignisse erfüllt ist. Zu diesem Zweck gehen wir etwas allgemeiner vor. Es sei

$$\Omega := \{\omega_1, \ldots, \omega_n\}, \quad P(\{\omega_i\}) := p_i, \quad p_i \in [0,1], \quad \sum_{i=1}^n p_i = 1$$

und

$$(A_i)_{i=1}^m, \quad A_i \subseteq \Omega$$

eine Folge sich paarweise gegenseitig ausschließender Ereignisse. Da Ω endlich ist, reicht die Betrachtung solcher Folgen endlicher Länge aus. Wegen $A_i \subseteq \Omega$, gestattet jedes A_i die Darstellung

$$A_i = \{\omega_j^i : j = 1, \ldots, \ell_i\}, \quad \omega_j^i \in \Omega.$$

Wegen $A_i \cap A_j = \emptyset$ haben wir

$$A_1 \cup A_2 \cup \cdots \cup A_m = \{\omega_1^1, \ldots, \omega_{\ell_1}^1, \ldots, \omega_1^m, \ldots, \omega_{\ell_m}^m\}$$

d.h.

$$P(A_1 \cup A_2 \cup \cdots \cup A_m) = \sum_{i=1}^{m} \underbrace{\sum_{j=1}^{\ell_i} P(\{\omega_j^i\})}_{= P(A_i)} = \sum_{i=1}^{m} P(A_i).$$

◇

2.6.3 Erwartungswert

Wir betrachten die mittels einer diskreten Zg $X : \Omega \mapsto \{x_0, x_1, \ldots\}$ erzeugten Urbilder

$$A_k := \{X = x_k\}.$$

Da X auf ganz Ω eindeutig definiert ist, wird jedem $\omega \in \Omega$ genau eine reelle Zahl durch X zugeordnet. Daher bildet die Ereignisfamilie

$$\{A_k : k \in \mathbb{N}_0\}$$

eine disjunkte Zerlegung von Ω:

$$\Omega = \bigcup_{k=0}^{\infty} A_k, \quad A_i \cap A_j = \emptyset, \quad i \neq j.$$

Unter Ausnutzung dieser Zerlegung wird für eine diskrete Zg der **Erwartungswert** EX durch

$$\boxed{EX := \sum_{k=0}^{\infty} x_k P(A_k) = \sum_{k=0}^{\infty} x_k P(X = x_k)}$$

erklärt, vorausgesetzt, daß die Summe konvergiert. Falls das nicht der Fall ist, sagt man, daß der Erwartungswert für X nicht existiert.

Beim Erfülltsein der Laplaceschen Annahme und $\#\Omega < \infty$ ist die Zahl $P(A_k)$ gleich dem Quotienten aus der Anzahl wie oft X den Wert x_k annimmt und der Anzahl der Elemente von Ω. Daher läßt sich die Erwartung EX als der gewichtete Mittelwert aller Werte, die X annehmen kann, interpretieren. Falls ein x_k von X sehr oft angenommen wird, geht dieses x_k entsprechend oft in EX ein.

Beispiele für diskrete Zg, die mehr als endlich viele Werte annehmen können, liefern alle poissonverteilten Zg. Solche Zg kommen u.a. in der Physik bei der Beschreibung von radioaktiven Zerfallsprozessen vor, vgl. Beispiel 16.

2.6.4 Wahrscheinlichkeitsfunktion

[13] S.318, [18] S.310

Setzt man

$$f: \underbrace{X(\Omega)}_{\substack{\text{Wertevorrat} \\ \text{von } X}} \mapsto [0,1], \quad f(x_k) := P(A_k) \tag{11}$$

so nennen wir f die **Wahrscheinlichkeitsfunktion** oder die **Zähldichte** der diskreten Zg X. Falls $\#\,\Omega = +\infty$, verlangt man von den Zahlen $P(A_k)$ nur, daß sie zu $[0,1]$ gehören und daß ihre Summe gleich Eins ist.

Unter Benutzung von f kann man auch

$$\boxed{EX = \sum_{k=0}^{\infty} x_k f(x_k)} \tag{12}$$

schreiben.

Für stetige Zg ist es üblich, f als **Dichtefunktion** zu bezeichnen. Die rechte Seite von (12) geht dann in ein Integral

$$\boxed{EX = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x}$$

über.

Mechanische Interpretation von EX als Massenschwerpunkt: Man betrachtet einen homogenen Balken $\mathbb{R}$, auf dem in den Positionen $x_1, \ldots, x_n$ die punktförmigen Massen $m_1, \ldots, m_n$ sitzen. Der Schwerpunkt dieser Massenverteilung befindet sich dann auf dem Balken in der Position

$$\frac{\sum_{k=1}^{n} x_k m_k}{\sum_{k=1}^{n} m_k}.$$

Man kann $P(A_k)$ als Masse m_k interpretieren, die sich auf $\mathbb{R}$ im Punkt x_k befindet. Da das Mengensystem $\{A_k : k \in \mathbb{N}\}$ eine disjunkte Zerlegung von Ω bildet, haben wir

$$1 = P(\Omega) = P(\cup_{k=0}^{\infty} A_k) = \sum_{k=0}^{\infty} P(A_k) = \sum_{k=0}^{\infty} f(x_k)$$

d.h. die Summe aller Massen ist gleich 1:

$$\boxed{\sum_{k=0}^{\infty} f(x_k) = 1}$$

Daher stellt EX den **Schwerpunkt** der Massenverteilung

$$\{\underbrace{(x_1, f(x_1))}_{\substack{\text{In } x_1 \text{ befindet} \\ \text{sich die Masse} \\ f(x_1)}}, \ldots, (x_\infty, f(x_\infty))\}$$

dar. Setzt man

$$x_1 < x_2 < x_3 \ldots$$

voraus, so gilt

$$x_1 \le EX \le x_\infty.$$

In vielen Anwendungen ist der Wertevorrat endlich, sodaß die Anzahl der Summanden endlich ist und Konvergenzfragen sich nicht stellen. Falls x_k nicht im Wertevorrat von X liegt, ergibt sich für A_k die leere Menge $\emptyset$, so daß $P(A_k) = 0$ und in der Summe EX weggelassen werden kann.

Fortsetzung Beispiel 8 Erwartungswert

[13] S.339

Für das Würfelbeispiel hat man

$$P(A_k), \quad k = 2, \ldots, 12$$

zu berechnen, falls man EX wissen will.

Jede Antidiagonale von

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

korrespondiert mit einem Ereignis A_k:

$$A_2 = \{(1,1)\},\ A_3 = \{(1,2),(2,1)\},\ \ A_4 = \{(1,3),(2,2),(3,1)\}, \ldots, A_{12} = \{(6,6)\}.$$

Demnach gilt zusammen mit der Definition $x_k := k$

$$\begin{gathered} f(x_2) = P(A_2) = 1/36 = P(A_{12}) = f(x_{12}) \\ f(x_3) = P(A_3) = 2/36 = P(A_{11}) = f(x_{11}) \\ f(x_4) = P(A_4) = 3/36 = P(A_{10}) = f(x_{10}) \\ f(x_5) = P(A_5) = 4/36 = P(A_9) = f(x_9) \\ f(x_6) = P(A_6) = 5/36 = P(A_8) = f(x_8) \end{gathered}$$

$$f(x_7) = P(A_7) = 6/36$$

sodaß sich

$$EX = \sum_{k=2}^{12} x_k f(x_k) = \frac{1}{36}(2+6+12+20+30+42+40+36+30+22+12) = 7$$

ergibt, d.h. wenn man lange mit zwei Würfeln würfelt und den Durchschnitt aller so erhaltenen Augenanzahlsummen bildet, bekommt man ungefähr 7. $\diamond$

Bedingter Erwartungswert

Es sei

$$A \subseteq \Omega$$

ein Ereignis und X eine diskrete Zg über Ω. Dann ist der bedingte Erwartungswert $E(X|A)$ von X unter der Bedingung A durch die Summe

$$\boxed{E(X|A) := \sum_{k=0}^{\infty} x_k P(A_k|A)}$$

definiert, wobei $P(A_k|A)$ durch (7) erklärt ist. Mit (7) folgt unmittelbar

$$\boxed{E(X|A) = \sum_{k=0}^{\infty} x_k \frac{P(A_k \cap A)}{P(A)}}$$

2.6.5 Eigenschaften des Erwartungswerts

1. Darstellung von EX durch gewichtete Summe bedingter Erwartungswerte

Es sei $\{A_1, \ldots, A_n\}$ eine Familie von sich gegenseitig ausschließenden Ereignissen, die den gesamten Ereignisraum Ω erzeugen:

$$\Omega = \bigcup_{k=1}^{n} A_k, \quad A_k \neq \emptyset, \quad i \neq j \Rightarrow A_i \cap A_j = \emptyset$$

d.h. wir betrachten eine Zerlegung der Menge Ω in disjunkte Teilmengen. Solch eine Ereignisfamilie heißt auch **Erzeugendensystem** für Ω. Dann gilt

$$\begin{aligned}
EX &= \sum_{k=0}^{\infty} x_k P(X = x_k) = \sum_{k=0}^{\infty} x_k P(\{X = x_k\} \cap \Omega) \\
&= \sum_{k=0}^{\infty} x_k P(\{X = x_k\} \cap (\cup_{i=1}^{n} A_i)) = \sum_{k=0}^{\infty} x_k P\left(\cup_{i=1}^{n}(\{X = x_k\} \cap A_i)\right) \\
&= \sum_{k=0}^{\infty} x_k \sum_{i=1}^{n} P\left(\{X = x_k\} \cap A_i\right) = \sum_{i=1}^{n} \sum_{k=0}^{\infty} x_k P\left(\{X = x_k\} \cap A_i\right) \\
&= \sum_{i=1}^{n} \sum_{k=0}^{\infty} x_k \frac{P\left(\{X = x_k\} \cap A_i\right)}{P(A_i)} P(A_i) = \sum_{i=1}^{n} E(X|A_i) P(A_i).
\end{aligned}$$

Demnach kann die Erwartung von X mittels eines Erzeugendensystem $\{A_i, \ldots, A_n\}$ von Ω als gewichtete Summe der bedingten Erwartungen $E(X|A_i)$ dargestellt werden:

$$\boxed{EX = \sum_{i=1}^{n} E(X|A_i)P(A_i)} \tag{13}$$

2. Darstellung von $E(aX + b)$

Bildet man aus der Zg X entsprechend

$$Y := aX + b, \quad a, b \in \mathbb{R}$$

eine neue Zg Y, so ergibt sich

$$EY = \sum_{k=0}^{\infty}(ax_k + b)P(A_k) = a\underbrace{\sum_{k=0}^{\infty} x_k P(A_k)}_{= EX} + b \underbrace{\sum_{k=0}^{\infty} P(A_k)}_{\substack{= P(\bigcup_{k=0}^{\infty} A_k) \\ = P(\Omega) = 1}} = aEX + b$$

d.h.

$$\boxed{E(aX + b) = aEX + b}$$

3. Darstellung von $E(X + Y)$ durch die Summe der Erwartungen

Betrachtet man zwei Zg

$$X : \Omega \mapsto \mathbb{R}, \quad Y : \Omega \mapsto \mathbb{R}$$

über dem selben Ereignisraum, so kann man ihre Summe bilden und erhält eine neue Zg Z über Ω. Erwartungsgemäß gilt

$$\boxed{E(X + Y) = EX + EY} \tag{14}$$

Zur Begründung bilden wir die Folge $(z_k)_{k=0}^{\infty}$, wobei jedes Glied Summe von x_i und y_j ist (Cantorsches Diagonalverfahren):

	x_0	x_1	x_2	x_3	x_4	$\ldots$
y_0	z_0	z_1	z_5	z_6	z_{14}	
y_1	z_2	z_4	z_7	z_{13}		
y_2	z_3	z_8	z_{12}			
y_3	z_9	z_{11}				
y_4	z_{10}					
$\vdots$						

$$\begin{aligned} z_0 &:= x_0 + y_0 \\ z_1 &:= x_1 + y_0 \\ z_2 &:= x_0 + y_1 \\ z_3 &:= x_0 + y_2 \\ z_4 &:= x_1 + y_1 \\ &\vdots \end{aligned}$$

und betrachten die Ereignisse

$$A_{ij} := \{X = x_i\} \cap \{Y = y_j\}, \quad i, j \in \mathbb{N}_0.$$

Offenbar bilden die Glieder der Folge $(z_k)_{k=0}^{\infty}$ den Wertevorrat von Z und auf A_{ij} nimmt Z nur den Wert $x_i + y_j$ an. Außerdem bilden für konstant gehaltenes j bzw. i die Familien

$$\{A_{ij} : i \in \mathbb{N}_0\}, \quad \{A_{ij} : j \in \mathbb{N}_0\}$$

disjunkte Zerlegungen der Ereignisse $\{Y = y_j\}$ bzw. $\{X = x_i\}$:

$$\bigcup_{i=0}^{\infty} A_{ij} = \{Y = y_j\}, \qquad \bigcup_{j=0}^{\infty} A_{ij} = \{X = x_i\},$$

$$i \neq k \Rightarrow A_{ij} \cap A_{kj} = \emptyset, \qquad j \neq k \Rightarrow A_{ij} \cap A_{ik} = \emptyset.$$

Daher gilt

$$\begin{aligned}
EZ &= z_0 P(A_{0,0}) + z_1 P(A_{1,0}) + z_2 P(A_{0,1}) + z_3 P(A_{0,2}) + z_4 P(A_{1,1}) + \ldots \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (x_i + y_j) P(A_{ij}) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} x_i P(A_{ij}) + y_j P(A_{ij}) \\
&= \sum_{i=0}^{\infty} x_i \sum_{j=0}^{\infty} P(A_{ij}) + \sum_{j=0}^{\infty} y_j \sum_{i=0}^{\infty} P(A_{ij}) \\
&= \sum_{i=0}^{\infty} x_i P\left(\bigcup_{j=0}^{\infty} A_{ij} \right) + \sum_{j=0}^{\infty} y_j P\left(\bigcup_{i=0}^{\infty} A_{ij} \right) \\
&= \sum_{i=0}^{\infty} x_i P(X = x_i) + \sum_{j=0}^{\infty} y_j P(Y = y_j) = EX + EY.
\end{aligned}$$

Die Unabhängigkeit von X und Y ist für die Gültigkeit dieser Gleichungen nicht notwendig, im Gegensatz bei der Erwartung des Produkts XY.

4. Darstellung von $E(XY)$ durch Produkt von Erwartungen

Falls X und Y **unabhängig** sind, so gilt

$$\boxed{E(XY) = EXEY}$$

(Vgl. [15], S.105). Diese Gleichung läßt sich mit Hilfe des bedingten Erwartungswertes plausibel machen. Um Konvergenzfragen nicht beachten zu müssen, gelte

$$Z := XY, \quad X : \Omega \mapsto \{x_0, \ldots, x_n\}, \quad Y : \Omega \mapsto \{y_0, \ldots, y_m\}$$

d.h. wir haben es nur mit endlichen Wertevorräten zu tun. Die Urbildmengen

$$A_i := \{X = x_i\}$$

werden als Erzeugendensystem für Ω benutzt, sodaß

$$EZ = \sum_{i=0}^{n} E(Z|A_i)P(A_i).$$

Wegen der Unabhängigkeit von X und Y gilt

$$P(\{Y = y_j\} \cap \{X = x_i\}) = P(Y = y_j)P(X = x_i)$$

resultierend in

$$\begin{aligned} P(Z = x_i y_j | A_i) &= P(XY = x_i y_j | X = x_i) = P(Y = y_j | X = x_i) \\ &= \frac{P(\{Y = y_j\} \cap \{X = x_i\})}{P(X = x_i)} = \frac{P(Y = y_j)P(X = x_i)}{P(X = x_i)} \\ &= P(Y = y_j). \end{aligned}$$

Kombination mit der Darstellung von EZ als gewichtete Summe bedingter Erwartungen liefert schließlich

$$\begin{aligned} EZ &= \sum_{i=0}^{n} E(Z|A_i)P(A_i) = \sum_{i=0}^{n} \underbrace{\left(\sum_{j=0}^{m} x_i y_j P(Z = x_i y_j | A_i) \right)}_{\substack{\text{Aufgrund der Bedingung } A_i \\ \text{nimmt } Z \text{ nur Werte} \\ \text{der Form } x_i y_j \text{ an.}}} P(A_i) \\ &= \sum_{i=0}^{n} \sum_{j=0}^{m} x_i y_j P(Y = y_j)P(X = x_i) = EXEY \end{aligned}$$

die Behauptung.

2.6.6 Varianz, Dispersion

Die **Varianz** oder die **Dispersion** von X mißt die durchschnittliche Streuung einer Zg X um ihren Erwartungswert:

$$\boxed{\mathrm{Var}(X) := E(X - \mu)^2, \quad \mu := EX}.$$

Da der Durchschnitt über nichtnegative Werte gebildet wird, gilt immer

$$0 \leq \mathrm{Var}(X).$$

Die nichtnegative Quadratwurzel

$$\boxed{\sigma := \sqrt{\mathrm{Var}(X)}}$$

aus der Varianz von X wird die **Standardabweichung** von X genannt.

Eigenschaften der Varianz

Für praktische Berechnungen verwendet man

$$\mathrm{Var}(X) = E(X^2 - 2\mu X + \mu^2) = EX^2 - 2\mu \underbrace{EX}_{=\mu} + \mu^2 = EX^2 - \mu^2$$

d.h.

$$\boxed{\mathrm{Var}(X) = EX^2 - \mu^2}$$

Es folgt unmittelbar die Ungleichung

$$(EX)^2 \leq EX^2.$$

Wie bei der Erwartung bringt man $\mathrm{Var}(X)$ und $\mathrm{Var}(Y)$ für $Y := aX + b$ in Zusammenhang:

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}(aX+b) = E(aX+b)^2 - (E(aX+b))^2 \\
&= E(a^2X^2 + 2abX + b^2) - (aE(X)+b)^2 \\
&= a^2E(X^2) + 2abE(X) + b^2 - a^2(E(X))^2 - 2abE(X) - b^2 \\
&= a^2E(X^2) - a^2(E(X))^2 = a^2\mathrm{Var}(X)
\end{aligned}$$

d.h.

$$\boxed{\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)} \tag{15}$$

Falls X und Y unabhängig sind, so gilt

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E(X+Y)^2 - (E(X+Y))^2 \\
&= E(X^2 + 2XY + Y^2) - (EX+EY)^2 \\
&= EX^2 + 2\underbrace{E(XY)}_{=EXEY} + EY^2 - (EX)^2 - 2EXEY - (EY)^2 \\
&= EX^2 - (EX)^2 + EY^2 - (EY)^2 = \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}$$

d.h. es gilt der **Satz über die Summe der Varianzen**

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)} \tag{16}$$

vorausgesetzt, daß X und Y unabhängig sind. In derselben Weise schließt man für n paarweise unabhängige Zg $X_1, \ldots, X_n$ auf

$$\boxed{\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k)}$$

Zur Begründung für $n=3$ betrachten wir eine Familie von Zg X, Y, Z, deren Mitglieder paarweise unabhängig sind, Dann gilt

$$\begin{aligned}
E(X+Y+Z)^2 &= E\left(\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\begin{bmatrix} X & Y & Z \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) \\
&= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} EX^2 & E(XY) & E(XZ) \\ E(YX) & EY^2 & E(YZ) \\ E(ZX) & E(ZX) & EZ^2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} EX^2 & EXEY & EXEZ \\ EYEX & EY^2 & EYEZ \\ EZEX & EZEX & EZ^2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
(E(X+Y+Z))^2 &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} EX \\ EY \\ EZ \end{bmatrix}\begin{bmatrix} EX & EY & EZ \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} (EX)^2 & EXEY & EXEZ \\ EYEX & (EY)^2 & EYEZ \\ EZEX & EZEX & (EZ)^2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\end{aligned}$$

Differenzbildung führt auf

$$\begin{aligned}
\operatorname{Var}(X+Y+Z) &= E(X+Y+Z)^2 - (E(X+Y+Z))^2 \\
&= EX^2 - (EX)^2 + EY^2 - (EY)^2 + EZ^2 - (EZ)^2 \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + \operatorname{Var}(Z).
\end{aligned}$$

2.6.7 Verteilungsfunktion

Die **Verteilungsfunktion** $F : \mathbb{R} \mapsto [0,1]$ der Zg X wird durch

$$\boxed{F(x) := P(X \leq x)}$$

erklärt, d.h. die zwischen 0 und 1 liegende Zahl $F(x)$ wird durch die Wahrscheinlichkeit, daß beim Experimentieren die Zg X die Schranke x nicht überschreitet, erklärt. Offenbar folgen aus der Ungleichung $x \leq y$ die Mengengleichungen

$$\{X \leq y\} = \{X \leq x\} \cup \{x < X \leq y\}, \quad \{X \leq x\} \cap \{x < X \leq y\} = \emptyset.$$

Demnach gilt entsprechend Zeile 6 aus der Tabelle oberhalb von Beispiel 3

$$P(X \leq y) = P(X \leq x) + P(x < X \leq y)$$

sodaß

$$\boxed{P(x < X \leq y) = F(y) - F(x)}$$

d.h. die Differenz $F(y) - F(x)$ ist gleich zur Wahrscheinlichkeit, daß die Zg X Werte annimmt, die größer als x und kleiner oder gleich y sind.

Falls X diskret ist, gilt

$$\boxed{F(x) = \sum_{x_k \leq x} f(x_k)} \tag{17}$$

Für stetige Zg geht die rechte Seite von (17) in ein Integral

$$\boxed{F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t}$$

der Dichtefunktion über.

Fortsetzung Beispiel 8 Summe der Erwartungswerte und Varianzen, Verteilungsfunktion

Für das Würfelbeispiel (X ist gleich der Augensumme) hatten wir $EX = 7$ festgestellt. Führt man X_1 und X_2 jeweils als die Augenanzahl von Würfel 1 und Würfel 2 ein, so gilt

$$X = X_1 + X_2.$$

Die zugehörigen Erwartungen sind gleich

$$EX_i = \frac{1}{6}\sum_{k=1}^{6} k = \frac{21}{6} = 3.5$$

sodaß

$$EX = EX_1 + EX_2 = 3.5. + 3.5 = 7$$

in Übereinstimmung mit dem Ergebnis von vorn. Für die Varianzen bekommen wir

$$\mathrm{Var}(X_i) = \frac{1}{6}\sum_{k=1}^{6} k^2 - 3.5^2 = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12}. \tag{18}$$

Da X_1 und X_2 unabhängig, gilt

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \frac{35}{6}.$$

Wir hatten

$$\begin{gathered} f(2) = 1/36 = f(12) \\ f(3) = 2/36 = f(11) \\ f(4) = 3/36 = f(10) \\ f(5) = 4/36 = f(9) \\ f(6) = 5/36 = f(8) \\ f(7) = 6/36 \end{gathered}$$

festgestellt. Daher sieht F wiefolgt aus:

$$F(x) = \begin{cases} 0 \\ f(2) \\ f(2)+f(3) \\ f(2)+f(3)+f(4) \\ f(2)+f(3)+f(4)+f(5) \\ f(2)+f(3)+f(4)+f(5)+f(6) \\ f(2)+\cdots+f(7) \\ f(2)+\cdots+f(8) \\ f(2)+\cdots+f(9) \\ f(2)+\cdots+f(10) \\ f(2)+\cdots+f(11) \\ f(2)+\cdots+f(12) \end{cases} = \frac{1}{36}\begin{cases} 0, & \text{falls } x < 2 \\ 1, & \text{falls } 2 \le x < 3 \\ 3, & \text{falls } 3 \le x < 4 \\ 6, & \text{falls } 4 \le x < 5 \\ 10, & \text{falls } 5 \le x < 6 \\ 15, & \text{falls } 6 \le x < 7 \\ 21, & \text{falls } 7 \le x < 8 \\ 26, & \text{falls } 8 \le x < 9 \\ 30, & \text{falls } 9 \le x < 10 \\ 33, & \text{falls } 10 \le x < 11 \\ 35, & \text{falls } 11 \le x < 12 \\ 36, & \text{falls } 12 \le x \end{cases}$$

d.h. F ist eine rechtsseitig stetige Treppenfunktion mit den Sprungstellen $2, 3, \dots, 12$ mit den Sprunghöhen

$$[f(2), \dots, f(12)] = \frac{1}{36}[1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1].$$

Im linken Teil von Abbildung 1 wird der Graph von F gezeigt. Beispielsweise gilt

$$F(7.8) - F(3.4) = (21 - 3)/36 = 18/36 = 1/2$$

d.h. in 50 % aller Fälle liegt die Augensumme zwischen 3.4 und 7.8.

Ein Blick auf

	1	2	3	4	5	6
1	2	3	(**4**)	(**5**)	(**6**)	(**7**)
2	3	(**4**)	(**5**)	(**6**)	(**7**)	8
3	(**4**)	(**5**)	(**6**)	(**7**)	8	9
4	(**5**)	(**6**)	(**7**)	8	9	10
5	(**6**)	(**7**)	8	9	10	11
6	(**7**)	8	9	10	11	12

bestätigt dieses Resultat. Tatsächlich nimmt in 18 (fettgedruckte Summe) von insgesamt 36 möglichen Fällen X einen Wert zwischen 3.4 und 7.8 an. ⋄

2.6.8 Tschebyschewsche **Ungleichung**

[20] S.1025, [6] S.98, [8] S.118

Die nach dem russischen Mathematiker Pafnuti Lwowitsch Tschebyschew (1821-1894) benannte Ungleichung kann man sich wiefolgt plausibel machen. Es sei Y eine Zg, die

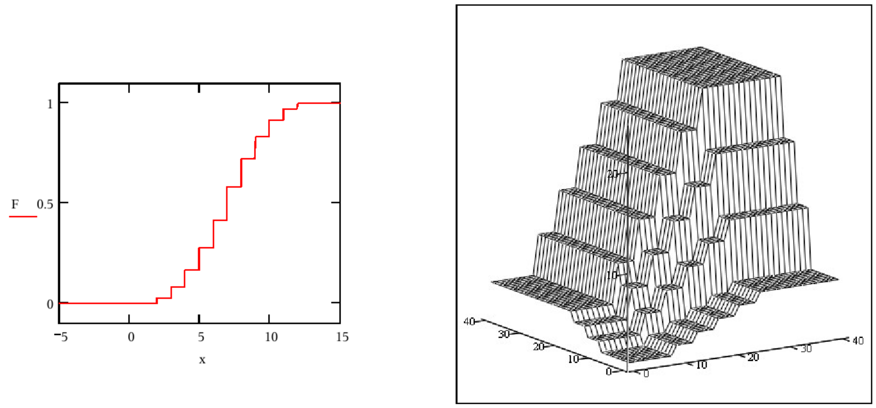


Abbildung 1: Im linken Teil wird die Verteilungsfunktion F der Augensumme beim Würfeln mit zwei Würfeln gezeigt. Eine Funktion, deren Graph diese Gestalt besitzt, heißt Treppenfunktion. Gesprungen wird an den Stellen $2, 3, \dots, 12$. In 12 wird endgültig auf 1 gesprungen. F ist rechtsstetig, d.h. F nimmt an einer Sprungstelle den Wert der oberen Stufe an. Im rechten Teil wird die Verteilungsfunktion des zweidimensionalen diskreten Zufallsvektors gezeigt, den man beim Würfeln mit einem Hexaeder und einem Tetraeder erhält. Um ganz korrekt zu sein, müßten die Verbindungsgeraden der einzelnen Plateaus im rechten Bild senkrecht sein. Einzelheiten dazu werden in Beispiel 10 erläutert.

nur nichtnegative Werte annehmen kann und die entsprechend der Dichtefunktion f verteilt ist. Dann gilt

$$EY = \int_0^{+\infty} y f(y)\mathrm{d}y.$$

Für jedes $K > 0$ hat man daher die Abschätzung

$$\begin{aligned} EY = & \int_0^{+\infty} y f(y)\mathrm{d}y = \int_0^K y f(y)\mathrm{d}y + \int_K^{+\infty} y f(y)\mathrm{d}y \\ \geq & \int_K^{+\infty} y f(y)\mathrm{d}y \geq K \int_K^{+\infty} f(y)\mathrm{d}y = KP(Y \geq K). \end{aligned}$$

Es gelte nun wie üblich $\mu := EX$. Dann folgt mit der obigen Abschätzung

$$\sigma^2 = \mathrm{Var}(X) = E(X-\mu)^2 \geq KP((X-\mu)^2 \geq K).$$

Die Wahl $K := k^2\sigma^2$, wobei $k > 0$ beliebig ist, führt auf

$$\sigma^2 \geq k^2\sigma^2 P((X-\mu)^2 \geq k^2\sigma^2).$$

Wegen der Mengengleichheit

$$\{\omega \in \Omega : (X(\omega)-\mu)^2 \geq K\} = \{\omega \in \Omega : |X(\omega)-\mu| \geq \sqrt{K}\}$$

führt die Division mit σ^2 schließlich auf eine Form der Tschebyschewschen Ungleichung, nämlich

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

Setzt man $k\sigma = c$, so kann man auch

$$\boxed{P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}}$$

schreiben, was sich wiefolgt interpretieren läßt: die Wahrscheinlichkeit, daß X Werte annimmt, die vom Erwartungswert μ um mehr als c entfernt sind, ist durch σ^2/c^2 beschränkt. Der Vorteil dieser Ungleichung besteht darin, daß man eine solche Aussage treffen kann, ohne die Verteilungsfunktion von X zu kennen.

Die Tschebyschewsche Ungleichung gilt auch für diskrete Zg. Der aufsteigend geordnete Wertevorrat von Y sei $(y_k)_{k=1}^{\infty} \subset \mathbb{R}_+$ und die zugehörigen Wahrscheinlichkeiten seien durch $f(y_k)$ gegeben. Dann gilt

$$EY = \sum_{k=1}^{\infty} y_k f(y_k) = \sum_{k=1}^{\ell-1} y_k f(y_k) + \sum_{k=\ell}^{\infty} y_k f(y_k) \geq \sum_{k=\ell}^{\infty} y_k f(y_k) \geq y_\ell \sum_{k=\ell}^{\infty} f(y_k)$$

d.h. $EY \geq y_\ell P(Y \geq y_\ell)$. Man kann nun y_ℓ durch einen beliebigen positiven Wert K ersetzen, ohne daß die Ungleichung ihre Gültigkeit verliert. Der Rest der Herleitung bleibt unverändert.

2.6.9 Diskrete Zufallsvektoren

Es seien

$$X_i : \Omega_i \mapsto W(X_i) := \{x_1^i, x_2^i, \ldots\}, \quad i = 1, \ldots, n$$

diskrete Zg. Ein diskreter Zufallsvektor ist dann durch die Abbildung

$$V := \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} : \Omega_1 \times \ldots \times \Omega_n \mapsto W(X_1) \times \ldots \times W(X_n)$$

gegeben. Faßt man $\Omega_1 \times \ldots \times \Omega_n =: \mathit{\Omega}$ zu einem neuen Ereignisraum zusammen, so stellt $V : \mathit{\Omega} \mapsto \mathbb{R}^n$ eine diskrete vektorwertige Zg dar. Die zugehörige Wahrscheinlichkeitsfunktion ist nun auf $W := W(X_1) \times \ldots \times W(X_n)$ durch

$$f_V(x_1, \ldots, x_n) := P(\{X_1 = x_1\} \cap \ldots \cap \{X_n = x_n\})$$

gegeben. Falls die Zg $X_1, \ldots, X_n$ unabhängig sind, gestattet f_V die Faktorisierung

$$f_V(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

wobei f_{X_i} die Wahrscheinlichkeitsfunktion von X_i darstellt.

Die Verteilungsfunktion

$$F_V : \mathbb{R}^n \mapsto [0,1]$$

ist durch

$$F_V(x_1, \ldots, x_n) := P(\{X_1 \le x_1\} \cap \ldots \cap \{X_n \le x_n\})$$

erklärt, also durch die Wahrscheinlichkeit, daß eine Realisierung des Zufallsvektors V in keiner Komponente größer wird als der Vektor $[x_1, \ldots, x_n]^{\mathrm{T}} \in \mathbb{R}^n$.

Beispiel 10 Diskreter Zufallsvektor

Wir würfeln gleichzeitig mit einem Hexaeder und einem Tetraeder (Vierflächner). Zg X_1 sei die obenaufliegende Augenanzahl des Hexaeders und X_2 die verdeckte Augenanzahl des Tetraeders. Offenbar gilt

$$X_1 : \{\omega_1, \ldots, \omega_6\} \mapsto \{1, 2, 3, 4, 5, 6\}, \quad X_2 : \{\omega_1, \ldots, \omega_4\} \mapsto \{1, 2, 3, 4\}.$$

Hier bezeichnet ω_i im Zusammenhang mit X_1 das Elementarereignis, bei dem die obenaufliegende Augenanzahl i ist, und im Zusammenhang mit X_2 das Elementarereignis, bei dem die verdeckte Augenanzahl i ist. Die zugehörigen Wahrscheinlichkeitsfunktionen lauten

$$f_{X_1}(x) := \frac{1}{6}\begin{cases} 1 & \text{falls } x \in \{1, \ldots, 6\} \\ 0 & \text{falls } x \notin \{1, \ldots, 6\} \end{cases}, \quad f_{X_1}(x) := \frac{1}{4}\begin{cases} 1 & \text{falls } x \in \{1, \ldots, 4\} \\ 0 & \text{falls } x \notin \{1, \ldots, 4\} \end{cases}$$

sodaß sich die Wahrscheinlichkeitsfunktion von $V := [X_1, X_2]^{\mathrm{T}}$, gemäß

$$f_V(x_1, x_2) := f_{X_1}(x_1) f_{X_2}(x_2)$$

ergibt. Definiert man $g(x)$ durch den ganzen Anteil von x, so gestattet die Verteilungsfunktion die Darstellung

$$F_V(x_1, x_2) = \sum_{i=1}^{g(x_1)} \sum_{j=1}^{g(x_2)} f_V(i, j).$$

Aufgrund der Faktorisierung von f_V bekommt man sofort

$$F_V(x_1, x_2) = \sum_{i=1}^{g(x_1)} f_{X_1}(i) \sum_{j=1}^{g(x_2)} f_{X_2}(j).$$

Demnach gilt für $6 \le x_1$ die Gleichung

$$F_V(x_1, x) = F_{X_1}(x)$$

und für $4 \le x_2$ die Gleichung

$$F_V(x, x_2) = F_{X_2}(x).$$

Im rechten Teil von Abbildung 1 wird F_V für $(x, y) \in [0, 7] \times [0, 7]$ gezeigt. $\diamond$

Die letzten beiden Gleichungen legen die Betrachtung der beiden Grenzwertfunktionen

$$f_1(x) := \lim_{y \to +\infty} f_V(x, y), \quad f_2(y) := \lim_{x \to +\infty} f_V(x, y)$$

nahe. Folgendes Beispiel zeigt, daß man so auch im abhängigen Fall die Wahrscheinlichkeitsfunktionen von X_1 und X_2 bekommt.

Beispiel 11 Diskreter Zufallsvektor, dessen Komponenten nicht unabhängig sind

Wir betrachten die Zg X und Y, die beide nur die Werte $1, 2, 3$ mit den Wahrscheinlichkeiten

$$\boldsymbol{p} := \frac{1}{27}\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

annehmen können. Dabei bedeutet

$$P(\{X = i\} \cap \{Y = j\}) = p_{ij}.$$

Da die Matrix $\boldsymbol{p}$ keine Rang-1-Matrix ist, sind X und Y abhängig.

Offenbar gilt

$$P(X = 1) = P\left(\{X = 1\} \cap (\{Y = 1\} \cup \{Y = 2\} \cup \{Y = 3\})\right)$$
$$= P\left((\{X = 1\} \cap \{Y = 1\}) \cup (\{X = 1\} \cap \{Y = 2\}) \cup (\{X = 1\} \cap \{Y = 3\})\right)$$
$$= p_{11} + p_{12} + p_{13} = \frac{1}{27}(1 + 2 + 3) = \frac{6}{27}.$$

Demnach hat man einfach die Matrix $\boldsymbol{p}$ mit dem Vektor $\mathbf{1} := [1, 1, 1]^{\mathrm{T}}$ von rechts zu multiplizieren, wenn man den Werteverlauf f_X wissen möchte, und von links mit $\mathbf{1}^{\mathrm{T}}$, wenn man den Werteverlauf f_Y wissen möchte. Offenbar erhält man f_X durch die Zeilensummen und f_Y durch die Spaltensummen von p. $\diamond$

2.7 Bernoulli-Experiment (Serien-Schema)

Das folgende von Jakob Bernoulli (1654-1705) stammende Modell läßt sich in sehr vielen Situationen der Praxis anwenden und gehört zu den wichtigsten Modellen der Wahrscheinlichkeitsrechnung. Insbesondere erlaubt dieses Modell eine Untersuchung des Zusammenhanges zwischen Wahrscheinlichkeit und relativer Häufigkeit. Bisher hatten wir jedem Elementarereignis ω die gleiche Eintrittswahrscheinlichkeit zugeordnet. Das ändert sich jetzt.

Es wird ein Grundversuch durchgeführt, der nur die beiden möglichen Ausgänge A und $\overline{A}$ besitzt. Hierbei stellt A irgend ein Ereignis dar, wie z.B. das Treffen einer Zielscheibe, die man trifft (A) oder nicht ($\overline{A}$). Die Wahrscheinlichkeit für das Eintreten von A sei

$$\boxed{p \in [0, 1]}$$

d.h.

$$P(A) = p, \quad P(\overline{A}) = 1 - p.$$

Wir nennen p die Grundwahrscheinlichkeit des Versuchs. Der Grundversuch wird n mal durchgeführt. Alle diese Versuche werden als voneinander unabhängig angesehen, d.h. ihre Ausgänge beeinflussen sich gegenseitig nicht. Das Elementarereignis ω besteht nun aus einer Folge

$$\omega := \underbrace{A\,A\,\overline{A}\,A\,\overline{A}\ldots A}_{n}$$

der Ereignisse A und $\overline{A}$ von der Länge n. Eine Realisierung wäre das n-malige Schießen auf eine Zielscheibe. Das durch ω dargestellte Elementarereignis wäre dann das Treffen mit den ersten beiden Schüssen, das Vorbeischießen beim drittenmal ... und das Treffen beim n-ten Schuß. Dem Elementarereignis ω wird unter Zuhilfenahme der Grundwahrscheinlichkeit p die Wahrscheinlichkeit

$$\boxed{P(\{\omega\}) := p^k(1-p)^{n-k}} \tag{19}$$

zugeordnet, wobei sich k aus der Anzahl des Eintretens des Ereignisses A ergibt. Offenbar variiert für $p \neq 1/2$ die Zahl $P(\omega)$ für variierendes ω.

Wir führen eine Zg $X : \Omega \mapsto \{0,\ldots,n\}$ gemäß

$$\boxed{X(\omega) := \text{Anzahl des Auftretens von } A \text{ in } \omega}$$

und die Zg $X_i : \Omega \mapsto \{0,1\}$ als sogenannte **Indikatorfunktion** gemäß

$$\boxed{X_i(\omega) := \begin{cases} 1, & \text{falls in } \omega \text{ an der Stelle } i \text{ ein A steht} \\ 0, & \text{falls in } \omega \text{ an der Stelle } i \text{ ein } \overline{A} \text{ steht} \end{cases}}$$

ein. Wegen

$$P(X_i = 1) = p, \quad P(X_i = 0) = q, \quad q := 1-p$$

gilt entsprechend (11) für die zu X_i gehörige Wahrscheinlichkeitsfunktion

$$\boxed{f : \underbrace{X_i(\Omega)}_{=\{0,1\}} \mapsto [0,1], \quad f(x) = \begin{cases} p & \text{falls } x = 1 \\ q & \text{falls } x = 0 \end{cases}}$$

d.h. sie ist für alle X_i gleich, wobei n verschiedene Indikatorfunktionen existieren.

Es zeigt sich, daß die Zg $X_1,\ldots,X_n$ unabhängig sind. [1]

Wegen

$$\forall \omega \in \Omega : X(\omega) = \sum_{i=1}^{n} X_i(\omega)$$

[1]Um die Unabhängigkeit der X_i plausibel zu machen, zeigen wir wenigstens ihre paarweise Unabhängigkeit. Dazu definieren wir für $i \in \{1,\ldots,n\}$ das Ereignis A_i durch das Urbild von 1 vermöge X_i, d.h. das Ereignis A_i tritt ein, falls im i-ten Versuch das Ereignis A eintritt. Es zeigt sich, daß für alle Indexpaare $(i,j) \in \{1,\ldots,n\}^2$ mit $i \neq j$ die Ereignisse A_i und A_j unabhängig sind. Dafür hat man

$$P(A_i \cap A_j) = P(A_i)P(A_j) \tag{20}$$

können wir

$$X = \sum_{i=1}^{n} X_i \tag{21}$$

schreiben und haben so eine Darstellung von X als Summe von unabhängigen und identisch verteilten Zg erhalten. Diese Darstellung ist nützlich für die Anwendung des **Zentralen Grenzwertsatzes** auf die Verteilung von X für große n und für die Berechnung von Erwartung und Varianz.

Erwartungswert: Nach Voraussetzung gilt

$$P(X_i = 1) = f(1) = p.$$

Da X_i als einzigen von Null verschiedenen Wert nur die Eins annehmen kann, bekommt man

$$EX_i = 1 \times P(X_i = 1) = p, \quad EX_i^2 = 1^2 \times P(X_i = 1) = p$$

und demzufolge

$$EX = E\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} EX_i = np$$

d.h.

$$\boxed{EX = np}$$

Varianz: Für $\mathrm{Var}(X)$ beobachtet man

$$\mathrm{Var}(X_i) = EX_i^2 - p^2 = p - p^2 = pq$$

sodaß aufgrund der paarweisen Unabhängigkeit der X_i auf

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = npq$$

---

nachzuprüfen. Für $n = 2$ gibt es nur die beiden Ereignisse A_1 und A_2, deren mengenmäßige Darstellung

$$A_1 = \{10, 11\}, \quad A_2 := \{01, 11\}$$

lautet, wenn man das Eintreten von A durch eine 1 und das Nichteintreten von A durch eine 0 codiert. Diese Art der Codierung besitzt den Vorteil, daß jedes Elementarereignis ω mit einer Binärzahl der Länge n identifiziert werden kann, sodaß deren Zahlenwert zur Anordnung aller Elemente von Ω benutzt werden kann, indem man der Größe nach sortiert.

Offenbar gilt

$$A_1 \cap A_2 = \{11\}$$

sodaß

$$P(A_1 \cap A_2) = P(\{11\}) = p^2.$$

Andererseits haben wir

$$\begin{aligned} P(A_1) &= P(\{10, 11\}) = P(\{10\}) + P(\{11\}) = p(1-p) + p^2 = p \\ P(A_2) &= P(\{01, 11\}) = P(\{01\}) + P(\{11\}) = (1-p)p + p^2 = p \end{aligned}$$

d.h. es gilt tatsächlich (20). Wegen

$$A_i = \{X_i = 1\}$$

und der paarweisen Unabhängigkeit der Ereignisse $A_1, \ldots, A_n$ sind die Zg X_i paarweise unabhängig.

geschlossen werden kann, d.h.

$$\boxed{\mathrm{Var}(X) = npq}$$

Beispiel 12 Bernoulli Schema

Es werden vier unabhängige Schüsse mit einer Trefferwahrscheinlichkeit von

$$p := 0.8$$

abgefeuert.

a) Mit welcher Wahrscheinlichkeit wird das Ziel wenigstens einmal getroffen?

Wir modellieren das Treffen der Zielscheibe, d.h. das Eintreten des Ereignisses A beim i-ten Schuß, mit der Vergabe einer 1 und das Vorbeischießen mit einer 0. Dann lautet der Ereignisraum

$$\begin{aligned} \Omega := & \ \{\underbrace{0000}_{0}, \underbrace{0001}_{1}, \underbrace{0010}_{2}, \underbrace{0011}_{3}, \underbrace{0100}_{4}, \underbrace{0101}_{5}, \underbrace{0110}_{6}, \underbrace{0111}_{7}, \underbrace{1000}_{8}, \ldots, \underbrace{1111}_{15}\} \\ =: & \ \{\omega_0, \ldots, \omega_{15}\} \end{aligned} \tag{22}$$

bestehend aus den Binärzahlen 0000 bis 1111, die mit den Dezimalzahlen 0 bis 15 identifiziert werden können. Dabei bedeutet eine 1 in Position i, daß mit dem i-ten Schuß getroffen wird und eine Null an dieser Stelle, daß der i-te Schuß daneben geht.

Wir definieren als Zg X die Anzahl von Einsen in ω_i. Offenbar wird nur bei dem Eintreten des Elementarereignisses $\omega_0 := 0000$ viermal vorbeigeschossen also das Ziel nicht getroffen. Die Voraussetzung über die Trefferwahrscheinlichkeit bedeutet

$$P(X = 0) = \binom{4}{0} p^0 (1-p)^4 = 0.2^4.$$

Daher bekommt man für die Wahrscheinlichkeit, das Ziel bei vier abgegebenen Schüssen wenigstens einmal zu treffen, in Hinblick auf Zeile 8 in der Ereignisalgebratabelle den Wert

$$P(\Omega \setminus \{\omega_0\}) = P(\Omega) - P(X = 0) = 1 - 0.2^4 = 0.9984.$$

Daher wird das Ziel mit 99.84 %-ger Wahrscheinlichkeit wenigstens einmal getroffen.

b) Es soll nun illustriert werden, daß X nicht immer die “Anzahl der Einsen” in ω_i sein muß. Als Zufallsgröße

$$X : \Omega \mapsto \mathbb{N}$$

wird die Anzahl der abgefeuerten Schüsse bis einschließlich zum ersten Treffer definiert, genauer

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X(\omega_i)$	4	4	3	3	2	2	2	2	1	1	1	1	1	1	1	1

X kann als die Anzahl der abgegebenen Schüsse interpretiert werden, wenn man nach dem ersten Treffer aufhört zu schießen. Man berechne den Erwartungswert von X.

Dazu hat man

$$EX := \sum_{i=0}^{15} X(\omega_i)P(\{\omega_i\})$$

auszuwerten. Aufgrund des Wertevorrates $\{1, 2, 3, 4\}$ von X bekommt man

$$EX = 4P(\{\omega_0, \omega_1\}) + 3P(\{\omega_2, \omega_3\}) + 2P(\{\omega_4, \dots, \omega_7\}) + P(\{\omega_8, \dots, \omega_{15}\}).$$

- Setzt man $q := 1 - p$, so gilt

  $$P(\{\omega_0, \omega_1\}) = P(\{\omega_0\}) + P(\{\omega_1\}) = q^4 + pq^3 = q^3(p + q) = q^3$$

  was mit der Wahrscheinlichkeit dafür übereinstimmt, daß man dreimal vorbeischießt.

- Die Wahrscheinlichkeit des Eintretens des Ereignisses $\{\omega_2, \omega_3\}$ beträgt

  $$q^2 p$$

  da man erst zweimal vorbeischießt und dann beim drittenmal trifft.

- Die Wahrscheinlichkeit des Eintretens des Ereignisses $\{\omega_4, \dots, \omega_7\}$ beträgt

  $$qp$$

  da man erst einmal vorbeischießt und dann beim zweitenmal trifft.

- Die Wahrscheinlichkeit des Eintretens des Ereignisses $\{\omega_8, \dots, \omega_{15}\}$ beträgt

  $$p$$

  da man beim erstenmal trifft.

Demnach gilt

$$EX = 4q^3 + 3q^2p + 2qp + p = 1.248, \quad p = 0.8,\ q = 0.2. \tag{23}$$

Dieser Wert läßt sich dahingehend interpretieren, daß man bei einer Trefferwahrscheinlichkeit von 0.8 nach 1.248 Schüssen getroffen hat.

Ersetzt man q durch $1 - p$, erhält man für EX das in $[0, 1]$ streng monoton fallende Polynom

$$EX(p) = -p^3 + 4p^2 - 6p + 4.$$

dessen Werteverlauf in Abbildung 2 gezeigt wird. Für $p = 0$ bekommt man $EX = 4$, was sinnvoll ist, da man mit Sicherheit nicht trifft und demzufolge auch viermal schießt. Für $p = 1$ bekommt man $EX = 1$, was sich mit der Tatsache deckt, daß nur einmal geschossen wird, wenn man mit Sicherheit beim Erstenmal trifft.

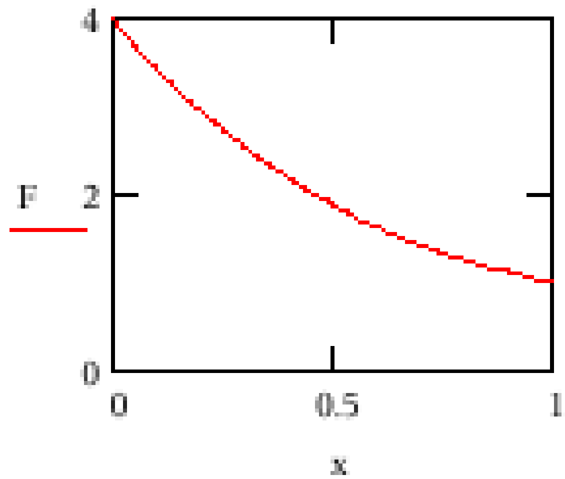


Abbildung 2: Erwartung $EX(p)$ von X für $p \in [0,1]$. Offensichtlich ist $EX(p)$ streng monoton fallend.

Wegen $EX(0) = 4$ und $EX(1) = 1$ existiert für jede Vorgabe von $\alpha \in [1,4]$ eine Wahrscheinlichkeit p, mit $EX = \alpha$. Z.B. liefert $\alpha = 2$ die Wahrscheinlichkeit $p = 0.4563$.

c) Man bestimme die Varianz $\mathrm{Var}(X)$ von X.

Nach Definition hat man

$$\mathrm{Var}(X) := E(X-\mu)^2, \quad \mu := EX$$

auszuwerten. Dazu benutzt man praktischerweise die Identität

$$\mathrm{Var}(X) = EX^2 - \mu^2.$$

Offenbar hat man in (23) nur die Koeffizienten zu quadrieren, daher gilt

$$EX^2 = 16q^3 + 9pq^2 + 4pq + p$$

sodaß mit $q = 1 - p$ die Gleichungen

$$\begin{aligned} \mathrm{Var}(X) &= 16q^3 + 9pq^2 + 4pq + p - (-p^3 + 4p^2 - 6p + 4)^2 \\ &= -p^6 + 8p^5 - 28p^4 + 49p^3 - 42p^2 + 14p \\ &=: v(p) \end{aligned}$$

folgen. Insbesondere gilt

$$v(0.8) = 0.298.$$

Der Verlauf von $\mathrm{Var}(X)$ für $p \in [0,1]$ wird in Abbildung 3 gezeigt.

- Für $p = 0.2714$ wird die Varianz maximal, d.h. für eine Trefferwahrscheinlichkeit von 0.2714 streut X um die zugehörige Erwartung $EX(0.2714)$ stärker als für jede andere Trefferwahrscheinlichkeit.

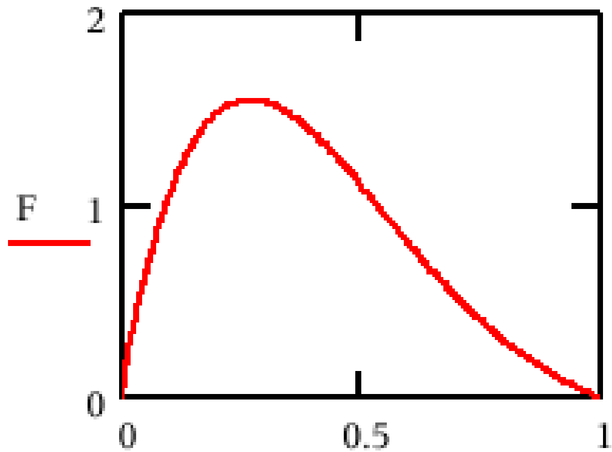


Abbildung 3: Varianz von X für $p \in [0, 1]$

- Erwartungsgemäß gilt $v(0) = 0$, da für eine Trefferwahrscheinlichkeit von 0 fast jedes Experiment (es wird tatsächlich viermal geschossen) das Ereignis 0000 liefert, sodaß die Zufallsgröße X fast immer den Wert 4 annimmt, d.h. X um 4 herum nicht streut.

- Erwartungsgemäß gilt $v(1) = 0$. Für eine Trefferwahrscheinlichkeit von 1 liefert fast jedes Experiment das Ereignis 1111, sodaß die Zufallsgröße X fast immer den Wert 1 annimmt, d.h. X um 1 herum nicht streut.

$\diamond$

2.8 Wichtige diskrete Verteilungen

2.8.1 Binomialverteilung $\mathscr{B}(n, p)$

[18] S.323

Definiert man Ω durch die Menge aller Binärzahlen ω der Länge n und $X : \Omega \mapsto \{0, \ldots, n\}$ durch die Anzahl der Einsen in ω, so bekommt man

$$\boxed{P(X = k) = f_{n,p}(k) := \binom{n}{k} p^k q^{n-k}, \quad q := 1 - p} \tag{24}$$

wenn man für das Auftreten einer 1 die Grundwahrscheinlichkeit

$$p \in [0, 1]$$

annimmt. Eine diskrete Zufallsgröße

$$X : \Omega \mapsto \{0, 1, \ldots, n\}$$

die (24) erfüllt, heißt mit den Parametern

$$n \in \mathbb{N}, \quad p \in [0,1]$$

binomialverteilt, was durch die Schreibweise

$$X \sim \mathscr{B}(n,p)$$

zum Ausdruck gebracht wird.

Beachte, daß

$$f_{n,p} : \underbrace{X(\Omega)}_{=\{0,1,\ldots,n\}} \mapsto [0,1]$$

die Wahrscheinlichkeitsfunktion von X darstellt und, daß die Summe aller Gewichte tatsächlich gleich 1 ist:

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

Es ist naheliegend, nach der Erwartung und der Varianz von $X \sim \mathscr{B}(n,p)$ zu fragen.

Erwartungswert. Es gilt

$$\begin{aligned} EX &= \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} \frac{n(n-1)!}{(n-k)!(k-1)!} p^k q^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} = np(p+q)^{n-1} = np \end{aligned}$$

d.h.

$$\boxed{EX = np}$$

Mit Hilfe der Darstellung (21) bekommt man das gleiche Ergebnis.

Varianz. Wegen

$$\mathrm{Var}(X) = EX^2 - \mu^2 = EX^2 - n^2p^2$$

genügt es

$$\begin{aligned} EX^2 &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = np \underbrace{\sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}}_{(n-1)p+1} \\ &= np((n-1)p+1) \end{aligned}$$

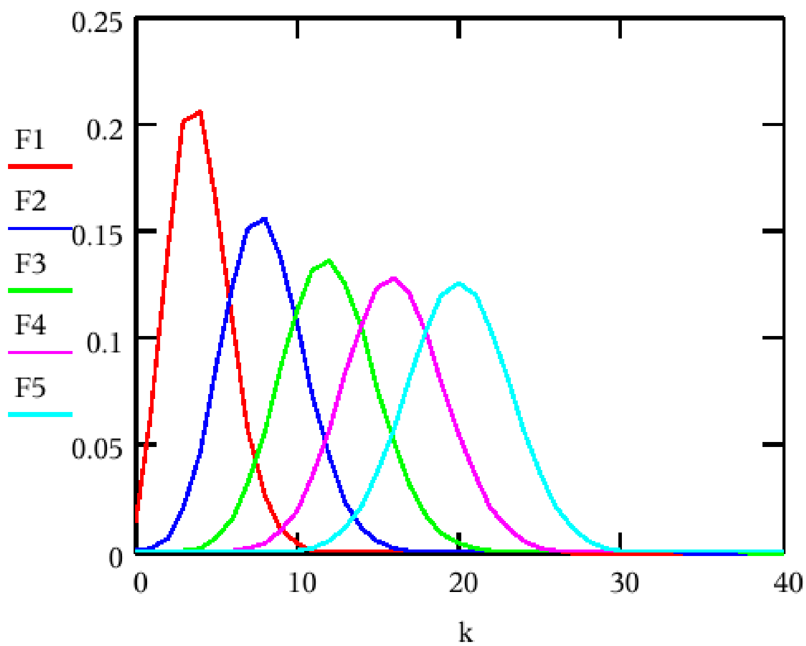


Abbildung 4: Es wird die Wahrscheinlichkeitsfunktion $f_{n,p}$ von $X \sim \mathscr{B}(n,p)$ für die Parameter $n = 40$ und $p = \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}, \frac{5}{10}$ gezeigt. Es ist zu beachten, daß $F1, \ldots, F5$ Mengen sind bestehend aus jeweils 40 Punkten, die als miteinander verbunden dargestellt werden. Die resultierenden Polygonzüge erinnern sehr an Gaußsche Glockenkurven, ein Umstand der durch den Satz von de-Moivre-Laplace erklärt wird.

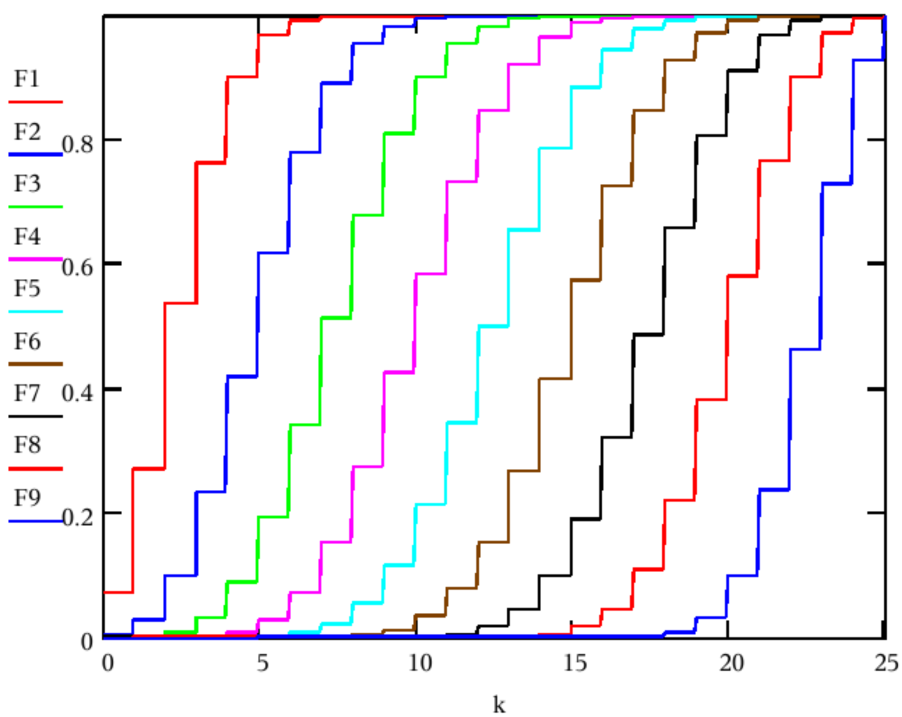


Abbildung 5: Es wird die Verteilungsfunktion $F(x) := \sum_{0 \le k \le x} f_{n,p}(k)$ von $X \sim \mathscr{B}(n,p)$ für die Parameter $n = 25$ und $p = \frac{1}{10}, \frac{2}{10}, \dots \frac{9}{10}$ gezeigt. Offensichtlich sind $F1, \dots, F9$ rechtsstetige Treppenfunktionen mit Sprüngen an den Stellen $k = 0, \dots, 25$. Die Sprunghöhen ergeben sich durch die Werte der Wahrscheinlichkeitsfunktion $f_{n,p}$. An der Stelle 25 springen alle Verteilungsfunktionen endgültig auf 1. Oft sind die Sprunghöhen so klein, daß sie im Bild nicht mehr zu erkennen sind.

festzustellen. Daher gilt

$$\mathrm{Var}(X) = np(np + 1 - p) - (np)^2 = npq$$

d.h.

$$\boxed{\mathrm{Var}(X) = npq}$$

Mit Hilfe von (16) und der Darstellung (21) auf Seite 42 von X als Summe von Indikatorfunktionen bekommt man dasselbe Ergebnis.

In **mathcad** liefert die Auswertung von

$$\mathrm{dbinom}(k, n, p), \quad \mathrm{pbinom}(k, n, p) \tag{25}$$

jeweils die Werte

$$f_{n,p}(k), \quad \sum_{i=0}^{k} f_{n,p}(i).$$

Die Abbildungen 4 und 5 wurden jeweils mit den Befehlen (25) erzeugt.

Beispiel 13 Binomialverteilte Zg, Knabengeburt

Die Wahrscheinlichkeit einer Knabengeburt sei

$$p = 0.515.$$

Eine Familie hat 4 Kinder. Wie groß ist die Wahrscheinlichkeit, daß unter diesen genau 2 Knaben sind ?

Zur Beantwortung definieren wir den Ereignisraum

$$\Omega := \left\{ \begin{array}{l} \underbrace{0000}_{\omega_0}, \underbrace{0001}_{\omega_1}, \underbrace{0010}_{\omega_2}, \underbrace{0011}_{\omega_3}, \underbrace{0100}_{\omega_4}, \underbrace{0101}_{\omega_5}, \underbrace{0110}_{\omega_6}, \underbrace{0111}_{\omega_7} \\ \underbrace{1000}_{\omega_8}, \underbrace{1001}_{\omega_9}, \underbrace{1010}_{\omega_{10}}, \underbrace{1011}_{\omega_{11}}, \underbrace{1100}_{\omega_{12}}, \underbrace{1101}_{\omega_{13}}, \underbrace{1110}_{\omega_{14}}, \underbrace{1111}_{\omega_{15}} \end{array} \right\}$$

wie in (22) und interpretieren jedes seiner Elemente wie folgt: die Anzahl der Einsen in ω_i ergibt die Anzahl der Knaben beim Eintreten des Elementarereignisses ω_i. Das Ereignis A_2 ist definiert durch den Fall, daß genau 2 Knaben unter den vier Kindern sind. Demzufolge besteht A_2 mengenmäßig aus allen Elementarereignissen ω_i, die genau zweimal die Eins enthalten:

$$A_2 = \{\omega_i \in \Omega : \text{in } \omega_i \text{ erscheint genau zweimal die Eins}\}.$$

Offenbar gilt

$$A_2 = \{\omega_3,\ \omega_5,\ \omega_6,\ \omega_9,\ \omega_{10},\ \omega_{12}\}.$$

Es wäre nun im allgemeinen falsch, $P(A_2)$ durch den Quotienten $\#A_2/\#\Omega$ zu berechnen. Das wird schon dadurch plausibel, daß in diese Berechnung garnicht die Grundwahrscheinlichkeit p eingehen würde. Vielmehr wird entsprechend (19) jedem ω_i eine Wahrscheinlichkeit für sein Eintreten unter Zuhilfenahme von $q := 1 - p$ zugeordnet:

$$\underbrace{0000}_{P(\omega_0):=qqqq} \quad \underbrace{0001}_{P(\omega_1):=qqqp} \quad \underbrace{0010}_{P(\omega_2):=qqpq} \quad \underbrace{0011}_{P(\omega_3):=qqpp} \quad \underbrace{0100}_{P(\omega_4):=qpqq}$$

$$\underbrace{0101}_{P(\omega_5):=qpqp} \quad \underbrace{0110}_{P(\omega_6):=qppq} \quad \underbrace{0111}_{P(\omega_7):=qppp} \quad \underbrace{1000}_{P(\omega_8):=pqqq} \quad \underbrace{1001}_{P(\omega_9):=pqqp} \quad \underbrace{1010}_{P(\omega_{10}):=pqpq}$$

$$\underbrace{1011}_{P(\omega_{11}):=pqpp} \quad \underbrace{1100}_{P(\omega_{12}):=ppqq} \quad \underbrace{1101}_{P(\omega_{13}):=ppqp} \quad \underbrace{1110}_{P(\omega_{14}):=pppq} \quad \underbrace{1111}_{P(\omega_{15}):=pppp}$$

d.h. die 0 wird durch q, die 1 durch p ersetzt und das Produkt gebildet. Demnach besitzt jedes $\omega_i \in A_2$ als Wahrscheinlichkeit für sein Eintreten den Wert p^2q^2. Außerdem enthält A_2 genau $\binom{4}{2}$ Elemente. Daher gilt

$$P(A_2) = \binom{4}{2} p^2 q^2.$$

Mit $p = 0.515$ erhält man

$$P(A_2) = 0.3743.$$

Antwort: Die Wahrscheinlichkeit, daß sich unter den vier Kindern genau zwei Knaben befinden, ist gleich 0.3743.

Bemerkungen:

- Für $p = \frac{1}{2}$ wird jedes ω_i gleich gewichtet mit $\frac{1}{2^4}$, sodaß in diesem und nur in diesem Fall $P(A_2) = 6/2^4 = \#A_2/\#\Omega$ gilt.
- Die Wahl $p = 1$ ($p = 0$) ordnet die gesamte Wahrscheinlichkeitsmasse dem Elementarereignis ω_{15} (ω_0) zu, alle anderen ω_i werden mit Null bewertet.
- Führt man die Zufallsgröße $X : \Omega \mapsto \{0, \ldots, 4\}$ gemäß

  $$X(\omega_i) := \text{Anzahl der Einsen in } \omega_i$$

  ein, so kann man

  $$P(X = 2) = \binom{4}{2} p^2 (1 - p)^2$$

  schreiben. Hier bedeutet die Gleichung $X = 2$ die Menge aller Elemente von Ω, die von X auf die 2 abgebildet werden, d.h. das Urbild von 2 vermöge X, welches mit der Menge A_2 übereinstimmt.

$\diamond$

2.8.2 Poisson-Verteilung $\mathscr{P}(\boldsymbol{\lambda})$

[18] S.335

Die Betrachtung der Poisson-Verteilung, benannt nach Siméon Poisson (1781-1840), kann durch Berechnungsschwierigkeiten von (24) für große n motiviert werden. Eine diskrete Zufallsgröße

$$X : \Omega \mapsto \mathbb{N}_0 \tag{26}$$

heißt mit dem nichtnegativen Parameter

$$\lambda \in \mathbb{R}, \quad 0 \leq \lambda$$

poissonverteilt, falls

$$\boxed{P(X = k) = f_\lambda(k) := \frac{\lambda^k}{k!} e^{-\lambda}}$$

was durch

$$X \sim \mathscr{P}(\lambda)$$

zum Ausdruck gebracht wird. Man beachte, daß wir es jetzt mit einer diskreten Zg zu tun haben, die unendlich viele Werte annehmen kann. Da X eindeutig ist, muß demzufolge auch Ω unendlich viele Elemente enthalten. Als Realisierung kann man sich die Menge aller unendlich langen Zeichenketten bestehend aus den zwei verschiedenen Zeichen 0 und 1 vorstellen. Als Zg X kann man die Anzahl des Erscheinens des Zeichens 1 verwenden. Offenbar gilt dann (26). Wie üblich stellt

$$e^x := \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

die wichtigste Funktion der Mathematik dar, die die **Eulersche e-Funktion** genannt wird (benannt nach Leonhard Euler (1707-1783)). Das Symbol e bezeichnet die Eulersche Zahl

$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = 2.71\,82\,81\ldots$$

Die e-Funktion erfüllt die Funktionalgleichung

$$e^{x+y} = e^x e^y.$$

Beachte, daß

$$f_\lambda : X(\Omega) \mapsto [0, 1]$$

die Wahrscheinlichkeitsfunktion von X darstellt, und daß die Summe aller Gewichte tatsächlich gleich 1 ist:

$$\sum_{k=0}^{\infty} f_\lambda(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = e^0 = 1.$$

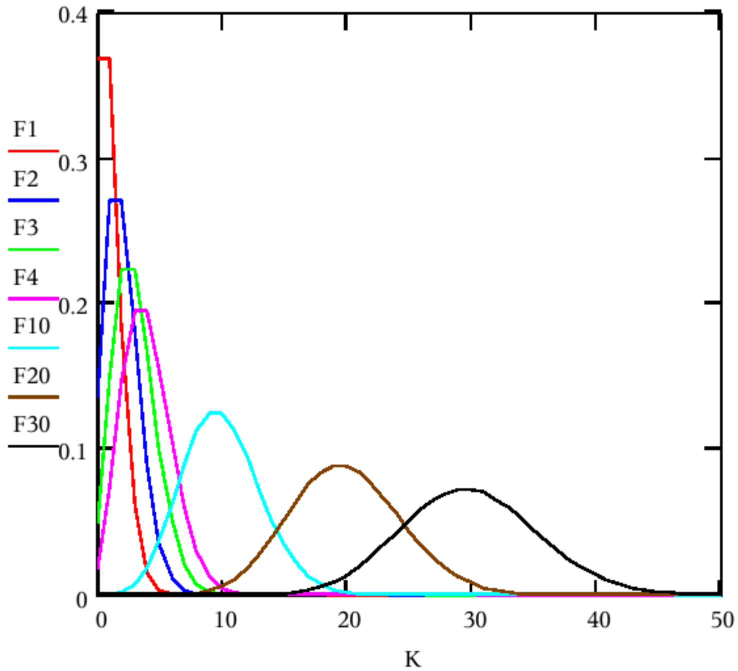


Abbildung 6: Die Wahrscheinlichkeitsfunktion $f_\lambda(k)$ einer Zg $X \sim \mathscr{P}(\lambda)$ über dem Bereich $K := [0, 1, \ldots, 50]$ für $\lambda = 1, 2, 3, 4, 10, 20, 30$. Offensichtlich nimmt $f_\lambda(k)$ für fixiertes λ sein Maximum in $k := \lambda$ an. Wie bei der Binomialverteilung hat man es mit Polygonzügen zu tun, die jeweils 51 Punkte miteinander verbinden. Auch hier fällt die Ähnlichkeit zur Gaußschen Glockenkurve auf.

Für $\lambda = 0$ bekommt man den Extremfall, der dem Ereignis $\{X = 0\}$ die Wahrscheinlichkeit 1 und allen anderen Ereignissen $\{X = k\}$, $k \neq 0$, die Wahrscheinlichkeit 0 erteilt. D.h. eine mit dem Parameter $\lambda = 0$ poissonverteilte Zg nimmt fast sicher nur den Wert 0 an. Eine Zg mit einer solchen Verteilung nennt man **ausgeartet**. Anwachsendes λ läßt auch die Ereignisse $\{X = k\}$ für nichtverschwindendes k wahrscheinlich werden. Dabei wird die größte Wahrscheinlichkeit dem Ereignis $\{X = k\}$ zugewiesen, dessen k mit dem ganzen Anteil von λ übereinstimmt. In Abbildung 6 wird das offensichtlich. Dort wird die Wahrscheinlichkeitsfunktion f_λ für verschiedene λ über der Menge $\{0, \ldots, 50\}$ gezeigt.

Erwartung. Für die Erwartung einer poissonverteilten Zg X bekommt man

$$EX = \sum_{k=0} k f_\lambda(k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda e^{\lambda} = \lambda e^{\lambda - \lambda} = \lambda e^0 = \lambda$$

d.h.

$$\boxed{EX = \lambda}$$

Varianz. Wegen

$$
\begin{aligned}
EX^2 &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} (k-1+1) \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \left(\sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) \\
&= \lambda e^{-\lambda} \left(\lambda \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) = \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) = \lambda(\lambda+1)
\end{aligned}
$$

bekommt man für die Varianz

$$
\mathrm{Var}(X) = EX^2 - (EX)^2 = \lambda(\lambda+1) - \lambda^2 = \lambda
$$

d.h.

$$
\boxed{\mathrm{Var}(X) = \lambda}
$$

Approximation der Binomial-Verteilung durch Poisson-Verteilung. Der Zusammenhang zur Binomial-Verteilung ergibt sich aus der Gleichung

$$
\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x .
$$

Setzt man voraus, daß die Zahlenfolge

$$
\lambda_n := n p_n, \quad n = 1, 2, \ldots
$$

den Grenzwert λ besitzt, so gilt

$$
\begin{aligned}
P(X=k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \frac{n!}{(n-k)!k!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\
&= \left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right) \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \qquad (27) \\
&= \underbrace{\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)}_{\to 1} \underbrace{\frac{\lambda_n^k}{k!}}_{\to \frac{\lambda}{k!}} \underbrace{\frac{\left(1-\frac{\lambda_n}{n}\right)^n}{\left(1-\frac{\lambda_n}{n}\right)^k}}_{\to e^{-\lambda}} \quad \to \frac{\lambda^k}{k!} e^{-\lambda}
\end{aligned}
$$

Numerische Analysen zeigen, daß für

$$\boxed{n \geq 10, \quad p \leq \frac{1}{10}, \quad \lambda := np}$$

die Approximation in der Regel ausreichend ist [15] S.115.

Beispiel 14 Approximation von Binomial- durch Poisson-Verteilung beim Versuch, ein feindliches Flugzeug vom Himmel zu holen, [9] S.93

Fünftausend Soldaten versuchen am Himmel ein feindliches Flugzeug abzuschießen. Jedem Schuß wird eine Trefferwahrscheinlichkeit von

$$p := \frac{1}{1000}$$

unterstellt. Außerdem sollen sich die Schüsse gegenseitig nicht beeinflussen. Wie groß ist die Wahrscheinlichkeit, daß das Flugzeug wenigstens zweimal getroffen wird?

Da 5000 mal geschossen wird, kann der Ereignisraum Ω mittels aller Binärzahlen der Länge 5000 modelliert werden. Es wird nach der Wahrscheinlichkeit des Ereignisses A gefragt, das aus allen solchen Binärzahlen besteht, die wenigstens zweimal die Ziffer 1 enthalten, wobei eine 1 an der Stelle i vergeben wird, falls der Soldat i trifft. Offenbar gilt

$$A = \Omega \setminus (A_0 \cup A_1)$$

wobei

- A_0 nur aus der Binärzahl der Länge 5000 besteht, die nur Nullen aufweist, und
- A_1 aus allen Binärzahlen der Länge 5000, bei denen die Ziffer 1 genau einmal erscheint.

Führt man die Zg

$$X : \Omega \mapsto \{0, \ldots, n\}$$

ein, die jedem $\omega \in \Omega$ die Anzahl des Erscheinens der Ziffer 1 zuordnet, so identifiziert man A_0 und A_1 wieder als die Urbildmengen von 0 und 1 vermöge X:

$$A_0 = \{X = 0\}, \quad A_1 = \{X = 1\}.$$

Wegen der Unabhängkeit aller abgegebenen Schüsse, kann die gesamte Salve als ein 5000 mal wiederholtes Bernoulli Experiment mit der Grundwahrscheinlichkeit $p = 1/1000$ aufgefaßt werden. Demnach gilt $X \sim \mathscr{B}(5000, 1/1000)$, d.h.

$$P(X=0) = \binom{5000}{0} p^0 (1-p)^{5000}, \quad P(X=1) = \binom{5000}{1} p^1 (1-p)^{4999}, \quad p = \frac{1}{1000}$$

sodaß

$$\begin{aligned} P(A) = {} & P(\Omega) - P(X=0) - P(X=1) \\ = {} & 1 - \binom{5000}{0} p^0 (1-p)^{5000} - \binom{5000}{1} p(1-p)^{4999} \\ = {} & 1 - \left(1 - \frac{1}{1000}\right)^{5000} - \frac{5000}{1000}\left(1 - \frac{1}{1000}\right)^{4999} = 0.959\,64 \end{aligned}$$

folgt.

Nimmt man nun

$$X \sim \mathscr{P}(\lambda), \quad \lambda := np = 5000 \times \frac{1}{1000} = 5$$

an, so bekommt man die Approximation

$$P(A) \approx 1 - \underbrace{\frac{\lambda^0}{0!} e^{-\lambda}}_{\approx P(X=0)} - \underbrace{\frac{\lambda^1}{1!} e^{-\lambda}}_{\approx P(X=1)} = 1 - (1+\lambda) e^{-\lambda} = 0.959\,56$$

die in den ersten drei Nachkommastellen mit der exakten Lösung übereinstimmt. ◇

Falls eine Zg X mit den Parametern p und n binomialverteilt ist, wird meistens nach dem Wert von

$$P := P(i \leq X \leq j) = \sum_{k=i}^{j} \binom{n}{k} p^k (1-p)^{n-k} \tag{28}$$

gefragt. Offenbar kann man sich aber auch die Wahrscheinlichkeiten

$$p, P \in (0,1)$$

und die Grenzen $i, j \in \{0, \ldots, n\}$ vorgeben und nach dem Parameter n fragen, sodaß Gleichung (28) gilt. In folgendem Beispiel wird diese Fragestellung behandelt und mit Hilfe der Poisson-Verteilung näherungsweise gelöst.

Beispiel 15 Inverse Fragestellung beantwortet mit Poisson-Verteilung, Mindestanzahl von Schrauben

Eine Firma stellt Schrauben her, wobei 2 % Ausschuss sind. Wieviele Schrauben gehören mindestens in eine Abpackung, damit die Firma mit wenigstens 95% Sicherheit garantieren kann, daß 100 gute Schrauben in jeder Packung enthalten sind?

In obigem Kontext erhalten wir $P = 0.95$ und $p = 0.02$. Da $P > 0$, müssen wenigstens 100 Schrauben in der Packung sein, da es sonst ausgeschlossen wäre, überhaupt einmal 100 gute Schrauben in einer Pachung zu haben. Daher gilt für den gesuchten Parameter n die Ungleichung $100 \leq n$. Die Zg X definieren wir durch die Anzahl der schlechten Schrauben in einer Packung von n Schrauben. Dann gilt

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

wobei n nicht bekannt ist. Offenbar hält die Firma ihr Garantieversprechen ein, falls

$$0.95 \leq P(X=0) + P(X=1) + \ldots + P(X=n-100).$$

Mit der letzten Gleichung geht diese Ungleichung in

$$0.95 \leq g(n)$$

über, wobei

$$g(n) := \binom{n}{0} p^0 (1-p)^{n-0} + \binom{n}{1} p^1 (1-p)^{n-1} + \ldots + \binom{n}{n-100} p^{n-100} (1-p)^{100}.$$

Demnach ist n ist so zu bestimmen, daß $g(n)$ nicht kleiner als 0.95 wird. Eine Auflösung nach n scheint ausgeschlossen, daher setzen wir einfach einige Werte für n ein und beobachten den Werteverlauf von g. Um nicht mit den Binomialkoeffizienten rechnen zu müssen, verwenden wir zur Approximation von g die Grenzwertbeziehung (27), was durch $n \geq 100$ gerechtfertigt ist:

$$g(n) \approx \left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^{n-100}}{(n-100)!} \right) e^{-\lambda}, \quad \lambda := n \times 0.02.$$

Auf diese Weise bekommen wir

$$g(103) = 0.8461, \quad g(104) = 0.9398, \quad g(105) = 0.9796, \quad g(106) = 0.9938$$

d.h. mit 104 Schrauben pro Packung wird das Garantieziel noch nicht erreicht, aber mit 105. $\diamond$

Das folgende Beispiel zeigt, daß $\mathscr{P}(\lambda)$ in der Physik direkt vorkommt.

Beispiel 16 Geiger-Müller-Zählrohr

Das Geiger-Müller-Zählrohr zählt bei einer radioaktiven Substanz wie oft in einer Zeitspanne der radioaktive Zerfall stattfindet (Einzelne ionisierende Teilchen erzeugen in einem gasgefüllten Gefäß durch Stoßionisation Ionenlawinen, die sich als Entladungsstoß äußern und so gezählt werden können). Auf Grundlage von längeren Beobachtungen, d.h. man zählt eine Stunde lang die Entladungsstöße, interessiert man sich für die Wahrscheinlichkeit, daß in einer vorgegebenen Zeitspanne $[t_0, t_1]$, z.B eine Minute, eine bestimmte Anzahl von Entladungsstößen vorsichgehen.

Es sei die Zg X durch die Anzahl der Entladungsstöße in $[t_0, t_1]$ erklärt.

Um die Verteilungsfunktion von X zu finden, teilen wir das Zeitintervall $[t_0, t_1]$ äquidistant in die Teilintervalle

$$[t_0, t_0 + \Delta], \quad [t_0 + \Delta, t_0 + 2\Delta] \quad \ldots \quad [t_0 + (n-1)\Delta, \underbrace{t_0 + n\Delta}_{t_1}], \quad \Delta := \frac{t_1 - t_0}{n}$$

ein und nehmen an, daß die Grundwahrscheinlichkeit, mit der in einem beliebigen Teilintervall genau ein Entladungsstoß vorsichgeht, proportional zur Teilintervallbreite Δ ist:

$$p_n := \Delta\varrho.$$

Es wird davon ausgegangen, daß entweder ein oder kein Entladungsstoß pro Zeitintervall stattfindet (Bernoullieexperiment). Da wir im nächsten Schritt die Intervallbreite gegen Null gehen lassen werden, ist das eine vernünftige Annahme.

Offenbar hängt p_n von einem Parameter ϱ ab, der von den Eigenschaften des radioaktiven Materials abhängt und der durch Beobachtung geschätzt wird.

Zur Approximation von X wählen wir die Zg X_n definiert durch die Anzahl der Entladungsstöße im diskretisierten Intervall $[t_0, t_1]$. Dann gilt

$$X_n \sim \mathscr{B}(n, p_n)$$

d.h.

$$P(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}.$$

Setzt man

$$\lambda_n := np_n$$

so gilt $\lambda_n = (t_1 - t_0)\varrho$, d.h.

$$\lim_{n\to\infty} \lambda_n = (t_1 - t_0)\varrho =: \lambda$$

und entsprechend (27)

$$P(X = k) = \lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

d.h. die Wahrscheinlichkeit, daß es während $[t_0, t_1]$ zu genau k Entladungsstößen kommt, stimmt mit der Zahl

$$\frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda = (t_1 - t_0)\varrho$$

überein, d.h. $X \sim \mathscr{P}(\lambda)$.

Es verbleibt, den Parameter ϱ zu schätzen. Zu diesem Zweck zählt man einfach für eine deutlich längere Zeitspanne T als $t_1 - t_0$ die Anzahl N_T der Entladungsstöße und setzt

$$\varrho := \frac{N_T}{T}.$$

Nach dem Bernouillschen Gesetz der großen Zahl ist diese Vorgehensweise gerechtfertigt.

Zahlenbeispiel: Bei einer radioaktiven Substanz werden mit einem GM-Zählrohr in einer Stunde 1961 Zerfallsakte registriert. Wie groß ist die Wahrscheinlichkeit, daß in einem 10 Sekunden Intervall genau 3 Zerfallsakte stattfinden? ([15], S.117)

Da eine Stunde aus 3600 Sekunden besteht, bekommen wir $T = 3600$ und $t = 10$. Daher gilt

$$\varrho \approx \frac{1961}{3600} = 0.5447, \quad \lambda := t\varrho = 10 \times 0.5447 = 5.447$$

und

$$P(X = 3) \approx \frac{5.447^3}{3!} \exp(-5.447) = 0.1159.$$

Antwort: Ungefähr mit einer Wahrscheinlichkeit von 11.59 % ist damit zu rechnen, daß es innerhalb von 10 Sekunden zu genau drei Entladungsstößen kommt. Da ϱ nur geschätzt wurde, läßt sich nichts genaueres sagen. ◇

Beispiel 17 Wieviel Geburten im 5-Minutenintervall

In einem statistikführenden Land werden pro Jahr 124 000 Geburten registriert. Wie groß ist ungefähr die Wahrscheinlichkeit, daß innerhalb eines beliebig ausgewählten 5-Minuten-Intervalls zwei oder mehr Geburten stattfinden?

Die Zg X sei die Anzahl der Geburten innerhalb von 5 Minuten. Lt. Aufgabenstellung ist nach

$$1 - P(X = 0) - P(X = 1)$$

gefragt, d.h. nach

$$1 - \frac{\lambda^0}{0!} e^{-\lambda} - \frac{\lambda}{1!} e^{-\lambda} = 1 - (1 + \lambda) e^{-\lambda}, \quad \lambda := t\varrho = 5\varrho.$$

Zur Schätzung von ϱ muß ein Jahr in Minuten umgerechnet werden

$$1 \text{ Jahr} = 365 \times 24 \times 60 = 525\,600 \text{ Minuten}$$

Demnach bekommen wir für ϱ die Schätzung

$$\varrho \approx \frac{124\,000}{525\,600} = 0.236$$

d.h. $\lambda = 1.18$ und

$$\begin{aligned} 1 - P(X = 0) - P(X = 1) &= 1 - (1 + \lambda) e^{-\lambda} = 1 - 2.18 \exp(-1.18) \\ &= 1 - 2.18 \times 0.31 = 1 - 0.68 = 0.32. \end{aligned}$$

Antwort: Ungefähr mit einer Wahrscheinlichkeit von 32 % ist damit zu rechnen, daß es innerhalb von 5 Minuten zu wenigstens zwei Geburten kommt. Da ϱ geschätzt werden muß, kann man nichts genaueres sagen. ◇

2.8.3 Geometrische Verteilung $\mathscr{G}(p)$

[5] S.234

Eine Zg $X : \Omega \mapsto \mathbb{N}$ heißt mit dem Parameter $p \in [0,1]$ **geometrisch** verteilt, falls für jedes $k \in \mathbb{N}$ die Gleichung

$$P(X = k) = p(1-p)^{k-1}$$

gilt, was durch

$$X \sim \mathscr{G}(p)$$

zum Ausdruck gebracht wird. Unter Zuhilfenahme der geometrischen Reihe

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}, \quad |q| < 1$$

erhält man sofort

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{k=0}^{\infty} (1-p)^k = \frac{p}{1-(1-p)} = 1.$$

Für die Erwartung bekommt man

$$EX = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = -p \sum_{k=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}p}(1-p)^k = -p \frac{\mathrm{d}}{\mathrm{d}p} \sum_{k=0}^{\infty} (1-p)^k$$

$$= -p \frac{\mathrm{d}}{\mathrm{d}p} \frac{1}{1-(1-p)} = -p \frac{\mathrm{d}}{\mathrm{d}p} \frac{1}{p} = \frac{1}{p}.$$

Zur Berechnung der Varianz benutzen wir

$$\mathrm{Var}(X) = EX^2 - (EX)^2 = EX^2 - \frac{1}{p^2}$$

wobei

$$EX^2 = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = \sum_{k=1}^{\infty} (k(k-1)+k) p(1-p)^{k-1}$$

$$= p \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-1} + \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-1} + \frac{1}{p}$$

$$= p(1-p) \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} + \frac{1}{p} = p(1-p) \frac{\mathrm{d}^2}{\mathrm{d}p^2} \sum_{k=0}^{\infty} (1-p)^k + \frac{1}{p}$$

$$= p(1-p) \frac{\mathrm{d}^2}{\mathrm{d}p^2} \frac{1}{p} + \frac{1}{p} = p(1-p) \frac{2}{p^3} + \frac{1}{p} = \frac{2(1-p)+p}{p^2} = \frac{2-p}{p^2}.$$

Daher gilt

$$\mathrm{Var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Zusammenfassend stellen wir

$$\boxed{EX = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}}$$

fest.

Beispiel 18 Würfeln bis eine 6 kommt

Wie groß ist die Wahrscheinlichkeit beim Würfeln, daß man genau beim dritten Wurf eine 6 bekommt?

Die Wahrscheinlichkeit für das Erwürfeln einer 6 beträgt $p := 1/6$. Daher bekommt man

$$p(1-p)^2 = \frac{1}{6} \times \frac{5^2}{6^2} = \frac{25}{216} = 0.1157.$$

$\diamond$

Das stetige Analogon zur geometrischen Verteilung ist die Exponentialverteilung.

2.8.4 Hypergeometrische Verteilung $\mathscr{H}(M, N, n)$

[18] S.331, [13] S.357, [15] S.117

Macht man sich von der Realisierung durch das Lottospiel '6 aus 49' frei und geht man in (6) zu 'M aus N' über, so bekommt man den Quotienten

$$\frac{\binom{M}{k}\binom{N-M}{M-k}}{\binom{N}{M}}$$

der die Wahrscheinlichkeit dafür angibt, in Bezug auf ein fixierte Teilmenge

$$\mathcal{D} := \{d_1, \ldots, d_M\} \subseteq \mathcal{N} := \{1, \ldots, N\}$$

von M Elementen aus $\mathcal{N}$ bei einer Ziehung von M Elementen genau k Elemente von $\mathcal{D}$ zu treffen.

Dieses Experiment läßt sich dahingehend verallgemeinern, daß man nach Fixierung von $\mathcal{D}$ nicht nur M Elemente aus $\mathcal{N}$ zieht, sondern n, die eine Menge $\mathcal{S}$ bilden. Dann gibt der Quotient

$$\boxed{f_{M,N,n}(k) := \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}}$$

die Wahrscheinlichkeit dafür an, daß $\mathcal{S}$ genau k Elemente von $\mathcal{D}$ enthält.

Eine diskrete Zg $X : \Omega \mapsto \{0, \ldots, M\}$ heißt **hypergeometrisch** verteilt mit den Parametern M, N, n, was durch

$$X \sim \mathscr{H}(M, N, n)$$

zum Ausdruck gebracht wird, falls sie die Werte $0, \ldots, M$ mit den Wahrscheinlichkeiten

$$P(X = k) = f_{M,N,n}(k)$$

annimmt.

Wir wissen, daß $\binom{n}{k}$ im kombinatorischen Sinn nur für $0 \le k \le n$ erklärt ist. Daher kann $f_{M,N,n}$ nur unter der Bedingung

$$0 \le k \le M, \quad 0 \le n-k \le N-M, \quad 0 \le n \le N$$

ausgewertet werden.

Insbesondere gilt entsprechend (4)

$$\sum_{k=0}^{M} f_{M,N,n}(k) = \binom{N}{n}^{-1} \underbrace{\sum_{k=0}^{M} \binom{M}{k}\binom{N-M}{n-k}}_{=\binom{N}{n}} = 1$$

d.h. $f_{M,N,n}$ ist tatsächlich eine Wahrscheinlichkeitsfunktion.

Aus Vollständigkeitsgründen stellen wir

$$\boxed{EX = np, \quad \mathrm{Var}(X) = np\,(1-p)\,\frac{N-n}{N-1}, \quad p := \frac{M}{N}}$$

fest.

Beispiel 19 Anwendung Hypergeom.-Verteilung bei der Qualitätskontrolle

In einer Kiste lagern $N := 1000$ Artikel $\{a_1, \dots, a_N\}$, wovon $M := 93$ Stück defekt sind. Der Qualitätskontrolleur nimmt eine Stichprobe $\mathcal{S}$ im Umfang von $n := 100$ Stück. Mit welcher Wahrscheinlichkeit befinden sich genau $k := 7$ defekte Stücke in dieser Stichprobe?

Der Ereignisraum Ω sei durch durch alle möglichen Stichproben $\mathcal{S}$ vom Umfang n erklärt, d.h. durch alle n-elementigen Teilmengen von $\{a_1, \dots, a_N\}$. $\mathcal{S}$ stellt somit ein Elementarereignis dar und es gilt $\#\Omega = \binom{N}{n}$.

Willkürlich setzen wir voraus, daß die Artikel $a_1, \dots, a_M$ defekt sind und bezeichnen diese Menge mit $\mathcal{D}$. Dann definieren wir

$$X : \Omega \mapsto \{0, \dots, M\}, \quad X(\mathcal{S}) := \#(\mathcal{S} \cap \mathcal{D})$$

d.h. durch die Anzahl defekter Elemente in der Stichprobe $\mathcal{S}$. Da sich genau M defekte Artikel in der Kiste befinden, kann X nur die Werte $0, 1, \dots, M$ annehmen.

Nach dem, was wir oben festgestellt hatten, gilt $X \sim \mathscr{H}(M, N, n)$, d.h.

$$P(X = k) = fM, N, n(k).$$

Numerische Spezifizierung der Argumente liefert

$$P(X=7) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{\binom{93}{7}\binom{1000-93}{100-7}}{\binom{1000}{100}} = 0.1106$$

d.h. man hat ungefähr eine Chance von 11 % genau 7 defekte Werkstücke zu erwischen, wenn man 100 Werkstücke auswählt. $\diamond$

In **mathcad** liefert die Auswertung von

$$\text{dhypergeom}(k, M, N-M, n)$$

den Wert von $f_{M,N,n}(k)$.

Anhand des Stichprobenbeispiels läßt sich $f_{M,N,n}(k)$ leicht interpretieren: sein Zähler stimmt mit der Anzahl L aller Stichproben vom Umfang n überein, die genau k defekte Artikel enthalten, sein Nenner mit der Anzahl aller möglichen Stichproben. Daß tatsächlich

$$L = \binom{M}{k}\binom{N-M}{n-k}$$

gilt, sieht man wiefolgt. Der Faktor $\binom{M}{k}$ liefert die Anzahl aller Möglichkeiten k defekte Artikel aus $\mathcal{D}$ auszuwählen, wobei $\#\mathcal{D} = M$ zu beachten ist. Der zweite Faktor liefert die Anzahl aller Möglichkeiten, die restlich verfügbaren Plätze in einer Stichprobe, in der sich schon k defekte Artikel befinden, noch mit $n-k$ Elementen aus $N-M$ schadlosen Artikeln aufzufüllen.

2.9 p-Quantile

2.9.1 Das p-Quantil für Zg mit stetiger streng monoton wachsender Verteilungsfunktion

Unter dem ***p*-Quantil** einer stetigen Zg X, die entsprechend F verteilt ist, versteht man das Minimum der Lösungen der Gleichung

$$F(z) = p, \quad p \in (0,1).$$

Falls die Dichtefunktion f immer positiv ist, verhält sich F streng monoton wachsend, sodaß F eine Umkehrfunktion F^{-1} besitzt und das p-Quantil mit der Zahl

$$F^{-1}(p)$$

übereinstimmt. [2]

[2] Generell existiert für eine streng monoton wachsende Funktion f eine Funktion g mit

$$g(f(x)) = f(g(x)) = x.$$

Solch eine Funktion wird mit f^{-1} bezeichnet und heißt die Umkehrfunktion von f. Sie ist im allgemeinen verschieden von $1/f$.

Wegen

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

gilt

$$P(X \leq z) = p$$

d.h. das p-Quantil liefert eine Schranke z, die die Zg X auf

$$100p\%$$

des Ereignisraumes Ω nicht überschreitet. Ist z.B. z das 0.6-Quantil der Zg X, so ist X auf 60 % von Ω durch z beschränkt.

Das p-Quantil ist geeignet, einen vorgeschriebenen Anteil von Ω abzutrennen: Wenn man beispielsweise 75 % von Ω abtrennen möchte, kann man das 0.75-Quantil z berechnen und die Feststellung treffen, daß 75% von Ω den Wert z nicht überschreiten.

2.9.2 Das p-Quantil für diskrete Zg

Für diskrete Zg sind die Verhältnisse etwas komplizierter, da die Verteilungsfunktionen treppenförmig sind, somit nicht streng monoton wachsen und daher keine Umkehrfunktion besitzen. Falls p im Wertevorrat von F liegt, besitzt die Gleichung $F(x) = p$ ein nach rechts halboffenes Intervall auf der x-Achse als Lösungsmenge. Darüberhinaus existieren $p \in [0,1]$, die von keinem $x \in \mathbb{R}$ als Bild vermöge F angenommen werden, d.h. die Lösungsmenge der Gleichung $F(x) = p$ ist leer. Um diese Defizite zu überwinden, kann man beispielsweise

$$G(p) := \min\{x \in \mathbb{R} : F(x) = \widehat{p}\}$$

setzen, wobei $\widehat{p}$ das kleinste Element aus dem Wertevorrat von F ist, das die Ungleichung

$$p \leq \widehat{p}$$

erfüllt. Dann ist G eine linksstetige Treppenfunktionen, die man als eine (verallgemeinerte) Inverse von F ansehen kann. Diese Definition von G folgt der **mathcad**-Implementation von

$$\text{qbinom}(x, n, p).$$

Beispiel 20 Folgendes Beispiel soll zeigen, was diese Konstruktion bedeutet. Wir betrachten $X \sim \mathscr{B}(n,p)$ für $n = 5$ und $p = 0.5$. Dann nimmt X nur die Werte $0, \ldots, 5$ jeweils mit der Wahrscheinlichkeit

$$\frac{1}{32}, \frac{5}{32}, \frac{10}{32}, \frac{10}{32}, \frac{5}{32}, \frac{1}{32}$$

an, so daß F von der Gestalt

$$F(x) = \begin{cases} 0, & \text{falls } x < 0 \\ \frac{1}{32}, & \text{falls } 0 \leq x < 1 \\ \frac{6}{32}, & \text{falls } 1 \leq x < 2 \\ \frac{16}{32}, & \text{falls } 2 \leq x < 3 \\ \frac{26}{32}, & \text{falls } 3 \leq x < 4 \\ \frac{31}{32}, & \text{falls } 4 \leq x < 5 \\ 1, & \text{falls } 5 \leq x \end{cases}$$

ist. Der Wertevorrat von F sind die kommulierten Wahrscheinlichkeiten

$$0,\ \frac{1}{32},\ \frac{6}{32},\ \frac{16}{32},\ \frac{26}{32},\ \frac{31}{32},\ 1$$

sodaß man entsprechend der Definition von G

$$G(p) = \begin{cases} 0, & \text{falls} \quad 0 \leq p \leq \frac{1}{32} \\ 1, & \text{falls} \quad \frac{1}{32} < p \leq \frac{6}{32} \\ 2, & \text{falls} \quad \frac{6}{32} < p \leq \frac{16}{32} \\ 3, & \text{falls} \quad \frac{16}{32} < p \leq \frac{26}{32} \\ 4, & \text{falls} \quad \frac{26}{16} < p \leq \frac{31}{32} \\ 5, & \text{falls} \quad \frac{31}{32} < p \leq 1 \end{cases}$$

erhält. Offenbar sind die p-Quantile für $p \in (\frac{6}{32}, \frac{16}{32}] = (18.75\%, 50\%]$ alle gleich 2. Daher hat man mit der Schranke 2 zu arbeiten, wenn man z.B. 40% aller Fälle abtrennen möchte: In 40% aller Fälle ist die Zg X durch 2 beschränkt. Tatsächlich gilt diese Feststellung aber für 50 % aller Fälle, was kein Widerspruch ist. Da die Verteilungsfunktion Sprünge aufweist, gibt es keine Schranke, die genau 40 % aller Fälle abtrennt.

◇

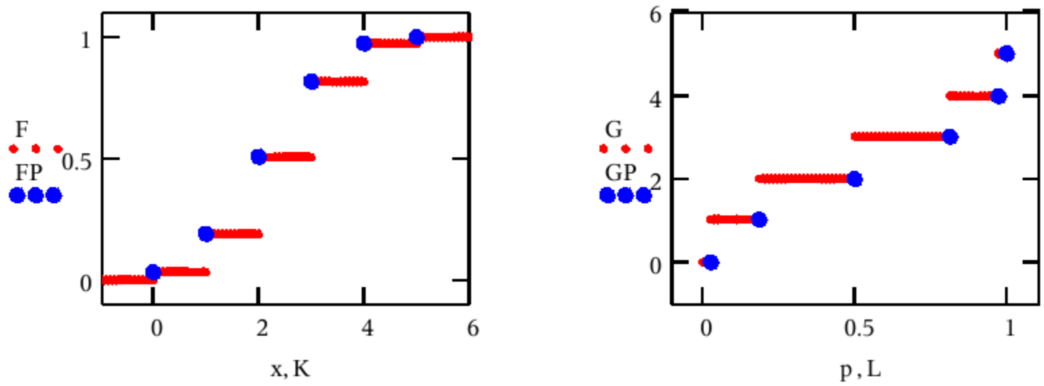


Abbildung 7: Die linke Seite zeigt F und die rechte G. Die blauen Punkte geben an, welche Funktionswerte an den Sprungstellen angenommen werden. Offensichtlich ist F rechtsstetig, d.h. durchläuft man eine Stufe von rechts nach links, so gehört die Stufenkante mit zur Stufe. Die Stufenhöhen stimmen mit den Wahrscheinlichkeiten $\frac{1}{32}$, $\frac{5}{32}$, $\frac{10}{32}$, $\frac{10}{32}$, $\frac{5}{32}$, $\frac{1}{32}$ überein. Offensichtlich ist G linksstetig, d.h. durchläuft man eine Stufe von rechts nach links, so gehört die Stufenkante nicht mit zur Stufe. Jede Stufe besitzt die Höhe 1.

2.10 Wichtige stetige Verteilungen

Eine Zg X über einem Ereignisraum heißt **stetig**, falls ihr Wertevorrat überabzählbar und ihre Verteilungsfunktion von der Gestalt

$$\boxed{F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t}$$

ist. Der Integrand f heißt die **Dichtefunktion** von X, besitzt die Eigenschaften

$$\boxed{0 \leq f(t), \qquad \int_{-\infty}^{+\infty} f(t)\mathrm{d}t = 1}$$

und ist stetig bis auf endlich viele Punkte, d.h. f kann in endlich vielen Punkten Sprung- oder auch Polstellen besitzen. Ein Beispiel für eine Dichtefunktion mit einer Polstelle ist durch die Dichte der χ^2-Verteilung zum Freiheitsgrad 1 gegeben. Beispiele für Dichtefunktionen mit Sprungstellen sind durch die Gleichverteilung und durch die Exponentialverteilung gegeben.

Die Wahrscheinlichkeit $P(a \leq X \leq b)$, daß X im Intervall $[a, b]$ Werte annimmt, erfüllt die Gleichung

$$P(a \leq X \leq b) = F(b) - F(a) = \int_{a}^{b} f(t)\mathrm{d}t.$$

Es folgt unmittelbar, daß bei einer Zg mit einer Dichtefunktion, die keine Sprung- und Polstellen aufweist, die Gleichung

$$P(X = a) = 0$$

gilt, d.h. die Wahrscheinlichkeit, daß X einen diskreten Wert annimmt, ist Null.

In Anlehnung an (12) ist der **Erwartungswert** einer stetigen Zg X mit der Dichtefunktion f entsprechend

$$\boxed{\mu := EX := \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x}$$

definiert und die **Varianz** entsprechend

$$\boxed{\sigma^2 := \mathrm{Var}(X) := \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) \mathrm{d}x}$$

Die vom diskreten Fall bekannte Gleichung

$$\boxed{\mathrm{Var}(X) = EX^2 - (EX)^2}$$

bleibt bestehen.

2.10.1 Gleichverteilung

[18] S.318

Die Wahrscheinlichkeit beim Würfeln, eine bestimmte Augenanzahl zu erwürfeln, ist immer 1/6, d.h. die Zg 'Augenanzahl' ist gleichverteilt und die zugehörige Wahrscheinlichkeitsfunktion lautet

$$f(x) := \begin{cases} 1/6, & \text{falls } x \in \{1, \dots, 6\}, \\ 0, & \text{falls } x \in \mathbb{N} \setminus \{1, \dots, 6\}. \end{cases}$$

Offenbar ist $f : \mathbb{N} \mapsto \{0, \frac{1}{6}\}$ eine Treppenfunktion. Das stetige Analogon lautet wiefolgt.

Eine stetige Zg X heißt **identisch verteilt** oder **gleichverteilt** über dem Intervall $[a, b] \subset \mathbb{R}$, falls ihre Dichtefunktion von der Gestalt

$$f_{a,b}(x) = \begin{cases} \dfrac{1}{b-a}, & \text{falls } x \in [a, b] \\ 0, & \text{falls } x \in \mathbb{R} \setminus [a, b] \end{cases} \tag{29}$$

ist, d.h. die Wahrscheinlichkeit, daß X einen Wert innerhalb eines Intervalls $[x, y]$ annimmt stimmt mit der Länge des Durchschnitts $[x, y] \cap [a, b]$ dividiert durch $b - a$ überein. Die zugehörige Verteilungsfunktion lautet

$$F_{a,b}(x) := \int_{-\infty}^{x} f_{a,b}(t) \mathrm{d}t.$$

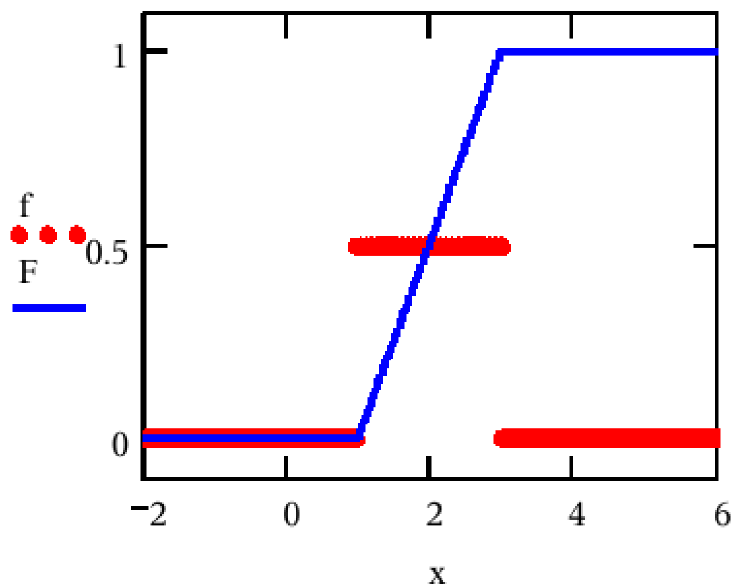


Abbildung 8: Es werden die Graphen der Dichte (29) (rot) und der zugehörigen Verteilungsfunktion $F_{a,b}(x) =: \int_{-\infty}^{x} f_{a,b}(t)\mathrm{d}t$ (blau) für $a = 1$ und $b = 3$ gezeigt.

Beispiel 21 Warten an der Straßenbahnhaltestelle

An einer Haltestelle kommt pünktlich aller 20 Minuten eine Straßenbahn. Eine Person kommt zufällig an die Haltestelle und nimmt die nächste Bahn. Wie groß ist die Wahrscheinlichkeit, daß die Person zwischen 5 und 10 Minuten warten muß?

Da die nächste Straßenbahn genommen wird, ist die Wartezeit durch 20 min beschränkt, da es keine negativen Wartezeiten gibt, nimmt unsere Zg W (Wartezeit) nur Werte im Intervall $[0, 20]$ an. Wir bestimmen nun die Spanne aller Zeitpunkte, bei denen man höchstens 10 und mindestens 5 Minuten auf die nächste Bahn warten muß. Diese Zeitpunkte liegen offenbar in dem Intervall, welches 10 min nach der Abfahrt der vorhergehenden Bahn beginnt und 5 min vor der Ankunft der nächsten Bahn endet. Die Breite dieses Intervall ist ins Verhältnis zu 20 zu setzen, d.h. die Wahrscheinlichkeit, daß man zwischen 5 und 10 min auf die nächste Bahn warten muß ist

$$P(5 \leq W \leq 10) = \frac{10-5}{20} = \frac{1}{4}.$$

Was hat das nun mit der Gleichverteilung zu tun? Falls man etwas allgemeiner nach der Wahrscheinlichkeit fragt, daß man wenigstens x und höchstens y Minuten zu warten hat, so bekommt man

$$P(x \leq W \leq y) = \frac{y-x}{b-a}, \quad b := 20, \quad a := 0.$$

als Antwort, wobei $a \le x \le y \le 20$ vorausgesetzt ist. Setzt man

$$F_{a,b}(x) := \begin{cases} 0 & \text{falls } x \le a \\ \dfrac{x-a}{b-a} & \text{falls } x \in [a,b] \\ 1 & \text{falls } b \le x \end{cases}$$

so gilt

$$P(x \le W \le y) = F_{a,b}(y) - F_{a,b}(x)$$

d.h. $F_{a,b}$ ist für $a = 0$ und $b = 20$ die Verteilungsfunktion von W. Schließlich gilt

$$\frac{\mathrm{d}}{\mathrm{d}x} F_{a,b}(x) := \begin{cases} 0 & \text{falls } x \le a \\ \dfrac{1}{b-a} & \text{falls } x \in [a,b] \\ 0 & \text{falls } b \le x \end{cases}$$

d.h.

$$\frac{\mathrm{d}}{\mathrm{d}x} F_{a,b}(x) = f_{a,b}(x)$$

was die Gleichverteilung der Zg W bedeutet. ◇

Erwartung und Varianz.

$$\boxed{EX = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}}$$

Zur Begründung haben wir

$$EX = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \frac{1}{b-a}\int_a^b x\mathrm{d}x = \frac{1}{b-a}\frac{x^2}{2}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

und

$$EX^2 = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \frac{1}{b-a}\int_a^b x^2\mathrm{d}x = \frac{1}{b-a}\frac{x^3}{3}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

d.h.

$$\begin{aligned} \mathrm{Var}(X) &= EX^2 - (EX)^2 = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} \\ &= \frac{1}{12}\big(4(b^2+ab+a^2) - 3(a^2+2ab+b^2)\big) = \frac{(b-a)^2}{12}. \end{aligned}$$

2.10.2 Dreiecksverteilung, Simpsonsches Verteilungsgesetz

[10] S.138

Eine Dreiecksverteilung erhält man bei der Betrachtung der Summe zweier gleichverteilter Zg. Einzelheiten dazu werden in Beispiel 33 bereitgestellt.

2.10.3 Exponentialverteilung $\mathscr{E}(\boldsymbol{\lambda})$

[5] S.279

Die Exponentialverteilung ist das stetige Analogon zur geometrischen Verteilung, wenn man die Wartezeit bis zum Eintreten eines bestimmten Ereignisses als stetige Zg auffaßt.

Eine Zg $X : \Omega \mapsto \mathbb{R}_+ \cup \{0\}$ heißt mit dem Parameter $\lambda \in \mathbb{R}_+$ exponentialverteilt, was durch

$$X \sim \mathscr{E}(\lambda)$$

ausgedrückt wird, falls ihre Dichtefunktion von der Gestalt

$$f(x) := \begin{cases} 0 & \text{falls } x < 0 \\ \lambda e^{-\lambda x} & \text{falls } 0 \leq x \end{cases}$$

ist. Wegen

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \lambda \int_{0}^{+\infty} e^{-\lambda x}\mathrm{d}x = -e^{-\lambda x}\Big|_0^{+\infty} = 1$$

hat man es tatsächlich mit einer Dichtefunktion zu tun.

Wir zeigen, daß sich die Exponentialverteilung als Grenzfall der geometrischen Verteilung interpretieren läßt. Zu diesem Zweck unterteilen wir das Zeitintervall $[0, T)$ in n Intervalle

$$\underbrace{\left[0, \frac{T}{n}\right)}_{=:\,T_1}, \underbrace{\left[\frac{T}{n}, \frac{2T}{n}\right)}_{=:\,T_2}, \underbrace{\left[\frac{2T}{n}, \frac{3T}{n}\right)}_{=:\,T_3} \cdots \underbrace{\left[\frac{(n-1)T}{n}, \frac{nT}{n}\right)}_{=:\,T_n}$$

und unterstellen für das Eintreten eines Ereignisses A während der Zeitspanne T_k die Wahrscheinlichkeit $\frac{\lambda T}{n}$, wobei λ ein zu schätzender Parameter ist. Demnach tritt das Ereignis A mit der Wahrscheinlichkeit

$$\left(1 - \frac{\lambda T}{n}\right)^n$$

nicht vor dem Zeitpunkt T ein. Demnach erhält man mit $n \to \infty$ als Wahrscheinlichkeit, daß A nicht vor dem Zeitpunkt T eintritt, den Wert

$$\lim_{n\to\infty} \left(1 - \frac{\lambda T}{n}\right)^n = e^{-\lambda T}.$$

Führt man die Zg X als die Zeit ein, die vergeht bis das Ereignis A das Erstemal eintritt, so kann man

$$P(X > T) = e^{-\lambda T}$$

schreiben, d.h.

$$P(X \leq T) = 1 - e^{-\lambda T} = \int_0^T \lambda e^{-\lambda x}\mathrm{d}x = \int_{-\infty}^T f(x)\mathrm{d}x.$$

Die Verteilungsfunktion der Exponentialverteilung lautet also

$$F(x) := \begin{cases} 0 & \text{falls } x < 0, \\ 1 - e^{-\lambda x} & \text{falls } 0 \leq x. \end{cases}$$

Offenbar bedeutet ein großer Parameterwert eine hohe Wahrscheinlichkeit für das Eintreten von A vor dem Zeitpunkt T.

Wir berechnen die Erwartung und die Varianz einer exponentiell verteilten Zg. Die Stammfunktion von xe^{-x} lautet

$$-(x+1)e^{-x} + C, \quad C \in \mathbb{R}.$$

Daher bekommen wir mittels der Substitution $y = \lambda x$ für die Erwartung

$$EX = \int_0^{+\infty} x\lambda e^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}\int_0^{+\infty} ye^{-y}\mathrm{d}y = \frac{1}{\lambda}\left(-(y+1)e^{-y}\right)\Big|_0^{+\infty} = \frac{1}{\lambda}.$$

Den Grenzwert

$$\lim_{y\to+\infty} \frac{y+1}{e^y} = 0$$

kann man mit der Regel von L'Hospital verifizieren.

Die Varianz berechnen wir mittels

$$\mathrm{Var}(X) = EX^2 - (EX)^2.$$

Durch Differentiation bestätigt man, daß

$$-(x^2 + 2x + 2)e^{-x} + C, \quad C \in \mathbb{R}$$

die Stammfunktion von x^2e^{-x} ist.

Daher gilt

$$\begin{aligned} EX^2 &= \int_0^{+\infty} x^2\lambda e^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}\int_0^{+\infty} x^2\lambda^2 e^{-\lambda x}\mathrm{d}x \\ &= \frac{1}{\lambda^2}\int_0^{+\infty} y^2e^{-y}\mathrm{d}y = \frac{1}{\lambda^2}\left(-(y^2+2y+2)e^{-y}\right)\Big|_0^{\infty} \\ &= \frac{2}{\lambda^2} \end{aligned}$$

sodaß

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Den Grenzwert

$$\lim_{y\to+\infty} \frac{y^2+2y+2}{e^y} = 0$$

kann man mit der Regel von L'Hospital verifizieren.

Zusammenfassend stellen wir

$$\boxed{EX = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}}$$

fest.

Die Exponentialverteilung wird vor allem im Zusammenhang mit der Lebenserwartung technischer Geräte und in der Theorie der Warteschlangen verwendet [10].

2.10.4 Normalverteilung $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$

[18] S.338

Wir kommen jetzt zur wichtigsten Verteilungsfunktion in der Wahrscheinlichkeitsrechnung. Die Dichtefunktion der sogenannten **Normalverteilung** $\Phi_{\mu,\sigma}$ mit den Parametern $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$ ist durch

$$\boxed{\varphi_{\mu,\sigma}(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}}$$

gegeben, d.h. die zugehörige Verteilungsfunktion lautet

$$\boxed{\Phi_{\mu,\sigma}(x) := \int_{-\infty}^{x} \varphi_{\mu,\sigma}(t)\mathrm{d}t}$$

Der Graph von $\varphi_{\mu,\sigma}$ ist die bekannte, nach C.F. Gauß (1777-1855) benannte Glockenkurve. Eine einfache Kurvendiskussion ergibt, daß $\varphi_{\mu,\sigma}$ ihr Maximum in μ annimmt, wobei

$$\varphi_{\mu,\sigma}(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

und je ein Wendepunkt bei $\mu \pm \sigma$ liegt. Daher ist σ ein Maß für die 'Breite' der Glockenkurve. Eine Vergrößerung von σ bewirkt ein Absenken des Kurvengipfels und ein Auseinanderrücken der Wendepunkte. Abbildung 9 zeigt den Graphen von $\varphi_{\mu,\sigma}$ für verschiedene Werte von μ und σ.

Wir identifizieren μ als die Erwartung und σ als die Standardabweichung einer normalverteilten Zg.

Erwartung und Varianz. Für eine Zg X, die entsprechend $\mathscr{N}(\mu, \sigma^2)$ verteilt ist, gelten die Gleichungen

$$\boxed{EX = \mu, \quad \mathrm{Var}(X) = \sigma^2}$$

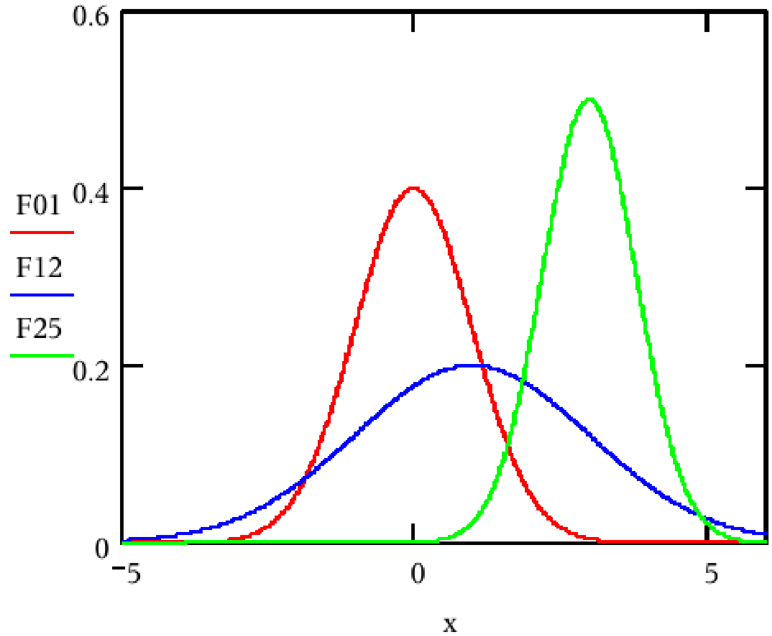


Abbildung 9: Es wird die Dichte der Normalverteilung für die Parameterwerte $\mu = 0, \sigma = 1$ (F01), $\mu = 1, \sigma = 2$ (F12) und $\mu = 2, \sigma = 0.5$ (F25) gezeigt. Je größer die Varianz σ, um so breiter wird die Kurve. Jeder Graph besitzt in den Punkten $\mu \pm \sigma$ jeweils einen Wendepunkt. Der Erwartungswert μ bestimmt die Lage ihres Gipfels. Die Höhe des Gipfels beträgt $\frac{1}{\sqrt{2\pi}\sigma}$. Demnach verursacht eine Vergrößerung von σ ein Absenken des Gipfels und ein Auseinanderrücken der Wendepunkte. Das nächste Bild zeigt den Meister persönlich, links von ihm befindet sich die Glockenkurve. In D. Struik, *Abriss der Geschichte der Mathematik* fällt auf S.161 der Satz: *Auf der Trennlinie der Mathematik des achtzehnten zum neunzehnten Jahrhundert erhebt sich die majestätische Gestalt von Carl Friedrich Gauß.* Die Einbindung in 'majestätische' Verhältnisse wird nur sehr wenigen Mathematikern zuteil. Auf die Frage: *Wen halten Sie für den größten Mathematiker aller Zeiten?* antwortet Prof. E. Zeidler (Leipzig): *Für den König der Mathematik halten viele Isaac Newton, ich persönlich favorisiere C. F. Gauß.*

Die Feststellung $EX = \mu$ folgt aus der Gleichung ([19], S.326)

$$\int_{-\infty}^{+\infty} e^{-x^2/2} \mathrm{d}x = \sqrt{2\pi} \tag{30}$$

die wiefolgt verifiziert werden kann. Offenbar gilt die Implikation

$$a := \int_{-\infty}^{+\infty} e^{-x^2/2} \mathrm{d}x \quad \Rightarrow \quad a^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2/2} e^{-y^2/2} \mathrm{d}x \mathrm{d}y = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2} \mathrm{d}x \mathrm{d}y.$$

Das rechtsstehende Bereichsintegral kann mittels der Variablentransformation

$$x(r,\varphi) = r\cos(\varphi), \quad y(r,\varphi) = r\sin(\varphi)$$

berechnet werden. Die Jakobische Funktionaldeterminante lautet

$$\det \begin{bmatrix} x_r & x_\varphi \\ y_r & y_\varphi \end{bmatrix} = \det \begin{bmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{bmatrix} = r(\underbrace{\cos^2(\varphi) + \sin^2(\varphi)}_{=1}) = r.$$

Daher gilt

$$a^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r \mathrm{d}r \mathrm{d}\varphi = -2\pi e^{-r^2/2}|_0^{\infty} = 2\pi.$$

Schließlich bekommt man mit der Variablentransformation $y = (x-\mu)/\sigma$ die Gleichung $\sigma \mathrm{d}y = \mathrm{d}x$, sodaß

$$\begin{aligned} EX &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma y + \mu) e^{-\frac{y^2}{2}} \sigma \mathrm{d}y \\ &= \frac{\sigma}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2}} \mathrm{d}y}_{=0} + \frac{\mu}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} \mathrm{d}y}_{=\sqrt{2\pi}} = \mu. \end{aligned}$$

Der erste Summand verschwindet, da der Integrand eine ungerade Funktion ist und das Integrationsintervall symmetrisch um den Nullpunkt liegt.

Unter Benutzung von

$$\int_0^{\infty} x^2 e^{-\alpha^2 x^2} \mathrm{d}x = \frac{\sqrt{\pi}}{4\alpha^3}$$

bekommt man mit $\alpha := 1/\sqrt{2}$ für die Varianz

$$\begin{aligned} \mathrm{Var}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 y^2 e^{-\frac{y^2}{2}} \sigma \mathrm{d}y = \sigma^2 \frac{1}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} \mathrm{d}y}_{=\sqrt{2\pi}} = \sigma^2. \end{aligned}$$

Es gilt der **Linearitätssatz**: Es sei X normalverteilt mit den Parametern μ und σ und die Zg Y sei entsprechend

$$Y := aX + b, \quad a \neq 0$$

geformt. Dann gilt

$$\boxed{Y \sim \mathscr{N}(a\mu + b, a^2\sigma^2)}$$

Es gilt der sogenannte **Additionssatz**: Gilt

$$X_i \sim \mathscr{N}(\mu_i, \sigma_i^2)$$

und sind X_1 und X_2 unabhängig voneinander, so folgt

$$\boxed{X_1 + X_2 \sim \mathscr{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)}$$

Beachte, daß entsprechend (14) und (16) Erwartung und Varianz von $X_1 + X_2$ wie behauptet sein müssen, das Wesentliche am Additionssatz ist die Eigenschaft der Summe, wieder normalverteilt zu sein. In Abschnitt 2.11 wird genauer auf den Additionssatz und seine Anwendungen eingegangen.

Die Verteilungsfunktion $\Phi_{0,1}$ heißt **Standard-Normalverteilung**. Es besteht die Beziehung

$$\boxed{\Phi_{\mu,\sigma}(x) = \Phi_{0,1}\left(\frac{x-\mu}{\sigma}\right)} \tag{31}$$

sodaß eine numerische Berechnung von $\Phi_{\mu,\sigma}(x)$ durch die numerische Berechnung von $\Phi_{0,1}\left(\frac{x-\mu}{\sigma}\right)$ ersetzt werden kann. Da der Werteverlauf von $\Phi_{0,1}$ in tabellierter Form vorliegt, vgl. Tabelle (78), reduziert sich die Berechnung von $\Phi_{\mu,\sigma}(x)$ auf das Heraussuchen eines Tabellenwertes an der Stelle $\frac{x-\mu}{\sigma}$.

Aufgrund der Symmetrie von $\varphi_{0,1}$ gilt

$$\boxed{\Phi_{0,1}(x) = 1 - \Phi_{0,1}(-x)} \tag{32}$$

Daher reicht es aus, $\Phi_{0,1}$ nur für nichtnegative Argumente zu tabellieren, was in Tabelle (78) bis zum Argument 3.09 realisiert wird. Für $x > 3.09$ setzen wir $\Phi_{0,1}(x) := 1$.

Gleichung (32) impliziert unmittelbar:

$$\left.\begin{array}{l} \Phi_{0,1}(x_0) = \alpha \\ \Phi_{0,1}(x_1) = 1 - \alpha \end{array}\right\} \quad \Rightarrow \quad \Phi_{0,1}(x_1) = 1 - \Phi_{0,1}(x_0) \quad \Leftrightarrow \quad x_0 = -x_1.$$

Außerdem gilt

$$\begin{array}{lcl} \Phi_{0,1}(x) \leq 0.5 & \Leftrightarrow & x \leq 0, \\ \Phi_{0,1}(x) \geq 0.5 & \Leftrightarrow & x \geq 0. \end{array}$$

In **mathcad** liefert die Auswertung von

$$\mathrm{dnorm}(x, \mu, \sigma), \quad \mathrm{pnorm}(x, \mu, \sigma), \quad \mathrm{qnorm}(x, \mu, \sigma)$$

('norm' entspricht $\Phi_{\mu,\sigma}$, d=density, p=probability, q=quantile) jeweils den Wert von

$$\varphi_{\mu,\sigma}(x), \qquad \Phi_{\mu,\sigma}(x), \qquad \Phi^{-1}_{\mu,\sigma}(x).$$

Man beachte, daß in dem letzten Befehl für x nur Werte zwischen 0 und 1 eingesetzt werden dürfen und daß $\Phi^{-1}_{\mu,\sigma}(x)$ die Umkehrfunktion von $\Phi_{\mu,\sigma}(x)$ darstellt.

Beispiel 22 *p*-Quantil bei Normalverteilung, Ableseübung in Tabelle (78)

Eine Zg X sei normalverteilt mit der Erwartung $\mu = 2$ und der Standardabweichung $\sigma = 1$. Gefragt wird nach einer oberen Schranke z, die X in 95 % aller Fälle nicht überschreitet.

Zur Berechnung von z hat man das 0.95-Quantil von X zu berechnen, d.h. die Gleichung

$$P(X \leq z) = \Phi_{2,1}(z) = 0.95$$

zu lösen. Da $\Phi_{2,1}$ invertierbar ist, kann die Umkehrfunktion dazu benutzt werden:

$$\Phi^{-1}_{2,1}(0.95) = \mathrm{qnorm}(0.95, 2, 1) = 3.645.$$

Daher ist in 95 % aller Fälle die Zg X durch 3.645 beschränkt.

Falls man nur eine Tabelle mit den Werten von $\Phi_{0,1}$ für positive Argumente zur Verfügung hat, geht man wie folgt vor: Normalisierung liefert

$$0.95 = \Phi_{2,1}(z) = \Phi_{0,1}\left(\frac{z-2}{1}\right).$$

Um die Gleichung $\Phi_{0,1}(z) = 0.95$ zu lösen, suchen wir im Inneren von Tabelle (78) den Wert, der am nächsten bei 0.95 liegt. Das sind die Werte 0.9495 und 0.9505. Auf dem Rand der Tabelle lesen wir den jeweils dazugehörigen Wert ab, in diesem Falle sind das 1.64 und 1.65. Um nun noch etwas genauer zu sein, nehmen wir den Mittelwert

$$z = 1.645 \quad \Rightarrow \quad \frac{z-2}{1} = 1.645 \quad \Rightarrow \quad z = 3.645.$$

◇

Beispiel 23 Christstollen der Fa. Dr. Quendt

Es hat sich herausgestellt, daß die Massen der Christstollen dieser Firma mit den Parametern $\mu := 1000$ und $\sigma = 15$ normalverteilt sind. Die Christstollen werden zu je 9 Stück in Kartons verpackt und verschickt. Das Verpackungsmaterial hat die konstante Masse von 200.

- Wie ist die Gesamtmasse der gefüllten Kartons verteilt, wenn man Unabhängigkeit voraussetzt?

- Bestimmen Sie die Wahrscheinlichkeit, daß ein Karton weniger als 9100 und zwischen 9155 und 9245 wiegt.

Es sei X_i die Zg, die durch die Masse der i-ten Stolle erklärt ist. Nach Voraussetzung gilt

$$X_i \sim \mathscr{N}(1000, 15^2).$$

Die Zg X sei durch die Massen der gefüllten Kartons erklärt. Dann gilt

$$X = \sum_{i=1}^{9} X_i + b, \quad b := 200.$$

Da Unabhängigkeit vorliegt, liefert die Kombination von Additions- und Linearitätssatz die Normalverteilung von X mit den Parametern

$$\mu_X := 9\mu + b = 9.200, \quad \sigma_X = 3\sigma = 45$$

womit die erste Frage beantwortet ist.

Als Folgerung erhält man

$$\begin{aligned} P(X \leq 9100) \; &= \Phi_{9200,45}(9100) = \Phi_{0,1}\left(\frac{9100 - 9200}{45}\right) \\ &= \Phi_{0,1}\,(-2.222) = 1 - \Phi_{0,1}\,(2.222) \\ &= 1 - 0.987 = 0.013 \end{aligned}$$

d.h. in 1.3 % aller Fälle wiegt ein Karton weniger oder gleich 9100.

Die Frage nach $P(9155 \leq X \leq 9245)$ zielt auf die sog. 1σ-Regel ab:

$$X \sim \mathscr{N}(\mu, \sigma^2) \Rightarrow P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6826.$$

Wegen $9155 = 9200 - 45 = \mu - \sigma$ und $9245 = 9200 + 45 = \mu + \sigma$ lautet die Antwort 0.6826, d.h. in 68.26 % aller Fälle wiegt ein Karton zwischen 9155 und 9245. ([15], S.124) $\diamond$

Beispiel 24 $k\sigma$-Regel, $k = 1, 2, 3$

Es wird nach der Wahrscheinlichkeit gefragt, daß eine normalverteilte Zg X Werte annimmt, die von ihrem Erwartungswert höchstens $k\sigma$ entfernt sind. Rechnerich erhält man

$$\begin{aligned} P(\mu - k\sigma \leq X \leq \mu + k\sigma) &= \Phi_{0,1}\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi_{0,1}\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\ &= \Phi_{0,1}(k) - \Phi_{0,1}(-k) = 2\,\Phi_{0,1}(k) - 1 \end{aligned}$$

sodaß man in Abhängigkeit von k folgende Wahrscheinlichkeiten bekommt:

k	1	2	3
P	0.6826	0.9544	0.9974

D.h. ca. 68% aller Werte von X liegen weniger als σ von μ entfernt und der Fall, daß X einen Wert annimmt, der mehr als 3σ von μ entfernt ist, ist fast ausgeschlossen. ◇

Beispiel 25 Schätzung von μ und σ mittels eines Gleichungssystems

[15] S.131 Aufg. 3.35

Der Durchmesser von serienmäßig hergestellten Kugeln sei normalverteilt. Die Parameter μ und σ können wie folgt geschätzt werden. Von zwei Sieben hat

- Sieb 1 Löcher mit Durchmesser 69 mm und
- Sieb 2 Löcher mit Durchmesser 72 mm.

Man beobachtet, daß

- 30.86 % der Kugeln durch Sieb 1 hindurchfallen und
- 15.87 % der Kugeln von Sieb 2 aufgehalten werden.

Wie groß sind μ und σ?

Es sei X die Zg, die durch die Kugeldurchmesser erklärt wird.

Da 30.86 % aller Kugeln einen kleineren Durchmesser als 69 mm haben, gilt

$$P(X < 69) = 0.3086.$$

Da 15.87 % einen Kugeldurchmesser von mehr als 72 mm besitzen, gilt

$$P(72 < X) = 1 - P(X < 72) = 0.1587.$$

Da $X \sim \mathcal{N}(\mu, \sigma^2)$, folgt

$$\Phi_{\mu,\sigma}(69) = 0.3086, \quad \Phi_{\mu,\sigma}(72) = 0.8413.$$

Wegen

$$\Phi_{\mu,\sigma}(x) = \Phi_{0,1}\left(\frac{x-\mu}{\sigma}\right)$$

bekommen wir

$$\Phi_{0,1}\left(\frac{69-\mu}{\sigma}\right) = 0.3086, \quad \Phi_{0,1}\left(\frac{72-\mu}{\sigma}\right) = 0.8413.$$

Wegen

$$\Phi_{0,1}(-0.5) = 0.3086, \quad \Phi_{0,1}(1) = 0.8413$$

und der Eineindeutigkeit von $\Phi_{0,1}$ folgt damit

$$\frac{69-\mu}{\sigma} = -0.5, \quad \frac{72-\mu}{\sigma} = 1$$

d.h.

$$69 - \mu = -\sigma/2, \quad 72 - \mu = \sigma.$$

Die Lösung dieses (2×2)-Gleichungssystem lautet

$$\mu = 70 \text{ mm}, \quad \sigma = 2 \text{ mm}.$$

Bemerkung zur Lösung der Gleichungen

$$\Phi_{0,1}(x) = 0.3086, \quad \Phi_{0,1}(x) = 0.8413. \tag{33}$$

mittels Tabelle, die nur für nichtnegative x ausgelegt ist

Zur Lösung der zweiten Gleichung reicht ein Blick in die Werteverlaufstabelle von $\Phi_{0,1}$. Man sucht im Inneren der Tabelle den Wert, der am nächsten bei 0.8413 liegt und erhält durch Zusammensetzung seiner Koordinaten die Lösung 1.00.

Da der Werteverlauf von $\Phi_{0,1}$ nur für nichtnegative x tabelliert ist, erscheinen in diesem Verlauf nur Werte, die größer oder gleich 0.5 sind. Da 0.3086 kleiner ist, muß die Lösung der ersten Gleichung negativ sein. An dieser Stelle benutzen wir

$$\Phi_{0,1}(x) = 1 - \Phi_{0,1}(-x)$$

nämlich

$$0.3086 = 1 - \Phi_{0,1}(-x) \quad \Rightarrow \quad \Phi_{0,1}(-x) = 1 - 0.3086 = 0.6914.$$

Blick in die Tabelle liefert $-x = 0.50$, d.h. $x = -0.5$. ◇

Beispiel 26 Bolzen und Löcher, Anwendung des Additionssatzes

[15] S.126

Eine Maschine produziert Bolzen, deren Durchmesser mit dem Mittelwert 9.8 mm und der Standardabweichung 0.1 mm normalverteilt sind. Eine andere Maschine bohrt Löcher, deren Durchmesser ebenfalls normalverteilt sind, aber mit dem Mittelwert 10 mm und der Standardabweichung 0.08 mm. Es darf Unabhängigkeit vorausgesetzt werden. Wie groß ist die Wahrscheinlichkeit, daß ein beliebig ausgewählter Bolzen in ein beliebig ausgewähltes Loch passt?

Den Ereignisraum Ω kann man sich als das Kartesische Produkt aus allen Bolzen und allen Löchern vorstellen:

$$\Omega := \left\{ (B,L) : \begin{array}{l} B \text{ ist ein von Maschine 1 hergestellter Bolzen} \\ L \text{ ist ein von Maschine 2 gebohrtes Loch} \end{array} \right\}$$

Die Zg $X, Y, Z : \Omega \mapsto \mathbb{R}$ agieren wiefolgt

$$\begin{aligned} X((B,L)) &:= \text{Durchmesser von } B, \\ Y((B,L)) &:= \text{Durchmesser von } L, \\ Z((B,L)) &:= X((B,L)) - Y((B,L)). \end{aligned}$$

Gilt für ein Elementarereignis ω die Ungleichung $Z(\omega) < 0$, so passt der Bolzen durch das Loch. Demnach wird lt. Aufgabenstellung nach

$$P(Z < 0)$$

gefragt. Nach Voraussetzung sind X und Y unabhängig und es gelten die Gleichungen

$$\mu_X = 9.8, \quad \mu_Y = 10, \quad \sigma_X = 0.1, \quad \sigma_Y = 0.08.$$

Da X und Y unabhängig und normalverteilt, gilt nach dem Additionssatz

$$Z \sim \mathcal{N}(\underbrace{\mu_X - \mu_Y}_{=-0.2}, \underbrace{\sigma_X^2 + \sigma_Y^2}_{=0.1281^2}).$$

Daher

$$P(Z < 0) = \Phi_{-0.2,\,0.1281}(0) = \text{pnorm}(0, -0.2, 0.1281) = 0.941$$

d.h. in 94 % aller Fälle passt ein zufällig ausgewählter Bolzen durch ein zufällig ausgewähltes Loch.

Falls man nur eine Tabelle für die Funktionswerte von $\Phi_{0,1}(x)$, $0 \leq x$, zur Verfügung hat, benutzt man Gleichung (31):

$$\Phi_{-0.2,\,0.1281}(0) = \Phi_{0,1}\left(\frac{0.2}{0.1281}\right) = \Phi_{0,1}(1.562) = 0.941.$$

◇

Beispiel 27 Schokoladentafeln

Von einer Kiste voller Schokoladentafeln ist bekannt, daß deren Gewichte mit den Parametern $\mu := 100$ und $\sigma := 2$ normalverteilt sind.

Wie muß man eine positive Konstante c wählen, damit 93 % aller Tafeln in Bezug auf ihr Gewicht im Intervall $[\mu - c, \mu + c]$ liegen?

Offenbar muß die Unbekannte c die Gleichungen

$$\begin{aligned} 0.93 &= P(\mu - c \leq X \leq \mu + c) = P(X \leq \mu + c) - P(X \leq \mu - c) \\ &= \Phi_{\mu,\sigma}(\mu + c) - \Phi_{\mu,\sigma}(\mu - c) \\ &= \Phi_{0,1}(c/\sigma) - \Phi_{0,1}(-c/\sigma) = 2\,\Phi_{0,1}(c/\sigma) - 1 \end{aligned}$$

erfüllen. Demnach hat man die Gleichung

$$\frac{1.93}{2} = 0.965 = \Phi_{0,1}(c/2)$$

zu lösen. Blick in die Werteverlaufstabelle von $\Phi_{0,1}$ liefert $c/2 = 1.812$, was schließlich auf

$$c = 3.624$$

führt. $\diamond$

2.10.5 χ^2-Verteilung

[9] S.143, [18] S.343

Bei der χ^2-Verteilung interessiert man sich für die Verteilung der Summe von Quadraten von **unabhängigen standardnormalverteilten** Zufallsgrößen: Falls

$$\chi^2 := X_1^2 + \ldots + X_n^2, \quad X_i \sim \mathcal{N}(0,1)$$

so gehorcht die Zg χ^2 einer sogenannten **χ^2-Verteilungsfunktion**

$$F_n(x) := \int_{-\infty}^{x} f_n(t)\mathrm{d}t$$

mit n Freiheitsgraden, wobei ihre Dichtefunktion

$$\boxed{f_n(x) := \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \begin{cases} 0 & \text{falls } x < 0 \\ x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & \text{falls } 0 \le x \end{cases}}$$

lautet. Hier bezeichnet

$$\Gamma(x) := \int_0^{\infty} t^{x-1}e^{-t}\mathrm{d}t, \quad 0 < x \tag{34}$$

die **Gammafunktion**. Beachte, daß Γ an den Stellen $n/2$, $n \in \mathbb{N}$, ausgewertet wird, so daß in (34) die Bedingung $0 < x$ erfüllt ist. Die Gammafunktion kann wegen

$$\Gamma(n+1) = n!, \quad n \in \mathbb{N}_0$$

als eine Verallgemeinerung von $n!$ aufgefaßt werden. Insbesondere gelten die Gleichungen

$$\Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(1) = 1, \quad \Gamma\left(n + \frac{1}{2}\right) = \frac{1 \times 3 \times 5 \times \ldots \times (2n-1)}{2^n}\sqrt{\pi}.$$

Erwartung und Varianz. Für eine mit n Freiheitsgraden χ^2-verteilte Zg X gilt

$$\boxed{EX = n,\ \mathrm{Var}(X) = 2n}$$

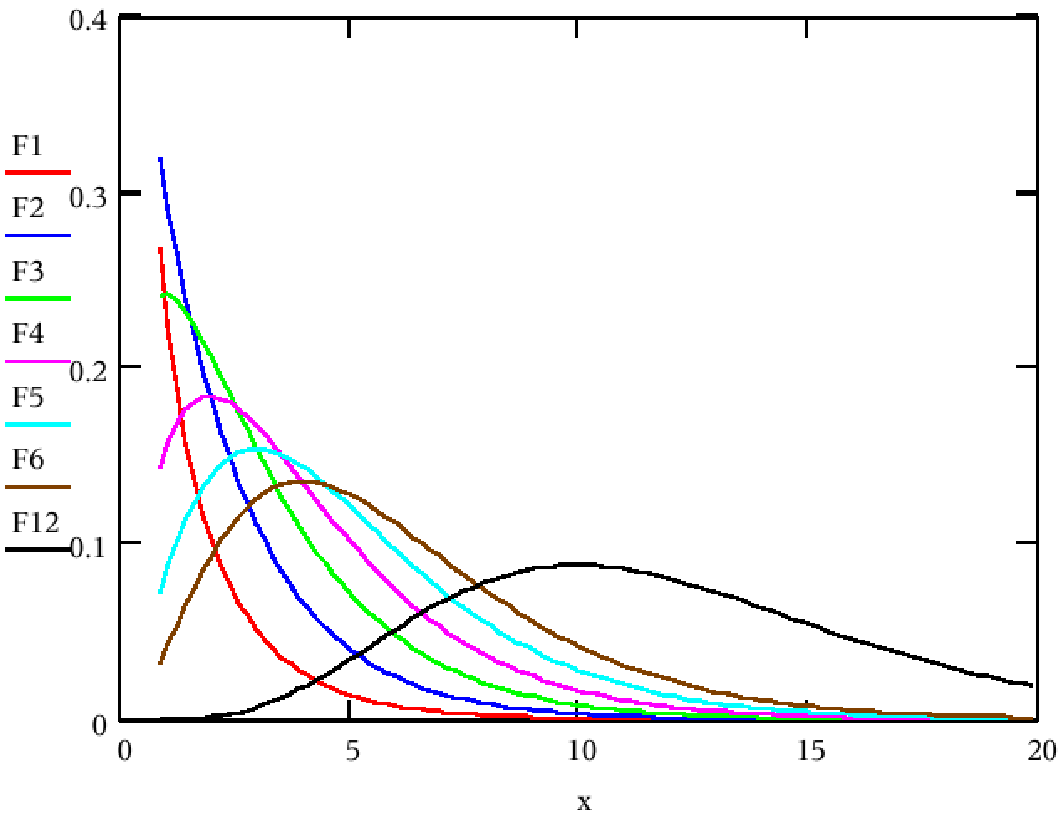


Abbildung 10: Graphen der Dichten der χ^2-Verteilung für die Freiheitsgrade $1, 2, 3, 4, 5, 6$ und 12.

Für $n = 1, 2, 3, 4$ bekommen wir

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi}} \begin{cases} 0, & \text{falls } x < 0 \\ x^{-\frac{1}{2}} e^{-\frac{x}{2}}, & \text{falls } 0 \le x \end{cases} & f_2(x) &= \frac{1}{2} \begin{cases} 0\,, & \text{falls } x < 0 \\ e^{-\frac{x}{2}}, & \text{falls } 0 \le x \end{cases} \\ f_3(x) &= \frac{1}{\sqrt{2\pi}} \begin{cases} 0\,, & \text{falls } x < 0 \\ x^{\frac{1}{2}} e^{-\frac{x}{2}}, & \text{falls } 0 \le x \end{cases} & f_4(x) &= \frac{1}{4} \begin{cases} 0\,, & \text{falls } x < 0 \\ x e^{-\frac{x}{2}}, & \text{falls } 0 \le x \end{cases} \end{aligned} \tag{35}$$

d.h. die ersten beiden sind in $\mathbb{R}_+$ streng monoton fallend in Übereinstimmung mit der roten und der blauen Kurve in Abbildung 10. Offensichtlich stimmt

$$F_2(x) = \begin{cases} 0, & \text{falls } x < 0 \\ 1 - e^{-\frac{x}{2}}, & \text{falls } 0 \le x \end{cases}$$

mit der Exponential-Verteilung für den Parameter $\lambda := \frac{1}{2}$ überein. Offensichtlich gilt

$$F_4(x) = \begin{cases} 0, & \text{falls } x < 0, \\ 1 - (1 + \frac{x}{2}) e^{-\frac{x}{2}}, & \text{falls } 0 \le x. \end{cases}$$

Für $2 < n$ nimmt $f_n(x)$ sein Maximum an der Stelle $n - 2$ an. Mit zunehmendem n wird die χ^2-Verteilung immer ähnlicher zur Normalverteilung mit $\mu := n$ und $\sigma^2 = 2n$. In Abbildung 10 ist zu erkennen, daß schon für $n = 12$ der Graph "Glockengestalt" annimmt mit einem Gipfel bei 10. In Abschnitt 2.11.2 überprüfen wir die Richtigkeit von (35).

In **mathcad** liefert die Auswertung von

$$\text{dchisq}(x, n), \quad \text{pchisq}(x, n), \quad \text{qchisq}(x, n)$$

(χ-square entspricht 'chisq', d=density, p=probability, q=quantile) jeweils den Wert von

$$f_n(x), \quad F_n(x), \quad F_n^{-1}(x)$$

an der Stelle $x \in \mathbb{R}_+$ für den Freiheitsgrad n.

Wie in Abschnitt 3.4.3 vorgeführt wird, benutzt man die χ^2-Verteilung unter anderem zur Berechnung von Vertrauensintervallen für die Varianz einer normalverteilten Grundgesamtheit. Daher zählt sie zur Gruppe der sogenannten Testverteilungen. Der Werteverlauf von F_n^{-1} ist für $n = 1, \ldots, 30$ und $p = 0.9, 0.95, 0.99$ in Abschnitt 6.3 tabelliert.

Für $n = 2$ ist die Umkehrfunktion F_2^{-1} leicht berechenbar, offensichtlich gilt für $p \in [0, 1)$

$$F_2^{-1}(p) = -2 \ln(1 - p).$$

Setzt man der Reihe nach für p die Werte 0.99, 0.95 und 0.90 ein, so erhält man Zeile 2 aus der Tabelle auf Seite 253.

Die vierte Zeile aus der Tabelle auf Seite 253 ergibt sich aus den Lösungen der Gleichung

$$p = 1 - \left(1 + \frac{x}{2}\right) e^{-\frac{x}{2}}$$

für $p \in \{0.99, 0.95, 0.90\}$.

Für $n = 1$ und $n = 3$ kann man F_n nicht explizit angeben, es gelten aber die Gleichungen

$$F_1(6.63) = \frac{1}{\sqrt{2\pi}} \int_0^{6.63} t^{-\frac{1}{2}} e^{-\frac{t}{2}} \mathrm{d}t = 0.99, \quad F_1(3.84) = 0.95, \quad F_1(2.71) = 0.9$$

$$F_3(11.35) = \frac{1}{\sqrt{2\pi}} \int_0^{11.35} t^{\frac{1}{2}} e^{-\frac{t}{2}} \mathrm{d}t = 0.99, \quad F_3(7.81) = 0.95, \quad F_3(6.25) = 0.9$$

in Übereinstimmung mit den Zeilen 1 und 3 in der Tabelle auf S.253.

Explizite Darstellung von F_n für gerades n. Für $n = 2k$ lautet die Dichtefunktionen

$$f_{2k}(x) = \frac{1}{2^k\,(k-1)!} x^{k-1} e^{-\frac{1}{2}x}.$$

Um die zugörige Verteilungsfunktion zu finden, stellen wir fest, daß für $\alpha \neq 0$ die Funktion

$$g(x, \alpha, k) := \left[1, x, \ldots, x^k\right] \begin{bmatrix} \alpha & 1 & & & \\ & \alpha & 2 & & \\ & & \ddots & \ddots & \\ & & & \alpha & k \\ & & & & \alpha \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} e^{\alpha x}$$

die Gleichungen

$$\frac{\mathrm{d}}{\mathrm{d}x} g(x, \alpha, k) = x^k e^{\alpha x}, \quad g(0, \alpha, k) = (-1)^k \frac{k!}{\alpha^{k+1}}$$

erfüllt. Daher gilt

$$\begin{aligned} F_{2k}(x) = &\ \frac{1}{2^k(k-1)!} \int_0^x t^{k-1} e^{-\frac{1}{2}t} \mathrm{d}t = \frac{1}{2^k(k-1)!} \left(g\left(x, -\frac{1}{2}, k-1\right) - g\left(0, -\frac{1}{2}, k-1\right)\right) \\ = &\ 1 + \frac{1}{2^k(k-1)!} g\left(x, -\frac{1}{2}, k-1\right). \end{aligned}$$

Wir werten für $k = 3$ diese Darstellung aus. Für postives x gilt

$$F_6(x) = 1 + \frac{1}{16}[1, x, x^2] \begin{bmatrix} -\frac{1}{2} & 1 & \\ & -\frac{1}{2} & 2 \\ & & -\frac{1}{2} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} e^{-\frac{x}{2}} = 1 - \left(1 + \frac{x}{2} + \frac{x^2}{8}\right) e^{-\frac{x}{2}}.$$

Im allgemeinen findet man

$$F_{2k}(x) = \begin{cases} 0 & \text{falls } x < 0, \\ 1 - e^{-\frac{x}{2}} p_{k-1}(x) & \text{falls } 0 \leq x. \end{cases}, \quad p_k(x) := \sum_{i=0}^{k} \frac{1}{i!} \left(\frac{x}{2}\right)^i$$

Offenbar stellt $p_{k-1}(x)$ den Anfangsabschnitt der Taylorreihe von $e^{-x/2}$ der Länge k dar.

Falls eine Zg X einer χ^2-Verteilung mit n Freiheitsgraden gehorcht, so drücken wir diesen Umstand durch $X \sim \chi^2(n)$ aus.

2.10.6 Studentsche Verteilung

[9] S.145, [13] S.440, [18] S.345

Es seien X und Y zwei unabhängige Zg, wobei

$$X \sim \mathcal{N}(0,1), \quad Y \sim \chi^2(n).$$

Dann ist die Zg

$$T := \frac{X}{\sqrt{Y}}\sqrt{n}$$

entsprechend der t-Verteilung mit n Freiheitsgraden verteilt, was durch

$$T \sim t_n$$

zum Ausdruck gebracht wird. Die Dichtefunktion der t-Verteilung lautet

$$\boxed{f_n(x) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}}$$

Wegen $\Gamma(1)=1$, $\Gamma(1/2)=\sqrt{\pi}$ und $\Gamma(3/2)=\frac{1}{2}\sqrt{\pi}$, erhält man

$$f_1(x) = \frac{1}{\pi}\frac{1}{(1+x^2)}, \quad f_2(x) = \frac{1}{2\sqrt{2}}\left(1+\frac{x^2}{2}\right)^{-\frac{3}{2}}. \tag{36}$$

In Abschnitt 2.12.2 überprüfen wir (36).

Erwartung und Varianz. Für eine mit n Freiheitsgraden t-verteilte Zg X gilt

$$\boxed{EX = \begin{cases} \text{existiert nicht,} & \text{falls } n=1 \\ 0, & \text{falls } 2 \le n \end{cases} \quad \operatorname{Var}(X) = \begin{cases} \text{existiert nicht,} & \text{falls } n<3 \\ \dfrac{n}{n-2}, & \text{falls } 3 \le n \end{cases}}$$

Die Betrachtung der Zg T und deren Verteilung wurde von dem englischen Mathematiker und Chemiker William Sealey Gosset (1876-1937) angeregt. Abbildung 11 zeigt die Graphen der Dichtefunktion $f_n(x)$ über dem Intervall $[-6,6]$ für verschiedene Freiheitsgrade.

In **mathcad** liefert die Auswertung von

$$\mathrm{dt}(x,n), \quad \mathrm{pt}(x,n), \quad \mathrm{qt}(x,n)$$

jeweils den Wert von

$$f_n(x), \quad F_n(x), \quad F_n^{-1}(x).$$

Wie in Abschnitt 3.4.2 vorgeführt wird, benutzt man die t-Verteilung unter anderem zur Berechnung von Vertrauensintervallen für die Erwartung μ einer normalverteilten

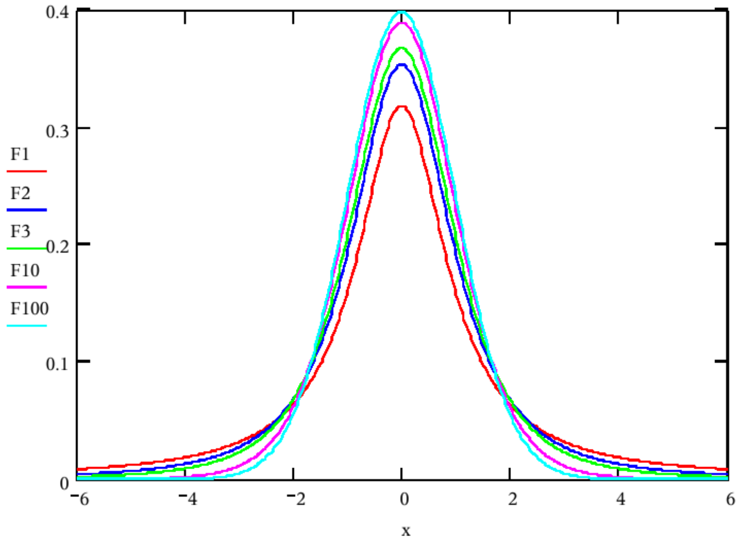


Abbildung 11: Die Dichte der t-Verteilung für die Freiheitsgrade $1, 2, 3, 10$ und 100. Offenbar sind alle Graphen symmetrisch bezüglich der y-Achse. Daher wird die t-Verteilung symmetrisch genannt.

Grundgesamtheit, wobei σ unbekannt ist. Daher zählt sie zur Gruppe der sogenannten Testverteilungen.

Die t-Verteilung wird auch STUDENTsche Verteilung genannt, da Gosset dazu unter dem Pseudonym STUDENT veröffentlicht hat. Gosset hat dieses Verteilungsgesetz empirisch gefunden.

In [9] S.145 wird als STUDENT-Verteilung die Verteilung eines Quotienten berechnet, dessen Zähler mit den Parametern $\mu := 0$ und $\sigma^2 := 1/n$ normalverteilt ist.

2.10.7 FISHERsche Verteilung

Die FISHERsche Verteilungfunktion beschreibt die Verteilung des Quotienten zweier χ^2-verteilter Zg, der bei der Überprüfung der Hypothese, ob zwei normalverteilte Zg gleiche Varianz besitzen, von Interesse ist.

2.10.8 WEIBULL-Verteilung

Diese Verteilungsfunktion wird bei Materialermüdungsvorgängen zur Vorhersage benutzt, wann mit dem Ausfall eines Bauteils zu rechnen ist. Sie ist nach dem schwedischen Ingenieur Wallodi Weibull (1887-1979) benannt.

2.11 Verteilung der Summe unabhängiger Zufallsgrößen

Gegeben sind zwei Zg X und Y über demselben Ereignisraum Ω mit den Verteilungsfunktionen F_X und F_Y. Gefragt wird nach der Verteilungsfunktion F_{X+Y} ihrer Summe.

2.11.1 Die Faltung

Falls X und Y diskret und unabhängig sind, so bekommt man die Wahrscheinlichkeitsfunktion von

$$Z := X + Y$$

folgendermaßen. Entsprechend der Definition (11) sind die Wahrscheinlichkeitsfunktionen f_X und f_Y nur auf den abzählbaren Wertevorräten von X und Y definiert. Ihre Fortsetzung auf gesamt $\mathbb{R}$ sei gemäß

$$f_X(x) := \begin{cases} P(X = x), & \text{falls } x \in X(\Omega) \\ 0, & \text{falls } x \notin X(\Omega) \end{cases}, \quad f_Y(x) := \begin{cases} P(Y = x), & \text{falls } x \in Y(\Omega) \\ 0, & \text{falls } x \notin Y(\Omega) \end{cases}$$

erklärt, d.h. man setzt sie an den Stellen auf Null, die nicht im Wertevorrat der betrachteten Zg liegen. Dann kann man für jedes $z \in \mathbb{R}$ aufgrund der Unabhängigkeit von X und Y

$$\begin{aligned} f_Z(z) &= P(Z = z) = \sum_{x_i + y_j = z} P(\{X = x_i\} \cap \{Y = y_j\}) = \sum_{x_i + y_j = z} P(X = x_i) P(Y = y_j) \\ &= \sum_{x_i + y_j = z} f_X(x_i) f_Y(y_j) = \sum_{j=0}^{\infty} f_X(z - y_j) f_Y(y_j) \end{aligned}$$

schreiben. Die Funktion

$$\boxed{h : \mathbb{R} \mapsto \mathbb{R}, \quad h(z) := \sum_{i=0}^{\infty} f_X(z - y_i) f_Y(y_i)}$$

heißt die **Faltung** von f_X und f_Y. Sie stellt für unabhängige diskrete Zg X und Y die Wahrscheinlichkeitsfunktion der Summe $X + Y$ dar.

Beispiel 28 Faltung diskreter Zg, Würfeln mit zwei Würfeln

Wir wählen als X_1 die Augenanzahl von Würfel 1 und als X_2 die Augenanzahl von Würfel 2. Die Wahrscheinlichkeitsfunktionen beider Zg stimmen überein, wobei

$$f(x) := f_{X_1}(x) = f_{X_2}(x) = \begin{cases} \dfrac{1}{6} & \text{falls } x \in \{1, \ldots, 6\}, \\ 0 & \text{falls } x \in \mathbb{R} \setminus \{1, \ldots, 6\}. \end{cases}$$

Dann besitzt die Summe $Z := X_1 + X_2$ die Wahrscheinlichkeitsfunktion

$$f_Z(z) = \sum_{i=1}^{6} f(z-i)f(i), \quad y_i = i.$$

Da Z den Wertevorrat $\{2,\dots,12\}$ besitzt, ist f_Z nur an diesen Stellen verschieden von Null. Im Einzelnen gilt

$$\begin{aligned}
f_Z(2) &= \textstyle\sum_{i=1}^{1} f(2-i)f(i) &&= f(1)f(1) &&= \tfrac{1}{36} \\
f_Z(3) &= \textstyle\sum_{i=1}^{2} f(3-i)f(i) &&= f(2)f(1)+f(1)f(2) &&= \tfrac{2}{36} \\
f_Z(4) &= \textstyle\sum_{i=1}^{3} f(4-i)f(i) &&= f(3)f(1)+f(2)f(2)+f(1)f(3) &&= \tfrac{3}{36} \\
f_Z(5) &= \textstyle\sum_{i=1}^{4} f(5-i)f(i) &&= f(4)f(1)+\ldots+f(1)f(4) &&= \tfrac{4}{36} \\
f_Z(6) &= \textstyle\sum_{i=1}^{5} f(6-i)f(i) &&= f(5)f(1)+\ldots+f(1)f(5) &&= \tfrac{5}{36} \\
f_Z(7) &= \textstyle\sum_{i=1}^{6} f(7-i)f(i) &&= f(6)f(1)+\ldots+f(1)f(6) &&= \tfrac{6}{36} \qquad (37) \\
f_Z(8) &= \textstyle\sum_{i=2}^{6} f(8-i)f(i) &&= f(6)f(2)+\ldots+f(2)f(6) &&= \tfrac{5}{36} \\
f_Z(9) &= \textstyle\sum_{i=3}^{6} f(9-i)f(i) &&= f(6)f(3)+\ldots+f(3)f(6) &&= \tfrac{4}{36} \\
f_Z(10) &= \textstyle\sum_{i=4}^{6} f(10-i)f(i) &&= f(6)f(4)+f(5)f(5)+f(4)f(6) &&= \tfrac{3}{36} \\
f_Z(11) &= \textstyle\sum_{i=5}^{6} f(11-i)f(i) &&= f(6)f(5)+f(5)f(6) &&= \tfrac{2}{36} \\
f_Z(12) &= \textstyle\sum_{i=6}^{6} f(12-i)f(i) &&= f(6)f(6) &&= \tfrac{1}{36}
\end{aligned}$$

in Übereinstimmung mit der Wahrscheinlichkeitsfunktion f in der Fortsetzung von Beispiel 8. $\diamond$

Für abhängige Zg stimmt die Faltung der Dichtefunktionen nicht notwendig mit der Dichtefunktion ihrer Summe überein.

Beispiel 29 Wie erhält man den Binomialkoeffizent durch Faltung?

Zu der auf Seite 41 definierten Indikatorfunktion X_i gehört die Wahrscheinlichkeitsfunktion

$$f(x) := \begin{cases} p & \text{falls } x = 1 \\ q & \text{falls } x = 0 \\ 0 & \text{sonst} \end{cases}$$

wobei $p + q = 1$. Daher ist zu erwarten, daß die n-malige Faltung von f mit sich selbst, an der Stelle $k \in \{0,\dots,n\}$ den Wert

$$\binom{n}{k} p^k q^{n-k}$$

besitzt. Setzt man rekursiv

$$f_{i+1}(x) := \sum_{k=0}^{1} f_i(x-k)f(k) = pf_i(x) + qf_i(x-1), \quad f_0 := f$$

so gilt tatsächlich

$$\binom{n}{k} = \frac{f_n(k)}{p^k q^{n-k}}.$$

Speziell für $p = q = \frac{1}{2}$ folgt

$$\binom{n}{k} = 2^n f_n(k).$$

$\diamond$

Falls X und Y stetig und unabhängig und sind, so ist $X+Y$ ebenfalls stetig und für die zugehörige Dichtefunktion h gilt

$$\boxed{h(x) = \int_{-\infty}^{+\infty} f(x-y)g(y)\mathrm{d}y, \quad x \in \mathbb{R}} \tag{38}$$

Dabei stellen f und g die Dichtefunktionen von X und Y dar. Die Funktion h heißt die **Faltung** von f und g und wird mit

$$f * g$$

bezeichnet. Vgl. [19] S.616.

Beispiel 30 Faltung der Gaußschen Glockenkurve mit sich selbst

Mittels (38) erhält man den Additionssatz für normalverteilte unabhängige Zg:

$$X_i \sim \mathscr{N}(\mu_i, \sigma_i^2),\ i = 1,2 \quad \Rightarrow$$

$$\begin{aligned}
h(x) &= \int_{-\infty}^{+\infty} \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\dfrac{((x-t)-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\dfrac{(t-\mu_2)^2}{2\sigma_2^2}} \mathrm{d}t \\
&= \frac{1}{\sigma_1\sigma_2 2\pi} \int_{-\infty}^{+\infty} e^{-\dfrac{((x-t)-\mu_1)^2}{2\sigma_1^2} - \dfrac{(t-\mu_2)^2}{2\sigma_2^2}} \mathrm{d}t \\
&= \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}\sqrt{2\pi}} \exp\left(-\frac{(x-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}\right)
\end{aligned}$$

$$\Rightarrow X_1 + X_2 \sim \mathscr{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2).$$

Da der Übergang von der vorletzten zur letzten Zeile viele Umformungen erfordert, verifizieren wir ihn nur für $\mu_1 = \mu_2 = 0$ und $\sigma_1 = \sigma_2 = 1$:

$$h(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-t)^2}{2} - \frac{t^2}{2} \right) \mathrm{d}t =$$

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left(\frac{-x^2 + 2xt - t^2 - t^2}{2} \right) \mathrm{d}t = \frac{1}{2\pi} e^{-x^2/2} \int_{-\infty}^{+\infty} \exp\left(xt - t^2 \right) \mathrm{d}t.$$

Wegen

$$xt - t^2 = x^2/4 - (t - x/2)^2$$

haben wir

$$h(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}} \int_{-\infty}^{+\infty} e^{(\frac{x^2}{4} - (t - \frac{x}{2})^2)} \mathrm{d}t = \frac{1}{2\pi} e^{-\frac{x^2}{4}} \underbrace{\int_{-\infty}^{+\infty} e^{-(t-x/2)^2} \mathrm{d}t}_{= \sqrt{\pi}} = \frac{1}{\sqrt{1+1}\sqrt{2\pi}} e^{-\frac{x^2}{2(1+1)}}$$

als Spezialfall von oben. $\diamond$

2.11.2 Berechnung der χ^2-Verteilungsfunktion

Wir berechnen die Dichte der χ^2-Verteilungsfunktion einerseits mittels Faltung und andererseits direkt.

Beispiel 31 Faltung der χ^2-Dichtefunktion mit sich selbst

In Abschnitt 2.10.5 über die χ^2-Verteilung wird behauptet, daß sie die Verteilung der Summe von Quadraten standardnormalverteilter unabhängiger Zg beschreibt. Wir verifizieren die Dichtefunktionen für die Freiheitsgrade $n = 1, 2, 3, 4$.

Als Spezialfall ergibt sich für $n = 1$ und $X \sim \mathscr{N}(0, 1)$ die Feststellung, daß X^2 der Verteilung

$$F_1(x) := \frac{1}{\sqrt{2}\underbrace{\Gamma(1/2)}_{\sqrt{\pi}}} \int_0^x t^{-1/2} e^{-t/2} \mathrm{d}t = \frac{1}{\sqrt{2\pi}} \int_0^x t^{-1/2} e^{-t/2} \mathrm{d}t$$

folgt. Wir überprüfen diese Behauptung. Offenbar gelten für nichtnegatives x aufgrund der Standardnormalverteilung von X die Gleichungen

$$\begin{aligned} P(X^2 \le x) = {} & P(-\sqrt{x} \le X \le \sqrt{x}) = \Phi_{0,1}(\sqrt{x}) - \Phi_{0,1}(-\sqrt{x}) \\ = {} & \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{t^2}{2}} \mathrm{d}t = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{t^2}{2}} \mathrm{d}t. \end{aligned}$$

Mit der Transformation $s = t^2$ folgt $\mathrm{d}s = 2t\mathrm{d}t$, sodaß man schließlich die gewünschte Gleichung

$$P(X^2 \leq x) = \frac{1}{\sqrt{2\pi}} \int_0^x s^{-1/2} e^{-\frac{s}{2}} \mathrm{d}s = F_1(x)$$

bekommt. Es ist zu beachten, daß die Dichtefunktion in 0 eine Polstelle besitzt.

Um nun die Dichtefunktion $f_2(x)$ der Verteilung von $X_1^2 + X_2^2$ zu bekommen, kann man die Dichtefunktion

$$f(x) := f_1(x) = \frac{1}{\sqrt{2\pi}} \begin{cases} x^{-1/2} \exp\left(-\dfrac{x}{2}\right), & \text{falls } 0 \leq x \\ 0, & \text{sonst} \end{cases}$$

von F_1 mit sich selbst falten:

$$(f * f)(x) = \int_{-\infty}^{+\infty} f(x-t) f(t) \mathrm{d}t.$$

Aus $f(x) = 0$ für $x \leq 0$ folgt

$$\begin{aligned} (f * f)(x) &= \int_0^x f(x-t) f(t) \mathrm{d}t \\ &= \frac{1}{2\pi} \int_0^x \frac{\exp(-(x-t)/2)}{\sqrt{x-t}} \frac{\exp(-t/2)}{\sqrt{t}} \mathrm{d}t \\ &= \frac{\exp(-x/2)}{2\pi} \underbrace{\int_0^x \frac{\mathrm{d}t}{\sqrt{(x-t)t}}}_{=\pi} = \frac{1}{2} e^{-x/2} = f_2(x). \end{aligned}$$

Die Ersetzung des Integrals durch π wird durch

$$\int_0^x \frac{\mathrm{d}t}{\sqrt{(x-t)t}} = \arcsin\left(\frac{t - x/2}{x/2}\right)\Bigg|_0^x = \arcsin(1) - \arcsin(-1) = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi$$

gerechtfertigt.

So fortfahrend bekommt man für nichtnegative $x \in \mathbb{R}$

$$\begin{aligned} f_3(x) = f_2(x) * f_1(x) &= \int_{-\infty}^{+\infty} f_2(x-t) f_1(t) \mathrm{d}t = \frac{1}{2} \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x-t}{2}} t^{-1/2} e^{-\frac{t}{2}} \mathrm{d}t \\ &= \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \int_0^x t^{-1/2} \mathrm{d}t = \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} 2t^{\frac{1}{2}} \Big|_0^x = \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} x^{\frac{1}{2}} \end{aligned}$$

in Übereinstimmung mit der Definition von $f_3(x)$ in Abschnitt 2.10.5 und

$$f_4(x) = f_2(x) * f_2(x) = \textstyle\int_{-\infty}^{+\infty} f_2(x-t) f_2(t) \mathrm{d}t = \frac{1}{4} \int_0^x e^{-\frac{x-t}{2}} e^{-\frac{t}{2}} \mathrm{d}t = \frac{1}{4} e^{-\frac{x}{2}} \int_0^x \mathrm{d}t = \frac{x}{4} e^{-\frac{x}{2}}.$$

Allgemein kann man

$$f_{n+1}(x) = f_n(x, n) * f_1(x)$$

zeigen, wobei $f_n(x)$ definiert ist wie in Abschnitt 2.10.5. ◇

Beispiel 32 Direkte Berechnung der χ^2-Dichtefunktion für $n = 2$ und $n = 3$
Unabhängig vom Faltungsgedanken kann man auch direkt vorgehen. Falls

$$g(s,t)$$

die Dichtefunktion der Zg

$$\chi^2 := X_1^2 + X_2^2$$

darstellt, so ist die Wahrscheinlichkeit

$$P(\chi^2 \le x)$$

gleich zur Wahrscheinlichkeit W dafür, daß die Werte von χ^2 in der Kreisscheibe

$$B_r(0) := \{[x,y]^{\mathrm{T}} \in \mathbb{R}^2 : x^2 + y^2 \le r^2\}, \quad r := \sqrt{x}$$

liegen. W kann man durch Integration von g über $B_{\sqrt{x}}(0)$ berechnen, d.h. durch

$$P(\chi^2 \le x) = \int_{B_{\sqrt{x}}(0)} g(s,t)\mathrm{d}s\mathrm{d}t.$$

Da X_1 und X_2 beide standardnormalverteilt und unabhängig sind, ist die Dichtefunktion $g(s,t)$ durch das Produkt

$$g(s,t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = \frac{1}{2\pi} e^{-\frac{s^2+t^2}{2}}$$

gegeben, sodaß

$$P(\chi^2 \le x) = \frac{1}{2\pi} \int_{B_{\sqrt{x}}(0)} e^{-\frac{s^2+t^2}{2}} \mathrm{d}s\mathrm{d}t.$$

Der Übergang zu Polarkoordinaten

$$s(\varrho,\varphi) = \varrho\cos(\varphi), \quad t(\varrho,\varphi) = \varrho\sin(\varphi)$$

wie in der Fußnote auf S.75 führt auf

$$P(\chi^2 \le x) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\sqrt{x}} e^{-\frac{\varrho^2}{2}} \varrho \mathrm{d}\varrho \mathrm{d}\varphi = \int_0^{\sqrt{x}} e^{-\frac{\varrho^2}{2}} \varrho \mathrm{d}\varrho = -e^{-\frac{\varrho^2}{2}} \Big|_0^{\sqrt{x}} = 1 - e^{-\frac{x}{2}}$$

d.h.

$$F_2(x) = \begin{cases} 0 & \text{falls } x < 0, \\ 1 - e^{-\frac{x}{2}} & \text{falls } 0 \le x. \end{cases}$$

Offenbar stimmt die Ableitung von $F_2(x)$ mit $f_2(x)$ wie gewünscht überein.

Für die Summe dreier Quadrate standardnormalverteilter unabhängiger Zg

$$\chi^2 := X_1^2 + X_2^2 + X_3^2$$

erfüllt die Dichtefunktion $g(s,t,u)$ die Gleichung

$$g(s,t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}} = \frac{1}{(2\pi)^{\frac{3}{2}}}e^{-\frac{s^2+t^2+u^2}{2}}$$

sodaß

$$P(\chi^2 \le x) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int_{B_{\sqrt{x}}(0)} e^{-\frac{s^2+t^2+u^2}{2}}\mathrm{d}s\mathrm{d}t\mathrm{d}u.$$

Der Übergang zu Polarkoordinaten ([20] S.356)

$$s(\varrho,\varphi,\psi) = \varrho\cos(\varphi)\cos(\psi), \quad t(\varrho,\varphi,\psi) = \varrho\sin(\varphi)\cos(\psi), \quad u(\varrho,\varphi,\psi) = \varrho\sin(\psi)$$

führt auf

$$P(\chi^2 \le x) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int_0^{\sqrt{x}}\int_{-\pi/2}^{+\pi/2}\int_{-\pi}^{+\pi} e^{-\frac{\varrho^2}{2}}\varrho^2\cos(\psi)\mathrm{d}\varphi\mathrm{d}\psi\mathrm{d}\varrho$$

$$= \frac{1}{(2\pi)^{\frac{3}{2}}}\int_{-\pi/2}^{+\pi/2}\cos(\psi)\mathrm{d}\psi\int_{-\pi}^{+\pi}\mathrm{d}\varphi\int_0^{\sqrt{x}}\varrho^2 e^{-\frac{\varrho^2}{2}}\mathrm{d}\varrho = \frac{4\pi}{(2\pi)^{\frac{3}{2}}}\int_0^{\sqrt{x}}\varrho^2 e^{-\frac{\varrho^2}{2}}\mathrm{d}\varrho.$$

Schließlich führt die Variablentransformation

$$\varrho^2 = v,\ \mathrm{d}v = 2\varrho\mathrm{d}\varrho = 2v^{\frac{1}{2}}\mathrm{d}\varrho$$

auf

$$P(\chi^2 \le x) = \frac{2}{(2\pi)^{\frac{1}{2}}}\int_0^x ve^{-\frac{v}{2}}\frac{1}{2v^{\frac{1}{2}}}\mathrm{d}v = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_0^x v^{\frac{1}{2}}e^{-\frac{v}{2}}\mathrm{d}v$$

d.h. $f_3(x) = x^{\frac{1}{2}}e^{-\frac{x}{2}}$. ◇

2.11.3 Zentraler Grenzwertsatz

[1] S.237, [13] S.430, [15] S.127, [18] S.356, [20] S.1051

Der Additionssatz für normalverteilte unabhängige Zg stellt fest, daß ihre Summe wieder normalverteilt ist. Es zeigt sich, daß näherungsweise diese Verteilungsverhältnisse auch gelten, wenn die einzelnen Summanden nicht normalverteilt sind. Das ist Gegenstand des **zentralen Grenzwertsatzes**.

Sei $(X_k)_{k=1}^\infty$ eine Folge identisch verteilter und unabhängiger Zufallsgrößen über demselben Ereignisraum Ω mit dem Erwartungswert μ und der Varianz σ^2. Bildet man die normierte Partialsummenfolge

$$Z_n := \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma}$$

so ist deren Grenzwert

$$\lim_{n\to\infty} Z_n$$

standardnormalverteilt.

Man beachte, daß nicht verlangt wird, daß alle X_k normalverteilt sind, wesentlich ist nur, daß

$$F_{X_k} = F$$

für alle $k \in \mathbb{N}$ und Unabhängigkeit gilt.

Fortsetzung Beispiel 28, Würfeln mit drei Würfeln, Dreiecksverteilung

Dort hatten wir die Wahrscheinlichkeitsfunktion für die Summe der Augenanzahlen beim Würfeln mit zwei Würfeln durch Faltung berechnet. Die roten Punkte in Abbildung 12 zeigen den Werteverlauf von

$$f_{X_1+X_2} := f * f.$$

Wir gehen nun zu drei Würfeln über und beobachten die Wahrscheinlichkeitsfunktion

$$f_{X_1+X_2+X_3}(x) := \sum_{k=1}^{6} f_{X_1+X_2}(x-k) f(k)$$

für die Summe ihrer Augenanzahlen. Der Wertverlauf von $f_{X_1+X_2+X_3}$ wird durch die blauen Punkte visualisiert. Offenbar gilt $EX_i = 3.5$. Daher gilt

$$E(X_1 + X_2 + X_3) = 10.5.$$

Weiter vorn hatten wir in (18) berechnet, daß $\mathrm{Var}(X_i) = \frac{35}{12}$. Da die Zg X_i unabhängig voneinander sind, kann auf

$$\mathrm{Var}(X_1 + X_2 + X_3) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) = \frac{35}{4}$$

geschlossen werden. Setzt man daher

$$\mu := 10.5, \quad \sigma := \sqrt{\frac{35}{4}}$$

so erwartet man entsprechend des zentralen Grenzwertsatzes eine Nähe zwischen den Werten von $f_{X_1+X_2+X_3}$ und $\varphi_{\mu,\sigma}$. Die grünen Kreise in Abbildung 12 stellen die Menge

$$\{(k, \varphi_{\mu,\sigma}(k)) : k = 2, \ldots, 18\}$$

dar. Offensichtlich gibt es eine annähernde Übereinstimmung mit

$$\{(k, f_{X_1+X_2+X_3}(k)) : k = 2, \ldots, 18\}.$$

◇

Das folgende Beispiel illustriert diese Verhältnisse für stetige Zg. Dazu falten wir die Rechteckverteilung (29) zweimal mit sich selbst und beobachten, daß das Faltungsresultat schon für $n = 3$ einer Gaußschen Glockenkurve sehr ähnelt.

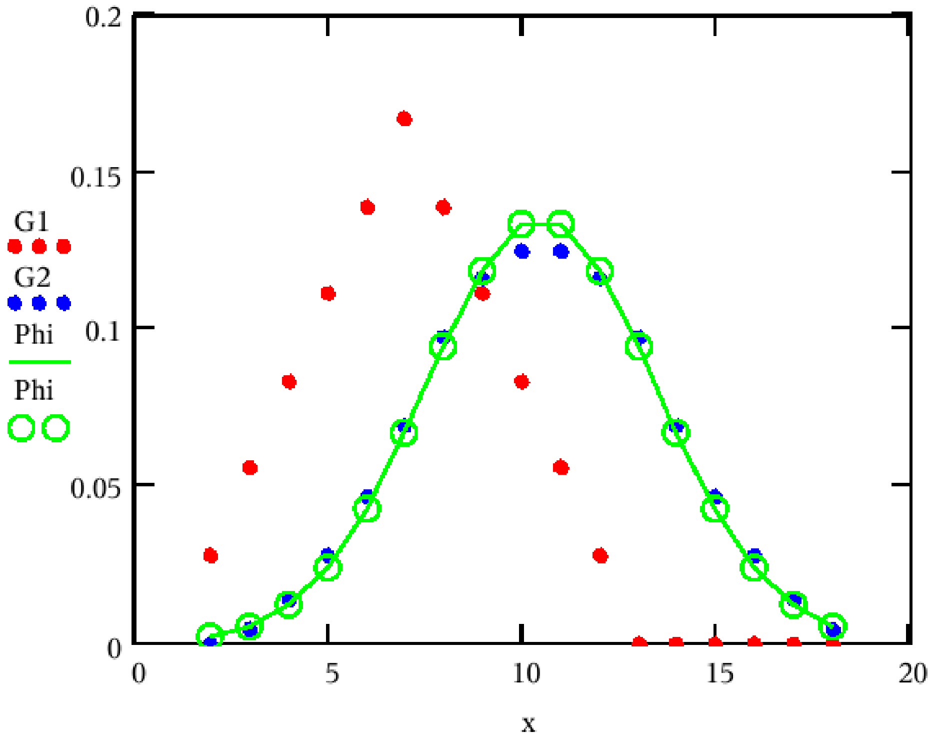


Abbildung 12: Die roten Punkte korrespondieren mit $f_{X_1+X_2}$, die blauen Punkte mit $f_{X_1+X_2+X_3}$ und die grünen Kreise mit $\varphi_{\mu,\sigma}$ für $\mu := 10.5$ und $\sigma = \sqrt{\frac{35}{4}}$. Offensichtlich bilden die roten Punkte mit der x-Achse ein Dreieck. Deswegen heißt die zu $f_{X_1+X_2}$ gehörende Verteilungsfunktion auch eine Dreiecksverteilung. Es zeigt sich, daß im allgemeinen die Summe zweier gleichverteilter Zg dreiecksverteilt ist. Die kontinuierliche Version zeigen die Abbildungen 13 und 14.

Beispiel 33 Faltung von Rechteckverteilungsdichte mit sich selbst, Approximation der Gaußschen Glockenkurve durch zusammengesetzte Parabeln

Wir betrachten die Verteilungsfunktion der Summe

$$Z := X_1 + X_2 + X_3$$

wobei die X_i unabhängig und ihre drei Verteilungsfunktionen gleich sein sollen. Der Einfachheit halber wählen wir Rechteckverteilungen:

$$F_{X_i}(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t, \quad f(t) = \begin{cases} 1, & \text{falls } t \in [0,1], \\ 0, & \text{falls } t \in \mathbb{R} \setminus [0,1]. \end{cases}$$

Entsprechend Abschnitt 2.10.1 gilt

$$EX_i = \frac{1}{2}, \quad \mathrm{Var}(X_i) = \frac{1}{12}$$

sodaß entsprechend (14) und (16) für die Erwartung und die Varianz von Z die Gleichungen

$$EZ = E(X_1 + X_2 + X_3) = \frac{3}{2}, \quad \mathrm{Var}(Z) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) = \frac{1}{4}$$

gelten und man erwartet, daß die Verteilungsfunktion von Z in etwa mit

$$\varphi_{\mu,\sigma}, \quad \mu := EZ = \frac{3}{2}, \quad \sigma := \sqrt{\mathrm{Var}(Z)} = \frac{1}{2}$$

übereinstimmt.

Die Dichtefunktion von $X_1 + X_2$ lautet

$$g(x) := \int_{-\infty}^{+\infty} f(x-t)f(t)\mathrm{d}t = \int_0^1 f(x-t)\mathrm{d}t.$$

Die Definition von f impliziert

$$f(x-t) = \begin{cases} 1 & \text{falls } t \in [x-1, x], \\ 0 & \text{sonst.} \end{cases}$$

Daher wird der Integrand von $\int_0^1 f(x-t)\mathrm{d}t$ nur für

$$t \in I(x) := [0,1] \cap [x-1, x].$$

verschieden von Null, sodaß

$$g(x) = \int_{I(x)} \mathrm{d}t = \text{Länge von } I(x)$$

was folgende Fallunterscheidung nahelegt:

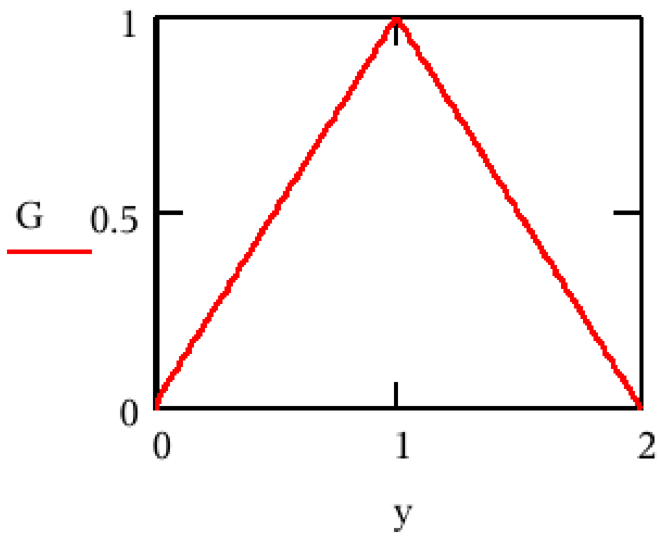


Abbildung 13: G stellt die Dichte g von $X_1 + X_2$ über $[0, 2]$ dar. Aufgrund ihrer Gestalt wird die zugehörige Verteilungsfunktion auch Dreiecksverteilung genannt. Vergleicht man mit dem Werteverlauf von f_Z in (37), so erkennt man in f_Z die diskrete Variante von g.

- Für negatives x ist $I(x)$ leer, sodaß $g(x) = 0$.
- Für $x \in [0, 1]$ bekommt man $I(x) = [0, x]$, sodaß $g(x) = x$.
- Für $x \in [1, 2]$ bekommt man $I(x) = [x - 1, 1]$, sodaß $g(x) = 1 - (x - 1) = 2 - x$.
- Schließlich ist für $2 \leq x$ das Intervall $I(x)$ wieder leer, sodaß $g(x) = 0$.

Zusammenfassend kann man schreiben:

$$g(x) = \begin{cases} 0 & \text{falls } x \leq 0, \\ x & \text{falls } 0 \leq x \leq 1, \\ 2 - x & \text{falls } 1 \leq x \leq 2, \\ 0 & \text{falls } 2 \leq x. \end{cases} \tag{39}$$

Schließlich stellt

$$h(x) := \int_0^1 g(x - t) \mathrm{d}t$$

die Dichtefunktion von Z dar. Die Auswertung liefert unter Zuhilfenahme von (39) für h

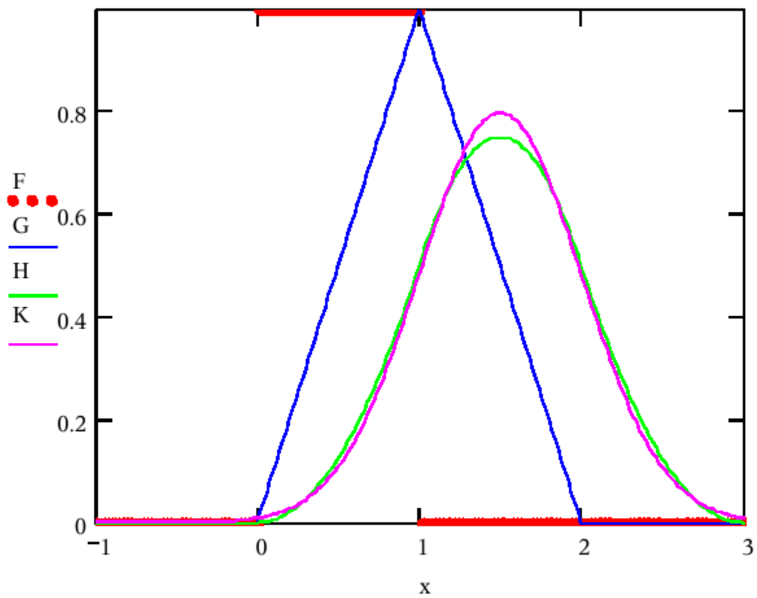


Abbildung 14: F=Graph von f, G=Graph von g, H=Graph von h, K=Graph von $\varphi_{\mu,\sigma}$ für $\mu = 3/2$ und $\sigma = 1/2$. Offenbar bestehen zwischen H und K nur geringe Unterschiede.

die Darstellung

$$h(x) = \begin{cases} 0 & \text{falls } x \leq 0, \\ \dfrac{x^2}{2} & \text{falls } 0 \leq x \leq 1, \\ -x^2 + 3x - \dfrac{3}{2} & \text{falls } 1 \leq x \leq 2, \\ \dfrac{x^2}{2} - 3x + \dfrac{9}{2} & \text{falls } 2 \leq x \leq 3, \\ 0 & \text{falls } 3 \leq x, \end{cases}$$

d.h. h ist verschieden von 0 nur in $[0, 3]$ und setzt sich dort stetig aus Parabeln zusammen. In Abbildung 14 wird diese Zusammensetzung mit $\varphi_{\mu,\sigma}$ für die Parameter $\mu = 3/2$ und $\sigma = 1/2$ verglichen. Ein ähnliches Beispiel ist in [18] auf S.357 zu finden. ◇

2.11.4 Satz von de Moivre-Laplace

[15] S.129, [18] S.358, [20] S.1054

Der Satz von de Moivre-Laplace bringt die Binomial- mit der Normalverteilung in Zusammenhang (Abraham de Moivre 1667-1754, Pierre Simon Laplace 1749-1827). Er stellt einen Spezialfall des zentralen Grenzwertsatzes dar.

Entsprechend (21) läßt sich eine binomialverteilte Zg X durch die Summe

$$X = \sum_{i=1}^{n} X_i$$

der Indikatorfunktionen X_i darstellen. Da diese unabhängig und identisch verteilt sind, muß als Konsequenz aus dem zentralen Grenzwertsatz X für große n annähernd normalverteilt sein. Genauer, falls

$$X \sim \mathscr{B}(n, p)$$

so gilt für große n

$$P(X \leq x) \approx \Phi_{\mu,\sigma}(x), \quad \mu := np, \quad \sigma := \sqrt{np(1-p)}.$$

Beispiel 34 Vergleich der Wahrscheinlichkeitsfunktion von $f(k, n, p)$ von $X \sim \mathscr{B}(n, p)$ mit $\varphi_{\mu,\sigma}(k)$

Für $n = 15$ und $p = 0.7$ setzen wir

$$\mu := np,\ \sigma := \sqrt{np(1-p)},\ F := \{(k, f_k) : k = 0, \ldots, n\}\,,\ f_k := f(k, n, p)$$

wobei wegen $X \sim \mathscr{B}(n, p)$ für die zugehörige Dichtefunktion

$$f(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

gilt. In Abbildung 15 wird die Punktmenge F mit dem Graphen G von $\varphi_{\mu,\sigma}$ über $[0, 15]$ verglichen. Offensichtlich liegen die Punkte von F fast auf G. ◇

Fortsetzung Beispiel 14. Dort hatten wir $\mathscr{B}(n, p)$ für $n = 5000$ und $p = 0.001$ mit $\mathscr{P}(\lambda)$ für $\lambda = 5$ verglichen. Wegen

$$P(A) = P(X \geq 2) = 1 - P(X < 2)$$

bekommen wir mit dem Satz von de Moivre-Laplace

$$P(A) \approx 1 - \Phi_{\mu,\sigma}(2), \quad \mu := 5, \quad \sigma = \sqrt{5 \times 0.999} = 2.235$$

d.h. $P(A) = 1 - 0.09 = 0.91$, was keine befriedigende Approximation des tatsächlichen Wertes von 0.96 darstellt.

Die Approximation der Verteilungsfunktion einer nach $\mathscr{B}(n, p)$ verteilten Zg durch $\Phi_{\mu,\sigma}$, $\mu := np$, $\sigma := \sqrt{np(1-p)}$ ist beim Erfülltsein der Ungleichung

$$\boxed{np(1-p) > 9}$$

dagegen akzeptabel [15] S.129.

Im Zusammenhang mit einem Testverfahren für einen Anteilswert macht sich noch folgende Betrachtung notwendig. Wir setzen

$$\phi_{\alpha,n}(p) := \Phi^{-1}_{np,\sqrt{np(1-p)}}(\alpha)$$

und fixieren $(\alpha, n) \in [0, 1] \times \mathbb{N}$. Dann hängt $\phi_{\alpha,n}$ nur noch von $p \in (0, 1)$ ab, sodaß man leicht Feststellungen hinsichtlich seines Monotonieverhalten machen kann.

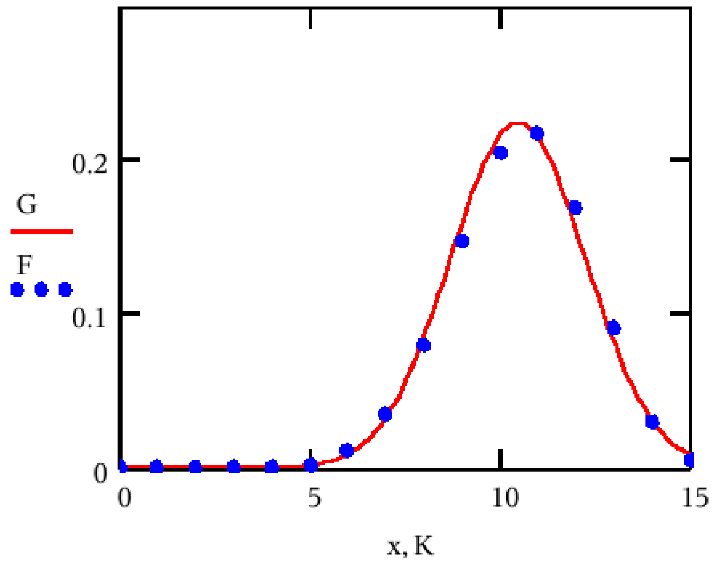


Abbildung 15: G (rote Linie) ist der Graph von $\varphi_{\mu,\sigma}$ über $[0, 15]$ für $\mu = 10.5$ und $\sigma = \sqrt{3.15}$. F (blaue Punkte) stellt die Menge $\{(0, f_0), \ldots, (15, f_{15})\}$ aus Beispiel 34 dar.

2.11.5 Skizze zum Beweis des zentralen Grenzwertsatzes

Für den Beweis des zentralen Grenzwertsatzes bietet sich die Fouriertransformierte der Dichte- oder der Wahrscheinlichkeitsfunktion einer Zg X an. Im Bereich der Wahrscheinlichkeitsrechnung werden die Resultate der Fouriertransformation als **charakteristische Funktionen** bezeichnet.

2.11.6 Berechnung der Wahrscheinlichkeitsfunktion der Summe zweier diskreter Zg ohne Faltung

Wie wir bald sehen werden, ist man nicht nur an der Verteilung der Summe zweier Zg $X_1 + X_2$ interessiert, sondern auch an der Verteilung ihres Quotienten $Z := X_1/X_2$. Zu diesem Zweck machen wir uns von dem Faltungsgedanken frei und berechnen die Wahrscheinlichkeitsfunktion der Summe $X_1 + X_2$ mit einer Methode, die auch für den Quotienten erfolgreich ist. Offenbar reicht es im diskreten Fall zur Bestimmung von

$$P(X_1 + X_2) = z$$

aus, die Wahrscheinlichkeiten aufzuaddieren, mit denen $X_1 + X_2$ einen Wert auf der Geraden $z = x + y$ annimmt. Die roten Punkte in Abbildung 16 illustrieren den Ereignisraum $\Omega := \{1, 2, 3, 4, 5, 6\}^2$ und die Gerade bestehend aus schwarzen Kreisen die Lösungen der Gleichung $y + x = z$ für $z = 5$. Offenbar liegen genau 4 Elementarereignisse auf dieser Geraden, deshalb gilt $P(X_1 + X_2 = 5) = \frac{4}{36}$. Läßt man daher z innerhalb von 2 bis 12

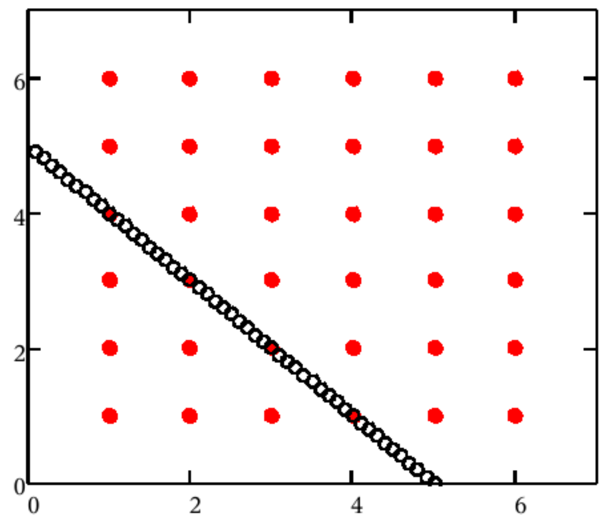


Abbildung 16: Es wird gezeigt für welche Elemantarereignisse die Zg 'Summe der Augenanzahlen' den Wert 5 annimmt.

variieren, so erhält man durch Auszählen bis auf den Faktor $\frac{1}{36}$ den Werteverlauf der Dreiecksverteilungsdichte (37).

Diese Herangehensweise läßt sich auch noch für drei Dimensionen illustrieren.

Beispiel 35 Würfeln mit drei Tetraedern

Wir überprüfen unser Vorgehen anhand des Würfelns mit drei Tetradern (Vieflächner), wobei die verdeckten Augenanzahlen aufaddiert werden. Es sei

$$Z : \Omega := \{1, 2, 3, 4\}^3 \mapsto \{3, \ldots, 12\}$$

durch die Summe der verdecktliegenden Augenanzahlen erklärt. Offensichtlich besteht Ω aus $4^3 = 64$ Elementarereignissen, wobei jedem die gleiche Eintrittswahrscheinlichkeit von $\frac{1}{64}$ zugestanden wird. Durch Faltung gewinnt man

$$f_Z(z) = \frac{1}{4}\sum_{k=1}^{4} f_1(z-k),\ f_1(z) = \frac{1}{4}\sum_{k=1}^{4} f_0(z-k),\ f_0(z) := \frac{1}{4}\begin{cases}1 & \text{falls } k \in \{1, \ldots, 4\},\\ 0 & \text{falls } k \notin \{1, \ldots, 4\}.\end{cases}$$

Die Auswertung liefert

k	3	4	5	6	7	8	9	10	11	12
$f_Z(k)$	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{6}{64}$	$\frac{10}{64}$	$\frac{12}{64}$	$\frac{12}{64}$	$\frac{10}{64}$	$\frac{6}{64}$	$\frac{3}{64}$	$\frac{1}{64}$

Das Zählen der roten Kugeln in Abbildung 17, die in den Ebenen E_k liegen, liefert das gleiche Ergebnis. $\diamond$

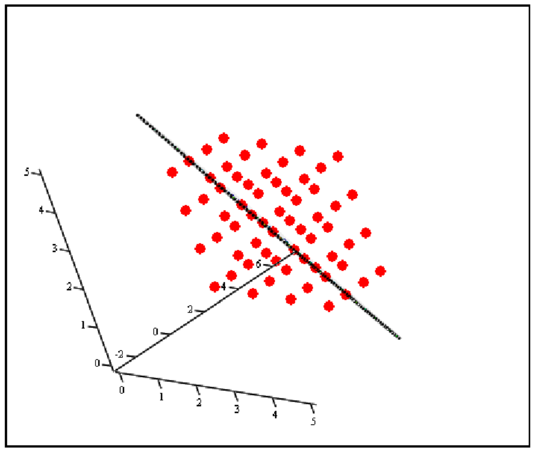


Abbildung 17: Der aus roten Kugeln bestehende Quader visualisiert Ω aus Beispiel 35. Das Koordinatensystem wurde so gedreht, daß keine der roten Kugeln durch eine andere verdeckt wird, d.h. es sind alle 64 Elementarereignisse erkennbar. Die schwarze Gerade liegt in der Ebene E_7, wobei $E_k = \{[x, y, z]^{\mathrm{T}} \in \mathbb{R}^3 : x + y + z = k\}$. Offenbar liegen 12 Kugeln in E_7. Daher gilt $P(Z = 7) = 12/64$.

2.12 Verteilung des Quotienten zweier unabhängiger Zufallsgrößen

Im Abschnitt 2.10.6 über die STUDENTsche Verteilung hatten wir den Quotienten

$$T := \frac{X}{\sqrt{Y}}\sqrt{n}$$

betrachtet, wobei $X \sim \mathcal{N}(0, 1)$ und $Y \sim \chi^2(n)$. In Abschnitt 2.10.7 ist man ebenfalls an der Verteilung eines Quotienten interessiert. Es stellt sich daher die Frage, wie man die Dichtefunktion von

$$Z := \frac{X_1}{X_2}$$

berechnen kann, falls man die Dichtefunktionen f_{X_1} und f_{X_2} kennt.

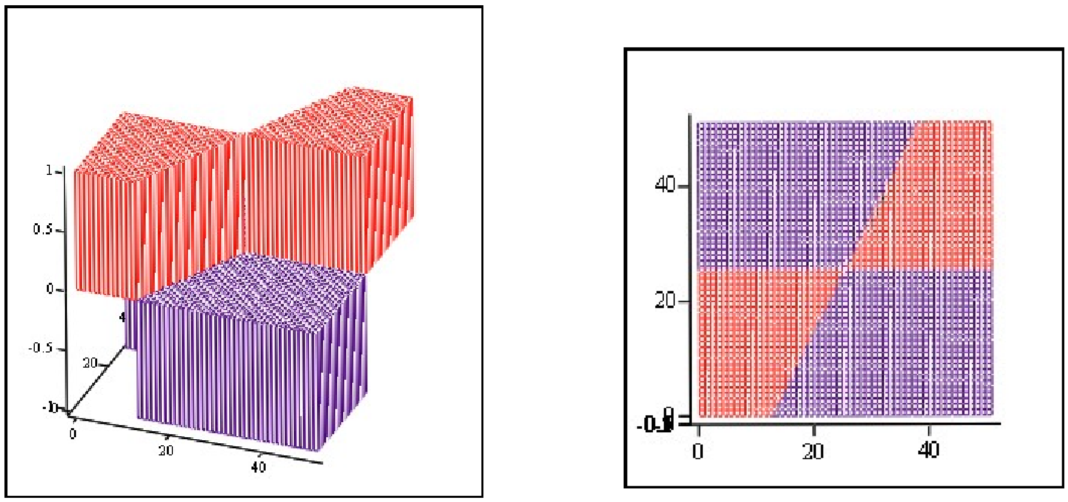
2.12.1 Allgemeiner Fall

Offenbar hat man zur Berechnung von

$$F_Z(z) := P\left(Z \leq z\right), \quad Z := \frac{X_1}{X_2}, \quad X_i : \Omega_i \mapsto \mathbb{R}$$

die Dichtefunktion $f : \mathbb{R}^2 \mapsto \mathbb{R}_+$ des Zufallsvektors

$$\left[\begin{array}{c} X_1(\omega_1) \\ X_2(\omega_2) \end{array} \right] : \Omega_1 \times \Omega_2 \mapsto \mathbb{R}^2$$



M

M

Abbildung 18: Im linken Teil wird der Werteverlauf von $g(x, y, z)$ für $x, y \in [0, 1]$ für den Parameterwert $z := \frac{1}{2}$ gezeigt. Im rechten Teil wird die Projektion in die (x, y)-Ebene gezeigt. Wegen $z := \frac{1}{2}$ besitzt die Grenzlinie, die der Graph von $y = \frac{1}{z}x$ ist, zwischen rotem und blauem Bereich den Anstieg 2.

über den Bereich B von $\mathbb{R}^2$ zu integrieren, in dem $Z \leq z$ gilt. Setzt man

$$g(x, y, z) := \begin{cases} \dfrac{\frac{x}{y} - z}{\left|\frac{x}{y} - z\right|} & \text{falls } x \neq yz \\ \\ 0 & \text{falls } x = yz \end{cases}$$

so gilt

$$g(x, y, z) = \begin{cases} -1 & \text{falls } \frac{x}{y} < z \\ 0 & \text{falls } \frac{x}{y} = z \\ 1 & \text{falls } \frac{x}{y} > z \end{cases}$$

Daher ist der Bereich B durch das Urbild von -1 vermöge g gegeben. Die Abbildung 18 zeigt den Werteverlauf von g für durch $1/2$ fixiertes z. Offensichtlich wird g im roten Bereich gleich 1 und im blauen Bereich gleich -1. An der Grenze beider Bereiche nimmt g den Wert Null an. Daher gilt

$$F_Z(z) = \int_B f(x, y)\mathrm{d}x\mathrm{d}y$$

wobei B das Urbild von -1 vermöge g darstellt. Für $z = 1/2$ stimmt B mit dem blauen Bereich im rechten Teil der Abbildung 18 überein. Der blaue und der rote Bereich werden durch die x-Achse und den Graphen der Geraden $y = \frac{1}{z}x$ voneinander getrennt.

Bezeichnet B_1 den Teil von B in der oberen Halbebene und B_2 den Teil von B in der unteren Halbebene, so läßt sich $\int_B f(x,y)\mathrm{d}x\mathrm{d}y$ zerlegen in

$$\int_B f(x,y)\mathrm{d}x\mathrm{d}y = \underbrace{\int_{B_1} f(x,y)\mathrm{d}x\mathrm{d}y}_{=:\, I_1(z)} + \underbrace{\int_{B_2} f(x,y)\mathrm{d}x\mathrm{d}y}_{=:\, I_2(z)}.$$

Offenbar gilt nun

$$I_1(z) = \int_0^{+\infty}\int_{-\infty}^{yz} f(x,y)\mathrm{d}x\mathrm{d}y, \quad I_2(z) = \int_{-\infty}^{0}\int_{yz}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y.$$

Die Variablentransformation $u := \frac{x}{y}$ liefert

$$\mathrm{d}u = \frac{1}{y}\mathrm{d}x, \quad x \in [-\infty, yz] \Leftrightarrow u \in [-\infty, z], \quad x \in [yz, +\infty] \Leftrightarrow u \in [z, +\infty]$$

sodaß

$$I_1(z) = \int_0^{+\infty}\int_{-\infty}^{z} yf(uy,y)\mathrm{d}u\mathrm{d}y, \quad I_2(z) = \int_{-\infty}^{0}\int_{z}^{+\infty} yf(uy,y)\mathrm{d}u\mathrm{d}y.$$

Differentiation nach z liefert

$$\frac{\mathrm{d}}{\mathrm{d}z}I_1(z) = \int_0^{+\infty} yf(zy,y)\mathrm{d}y, \quad \frac{\mathrm{d}}{\mathrm{d}z}I_2(z) = -\int_{-\infty}^{0} yf(zy,y)\mathrm{d}u\mathrm{d}y$$

sodaß

$$\frac{\mathrm{d}}{\mathrm{d}z}F(z) = \int_0^{+\infty} yf(zy,y)\mathrm{d}y - \int_{-\infty}^{0} yf(zy,y)\mathrm{d}u\mathrm{d}y = \int_{-\infty}^{+\infty} |y|f(zy,y)\mathrm{d}y.$$

Demnach ist die Dichtefunktion des Quotienten Z durch

$$f_Z(z) := \int_{-\infty}^{+\infty} |y|f(zy,y)\mathrm{d}y$$

gegeben in Übereinstimmung mit [6] S.85 und [9] S.144.

Falls X_1 und X_2 unabhängig sind, so haben wir wegen

$$f(zy,y) = f_{X_1}(zy)f_{X_2}(y)$$

die Gleichung

$$f_Z(z) := \int_{-\infty}^{+\infty} |y|f_{X_1}(zy)f_{X_2}(y)\mathrm{d}y. \tag{40}$$

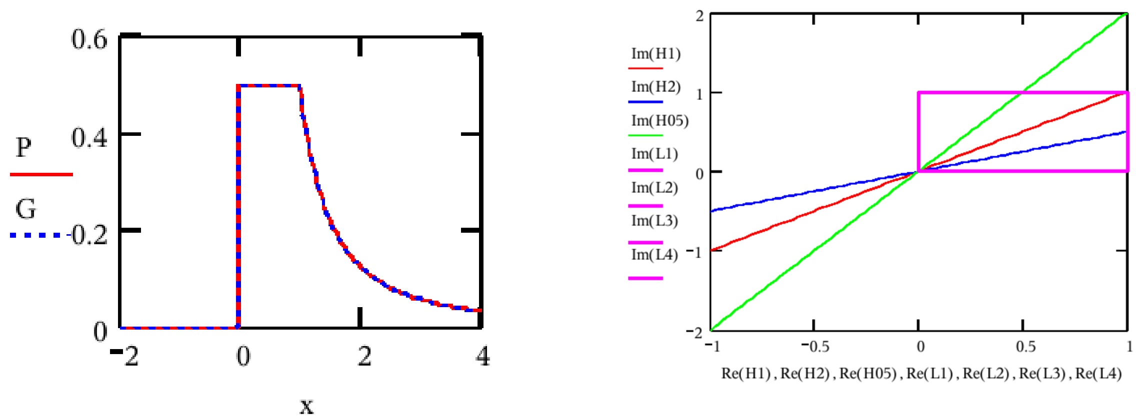


Abbildung 19: Im linken Teil wird der Graph der Dichtefunktion des Quotienten zweier in $[0,1]$ gleichverteilter Zg gezeigt. Im rechten Teil wird der Zusammenhang zu den Flächeninhalten von Abschnitten des Einheitsquadrates hergestellt.

Beispiel 36 Quotient zweier gleichverteilter Zg

Es seien X_1 und X_2 beide entsprechend der Dichtefunktion

$$f(x) := \begin{cases} 1 & \text{falls } x \in [0,1] \\ 0 & \text{falls } x \notin [0,1] \end{cases}$$

verteilt. Wir berechnen die Dichtefunktion der Verteilung von $Z := \frac{X_1}{X_2}$. Den Ausführungen von oben folgend gilt

$$f_Z(z) = \int_{-\infty}^{+\infty} |y| f(zy) f(y) \mathrm{d}y.$$

Aufgrund der speziellen Gestalt von f ergibt sich die Vereinfachung

$$f_Z(z) = \int_0^1 y f(zy) \mathrm{d}y.$$

- Für $z < 0$ wird das Argument von $f(zy)$ negativ. Da f an solchen Stellen verschwindet, ist der Integrant überall Null, sodaß man

  $$f_Z(z) = 0$$

  bekommt.

- Für $0 \leq z < 1$ liegen alle Argumente von $f(zy)$ in $[0,1]$, sodaß wir für solche z die Gleichungen

  $$f_Z(z) = \int_0^1 y \mathrm{d}y = \left.\frac{y^2}{2}\right|_0^1 = \frac{1}{2}$$

  erhalten.

- Für $1 \leq z$, verschwindet der Integrant $f(zy)$ für alle y, die größer als $1/z$ sind. Daher gelten für solche z die Gleichungen

$$f_Z(z) = \int_0^{1/z} y \mathrm{d}y = \left. \frac{y^2}{2} \right|_0^{\frac{1}{z}} = \frac{1}{2z^2}.$$

Zusammenfassend können wir

$$f_Z(z) = \frac{1}{2} \begin{cases} 0 & \text{falls } z < 0 \\ 1 & \text{falls } 0 \leq z < 1 \\ z^{-2} & \text{falls } 1 \leq z \end{cases}$$

schreiben. Im linken Teil von Abbildung 19 wird der Graph von f_Z im Bereich $[-2, 4]$ gezeigt.

Wir überprüfen das Resultat durch direkte Anschauung. Die Dichtefunktion des Zufallsvektors $[X_1, X_2]^{\mathrm{T}} : \Omega^2 \mapsto \mathbb{R}^2$ lautet offenbar

$$g(x, y) = f(x)f(y) = \begin{cases} 1 & \text{falls } (x, y) \in Q \\ 0 & \text{falls } (x, y) \notin Q \end{cases}, \quad Q := [0, 1]^2.$$

Daher gilt

$$F_Z(z) = P(Z \leq z) = \int_{B \cap Q} \mathrm{d}x\mathrm{d}y$$

wobei B in Abhängigkeit von z wieder den blauen Bereich aus dem rechten Teil von Abbildung 18 bezeichnet. Demnach stimmt $F(z)$ mit dem Inhalt des Teils von B überein, der in Q liegt. Für $z \in \{\frac{1}{2}, 1.2\}$ werden im rechten Teil von Abbildung 19 diese Teilbereiche gezeigt. Das Quadrat Q wird durch das magendafarbene Quadrat illustriert. Für $z = \frac{1}{2}$ stimmt der Durchschnitt $B \cap Q$ mit dem Dreieck überein, welches durch die grüne Gerade von Q abgeschnitten wird. Offenbar schneidet die grüne Gerade die Oberkante von Q im Punkt $(z, 1)$. Daher gilt für $z \in (0, 1]$ die Gleichung

$$\int_{B \cap Q} \mathrm{d}x\mathrm{d}y = \frac{z}{2}.$$

Für $z \in [1, +\infty)$ stimmt $B \cap Q$ mit Q vermindert um den Abschnitt, den die blaue Gerade liefert, überein. Offenbar schneidet die blaue Gerade die rechte Kante von Q im Punkt $(1, \frac{1}{z})$. Daher gilt für solche z die Gleichung

$$\int_{B \cap Q} \mathrm{d}x\mathrm{d}y = 1 - \frac{1}{2z}.$$

Schließlich ist für negative z der Durchschnitt $B \cap Q$ leer, sodaß wir zusammenfassend

$$F_Z(z) = \begin{cases} 0 & \text{falls } z < 0 \\ \frac{z}{2} & \text{falls } 0 \leq z < 1 \\ 1 - \frac{1}{2z} & \text{falls } 1 \leq z \end{cases}$$

schreiben können. Die Ableitung von F_Z nach z liefert wie gewünscht die Funktion f_Z. $\diamond$

2.12.2 Berechnung der Studentschen Dichtefunktion für einen und zwei Freiheitsgrade

Die Student-Verteilung beschreibt die Verteilung des Quotienten

$$T := \frac{X}{Y}.$$

Dabei gilt $X \sim \mathscr{N}(0,1)$ und Y ist durch

$$Y := \sqrt{\frac{1}{n}\sum_{k=1}^{n} X_k^2}, \quad X_k \sim \mathscr{N}(0,1)$$

gegeben, wobei die Unabhängigkeit von $X_1, \ldots, X_n$ und die Unabhängigkeit von X und Y vorausgesetzt wird.

Entsprechend (40) ergibt sich die Dichtefunktion von T durch das Integral

$$f_T(z) = \int_{-\infty}^{+\infty} |y| f_X(zy) f_Y(y) \mathrm{d}y.$$

Da X standardnormalverteilt ist, gilt

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Hinsichtlich der Verteilung von Y/n hat man die rechte Seite von

$$P(X_1^2 + \ldots + X_n^2 \leq nx^2) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{B_{\sqrt{n}x}(0)} \exp\left(-\frac{1}{2}\sum_{k=1}^{n} x_k^2\right) \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

in Abhängigkeit von x und n auszuwerten, wobei wie üblich

$$B_r(0) := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \leq r^2\}$$

gesetzt wird. Im einfachsten Fall erhält man für $n = 1$ und $0 \leq x$

$$P(X_1^2 \leq x^2) = P(-x \leq X_1 \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \mathrm{d}t = 2\int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \mathrm{d}t.$$

Demnach gilt

$$f_Y(x) = \begin{cases} 0, & \text{falls } x \leq 0, \\ \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, & \text{falls } 0 < x. \end{cases}$$

Nach Voraussetzung gilt

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

sodaß

$$f_T(z) = \int_{-\infty}^{+\infty} |y| f_X(zy) f_Y(y) \mathrm{d}y = \int_0^{+\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(zy)^2} \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \mathrm{d}y$$

$$= \frac{1}{\pi} \int_0^{+\infty} y e^{-\frac{1}{2}(z^2+1)y^2} \mathrm{d}y.$$

Mit der Variablentransformation

$$u := \frac{1}{2}(z^2+1)y^2$$

bekommt man $\mathrm{d}u = (z^2+1)y\mathrm{d}y$, sodaß in Übereinstimmung mit (36)

$$f_T(z) = \frac{1}{\pi} \int_0^{+\infty} e^{-u} \frac{\mathrm{d}u}{1+z^2} = \frac{1}{\pi} \frac{1}{(1+z^2)}$$

folgt.

Im Fall $n = 2$ haben wir

$$P(X_1^2 + X_2^2 \leq 2x^2) = \frac{1}{2\pi} \int_{B_{\sqrt{2}x}(0)} e^{-\frac{1}{2}(x_1^2+x_2^2)} \mathrm{d}x_1 \mathrm{d}x_2.$$

Der Übergang zu Polarkoordinaten

$$x_1(\varrho, \varphi) = \varrho \cos(\varphi), \quad x_2(\varrho, \varphi) = \varrho \sin(\varphi)$$

wie in der Fußnote auf S.75 führt auf

$$P(Y \leq x) = P\left(\sqrt{\frac{1}{2}(X_1^2 + X_2^2)} \leq x\right) = P(X_1^2 + X_2^2 \leq 2x^2)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\sqrt{2}x} e^{-\frac{\varrho^2}{2}} \varrho \mathrm{d}\varrho \mathrm{d}\varphi = \int_0^{\sqrt{2}x} e^{-\frac{\varrho^2}{2}} \varrho \mathrm{d}\varrho = 1 - e^{-x^2}.$$

Daher gilt

$$f_Y(x) = \begin{cases} 0, & \text{falls } x \leq 0 \\ 2xe^{-x^2}, & \text{falls } 0 < x \end{cases}$$

woraus zusammen mit der Kenntnis von f_X für f_T die Gleichungen

$$\begin{aligned} f_T(z) = {} & \frac{1}{\sqrt{2\pi}} \int_0^{\infty} y e^{-\frac{1}{2}(zy)^2} 2y e^{-y^2} \mathrm{d}y \\ = {} & \frac{2}{\sqrt{2\pi}} \int_0^{\infty} y^2 e^{-\frac{1}{2}(zy)^2 - y^2} \mathrm{d}y \\ = {} & \frac{2}{\sqrt{2\pi}} \int_0^{\infty} y^2 e^{-(1+\frac{z^2}{2})y^2} \mathrm{d}y \end{aligned}$$

folgen. Mit der Transformation

$$u := \left(1 + \frac{z^2}{2}\right)^{\frac{1}{2}} y$$

bekommt man

$$\mathrm{d}u = \left(1 + \frac{z^2}{2}\right)^{\frac{1}{2}} \mathrm{d}y, \quad y^2 = u^2 \left(1 + \frac{z^2}{2}\right)^{-1}$$

sodaß

$$f_T(z) = \left(1 + \frac{z^2}{2}\right)^{-\frac{3}{2}} \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-u^2} u^2 \mathrm{d}u.$$

Schließlich folgt mit

$$\int_0^\infty e^{-u^2} u^2 \mathrm{d}u = \frac{\sqrt{\pi}}{4}$$

in Übereinstimmung mit (36) die Gleichung

$$f_T(z) = \frac{\sqrt{\pi}}{4} \frac{2}{\sqrt{2\pi}} \left(1 + \frac{z^2}{2}\right)^{-\frac{3}{2}} = \frac{1}{2\sqrt{2}} \left(1 + \frac{z^2}{2}\right)^{-\frac{3}{2}}.$$

2.12.3 Berechnung der Fisherschen Dichtefunktion

3 Schließende Statistik

3.1 Stichproben und Stichprobenfunktionen

Eine **Zufallsstichprobe** vom Umfang n ist eine Folge von unabhängigen und identisch verteilten Zufallsgrößen $X_1, \ldots, X_n$, wobei X_i dem i-ten Element der Stichprobe eine reelle Zahl zuordnet. Identische Verteilung bedeutet

$$F_{X_1} = \cdots = F_{X_n}$$

wobei F_{X_i} die Verteilungsfunktion von X_i bezeichnet.

3.2 Punktschätzungen mittels Maximum-Likelihood-Methode

[13] S.498

Punktschätzungen zielen darauf ab, einen in einer Verteilungsfunktion einer Zg X auftretenden Parameter θ wie

- p in der Binomialverteilung $\mathscr{B}(n, p)$
- λ in der Poisson-Verteilung $\mathscr{P}(\lambda)$

- μ und σ in der Normal-Verteilung $\mathscr{N}(\mu, \sigma^2)$

durch die Angabe eines numerischen Wertes zu schätzen.

Ein **Schätzer** für θ ist eine Funktion der Gestalt

$$g(X_1(\omega_1), \ldots, X_n(\omega_n)) : \Omega^n \mapsto \mathbb{R}$$

die wiefolgt arbeitet:

- Man wählt eine Stichprobe
$$\mathcal{S} := [\omega_k]_{k=1}^n \in \Omega^n$$
vom Umfang n.

- Der k-ten Komponente von $\mathcal{S}$ wird durch X_k ein numerischer Wert x_k zugeordnet.

- Die Fixierung der Argumente von g durch $\boldsymbol{x} := [x_k]_{k=1}^n$ liefert einen numerischen wert, der eine Schätzung für θ abgibt.

Hierbei sind $X_1, \ldots, X_n$ unabhängige Zg, die alle die gleiche Verteilung wie X besitzen. Insbesondere kann
$$g(X_1, \ldots, X_n)$$
selbst als Zg über Ω^n aufgefasst werden. Die Funktion g wird in diesem Zusammenhang **Stichprobenfunktion** genannt. Einfache Beispiele sind die Funktionen

$$g(x_1, \ldots, x_n) := \frac{1}{n}\sum_{k=1}^n x_k, \quad g(x_1, \ldots, x_n) := \sum_{k=1}^n x_k^2$$

die wir gleich wieder antreffen werden. Die korrespondierenden Zg lauten

$$g(X_1, \ldots, X_n) := \frac{1}{n}\sum_{k=1}^n X_k, \quad g(X_1, \ldots, X_n) = \sum_{k=1}^n X_k^2.$$

Einen Ansatz, geeignete Stichprobenfunktion g zu finden, bietet die Maximum-Likelihood-Methode (Likelihood = Wahrscheinlichkeit).

- Man setzt voraus, daß eine diskrete Zg X mit dem Wertevorrat $x_1, x_2, \ldots$ entsprechend der Wahrscheinlichkeitsfunktion f verteilt ist, d.h. die Wahrscheinlichkeitsfunktion f ordnet jedem x_k die Wahrscheinlichkeit zu, mit der X den Wert x_k beim Experimentieren annimmt.

- Man setzt voraus, daß die Gestalt von f bekannt ist, aber von einem unbekannten Parameter θ abhängt, was durch $f(., \theta)$ ausgedrückt werden soll.

- Man führt das Experiment n mal durch und gewinnt so eine 'Stichprobe'

  $$\boldsymbol{x} := [x_i]_{i=1}^n \in \mathbb{R}^n$$

  im Umfang n. Genauer, man wählt einen Stichprobenvektor $\mathcal{S} \in \Omega^n$, ordnet mittels X jeder Komponente ω_i von $\mathcal{S}$ einen numerischen Wert $x_i := X(\omega_i)$ zu und kommt so zu $\boldsymbol{x}$. Diese Zuordnung führt auf die Zg X_i, $i = 1, \ldots, n$, über Ω^n:

  $$X_i(\mathcal{S}) := X(\omega_i), \quad \mathcal{S} = [\omega_i]_{i=1}^n.$$

  Nach Konstruktion sind $X_1, \ldots, X_n$ identisch verteilt, d.h. jedes X_i ist so verteilt wie X.

- Man setzt voraus, daß diese n Experimente unabhängig voneinander sind. Das hat zur Folge, daß die Zg $X_1, \ldots, X_n$ unabhängig sind, sodaß die Wahrscheinlichkeit

  $$L(\theta, \boldsymbol{x}) := P(\{X_1 = x_1\} \cap \ldots \cap \{X_n = x_n\})$$

  für das Auftreten von $\boldsymbol{x}$ die Gleichung

  $$L(\theta, \boldsymbol{x}) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} f(x_i, \theta)$$

  erfüllt.

- Schließlich bestimmt man nach dem **Prinzip von Gauß** die Stichprobenfunktion g so, daß der durch g zugewiesene Parameter

  $$\hat{\theta} := g(\boldsymbol{x})$$

  der gezogenen Stichprobe maximale Wahrscheinlichkeit verleiht, d.h.

  $$\forall \theta \in I : L(\theta, \boldsymbol{x}) \leq L(g(\boldsymbol{x}), \boldsymbol{x}).$$

Hierbei bezeichnet I einen zulässigen Parameterbereich.

In berechnungstechnischer Hinsicht hat man das Maximum von L für fixiertes $\boldsymbol{x}$ über I zu bestimmen.

3.2.1 ML-Schätzer für p in $\mathscr{B}(n, p)$, Gesetz der großen Zahl von Bernoulli

Wir führen ein Bernouilli Experiment n mal aus und definieren die Zg X_i wie die Indikatorfunktion in Abschnitt 2.7, d.h. durch 1 falls das Experiment beim i-ten Mal gelingt und durch Null falls nicht. Die zugehörige vom zu schätzenden Parameter p abhängende Wahrscheinlichkeitsfunktion f ist für alle X_i gleich, nämlich

$$f(x, p) = \begin{cases} p, & \text{falls } x = 1, \\ 1 - p, & \text{falls } x = 0. \end{cases}$$

Der Ausgang des n-maligen Experimentierens liefert einen $(0,1)$-Vektor

$$\boldsymbol{x} := [x_i]_{i=1}^n \in \{0,1\}^n$$

sodaß die Likelihood-Funktion von der Form

$$L(p,\boldsymbol{x}) := \prod_{i=1}^n f(x_i,p) = p^{k(\boldsymbol{x})}(1-p)^{n-k(\boldsymbol{x})}$$

ist, wobei

$$k(\boldsymbol{x}) := \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \sum_{i=1}^n x_i^2. \quad [3]$$

Offenbar stimmt $k(\boldsymbol{x})$ mit der Anzahl der Komponenten von $\boldsymbol{x}$ überein, die gleich Eins sind.

Der unbekannte Parameter p wird durch das Maximum von $L(.,\boldsymbol{x})$ über $[0,1]$ für fixiertes $\boldsymbol{x}$ geschätzt. Offenbar gilt

$$\begin{aligned}\frac{\partial}{\partial p}L(p,\boldsymbol{x}) &= k(\boldsymbol{x})p^{k(\boldsymbol{x})-1}(1-p)^{n-k(\boldsymbol{x})} - p^k(n-k(\boldsymbol{x}))(1-p)^{n-k(\boldsymbol{x})-1}\\ &= p^{k(\boldsymbol{x})-1}(1-p)^{n-k(\boldsymbol{x})-1}\left(k(\boldsymbol{x})(1-p) - p(n-k(\boldsymbol{x}))\right)\\ &= p^{k(\boldsymbol{x})-1}(1-p)^{n-k(\boldsymbol{x})-1}(k(\boldsymbol{x}) - k(\boldsymbol{x})p - pn + pk(\boldsymbol{x}))\\ &= p^{k(\boldsymbol{x})-1}(1-p)^{n-k(\boldsymbol{x})-1}(k(\boldsymbol{x}) - pn)\end{aligned}$$

d.h.

$$\frac{\partial}{\partial p}L(p,\boldsymbol{x}) = 0 \quad \Leftrightarrow \quad p \in \left\{0, \frac{k(\boldsymbol{x})}{n}, 1\right\}$$

Wegen

$$L(0,\boldsymbol{x}) = L(1,\boldsymbol{x}) = 0$$

liegt das Maximum bei $k(\boldsymbol{x})/n$, d.h. als Schätzer g für den Parameter p bekommt man mit der Maximum-Likelihood Methode

$$g(x_1,\ldots,x_n) := H_n := \frac{k(\boldsymbol{x})}{n}.$$

Demnach ist unsere Stichprobenfunktion von der Gestalt $\dfrac{1}{n}\sum_{i=1}^n x_i^2$.

Der Quotient H_n wird als **relative Häufigkeit** des Gelingens des Experiments bezeichnet.

Darüber hinaus läßt sich für jedes positive ε

$$\lim_{n\to\infty} P(|H_n - p| < \varepsilon) = 1$$

zeigen, d.h. fast sicher konvergiert die relative Häufigkeit H_n gegen den Parameter p. Diese stochastische Konvergenzeigenschaft von H_n heißt das **(schwache) Gesetz der großen Zahl von Bernoulli** ([20] S. 1054).

[3] Die Anwendung des Operators T macht aus einem Spaltenvektor einen Zeilenvektor, so daß $\boldsymbol{x}^{\mathrm{T}}$ und $\boldsymbol{x}$ im Sinne der Matrizenrechnung miteinander multipliziert werden können.

3.2.2 ML-Schätzer für $\boldsymbol{\lambda}$ in $\mathscr{P}(\boldsymbol{\lambda})$

Wir setzen voraus, daß $X \sim \mathscr{P}(\lambda)$ und führen das Experiment n mal durch, was zu einer Stichprobe

$$[k_i]_{i=1}^n \in \mathbb{N}_0^n$$

führt. Wegen $X \sim \mathscr{P}(\lambda)$ gilt

$$f(k_i, \lambda) = P(X = k_i) = \frac{\lambda^{k_i}}{k_i!} e^{-\lambda}.$$

Demzufolge lautet die Maximum-Likelihood-Funktion

$$L(\lambda, k_1, \ldots, k_n) = \prod_{i=1}^n f(k_i, \lambda) = \frac{1}{\prod_{i=1}^n k_i!} \lambda^s e^{-n\lambda}, \quad s := \sum_{i=1}^n k_i.$$

Wegen

$$\frac{\partial}{\partial \lambda} L(\lambda, k_1, \ldots, k_n) = \frac{1}{\prod_{i=1}^n k_i!} (s\lambda^{s-1} e^{-n\lambda} - \lambda^s n e^{-n\lambda}) = \frac{\lambda^{s-1} e^{-n\lambda}}{\prod_{i=1}^n k_i!} (s - \lambda n)$$

haben wir

$$\frac{\partial}{\partial \lambda} L(\lambda, k_1, \ldots, k_n) = 0 \quad \Leftrightarrow \quad s = \lambda n$$

d.h. als Maximum-Likelihood-Schätzer für λ bekommen wir

$$g(k_1, \ldots, k_n) := \frac{1}{n} \sum_{i=1}^n k_i.$$

Beispiel 37 Schätzung von $\boldsymbol{\lambda}$ in poissonverteiltem Datenvektor

Wir setzen $\lambda := 8$, $n := 200$ und erzeugen mittels

$$\boldsymbol{x} := \mathrm{rpois}(n, \lambda)$$

einen n-dimensionalen Vektor mit nichtnegativen ganzzahligen Komponenten, die zum Parameter λ poissonverteilt sind. Um uns zu veranschaulichen, was das bedeutet, berechnen wir

$$\max(\boldsymbol{x}) = 18$$

sodaß

$$\boldsymbol{x} \in \{0, \ldots, 18\}^n.$$

Wir erklären die natürliche Zahl h_k durch die Häufigkeit, mit der $\boldsymbol{x}$ den Wert k annimmt, d.h. mittels der Funktion

$$f(x, a) := \begin{cases} 1 & \text{falls } x = a \\ 0 & \text{falls } x \neq a \end{cases}$$

gilt

$$h_k = \sum_{i=1}^{n} f(x_i, k).$$

Die Division mit n liefert die entsprechenden relativen Häufigkeiten

$$f_k := \frac{h_k}{n}.$$

Die Daten $\{x_1, \ldots, x_n\}$ können nun als zum Parameter λ poissonverteilt angesehen werden, falls

$$f_k \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

In Abbildung 20 werden die Punktmengen

$$\underbrace{\{(k, f_k) : k = 0, \ldots, 18\}}_{\text{blaue Punkte}}, \quad \underbrace{\left\{\left(k, \frac{\lambda^k}{k!} e^{-\lambda}\right) : k = 0, \ldots, 18\right\}}_{\text{rote Punkte}}$$

miteinander verglichen. Mit einigem Wohlwollen läßt sich für $\boldsymbol{x}$ Poisson-Verteilung konstatieren. Es gibt eine wesentliche Abweichung bei der Häufigkeit mit der $\boldsymbol{x}$ den Wert 9 annimmt, sie ist augenscheinlich höher als $\frac{\lambda^9}{9!} e^{-\lambda}$. Nichtsdestotrotz bekommt man

$$\overline{\boldsymbol{x}} = \frac{1}{n} \sum_{k=1}^{n} x_k = 7.965$$

als Schätzung für λ, von dem wir wissen, daß es in Wirklichkeit 8 ist. Die Zahl $\overline{\boldsymbol{x}}$ heißt **arithmetisches Mittel** oder **Mittelwert** der Zahlen $x_1, \ldots, x_n$.

Zur Ergänzung wird in Abbildung 21 noch das **Stabdiagramm** für $\boldsymbol{x}$ gezeigt. Der rote Stab (Quader) korrespondiert mit dem 'Ausreißer' bei 9.

Wir kehren nocheinmal zum Beispiel 16 zurück, bei dem es um das Geiger-Müller-Zählrohr ging. Dort war der Parameter ϱ zu schätzen, was durch das Zählen der Entladungsstöße (es knackt im Zählrohr) während der Zeitspanne T geschah. Bezeichnet N_T diese Anzahl, so setzt man

$$\varrho := \frac{N_T}{T}.$$

Diese Vorgehensweise deckt sich mit der Verwendung des ML-Schätzers für λ. Entsprechend diesem Ansatz hat man n-mal die Anzahl x_i von Entladungsstößen während der Zeitspanne t zu ermitteln und dann das arithmetische Mittel $\overline{x}$ als Schätzung für λ zu verwenden. Offenbar kann man die Summe $x_1 + \ldots + x_n$ auch als die Anzahl der Entladungsstöße im Zeitintervall $T := nt$ auffassen, was $N_T = x_1 + \ldots + x_n$ bedeutet, sodaß

$$t\varrho = t\frac{N_T}{T} = t\frac{x_1 + \ldots + x_n}{nt} = \overline{\boldsymbol{x}}.$$

◇

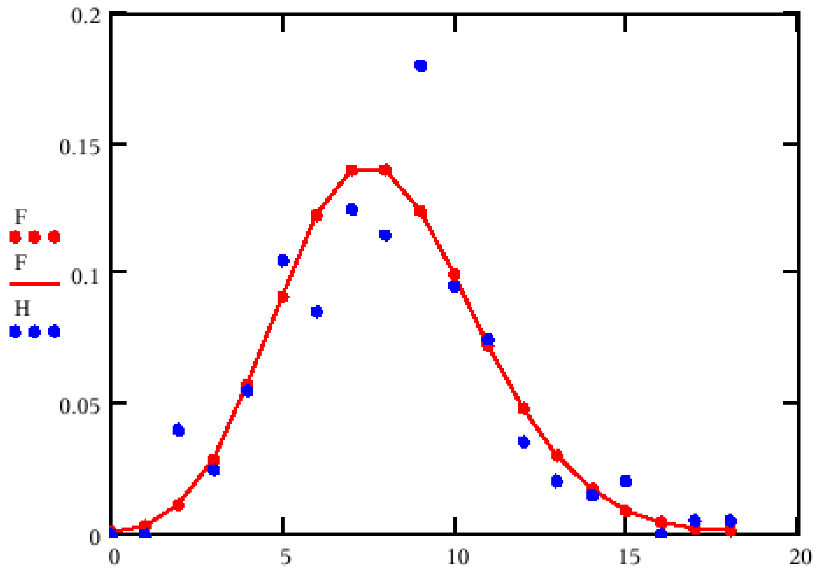


Abbildung 20: Es werden die relativen Häufigkeiten, mit denen $\boldsymbol{x}$ die Werte $0, \ldots, 18$ annimmt (blaue Punkte), mit den Werten der Wahrscheinlichkeitsfunktion $f(k,\lambda) := \frac{\lambda^k}{k!}e^{-\lambda}$ an den Stellen $k = 0, \ldots, 18$ für $\lambda = 8$ (rote Punkte) verglichen. Der Übersichtlichkeit wegen, sind die roten Punkte durch Geraden zu einem Polygonzug miteinander verbunden wurden.

3.2.3 ML-Schätzer für μ und σ in $\mathscr{N}(\mu, \sigma^2)$

Wir setzen $X \sim \mathscr{N}(\mu, \sigma^2)$ voraus. In Anlehnung an die Vorgehensweise für diskrete Zg, setzt man auch für stetige Zg die Likelihood-Funktion entsprechend

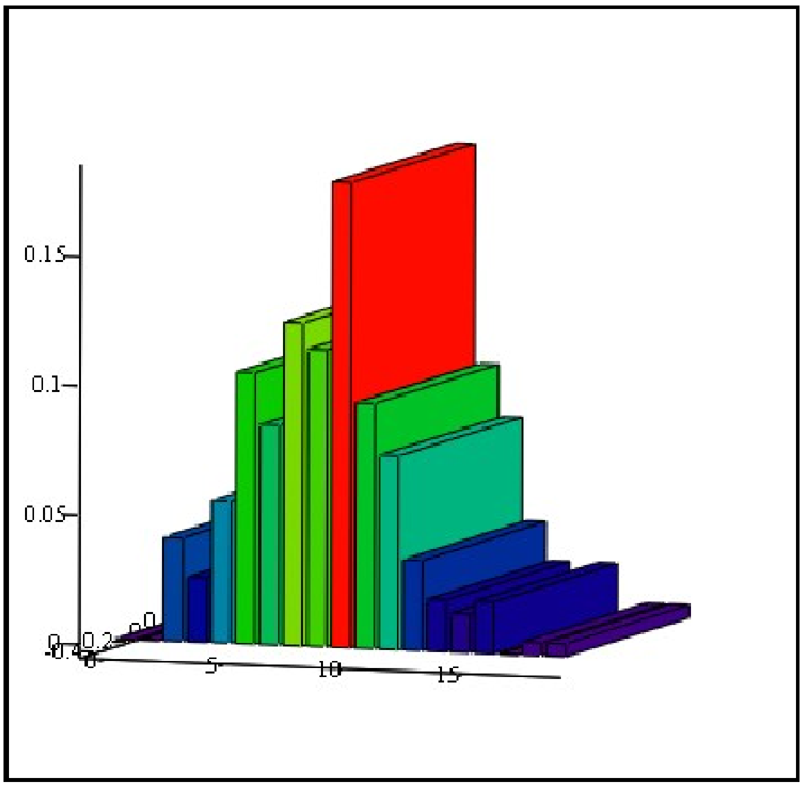
$$\begin{aligned} L(\mu, \sigma, x_1, \ldots, x_n) &= \prod_{k=1}^{n} \varphi_{\mu,\sigma}(x_k) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} \prod_{k=1}^{n} \exp\left(-\frac{(x_k - \mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2\right) \end{aligned}$$

an. Wir haben nun μ und σ so zu bestimmen, daß L für fixierte x_k maximal wird. Offenbar ist L positiv für alle $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$. Daher wird

$$L^*(\mu, \sigma, x_1, \ldots, x_n) := \ln L(\mu, \sigma, x_1, \ldots, x_n)$$

maximal genau dann, wenn L maximal wird. Unter Beachtung von

$$\ln(ab) = \ln(a) + \ln(b), \quad \ln(a^n) = n \ln(a)$$



H

Abbildung 21: Stabdiagramm des Datenvektors $\boldsymbol{x}$. Der rote Stab (Quader) korrespondiert mit dem Ausreißer bei 9 in Abbildung 20.

erhält man

$$\begin{aligned} L^*(\mu,\sigma,x_1,\ldots,x_n) &= \ln\left(\frac{1}{(2\pi)^{n/2}}\frac{1}{\sigma^n}\right) - \frac{1}{2\sigma^2}\sum_{k=1}^n (x_k-\mu)^2 \\ &= \ln\left(\frac{1}{(2\pi)^{n/2}}\right) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{k=1}^n (x_k-\mu)^2. \end{aligned}$$

Eine notwendige Bedingung für die Annahme eines Maximums ist das simultane Verschwinden der partiellen Ableitungen von L^* nach μ und σ:

$$\begin{aligned} \frac{\partial L^*(\mu,\sigma,x_1,\ldots,x_n)}{\partial\mu} &= \frac{1}{\sigma^2}\sum_{k=1}^n (x_k-\mu) \\ \frac{\partial L^*(\mu,\sigma,x_1,\ldots,x_n)}{\partial\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{k=1}^n (x_k-\mu)^2. \end{aligned}$$

Die Forderung $L^*_\mu = 0 = L^*_\sigma$ führt auf

$$\mu = \frac{1}{n}\sum_{k=1}^n x_k, \quad \sigma^2 = \frac{1}{n}\sum_{k=1}^n (x_k-\mu)^2$$

d.h. der ML-Schätzer für μ ist das **Stichprobenmittel** und der ML-Schätzer für σ die Stichprobenfunktion

$$s^*(x_1,\ldots,x_n) := \sqrt{\frac{1}{n}\sum_{k=1}^n (x_k-\mu)^2}$$

die mit der Stichprobenvarianz $s^2(x_1,\ldots,x_n)$, wobei

$$s(x_1,\ldots,x_n) := \sqrt{\frac{1}{n-1}\sum_{k=1}^n (x_k-\mu)^2}$$

im Verhältnis

$$s^* = \sqrt{\frac{n-1}{n}}s$$

steht.

Weshalb wird bei s anstatt durch n nur durch $n-1$ geteilt? Das hat mit dem Wunsch nach Erwartungstreue zu tun.

3.3 Erwartungstreue und Konsistenz

Eine Schätzfunktion $g(X_1,\ldots,X_n)$ für den Parameter θ heißt **erwartungstreu**, falls $g(X_1,\ldots,X_n)$ aufgefaßt als Zg über Ω^n die Bedingung

$$Eg(X_1,\ldots,X_n) = \theta$$

erfüllt.

Das Stichprobenmittel $\overline{X} := \frac{1}{n}\sum_{k=1}^{n} X_k$ ist wegen

$$E\overline{X} = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} EX_k = \frac{1}{n}\sum_{k=1}^{n} \mu = \mu$$

erwartungstreu.

Hinsichtlich der Erwartungstreue der Zufallsgröße **Stichprobenvarianz** S^2, wobei

$$S := \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2} \tag{41}$$

aus n unabhängigen identisch verteilten Zg gebildet wird, stellen wir

$$ES^{*2} = \frac{n-1}{n}\sigma^2, \quad S^* := \sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2}$$

fest, so daß

$$ES^2 = E\frac{n}{n-1}S^{*2} = \sigma^2.$$

Für drei identisch verteilte und unabhängige Zg X, Y, Z gilt

$$\begin{aligned}
E(2X-Y-Z)^2 &= E\left(\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 2X \\ -Y \\ -Z \end{bmatrix}\begin{bmatrix} 2X & -Y & -Z \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) \\
&= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 4EX^2 & -2E(XY) & -2E(XZ) \\ -2E(YX) & EY^2 & E(YZ) \\ -2E(ZX) & E(ZX) & EZ^2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 4EX^2 & -2EXEY & -2EXEZ \\ -2EYEX & EY^2 & EYEZ \\ -2EZEX & EZEX & EZ^2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 4EX^2 & -2\mu^2 & -2\mu^2 \\ -2\mu^2 & EY^2 & \mu^2 \\ -2\mu^2 & \mu^2 & EZ^2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
&= 4EX^2 + EY^2 + EZ^2 - 6\mu^2 \\
&= 4(EX^2 - \mu^2) + (EY^2 - \mu^2) + (EZ^2 - \mu^2) = 6\sigma^2.
\end{aligned}$$

Daher bekommen wir für $n = 3$ und $\overline{X} := \frac{1}{n}\sum_{k=1}^{n} X_k$

$$\begin{aligned} ES^{*2} &= E\left(\frac{1}{3}\sum_{k=1}^{3}(X_k - \overline{X})^2\right) = \frac{1}{3}\sum_{k=1}^{3} E(X_k - \overline{X})^2 = \frac{1}{3}\sum_{k=1}^{3} E\left(X_k - \frac{1}{3}\sum_{k=1}^{3} X_k\right)^2 \\ &= \frac{1}{3}\sum_{k=1}^{3} E\left(\frac{1}{3}\left(3X_k - \sum_{k=1}^{3} X_k\right)\right)^2 = \frac{1}{3^3}\sum_{k=1}^{3} \underbrace{E\left(3X_k - \sum_{k=1}^{3} X_k\right)^2}_{6\sigma^2} = \frac{18}{3^3}\sigma^2 = \frac{2}{3}\sigma^2. \end{aligned}$$

Daher ist die Zufallsgröße S^2 erwartungstreu im Gegensatz zum ML-Schätzer S^*. Die Verwendung von S^* liefert eine systematische Unterschätzung von σ, die allerdings für große Stichprobenumfänge bedeutungslos wird.

Eine Schätzfunktion g aufgefaßt als Zg, d.h. ihre Argumente werden durch die Zg X_1, …, X_n bestimmt, heißt **konsistent**, falls ihre Varianz für $n \to \infty$ gegen Null strebt:

$$\lim_{n\to\infty} \mathrm{Var}(g(X_1, \ldots, X_n)) = 0.$$

Das Stichprobenmittel ist konsistent, denn mit (15) und (16) gilt

$$\lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \lim_{n\to\infty} \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{Var}(X_k) = \lim_{n\to\infty} \frac{n\sigma^2}{n^2} = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

Selbstverständlich wurde die Unabhängigkeit aller X_k und ihre identische Verteilung vorausgesetzt (Vgl. [15] S.137, [13] S.491). Demnach ist durch Stichprobenvergrößerung ein geringeres Streuen um den zu schätzenden Mittelwert zu erwarten.

3.4 Intervallschätzungen

Durch eine Punktschätzung bekommt man einen numerischen Wert, der in der Nähe des zu schätzenden Parameters liegt. Falls man mit einem konsistenten Schätzer arbeitet, wird die Schätzung um so genauer, je größer der Stichprobenumfang n ist. Es läßt sich aber nicht beschreiben, wie weit man noch, abhängend von n, vom tatsächlichen Wert des zu schätzenden Parameters entfernt ist. Dieses Problem kann mit der Intervallschätzung erledigt werden.

Zu diesem Zweck wird der Begriff des **Vertrauens- oder Konfidenzintervalls zum Niveau $1 - \alpha$** eingeführt, wobei $\alpha \in [0, 1]$. Bei den Anwendungen wird tatsächlich

$$\alpha \in \{0.01, \quad 0.05, \quad 0.1\}$$

gewählt, sodaß man es mit Vertrauensintervallen zu den Niveaus

$$99\,\% \quad 95\,\% \quad 90\,\%$$

zu tun hat. Die Zahl α wird auch als **Irrtumswahrscheinlichkeit** oder **Signifikanzzahl** bezeichnet.

Wir sagen, daß das Intervall $[a, b] \subset \mathbb{R}$ für den zu schätzenden Parameter $\theta \in \mathbb{R}$ ein **Vertrauensintervall zum Niveau** $1-\alpha$ ist, falls $[a, b]$ mit der Wahrscheinlichkeit $1-\alpha$ den Parameter θ überdeckt. Die Wahl eines kleinen α schafft großes Vertrauen. Wir werden sehen, daß bei Beibehaltung des Stichprobenumfanges n die Verkleinerung von α mit dem Auseinanderrücken von $[a, b]$ einher geht, sodaß eine hohes Vertrauensniveau mit einem Verlust an Schätzgenauigkeit verbunden ist. Falls man α konstant hält und n vergrößert, rücken a und b enger zusammen.

Man beachte, daß die Sprechweise

θ liegt mit einer Wahrscheinlichkeit von $1-\alpha$ in $[a, b]$

nicht korrekt ist, da es sich bei θ um eine fixierte reelle Zahl und um keine Zufallsgröße handelt. Die Zufallsgröße ist vielmehr das Intervall $[a, b]$, das zufällig mit Hilfe einer zufällig ausgewählten Stichprobe erzeugt wird, d.h. a und b sind Zufallsgrößen, deren Fixierung durch Ziehen einer Stichprobe das Vertrauensintervall liefert. Die folgenden Abschnitte illustrieren diese Verhältnisse.

3.4.1 Vertrauensintervall für $\boldsymbol{\mu}$ einer Zg, die entsprechend $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ verteilt ist mit bekanntem $\boldsymbol{\sigma}$

[5] S.388, [13] S.515, [15] S.148

Das Vertrauensintervall für μ suchen wir in der Form

$$\left[\overline{X}-c, \overline{X}+c\right], \quad \overline{X} := \frac{1}{n}\sum_{k=1}^{n} X_k, \quad c \in \mathbb{R}_+$$

wobei X_k die Zg bezeichnet, die zum Ziehen der k-ten Komponente eines Stichprobenvektors aus einer entsprechend $\mathscr{N}(\mu, \sigma^2)$ verteilten Grundgesamtheit gehört.

Wir bestimmen die positive Konstante c so, daß für $\alpha \in [0, 1]$ die Gleichung

$$P(\overline{X}-c \le \mu \le \overline{X}+c) = 1-\alpha$$

gilt. Einfache Umformungen führen auf

$$\begin{aligned} 1-\alpha &= P(\overline{X}-c \le \mu \le \overline{X}+c) \\ &= P(\overline{X}-\mu-c \le 0 \le \overline{X}-\mu+c) \\ &= P\left(\frac{\overline{X}-\mu}{\sigma}\sqrt{n} - \frac{c}{\sigma}\sqrt{n} \le 0 \le \frac{\overline{X}-\mu}{\sigma}\sqrt{n} + \frac{c}{\sigma}\sqrt{n}\right). \end{aligned}$$

Setzt man

$$Y := \frac{\overline{X}-\mu}{\sigma}\sqrt{n}, \quad z := \frac{c\sqrt{n}}{\sigma} \tag{42}$$

so bekommt man

$$1-\alpha = P\left(Y - z \le 0 \le Y + z\right)$$

und entsprechend (14), (15) und (16) die Gleichungen

$$EY = \frac{E\overline{X} - \mu}{\sigma}\sqrt{n} = 0, \quad \mathrm{Var}(Y) = \frac{n}{\sigma^2}\mathrm{Var}(\overline{X}) = \frac{1}{n\sigma^2}\sum_{k=1}^{n}\underbrace{\mathrm{Var}(X_k)}_{=\sigma^2} = 1.$$

Da $X_k \sim \mathcal{N}(\mu, \sigma^2)$, gilt entsprechend des **Additionssatzes**

$$\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$$

d.h.

$$Y \sim \mathcal{N}(0,1).$$

Demnach hat man für standard-normalverteilte Zg Y die Gleichung

$$P(Y - z \le 0 \le Y + z) = 1 - \alpha$$

zu lösen, wobei positives z gesucht ist. Offenbar gilt

$$P(Y - z \le 0 \le Y + z) = P(\{Y \le z\} \cap \{-z \le Y\}) = P(-z \le Y \le z).$$

Wegen $Y \sim \mathcal{N}(0,1)$, können wir die Symmetrie von

$$\varphi_{0,1}$$

bezüglich der y-Achse wiefolgt ausnutzen:

$$P(-z \le Y \le z) = \Phi_{0,1}(z) - \Phi_{0,1}(-z) = \Phi_{0,1}(z) - (1 - \Phi_{0,1}(z)) = 2\Phi_{0,1}(z) - 1.$$

Daher bekommt man

$$1 - \alpha = 2\Phi_{0,1}(z) - 1$$

d.h.

$$1 - \alpha/2 = \Phi_{0,1}(z).$$

Aus dieser Gleichung kann schließlich die gesuchte Zahl z bestimmt werden. Die Lösung dieser Gleichung wird oft mit $z_{1-\frac{\alpha}{2}}$ bezeichnet.

Algorithmus zur Berechnung eines Vertrauensintervalls für $\boldsymbol{\mu}$ bei bekannter Standardabweichung $\boldsymbol{\sigma}$

- Wähle $\alpha \in (0,1)$.
- Bestimme z aus der Gleichung

$$1 - \frac{\alpha}{2} = \Phi_{0,1}(z).$$

- Wähle eine Stichprobe $x_1, \ldots, x_n$ im Umfang n und bilde das Stichprobenmittel $\boldsymbol{\overline{x}}$.
- Berechne $c := \dfrac{\sigma z}{\sqrt{n}}$.

Statistische Aussage: Das Intervall

$$[\boldsymbol{\overline{x}} - c, \boldsymbol{\overline{x}} + c]$$

überdeckt den Erwartungswert μ mit einer Wahrscheinlichkeit von $1 - \alpha$.

Bemerkung: Offenbar stimmt die Konstante c mit dem Produkt aus $\mathrm{Var}(\overline{X})$ und $\Phi_{0,1}^{-1}\left(1 - \frac{\alpha}{2}\right)$ überein.

Da der **mathcad**-Befehl 'qnorm' die Gleichung

$$\Phi_{\mu,\sigma}^{-1}(x) = \mathrm{qnorm}(x, \mu, \sigma).$$

erfüllt, bekommt man mit seiner Hilfe für $\alpha \in \{0.01,\ 0.05,\ 0.1\}$ den Werteverlauf

$$\begin{aligned} \Phi_{0,1}^{-1}(1 - \tfrac{0.01}{2}) &= \mathrm{qnorm}(1 - \tfrac{0.01}{2}, 0, 1) = 2.567 \\ \Phi_{0,1}^{-1}(1 - \tfrac{0.05}{2}) &= \mathrm{qnorm}(1 - \tfrac{0.05}{2}, 0, 1) = 1.960 \\ \Phi_{0,1}^{-1}(1 - \tfrac{0.10}{2}) &= \mathrm{qnorm}(1 - \tfrac{0.10}{2}, 0, 1) = 1.645 \end{aligned}$$

d.h.

$$\begin{array}{c|c|c|c} \alpha & 0.010 & 0.050 & 0.100 \\ \hline z & 2.576 & 1.960 & 1.645 \end{array}, \quad z = \Phi_{0,1}^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{43}$$

Es ist zu beachten, daß man auf dem 'Rand' der Tabelle den Wert $\Phi_{0,1}^{-1}(1 - \alpha/2)$ abliest, wobei $1 - \alpha$ in ihrem Inneren verteilt ist.

Abhängigkeit des Konfidenzintervalls von der Variation von n und α:

- Da man durch $\sqrt{n}$ teilt, sorgt eine Vergrößerung von n für eine Verkleinerung von c, so daß das Überdeckungsintervall kleiner wird, was eine Präzisierung der Schätzung bei konstant gehaltenem Vertrauensniveau bedeutet.
- Hält man dagegen n konstant und vergrößert man α im Rahmen $(0, 1)$, so wird z kleiner, was mit einer Verkleinerung von c einhergeht, so daß das Überdeckungsintervall zusammenschrumpft, was aber mit einem Vertrauensverlust bezahlt wird.
- Im Extremfall $\alpha = 1$ erhält man $z = 0$, sodaß c verschwindet und das Überdeckungsintervall nur aus $\overline{x}$ besteht. Die Wahrscheinlichkeit μ mit $\overline{x}$ zu treffen, ist tatsächlich Null.
- Im Extremfall $\alpha = 0$ bekommt man $z = \infty$, sodaß das Überdeckungsintervall mit $\mathbb{R}$ übereinstimmt. Da $\mu \in \mathbb{R}$, überdeckt man mit $\mathbb{R}$ sicher μ.

Beispiel 38 Eine Grundgesamtheit sei normalverteilt und die Komponenten des Vektors

$$\boldsymbol{x} := [17.21,\ 17.22,\ 17.22,\ 17.20,\ 17.19]^{\mathrm{T}}$$

seien durch eine Stichprobe vom Umfang $n := 5$ zustande gekommen. Wie lautet das aus dieser Stichprobe resultierende Vertrauensintervall für μ zum Vertrauensniveau 99 %, wobei die Standardabweichung

$$\sigma := 0.01304$$

als bekannt vorausgesetzt wird.

- Da mit einem Vertrauensniveau von 99 % gearbeitet werden soll, haben wir

  $$\alpha = 0.01$$

  zu wählen.
- Demnach bekommen wir aus Tabelle (43) als Lösung von $1 - \dfrac{\alpha}{2} = \Phi_{0,1}(z)$

  $$z = 2.576.$$

- Das Stichprobenmittel lautet

  $$\overline{\boldsymbol{x}} = \frac{1}{5}(17.21 + 17.22 + 17.22 + 17.20 + 17.19) = 17.208.$$

- Die Konstante c lautet daher

  $$c = \frac{\sigma z}{\sqrt{n}} = \frac{0.01304 \times 2.576}{\sqrt{5}} = 0.01502.$$

Antwort: Das Vertrauensintervall lautet

$$[17.208 - 0.01502, 17.208 + 0.01502].$$

◇

Beispiel 39 Vertrauensintervall für $\boldsymbol{\mu}$ bei bekanntem $\boldsymbol{\sigma}$ von zufällig erzeugtem Datenvektor $\boldsymbol{v}$ mit normalverteilten Komponenten, Einführung einiger Begriffe aus der beschreibenden Statistik

Wir wählen als Datenumfang

$$m := 3000$$

und erzeugen mittels des **mathcad**-Befehls

$$v := [v_k]_{k=0}^{2999} := \mathrm{rnorm}(m, \mu, \sigma), \quad \mu := 2, \quad \sigma := 0.3$$

einen Zufallsvektor, dessen Komponenten mit den Parametern μ und σ normalverteilt sind.

Es bietet sich an, in diesem Zusammenhang einige Begriffe aus der **beschreibenden Statistik** einzuführen.

- Es sei
$$A := \min(v), \quad B := \max(v).$$
Dann ist die **Spannweite** der durch v gegebenen Daten durch
$$S := B - A$$
erklärt.

- Entsprechend
$$\Delta := \frac{S + 2\varepsilon}{N}, \quad 0 < \varepsilon, \quad A_i := (A - \varepsilon) + i\Delta, \quad i := 0, \dots, N,$$
bekommen wir eine äquidistante Einteilung des nach rechts halboffenen Intervalls
$$[A - \varepsilon, B + \varepsilon)$$
in N **Klassen**
$$K_0 := [A_0, A_1), \quad K_1 := [A_1, A_2) \quad \dots \quad K_{N-1} := [A_{N-1}, A_N).$$
Die Zahl Δ wird die **Klassenbreite** genannt.

- Wir wollen nun die Anzahl aller Komponenten des Datenvektors v ermitteln, die in der Klasse K_i liegen. Zu diesem Zweck führen wir die Indikatorfunktion
$$f(x, a, b) := \begin{cases} 0 & \text{falls } x < a \\ 1 & \text{falls } a \leq x < b \\ 0 & \text{falls } b \leq x \end{cases}$$
ein. Offenbar überprüft f, ob x in $[a, b)$ liegt oder nicht, f liefert je nachdem den Funktionswert 1 oder 0. Daher gibt die Summe
$$h_i := \sum_{k=0}^{m-1} f(v_k, A_i, A_{i+1})$$
an, wieviele Einträge von v zu K_i gehören. Die natürliche Zahl h_i heißt **absolute Häufigkeit** der Klasse K_i.

- Normalverteilung der Einträge von v bedeutet nun, daß die **relativen Häufigkeiten**
$$f_i := h_i/m$$
die Approximation
$$f_i \approx \Phi_{\mu,\sigma}(A_{i+1}) - \Phi_{\mu,\sigma}(A_i) = \int_{A_i}^{A_{i+1}} \varphi_{\mu,\sigma}(x)\mathrm{d}x, \quad \mu = 2, \sigma = 0.3$$
erfüllen.

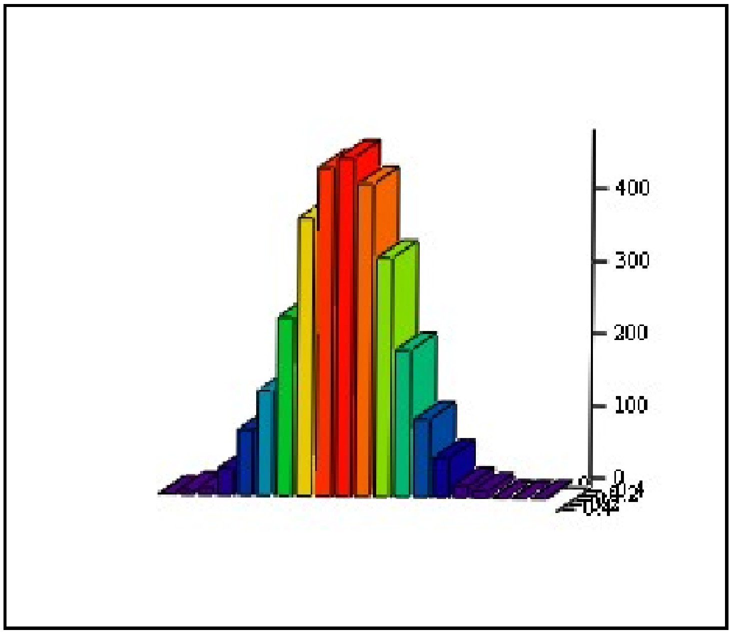


Abbildung 22: Absolute Häufigkeiten h_i dargestellt über den Klassen K_i, $i = 0, \ldots, 19$. Eine solche Darstellung heißt **Histogramm**. Es gilt $\sum_{i=0}^{19} h_i = m = 3000$.

Die graphische Zuordnung von h_i zur Klasse K_i in einem Balkendiagramm wird **Histogramm** genannt. Abbildung 22 zeigt ein solches Diagramm für $N = 20$. Da eine Ähnlichkeit zur Gestalt von $\varphi_{\mu,\sigma}$ zu erkennen ist, scheinen die Einträge von v tatsächlich normalverteilt zu sein.

Wir wählen als Stichprobenumfang $n := 30$, $\alpha := 0.2$ und bilden

$$\overline{\boldsymbol{x}}_i := \frac{1}{n} \sum_{k=(i-1)n}^{in-1} v_k, \quad i = 1, \ldots, 100$$

d.h. wir berechnen 100 Stichprobenmittel der Form

$$\overline{\boldsymbol{x}}_1 := \frac{1}{n} \sum_{k=0}^{n-1} v_k, \quad \overline{\boldsymbol{x}}_2 := \frac{1}{n} \sum_{k=n}^{2n-1} v_k \quad \ldots \quad \overline{\boldsymbol{x}}_{100} := \frac{1}{n} \sum_{k=99n}^{100n-1} v_k.$$

Wegen $c = \sigma z / \sqrt{n}$ bekommen wir mit

$$z = \Phi_{0,1}^{-1}(1 - \alpha/2) = \text{qnorm}(1 - \alpha/2, 0, 1) = \text{qnorm}(1 - 0.1, 0, 1) = 1.282$$

für die Konstante c den Wert

$$c = \sigma z / \sqrt{n} = \frac{0.3 \times 1.282}{\sqrt{30}} = 0.070.$$

Die 100 verschiedenen Stichprobenmittel $\overline{\boldsymbol{x}}_i$ liefern 100 Konfidenzintervalle

$$[\overline{\boldsymbol{x}}_i - c, \overline{\boldsymbol{x}}_i + c]$$

für den Erwartungswert μ. Da wir mit einer Irrtumswahrscheinlichkeit von $\alpha = 0.2$ die Konstante c berechnet haben, erwarten wir, daß ca. 80 Intervalle μ überedecken.

Schließlich zählen wir das Auftreten der Eins in der $(0,1)$-Folge

$$f(\mu, \overline{\boldsymbol{x}}_1 - c, \overline{\boldsymbol{x}}_1 + c), \ldots, f(\mu, \overline{\boldsymbol{x}}_{100} - c, \overline{\boldsymbol{x}}_{100} + c).$$

Die vielfältige Generation von v zeigt, daß die Anzahl der Einsen für jede Generation ungefähr mit 80 übereinstimmt. $\diamond$

3.4.2 Vertrauensintervall für $\boldsymbol{\mu}$ einer Zg, die entsprechend $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ mit unbekanntem $\boldsymbol{\sigma}$ verteilt ist

[5] S.389,, [13] S.519, [15] S.150

Falls σ nicht bekannt ist, ersetzt man in (42) den Parameter σ durch die Wurzel aus der Stichprobenvarianz s^2 definiert wie in (41). Auf diese Weise kommt man zu der Zg

$$T := \frac{\overline{X} - \mu}{S}\sqrt{n}, \quad S := \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2}.$$

Es stellt sich heraus, daß

$$T \sim t_{n-1}$$

wobei t_n die Student-Verteilung mit n Freiheitsgraden darstellt. Die zugehörige Verteilungsfunktion wird mit $F_n(x)$ bezeichnet. Man beachte, daß bei der Einführung der Student-Verteilung nur standardnormalverteilte Zg zugelassen wurden, sodaß man nicht so einfach auf $T \sim t_{n-1}$ schließen kann. In [6] S.408 wird gezeigt, daß die Verteilungsbehauptung auch für $\sigma \neq 1$ gilt.

In Abschnitt 3.4.3 wird es plausibel gemacht, daß man $\sum_{k=1}^{n}(X_k - \overline{X})^2$ als Summe von $n-1$ Quadraten schreiben kann, sodaß man nur mit $n-1$ Freiheitsgraden zu arbeiten hat.

Algorithmus zur Berechnung eines Vertrauensintervalls für $\boldsymbol{\mu}$ bei unbekannter Standardabweichung

- Wähle $\alpha \in (0,1)$.
- Bestimme z aus der Gleichung
$$1 - \frac{\alpha}{2} = F_{n-1}(z).$$
Die Lösung wird oft mit $t_{1-\frac{\alpha}{2}}(n-1)$ bezeichnet.
- Ziehe Stichprobe $\boldsymbol{x} := [x_k]_{k=1}^n$ im Umfang n und berechne
$$\overline{\boldsymbol{x}} := \frac{1}{n}\sum_{k=1}^{n} x_k, \quad s := \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2}.$$

Für Berechnungszwecke kann man

$$s = \sqrt{\frac{1}{n-1}\left(\sum_{k=1}^{n} x_k^2 - n\overline{\boldsymbol{x}}^2\right)}$$

verwenden.

- Setze
$$c := \frac{sz}{\sqrt{n}}.$$

Statistische Aussage: Das Intervall

$$[\overline{\boldsymbol{x}} - c, \overline{\boldsymbol{x}} + c]$$

überdeckt den Erwartungswert μ mit einer Wahrscheinlichkeit von $1 - \alpha$.

Mittels des **mathcad**-Befehls $\mathrm{qt}(x, n)$ bekommt man den in Tabelle (79) auf S.252 aufgezeichneten Werteverlauf von $F_n^{-1}(x)$ für

$$(n, x) \in \{1, \ldots, 30\} \times \{0.900,\ 0.950,\ 0.975,\ 0.990,\ 0.995,\ 0.999\}, \quad x := 1 - \frac{\alpha}{2}$$

wobei F_n die Verteilungsfunktion der t-Verteilung für n Freiheitsgrade darstellt. Die erste Zeile sind dabei die Funktionswerte von

$$\tan(\pi(x - 0.5))$$

was sich sofort aus (36) ergibt.

Ablesebeispiel: Wir wollen den Wert z von $F_6^{-1}(0.990)$ wissen. Zu diesem Zweck liest man in Zeile 6 und Spalte 4 den Wert 3.143 ab. Die für uns in diesem Zusammenhang wichtigen Werte von z befinden sich im Inneren der Tabelle.

Beispiel 40 Wir benutzen dieselbe Stichprobe wie in Beispiel 38 einer normalverteilten Grundgesamtheit, d.h.

$$\boldsymbol{x} := [17.21,\ 17.22,\ 17.22,\ 17.20,\ 17.19]^{\mathrm{T}}$$

gehen aber jetzt von unbekanntem σ aus.

Wie lautet das aus $\boldsymbol{x}$ resultierende Vertrauensintervall für μ mit einem Vertrauensniveau von 99 %?

- Da das Vertrauensniveau 99 % sein soll, ergibt sich für α der Wert
$$\alpha = 0.01.$$

- Da $n = 5$ gilt, hat man mit der t-Verteilung von 4 Freiheitsgraden zu arbeiten. Entsprechend Schritt 2 aus obigem Algorithmus hat man die Gleichung

$$F_4(z) = 1 - \frac{\alpha}{2} = 1 - 0.005 = 0.995$$

zu lösen. Offenbar gilt

$$z = F_4^{-1}(0.995).$$

Da in Tabelle (79) Spalte 5 zu 0.995 und Zeile 4 zum Freiheitsgrad 4 jeweils gehört, liefert der Eintrag an der Stelle $(4, 5)$ das gewünschte z, nämlich

$$z = 4.604.$$

- Entsprechend Schritt 3 berechnen wir

$$\begin{aligned}
\overline{\boldsymbol{x}} &= \frac{1}{5}(17.21 + 17.22 + 17.22 + 17.20 + 17.19) \\
&= 17.208 \\
s &= \sqrt{\frac{(17.21-\overline{\boldsymbol{x}})^2 + (17.22-\overline{\boldsymbol{x}})^2 + (17.22-\overline{\boldsymbol{x}})^2 + (17.20-\overline{\boldsymbol{x}})^2 + (17.19-\overline{\boldsymbol{x}})^2}{4}} \\
&= 0.01304.
\end{aligned}$$

- Entsprechend Schritt 4 bekommen wir

$$c = \frac{sz}{\sqrt{n}} = \frac{0.01304 \times 4.604}{\sqrt{5}} = 0.02685.$$

Antwort: Mit einer Wahrscheinlichkeit von 99 % überdeckt das Intervall

$$[17.208 - 0.02685, 17.208 + 0.02685]$$

den Erwartungswert μ der betrachteten Grundgesamtheit.

Erwartungsgemäß ist dieses Intervall etwas breiter als das Intervall aus Beispiel 38, da man mit weniger Information über die Grundgesamtheit auskommen muß und so die Schätzung unter Beibehaltung des Vertrauensniveaus etwas ungenauer ausfällt.

Das **mathcad**-Programm

$$\overline{\boldsymbol{x}} := \mathrm{mittelwert}(\boldsymbol{x}), \quad z := \mathrm{qt}(0.995, 4), \quad s := \mathrm{Stdev}(\boldsymbol{x}), \quad c := \frac{sz}{\sqrt{5}}$$

liefert den Mittelwert $\overline{\boldsymbol{x}}$ und die gewünschte Konstante c. $\diamond$

3.4.3 Vertrauensintervall für $\boldsymbol{\sigma}$ einer entsprechend $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ verteilten Zg

[5] S.391, [13] S.524, [15] S.160

Unter der Voraussetzung $X_k \sim \mathscr{N}(\mu, \sigma^2)$ bilden wir die Zg

$$U := \frac{1}{\sigma^2} \sum_{k=1}^{n} (X_k - \overline{X})^2, \quad \overline{X} := \frac{1}{n} \sum_{k=1}^{n} X_k.$$

Dann ist U die Summe von $n-1$ Quadraten standardnormalverteilter Zg. Nimmt man noch deren Unabhängigkeit an, so gehorcht U der χ^2-Verteilung mit $n-1$ Freiheitsgraden:

$$\boxed{U \sim \ \chi^2\text{-Verteilung mit } n-1 \text{ Freiheitsgraden}}$$

Um diese Feststellung plausibel zu machen, berechnen wir U für $n=2$. Offenbar gilt

$$\begin{aligned} U = \frac{1}{\sigma^2} \sum_{k=1}^{2} (X_k - \overline{X})^2 &= \frac{1}{\sigma^2} \left(\left(X_1 - \frac{1}{2}(X_1 + X_2) \right)^2 + \left(X_2 - \frac{1}{2}(X_1 + X_2) \right)^2 \right) \\ &= \frac{1}{\sigma^2} \left(\left(\frac{1}{2}(X_1 - X_2) \right)^2 + \left(\frac{1}{2}(X_2 - X_1) \right)^2 \right) \\ &= \left(\frac{X_1 - X_2}{\sqrt{2}\sigma} \right)^2. \end{aligned}$$

Aufgrund der Unabhängigkeit von X_1 und X_2 liefert die Anwendung des Additionssatzes die Normalverteilung von

$$\frac{X_1 - X_2}{\sqrt{2}\sigma}.$$

Es bleibt zu klären, wie deren Parameter aussehen. Aus $EX_1 = EX_2$ folgt

$$E\left(\frac{X_1 - X_2}{\sqrt{2}\sigma} \right) = 0.$$

Aufgrund der Unabhängigkeit von X_1 und X_2 gilt

$$\operatorname{Var}\left(\frac{X_1 - X_2}{\sqrt{2}\sigma} \right) = \frac{1}{2\sigma^2} \left(\operatorname{Var}(X_1) + \operatorname{Var}(X_2) \right) = \frac{2\sigma^2}{2\sigma^2} = 1.$$

Demnach ist für $n=2$ die Zg U das Quadrat einer standard-normalverteilten Zg, d.h. U gehorcht der χ^2-Verteilung zum Freiheitsgrad 1.

Für $n=3$ beobachtet man

$$U = \frac{1}{\sigma^2} \sum_{k=1}^{3} (X_k - \overline{X})^2 = \left(\frac{X_1 - X_2}{\sigma\sqrt{2}} \right)^2 + \left(\frac{X_1 + X_2 - 2X_3}{\sigma\sqrt{6}} \right)^2.$$

Wie kommt man auf diese Darstellung? Diese Frage kann man mittels der Transformation einer quadratischen Form auf Hauptachsen beantworten. Ausmultiplizieren von $\sum_{k=1}^{3}(X_k - \overline{X})^2$ führt auf

$$\sigma^2 U = \frac{2}{3}(X_1^2 + X_2^2 + X_3^2 - X_1X_2 - X_2X_3 - X_1X_3).$$

Mittels der Matrix

$$Q := \frac{1}{2}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

kann man daher

$$\sigma^2 U = \frac{2}{3}\begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix} Q \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

schreiben. Man prüft leicht nach, daß die Eigenwerte von Q durch $\{0, 3/2, 3/2\}$ gegeben sind. Da ein Eigenwert verschwindet, läßt sich U als Summe zweier Quadrate schreiben, was dazu führt, daß U nur mit zwei Freiheitsgraden verteilt ist. Die Transformation von Q auf Hauptachsen geht einher mit der Faktorisierung

$$Q = TDT^{\mathrm{T}}, \quad D := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3/2 & 0 \\ 0 & 0 & 3/2 \end{bmatrix}, \quad T := \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{bmatrix}$$

wobei sich die Diagonalmatrix D aus den Eigenwerten von Q und die Spalten von T sich aus den zugehörigen orthonormierten Eigenvektoren ergeben. Ersetzung von Q durch TDT^{T} führt auf

$$\begin{aligned} \sigma^2 U &= [X_1\ X_2\ X_3] \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-2}{\sqrt{6}} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \\ &= \left(\frac{X_1 - X_2}{\sqrt{2}}\right)^2 + \left(\frac{X_1 + X_2 - 2X_3}{\sqrt{6}}\right)^2 \end{aligned}$$

d.h. U gestattet tatsächlich die Darstellung als Summe zweier quadrierter Zg, wobei wir von der ersten Zg schon wissen, daß sie standardnormalverteilt ist. Für die zweite Zg haben wir

$$E(X_1 + X_2 - 2X_3) = 0$$

$$\mathrm{Var}\left(\frac{X_1 + X_2 - 2X_3}{\sigma\sqrt{6}}\right) = \frac{1}{6\sigma^2}(\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 4\mathrm{Var}(X_3)) = 1$$

d.h. auch die zweite Zg ist standard-normalverteilt. Demnach gehorcht für $n = 3$ die Zg U einer χ^2-Verteilung mit 2 Freiheitsgraden.

Setzen wir

$$F_n(x) := \int_{-\infty}^{x} f_n(t)\mathrm{d}t$$

wobei f_n die Dichtefunktion der χ^2-Verteilung mit n Freiheitsgraden bezeichnet, so gilt

$$P(c_1 \le U \le c_2) = F_{n-1}(c_2) - F_{n-1}(c_1).$$

In Hinblick auf die Definition (41) der Stichprobenvarianz S^2 stehen U und S^2 im Verhältnis

$$U = \frac{n-1}{\sigma^2} S^2$$

sodaß

$$\begin{aligned} F_{n-1}(c_2) - F_{n-1}(c_1) &= P\left(c_1 \le \frac{n-1}{\sigma^2} S^2 \le c_2\right) \\ &= P\left(\frac{c_1}{(n-1)S^2} \le \frac{1}{\sigma^2} \le \frac{c_2}{(n-1)S^2}\right) \\ &= P\left(\frac{(n-1)S^2}{c_2} \le \sigma^2 \le \frac{(n-1)S^2}{c_1}\right). \end{aligned}$$

Daher überdecken die Realisierungen des Zufallsvektors

$$\left[\sqrt{\frac{n-1}{c_2}} S, \sqrt{\frac{n-1}{c_1}} S\right]$$

den unbekannten Parameter σ mit der Wahrscheinlichkeit $F_{n-1}(c_2) - F_{n-1}(c_1)$.

Es verbleibt, c_1 und c_2 so zu bestimmen, daß

$$F_{n-1}(c_2) - F_{n-1}(c_1) = 1 - \alpha$$

für ein vorgegebenes $\alpha \in [0, 1]$ gilt. Da diese Gleichung zwei Unbekannte hat, ist die Lösung nicht eindeutig. Um zu einem Algorithmus für die Bestimmung eines Konfidenzintervalls für σ zu kommen, ersetzt man diese unterbestimmte Gleichung durch das Gleichungssystem

$$F_{n-1}(c_1) = \frac{\alpha}{2}, \quad F_{n-1}(c_2) = 1 - \frac{\alpha}{2}.$$

Eine geometrische Interpretation liefert Abbildung 23.

Algorithmus zur Berechnung eines Vertrauensintervalls für σ einer normalverteilten Zg

- Wähle $\alpha \in (0, 1)$.
- Bestimme c_1 und c_2 aus den Gleichungen

$$F_{n-1}(c_1) = \frac{\alpha}{2}, \quad F_{n-1}(c_2) = 1 - \frac{\alpha}{2}$$

wobei F_n die Verteilungsfunktion der χ^2-Verteilung mit n Freiheitsgraden darstellt. Falls

$$\alpha \in \{0.01, 0.05, 0.10\}, \quad n \in \{3, 4, 5, 10, 15, 30\}$$

kann dazu Tabelle (45) benutzt werden. Die Zahlen c_1 und c_2 werden auch oft mit $\chi^2_{\frac{\alpha}{2}}(n-1)$ und $\chi^2_{1-\frac{\alpha}{2}}(n-1)$ bezeichnet.

- Wähle eine Stichprobe $\boldsymbol{x} := [x_k]_{k=1}^n$ im Umfang n und bilde die Stichprobenvarianz

$$s^2 := \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{\boldsymbol{x}})^2, \quad \overline{\boldsymbol{x}} := \frac{1}{n}\sum_{k=1}^{n} x_k. \tag{44}$$

Statistische Aussage: $(1-\alpha)\%$ aller so konstruierten Intervalle

$$\left[\sqrt{\frac{n-1}{c_2}}s, \sqrt{\frac{n-1}{c_1}}s\right]$$

überdecken σ.

Die folgenden Tabellen geben die Konstanten c_1 und c_2 für die Irrtumswahrscheinlichkeiten

$$\alpha = 0.01, \quad 0.05, \quad 0.10$$

und die Stichprobenumfänge

$$n = 3, 4, 5, 10, 15, 30$$

an:

$\boldsymbol{c}_1$ $\left(F_{n-1}(c_1) = \frac{\alpha}{2}\right)$

$\alpha \setminus n-1$	2	3	4	9	14	29
0.01	00.010	00.072	00.207	01.735	04.075	13.121
0.05	00.051	00.216	00.484	02.700	05.629	16.047
0.10	00.103	00.352	00.711	03.325	06.571	17.708

$\boldsymbol{c}_2$ $\left(F_{n-1}(c_2) = 1 - \frac{\alpha}{2}\right)$

$\alpha \setminus n-1$	2	3	4	9	14	29
0.01	10.597	12.838	14.860	23.589	31.319	52.336
0.05	07.378	09.348	11.143	19.023	26.119	45.722
0.10	05.991	07.815	09.488	16.919	23.685	42.557

(45)

Vgl. Tabelle A.3 in [15], S. 181 und Tabelle 0.4.6.4 in [20], S. 93. F_n bezeichnet die Verteilungsfunktion der χ^2-Verteilung mit n Freiheitsgraden.

Ablesebeispiel: Für $\alpha = 0.05$ und $n = 15$ hat man jeweils den Wert aus der zweiten Zeile und der fünften Spalte zu verwenden

$$c_1 = 5.629, \quad c_2 = 26.119.$$

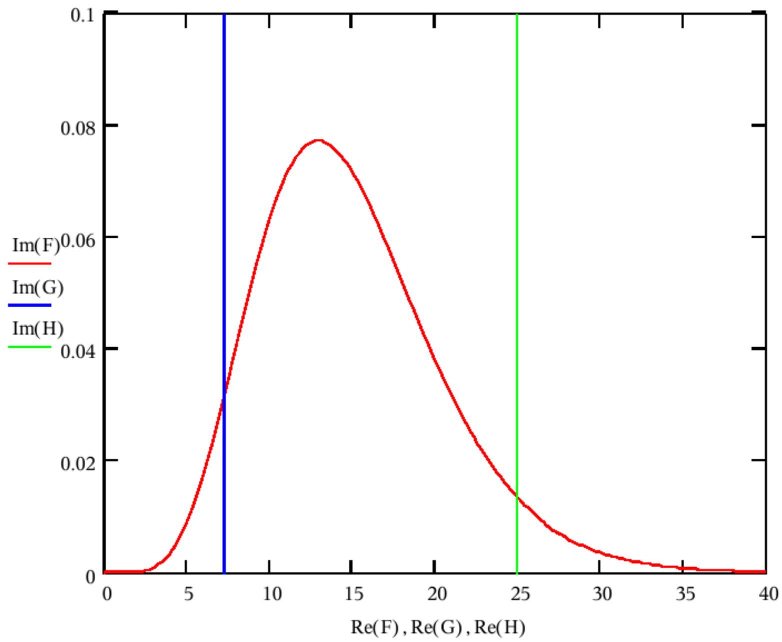


Abbildung 23: Die 'Glockenkurve' F (rot) gehört zur Dichtefunktion der χ^2-Verteilung für 15 Freiheitsgrade. Die Konstanten c_1 und c_2 erfüllen $F_{15}(c_1) = \alpha/2$ und $F_{15}(c_2) = 1-\alpha/2$ für $\alpha = 0.1$. Die Geraden G (blau) und H (grün) schneiden die x-Achse in $c_1 = 7.261$ bzw. $c_2 = 24.996$. Der Inhalt der von F,G,H und der x-Achse eingeschlossenen Fläche ist gleich $1-\alpha$, d.h. sie nimmt 90 % der gesamten Fläche unter F ein. Die Inhalte der sich links und rechts anschließenden Flächen unterhalb von F sind jeweils gleich $\alpha/2$, d.h. sie umfassen jeweils 5 % der Gesamtfläche. Dabei ist die linke Fläche beschränkt und die rechte nicht.

Die Tabellen (45) wurden mit dem **mathcad**-Befehl

$$\mathrm{qchisq}(x, f)$$

erzeugt, wobei für x der Reihe nach die Werte $\alpha/2$ und $1-\alpha/2$ und für f die Werte $2, 3, 4, 9, 14, 29$ eingesetzt wurden. Da $F_n(x)$ streng monoton wachsend ist, wachsen in jeder Spalte die Werte für c_1 an, wohingegen die Werte für c_2 in jeder Spalte fallen.

Beispiel 41 Vertrauensintervall für σ, wobei sich $X \sim \mathcal{N}(\mu, \sigma^2)$ aus der Messung von elektrischen Widerständen ergibt

Die Messung von 10 Widerständen, die rein zufällig aus einem großen Sortiment von Widerständen entnommen wurden, liefert folgendes Messprotokoll:

k	1	2	3	4	5	6	7	8	9	10
x_k	100	104	98	96	101	104	98	99	94	101

Es wird vorausgesetzt, daß die Grundgesamtheit normalverteilt ist. Es soll ein Vertrauensintervall für σ mit dem Vertrauensniveau 0.9 angegeben werden.

- Offenbar ist die Stichprobe $\boldsymbol{x}$ von der Größe $n := 10$. Da ein Vertrauensniveau von 0.9 gelten soll, ergibt sich
$$\alpha := 0.1.$$
- Damit sind die Konstanten c_1 und c_2 eindeutig bestimmt.

  a) Wegen $n - 1 = 9$ hat man Spalte '9' und

  b) wegen $\alpha = 0.1$ hat man Zeile 3

  in den Tabellen (45) zu verwenden, d.h. man liest
$$c_1 = 3.325, \quad c_2 = 16.919$$
  ab.
- Es verbleibt die Berechnung von s:
$$s = \sqrt{\frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{\boldsymbol{x}})^2} = 3.206, \quad \overline{\boldsymbol{x}} = 99.5.$$

Daher ergeben sich die Grenzen des Vertrauensintervalls für σ entsprechend
$$\sqrt{\frac{9}{16.919}}3.206 = 2.338, \qquad \sqrt{\frac{9}{3.325}}3.206 = 5.274$$
d.h. mit einer Wahrscheinlichkeit von 90 % überdeckt das Intervall
$$[2.338, 5.274]$$
die Varianz aller betrachteten Widerstände. $\diamond$

Abhängigkeit des Vertrauensintervalls von der Variation von α:

Eine Vergrößerung des Vertrauensniveaus geht einher mit der Verkleinerung von α, was eine Verkleinerung von c_1 und eine Vergrößerung von c_2 nachsichzieht. Eine Vergrößerung von c_2 bedeutet ein Nachlinksrücken der linken Intervallseite und eine Verkleinerung von c_1 ein Nachrechtsrücken der rechten Intervallseite, sodaß das Vertrauensintervall größer wird.

Hinsichtlich des Einflußes von n auf das Vertrauensintervall siehe Abschnitt 3.8.8, der sich mit der Berechnung von Stichprobenumfängen beschäftigt.

3.4.4 Näherungsweises Vertrauensintervall für einen unbekannten Anteilswert p

Wir suchen ein Vertrauensintervall, das mit einer vorgegebenen Vertrauenswahrscheinlichkeit die Wahrscheinlichkeit

$$p \in (0,1)$$

überdeckt, mit der ein aus einer Grundgesamtheit beliebig ausgewähltes Element eine bestimmte Eigenschaft besitzt. Die Grundgesamtheit G kann eine Menge von Gegenständen sein, die man nach funktionstüchtig und defekt unterscheidet. Falls die Menge endlich ist, d.h. aus N Elementen besteht, ist p offenbar gleich zur Anzahl

- aller funktionstüchtigen
- oder aller defekten

Elemente dividiert durch N, jenachdem wie man die interessierende Eigenschaft spezifiziert. Falls man in der Lage ist, alle Elemente von G zu überprüfen, kann man p exakt berechnen. Falls man nur eine Stichprobe $\mathcal{S}$ im Umfang n überprüfen kann, so bekommt man ein mit einem $(0,1)$-Vektor $\boldsymbol{x}$ der Dimension n identifizierbares Prüfprotokoll, welches entsprechend Abschnitt 3.2.1 zur Punktschätzung von p benutzt werden kann: Man bestimmt die Anzahl der Einsen oder der Nullen und dividiert durch den Stichprobenumfang n. Der resultierende Quotient $\widehat{p}$ stimmt nach dem Gesetz der großen Zahl von Bernoulli in etwa mit p überein. Modelliert man die interessierende Eigenschaft durch 1, so gilt

$$\widehat{p} = \frac{k(\boldsymbol{x})}{n}, \quad k(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}.$$

Wegen der Unabhängigkeit der Ziehung der Komponenten von $\mathcal{S}$ für großes N, ist die Zg

$$X := \text{'Anzahl der Einsen'}$$

binomialverteilt mit den Parametern n und p. Konsequenterweise gilt nach dem zentralen Grenzwertsatz für große Stichprobenumfänge n

$$X \sim \mathcal{N}(np, np(1-p)).$$

Mit dem Linearitätssatz bekommt man

$$H := \frac{1}{n}X \sim \mathcal{N}(\mu, \sigma^2), \quad \mu := p, \quad \sigma := \sqrt{\frac{p(1-p)}{n}}.$$

Wir suchen das Vertrauensintervall in der Form

$$[\widehat{p} - c, \widehat{p} + c], \quad c \in \mathbb{R}_+$$

d.h. wir bestimmen c so, daß

$$P(H - c \leq \mu \leq H + c) = 1 - \alpha.$$

Hier ist zu beachten, daß bei Ziehen einer Stichprobe die Zg H in den Schätzwert $\widehat{p}$ übergeht. Offenbar gilt die Äquivalenz

$$H - c \leq \mu \leq H + c \quad \Leftrightarrow \quad \frac{H-\mu}{\sigma} - \frac{c}{\sigma} \leq 0 \leq \frac{H-\mu}{\sigma} + \frac{c}{\sigma}$$

so, daß wir für c die Gleichung

$$1 - \alpha = P\left(\frac{H-\mu}{\sigma} - \frac{c}{\sigma} \leq 0 \leq \frac{H-\mu}{\sigma} + \frac{c}{\sigma}\right)$$

bekommen. Setzen wir

$$Y := \frac{H-\mu}{\sigma}, \quad z := \frac{c}{\sigma}$$

so ist wegen $H \sim \mathscr{N}(\mu, \sigma^2)$ die Zg Y standard-normalverteilt und wir bekommen die Gleichungen

$$\begin{aligned} 1-\alpha &= P(Y - z \leq 0 \leq Y + z) = P(-z \leq Y \leq z) \\ &= \Phi_{0,1}(z) - \Phi_{0,1}(-z) = \Phi_{0,1}(z) - (1 - \Phi_{0,1}(z)) = 2\Phi_{0,1}(z) - 1 \end{aligned}$$

d.h.

$$1 - \frac{\alpha}{2} = \Phi_{0,1}(z).$$

Ersetzt man schließlich in der Definition von σ den zu schätzenden Parameter p durch die Punktschätzung $\widehat{p}$, so ergibt sich folgende Konstruktionsvorschrift für ein Vertrauensintervall für p.

Algorithmus zur Berechnung eines näherungsweisen Vertrauensintervalls für einen unbekannten Anteilswert p, [15] S.152

- Wähle $\alpha \in (0, 1)$.
- Bestimme z aus der Gleichung

  $$1 - \frac{\alpha}{2} = \Phi_{0,1}(z).$$

  Für die praxisrelevanten $\alpha \in \{0.01, 0.05, 0.1\}$ kann die Tabelle

α	0.010	0.050	0.100
z	2.576	1.960	1.645

  benutzt werden.
- Wähle eine Stichprobe $\mathcal{S}$ im Umfang n und bilde den Quotienten $\widehat{p}$ aus der Anzahl der Elemente von $\mathcal{S}$, die eine bestimmte Eigenschaft besitzen, und n.
- Berechne $c := z\sqrt{\dfrac{\widehat{p}(1-\widehat{p})}{n}}$.

Statistische Aussage: Das Intervall

$$[\widehat{p} - c, \widehat{p} + c]$$

überdeckt p mit einer Wahrscheinlichkeit von $1 - \alpha$.

Beispiel 42 Anteil der Autofahrer, die sich um Geschwindigkeitsbegrenzungen nicht kümmern, [15], S.153.

Eine Stichprobe im Umfang von $n := 1000$ Pkw, die eine Radarfalle durchfuhren, ergab, daß 135 Fahrzeuge die erlaubte Höchstgeschwindigkeit überschritten. Man berechne ein Vertrauensintervall für den Anteil p der Raser zu einem 95 %-tigen Vertrauensniveau.

- Da mit einem 95 %-tigen Vertrauensniveau gearbeitet werden soll, ergibt sich

$$\alpha := 0.05.$$

- Demnach gilt $z = 1.96$.
- Lt. Aufgabenstellung hat man als Punktschätzung für p

$$\widehat{p} := \frac{135}{1000} = 0.135.$$

- Schließlich bekommen wir für die Konstante c

$$c := z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.96\sqrt{\frac{0.135(1-0.135)}{1000}} = 0.0212.$$

Näherungsweise überdeckt das Intervall

$$[0.135 - 0.0212, 0.135 + 0.0212]$$

mit einer Wahrscheinlichkeit von 0.95 den wahren Anteil p der Raser, d.h. mit 95 % Sicherheit liegt der wahre Anteil der Raser zwischen 11.38 % und 15.62 %, wenn man den Fehler, den man bei der Berechnung von c durch Ersetzung von p durch die Punktschätzung $\widehat{p}$ macht, nicht weiter berücksichtigt. $\diamond$

3.5 Lineare Regressionsrechnung

3.5.1 Kleinste Quadrate Lösungen überbestimmter linearer Gleichungssysteme

Gegeben sei eine Matrix $A \in \mathbb{R}^{n \times m}$ und ein Vektor $b \in \mathbb{R}^n$, wobei $m \leq n$. Gesucht ist ein Vektor $v \in \mathbb{R}^m$, der das lineare Gleichungssystem

$$Ax = b \tag{46}$$

löst. Da die Anzahl n der Gleichungen die Anzahl m der Variablen übersteigt, wird (46) im allgemeinen keine Lösung besitzen. Wie üblich definiert man die **Euklidische Länge** eines Vektors v durch

$$\|v\| := \sqrt{\sum_{k=1}^{m} v_k^2} = \sqrt{v^{\mathrm{T}} v}.$$

Da es im allgemeinen keinen Vektor $v \in \mathbb{R}^m$ gibt, für den die Differenz

$$Av - b$$

der Nullvektor wird, versucht man v so zu bestimmen, daß die Euklidische Länge der Differenz $Av - b$ möglichst klein wird, d.h. man sucht das Minimum der Funktion

$$\|Ax - b\| : \mathbb{R}^n \mapsto \mathbb{R}_+.$$

Offenbar wird $\|Ax - b\|$ genau dann minimal, wenn $\|Ax - b\|^2$ minimal wird. Daher braucht man nur die Funktion

$$\begin{aligned} G(x) &:= \|Ax - b\|^2 = (Ax - b)^{\mathrm{T}}(Ax - b) = (x^{\mathrm{T}} A^{\mathrm{T}} - b^{\mathrm{T}})(Ax - b) \\ &= x^{\mathrm{T}} A^{\mathrm{T}} A x - 2x^{\mathrm{T}} A^{\mathrm{T}} b + b^{\mathrm{T}} b \end{aligned} \tag{47}$$

zu minimieren. Notwendig dafür, daß G in v ein Minimum annimmt, ist die Lösungseigenschaft v bezüglich

$$\frac{\partial G(x)}{\partial x_k} = 0, \quad k = 1, \ldots, m.$$

Die Berechnung der partiellen Ableitungen liefert für v das lineare $(m \times m)$-Gleichungssystem

$$A^{\mathrm{T}} A x = A^{\mathrm{T}} b.$$

Unter der Voraussetzung, daß die Spalten von A linear unabhängig sind, wird die Matrix $A^{\mathrm{T}} A$ invertierbar, so daß an der Stelle

$$v := (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} b$$

alle partiellen Ableitungen von G verschwinden. Einsetzen von v in G zeigt, daß G in v ein Minimum annimmt. Die Matrix

$$A^{\dagger} := (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}}$$

heißt **Moore-Penrose Inverse** von A. Sie ist eine verallgemeinerte Inverse von A und geht für quadratisches A in die gewöhnliche Inverse über.

3.5.2 Regressionsgerade für ebene Punktwolke

Gegeben sei eine Punktwolke

$$W := \{(x_k, y_k) : k = 1, \ldots, n\} \subset \mathbb{R}^2 \tag{48}$$

bestehend aus n Punkten in der Ebene. Gesucht ist eine Gerade

$$y(x) := ax + b, \quad a, b \in \mathbb{R} \tag{49}$$

die möglichst 'gut' durch W hindurchführt. Möglichst 'gut' soll heißen

$$y_k \approx y(x_k), \quad k = 1, \ldots, n$$

was sich dahingehend verschärfen läßt, daß man die Koeffizienten a und b so wählt, daß

$$G(a, b) := \sum_{k=1}^{n} (y(x_k) - y_k)^2$$

minimal wird. Die Gerade (49) heißt dann die **Regressionsgerade** für W. Setzt man

$$A := \left[\begin{array}{cc} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{array} \right], \quad \boldsymbol{y} := [y_k]_{k=1}^n$$

so gilt

$$G(a, b) = \left\| A \left[\begin{array}{c} a \\ b \end{array} \right] - \boldsymbol{y} \right\|^2$$

d.h. $G(a, b)$ ist von der Gestalt (47), wobei $m = 2$. Demzufolge ergeben sich die gesuchten Koeffizienten entsprechend

$$\left[\begin{array}{c} a \\ b \end{array} \right] = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} \boldsymbol{y}. \tag{50}$$

Die Auswertung der rechten Seite liefert unter Benutzung von

$$\mathbf{1} := [1, \ldots, 1]^{\mathrm{T}}, \quad \boldsymbol{x} := [x_k]_{k=1}^n, \quad A = \left[\begin{array}{cc} \boldsymbol{x} & \mathbf{1} \end{array} \right]$$

die Gleichungen

$$A^{\mathrm{T}} A = \left[\begin{array}{cc} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} & \boldsymbol{x}^{\mathrm{T}} \mathbf{1} \\ \mathbf{1}^{\mathrm{T}} \boldsymbol{x} & n \end{array} \right], \quad A^{\mathrm{T}} \boldsymbol{y} = \left[\begin{array}{c} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y} \\ \mathbf{1}^{\mathrm{T}} \boldsymbol{y} \end{array} \right].$$

Wegen

$$\left[\begin{array}{cc} a & b \\ c & d \end{array} \right]^{-1} = \frac{1}{ad - cb} \left[\begin{array}{rr} d & -b \\ -c & a \end{array} \right]$$

bekommt man

$$(A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} \boldsymbol{y} = \frac{1}{n \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} - (\boldsymbol{x}^{\mathrm{T}} \mathbf{1})^2} \left[\begin{array}{cc} n & -\boldsymbol{x}^{\mathrm{T}} \mathbf{1} \\ -\mathbf{1}^{\mathrm{T}} \boldsymbol{x} & \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} \end{array} \right] \left[\begin{array}{c} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y} \\ \mathbf{1}^{\mathrm{T}} \boldsymbol{y} \end{array} \right]$$

d.h.

$$\left[\begin{array}{c} a \\ b \end{array} \right] = \frac{1}{n \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} - (\boldsymbol{x}^{\mathrm{T}} \mathbf{1})^2} \left[\begin{array}{c} n \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y} - (\boldsymbol{x}^{\mathrm{T}} \mathbf{1})(\mathbf{1}^{\mathrm{T}} \boldsymbol{y}) \\ \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} (\mathbf{1}^{\mathrm{T}} \boldsymbol{y}) - \mathbf{1}^{\mathrm{T}} \boldsymbol{x} (\boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}) \end{array} \right].$$

Wegen $\mathbf{1}^{\mathrm{T}}\boldsymbol{x} = n\overline{\boldsymbol{x}}$ und $\mathbf{1}^{\mathrm{T}}\boldsymbol{y} = n\overline{\boldsymbol{y}}$ kann man auch

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - n^2\overline{\boldsymbol{x}}^2} \begin{bmatrix} n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} - n^2\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}} \\ n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\overline{\boldsymbol{y}} - n\overline{\boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}) \end{bmatrix} = \frac{1}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - n\overline{\boldsymbol{x}}^2} \begin{bmatrix} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}} \\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\overline{\boldsymbol{y}} - \overline{\boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}) \end{bmatrix}$$

schreiben, d.h. der Anstieg a und das Absolutglied b der Regressionsgeraden sind jeweils durch

$$\boxed{a := \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - n\overline{\boldsymbol{x}}^2} = \frac{\sum_{k=1}^{n} x_k y_k - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}}{\sum_{k=1}^{n} x_k^2 - n\overline{\boldsymbol{x}}^2}} \tag{51}$$

$$\boxed{b := \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\overline{\boldsymbol{y}} - \overline{\boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y})}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - n\overline{\boldsymbol{x}}^2} = \frac{\overline{\boldsymbol{y}}\sum_{k=1}^{n} x_k^2 - \overline{\boldsymbol{x}}\sum_{k=1}^{n} x_k y_k}{\sum_{k=1}^{n} x_k^2 - n\overline{\boldsymbol{x}}^2}} \tag{52}$$

gegeben. Falls a schon bekannt ist, kann b entsprechend

$$\boxed{b = \overline{\boldsymbol{y}} - a\overline{\boldsymbol{x}}}$$

berechnet werden.

Beispiel 43 Wir wollen ein überschaubares Beispiel für die Anwendung der Methode der kleinsten Quadrate geben. Gesucht ist die Regressionsgerade für die dreielementige Punktwolke, die durch

$$\boldsymbol{x} = [1, 2, 3]^{\mathrm{T}}, \quad \boldsymbol{y} = [2, 4, 3]^{\mathrm{T}}$$

gegeben ist.

Es gilt

$$\overline{\boldsymbol{x}} = 2, \quad \overline{\boldsymbol{y}} = 3, \quad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 19, \quad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 14$$

sodaß

$$a = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - n\overline{\boldsymbol{x}}^2} = \frac{19 - 3 \times 2 \times 3}{14 - 3 \times 4} = \frac{1}{2}, \quad b = \overline{\boldsymbol{y}} - a\overline{\boldsymbol{x}} = 3 - \frac{1}{2}2 = 2.$$

Die Regressionsgerade lautet schließlich

$$y = \frac{1}{2}x + 2.$$

◇

Für Berechnungszwecke benutzt man am einfachsten (50).

Um die Approximationseigenschaften von $y = ax + b$ zu messen, kann man den **Residualstandardfehler** gemäß

$$S_u := \sqrt{\frac{1}{n}\sum_{k=1}^{n}(ax_k + b - y_k)^2} = \frac{1}{\sqrt{n}}\|a\boldsymbol{x} + b\mathbf{1} - \boldsymbol{y}\|$$

einführen. (Residual: 'als Rest zurückbleibend')

Die Formeln (51) und (52) zeigen den Zusammenhang zwischen den Koeffizienten der Regressionsgeraden $y(x)$ und der Kovarianz der Punktwolke W.

3.5.3 Kovarianz ebener Punktwolke

Die **Kovarianz** von

$$W := \{(x_k, y_k) : k = 1, \ldots, n\} \subset \mathbb{R}^2$$

ist durch die Zahl

$$\boxed{s_{\boldsymbol{xy}} := \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{\boldsymbol{x}})(y_k - \overline{\boldsymbol{y}}), \quad \overline{\boldsymbol{x}} := \frac{1}{n}\sum_{k=1}^{n} x_k, \quad \overline{\boldsymbol{y}} := \frac{1}{n}\sum_{k=1}^{n} y_k}$$

erklärt ([15] S.49). Wegen

$$s_{\boldsymbol{xx}} = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{\boldsymbol{x}})^2$$

stimmt $s_{\boldsymbol{xx}}$ mit einer Realisierung der wie in (41) definierten Zg 'Stichprobenvarianz' S^2 überein.

Für den Zusammenhang zur Regressionsgeraden stellen wir

$$\begin{aligned}
s_{\boldsymbol{xy}} &= \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{\boldsymbol{x}})(y_k - \overline{\boldsymbol{y}}) \\
&= \frac{1}{n-1}\sum_{k=1}^{n}(x_k y_k - \overline{\boldsymbol{x}} y_k - \overline{\boldsymbol{y}} x_k + \overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}) \\
&= \frac{1}{n-1}\left(\sum_{k=1}^{n} x_k y_k - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}} - n\overline{\boldsymbol{y}}\,\overline{\boldsymbol{x}} + n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}\right) \\
&= \frac{1}{n-1}\left(\sum_{k=1}^{n} x_k y_k - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}\right)
\end{aligned}$$

fest. Daher gilt

$$a = \frac{\sum_{k=1}^{n} x_k y_k - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}}{\sum_{k=1}^{n} x_k^2 - n\overline{\boldsymbol{x}}^2} = \frac{\frac{1}{n-1}\left(\sum_{k=1}^{n} x_k y_k - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}\right)}{\frac{1}{n-1}\left(\sum_{k=1}^{n} x_k^2 - n\overline{\boldsymbol{x}}^2\right)} = \frac{s_{\boldsymbol{xy}}}{s_{\boldsymbol{xx}}}.$$

Demnach stehen der Anstieg a der Regressionsgeraden von W und ihre Kovarianz s_{xy} in der Beziehung

$$\boxed{a = \frac{s_{\boldsymbol{xy}}}{s_{\boldsymbol{xx}}}}$$

Falls man $s_{\boldsymbol{xy}}$ berechnen möchte, verwendet man

$$s_{\boldsymbol{xy}} = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}}{n-1}.$$

Die Definitionsgleichung

$$s_{\boldsymbol{xy}} := \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{\boldsymbol{x}})(y_k - \overline{\boldsymbol{y}})$$

gestattet eine geometrische Interpretation von $s_{\boldsymbol{xy}}$. Man legt den Koordinatenursprung eines Koordinatensystems für $\mathbb{R}^2$ in den Punkt $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ und betrachtet die Anzahl der Punkte (x_k, y_k) in jedem Quadranten. Dann gilt folgendes.

- Der Punkt $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}) \in \mathbb{R}^2$ ist der Schwerpunkt der Punktwolke W.
- Falls $s_{\boldsymbol{xy}}$ positiv, liegen verhältnismäßig viele Punkte in $Q_{\mathrm{I}} \cup Q_{\mathrm{III}}$.
- Falls $s_{\boldsymbol{xy}}$ negativ, liegen verhältnismäßig viele Punkte in $Q_{\mathrm{II}} \cup Q_{\mathrm{IV}}$.
- Falls $s_{\boldsymbol{xy}}$ fast Null, liegen alle Punkte in $Q_{\mathrm{I}} \cup Q_{\mathrm{II}} \cup Q_{\mathrm{III}} \cup Q_{\mathrm{IV}}$ ziemlich gleichmäßig verteilt.

Dieser Zusammenhang rührt daher, daß ein Punkt aus $Q_{\mathrm{I}} \cup Q_{\mathrm{III}}$ positiv und ein Punkt aus $Q_{\mathrm{II}} \cup Q_{\mathrm{IV}}$ negativ zu $s_{\boldsymbol{xy}}$ beiträgt.

Beispiel 44 Interpretation Kovarianz $s_{\boldsymbol{xy}}$

Abbildung 20 wurde durch gleichverteilte $\boldsymbol{x} \in \mathbb{R}^{50}$ und $\boldsymbol{y} \in \mathbb{R}^{50}$ erzeugt. Es erscheinen in jedem Bild $n := 50$ rote Punkte. In allen drei Bildern wurde der Vektor $\boldsymbol{x}$ durch runif$(n, 1, 2)$ erzeugt, d.h. seine Komponenten liegen zwischen 1 und 2 gleichverteilt.

- Im linken Bild wurde der Vektor $\boldsymbol{y}$ durch

$$y_k := 2x_k + 1 + \varepsilon_k, \quad \varepsilon := \text{runif}(\text{n}, -1, 1)$$

erzeugt, d.h. y_k ergibt sich durch den gestörten Wert $2x_k + 1$, wobei die Störung gleichverteilt zwischen -1 und 1 liegt. Die Lage des blauen Kreuzes ergibt sich aus den Mittelwerten

$$\overline{\boldsymbol{x}} = 1.448, \quad \overline{\boldsymbol{y}} = 3.769.$$

Offenbar liegen die meisten Punkte im 1. und im 3. Quadranten, was zu

$$s_{\boldsymbol{xy}} = 0.206$$

passt, da die Regressionsgerade positiven Anstieg hat.

- Im mittleren Bild wurde der Vektor $\boldsymbol{y}$ durch

$$y_k := -2x_k + 1 + \varepsilon_k, \quad \varepsilon := \text{runif}(\text{n}, -1, 1)$$

erzeugt, d.h. y_k ergibt sich durch den gestörten Wert $-2x_k + 1$, wobei die Störung gleichverteilt zwischen -1 und 1 liegt. Die Lage des blauen Kreuzes ergibt sich aus den Mittelwerten

$$\overline{\boldsymbol{x}} = 1.448, \quad \overline{\boldsymbol{y}} = -2.022.$$

Offenbar liegen die meisten Punkte im 2. und im 4. Quadranten, was zu

$$s_{\boldsymbol{xy}} = -0.139$$

passt, da die Regressionsgerade negativen Anstieg hat.

- Im rechten Bild wurde der Vektor $\boldsymbol{y}$ durch

$$\boldsymbol{y} := \text{runif}(\text{n}, -1, 1)$$

erzeugt, d.h. $\boldsymbol{y}$ ergibt sich allein aus einer Störung, die gleichverteilt zwischen -1 und 1 ist. Die Lage des blauen Kreuzes ergibt sich aus den Mittelwerten

$$\overline{\boldsymbol{x}} = 1.448, \quad \overline{\boldsymbol{y}} = -0.126.$$

Offenbar liegen in allen Quadranten ungefähr gleich viele Punkte, was zu

$$s_{\boldsymbol{xy}} = 0.036$$

passt, da die Regressionsgerade in etwa parallel zur x-Achse verläuft (Anstieg 0).

(Beachte, daß in **mathcad** beim Befehl kvar$(\boldsymbol{x}, \boldsymbol{y})$ durch n anstatt durch $n-1$ geteilt wird.) $\diamond$

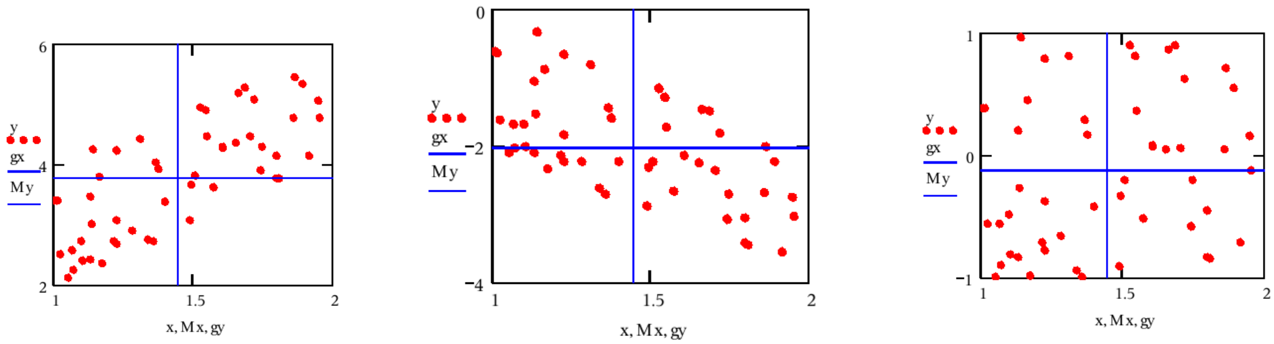


Abbildung 24: Es werden Punktwolken W mit den Kovarianzen 0.206, -0.139 und 0.036 gezeigt. Diese Darstellungen von W heißen auch **Streudiagramm** von $\boldsymbol{x}$ und $\boldsymbol{y}$.

3.5.4 Korrelationskoeffizient ebener Punktwolke

Zur Abschätzung der Kovarianz $s_{\boldsymbol{xy}}$ von $\boldsymbol{x}$ und $\boldsymbol{y}$ mittels der Varianzen $s_{\boldsymbol{xx}}$ und $s_{\boldsymbol{yy}}$ erinnern wir uns an die **Cauchy-Schwarzsche Ungleichung** (CSU) [7] S.237

$$|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}|^2 \le (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}).$$

Dabei wird die Gleichheit genau dann angenommen, wenn die Vektoren $\boldsymbol{x}$ und $\boldsymbol{y}$ linear abhängig sind, d.h. es gibt eine Zahl λ mit $\boldsymbol{x} = \lambda\boldsymbol{y}$ oder $\boldsymbol{y} = \lambda\boldsymbol{x}$.

Die Anwendung der CSU liefert

$$\left|\sum_{k=1}^{n}(x_k - \overline{\boldsymbol{x}})(y_k - \overline{\boldsymbol{y}})\right|^2 \le \sum_{k=1}^{n}(x_k - \overline{\boldsymbol{x}})^2 \sum_{k=1}^{n}(y_k - \overline{\boldsymbol{y}})^2$$

sodaß die Division mit $n-1$ auf

$$\boxed{|s_{\boldsymbol{xy}}|^2 \le s_{\boldsymbol{xx}} s_{\boldsymbol{yy}}}$$

führt. Der Quotient

$$\boxed{r_{\boldsymbol{xy}} := \frac{s_{\boldsymbol{xy}}}{s_{\boldsymbol{x}} s_{\boldsymbol{y}}}, \quad s_{\boldsymbol{x}} := \sqrt{s_{\boldsymbol{xx}}}, \quad s_{\boldsymbol{y}} := \sqrt{s_{\boldsymbol{yy}}}}$$

heißt der **Korrelationskoeffizient** von $\boldsymbol{x}$ und $\boldsymbol{y}$. Die Zahl $r_{\boldsymbol{xy}}$ wird zu Unterscheidungszwecken auch manchmal Pearsonscher Korrelationskoeffizient genannt (Karl Pearson, 1857-1936, ein Mitbegünder der modernen Statistik, sein Sohn Egon Pearson, 1895-1980, ebenfalls Statistiker).

Eigenschaften:

- Entsprechend der Folgerung aus der CSU gilt

$$|r_{\boldsymbol{xy}}| \le 1. \tag{53}$$

- Die Gleichheitsbedingung führt auf

  $$|s_{\boldsymbol{xy}}|^2 = s_{\boldsymbol{xx}} s_{\boldsymbol{yy}} \quad \Leftrightarrow \quad \exists a \in \mathbb{R} : \boldsymbol{y} - \overline{\boldsymbol{y}}\mathbf{1} = a(\boldsymbol{x} - \overline{\boldsymbol{x}}\mathbf{1}) \text{ oder } a(\boldsymbol{y} - \overline{\boldsymbol{y}}\mathbf{1}) = \boldsymbol{x} - \overline{\boldsymbol{x}}\mathbf{1}$$

  wobei $\mathbf{1} = [1, \ldots, 1]^{\mathrm{T}}$. Die Gleichung

  $$\boldsymbol{y} - \overline{\boldsymbol{y}}\mathbf{1} = a(\boldsymbol{x} - \overline{\boldsymbol{x}}\mathbf{1})$$

  bedeutet ausführlich

  $$y_k = ax_x + b, \quad b := \overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}.$$

  Demnach gilt $|r_{\boldsymbol{xy}}| = 1$ genau dann, wenn alle alle Punkte (x_k, y_k) auf einer Geraden liegen. [4]

- Falls $r_{\boldsymbol{xy}} = 1$, liegen alle (x_k, y_k) auf einer Geraden mit positivem Anstieg.

- Falls $r_{\boldsymbol{xy}} = -1$, liegen alle (x_k, y_k) auf einer Geraden mit negativem Anstieg.

3.5.5 Erwartungstreue Punktschätzer für die Koeffizienten der Regressionsgeraden

Wir gehen davon aus, daß zwischen zwei Größen x und y ein linearer Zusammenhang besteht, d.h. ihre Realisierungen (x_k, y_k) erfüllen die Gleichung

$$y_k = ax_k + b$$

für zwei Parameter $a, b \in \mathbb{R}$.

Wir gehen weiter davon aus, daß die Zahlen x_k exakt bekannt sind und die zugehörigen y_k nur mit einem gewissen Fehler entsprechend

$$\widehat{y}_k = y_k + \varepsilon_k$$

gemessen werden können. Es stellt sich nun die Frage, ob die Punktwolke

$$W := \{(x_1, \widehat{y}_1), \ldots, (x_n, \widehat{y}_n)\}$$

tauglich ist, die Parameter a und b zu schätzen.

Als Schätzer für den Geradenanstieg a bietet sich natürlich

$$\widehat{a} := \frac{s_{x\widehat{y}}}{s_{xx}}$$

an.

---

[4] Der Fall

$$a(\boldsymbol{y} - \overline{\boldsymbol{y}}\mathbf{1}) = \boldsymbol{x} - \overline{\boldsymbol{x}}\mathbf{1}, \quad a = 0$$

bedeutet $x_k = \overline{\boldsymbol{x}}$, $k = 1, \ldots, n$, d.h. alle x_k sind gleich, d.h. alle Punkte (x_k, y_k) liegen auf der Geraden, die parallel zur y-Achse verläuft und die x-Achse im Punkt $(\overline{\boldsymbol{x}}, 0)$ schneidet.

Aufgrund des linearen Zusammenhangs gilt

$$\widehat{y}_k = ax_k + b + \varepsilon_k.$$

Setzt man voraus, daß der dem Messen innewohnende Zufallsprozess, bei der k-ten Messung einen Fehler ε_k mit verschwindender Erwartung produziert, d.h.

$$E\varepsilon_k = 0$$

so erweist sich $\widehat{a}$ als erwartungstreuer Punktschätzer für a:

$$E\widehat{a} = E\frac{s_{x\widehat{y}}}{s_{xx}} = \frac{Es_{x\widehat{y}}}{s_{xx}} = \frac{1}{s_{xx}}\frac{1}{n-1}E\sum_{k=1}^{n}(x_k-\overline{x})(\widehat{y}_k-\overline{\widehat{y}}) = \frac{1}{s_{xx}}\frac{1}{n-1}\sum_{k=1}^{n}(x_k-\overline{x})(E\widehat{y}_k-E\overline{\widehat{y}}).$$

Wegen

$$E\widehat{y}_k = E(ax_k + b + \varepsilon_k) = ax_k + b, \quad E\overline{\widehat{y}} = \frac{1}{n}\sum_{k=1}^{n}E\widehat{y}_k = a\overline{x} + b$$

bekommt man

$$E\widehat{a} = \frac{1}{s_{xx}}\frac{1}{n-1}\sum_{k=1}^{n}(x_k-\overline{x})(ax_k + b - a\overline{x} - b) = a\frac{1}{s_{xx}}\underbrace{\frac{1}{n-1}\sum_{k=1}^{n}(x_k-\overline{x})(x_k-\overline{x})}_{=s_{xx}} = a$$

sodaß man sich sicher sein kann, mit der Methode der kleinsten Quadrate, den Anstieg a nicht systematisch falsch zu schätzen, wie das beim ML-Schätzer S^* für die Varianz einer normalverteilten Zg der Fall ist.

Um einen Punktschätzer für b in Abhängigkeit von $\widehat{a}$ zu bekommen, beobachten wir, daß

$$f(a,b) := \sum_{k=1}^{n}(y_k - ax_k - b)^2$$

minimal wird, falls f_a und f_b verschwinden:

$$f_b(a,b) = \sum_{k=1}^{n}2(y_k - ax_k - b)(-1) = 0 \Rightarrow b = \overline{y} - a\overline{x}.$$

Daher benutzt man

$$\widehat{b} := \overline{\widehat{y}} - \widehat{a}\overline{x}$$

als Punktschätzer für das Absolutglied b. Wegen

$$E\widehat{b} = E(\overline{\widehat{y}} - \widehat{a}\overline{x}) = E\overline{\widehat{y}} - \overline{x}E\widehat{a} = a\overline{x} + b - \overline{x}a = b$$

ist auch dieser Punktschätzer erwartungstreu.

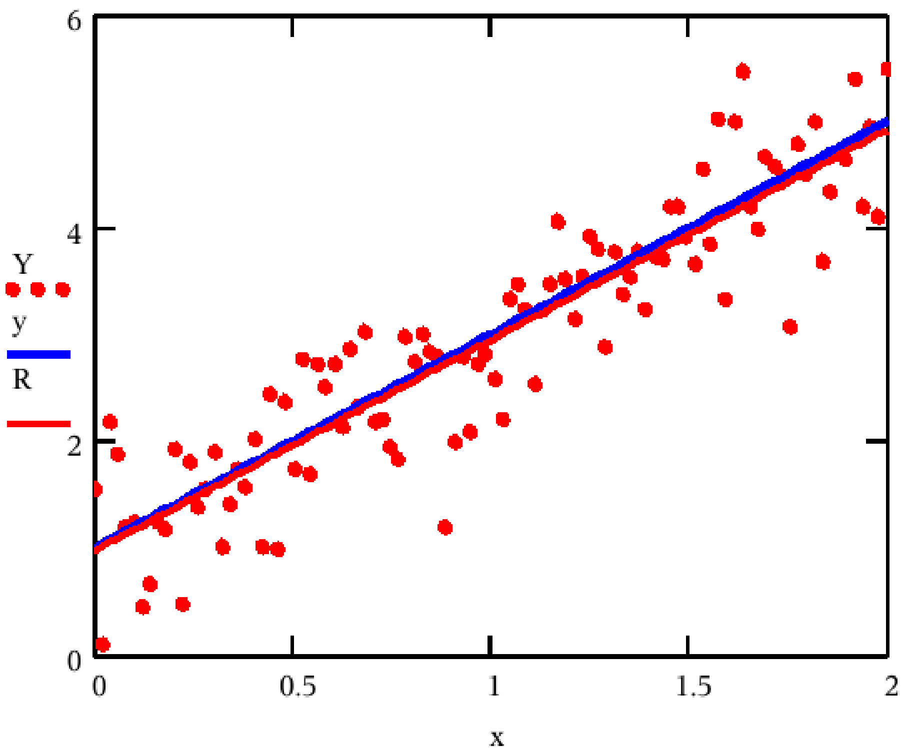


Abbildung 25: Punktwolke und Regressionsgerade zu Beispiel 45

Beispiel 45 Wir bilden die Punktwolke

$$W := \left\{(x_k, 2x_k + 1 + \varepsilon_k) : x_k = 0, \dots, \frac{2k}{100},\ k = 0, \dots, 100\right\}$$

wobei die y-Werte mit einem Fehler ε_k behaftet sind, der mit der Erwartung $\mu = 0$ und der Standardabweichung $\sigma := 0.5$ normalverteilt ist. Die blaue Gerade in Abbildung 21 ist der Graph von

$$y = 2x + 1$$

und die rote Gerade ist die berechnete Regressionsgerade

$$y = 1.971x + 0.977.$$

Demnach bekommt man mit der Methode der kleinsten Quadrate als Schätzung für den Anstieg 2 den Wert 1.971 und für das Absolutglied 1 den Wert 0.977. $\diamond$

Beispiel 46 In Abbildung 26 werden acht Punkte mittels der Geradengleichung $y = 3.5x - 0.4$ erzeugt erzeugt und diese leicht gestört. Aufgabe ist es, aus diesen gestörten Daten die Parameter $a = 3.5$ und $b = -0.4$ zu rekonstruieren. $\diamond$

3.5.6 Regressionsparabel für ebene Punktwolke

Gegeben sei eine Punktwolke W wie in (48). Gesucht sind nun die Koeffizienten a, b und c der Parabel $p(x) := ax^2 + bx + c$, sodaß

$$\sum_{k=1}^{n} (y_k - p(x_k))^2$$

möglichst klein wird. Bildet man die Matrix

$$A := \begin{bmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{bmatrix} \in \mathbb{R}^{n\times 3}$$

so bekommt man die gewünschten Koeffizienten durch

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} := (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{y}.$$

Beispiel 47 Wir rekonstruieren die Koeffizienten von $q(x) := 2x^2 - 4x + 3$ aus den gestörten Funktionswerten $y_k := q(x_k) + \varepsilon_k$, wobei der Fehler ε_k in $[-10, +10]$ gleichverteilt ist. Als Rekonstruktionsresultat erhalten wir

$$a = 2.067, \quad b := 4.681, \quad c = 5.254$$

Die roten Punkte in Abbildung 27 visualisieren die Punktwolke und die blaue Kurve den Graphen der zugehörigen Regressionsparabel. $\diamond$

Regressionsgerade $\quad a := 3.5 \quad b := -0.4 \quad y(x) := a\cdot x + b$

$k := 0..7 \qquad x_k := \frac{4}{8}\cdot k \qquad Y_k := y(x_k) \qquad E_k := 1$

$e := \frac{\text{stapeln}(1,-1,2,1,-3,4,1,-2)}{10} \qquad Y := Y + e \qquad \text{mittelwert}(e) = 0.038$

$$Y = \begin{pmatrix} -0.3 \\ 1.25 \\ 3.3 \\ 4.95 \\ 6.3 \\ 8.75 \\ 10.2 \\ 11.65 \end{pmatrix} \qquad x = \begin{pmatrix} 0 \\ 0.5 \\ 1 \\ 1.5 \\ 2 \\ 2.5 \\ 3 \\ 3.5 \end{pmatrix}$$

Man schätze die Parameter a und b

$\text{xquer} := \text{mittelwert}(x) \qquad \text{xquer} = 1.75$

$\text{yquer} := \text{mittelwert}(Y) \qquad \text{yquer} = 5.763$

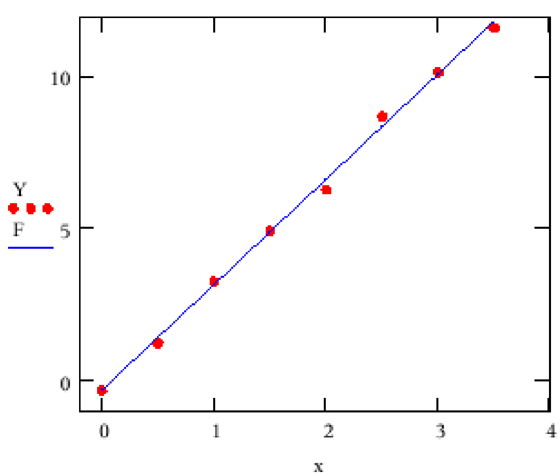
$x^T\cdot Y = 117.2$

$x^T\cdot x = 35$

$$\text{ahat} := \frac{x^T\cdot Y - 8\cdot\text{xquer}\cdot\text{yquer}}{x^T\cdot x - 8\cdot\text{xquer}^2} \qquad \text{ahat} = 3.479$$

$\text{bhat} := \text{yquer} - \text{ahat}\cdot\text{xquer} \qquad \text{bhat} = -0.325$

$F_k := \text{ahat}\cdot x_k + \text{bhat}$



$A := \text{erweitern}(x, E) \qquad \left(A^T\cdot A\right)^{-1}\cdot A^T\cdot Y = \begin{pmatrix} 3.479 \\ -0.325 \end{pmatrix}$

Abbildung 26: Rekonstruktion von a und b in Beispiel 46. Als Rekonstruktion bekommt man $\hat{a} = 3.479$ und $\hat{b} = -0.325$.

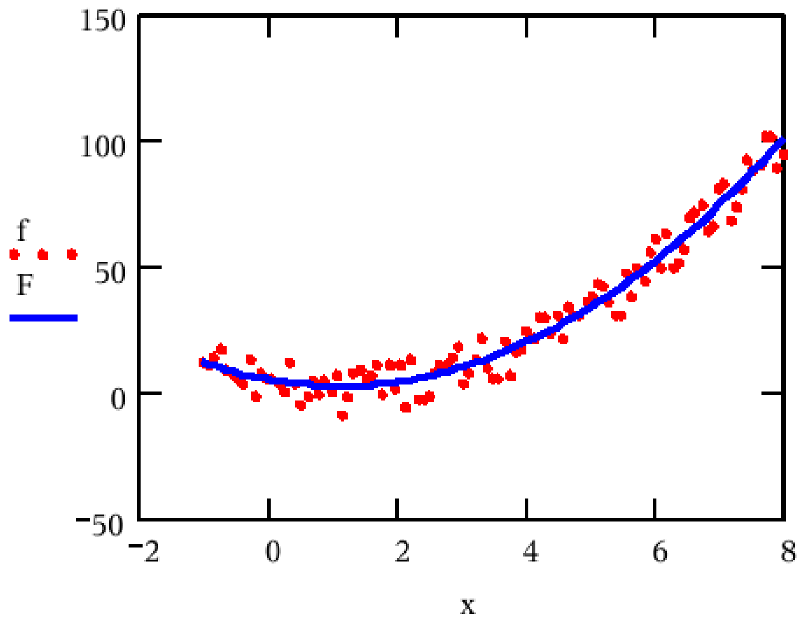


Abbildung 27: Es wird die Punktwolke W (rote Punkte) und der Graph ihrer Regressionsparabel (blaue Kurve) aus Beispiel 47 gezeigt.

3.6 Kovarianz und Korrelationskoeffizient zweier stetiger Zg

Im vorhergehenden Abschnitt wurde behandelt, wie man anhand von Stichproben etwas über die Unabhängigkeit zweier Zg X und Y aussagen kann. Jetzt soll diese Thematik auf dem Niveau der Dichtefunktionen des Zufallsvektors

$$\begin{bmatrix} X \\ Y \end{bmatrix} : \Omega \mapsto \mathbb{R}^2$$

behandelt werden. In Abschnitt 3.5.2 hatten wir die Kovarianz zweier Vektoren $\boldsymbol{x}$ und $\boldsymbol{y}$, deren Komponenten sich aus den Werten zweier Zg X und Y beim Experimentieren ergeben, eingeführt:

$$s_{\boldsymbol{xy}} := \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \overline{\boldsymbol{x}})(y_k - \overline{\boldsymbol{y}}).$$

Diese Definition geht für zwei Zufallsgrößen X und Y in

$$\boxed{\mathrm{Cov}(X,Y) := E((X-EX)(Y-EY))}$$

über. Die Zahl $\mathrm{Cov}(X,Y)$ heißt **Kovarianz** von X und Y. Offenbar gilt

$$\mathrm{Cov}(X,X) = E(X-EX)^2 = \mathrm{Var}(X).$$

Offenbar erhält man durch Ausmultiplizieren

$$\begin{aligned} \mathrm{Cov}(X,Y) = \; & E((X-EX)(Y-EY)) = E(XY - (EX)Y - X(EY) + EXEY) \\ = \; & E(XY) - 2EXEY + EXEY = E(XY) - EXEY \end{aligned}$$

d.h.

$$\boxed{\mathrm{Cov}(X,Y) = E(XY) - EXEY}$$

Entsprechend Abschnitt 2.6.5 wissen wir, daß für unabhängige disktrete Zg X und Y die Gleichung $E(XY) = EXEY$ gilt. Demnach verschwindet für unabhängige diskrete Zg die Kovarianz.

Für zwei stetige unabhängige Zg X und Y erfüllt die Dichtefunktion des entsprechenden Zufallsvektors $[X,Y]^{\mathrm{T}}$ die Gleichung

$$\varphi(x,y) = \varphi_X(x)\varphi_Y(y).$$

Daher gilt

$$\begin{aligned} E(XY) = & \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\varphi(x,y)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\varphi_X(x)\varphi_Y(y)\mathrm{d}x\mathrm{d}y \\ = & \int_{-\infty}^{+\infty} x\varphi_X(x)\mathrm{d}x \int_{-\infty}^{+\infty} y\varphi_Y(y)\mathrm{d}y = EXEY \end{aligned}$$

sodaß auch im stetigen Fall die Kovarianz zweier unabhängiger Zg verschwindet. Demnach eignet sich nach geeigneter Normierung die Kovarianz als Maß für den Grad der Abhängigkeit zwischen zwei Zg. Die Normierung geschieht wie beim Übergang von der Kovarianz einer Punktwolke zu ihrem Korrelationskoeffizienten in Abschnitt 3.5.4. In diesem Sinne setzen wir

$$\boxed{\varrho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}}$$

Mit Hilfe der Schwarzschen Ungleichung machen wir uns die Abschätzung

$$|\varrho(X,Y)| \leq 1 \tag{54}$$

plausibel. Ferner ist $|\varrho(X,Y)| = 1$ genau dann, wenn für gewisse Konstanten a und b die Gleichung

$$P(X = aY + b) = 1$$

gilt, wobei die Vorzeichen von $\varrho(X,Y)$ und a übereinstimmen. Die **Schwarzsche Ungleichung** lautet

$$\boxed{(EXY)^2 \leq EX^2EY^2} \tag{55}$$

die man bei der Annahme $EY^2 \neq 0$ wiefolgt einsieht: für alle λ gilt

$$0 \leq E(X+\lambda Y)^2 = EX^2 + \lambda E(XY) + \lambda(E(XY) + \lambda EY^2).$$

Die Wahl $\lambda := -\dfrac{E(XY)}{EY^2}$ führt auf

$$0 \leq EX^2 - \frac{(E(XY))^2}{EY^2} - \frac{E(XY)}{EY^2}\underbrace{\left(E(XY) - \frac{E(XY)}{EY^2}EY^2\right)}_{=0} = EX^2 - \frac{(E(XY))^2}{EY^2}$$

d.h. (55) gilt. Ihre Anwendung liefert

$$|\mathrm{Cov}(X,Y)|^2 = (E((X-EX)(Y-EY)))^2 \le E(X-EX)^2 E(Y-EY)^2 = \mathrm{Var}(X)\mathrm{Var}(Y)$$

d.h. (54) gilt tatsächlich.

Man sagt, daß X und Y **unkorreliert** sind, falls $\varrho(X,Y)=0$ gilt. Aus der Unkorreliertheit folgt aber nicht ihre Unabhängigkeit. Falls $\varrho(X,Y)\neq 0$ heißen X und Y korreliert.

3.7 Korrelierte normalverteilte Zufallsvektoren

Ein stetiger zweidimensionaler Zufallsvektor $V := [X,Y]^{\mathrm{T}}$ heißt mit den Parametern

$$\mu_1,\mu_2,\sigma_1,\sigma_2,\varrho, \quad 0<\sigma_1,\sigma_2,\ |\varrho|<1$$

normalverteilt, falls seine Dichtefunktion von der Gestalt

$$\varphi(x,y) := \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} e^{-\frac{1}{2(1-\varrho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\varrho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)}$$

ist [5] S.358. In Hinblick auf die Definition von $\varrho(X,Y)$ erwartet man natürlich, daß der Parameter ϱ mit dem Wert des Korrelationskoeffizienten von X und Y übereinstimmt. Zur Überprüfung dieser Vermutung berechnen wir zuerst die beiden Randdichten:

$$\varphi_X(x) := \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\varrho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\varrho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)} \mathrm{d}y$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\varrho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2(1-\varrho^2) + \left(\frac{x-\mu_1}{\sigma_1}\right)^2\varrho^2 - 2\varrho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)} \mathrm{d}y$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\varrho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2(1-\varrho^2) + \left(\left(\frac{x-\mu_1}{\sigma_1}\right)\varrho - \left(\frac{y-\mu_2}{\sigma_2}\right)\right)^2\right)} \mathrm{d}y$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\varrho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)\varrho - \left(\frac{y-\mu_2}{\sigma_2}\right)\right)^2} \mathrm{d}y.$$

Mit der Variablentransformation

$$z := \frac{1}{\sqrt{1-\varrho^2}}\left(\left(\frac{y-\mu_2}{\sigma_2}\right) - \left(\frac{x-\mu_1}{\sigma_1}\right)\varrho\right)$$

bekommen wir

$$\mathrm{d}z = \frac{1}{\sqrt{1-\varrho^2}}\frac{1}{\sigma_2}\mathrm{d}y$$

sodaß wir auch

$$\varphi_X(x) := \frac{1}{2\pi\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} \mathrm{d}z$$

schreiben können. Auf Seite 74 wird $\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} \mathrm{d}z = \sqrt{2\pi}$ gezeigt, sodaß schließlich

$$\varphi_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}$$

folgt, d.h. $X \sim \mathscr{N}(\mu_1, \sigma_1^2)$. In der gleichen Weise erhält man $Y \sim \mathscr{N}(\mu_2, \sigma_2^2)$.

Demnach gilt

$$\varrho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_1^2\sigma_2^2} = \frac{EXY - \mu_1\mu_2}{\sigma_1^2\sigma_2^2}$$

wobei

$$EXY = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy e^{-\frac{1}{2(1-\varrho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\varrho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)} \mathrm{d}x\mathrm{d}y.$$

Es folgt

$$EXY = \mu_1\mu_2 + \varrho\sigma_1^2\sigma_2^2$$

sodaß tatsächlich $\varrho(X,Y) = \varrho$ gilt. Wir überprüfen die Gleichung für EXY für den Fall $\mu_1 = \mu_2 = 0$ und $\sigma_1 = \sigma_2 = 1$:

$$\begin{aligned} EXY &= \frac{1}{2\pi\sqrt{1-\varrho^2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy e^{-\frac{1}{2(1-\varrho^2)}\left(x^2 - 2\varrho xy + y^2\right)} \mathrm{d}x\mathrm{d}y \\ &= \frac{1}{2\pi\sqrt{1-\varrho^2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy e^{-\frac{1}{2(1-\varrho^2)}\left(x^2(1-\varrho^2) + (y-\varrho x)^2\right)} \mathrm{d}x\mathrm{d}y \\ &= \frac{1}{2\pi\sqrt{1-\varrho^2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy e^{-\frac{x^2}{2}} e^{-\frac{(y-\varrho x)^2}{2(1-\varrho^2)}} \mathrm{d}x\mathrm{d}y. \end{aligned}$$

Mit der Variablentransformation

$$z := \frac{y - \varrho x}{\sqrt{1-\varrho^2}}$$

gilt

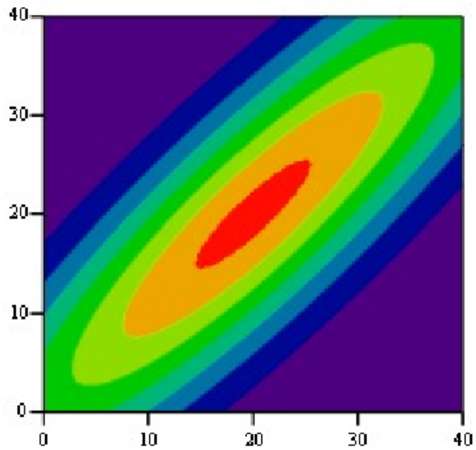
$$\mathrm{d}z = \frac{1}{\sqrt{1-\varrho^2}}\mathrm{d}y, \quad y = \sqrt{1-\varrho^2}z + \varrho x$$

sodaß

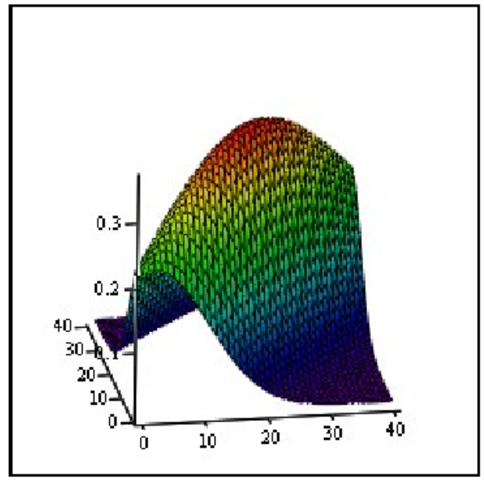
$$EXY = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(xz\sqrt{1-\varrho^2} + \varrho x^2\right) e^{-\frac{x^2}{2}} e^{-\frac{z^2}{2}} \mathrm{d}x\mathrm{d}z$$

$$= \frac{\sqrt{1-\varrho^2}}{2\pi} \underbrace{\int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} \mathrm{d}x}_{=0} \underbrace{\int_{-\infty}^{+\infty} z e^{-\frac{z^2}{2}} \mathrm{d}z}_{=0} + \frac{1}{2\pi}\varrho \underbrace{\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} \mathrm{d}x}_{=\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} \mathrm{d}z}_{=\sqrt{2\pi}} = \varrho.$$

Da für $\varrho = 0$ die beiden Komponenten eines normalverteilten zweidimensionalen Zufallvektors unabhängig werden, sind im Falle der Normalverteilung die Begriffe Korrelation und Unabhängigkeit äquivalent [17] S.360. Im allgemeinen ist das nicht so, wie das im folgenden Beispiel gezeigt wird.



G1



G1

Abbildung 28: Es wird der Werteverlauf von $f(x,y) := \frac{1}{2\pi\sqrt{1-\varrho^2}} e^{-\frac{1}{2(1-\varrho^2)}(x^2-2\varrho xy+y^2)}$ innerhalb von $[-1,1]\times[-1,1]$ für $\varrho := 0.9$ gezeigt. Im roten Bereich ist die Wahrscheinlichkeit am größten, daß dort eine Realisierung von $[X,Y]^{\mathrm{T}}$ liegt. Die Regressionsgerade einer durch solche Realisierungen erzeugten Punktwolke hätte in etwa den Anstieg ϱ.

Beispiel 48 ([9] S.166) Wir wählen $Y := X^2$. Offenbar sind dann X und Y abhängig, Setzt man von der Dichtefunktion f von X die Symmetrie zur 0 voraus, d.h. $f(-x) = f(x)$ für alle reellen x, so gilt

$$EX = 0, \quad EXY = EX^3 = 0$$

d.h. $\mathrm{Cov}(X,Y) = EXY - EXEY = 0$. ◇

3.8 Testverfahren für Parameter in Verteilungsfunktion (Parametertests)

Eine Hypothese ist eine unbewiesene Annahme von Gesetzlichkeiten oder Tatsachen mit dem Ziel, sie durch Beweise zu verifizieren oder zu falsifizieren. Um eine Hypothese zu falsifizieren, genügt es, ein Beispiel, das sog. Gegenbeispiel, anzugeben, für welches sie nicht stimmt. So streng soll hier aber nicht vorgegangen werden. Die folgenden Betrachtungen sollen die Abschwächung illustrieren.

Es wird der Ausgang eines Bernoulliexperiment der Länge n beobachtet und daraus geschlossen, daß die dem Experiment innewohnende Grundwahrscheinlichkeit

$$p \in [0,1]$$

mit der der Grundversuch positiv ausgeht, einen gewissen Wert p_0 nicht unterschreitet. Es soll getestet werden, ob diese Behauptung mit einer gewissen Irrtumswahrscheinlichkeit

$$\alpha \in [0, 1]$$

haltbar ist oder abgeleht werden kann. Daher ergeben sich die sog. Null- und die Alternativhypothese H_1 gemäß

$$H_0 : p_0 \leq p, \quad H_1 : p < p_0.$$

Unsere Testgröße X ist selbstverständlich die Anzahl k des Gelingens des Versuchs.

Unter Einbeziehung der Irrtumswahrscheinlichkeit α berechnen wir aus der Bedingung

$$P(X \leq c_0) = \alpha$$

eine Konstante c_0. Die Menge

$$A_{c_0} := \{X \leq c_0\} = \{\omega \in \Omega : X(\omega) \leq c_0\}$$

wird uns als sogenannter **Ablehnungsbereich** dienen, das soll heißen, falls unsere Stichprobe ein Element aus A_{c_0} liefert, werden wir H_0 mit einer Irrtumswahrscheinlichkeit von $100\alpha\,\%$ ablehnen.

Zur Berechnung von c_0 beobachten wir, daß X entsprechend Abschnitt 2.7 mit den Parametern n und p binomialverteilt

$$X \sim \mathscr{B}(n, p)$$

ist. Da wir das Bernoulliexperiment sehr oft ausführen wollen, also eine große Stichprobe ziehen, darf man entsprechend des Zentralen Grenzwertsatzes zur Normalverteilung mit geeigneten Parametern μ und σ übergehen:

$$X \sim \mathscr{N}(\mu, \sigma^2), \quad \mu := np, \quad \sigma := \sqrt{np(1-p)}.$$

Daher geht unsere Bedingung für c_0 in die Gleichung

$$\alpha = \Phi_{\mu,\sigma}(c_0)$$

über, die allerdings von der unbekannten Grundwahrscheinlichkeit p abhängt. Jetzt kommt unsere Nullhypothese H_0 ins Spiel. Wir ersetzen in der Definition von μ und σ das unbekannte p, von dem behauptet wird, daß es wenigstens so groß ist wie p_0, durch dieses p_0 und können schließlich c_0 in Abhängigkeit von n und p_0 ausrechnen:

$$\alpha = \Phi_{\mu,\sigma}(c_0), \quad \mu := np_0, \quad \sigma := \sqrt{np_0(1-p_0)}.$$

Wenn p tatsächlich nicht kleiner als p_0 ist, so nimmt die Menge

$$\{X \leq c_0\}$$

höchstens $100\alpha\,\%$ der gesamten statistischen Masse ein und in mindestens $100(1-\alpha)\%$ aller Fälle gilt

$$c_0 < X.$$

Um das einzusehen, genügt ein Blick auf Abbildung 22. Dort wird versichert, daß für fixierten Stichprobenumfang n und fixierte Irrtumswahrscheinlichkeit α die Funktion

$$\phi_{\alpha,n}(p) := \Phi^{-1}_{np,\,\sqrt{np(1-p)}}(\alpha)$$

ab einem bestimmten $\check{p}$ streng monoton wachsend ist. Falls

$$\check{p} \leq p_0 \leq p$$

so fällt das zu p gehörende c größer aus als c_0, sodaß für die resultierenden Ablehnungsbereiche die Enthaltenseinsrelation

$$\{X \leq c_0\} \subseteq \{X \leq c\}$$

gilt, und demzufolge

$$P(X \leq c_0) \leq \alpha$$

weshalb wir oben von höchstens $100\alpha\,\%$ der statistischen Masse sprechen. Wegen $\alpha \in \{0.01, 0.05, 0.1\}$ beträgt die Wahrscheinlichkeit

$$P(X \leq c_0)$$

höchstens 1%, 5% oder 10% vorausgesetzt, daß unsere Hypothese $H_0 : p_0 \leq p$ gilt. Deshalb ist es um ihre Glaubwürdigkeit schlecht bestellt, falls beim Ziehen einer Stichprobe die Testgröße X kleiner als c_0 ausfällt, d.h. im Ablehnungsbereich

$$\{0, 1, \ldots, [c_0]\}$$

liegt. Hier bezeichnet $[c_0]$ den ganzen Anteil von c_0. Wegen

$$X(A_{c_0}) = \{0, 1, \ldots, [c_0]\}$$

ist auch für $\{0, 1, \ldots, [c_0]\}$ der Begriff 'Ablehnungsbereich' gerechtfertigt.

Das liefert uns die Entscheidungsregel:

- Falls bei einem Bernoulliexperiment der Länge n die Anzahl X des Gelingens des Grundversuchs kleiner ist als die Schranke c_0, so kann die Nullhypothese mit der Irrtumswahrscheinlichkeit $100\alpha\,\%$ abgelehnt werden.

- Andernfalls wird H_0 mit der Irrtumswahrscheinlichkeit $100\alpha\,\%$ akzeptiert.

- Die Wahrscheinlichkeit, daß $X(\omega) \in \{0, 1, \ldots, [c_0]\}$ beträgt höchstens α. Auf diese Weise ordnet man dem Ablehnungsbereich die Zahl α und dem Annahmebereich $\{[c_0] + 1, \ldots, n\}$ die Zahl $1 - \alpha$ zu.

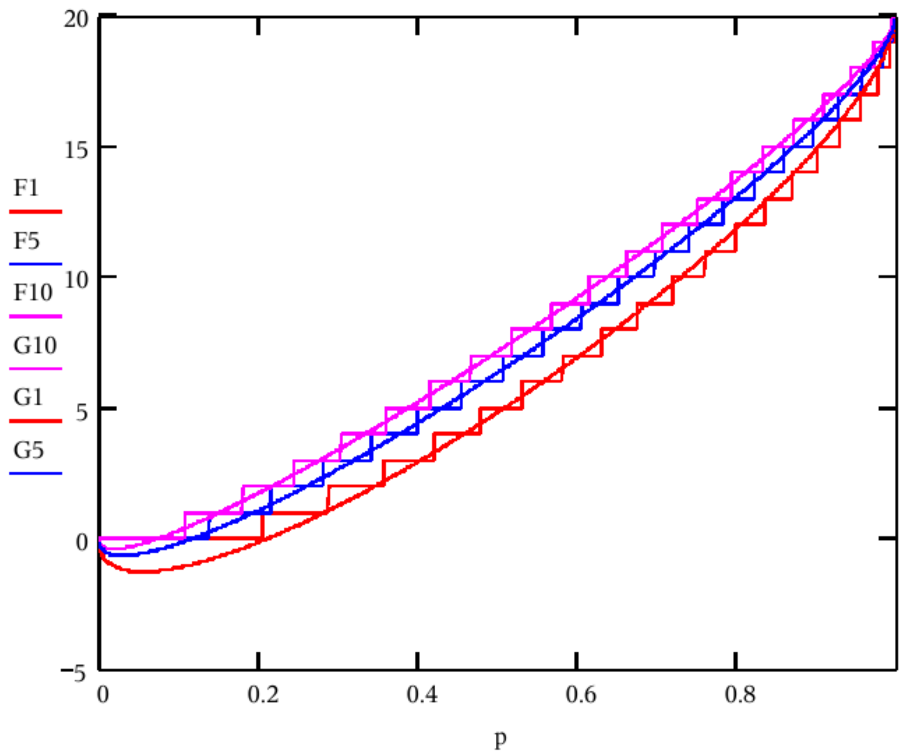


Abbildung 29: Es werden über $p \in (0,1)$ für $n = 20$ und für $\alpha = 0.01$ (rot), $\alpha = 0.05$ (blau), $\alpha = 0.1$ (violett) der Graph von $\phi_{\alpha,n}(p) := \Phi^{-1}_{np,\sqrt{np(1-p)}}(\alpha)$ (G1, G5, G10) mit dem Graphen (F1,F5,F10) der aus $F(x) = \alpha$, $F(x) := \sum_{0 \leq k \leq x} \binom{n}{k} p^k (1-p)^{n-k}$ resultierenden Inversen F^{-1} im Sinne von Abschnitt 2.9.2 verglichen. Die Treppenfunktionen sind offensichtlich monoton wachsend, was für $\phi_{\alpha,n}(p)$ erst ab einem gewissen Wert $\check{p}$ gilt. Daher muß man bei der Ersetzung von p durch p_0 auf $\check{p} \leq p_0$ achten.

Beispiel 49 Überraschungseier[15], S.163

Ein Hersteller von Überraschungseiern versichert, daß in mindestens 14 % der Eier Figuren aus einem beliebten Fantasy-Abenteuer stecken. Eine Verbraucherorganisation ist mißtrauisch und möchte diese Aussage überprüfen. Zu diesem Zweck wird eine Stichprobe von $n := 1000$ Eiern genommen und festgestellt, daß sich 130 Figuren aus dem beliebten Fantasy-Abenteuer darunter befinden, was nur einem Anteil von 13 % entspricht. Genügt diese Feststellung, um die Behauptung des Herstellers zu widerlegen?

Die Beantwortung hängt von der gewählten Irrtumswahrscheinlichkeit ab. Wir wählen

$$\alpha := 0.05.$$

Laut Aufgabenstellung befinden sich unter 100 Eiern wenigstens 14 'gute' Eier. Entsprechend des ML-Schätzers für den Anteilwert, liefert diese Aussage eine untere Schranke

$$p_0 = 0.14$$

für den Anteilswert p.

Daher lauten die Null- und die Ablehnungshypothese

$$H_0 : 0.14 \leq p, \quad H_1 : p < 0.14.$$

Der Stichprobenumfang ist $n = 1000$, sodaß die Konstante c_0 entsprechend

$$c_0 = \Phi^{-1}_{1000\times 0.14,\sqrt{1000\times 0.14(1-0.14)}}(0.05) = \Phi^{-1}_{140,\sqrt{140\times 0.86}}(0.05) = \Phi^{-1}_{140,\sqrt{120.4}}(0.05) = 121.95$$

berechnet werden kann, [5] d.h. der Ablehnungsbereich lautet

$$[0, \ldots, 121].$$

Demnach kann H_0 mit einer Irrtumswahrscheinlichkeit von 5 % abgelehnt werden, falls eine Stichprobe im Umfang von 1000 einen Anteil k von Eiern mit der richtigen Figur von weniger als 121 liefert.

[5] Falls man keine Software zur Berechnung von $\Phi^{-1}_{\mu,\sigma}(x) = \mathrm{qnorm}(x, \mu, \sigma)$ zur Verfügung hat, kann man die Gleichung

$$0.05 = \Phi_{0,1}\left(\frac{c_0 - 140}{\sqrt{120.4}}\right)$$

mittels Tabelle

$$\begin{array}{c|c|c|c} \alpha & 0.01 & 0.05 & 0.10 \\ \hline x & -2.326 & -1.645 & -1.282 \end{array}, \quad \Phi_{0,1}(x) = \alpha$$

lösen. Dann gilt

$$\frac{c_0 - 140}{\sqrt{120.4}} = -1.645$$

und Auflösung nach c_0 liefert

$$c_0 = 140 - 1.645 \times \sqrt{120.4} = 140 - 1.645 \times 10.972 = 140 - 18.0419 = 121.95.$$

Testergebnis: Da $k = 130$ mal die richtige Figur erscheint, wird H_0 mit einer Irrtumswahrscheinlichkeit von 5 % akzeptiert.

Variation von α: Die Erhöhung der Irrtumswahrscheinlichkeit auf 10 % vergrößert den Ablehnungsbereich, d.h. die Chance abzulehnen. Mit $\alpha = 10\%$ bekommen wir

$$c_0 = 140 - 1.282 \times \sqrt{120.4} = 140 - 1.282 \times 10.972 = 125.934$$

d.h. der Ablehnungsbereich lautet

$$[0, \dots, 125].$$

Falls die Stichprobe $k = 123$ ergibt, würde man mit 5 % Irrtumswahrscheinlichkeit akzeptieren und mit 10 % Irrtumswahrscheinlichkeit ablehnen. $\diamond$

3.8.1 Tests für den Anteilswert p

[15] S.163, [13] S.587

Wir fassen die Vorgehensweise aus der Einleitung in einem Algorithmus zusammen.

Test 1. Es gelte $np_0(1 - p_0) > 9$.

- Setze $H_0 : p_0 \leq p$, $H_1 : p < p_0$.
- Wähle Irrtumswahrscheinlichkeit $\alpha \in \{0.01, 0.05, 0.1\}$.
- Bestimme zugehöriges x aus

$$\begin{array}{c|r|r|r} \alpha & 0.010 & 0.050 & 0.100 \\ \hline x & -2.326 & -1.645 & -1.282 \end{array}, \qquad \Phi_{0,1}(x) = \alpha.$$

- Ziehe Stichprobe im Umfang n und bestimme relative Häufigkeit $\widehat{p} := \dfrac{k}{n}$, wobei k die Anzahl des Gelingens des Grundversuchs bezeichnet.
- Entscheidungsregel: Falls

$$\widehat{p} \leq c_u, \quad c_u := p_0 + x\sqrt{\frac{p_0(1-p_0)}{n}}$$

  kann man H_0 mit einer Irrtumswahrscheinlichkeit von α ablehnen. [6]

[6] Diese Ungleichung folgt aus der Bedingung

$$k \leq c_u, \quad \frac{c_u - \mu}{\sigma} = x, \quad \mu = np_0, \quad \sigma = \sqrt{np_0(1-p_0)}.$$

Dazu äquivalent ist nämlich

$$\widehat{p} = \frac{k}{n} \leq \frac{c_u}{n} = \frac{\mu + x\sigma}{n} = p_0 + x\frac{\sqrt{np_0(1-p_0)}}{n} = p_0 + x\sqrt{\frac{p_0(1-p_0)}{n}}.$$

Bemerkungen:

- Die Voraussetzung $np_0(1-p_0) > 9$ hängt mit dem monotonen Fallen von

$$\Phi_{np,\sqrt{np(1-p)}}(\alpha), \alpha \text{ und } n \text{ fixiert}$$

bei Variation von p in der Nähe von Null zusammen. Da p_0 zu Testbeginn bekannt ist, kann man $n \in \mathbb{N}$ immer so groß wählen, daß die geforderte Ungleichung erfüllt wird.

- Da man es mit der Begrenzung von p nach nur einer Seite zu tun hat, spricht man von einem einseitigen Test.

- Die Wahrscheinlichkeit, daß $\widehat{p} \in [0, c_u]$ beträgt näherungsweise α. Genaueres läßt sich ohne Weiteres nicht sagen, da die Binomialverteilung durch die Normalverteilung ersetzt wurde.

- Dem Ablehnungsbereich $[0, c_u]$ wird die Wahrscheinlichkeit α und dem Annahmebereich $[c_u, 1]$ die Wahrscheinlichkeit $1-\alpha$ zugeordnet.

Beispiel 50 Wir wenden Test 1 auf die Überraschungseier an. Wir überprüfen

$$np_0(1-p_0) = 1000 \times 0.14 \times 0.86 = 120,4 > 9$$

d.h. der Test kann angewendet werden.

- $H_0 : p_0 \leq p$, $H_1 : p < p_0$, $p_0 = 0.14$
- Wir wählen $\alpha = 0.05$.
- Wir bestimmen $x := \Phi_{0,1}^{-1}(0.05) = -1.645$.
- Wir wählen 1000 Eier aus und entdecken $k := 130$ 'gute' Eier. Daher gilt $\widehat{p} := 0.13$.
- Wir berechnen

$$\begin{aligned} c_u &:= p_0 + x\sqrt{\frac{p_0(1-p_0)}{n}} \\ &= 0.14 - 1.645\sqrt{\frac{0.14(1-0.14)}{1000}} \\ &= 0.14 - 1.645\sqrt{0.0001204} = 0.14 - 0.011 = 0.129. \end{aligned}$$

- Wegen $0.129 < 0.13 = \widehat{p}$ kann H_0 mit einer Irrtumswahrscheinlichkeit von 5 % akzeptiert werden.

$\diamond$

Wir hatten bisher getestet, ob ein Anteilswert p nicht unter eine Schranke p_0 fällt. Selbstverständlich kann man auch testen, ob p eine Schranke p_0 nicht überschreitet. Zu diesem Zweck, hat man c so zu bestimmen, daß

$$\alpha = P(c \leq X) = 1 - P(X \leq c) = 1 - \Phi_{\mu,\sigma}(c), \quad \mu := np_0, \quad \sigma := \sqrt{np_0(1-p_0)}.$$

Normierung führt auf

$$1 - \alpha = \Phi_{0,1}\left(\frac{c-\mu}{\sigma}\right)$$

sodaß aus $1 - \alpha = \Phi_{0,1}(x)$ die Gleichung

$$c = \mu + x\sigma$$

folgt. Beachtet man

$$1 - \alpha = \Phi_{0,1}(x) \quad \Leftrightarrow \quad \alpha = \Phi_{0,1}(-x)$$

so kann man auch die Gleichung $\alpha = \Phi_{0,1}(x)$ lösen und

$$c = \mu - x\sigma$$

setzen. Diese Feststellungen fassen wir zusammen in

Test 2. Es gelte $np_0(1-p_0) > 9$.

- Setze $H_0 : p \leq p_0$, $H_1 : p_0 < p$.
- Wähle Irrtumswahrscheinlichkeit $\alpha \in \{0.01, 0.05, 0.1\}$.
- Bestimme zugehöriges x aus

α	0.010	0.050	0.100
x	-2.326	-1.645	-1.282

$$\Phi_{0,1}(x) = \alpha.$$

- Ziehe Stichprobe im Umfang n und bestimme relative Häufigkeit $\widehat{p} := \dfrac{k}{n}$.
- Entscheidungsregel: Falls

$$c_o \leq \widehat{p}, \quad c_o := p_0 - x\sqrt{\frac{p_0(1-p_0)}{n}}$$

  kann man H_0 mit einer Irrtumswahrscheinlichkeit von α ablehnen.

Bemerkungen:

- Da man es mit der Begrenzung von p nach nur einer Seite zu tun hat, haben wir wieder einen einseitigen Test.
- Die Wahrscheinlichkeit, daß $\widehat{p} \in [c_o, 1]$, beträgt näherungsweise α.

- Dem Ablehnungsbereich $[c_o, 1]$ wird die Wahrscheinlichkeit α und dem Annahmebereich $[0, c_o]$ die Wahrscheinlichkeit $1-\alpha$ zugeordnet.

Schließlich soll getestet werden, ob $p = p_0$ mit einer gewissen Irrtumswahrscheinlichkeit gilt. Der Test dazu ergibt sich einfach aus Kombination der beiden vorhergehenden.

Führen beide Tests zur Annahme, so gelten mit der Irrtumswahrscheinlichkeit α die beiden Hypothesen

$$p_0 \leq p \text{ und } p \leq p_0$$

falls $\widehat{p} \in [c_u, 1]$ und $\widehat{p} \in [0, c_o]$, d.h.

$$p_0 = p$$

falls $\widehat{p} \in [c_u, c_o]$.

Test 3. Es gelte $np_0(1-p_0) > 9$.

- Setze $H_0 : p = p_0$, $H_1 : p \neq p_0$.
- Wähle Irrtumswahrscheinlichkeit $\alpha \in \{0.01, 0.05, 0.1\}$.
- Bestimme zugehöriges x aus

$$\begin{array}{c|r|r|r} \alpha & 0.010 & 0.050 & 0.100 \\ \hline x & -2.326 & -1.645 & -1.282 \end{array}, \quad \Phi_{0,1}(x) = \alpha$$

- Berechne die Grenzen des Annahmebereichs entsprechend

$$c_u := p_0 + x\sqrt{\frac{p_0(1-p_0)}{n}}, \quad c_o := p_0 - x\sqrt{\frac{p_0(1-p_0)}{n}}$$

- Ziehe Stichprobe im Umfang n und bestimme relative Häufigkeit $\widehat{p} := \dfrac{k}{n}$.
- Entscheidungsregel: Falls

$$\widehat{p} \leq c_u \text{ oder } c_o \leq \widehat{p}$$

  kann man H_0 mit einer Irrtumswahrscheinlichkeit von α ablehnen.

Bemerkung: Die letzte Entscheidungsregel kann man durch

“H_0 kann mit einer Irrtumswahrscheinlicheit
von α angenommen werden, falls $\widehat{p} \in (c_u, c_o)$”

ersetzen. Das Intervall (c_u, c_o) heißt in diesem Zusammenhang der **Annahmebereich** und die zweiteilige Menge

$$[0, 1] \setminus (c_u, c_o)$$

der **Ablehnungsbereich**.

Beispiel 51 Ein Hersteller behauptet, daß 18 % seiner hergestellten Artikel eine bestimmte Eigenschaft besitzen. Zu Überprüfungszwecken wurde eine Stichprobe im Umfang $n = 90$ gezogen, bei der 24 Artikel die in Frage stehende Eigenschaft besitzen. Man teste mit den üblichen Irrtumswahrscheinlichkeiten 1 %, 5 % und 10 %, ob der wahre Anteilswert p, der dem Herstellungsprozess innewohnt, mit 0.18 übereinstimmt.

Wir arbeiten Test 3 ab. Wegen $90 \times 0.18 \times (1 - 0.18) = 13.28 > 9$ ist seine Anwendung statthaft.

- $H_0 : p = p_0$, $H_1 : p \neq p_0$, $p_0 := 0.18$.
- Wir wählen der Reihe nach $\alpha = 0.01, 0.05, 0.10$.
- Entsprechend dieser Reihenfolge bekommen wir $x = -2.326, -1.645, -1.282$.
- Die Grenzen der entsprechenden Annahmebereiche lauten

α	c_u	c_o
0.01	0.086	0.274
0.05	0.113	0.247
0.10	0.128	0.232

- Für die relative Häufigkeit bekommen wir

$$\widehat{p} = \frac{24}{90} = 0.267.$$

- Offenbar gilt für $\alpha = 0.01$ die Relation $\widehat{p} \in [c_u, c_o]$ und für $\alpha = 0.05$ oder $\alpha = 0.1$ nicht, so daß mit der Irrtumswahrscheinlichkeit 1 % die Hypothese H_0 angenommen, mit den Irrtumswahrscheinlichkeiten 5 % und 10 % aber abgelehnt werden kann.

$\diamond$

3.8.2 Test für die Erwartung $\boldsymbol{\mu}$ von $\boldsymbol{X} \sim \mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ bei bekannter Standardabweichung, Gauss-Test

[13] S.552

Es soll geprüft werden, ob die Erwartung μ einer normalverteilten Grundgesamtheit, von der die Standardabweichung σ bekannt ist, mit einem gewissen Wert μ_0 übereinstimmt. Dieser Test korrespondiert mit der Konstruktion eines Vertrauensintervalls für μ bei bekanntem σ wie sie in Abschnitt 3.4.1 vorgenommen wurde. Dort hatten wir festgestellt, daß μ mit einer Wahrscheinlichkeit von $1 - \alpha$ vom Intervall

$$[\overline{\boldsymbol{x}} - c, \overline{\boldsymbol{x}} + c], \quad c := \frac{\sigma z}{\sqrt{n}}, \quad z := \Phi_{0,1}^{-1}(1 - \alpha/2)$$

überdeckt wird. Nimmt man für μ den hypothetischen Wert μ_0 an, so bedeutet diese Überdeckungswahrscheinlichkeit

$$1-\alpha = P(\overline{X}-c \le \mu_0 \le \overline{X}+c) = P(\mu_0 - c \le \overline{X} \le \mu_0 + c)$$

d.h. falls wirklich $\mu = \mu_0$ gilt, so ist die Wahrscheinlichkeit, daß das Stichprobenmittel $\overline{\boldsymbol{x}}$ einen Wert in $[\mu_0 - c, \mu_0 + c]$ annimmt, gleich $1-\alpha$, also hoch, wenn wir α klein wählen. Daher wird die Hypothese $\mu = \mu_0$ wahrscheinlich stimmen, wenn sich

$$\overline{\boldsymbol{x}} \in [\mu_0 - c, \mu_0 + c]$$

feststellen läßt. Zusammenfassend formulieren wir

Test 4.

- Setze $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$.
- Wähle Irrtumswahrscheinlichkeit $\alpha \in \{0.01, 0.05, 0.10\}$.
- Bestimme z aus der Gleichung

  $$\Phi_{0,1}(z) = 1 - \frac{\alpha}{2}.$$

  d.h. z kann der Tabelle

α	0.010	0.050	0.100
z	2.576	1.960	1.645

  $$z = \Phi_{0,1}^{-1}\left(1-\frac{\alpha}{2}\right)$$

  entnommen werden.
- Wähle eine Stichprobe im Umfang von n und bilde den Mittelwert $\overline{\boldsymbol{x}}$.
- Berechne die Grenzen

  $$c_u := \mu_0 - \frac{\sigma z}{\sqrt{n}}, \quad c_o := \mu_0 + \frac{\sigma z}{\sqrt{n}}$$

  des Annahmebereichs.
- Entscheidungsregel: Falls $\overline{\boldsymbol{x}} \in [c_u, c_o]$, kann H_0 mit einer Irrtumswahrscheinlichkeit α angenommen werden.

Bemerkungen:

- Die Wahl von $\overline{\boldsymbol{x}} \in [c_u, c_o]$ als Entscheidungsregel ist durch die Eigenschaft von $\overline{\boldsymbol{x}}$, der ML-Schätzer von μ zu sein, gerechtfertigt.
- Rechnerich gibt es zur Konstruktion des Vertrauensintervalls keine Unterschiede, allerdings werden unterschiedliche Aussagen getroffen.

- Offenbar gilt

$$\overline{\boldsymbol{x}} \in [c_u, c_o] \;\Leftrightarrow\; -z \leq \frac{\overline{\boldsymbol{x}} - \mu_0}{\sigma}\sqrt{n} \leq z \;\Leftrightarrow\; \left|\frac{\overline{\boldsymbol{x}} - \mu_0}{\sigma}\sqrt{n}\right| \leq z.$$

  Führt man daher $t := \frac{\overline{\boldsymbol{x}}-\mu_0}{\sigma}\sqrt{n}$ als 'Prüfgröße' ein, so hat man danach zu entscheiden, ob $|t|$ kleiner oder größer ist als z.

Beispiel 52 Nägel mit Köpfen aber unterschiedlichen Längen [13], S.554.

Zehn Nägel aus einem bestimmten Sortiment haben die Länge in mm:

i	1	2	3	4	5	6	7	8	9	10
x_i	20	22	19	22	18	21	18	19	21	20

Sie stammen aus einer normalverteilten Grundgesamtheit mit der Standardabweichung $\sigma = 3$ mm. Man teste zur Irrtumswahrscheinlichkeit 1 % die Hypothese, daß diese Stichprobe einer Grundgesamtheit mit dem Mittelwert $\mu_0 = 22$ mm entnommen wurde.

Wir arbeiten Test 4 ab. Die Einheit mm wird immer weggelassen.

- $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$, $\mu_0 := 22$.
- Als Irrtumswahrscheinlichkeit wählen wir $\alpha = 0.01$.
- Das zugehörige z lautet $z = 2.576$.
- Offenbar gilt $n = 10$ und

$$\overline{\boldsymbol{x}} = 20.$$

- Für die Grenzen des Annahmebereichs bekommen wir

$$c_u = \mu_0 - \frac{z\sigma}{\sqrt{n}} = 22 - 2.576\frac{3}{\sqrt{10}} = 19.56,\; c_o := \mu_0 + \frac{z\sigma}{\sqrt{n}} = 22 + 2.576\frac{3}{\sqrt{10}} = 24.44.$$

- Wegen

$$\overline{\boldsymbol{x}} = 20 \in [19.56, 24.44]$$

  kann mit einer Irrtumswahrscheinlichkeit von 1 % die Behauptung $\mu = 22$ angenommen werden.

$\diamond$

3.8.3 Test für die Erwartung $\boldsymbol{\mu}$ von $\boldsymbol{X \sim \mathcal{N}(\mu, \sigma^2)}$ bei unbekannter Standardabweichung, einfacher $\boldsymbol{t}$-Test

[15] Beispiel 4.17, [13] S.559

Es soll geprüft werden, ob die Erwartung μ einer normalverteilten Grundgesamtheit, von der die Varianz σ^2 unbekannt ist, mit einem gewissen Wert μ_0 übereinstimmt. Dieser Test korrespondiert mit der Konstruktion eines Vertrauensintervalls für μ bei unbekanntem σ entsprechend Abschnitt 3.4.2.

Test 5.

- $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$.
- Wähle Irrtumswahrscheinlichkeit $\alpha \in \{0.01, 0.05, 0.10\}$.
- Bestimme z aus der Gleichung
$$F_{n-1}(z) = 1 - \frac{\alpha}{2}$$
wobei F_n die t-Verteilungsfunktion für n Freiheitsgrade darstellt und n den Stichprobenumfang. Falls $n \in \{1, \ldots, 30\}$ kann Tabelle (79) benutzt werden.
- Wähle eine Stichprobe $x_1, \ldots, x_n$ und berechne die Prüfgröße
$$t := \frac{\overline{\boldsymbol{x}} - \mu_0}{s}\sqrt{n}, \quad \overline{\boldsymbol{x}} := \frac{1}{n}\sum_{k=1}^{n} x_k, \quad s := \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{\boldsymbol{x}})^2}.$$
- Entscheidungsregel: Falls $|t| \leq z$, kann H_0 mit einer Irrtumswahrscheinlichkeit von α akzeptiert werden.

Beispiel 53 Zylinderscheiben [13], S.561.

Ein Hersteller produziert in großen Mengen Zylinderscheiben mit einem *Solldurchmesser* von 20.2 mm. Um die Einhaltung des Sollwertes zu überprüfen, wird aus der laufenden Produktion eine Stichprobe im Umfang von $n = 16$ entnommen. Die Auswertung der Stichprobe ergibt dabei einen mittleren Durchmesser von $\overline{\boldsymbol{x}} = 20.6$ mm mit einer empirischen Standardabweichung $s = 0.5$ mm. Es soll mit einer Irrtumswahrscheinlichkeit von 5 % überprüft werden, ob bei diesem Produktionsprozess der Sollwert eingehalten wird. Es darf Normalverteilung der Zylinderscheibendurchmesser vorausgesetzt werden.

Wir arbeiten Test 5 ab. Die Einheit mm wird immer weggelassen.

- $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$, $\mu_0 := 20.2$.
- Als Irrtumswahrscheinlichkeit wählen wir $\alpha = 0.05$.

- Den zugehörigen z-Wert lesen wir in Tabelle (79) ab. Wegen $n = 16$ haben wir mit Zeile 15 und wegen $\alpha = 0.05$ mit Spalte 3 zu arbeiten. Dort lesen wir ab:
$$z = 2.131.$$
- Unsere Prüfgröße nimmt folgenden Wert an:
$$t := \frac{\overline{\boldsymbol{x}} - \mu_0}{s}\sqrt{n} = \frac{20.6 - 20.2}{0.5}\sqrt{16} = 3.2.$$
- Demnach gilt $t > z$, sodaß die Hypothese, der Produktionsvorgang würde den Solldurchmesser von 20.2 mm einhalten, mit einer Irrtumswahrscheinlichkeit von 5 % abgelehnt werden kann.

$\diamond$

3.8.4 Test für die Varianz $\boldsymbol{\sigma^2}$ von $\boldsymbol{X \sim \mathcal{N}(\mu, \sigma^2)}$, $\boldsymbol{\chi^2}$-Streuungstest

[13] S.583

Es soll geprüft werden, ob die Varianz σ^2 einer normalverteilten Grundgesamtheit mit einem gewissen Wert σ_0^2 übereinstimmt. Dieser Test korrespondiert mit der Konstruktion eines Vertrauensintervalls für σ entsprechend Abschnitt 3.4.3. Dort hatten wir mittels der χ^2-Verteilung zwei Konstanten c_1 und c_2 so bestimmt, daß
$$P\left(\frac{n-1}{c_2}S^2 \leq \sigma^2 \leq \frac{n-1}{c_1}S^2\right) = 1 - \alpha$$
gilt, wobei S^2 die Stichprobenvarianz darstellt. Falls die Hypothese $\sigma = \sigma_0$ aufgestellt wird, bekommt man
$$\begin{aligned}
1 - \alpha \quad &= P\left(\frac{n-1}{c_2}S^2 \leq \sigma_0^2 \leq \frac{n-1}{c_1}S^2\right) \\
&= P\left(\frac{n-1}{c_2\sigma_0^2} \leq \frac{1}{S^2} \leq \frac{n-1}{c_1\sigma_0^2}\right) \\
&= P\left(\frac{c_1\sigma_0^2}{n-1} \leq S^2 \leq \frac{c_2\sigma_0^2}{n-1}\right)
\end{aligned}$$
d.h. die Wahrscheinlichkeit mit der die Stichprobenvarianz s^2 im Intervall
$$\left[\frac{c_1\sigma_0^2}{n-1}, \frac{c_2\sigma_0^2}{n-1}\right]$$
liegt ist $1 - \alpha$, falls die Hypothese $\sigma = \sigma_0$ stimmt. Diese Beobachtung gibt Anlass zu folgendem Test.

Test 6.

- $H_0 : \sigma = \sigma_0,\ H_1 : \sigma \neq \sigma_0$.
- Wähle Irrtumswahrscheinlichkeit $\alpha \in (0, 1)$.
- Berechne aus den Gleichungen

$$F_{n-1}(c_1) = \frac{\alpha}{2}, \quad F_{n-1}(c_2) = 1 - \frac{\alpha}{2}$$

  die Konstanten c_1 und c_2, wobei n den Stichprobenumfang und F_n die χ^2-Verteilungsfunktion mit n Freiheitsgraden darstellen. Falls

$$\alpha \in \{0.01, 0.05, 0.10\}, \quad n \in \{3, 4, 5, 10, 15, 30\}$$

  kann dazu Tabelle (45) benutzt werden.

- Wähle eine Stichprobe $\boldsymbol{x} := [x_k]_{k=1}^n$ im Umfang n und bilde die Stichprobenvarianz

$$s^2 := \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \overline{\boldsymbol{x}})^2, \quad \overline{\boldsymbol{x}} := \frac{1}{n} \sum_{k=1}^{n} x_k.$$

- Berechne die Grenzen des Annahmebereichs unter Verwendung des hypothetischen Wertes σ_0:

$$c_u := \frac{c_1 \sigma_0^2}{n-1}, \quad c_o := \frac{c_2 \sigma_0^2}{n-1}.$$

- Entscheidungsregel: Falls $s^2 \in [c_u, c_o]$, kann H_0 mit der Irrtumswahrscheinlichkeit α angenommen werden.

Bemerkungen:

- In die Konstruktion des Vertrauensintervalls für σ geht die Stichprobenvarianz s^2 ein. Bei Test 6 wird der Annahmebereich mittels σ_0 konstruiert und überprüft, ob s^2 in diesem Annahmebereich liegt.
- Offenbar gilt

$$s^2 \in [c_u, c_o] \ \Leftrightarrow\ \frac{c_1 \sigma_0^2}{n-1} \leq s^2 \leq \frac{c_2 \sigma_0^2}{n-1} \ \Leftrightarrow\ c_1 \leq \frac{(n-1)s^2}{\sigma_0^2} \leq c_2.$$

  Führt man die Prüfgröße

$$t := (n-1)\frac{s^2}{\sigma_0^2}$$

  ein, so kann H_0 angenommen werden, falls $c_1 \leq t \leq c_2$.

Beispiel 54 Schraubenlänge [13], S.586

Bei der Serienherstellung von Schrauben mit einer bestimmten Länge kann die Zufallsgröße

$$X := \text{ Länge einer Schraube}$$

als normalverteilt angesehen werden. Aufgrund von Beobachtung behauptet jemand, daß die Standardabweichung $\sigma = 1.2$ beträgt. Eine zu Kontrollzwecken entnommene Stichprobe im Umfang $n = 15$ ergab eine empirische Standardabweichung von $s = 1.5$. Man teste die Behauptung mit den Irrtumswahrscheinlichkeiten 1 % und 10 %.

Wir arbeiten Test 6 ab.

- $H_0 : \sigma = \sigma_0$, $H_1 : \sigma \neq \sigma_0$, $\sigma_0 := 1.2$
- Wir wählen Irrtumswahrscheinlichkeit $\alpha = 0.01$.
- Wir berechnen aus den Gleichungen

$$F_{n-1}(c_1) = \frac{\alpha}{2}, \quad F_{n-1}(c_2) = 1 - \frac{\alpha}{2}$$

  die Konstanten c_1 und c_2. Wegen $n = 15$ und $\alpha = 0.01$ lesen wir in Tabelle (45) ab:

$$c_1 = 4.075 \quad c_2 = 31.319.$$

- Wir berechnen die Grenzen des Annahmebereichs unter Verwendung des hypothetischen Wertes σ_0:

$$c_u := \frac{c_1\sigma_0^2}{n-1} = \frac{4.075 \times 1.2^2}{14} = 0.41, \quad c_o := \frac{c_2\sigma_0^2}{n-1} = \frac{31.319 \times 1.2^2}{14} = 3.22.$$

- Wegen $s^2 = 2.25 \in [0.41, 3.22]$ kann die Hypothese $\sigma = 1.2$ mit einer Irrtumswahrscheinlichkeit von 1 % akzeptiert werden.

Für 10 % Irrtumswahrscheinlichkeit erhalten wir:

- $H_0 : \sigma = \sigma_0$, $H_1 : \sigma \neq \sigma_0$, $\sigma_0 := 1.2$
- Wir wählen Irrtumswahrscheinlichkeit $\alpha = 0.1$.
- Wir berechnen aus den Gleichungen

$$F_{n-1}(c_1) = \frac{\alpha}{2}, \quad F_{n-1}(c_2) = 1 - \frac{\alpha}{2}$$

  die Konstanten c_1 und c_2. Wegen $n = 15$ und $\alpha = 0.1$ lesen wir in Tabelle (45) ab:

$$c_1 = 6.571 \quad c_2 = 23.685.$$

- Wir berechnen die Grenzen des Annahmebereichs unter Verwendung des hypothetischen Wertes σ_0:

$$c_u := \frac{c_1\sigma_0^2}{n-1} = \frac{6.571 \times 1.2^2}{14} = 0.68 \quad c_o := \frac{c_2\sigma_0^2}{n-1} = \frac{23.685 \times 1.2^2}{14} = 2.44$$

- Wegen $s^2 = 2.25 \in [0.68, 2.44]$ kann die Hypothese $\sigma = 1.2$ mit einer Irrtumswahrscheinlichkeit von 10 % akzeptiert werden.

◇

3.8.5 Test für den Mittelwertvergleich bei bekannten Standardabweichungen σ_1 und σ_2

[13] S.570

Es soll geprüft werden, ob die Mittelwerte zweier normalverteilten Grundgesamtheiten übereinstimmen. Dabei wird die Kenntnis der Standardabweichungen σ_1 und σ_2 vorausgesetzt. Dieser Test korrespondiert mit der Konstruktion eines Vertrauensintervalls für μ entsprechend Abschnitt 3.4.1.

Test 7.

- $H_0 : \mu_0 = \mu_1$, $H_1 : \mu_0 \neq \mu_1$.
- Wähle Irrtumswahrscheinlichkeit $\alpha \in \{0.01, 0.05, 0.10\}$ und bestimme zugehöriges z anhand von

α	0.010	0.050	0.100
z	2.576	1.960	1.645

$$z = \Phi_{0,1}^{-1}\left(1 - \frac{\alpha}{2}\right).$$

- Berechne die Grenzen des Annahmebereichs unter Verwendung der Stichprobenumfänge n_1 und n_2 und der als bekannt vorausgesetzten Varianzen σ_1^2 und σ_2^2

$$c_u := -z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \quad c_o := z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

- Berechne die Stichprobenmittel $\overline{\boldsymbol{x}}$ und $\overline{\boldsymbol{y}}$ der Stichproben $x_1, \ldots, x_{n_1}$ und $y_1, \ldots, y_{n_2}$

$$\overline{\boldsymbol{x}} := \frac{1}{n_1}\sum_{i=1}^{n_1} x_i, \quad \overline{\boldsymbol{y}} := \frac{1}{n_2}\sum_{i=1}^{n_2} y_i.$$

- Entscheidungsregel: Falls $\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}} \in [c_u, c_o]$, kann H_0 mit der Irrtumswahrscheinlichkeit α angenommen werden.

Bemerkung: Falls X die zur ersten Grundgesamtheit gehörende Zg und Y die zur zweiten Grundgesamtheit gehörende Zg darstellt, so stimmt die Größe

$$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

mit der Standardabweichung von $\overline{X}-\overline{Y}$ überein.

Beispiel 55 Lebensdauer von Glühbirnen, [13], S.575

Es soll mit 1 % Irrtumswahrscheinlichkeit geprüft werden, ob die auf zwei verschiedenen Maschinen hergestellten Glühbirnen im Mittel die gleiche Lebensdauer besitzen. Dabei darf davon ausgegangen werden, daß die beiden Zufallsgrößen

$X :=$ Lebensdauer einer auf Maschine 1 hergestellten Glühbirne
$Y :=$ Lebensdauer einer auf Maschine 2 hergestellten Glühbirne

normalverteilt sind. Es werden zwei Stichproben von unterschiedlichem Umfang gezogen und die Stichprobenmittel berechnet:

$$\begin{array}{lll} \text{Maschine 1:} & n_1 = 80 & \overline{\boldsymbol{x}} = 520 \\ \text{Maschine 2:} & n_2 = 50 & \overline{\boldsymbol{y}} = 500 \end{array}$$

Die zu Maschine 1 und 2 jeweils dazugehörenden Standardabweichungen werden als bekannt vorausgesetzt:

$$\sigma_1 = 50, \quad \sigma_2 = 45.$$

Alle Angaben sind in h, die Einheit lassen wir aber weg.

Wir arbeiten Test 7 ab.

- $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$.
- Wir wählen Irrtumswahrscheinlichkeit $\alpha = 0.01$ und bestimmen zugehöriges z entsprechend $z = 2.576$.
- Wir berechnen die Standardabweichung von $\overline{X}-\overline{Y}$:

  $$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = \sqrt{\frac{2500}{80}+\frac{2025}{50}} = 8.471$$

  Daher gilt für die Grenzen des Annahmebereichs

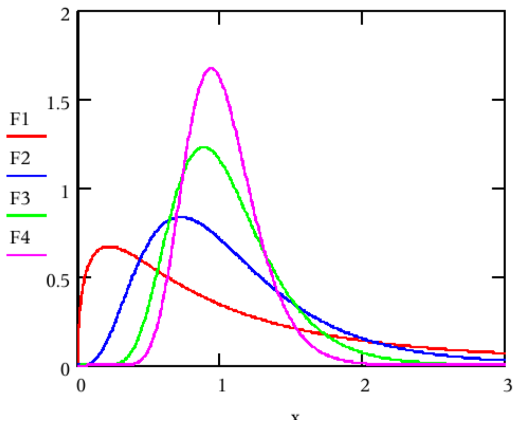
  $$c_u = -2.576 \times 8.471 = -21.82, \quad c_o = 21.82.$$

- Wegen

  $$\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}} = 20 \in [c_u, c_o]$$

  kann mit einer Irrtumswahrscheinlichkeit von 1 % die Gleichheit der mittleren Lebensdauern angenommen werden.

$\diamond$



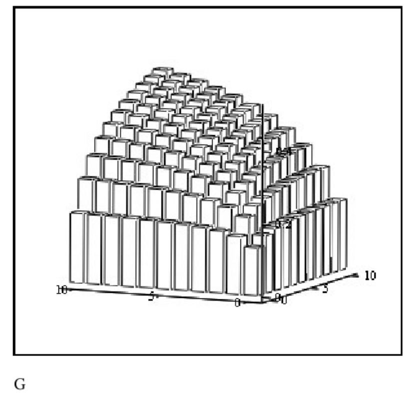


Abbildung 30: Das linke Bild zeigt $f(x, m_1, m_2)$ für $x \in [0, 3]$ und für $m_1 = 3$, $m_2 = 4$ (F1), $m_1 = 10$, $m_2 = 20$ (F2), $m_1 = 30$, $m_2 = 40$ (F3), $m_1 = 50$, $m_2 = 100$ (F4). Das rechte Bild zeigt $f(x, m_1, m_2)$ für $x = 1.2$ und $m_1, m_2 \in \{1, \ldots, 11\}$. Jede Säule korrespondiert mit einem Paar $(m_1, m_2) \in \{1, \ldots, 11\} \times \{1, \ldots, 11\}$. Beide Bilder wurden mit dem mathcad-Befehl dF(x, m_1, m_2) erzeugt.

3.8.6 *F*-Test

[20] S.1064

Ziel dieses Tests ist es festzustellen, ob zwei normalverteilte Grundgesamtheiten unterschiedliche Streuungen besitzen. Der Test benutzt die F-Verteilungsfunktion, deren Dichte durch

$$f(x, m_1, m_2) = \begin{cases} 0, & \text{falls } x < 0 \\ m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}} \dfrac{\Gamma(\frac{m_1}{2} + \frac{m_2}{2})}{\Gamma(\frac{m_1}{2})\Gamma(\frac{m_2}{2})} \dfrac{x^{\frac{m_1}{2}-1}}{(m_1 x + m_2)^{\frac{m_1+m_2}{2}}}, & \text{falls } 0 \leq x \end{cases}$$

gegeben ist (Vgl. [2], S. 706). Die natürlichen Zahlen m_1 und m_2 heißen die Freiheitsgrade von

$$F_{m_1,m_2}(x) := \int_{-\infty}^{x} f(t, m_1, m_2) \mathrm{d}t.$$

Tabelle (56) zeigt den Werteverlauf der Inversen F-Verteilungsfunktion, genauer

$$F^{-1}_{m_1,m_2}(x), \quad x := 1 - \alpha, \quad \alpha := 0.05, \quad m_1, m_2 \in \{1, \ldots, 10\}.$$

Auf dem Rand verlaufen die Freiheitsgrade, im Inneren steht der Wert von $F^{-1}_{m_1,m_2}(1-\alpha)$. Die Tabelle wurde mit dem mathcad-Befehl qF$(1 - \alpha, m_1, m_2)$, $\alpha = 0.05$ erzeugt.

Ausführlicheres Zahlenmaterial findet man in [20] S.94.

$m_2 \diagdown m_1$	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.543	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.384	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.84	8.812	8.78
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.998	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.772	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.099	4.06
7	5.59	4.74	4.35	4.12	3.97	3.86	3.79	3.73	3.676	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.388	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.178	3.14
10	4.96	4.10	3.71	3.48	3.32	3.22	3.14	3.07	3.020	2.98

(56)

Worin besteht nun der F-Test?

Test 8.

- Die Hypothese lautet $H_0 : \sigma_1 = \sigma_2$.
- Zur Überprüfung zieht man aus der ersten normalverteilten Grundgesamtheit eine Stichprobe $x_1, \ldots, x_{n_1}$ und aus der zweiten normalverteilten Grundgesamtheit eine Stichprobe $y_1, \ldots, y_{n_2}$. Dann bildet man die empirischen Varianzen

  $$s_x^2, \quad s_y^2$$

  wobei ohne Beschränkung der Allgemeinheit

  $$s_x^2 > s_y^2$$

  vorausgesetzt wird.
- Als Prüfgröße benutzt man den Quotienten

  $$F := \frac{s_x^2}{s_y^2}.$$

- Die Entscheidungsregel lautet dann: Falls

  $$F \leq F_{m_1,m_2}^{-1}(1-\alpha), \quad m_1 := n_1 - 1, \quad m_2 := n_2 - 1$$

  so kann H_0 mit der Irrtumswahrscheinlichkeit α angenommen werden.

Bemerkungen

- Es zeigt sich, daß für übereinstimmende Varianzen der Quotient S_x^2/S_y^2 entsprechend F_{n_1-1,n_2-1} verteilt ist, d.h.

$$P\left(\frac{S_x^2}{S_y^2} \le x\right) = F_{m_1,m_2}(x), \quad S_x^2 = \frac{1}{m_1}\sum_{k=1}^{n_1}(X_k - \overline{X})^2, \quad S_y^2 = \frac{1}{m_2}\sum_{k=1}^{n_2}(Y_k - \overline{Y})^2.$$

  Bestimmt man daher x so, daß

$$P(S_x^2/S_y^2 \le x) = 1-\alpha$$

  so gilt $x = F^{-1}_{m_1,m_2}(1-\alpha)$. Da α klein gewählt wird, ist die Wahrscheinlichkeit dafür, daß die Ungleichung

$$F \le F^{-1}_{m_1,m_2}(1-\alpha)$$

  erfüllt wird, ziemlich groß, sofern tatsächlich $\sigma_1 = \sigma_2$ gilt. Daher akzeptiert man mit der Irrtumswahrscheinlichkeit α, falls sich $F \le F^{-1}_{m_1,m_2}(1-\alpha)$ feststellen läßt.

- Falls $n_1, n_2 \le 11$ und $\alpha = 0.05$, kann Tabelle (56) für die Bestimmung von $F^{-1}_{m_1,m_2}(1-\alpha)$ benutzt werden.

- Falls beide Grundgesamtheiten gleiche Varianz besitzen, so ist für große Stichprobenumfänge der Quotient F ungefähr 1. Offenbar sind in Tabelle (56) alle Einträge größer als 1, was die Akzeptanz von H_0 bei $F \approx 1$ zur Folge hat.

Beispiel 56 Mittels des mathcad-Befehls rnorm(n, μ, σ) werden für

$$n_1 := 10,\ \mu_1 := 1,\ \sigma_1 := 3, \quad n_2 := 7,\ \mu_2 := 2,\ \sigma_2 := 2$$

zwei normalverteilte Zufallsvektoren erzeugt mit dem Ergebnis

$$x^{\mathrm{T}} = [-1.018,\ -0.843,\ 2.571,\ -1.232,\ 5.588,\ 0.158,\ -1.228,\ 0.322,\ -3.156,\ -2.301]$$

$$y^{\mathrm{T}} = [\quad 0.458, \quad 2.252,\ 1.041,\ -0.574,\ 0.682,\ 0.935, \quad 3.253].$$

Man bestätige mittels des F-Tests $\sigma_1 \ne \sigma_2$ mit einer Irrtumswahrscheinlichkeit von 5 %.

Laut Aufgabenstellung hat man $\alpha = 0.05$ zu wählen. Aufgrund der vorliegenden Daten bekommt man $m_1 = 9$, $m_2 = 6$ und

$$\overline{\boldsymbol{x}} = -0.114, \quad s_x^2 = 6.429, \quad \overline{\boldsymbol{y}} = 1.153, \quad s_y^2 = 1.554.$$

Offenbar gilt $s_x^2 > s_y^2$, so daß Test 8 in der Form angewendet werden kann, wie oben angegeben. Die Prüfgröße F erfüllt

$$F = \frac{6.429}{1.554} = 4.137.$$

Schließlich liest man aus Zeile 6 und Spalte 9 in Tabelle (56) den Wert 4.099 ab, sodaß die Hypothese $\sigma_1 = \sigma_2$ mit einer Irrtumswahrscheinlichkeit von 5 % abgelehnt werden kann. ◇

3.8.7 Test für den Mittelwertvergleich bei unbekannten aber gleichen Varianzen, Doppelter t-Test

[13] S.578

In Abschnitt 3.8.5 wurde die Hypothese getestet, ob zwei normalverteilte Grundgesamtheiten gleichen Mittelwert besitzen, wobei die Kenntnis der zugehörigen Varianzen vorausgesetzt wurde. Jetzt wird die Hypothese $\mu_1 = \mu_2$ ohne diese Voraussetzung getestet, allerdings geht man von $\sigma_1 = \sigma_2$ aus, was mit dem F-Test (Test 8) überprüft werden kann. Außerdem setzt man voraus, daß die Betrachtung von Grundgesamtheit 1 unabhängig ist von der Betrachtung der Grundgesamtheit 2.

Test 9.

- $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$.
- Wähle Irrtumswahrscheinlichkeit $\alpha \in \{0.01, 0.05, 0.10\}$.
- Ziehe aus Grundgesamtheit 1 eine Stichprobe im Umfang n_1 und aus Grundgesamtheit 2 eine Stichprobe im Umfang n_2.
- Berechne die empirischen Mittelwerte $\overline{\boldsymbol{x}}_1$ und $\overline{\boldsymbol{x}}_2$ und die Stichprobenvarianzen s_1^2 und s_2^2.
- Bestimme z aus der Gleichung

  $$F_{n_1+n_2-2}(z) = 1 - \frac{\alpha}{2}$$

  wobei F_n die t-Verteilungsfunktion für n Freiheitsgrade darstellt. Falls $n \in \{1, \ldots, 30\}$ kann Tabelle (79) benutzt werden.
- Als Prüfgröße setzen wir

  $$t := \frac{\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2}{s\sqrt{n_1^{-1} + n_2^{-1}}}, \quad s := \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

- Entscheidungsregel: Falls $|t| \leq z$, kann H_0 mit der Irrtumswahrscheinlichkeit α angenommen werden.

Bemerkungen:

- Ersetzt man in der Definition von t die Stichprobenmittel und -varianzen durch die zugehörigen Zg, so kommt man zu einer Zg T, die mit $n := n_1 + n_2 - 2$ Freiheitsgraden t-verteilt ist.

- Die Konstante z resultiert aus der Bedingung

$$P(-z \leq T \leq z) = 1 - \alpha.$$

Demnach bekommt man aus der Symmetrie der Dichtefunktion der t-Verteilung

$$1 - \alpha = F_n(z) - F_n(-z) = F_n(z) - (1 - F_n(z)) = 2F_n(z) - 1 \Rightarrow 1 - \frac{\alpha}{2} = F_n(z).$$

Beispiel 57 Als Stichproben haben wir

$$\boldsymbol{x} := [2.517, 2.549, 2.404, 1.372]^{\mathrm{T}}, \quad \boldsymbol{y} := [1.574, 2.418, 1.891, 1.613, 1.566, 1.690]^{\mathrm{T}}$$

die zwei normalverteilten Grundgesamtheiten mit unbekannter aber gleicher Varianz entstammen. Mit $\alpha = 0.05$ ist zu überprüfen, ob die Mittelwerte ebenfalls übereinstimmen.

Zu diesem Zweck arbeiten wir Test 9 ab.

Offenbar gilt $n_1 = 4$ und $n_2 = 6$. Direkte Berechnung liefert

$$\overline{\boldsymbol{x}} = 2.177, \quad \overline{\boldsymbol{y}} = 1.788, \quad s_x^2 = 0.243, \quad s_y^2 = 0.111.$$

Daher gilt

$$s = \sqrt{\frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}} = 0.412$$

sodaß

$$t = \frac{\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}}{s\sqrt{n_1^{-1} + n_2^{-1}}} = 1.563.$$

Wegen $1 - \alpha/2 = 0.975$ und $n_1 + n_2 - 2 = 9$ hat man in Tabelle (79) mit Zeile 9 und Spalte 3 zu arbeiten. Dort liest man für z den Wert 2.262 ab. Da

$$1.563 = |t| < 2.262 = z$$

kann man mit einer Irrtumswahrscheinlichlkeit von 5% die Gleichheit der Mittelwerte akzeptieren. $\diamond$

3.8.8 Berechnung notwendiger Stichprobenumfänge

Entsprechend des Abschnitts *Intervallschätzungen* beeinflußt der gewählte Stichprobenumfang die Breite des Überdeckungsintervall wiefolgt

zu überdeckender Parameter	Vertrauensintervall	Testverteilung
μ von $X \sim \mathcal{N}(\mu, \sigma^2)$ bei bekanntem σ	$[\overline{\boldsymbol{x}} - c, \overline{\boldsymbol{x}} + c]$, $c = \frac{\sigma z}{\sqrt{n}}$	Normalverteilung
μ von $X \sim \mathcal{N}(\mu, \sigma^2)$ bei unbekanntem σ	$[\overline{\boldsymbol{x}} - c, \overline{\boldsymbol{x}} + c]$, $c = \frac{sz}{\sqrt{n}}$	t-Verteilung
σ von $X \sim \mathcal{N}(\mu, \sigma^2)$	$\left[\sqrt{\frac{n-1}{c_2}}s, \sqrt{\frac{n-1}{c_1}}s\right]$	χ^2-Verteilung
Anteilswert p	$[\widehat{p} - c, \widehat{p} - c]$, $c = z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$	Normalverteilung

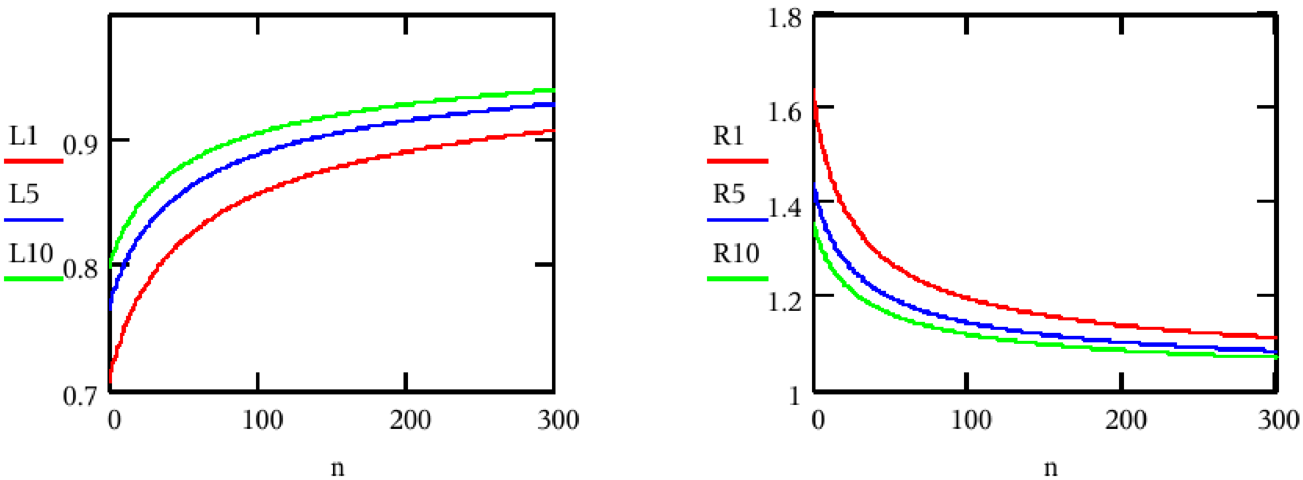


Abbildung 31: Das linke Bild zeigt den Werteverlauf von $\sqrt{\dfrac{n-1}{c_2}}$ und das rechte Bild den Werteverlauf von $\sqrt{\dfrac{n-1}{c_1}}$ für $n = 21, \ldots 321$, wobei $c_1 = F_{n-1}^{-1}(\alpha/2)$, $c_2 = F_{n-1}^{-1}(1 - \alpha/2)$ und F_n die χ^2-Verteilungsfunktion für n Freiheitsgrade darstellt. Die roten Kurven korrespondieren mit $\alpha = 0.01$, die blauen mit $\alpha = 0.05$ und die grünen mit $\alpha = 0.1$. Wie erwartet nehmen die Intervallbreiten mit zunehmender Irrtumswahrscheinlichkeit ab. Offenbar driftet mit zunehmendem Stichprobenumfang das Konfindenzintervall für σ nach links und wird dabei kleiner. Es schrumpft allerdings nicht auf beliebig kleine Größe zusammen, sondern für große n liest man unabhängig von $\alpha \in \{0.01, 0.05, 0.1\}$ eine Breite von etwa $s(1.1 - 0.95)$ ab.

Da der Mittelwert $\overline{\boldsymbol{x}}$ und die die Stichprobenvarianz s^2 erwartungstreu sind und entsprechend des Gesetzes der großen Zahl von Bernoulli $\widehat{p}$ fast sicher gegen p konvergiert, kann man in den Zeilen 1,2 und 4 den Stichprobenumfang n so groß wählen, daß bei konstant gehaltener Irrtumswahrscheinlichkeit α das Überdeckungsintervall beliebig klein wird. Bei der Variation von n in Zeile 3 kann man sich an der Abbildung 31 orientieren.

3.9 Statistische Prüfverfahren für die unbekannte Verteilungsfunktion einer Grundgesamtheit (nichtparametrische Tests)

3.9.1 χ^2-Anpassungstest für hypothetisch normalverteilte Zg

[20] S.1069

Der Test soll eine Entscheidung liefern, ob eine Grundgesamtheit normalverteilt ist oder nicht. Eine statistische Hypothese, die die unbekannte Gestalt der Verteilungsfunktion einer Zg festlegt, nennt man **nichtparametrisch**. Ein Verfahren, das zur Verifizierung einer nichtparametrischen Hypothese dient, nennt man einen **Anpassungstest** oder nichtparametrischen Test. Der folgende Test gehört in diese Klasse.

Test 10.

- H_0 : Grundgesamtheit ist normalverteilt.
- Zur Überprüfung von H_0 zieht man einen Stichprobenvektor

  $$\boldsymbol{x} := [x_1, \ldots, x_n]^{\mathrm{T}}$$

  vom Umfang n, bestimmt den empirischen Mittelwert und die empirische Streuung

  $$\overline{\boldsymbol{x}} := \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{\boldsymbol{x}})^2}.$$

- Dann nimmt man eine Klasseneinteilung vom Umfang k der Stichprobenwerte vor, d.h. man wählt Zahlen A_i mit

  $$A_1 < \ldots < A_k < A_{k+1}$$

  und bestimmt für $r = 1, \ldots, k$ die Anzahl h_r der Komponenten von $\boldsymbol{x}$, die sich in der Klasse

  $$K_r := [A_r, A_{r+1})$$

  befinden. Die Zahlen h_r werden **absolute Häufigkeiten** genannt.

- Dann berechnet man die hypothetischen Wahrscheinlichkeiten

  $$p_r := \Phi_{0,1}\left(\frac{A_{r+1} - \overline{\boldsymbol{x}}}{s}\right) - \Phi_{0,1}\left(\frac{A_r - \overline{\boldsymbol{x}}}{s}\right), \quad r = 1, \ldots, k.$$

  An dieser Stelle geht die Hypothese ein, da man die hypothetischen Wahrscheinlichkeiten mittels $\Phi_{0,1}$ berechnet.

- Dann bildet man die aus der Klasseneinteilung resultierende Testgröße

  $$c^2 := \sum_{r=1}^{k} \frac{(h_r - np_r)^2}{np_r}.$$

  Die rechte Seite dieser Definition wird die **Pearsonsche Stichprobenfunktion** genannt. Sie hängt offenbar von der Klasseneinteilung, den Meßdaten und derem Umfang ab. In der Literatur wird anstatt c^2 auch oft χ^2 geschrieben. Die Zahlen np_r heißen **theoretische Besetzungszahlen**. Die Testgröße c^2 wird offenbar klein, falls die beobachteten absoluten Häufigkeiten mit den theoretischen Besetzungszahlen in etwa übereinstimmmen.

- Man löst zu vorgegebener Irrtumswahrscheinlichkeit α die Gleichung
$$F_m(x) = 1 - \alpha$$
wobei
$$m := k - 3$$
und $F_m(x)$ die χ^2-Verteilungsfunktion mit m Freiheitsgraden darstellt.

- Entscheidungsregel: Falls die Ungleichung
$$c^2 \leq F_m^{-1}(1 - \alpha)$$
erfüllt wird, kann man die Hypothese, daß die Stichprobe einer normalverteilten Grundgesamtheit entstammt, mit einer Irrtumswahrscheinlichkeit von α annehmen.

Bemerkungen:

- Damit der Freiheitsgrad m positiv wird, ist wenigstens mit einer Einteilung in vier Klassen zu arbeiten.

- Es ist darauf zu achten, daß für $r = 1, \ldots, k$ die Ungleichung
$$np_r \geq 5$$
erfüllt wird [20], S.1070, [15], S.177.

- Offenbar kann durch Klassenzusammenlegung die Bedingung $np_r \geq 5$ erreicht werden, da
$$[A_{r+1}, A_r) \cup [A_r, A_{r-1}) = [A_{r+1}, A_{r-1})$$
mit der hypothetischen Wahrscheinlichkeit $p_r + p_{r-1}$ korrespondiert.

- Für die gebräuchlichen Irrtumswahrscheinlichkeiten 1 %, 5% und 10 % und Klasseneinteilungen bis ca. $k = 30$ kann Tabelle (80) zur Lösung von $F_m(x) = 1-\alpha$ benutzt werden.

- In Beispiel 39 wurde schon einmal eine Klasseneinteilung vorgenommen, um optisch zu prüfen, ob Normalverteilung vorliegt. Für die **relativen Häufigkeiten**
$$f_r := h_r/n$$
hatten wir die Approximation
$$f_r \approx \Phi_{\mu,\sigma}(A_{r+1}) - \Phi_{\mu,\sigma}(A_r)$$
festgestellt. Übergang zur standardisierten Normalverteilung liefert
$$f_r \approx \Phi_{0,1}\left(\frac{A_{r+1} - \mu}{\sigma}\right) - \Phi_{0,1}\left(\frac{A_r - \mu}{\sigma}\right)$$
wodurch der Zusammenhang jeweils zwischen den Zahlen f_r und p_r und den Zahlen h_r und np_r klar wird. Offenbar wird die Testgröße c^2 klein, falls die Vektoren $[h_r]_{r=1}^k$ und $[np_r]_{r=1}^n$ geringen Abstand besitzen. Die Division mit np_r bringt die χ^2-Verteilung ins Spiel.

- Für die zu wählende Anzahl k von Klassen gibt es verschiedene Empfehlungen: falls man mit konstant breiten Klassen arbeiten möchte, soll man $k \approx \sqrt{n}$ wählen, eine andere Empfehlung geht auf Brooks/Carruther zurück, die $k \leq 5 \log n$ nahelegen.

Beispiel 58 Es ist mit einer Irrtumswahrscheinlichkeit von 5 % zu testen, ob folgende aufsteigend sortierte Meßwerte einer normalverteilten Grundgesamtheit entstammen:

$$\begin{array}{lllllllll} 0.360 & 0.398 & 0.482 & 0.652 & 0.663 & 0.762 & 0.795 & 0.804 & \\ 0.857 & 0.934 & 0.990 & 0.992 & 1.016 & 1.081 & 1.090 & 1.119 & \\ 1.132 & 1.139 & 1.214 & 1.246 & 1.257 & 1.370 & 1.380 & 1.509 & 2.524. \end{array} \tag{57}$$

Offenbar gilt $n = 25$, da man 25 Meßwerte zur Verfügung hat. Man berechnet

$$\overline{\boldsymbol{x}} = 1.031, \quad s = 0.434.$$

Unter Berücksichtigung von $np_r \geq 5$ wählen wir als Klasseneinteilung

$$K_1 := [0, 0.7), \quad K_2 := [0.7, 0.95), \quad K_3 := [0.95, 1.2), \quad K_4 := [1.2, 2.6)$$

d.h. $k = 4$ und

$$A := [A_i]_{i=0}^4 := [0,\ 0.7,\ 0.95,\ 1.2,\ 2.6]^{\mathrm{T}}.$$

In Abhängigkeit von $\overline{\boldsymbol{x}}$, s und A bekommen wir

$$p_1 = 0.214, \quad p_2 = 0.203, \quad p_3 = 0.226, \quad p_4 = 0.348.$$

Multiplikation mit $n = 25$ zeigt, daß die Bedingung $np_r \geq 5$ eingehalten wird.

Zu Übungszwecken berechnen wir die theoretische Besetzungszahl np_1 ausführlich. Nach Definition gilt

$$\begin{aligned} p_1 = \quad & \Phi_{0,1}\left(\frac{A_1 - \overline{\boldsymbol{x}}}{s}\right) - \Phi_{0,1}\left(\frac{A_0 - \overline{\boldsymbol{x}}}{s}\right) = \Phi_{0,1}\left(\frac{0.7 - 1.031}{0.434}\right) - \Phi_{0,1}\left(\frac{-1.031}{0.434}\right) \\ = \quad & \Phi_{0,1}\left(-0.763\right) - \Phi_{0,1}\left(-2.376\right). \end{aligned}$$

Wegen $\Phi_{0,1}(x) = 1 - \Phi_{0,1}(-x)$ bekommt man

$$p_1 = \Phi_{0,1}\left(2.376\right) - \Phi_{0,1}\left(0.763\right) \approx 0.9911 - 0.7764 = 0.2147.$$

Die letzte Rechenschritt wurde mit Tabelle (78) ausgeführt. Das kostenlos verfügbare Berechnungssystem **octave** bietet die Befehle

$$\text{mean}(x), \quad \text{var}(x), \quad \text{normal\_cdf}(x)$$

an, deren Auswertung jeweils $\overline{\boldsymbol{x}}$, s und $\Phi_{0,1}(x)$ liefert. Der Anteil 'normal' deutet auf die Normalverteilung, der Anteil 'cdf' auf 'cumulative distribution function'. Die Kombination dieser drei Befehle gestattet eine mühelose Berechnung der Zahlen p_r. Die Argumente von 'mean' und 'var' sind Vektoren. Auch das Argument von normal_cdf kann ein Vektor sein: für $\boldsymbol{x} = [x_i]_{i=1}^n$ erhält man

$$[\Phi_{0,1}(x_i)]_{i=1}^n := \text{normal\_cdf}(\boldsymbol{x}).$$

Setzt man daher

$$a := \frac{1}{s}\left(\begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\overline{\boldsymbol{x}}\right) \in \mathbb{R}^4, \quad b := \frac{1}{s}\left(\begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\overline{\boldsymbol{x}}\right) \in \mathbb{R}^4$$

so bekommt man

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = \text{normal\_cdf}(a) - \text{normal\_cdf}(b).$$

Für die zugehörigen absoluten Häufigkeiten gilt

$$h_1 = 5, \quad h_2 = 5, \quad h_3 = 8, \quad h_4 = 7.$$

Unter Benutzung der Zahlen h_r und np_r berechnen wir die Prüfgröße c^2 entsprechend

$$c^2 = \sum_{r=1}^{4} \frac{(h_r - np_r)^2}{np_r} = 1.343.$$

Schließlich ist $F_{k-3}^{-1}(1-0.05) = F_1^{-1}(1-0.05)$ auszuwerten. Blick in Tabelle (80) Zeile 1 Spalte 2 liefert den Wert 3.84. Da

$$c^2 < 3.84$$

kann mit einer Irrtumswahrscheinlichkeit von 5 % davon ausgegangen werden, daß die Daten (57) einer normalverteilten Grundgesamtheit angehören. $\diamond$

Die Tabelle aus Abschnitt 6.3 zeigt den Werteverlauf von $F_m^{-1}(1-\alpha)$, für $\alpha = 0.01, 0.05, 0.10$ und $m = 1, \ldots, 30$. Die Tabelle wurde mit dem mathcad-Befehl qchisq(x, m) erzeugt.

Ergänzug: Die Prüfgröße c^2 kann man als die Realisierung einer Zufallsgröße χ^2 auffassen, die die quadrierten Differenzen zwischen den theoretischen Besetzungszahlen und den tatsächlich gezählten Werten einer Zg X in einer bestimmten Klasse K_r aufsummiert:

$$\chi^2 := \sum_{r=1}^{k} \frac{(h_r - np_r)^2}{np_r}, \quad n = n_1 + \ldots + n_k.$$

Dabei bezeichnet h_r die Anzahl der Ausprägungen von X, die in der Klasse K_r beobachtet wurden. Außerdem können die theoretischen Besetzungszahlen zu einer beliebigen hypothetischen Verteilung gehören, jenachdem auf welche Verteilungsfunktion getestet wird.

Bezeichnet man mit

$$F(x,n) := P(\chi^2 \leq x)$$

die Verteilungsfunktion von χ^2, so gilt entsprechend [6] S.508 im Falle der exakten Kenntnis der Zahlen p_k die Grenzwertbeziehung

$$\lim_{n\to\infty} F(x,n) = F_{k-1}(x) := \begin{cases} 0 & \text{falls } x < 0, \\ \dfrac{1}{2^{\frac{k-1}{2}}\,\Gamma\left(\frac{k-1}{2}\right)} \displaystyle\int_0^x t^{\frac{k-3}{2}} e^{-\frac{t}{2}} \mathrm{d}t & \text{falls } 0 \leq x. \end{cases} \tag{58}$$

Durch einen Blick in Abschnitt 2.10.5 überzeugt man sich davon, daß für $n \to \infty$ die Zg χ^2 einer χ^2-Verteilung mit $k-1$ Freiheitsgraden gehorcht, wobei k die Anzahl der benutzten Klassen bezeichnet. Jenachdem mit welcher Irrtumswahrscheinlichkeit α man nun arbeiten möchte, berechnet man eine Konstante z, für die

$$F_{k-1}(z) = 1-\alpha$$

gilt. D.h. für ausreichend große Stichprobenumfänge n gilt bei Korrektheit der hypthetischen Besetzungszahlen die Gleichung

$$P(\chi^2 \leq z) = 1-\alpha.$$

Wählt man beispielsweise eine Irrtumswahrscheinlichkeit von 1 %, so fällt bei richtiger Hypothese die Prüfgröße c^2 mit einer Wahrscheinlichkeit von 99 % nicht größer aus als z. Demnach ist im Falle

$$c^2 > z$$

davon auszugehen, daß die theoretischen Besetzungszhalen nicht zur tatsächllichen Verteilung der Grundgesamtheit passen und man wahrscheinlich von einer falschen Verteilungshypothese ausgegangen ist. Die Bedingung $np_r \geq 5$ für $r = 1,\ldots,k$ hängt mit der Grenzwertbeziehung (58) zusammen. Erst wenn diese Ungleichungen erfüllt sind, kann $F(x,n)$ durch $F_{k-1}(x)$ ersetzt werden.

Falls die hypothetische Verteilungsfunktion von Parametern abhängt, die man nicht kennt und daher geschätzt werden müssen, vermindert sich die zu benutzende Anzahl von Freiheitsgraden um die Anzahl dieser Parameter. Bei Test 10 haben wir das unbekannte μ und σ durch $\overline{\boldsymbol{x}}$ und s geschätzt. Daher wird in Test 10 mit $k-3$ Freiheitsgraden gearbeitet. Falls man testen möchte, ob F die zur Grundgesamtheit gehörende Verteilungsfunktion ist, müssen die theoretischen Besetzungszahlen mittels F berechnet werden. Der Rest vom Test bleibt unverändert.

3.9.2 χ^2-Anpassungstest für hypothetisch exponentialverteilte Zg

Durch Beobachtung hat sich herausgestellt, daß eine Zg X innerhalb des Wertevorrats $[0,16]$ mit folgenden absoluten Häufigkeiten Werte angenommen hat:

Teilbereich K_i	$(0,1]$	$(1,2]$	$(2,3]$	$(3,4]$	$(4,6]$	$(6,8]$	$(8,16]$
absolute Häufigkeit h_i	15	14	10	13	15	14	19

Z.B. nahm X dreizehnmal einen Wert innerhalb von $(3,4]$ und fünzehnmal einen Wert innerhalb von $(4,6]$ an. Mit einer Irrtumswahrscheinlichkeit von 5% soll getestet werden, ob X exponentialverteilt ist. Als Realisierung von X kann man sich die Dauer eines Telefongesprächs vorstellen.

Entsprechend Abschnitt 2.10.3 bedeutet die aufgestellte Hypothese, daß die Dichtefunktion von der Gestalt

$$f(x) := \begin{cases} 0 & \text{falls } x < 0 \\ \lambda e^{-\lambda x} & \text{falls } 0 \leq x \end{cases}$$

ist, wobei λ einen zu schätzenden Parameter darstellt. Wegen

$$EX = \frac{1}{\lambda}$$

bietet sich als Schätzung der Kehrwert der durchschnittlichen Gesprächsdauer an:

$$\widehat{\lambda} := \frac{100}{(15\times0.5)+(14\times1.5)+(10\times2.5)+(13\times3.5)+(15\times5)+(14\times7)+(19\times12)} = 0.2.$$

Als Klasseneinteilung kann man die einzelnen Teilbereiche verwenden, sodaß man unmittelbar zu folgenden theoretischen Besetzungszahlen kommt: wegen

$$F(x) := \begin{cases} 0 & \text{falls } x < 0 \\ 1 - e^{-\lambda x} & \text{falls } 0 \leq x \end{cases}$$

gilt

$$\begin{array}{llll} F(0) = 0, & F(1) = 0.181, & F(2) = 0.330, & F(3) = 0.451 \\ F(4) = 0.551, & F(6) = 0.699, & F(8) = 0.789, & F(16) = 0.959 \end{array}$$

sodaß

$$\begin{array}{ll} p_1 := P(X \in K_1) = F(1) - F(0) = 0.181 & p_2 := P(X \in K_2) = F(2) - F(1) = 0.148 \\ p_3 := P(X \in K_3) = F(3) - F(2) = 0.122 & p_4 := P(X \in K_4) = F(4) - F(3) = 0.099 \\ p_5 := P(X \in K_5) = F(6) - F(4) = 0.148 & p_6 := P(X \in K_6) = F(8) - F(6) = 0.099 \\ p_7 := P(X \in K_7) = F(16) - F(8) = 0.161 & \end{array}$$

Wegen $n := 100$, ist die Ungleichung $np_i > 5$ für alle i erfüllt, sodaß die Klasseneinteilung benutzt werden kann. Als Prüfgröße bekommt man mit $n := 100$ den Wert

$$c^2 := \sum_{k=1}^{7} \frac{(h_i - np_i)^2}{np_i} = 4.092.$$

Da wir mit 7 Klassen und einem geschätztem Parameter gearbeitet haben, kommt die χ^2-Verteilung mit 5 Freiheitsgraden in Ansatz. Da mit $\alpha := 5\%$ gearbeitet werden soll, hat man in Tabelle (80) mit Zeile 5 und Spalte 2 zu arbeiten. Dort findet man den Wert 11.07. Demnach überschreitet die der Prüfgröße c^2 zugeordnete Zg χ^2 mit 95 % Wahrscheinlichkeit den Wert 11.07 nicht. Da schließlich

$$c^2 = 4.092 < 11.07$$

gilt, erfüllt c^2 diese Erwartung, sodaß die Hypothese der Exponentialverteilung von X mit einer Irrtumswahrswcheinlichkeit von 5% aufrecht erhalten werden kann.

3.9.3 Kolmogorow-Smirnow-Test

[20] S.1067

Zu seiner Beschreibung benötigen wir den Begriff der empirischen Verteilungsfunktion. Von einer Grundgesamtheit, die entsprechend F verteilt ist, sei eine Stichprobe

$$\boldsymbol{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$$

im Umfang n bekannt. Dann bekommt man die empirische Verteilungsfunktion entsprechend

$$F_n(x) := \frac{\text{Anzahl der Komponenten von } \boldsymbol{x}\text{, die kleiner oder gleich } x \text{ sind}}{n}.$$

Offenbar ist F_n eine Treppenfunktion. Um den Abstand zwischen F und F_n zu messen, setzen wir

$$d_n := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|.$$

Der Hauptsatz der Statistik stellt fest, daß für zunehmendes n die empirische Verteilungsfunktion F_n die Funktion F immer besser approximiert, man kann sogar zeigen, daß fast sicher

$$\lim_{n \to \infty} d_n = 0$$

gilt (W.I. Gliwenko, 1933, [6], S. 456). Der Test beruht nun auf dem Zusammenhang

$$\lim_{n \to \infty} P(\sqrt{n} d_n < \lambda) = Q(\lambda), \quad Q(\lambda) := \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2} \tag{59}$$

der ausführlich ab S.459 in [6] beschrieben wird. Dabei wird stetiges F vorausgesetzt und festgestellt, daß für unstetiges F der Zusammenhang (59) nicht gilt. Offenbar hängt der Grenzwert Q nicht von F ab, was für unstetiges F nicht stimmt.

Zu frei gewählter Irrtumswahrscheinlichkeit α bestimmt man λ_0 nun so, daß

$$Q(\lambda_0) = 1 - \alpha.$$

Wie Abbildung 32 zeigt, ist diese Gleichung nicht eindeutig lösbar. Entsprechend der Tabelle [20] S.100, wird die Lösung, die rechts von 0.3 liegt verwendet, sodaß man

α	0.01	0.05	0.10
λ_0	1.63	1.26	1.22

, $\quad Q(\lambda_0) = 1 - \alpha$

erhält. Demnach wird die Ungleichung

$$\sqrt{n} d_n < \lambda_0$$

mit 90%, 95% oder 99%-ger Wahrscheinlichkeit erfüllt, sofern die Hypothese, daß X entsprechend F verteilt ist, gilt. Hat man nun eine Stichprobe $\boldsymbol{x}$ gezogen, deren zugehöriges

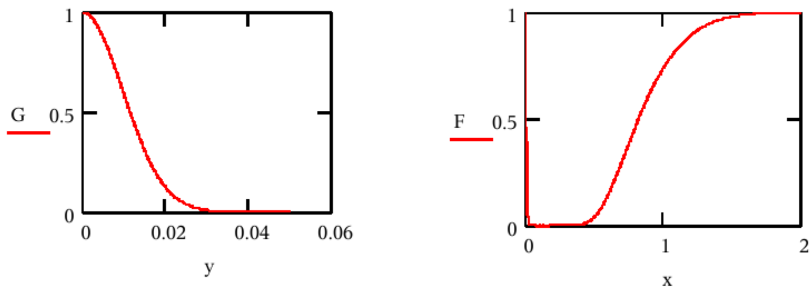


Abbildung 32: Es wird der Graph von $Q(\lambda)$ im Bereich $[0,2]$ gezeigt. Offenbar gilt $Q(0) = \sum_{k=-\infty}^{+\infty}(-1)^k$, was keine konvergente Reihe darstellt. Jenachdem, wie man die einzelnen Summanden zusammenfaßt, erhält man verschiedene 'Grenzwerte'. Aus Stetigkeitsgründen setzt man $Q(0) = 1$, was durch den linken Teil der Abbildung gerechtfertigt wird. Da Q eine gerade Funktion ist, d.h. $Q(-\lambda) = Q(\lambda)$, ist ihr Graph symmetrisch bezüglich der y-Achse. Augenscheinlich nimmt Q daher nur Werte innerhalb von $[0,1]$ an.

d_n diese Ungleichung nicht erfüllt, so weckt das unseren Argwohn, und wir lehnen mit der Irrtumswahrscheinlichkeit α die Hypothese ab, daß die Grundgesamtheit, aus der die Stichprobe $\boldsymbol{x}$ stammt, entsprechend F verteilt ist. Offenbar wird mit zunehmender Irrtumswahrscheinlichkeit das zugehörige λ_0 immer kleiner, sodaß unsere Neigung zur Ablehnung wächst. Der Test kann aufgrund der Grenzwertbeziehung (59) nur für große Stichprobenumfänge Anwendung finden und ist nicht geeignet für parameterabhängige Verteilungsfunktionen, deren Parameter aus $\boldsymbol{x}$ geschätzt werden müssen.

Beispiel 59 $\diamond$

4 Beschreibende Statistik

4.1 Beschreibung empirischer Verteilungsfunktion

Um F_n zu beschreiben, können die **Lageparameter**, die sich aus $\boldsymbol{x}$ ergeben, benutzt werden:

- Spannweite $S := \max(\boldsymbol{x}) - \min(\boldsymbol{x})$.
- Mittelwert $\overline{\boldsymbol{x}}$
- Median oder Zentralwert (Komponenten von $\boldsymbol{x}$ aufsteigend sortiert)

$$z := \begin{cases} x_{\frac{n+1}{2}} & \text{falls } n \text{ ungerade,} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{falls } n \text{ gerade.} \end{cases}$$

- Quartile $q_{\frac{1}{4}}$, $q_{\frac{2}{4}}$, $q_{\frac{2}{4}}$: Falls $n = 4m$ und die Komponenten von $\boldsymbol{x}$ aufsteigend sortiert sind, so setzt man

$$q_{\frac{1}{4}} := \text{Zahl zwischen } x_m \text{ und } x_{m+1}$$

$$q_{\frac{2}{4}} := \text{Zahl zwischen } x_{2m} \text{ und } x_{2m+1}$$

$$q_{\frac{3}{4}} := \text{Zahl zwischen } x_{3m} \text{ und } x_{3m+1}$$

  Interpretation: $q_{\frac{1}{4}}$ trennt das untere Viertel von $\boldsymbol{x}$ ab. Für nicht durch 4 teilbares n hat man diese Schranken in diesem Sinne möglichst gut zu wählen.

- Modus oder häufigster Wert D: D ergibt sich aus der Komponente von $\boldsymbol{x}$, die am häufigsten auftritt.

Eine Verteilungsfunktion F heißt **symmetrisch**, falls es einen Punkt a gibt, für den die Wahrscheinlichkeits- oder Dichtefunktion für alle $x \in \mathbb{R}$ die Gleichung

$$f(a + x) = f(a - x) \tag{60}$$

[15], S.104, erfüllt. Der Punkt a heißt das **Symmetriezentrum** [6], S. 96. Beispiele für symmetrische Verteilungsfunktionen sind

- die Normalverteilung $\Phi_{\mu,\sigma}$, vgl. Abbildung 9
- die t-Verteilung, vgl. Abbildung 11
- die Binomialverteilung für $p = 0.5$, vgl. Abbildung 4. Die Symmetrie ergibt sich aus der Symmetrie des Binomialkoeffizienten, vgl. (3)
- gleich- und dreiecksverteilte Zg, vgl. Abbildungen 1 und 8.

Zur Illustration asymmetrischer Verteilungsfunktionen kann Abbildung 5 dienen. Die dort betrachteten Zg nehmen alle die Werte $0, \dots, 25$ an. Betrachtet man die Verteilungsfunktion F für $p = 1/10$, so wächst F in $[0, 12.5]$ von 0 bis auf fast 1 und in $[12.5, 25]$ kaum noch, d.h. F wächst links viel stärker als rechts. Wegen $EX = np$ bekommen wir für $n = 25$ und $p = 1/10$ als Erwartungswert μ die Zahl 2.5, d.h. eine Stichprobe $\boldsymbol{x}$ erfüllt sehr wahrscheinlich die Ungleichung $\overline{x} < 12.5$. Daher deutet diese Ungleichung daraufhin, daß F links viel stärker wächst als rechts.

Die Lageparameter geben Hinweise auf die Symmetrieeigenschaften der Verteilungsfunktion:

- Die Gleichung $\overline{x} = z = D$ deutet auf eine symmetrische Verteilungsfunktion hin.
- Falls $D < z < \overline{x}$ hat man es wahrscheinlich mit einer Verteilungsfunktion zu tun, die rechts von z stärker wächst als links von z.
- Falls $\overline{x} < z < D$ hat man es wahrscheinlich mit einer Verteilungsfunktion zu tun, die links von z stärker wächst als rechts von z.

4.2 Die Momente einer Verteilung

Es sei F eine Verteilungsfunktion mit der Dichte f. Dann bezeichnet man als das k-te Moment von F bezüglich $a \in \mathbb{R}$ die Zahl

$$m_k(a) := \int_{-\infty}^{+\infty} (x-a)^k f(x)\mathrm{d}x.$$

Offenbar gelten die Gleichungen

$$m_1(0) = EX, \quad m_2(\mu) = \mathrm{Var}(X) = \sigma^2$$

falls X entsprechend F verteilt ist und μ wie üblich die Erwartung bezeichnet. In Hinblick auf

$$\mathrm{Var}(X) = EX^2 - (EX)^2$$

gilt

$$\mathrm{Var}(X) = m_2(0) - m_1(0)^2.$$

Falls von X nur n Merkmalswerte bekannt sind, d.h. man kennt nur eine Stichprobe $\boldsymbol{x}$ im Umfang n einer Grundgesamtheit, die entsprechend F verteilt ist, so gilt

$$m_k(a) \approx \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^k. \tag{61}$$

Daher stimmt das erste Moment bezüglich 0 ungefähr mit $\overline{\boldsymbol{x}}$ und das zweite Moment bezüglich μ ungefähr mit der Stichprobenvarianz s^2 überein.

4.3 Schiefe Dichte- und Verteilungsfunktionen

[17] S.81, [18] S.164

Eine nichtsymmetrische Dichte- oder Wahrscheinlichkeitsfunktion f heißt **schief**. An dieser Stelle zeigt sich die Bedeutung des dritten Moments bezüglich μ. Für eine symmetrische Verteilungsfunktion verschwindet es:

$$\begin{aligned} m_3(\mu) &= \int_{-\infty}^{+\infty}(x-\mu)^3 f(x)\mathrm{d}x = \int_{-\infty}^{+\infty} y^3 f(y+\mu)\mathrm{d}y \\ &= \int_{-\infty}^{0} y^3 f(y+\mu)\mathrm{d}y + \int_{0}^{+\infty} y^3 f(y+\mu)\mathrm{d}y \\ &= \int_{0}^{+\infty}(-y)^3 f(\mu-y)\mathrm{d}y + \int_{0}^{+\infty} y^3 f(y+\mu)\mathrm{d}y \\ &= \int_{0}^{+\infty} y^3 \underbrace{\left(f(\mu+y) - f(\mu-y)\right)}_{=0 \text{ wegen } (60)} \mathrm{d}y = 0. \end{aligned}$$

Wir nennen

- f **rechtsschief**, falls $m_3(\mu) < 0$

- f **linksschief**, falls $0 < m_3(\mu)$.

In [17], S.81 ist die Definition gerade anders herum.

Das folgende Beispiel gibt eine geometrische Interpretation.

Beispiel 60 Schiefe Dichtefunktion

Wir verwenden

$$f(x) := \begin{cases} 0 & \text{falls } x < a \\ \dfrac{c_1}{b-a} & \text{falls } a \leq x < b \\ 0 & \text{falls } b \leq x < c \qquad c_1 + c_2 = 1, \quad a < b \leq c < d \\ \dfrac{c_2}{d-c} & \text{falls } c \leq x < d \\ 0 & \text{falls } d \leq x \end{cases} \tag{62}$$

als Dichtefunktion. Der Erwartungswert μ beträgt

$$\mu = c_1 \frac{a+b}{2} + c_2 \frac{c+d}{2}.$$

Die Spezifizierung

$$a := 0,\ b := 1,\ c := 2,\ d := 3,\ c_1 := \frac{1}{4},\ c_2 := \frac{3}{4}$$

liefert $\mu = 2$. Insbesondere bekommt man

$$m_3(\mu) = m_3(2) = -\frac{3}{4}$$

was rechtsschiefes f bedeutet. Abbildung 33 visualisiert diese Verhältnisse. ◇

Zur Ergänzung betrachten wir noch ein Beispiel, in dem eine Stichprobe vorkommt.

Beispiel 61 Linkssteile Verteilungsfunktion

Wir betrachten einen Zufallsvektor mit Werten zwischen 0 und 12, der im linken Bild von Abbildung 34 gezeigt wird. Entsprechend der Klasseneinteilung

$$K_1 := [0,1),\ K_2 := [1,2),\ \ldots\ K_{12} := [11,12)$$

ergibt sich der Vektor

$$h := [4,8,12,14,14,13,12,10,9,7,5,4]^{\mathrm{T}}$$

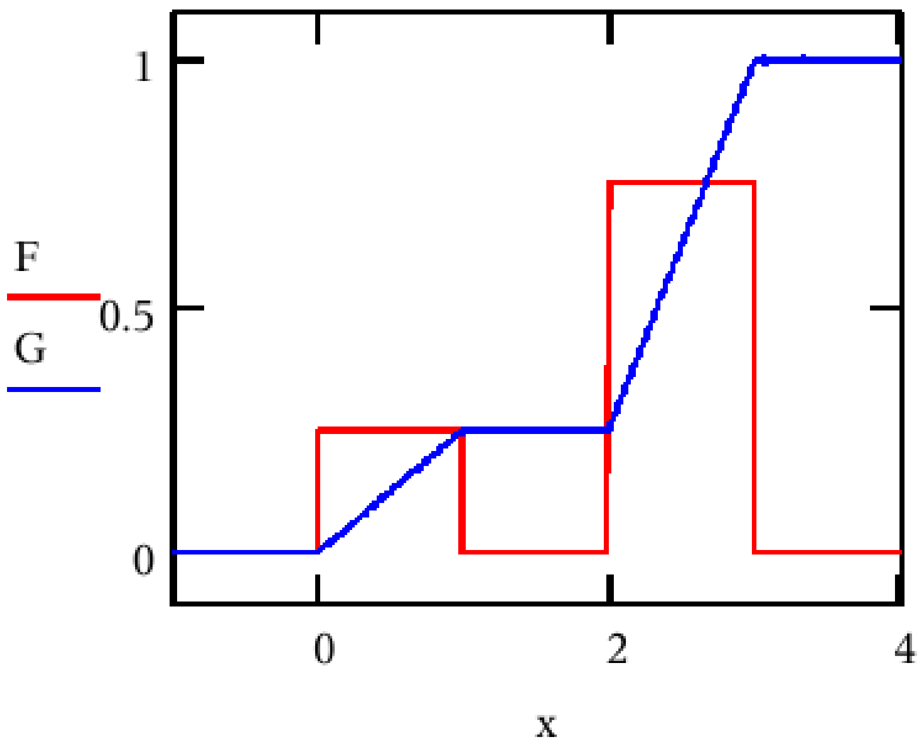


Abbildung 33: Die rote Treppenfunktion (F) stellt die Funktion (62) und die blaue Kurve (G) die zugehörige Verteilungsfunktion über $[-1, 4]$ dar. Offenbar wächst G innerhalb von $[0, 1.5]$ von 0 auf $c_1 = 0.25$ und in $[1.5, 3]$ von c_1 bis auf 1. Wir hatten $m_3(2) = -0.75$ festgestellt, d.h. f ist rechtsschief, was offensichtlich einer überwiegenden Ansiedlung der Wahrscheinlichkeitsmasse rechts bedeutet.

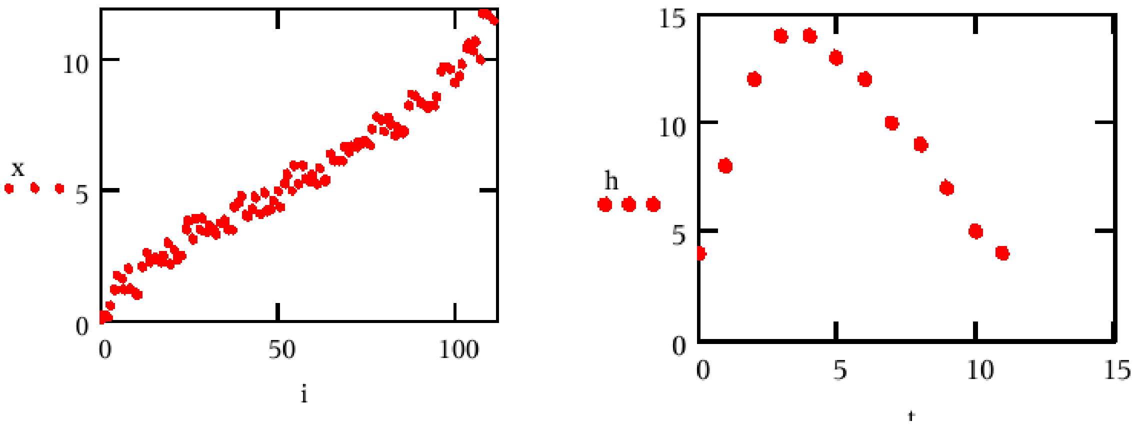


Abbildung 34: Das linke Bild zeigt einen Zufallsvektor der Länge 112 mit Werten zwischen 0 und 12. Das rechte Bild zeigt die zugehörigen Häufigkeiten für die Klasseneinteilung $[0,1),[1,2),\ldots,[11,12]$, die offensichtlich links größer sind als rechts. Die resultierende empirische Verteilungsfunktion bezeichnet man daher als linksschief.

bestehend aus den zugehörigen absoluten Häufigkeiten, der im rechten Bild gezeigt wird. Offenbar befinden sich die größeren Häufigkeiten im linken Bildbereich. Die resultierende empirische Verteilungsfunktion ist daher linksschief. Nun läßt sich diese Eigenschaft durch das Vorzeichen von $m_3(\mu)$ charakterisieren, wobei wir als Schätzung von μ den Mittelwert $\overline{\boldsymbol{x}}$ benutzen.

Für $m_3(\overline{\boldsymbol{x}})$ bekommt man

$$m_3(\overline{\boldsymbol{x}}) = \frac{1}{112}\sum_{k=1}^{112}(x_k - \overline{\boldsymbol{x}})^3 = 5.725, \quad \overline{\boldsymbol{x}} = 5.532.$$

Wegen $m_3(\overline{\boldsymbol{x}}) > 0$ kann auf eine linksschiefe Dichtefunktion geschlossen werden. $\diamond$

4.4 Wölbung

[18] S.165

Zur Beschreibung der **Wölbung** einer Verteilungsfunktion kann der Quotient

$$W := \frac{m_4(\mu)}{m_2^2(\mu)}$$

benutzt werden. Offenbar ergibt sich W aus dem vierten Moment bezüglich μ dividiert durch das Quadrat des zweiten Moments bezüglich μ. Es gilt die Relation: je größer W, desto aufgewölbter oder hochgipfliger ist die Verteilung. Dabei wird eine Verteilungsfunktion F_1 hochgipfliger als eine Verteilungsfunktion F_2 genannt, falls der Gipfel von f_1 höher ist als der Gipfel von f_2. Selbstverständlich bezeichnet f_k die Dichte- oder Wahrscheinlichkeitsfunktion von F_k. Wir nennen F_k aufgewölbt, falls der zugehörige Graph S-förmige Gestalt besitzt.

Beispiel 62 Mehr oder weniger gewölbte Verteilungsfunktion

Wir betrachten die Verteilungsfunktionen zu den Dichten

$$f(x,a) := \begin{cases} 0 & \text{falls } x < -a \\ -\dfrac{3}{4a^3}(x^2-a^2) & \text{falls } -a \le x < a \\ 0 & \text{falls } a \le x \end{cases}$$

$$g(x,a) := \begin{cases} 0 & \text{falls } x < -a \\ \dfrac{x}{a^2}+\dfrac{1}{a} & \text{falls } -a \le x < 0 \\ -\dfrac{x}{a^2}+\dfrac{1}{a} & \text{falls } 0 \le x < a \\ 0 & \text{falls } a \le x \end{cases}$$

und setzen wie üblich

$$F_a(x) := \int_{-\infty}^{x} f(t,a)\mathrm{d}t, \quad G_a(x) := \int_{-\infty}^{x} g(t,a)\mathrm{d}t.$$

Wir berechnen die Wölbung für die Dreicksversteilung G_a. Wegen

$$\int \left(\frac{x}{a^2}+\frac{1}{a}\right)x^4\mathrm{d}x = \frac{1}{30}x^5\left(\frac{5x+6a}{a^2}\right), \quad \int \left(\frac{-x}{a^2}+\frac{1}{a}\right)x^4\mathrm{d}x = \frac{-1}{30}x^5\left(\frac{5x-6a}{a^2}\right)$$

bekommen wir

$$\begin{aligned} \textstyle\int_{-\infty}^{+\infty} g(x,a)x^4\mathrm{d}x &= \textstyle\int_{-a}^{0} g(x,a)x^4\mathrm{d}x + \int_0^a g(x,a)x^4\mathrm{d}x \\ &= \textstyle\int_{-a}^{0}\left(\frac{x}{a^2}+\frac{1}{a}\right)x^4\mathrm{d}x + \int_0^a\left(\frac{-x}{a^2}+\frac{1}{a}\right)x^4\mathrm{d}x \\ &= \frac{1}{30}x^5\left(\frac{5x+6a}{a^2}\right)\Bigg|_{-a}^{0} + \frac{-1}{30}x^5\left(\frac{5x-6a}{a^2}\right)\Bigg|_{0}^{a} = \frac{a^4}{15}. \end{aligned}$$

Wegen

$$\int \left(\frac{x}{a^2}+\frac{1}{a}\right)x^2\mathrm{d}x = \frac{1}{12}x^3\left(\frac{3x+4a}{a^2}\right), \quad \int \left(\frac{-x}{a^2}+\frac{1}{a}\right)x^2\mathrm{d}x = \frac{-1}{12}x^3\left(\frac{3x-4a}{a^2}\right)$$

bekommen wir für das zweite Moment

$$\int_{-\infty}^{+\infty} g(x,a)x^2\mathrm{d}x = \frac{a^2}{6}$$

so daß

$$W(G_a) = \frac{a^4/15}{a^4/6^2} = \frac{36}{15} = \frac{12}{5}$$

d.h. die Wölbung von G_a ist unabhängig vom Parameter a und besitzt für alle a den Wert 2.4.

Für $W(F_a)$ bekommen wir ein ähnliches Ergebnis. Wegen

$$\int_{-\infty}^{+\infty} f(x,a)x^4 \mathrm{d}x = \int_{-a}^{+a} -\frac{3}{4a^3}(x^2-a^2)x^4 \mathrm{d}x = \frac{3}{35}a^4$$

und

$$\int_{-\infty}^{+\infty} f(x,a)x^2 \mathrm{d}x = \int_{-a}^{+a} -\frac{3}{4a^3}(x^2-a^2)x^2 \mathrm{d}x = \frac{1}{5}a^2$$

bekommen wir

$$W(F_a) = \frac{\int_{-\infty}^{+\infty} x^4 f(x,a)\mathrm{d}x}{\left(\int_{-\infty}^{+\infty} x^2 f(x,a)\mathrm{d}x\right)^2} = \frac{3}{35} \times \frac{25}{1} = \frac{15}{7}$$

d.h. die Wölbung von F_a ist unabhängig von a und beträgt 2.14.

Wegen

$$W(F_{0.6}) = \frac{15}{7} < \frac{12}{5} = W(G_{0.3})$$

erwarten wir, daß $G_{0.3}$ stärker gewölbt ist als $F_{0.6}$ oder das Maximum von $g(x, 0.3)$ größer ist als das Maximum von $f(x, 0.6)$. Es können auch beide Erwartungen erfüllt werden. Die Relation für die Maxima folgt sofort:

$$\max f(x,a)|_{a=0.6} = f(0,a) = \frac{3}{4a} = 1.25, \quad \max g(x,a)|_{a=0.3} = g(0,0.3) = \frac{10}{3}.$$

Der Blick auf Abbildung 35 bestätigt beide Erwartungen. Im Anwendungsfall kennt man f und g natürlich nicht, sodaß man das zweite und vierte Moment durch Stichproben mittels (61) schätzt. ◇

Beispiel 63 Wölbung der Normalverteilung

Da für das zweite und das vierte Moment der Normalverteilungsdichte $\varphi_{\mu,\sigma}$ bezüglich μ die Gleichungen gelten

$$\sigma^2 = m_2(\mu), \quad 3\sigma^4 = m_4(\mu)$$

gilt unabhängig von μ und σ

$$W(\Phi_{\mu,\sigma}) = 3$$

d.h. alle Normalverteilungen besitzen die gleiche Wölbung, nämlich 3. Die Gleichung

$$3\sigma^4 = m_4(\mu)$$

ergibt sich wie folgt. Entsprechend [20], S.185 oben, gilt für positives α immer

$$\int_0^{\infty} x^{2k} e^{-\alpha x^2} \mathrm{d}x = \frac{1 \times 3 \times \ldots \times (2k-1) \times \sqrt{\pi}}{2^{k+1}\alpha^{k+\frac{1}{2}}}.$$

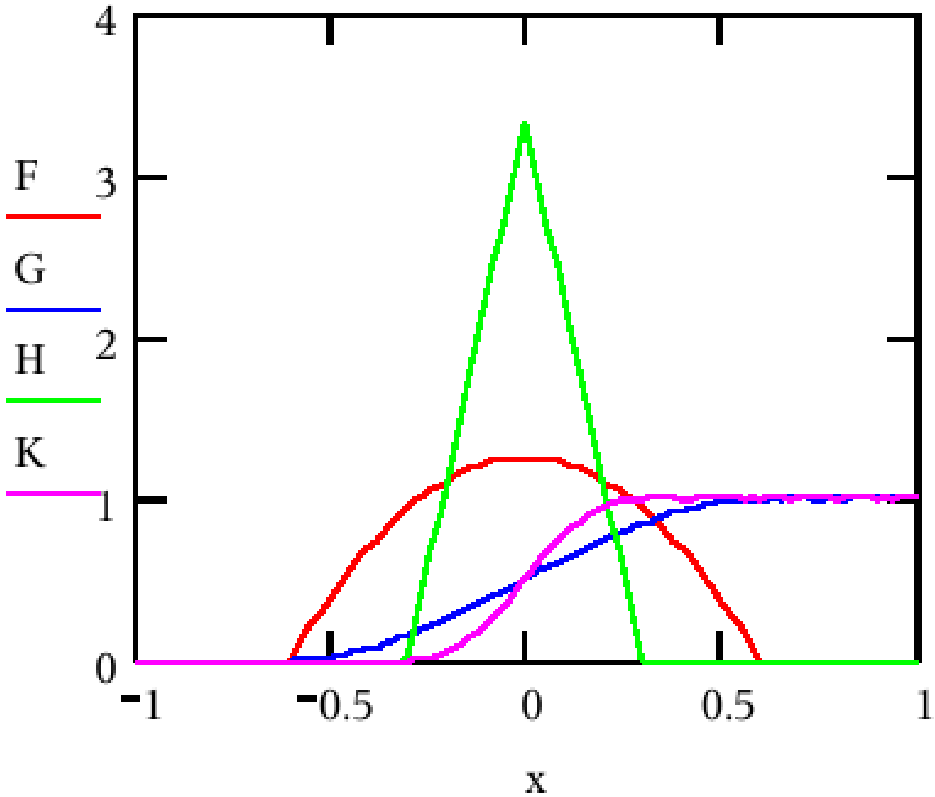


Abbildung 35: Die rote Kurve stellt $f(x, 0.6)$ und die blaue Kurve die Verteilungsfunktion $F_{0.6}$ dar. Die grüne Kurve gehört zur Dreiecksverteilung $g(x, 0.3)$ und die violette Kurve zur Verteilungsfunktion $G_{0.3}$. Offenbar ist der Graph von $G_{0.3}$ stärker gewölbt (S-förmiger) als der Graph von $F_{0.6}$.

Anwendung dieser Identität liefert

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}}\mathrm{d}x = 2\frac{1}{\sqrt{2\pi}}\int_0^{\infty} x^4 e^{-\frac{x^2}{2}}\mathrm{d}x = 2\frac{1}{\sqrt{2\pi}}\frac{1\times 3\times\sqrt{\pi}}{2^3\frac{1}{2^{2+\frac{1}{2}}}} = 3.$$

Schließlich bekommen wir mit der Variablentransformation $y = (x-\mu)/\sigma$ das gewünschte Resultat

$$m_4(\mu) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}(x-\mu)^4 e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma y)^4 e^{-\frac{y^2}{2}}\sigma\mathrm{d}y = 3\sigma^4.$$

Entgegen der Erwartung hängt die Wölbung von $\varphi_{\mu,\sigma}$ nicht von σ ab. Auch im Beispiel 62 kürzt sich der Parameter a heraus, sodaß man für alle a dieselbe Wölbung bekommt. Demnach kann man hinsichtlich der Wölbung nur Klassen von Verteilungsfunktionen miteinander vergleichen, z.B. die Klasse der Gleichverteilungen mit der Klasse der Normalverteilungen. $\diamond$

Die Wölbung wird auch **Kurtosis** genannt.

4.5 Konzentrationsmaß für diskrete Zg

[18] S.159

Es sei X eine diskrete Zg mit dem Wertevorrat $\{x_0, x_1, \ldots, x_n\}$ und den zugehörigen Wahrscheinlichkeiten $f(x_k)$, wobei f die Wahrscheinlichkeitsfunktion von X darstellt. Dann kann man die **Konzentration** von X durch

$$C(X) := \sqrt{\frac{(n+1)\sum_{k=0}^{n} f(x_k)^2 - 1}{n}}$$

erklären. Wegen

$$\sum_{k=0}^{n} f(x_k) = 1, \quad 0 \le f(x_k)$$

gilt

$$\frac{1}{n+1} \le \sum_{k=0}^{n} f(x_k)^2 \le 1 \quad ^7$$

sodaß sich C nur zwischen Null und Eins bewegt. Folgende beiden Verteilungen der statistischen Masse stellen Extremfälle dar:

[7] Diese Ungleichung sieht man ein, in dem man nach dem kürzesten Vektor fragt, der in der Ebene

$$E := \left\{\boldsymbol{x} \in \mathbb{R}^{n+1} : \sum_{k=0}^{n} x_k = 1\right\}$$

liegt. Das ist offenbar der Vektor $v := \frac{1}{n+1}[1, \ldots, 1]^{\mathrm{T}}$, für den $\|v\|^2 = \sum_{k=0}^{n+1}\frac{1}{(n+1)^2} = \frac{1}{n+1}$ gilt. Da jeder andere Vektor aus E länger ist, gilt die Ungleichung.

- Die Zg X ist gleichverteilt, d.h. $f(x_k) = \frac{1}{n+1}$ für $k = 0, \ldots, n$. Offenbar gilt dann $C = 0$.
- Es gilt $f(x_k) = 1$ und allen anderen Werten wird die Wahrscheinlichkeit Null zugeordnet. Offenbar gilt dann $C = 1$.

Demnach ist $C(X)$ geeignet, die Verteilung der statistischen Masse bei X zu messen: Falls einem Wert x_k die Wahrscheinlichkeit Eins zugeteilt wird, konzentriert sich die Wahrscheinlichkeitsmasse an der Stelle x_k und es gilt $C(X) = 1$. Falls X alle Werte mit gleicher Wahrscheinlichkeit annimmt, ist die Wahrscheinlichkeitsmasse nirgends besonders konzentriert und man bekommt $C(X) = 0$.

Beispiel 64 Konzentration binomialverteilter Zg in Abhängigkeit von p

Wir visualisieren $C(X)$ für $X \sim \mathscr{B}(n,p)$ in Abhängigkeit von $p \in [0,1]$ für verschiedene n. Offenbar gilt

$$C(X) = c(n,p) := \sqrt{\frac{(n+1)\sum_{k=0}^{n}\binom{n}{k}^2 p^{2k}(1-p)^{2(n-k)} - 1}{n}}.$$

Abbildung 36 zeigt, daß für $p = 1/2$ für jedes n die Konzentration minimal wird. Wegen

$$\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}$$

bekommen wir

$$c(n,1/2) = \sqrt{\frac{(n+1)\sum_{k=0}^{n}\binom{n}{k}^2 \frac{1}{2^{2n}} - 1}{n}} = \sqrt{\frac{n+1}{n4^n}\binom{2n}{n} - \frac{1}{n}}.$$

Entsprechend Abbildung 36 nimmt für $n \leq 20$ die Konzentration $c(n,1/2)$ ungefähr den Wert 0.3 an. $\diamond$

4.6 Lorenzkurve und Gini-Koeffizient

[5] S.78, [17] S.86

Es liegt uns ein numerischer Vektor

$$\boldsymbol{x} := [x_1 \ldots x_n]^{\mathrm{T}} \in \mathbb{R}^n \setminus \{o\}, 0 \leq x_k$$

mit nichtnegativen Komponenten vor, wobei wenigstens eine Komponente positiv ist. Es soll die zugehörige Lorenzkurve konstruiert werden. Zu diesem Zweck nehmen aufsteigende Sortierung wir

$$x_1 \leq x_2 \leq \ldots \leq x_n$$

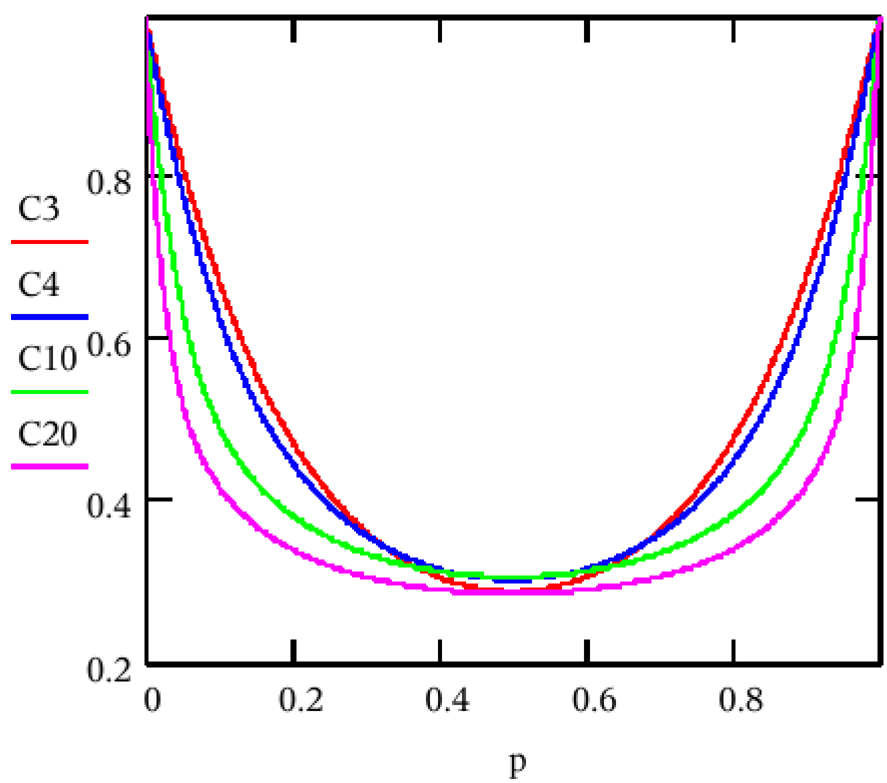


Abbildung 36: Es wird die Konzentration $c(n,p)$ von $X \sim \mathscr{B}(n.p)$ für $n = 3, 4, 10, 20$ und $p \in [0,1]$ gezeigt. Offenbar nimmt für jedes n die Konzentration in $p = 1/2$ ihr Minimum an, welches allerdings verschieden von Null ist. Außerdem ist zu erkennen, daß sich die Konzentrationskurven schneiden, d.h. es existieren Grundwahrscheinlichkeiten p, die für verschiedene n die gleiche Konzentration liefern.

an und bilden die Zahlen

$$u_k := \frac{k}{n}, \quad v_k := \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i}. \tag{63}$$

Der Polygonzug, der die Punkte

$$(u_0, v_0) := (0, 0), (u_1, v_1), (u_2, v_2), \ \ldots, \ (u_{n-1}, v_{n-1}), (u_n, v_n) = (1, 1)$$

der Reihe nach miteinander verbindet, heißt Lorenzkurve von $\boldsymbol{x}$, benannt nach Max Otto Lorenz (1905). Offenbar gibt es zwei Extremfälle:

- Die Stichprobenwerte sind konstant, d.h. $x_1 = \ldots = x_n$. Dann bekommt man

  $$v_k = u_k$$

  und die Lorenzkurve stimmt mit der Geraden g, die den Koordinatenursprung mit dem Punkt $(1, 1)$ verbindet, überein.

- Die ersten $n - 1$ Stichprobenwerte verschwinden, d.h. $x_1 = \ldots = x_{n-1} = 0$ und der n-te Wert nicht, d.h. $x_n \neq 0$. Dann bekommt man für $k = 0, \ldots, n - 1$

  $$v_k = 0$$

  sodaß die Lorenzkurve aus zwei Geraden besteht, wobei die erste auf der x-Achse liegend, den Koordinatenursprung mit $(0, 1 - \frac{1}{n})$ und die zweite $(0, 1 - \frac{1}{n})$ mit $(1, 1)$ verbindet.

Die Lorenzkurve ist monoton wachsend und konvex. Dabei heißt eine Abbildung $f : [a, b] \mapsto \mathbb{R}$ **konvex**, falls für alle $x, y \in [a, b]$ und alle $t \in [0, 1]$ die Ungleichung

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$$

erfüllt ist.

Der Gini-Koeffizient des Vektors $\boldsymbol{x}$, eingeführt von dem italienischen Statistiker Corrado Gini (1884-1965), ist durch

$$G(\boldsymbol{x}) := \text{das Doppelte der Fläche zwischen } g \text{ und Lorenzkurve von } \boldsymbol{x}$$

definiert. Der erste Extremfall liefert offenbar $G(\boldsymbol{x}) = 0$, da die Lorenzkurve von $\boldsymbol{x}$ in diesem Fall mit g übereinstimmt. Der zweite Extremfall liefert

$$G(\boldsymbol{x}) = 2\left(\frac{1}{2} - \frac{1}{2}\frac{1}{n}\right) = 1 - \frac{1}{n}$$

d.h. für große n ist $G(\boldsymbol{x})$ ungefähr gleich 1. Anschaulich ist klar, daß stets

$$0 \leq G(\boldsymbol{x}) < 1$$

gelten muß.

Statistische Interpretation: Wir betrachten n Personen, von denen uns deren Vermögensverhältnisse bekannt sind:

$$x_1 \leq x_2 \leq \cdots \leq x_n.$$

Die zwischen 0 und 100 liegende Zahl $100v_k$ bedeutet dann, wieviel Prozent des Gesamtvermögens

$$x_1 + \ldots + x_n$$

sich in Händen von $\frac{100k}{n}$% Personen befindet. Der Extremfall $x_1 = \ldots = x_n \neq 0$ bedeutet die Übereinstimmung beider Prozentsätze, sodaß sich nirgends besonders viel Vermögen konzentriert, sodaß man von keiner besonderen Vermögenskonzentration sprechen kann. Der entsprechende Gini-Koeffizient spiegelt dieses Verhältnis durch den Wert Null wider. Der andere Extremfall bedeutet, daß die Personen 1 bis $n-1$ ohne Vermögen sind und sich das Gesamtvermögen der Personengruppe in den Händen von Person n befindet, d.h. das Gesamtvermögen konzentriert sich auf eine Person. Der Gini-Koeffizient spiegelt dieses Verhältnis durch den Wert $1 - \frac{1}{n}$, d.h. bei einigermaßen großem n liegt der Wert fast bei Eins, d.h. eine starke Konzentration des Vermögensanteile ist äquivalent zu einem Gini-Keoffizienten in der Nähe von Eins. Alle anderen nicht extremen Fälle wird durch $G(\boldsymbol{x})$ ein zwischen 0 und 1 liegendes Konzentrationsmaß zugeordnet.

Unmittelbar aus der Definition von $G(\boldsymbol{x})$ und der Definition (63) der Zahlen v_k kommt man nach einigen Umformungen auf die Gleichung

$$G(\boldsymbol{x}) = 1 - \frac{1}{n}\left(1 + 2\sum_{k=1}^{n-1} v_k\right).$$

Unter Beachtung von $v_0 = 0$ und $v_n = 1$ findet man

$$G(\boldsymbol{x}) = 1 - \frac{1}{n}\left(v_0 + v_1 + v_1 + v_2 + v_2 + \ldots + v_{n-1} + v_1\right) = 1 - \frac{1}{n}\sum_{k=1}^{n}(v_{k-1} + v_k)$$

in Übereinstimmung mit [17] S.88.

Beispiel 65 Abbildung 37 zeigt für

$$\boldsymbol{x} = [0.3, 1.5, 2.5, 5, 7]^{\mathrm{T}}$$

die Lorenzkurve. Insbesondere gilt $G(\boldsymbol{x}) = 0.415$. ◇

4.7 Kontingenztafeln, Pearsons χ^2-Statistik, χ^2- Unabhängigkeitstest, Zufallsvektoren

Aus einer Grundgesamtheit werden n Individuen zufällig ausgewählt und in Hinsicht auf den Besitz der Eigenschaften $E_1, \ldots, E_m$ untersucht. Das Untersuchungsergebnis kann man in Form einer $0,1$-Kette der Länge m kodieren, wobei an der Stelle $i \in \{1, \ldots, m\}$ eine Eins vergeben wird, falls das gezogene Individuum die Eigenschaft E_i besitzt, und falls nicht, vergibt man eine Null. Auf diese Weise erhält man n nicht notwendig verschiedene $0,1$-Ketten der Länge m.

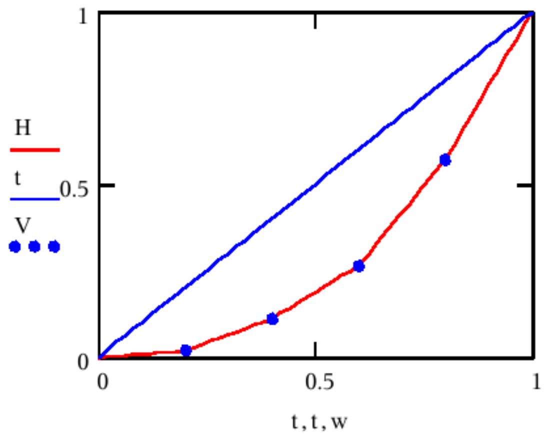


Abbildung 37: Es wird die Lorenzkurve (rot) zum Beispiel 65 gezeigt. Die blauen Punkte korrespondieren mit (u_k, v_k) für $k = 1, 2, 3, 4$. Der mit Zwei multiplizierte Inhalt der Fläche zwischen der blauen Gerade und der Lorenzkurve stimmt mit $G(\boldsymbol{x}) = 0.415$ überein.

Beispiel 66 Man wählt aus einer Menschengruppe einzelne Personen aus und befragt sie nach Alter und monatlichem Einkommen. Als Eigenschaften führen wir in willkürlicher Weise die Altersklassen

$$E_1 := [1, 30),\ E_2 := [31, 60),\ E_3 := [61, 90)$$

und die Einkommensklassen

$$E_4 := [0, 1000),\ E_5 := [1000, +\infty)$$

ein. Das Befragungsresultat ergibt dann eine Binärzahl der Länge 5 wie z.B.

$$00101$$

d.h. die Person ist zwischen 61 und 90 Jahren alt und hat ein monatliches Einkommen von wenigstens 1000 Euro. $\diamond$

Zur Kontingenztafel gelangt man, in dem man zählt, wieviele Individuen eine Kombination zweier bestimmter Eigenschaften besitzen:

$$\begin{array}{c|cccc}
 & E_1 & E_2 & \ldots & E_m \\
\hline
E_1 & n_{11} & n_{12} & \ldots & n_{1m} \\
E_2 & n_{21} & n_{22} & \ldots & n_{2m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
E_m & n_{m1} & n_{m2} & \ldots & n_{mm}
\end{array}$$

Dabei ergibt sich n_{ij} aus der Anzahl aller $0,1$-Ketten, die sowohl in Position i als auch in Position j eine Eins besitzen. Offenbar gilt $n_{ij} = n_{ji}$. Im Allgemeinen ist es nun nicht sinnvoll, alle Eigenschaften miteinander zu kombinieren, wie das Beispiel der Alters- und Einkommensklassen zeigt: jede Person liegt genau in einer Alters- und in einer Einkommensklasse, sodaß die Kombination zweier Altersklassen oder die Kombination zweier Einkommensklassen zu den Einträgen $n_{ij} = 0$ führt. Für das obige Beispiel wäre

	E_1	E_2	E_3
E_4	n_{41}	n_{42}	n_{43}
E_5	n_{51}	n_{52}	n_{53}

sinnvoll. In diesem Zusammenhang ist die Bezeichnung der Altersklassen mit A_1, A_2 und A_3 und die Bezeichnung der Einkommensklassen mit E_1 und E_2 naheliegend, sodaß man zu

	A_1	A_2	A_3
E_1	n_{11}	n_{12}	n_{13}
E_2	n_{21}	n_{22}	n_{23}

gelangt und die natürliche Zahl n_{ij} gibt die Anzahl der Personen an, die sowohl zur Altersklasse j als auch zur Einkommensklasse i gehören.

Wir behalten die Einteilung

$$A_1, \ldots, A_{m_1}, \quad E_1, \ldots, E_{m_2}, \quad m_1 + m_2 = m$$

bei und setzen voraus, daß jedes der betrachteten n Individuen in genau einer der Klassen $A_1, \ldots, A_{m_1}$ und in genau einer der Klassen $E_1, \ldots, E_{m_2}$ liegt, d.h. falls

- n_{i+} die Anzahl aller ausgewählten Individuen mit der Eigenschaft E_i und
- n_{+j} die Anzahl aller ausgewählten Individuen mit der Eigenschaft A_j

bezeichnet, so gelten die Gleichungen

$$\sum_{j=1}^{m_1} n_{+j} = \sum_{i=1}^{m_2} n_{i+} = \sum_{i=1}^{m_2} \sum_{j=1}^{m_1} n_{ij} = n.$$

Darüberhinaus erfüllen die Zeilen- und die Spaltensummen der Zahlen n_{ij} aus

	A_1	A_2	$\ldots$	A_{m_1}
E_1	n_{11}	n_{12}	$\ldots$	n_{1m_1}
E_2	n_{21}	n_{22}	$\ldots$	n_{2m_1}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
E_{m_2}	n_{m_21}	n_{m_22}	$\ldots$	$n_{m_2m_1}$

die Gleichungen

$$n_{i+} := \sum_{j=1}^{m_1} n_{ij} \quad \text{(Zeilensummen)}, \quad n_{+j} := \sum_{i=1}^{m_2} n_{ij} \quad \text{(Spaltensummen)}.$$

Setzt man

$$f_{ij} := \frac{n_{ij}}{n}$$

so erhält man die relative Häufigkeit, mit der ein Individuum sowohl die Eigenschaft E_i als auch die Eigenschaft A_j besitzt.

Die Zahlen f_{ij}, n_{i+}, n_{+j} können zur Entscheidung benutzt werden, ob zwei Zg X und Y unabhängig voneinander sind oder nicht. Zu diesem Zweck erinnern wir uns, daß entsprechend Abschnitt 2.6.2 zwei Zg X und Y unabhängig genannt werden, falls für alle $x, y \in \mathbb{R}$ die Gleichung

$$P(\{X \leq x\} \cap \{Y \leq y\}) = P(\{X \leq x\})P(\{Y \leq y\})$$

erfüllt wird. Zur Vereinfachung setzen wir

$$F(x,y) := P(\{X \leq x\} \cap \{Y \leq y\})$$

und nehmen für F die Existenz einer Dichtefunktion

$$\varphi(x,y)$$

an, sodaß

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} \varphi(s,t)\mathrm{d}s\mathrm{d}t. \tag{64}$$

Dann gilt folgende Äquivalenz: Unter der Voraussetzung (64) sind die Zg X und Y unabhängig genau dann, wenn $\varphi(x,y)$ eine Faktorisierung der Gestalt

$$\boxed{\varphi(x,y) = \varphi_X(x)\varphi_Y(y)}$$

gestattet, wobei φ_X und φ_Y die Dichtefunktionen der Verteilungsfunktionen von X und Y sind [20], S.1047.

Die einfache Richtung des Beweises geht von der Existenz einer solchen Faktorisierung aus:

$$\begin{aligned} P(\{X \leq x\} \cap \{Y \leq y\}) = & \int_{-\infty}^{y} \int_{-\infty}^{x} \varphi(s,t)\mathrm{d}s\mathrm{d}t = \int_{-\infty}^{y} \int_{-\infty}^{x} \varphi_X(s)\varphi_Y(t)\mathrm{d}s\mathrm{d}t \\ = & \int_{-\infty}^{x} \varphi_X(s)\mathrm{d}s \int_{-\infty}^{y} \varphi_Y(t)\mathrm{d}t = P(\{X \leq x\})P(\{Y \leq y\}) \end{aligned}$$

d.h. X und Y sind unabhängig. In der gleichen Weise bekommt man für alle $a, b, c, d \in \mathbb{R}$ mit $a \leq b$ und $c \leq d$ die Gleichungen

$$\begin{aligned} P(\{a \leq X \leq b\} \cap \{c \leq Y \leq d\}) = & \int_{c}^{d} \int_{a}^{b} \varphi(s,t)\mathrm{d}s\mathrm{d}t = \int_{c}^{d} \int_{a}^{b} \varphi_X(s)\varphi_Y(t)\mathrm{d}s\mathrm{d}t \\ = & \int_{a}^{b} \varphi_X(s)\mathrm{d}s \int_{c}^{d} \varphi_Y(t)\mathrm{d}t \\ = & P(\{a \leq X \leq b\})P(\{c \leq Y \leq d\}). \end{aligned}$$

Nun gehen wir von den Intervalleinteilungen

$$a_j := a + \frac{(b-a)j}{m_1},\ j = 0, \ldots, m_1, \quad c_i := c + \frac{(c-d)i}{m_2},\ i = 0, \ldots, m_2$$

aus und sagen, daß ein Individuum ω die Eigenschaft A_j besitzt, falls

$$X(\omega) \in [a_{j-1}, a_j)$$

und daß es die Eigenschaft E_i besitzt falls

$$Y(\omega) \in [c_{i-1}, c_i).$$

Entsprechend oben muß für voneinander unabhängige X und Y für alle $i \in \{1, \ldots, m_2\}$ und alle $j \in \{1, \ldots, m_1\}$ die Gleichung

$$P(\{a_{j-1} \leq X \leq a_j\} \cap \{c_{i-1} \leq Y \leq c_i\}) = P(\{a_{j-1} \leq X \leq a_j\})P(\{c_{i-1} \leq Y \leq c_i\}) \quad (65)$$

gelten. Offenbar stellen die Zahlen

$$f_{ij}, \quad \frac{n_{+j}}{n}, \quad \frac{n_{i+}}{n}$$

Schätzungen für die Wahrscheinlichkeiten

$$P(\{a_{j-1} \leq X \leq a_j\} \cap \{c_{i-1} \leq Y \leq c_i\}), \quad P(\{a_{j-1} \leq X \leq a_j\}), \quad P(\{c_{i-1} \leq Y \leq c_i\})$$

dar, sodaß sich (65) in

$$f_{ij} = \frac{n_{ij}}{n} \approx \frac{n_{+j}}{n}\frac{n_{i+}}{n}$$

niederschlägt, d.h. falls die entsprechend [17], S. 109, eingeführte Summe

$$\chi^2 := \sum_{i=1}^{m_2}\sum_{j=1}^{m_1} \frac{\left(n_{ij} - \frac{n_{+j}n_{i+}}{n}\right)^2}{\frac{n_{+j}n_{i+}}{n}} \quad (66)$$

'klein' ausfällt, kann man das als einen Hinweis auf Unabhängigkeit der Zg X und Y auffassen. Es ist zu beachten, daß diese Summe nur für Kontingenztafeln definiert ist, die keine vollständige Null-Zeile oder -Spalte besitzen, da ja in diesem Falle wenigstens eine der Zahlen n_{i+} oder n_{+j} zu Null wird. Man kann dann mit der Regel von L'Hospital den Grenzwert berechnen. Summen der Form χ^2 gehören zum Themenkreis der χ^2-Statistik im Sinne von Pearson. Um abzuschätzen, was 'klein' bedeutet, kann man die Ungleichung

$$0 \leq \chi^2 \leq n\,(\min\{m_1, m_2\} - 1) \quad (67)$$

benutzen, die wir an Hand des (2×2)-Falls plausibel machen werden. In [17], S. 109, wird diese Abschätzung im Text angedeutet. Die Kontingenztafel wird im (2×2)-Fall als Vier-Felder-Tafel bezeichnet.

4.7.1 Vier-Felder-Tafeln

Eine Vier-Felder-Tafel ist von der Gestalt

	A_1	A_2	
E_1	a	b	$n_{+1} := a + b$
E_2	c	d	$n_{+2} := c + d$
	$n_{1+} := a + c$	$n_{2+} := b + d$	$n := a + b + c + d$

(68)

wobei natürlich $0 < n$ vorausgesetzt wird. Für den zugehörigen χ^2-Wert folgt aus (66) die Gleichung

$$\chi^2 = \frac{\left(a - \frac{(a+b)(a+c)}{n}\right)^2}{\frac{(a+b)(a+c)}{n}} + \frac{\left(b - \frac{(a+b)(b+d)}{n}\right)^2}{\frac{(a+b)(b+d)}{n}} + \frac{\left(c - \frac{(c+d)(a+c)}{n}\right)^2}{\frac{(c+d)(a+c)}{n}} + \frac{\left(d - \frac{(c+d)(b+d)}{n}\right)^2}{\frac{(c+d)(b+d)}{n}}.$$

Mit $n = a + b + c + d$ prüft man leicht die Gültigkeit von

$$|ad - bc| = |a - n(a+b)(a+c)| = |b - n(a+b)(b+d)| = |c - n(c+d)(a+c)| = |d - n(c+d)(b+d)|$$

nach. Daher liefert die Erweiterung der Summanden von χ^2 mit n^2 die Gleichung

$$\begin{aligned}\chi^2 &= \frac{(ad-bc)^2}{n}\left(\frac{1}{(a+b)(a+c)} + \frac{1}{(a+b)(b+d)} + \frac{1}{(c+d)(a+c)} + \frac{1}{(c+d)(b+d)}\right) \\ &= \frac{(ad-bc)^2}{n}\left(\frac{(b+d)(c+d) + (a+c)(c+d) + (a+b)(b+d) + (a+b)(a+c)}{(a+b)(a+c)(b+d)(c+d)}\right)\end{aligned}$$

Schließlich gilt

$$(b+d)(c+d) + (a+c)(c+d) + (a+b)(b+d) + (a+b)(a+c) = n^2$$

sodaß wir

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

schreiben können. Ausmultiplizieren des Nenners liefert eine Summe, die die Summanden $(ad)^2$ und $(bc)^2$ enthält. Daher besteht die Differenz

$$(a+b)(a+c)(b+d)(c+d) - (ad-bc)^2 \tag{69}$$

aus Summanden der Form

$$a^{i_1} b^{i_2} c^{i_3} d^{i_4}, \quad i_k \in \{0, 1, 2\}.$$

Da es sich bei a, b, c, d um Häufigkeiten handelt, haben wir es mit nichtnegativen Zahlen zu tun, sodaß die Differenz (69) nicht negativ wird. Demnach gilt die Ungleichung

$$0 \leq (ad-bc)^2 \leq (a+b)(a+c)(b+d)(c+d) \tag{70}$$

d.h.

$$0 \le \chi^2 \le n.$$

Falls $0 < a, b, c, d$, gilt

$$(ad - bc)^2 < (a+b)(a+c)(b+d)(c+d)$$

sodaß in diesem Fall χ^2 kleiner wird als n. D.h. der χ^2-Wert einer Vier-Felder-Tafel, die keine Nullen aufweist, ist immer kleiner als $a + b + c + d$.

Wir untersuchen jetzt die Fälle, in denen χ^2 den Wert n annimmt. Nachdem was wir gerade festgestellt haben, muß dann wenigsten einer ihrer Einträge verschwinden.

- Wir betrachten den Fall, daß zwei Einträge verschwinden. Für

$$b = c = 0 \text{ oder } a = d = 0 \tag{71}$$

tritt in (70) Gleichheit ein, was sich wie folgt interpretieren läßt. Ein Blick auf die Tafel (68) zeigt, daß im ersten Fall, die Ereignisse A_1 und E_1 und im zweiten Fall die Ereignisse A_2 und E_2 nur simultan auftreten. Solche Ereignisse werden als im **exakten Zusammenhang** stehend referiert. Daher deutet die Gleichung

$$\chi^2 = n$$

auf einen exakten Zusammenhang hin. Ihre Gültigkeit impliziert aber nicht nur (71). Um alle Konsequenzen zu bekommen, beobachten wir

$$\chi^2 = n \Leftrightarrow \frac{(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} = 1.$$

Führt man das multivariate Polynom

$$p(a, b, c, d) := (a+b)(a+c)(b+d)(c+d) - (ad-bc)^2$$

ein, so liefern seine Nullstellen, die im nichtnegativen Quadranten von $\mathbb{R}^4$ liegen, Kandidaten für die Stellen, an denen $\chi^2 = a+b+c+d$ gilt. Offenbar gilt nicht nur $p(a, 0, 0, d) = p(0, b, c, 0) = 0$, sondern auch

$$p(a, b, 0, 0) = p(a, 0, c, 0) = p(0, b, 0, d) = p(0, 0, c, d) = 0. \tag{72}$$

Die erste Gleichung korrespondiert mit der Kontigenztafel

	A_1	A_2	
E_1	a	b	$n_{+1} := a+b$
E_2	0	0	$n_{+2} := 0$
	$n_{1+} := a$	$n_{2+} := b$	$n := a+b$

die sich dahingehend interpretieren läßt, daß mit Sicherheit das Ereignis E_1 eintritt und das Ereignis E_2 nicht, d.h. E_1 und E_2 sind unabhängig vom Eintreten von A_1 und A_2. Offenbar gelten für $b \neq 0$ die Grenzwertgleichungen

$$\begin{aligned}
&\lim_{d\to 0} \lim_{c\to 0} \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} = \\
&\lim_{d\to 0} \frac{(a+b+d)(ad)^2}{(a+b)(a)(b+d)(d)} = \\
&\lim_{d\to 0} \frac{(a+b+d)ad}{(a+b)(b+d)} = 0
\end{aligned} \tag{73}$$

d.h. diese Unabhängkeitssituation wird wie gewünscht mit Null bewertet. Die anderen Fälle aus (72) führen zu demselben Resultat.

- Wir betrachen jetzt den Fall, daß genau drei Einträge in der Vier-Felder-Tafel verschwinden. Zu diesem Zweck setzen wir in (73) den Wert von b auf Null. Dann geht der Grenzwert über

$$\lim_{d\to 0} \lim_{c\to 0} \frac{(a+d)(ad)^2}{(a)(a+c)(d)(c+d)} = \lim_{d\to 0} \frac{(a+d)(ad)^2}{(a)(a)(d)(d)} = \lim_{d\to 0}(a+d) = a$$

d.h. die Gleichung $p(a,0,0,0) = 0$ korrespondiert mit der Kontingenztafel

	A_1	A_2	
E_1	a	0	$n_{+1} := a$
E_2	0	0	$n_{+2} := 0$
	$n_{1+} := a$	$n_{2+} := 0$	$n := a$

und liefert für χ^2 den Wert a. Offenbar hängen E_1 und A_1 exakt zusammen, d.h. das eine Ereignis zieht das andere nach sich und es gilt, wie gewünscht, $\chi^2 = n$.

- Schließlich betrachten wir den Fall, bei dem genau ein Eintrag in (68) verschwindet, und überprüfen, ob χ^2 den Wert n annehmen kann. Zu diesem Zweck sei $d = 0 < a, b, c$ und $\chi^2 = a + b + c$. Dann muß

$$\frac{b^2c^2}{(a+b)(a+c)bc} = 1$$

gelten. Daraus folgt unmittelbar

$$bc = (a+b)(a+c) = a^2 + (b+c)a + bc \Rightarrow 0 = a^2 + (b+c)a.$$

Da $a \neq 0$ gelten soll, bekommen wir $0 = a + b + c$, was notwendigerweise auf negatives b oder negatives c führt. Da a, b, c, aber nach Voraussetzung positiv sind, ist das ein Widerspruch und unsere Annahme $\chi^2 = n$ ist unhaltbar. Offenbar haben wir es mit der Kontingenztafel

	A_1	A_2	
E_1	a	b	$n_{+1} := a + b$
E_2	c	0	$n_{+2} := c$
	$n_{1+} := a + c$	$n_{2+} := b$	$n := a + b + c$

zu tun.

Im allgemeinen $(m_1 \times m_2)$-Fall geht die Abschätzung in

$$0 \leq \chi^2 \leq n\left(\min\{m_1, m_2\} - 1\right).$$

über, wodurch die Einführung normierter Kontingenzkoeffizienten möglich wird, die mit Φ und C bezeichnet und jeweils der **Phikoeffizient** und der **Kontingenzkoeffizient** genannt werden.

Die vektorwertige Abbildung

$$\begin{bmatrix} X \\ Y \end{bmatrix} : \Omega \mapsto \mathbb{R}^2$$

heißt Zufallsvektor. Beachte, daß wir in vorhergehenden Beispielen Vektoren der Form

$$[X(\omega_i)]_{i=1}^n$$

auch als Zufallsvektoren bezeichnet hatten. Jetzt haben wir es aber mit Vektoren der Form

$$[X_i(\omega)]_{i=1}^n$$

zu tun, wobei ω ein fixiertes Individuum aus dem Ereignisraum Ω ist und $X_i : \Omega \mapsto \mathbb{R}$, $i = 1, \ldots, n$, Zg über Ω sind.

4.7.2 Zufallsvektoren

Falls $[X, Y]^{\mathrm{T}}$ ein stetiger Zufallsvektor ist, gilt

$$P(X \leq x) = P(\{X \leq x\} \cap \{Y \leq +\infty\}) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} \varphi(s,t) \mathrm{d}t \mathrm{d}s.$$

Setzt man daher

$$\varphi_X(s) := \int_{-\infty}^{+\infty} \varphi(s,t) \mathrm{d}t$$

kann man

$$P(X \leq x) = \int_{-\infty}^{x} \varphi_X(s) \mathrm{d}s$$

schreiben, d.h. φ_X stellt die Dichtefunktion der Zg X dar. Offenbar gilt

$$EX = \int_{-\infty}^{+\infty} x\varphi_X(x) \mathrm{d}x = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x\varphi(x,y) \mathrm{d}y \mathrm{d}x.$$

Mit der Definition

$$\varphi_Y(t) := \int_{-\infty}^{+\infty} \varphi(s,t) \mathrm{d}s$$

bekommt man die Dichtefunktion der Zg Y. Die Funktionen φ_X und φ_Y werden auch als **Randdichten** bezeichnet. Wie oben festgestellt wurde, sind X und Y unabhängig voneinander genau dann, wenn $\varphi(x,y) = \varphi_X(x)\varphi_Y(y)$.

4.8 Zeitreihen

[17] S.205, [18] S.243

Gegeben ist ein Beobachtungsvektor $\boldsymbol{X} := [x_0, x_1, \ldots, x_N]^{\mathrm{T}}$ dessen Komponenten sich zu den Zeitpunkten

$$\boldsymbol{t} := [t_0, t_1, \ldots, t_N]^{\mathrm{T}}, \quad t_k := k\Delta, \quad \Delta \in \mathbb{R}_+$$

ergeben haben. Einen solchen Vektor nennt man eine **Zeitreihe**. Gesucht wird eine Beziehung zwischen dem Zeitpunktvektor $\boldsymbol{t}$ und dem Beobachtungsvektor $\boldsymbol{X}$, die eine Vorhersage der zu erwartenden Beobachtungen $x_{N+1}, x_{N+2}, \ldots$ zu den Zeitpunkten $t_{N+1}, t_{N+2}, \ldots$ gestattet. Einen wichtigen Zugang bietet Abschnitt 3.5.4, in dem der Korrelationskoeffizient eingeführt wurde. Aus $\boldsymbol{X}$ bilden wir die beiden zeitlich zueinander verschobenen Teilreihen

$$\boldsymbol{x} := [x_0, \ldots, x_{N-k}]^{\mathrm{T}}, \quad \boldsymbol{y} := [x_k, \ldots, x_N]^{\mathrm{T}}$$

und aus ihnen den (empirischen) Korrelationskoeffizienten von $\boldsymbol{x}$ und $\boldsymbol{y}$:

$$\boxed{r_k := \frac{s_{\boldsymbol{xy}}}{s_{\boldsymbol{x}} s_{\boldsymbol{y}}} = \frac{\displaystyle\sum_{i=0}^{N-k} (x_i - \overline{\boldsymbol{x}})(y_j - \overline{\boldsymbol{y}})}{\sqrt{\displaystyle\sum_{i=0}^{N-k} (x_i - \overline{\boldsymbol{x}})^2} \sqrt{\displaystyle\sum_{i=0}^{N-k} (y_i - \overline{\boldsymbol{y}})^2}} \qquad \begin{array}{l} \overline{\boldsymbol{x}} := \frac{1}{N-k+1} \sum_{i=0}^{N-k} x_i \\ \overline{\boldsymbol{y}} := \frac{1}{N-k+1} \sum_{i=0}^{N-k} y_i \end{array}}$$

Die Zahl r_k heißt k-ter **Autokorrelationskoeffizient** von $\boldsymbol{X}$. Offenbar gilt $r_0 = 1$ und entsprechend Ungleichung (53)

$$|r_k| \le 1.$$

Eigenschaften: ([20], S.1079)

- Sind die Werte $|r_k|$ für $k = m, 2m, 3m \ldots$ fast 1, dann ist das ein Hinweis auf das Vorhandensein der Periode $T := 2m\Delta$ in $x_0, \ldots, x_N$.
- Sind alle $|r_k|$ $(k = 1, \ldots, N)$ klein, ist das ein Hinweis auf ein aperiodisches Verhalten der Zeitreihe.

Beispiel 67 Periodische Zeitreihe

Wir erzeugen eine Zeitreihe durch die Funktionswerte von

$$x_k := f(k) := \cos(2\pi\omega k),\ k := 0, \ldots, N,\ \omega := 1/16.$$

Ziel ist es, ω aus der Zeitreihe $x_0, \ldots, x_N$ zu rekonstruieren. Offenbar gilt die Periodizität

$$x_k = x_{k+16}$$

und $\Delta = 1$. Der Autokorrelationskoeffizient nimmt den Verlauf

$$r_0 = 1,\ r_1 = 0.924,\ r_2 = 0.707,\ r_3 = 0.383,\ r_4 = 0,\ r_5 = -0.383,$$
$$r_6 = -0.707,\ r_7 = -0.924,\ r_8 = -1, r_9 = -r_1, r_{10} = -r_2, \ldots$$

Es gilt

$$1 = |r_8| = |r_{16}| = |r_{24}| \ldots$$

Da alle natürlichen Vielfache von 8 als Index in dieser Gleichungsfolge erscheinen, ist das entsprechend der oberen Feststellung ein Hinweis darauf, daß

$$2 \times 8 \times \Delta$$

eine Periode der untersuchten Zeitreihe darstellt, was mit $\omega = 1/16$ übereinstimmt. $\diamond$

Gleitender Durchschnitt der Ordnung m:

$$\hat{x}_k := \frac{1}{m} \sum_{i=0}^{m-1} x_{k+i}, \quad k := 0, \ldots, N - (m-1).$$

Durch Ersetzung der Zeitreihe $x_0, \ldots, x_N$ durch $\hat{x}_0, \ldots, \hat{x}_{N-(m-1)}$ erhält man eine um $(m-1)$-Glieder kürzere Reihe, die aber glatter als die Originalreihe ist, und Aussagen über diese gestattet.

Beispiel 68 Gleitender Mittelwert, Vorhersagerekursion

Wir erzeugen durch

$$x_k := f(t_k) + e_k, \quad t_k := \frac{2\pi k}{N},\ N := 100,\ k := 0, \ldots, N, \quad f(t) := 20\cos(3t) + \sin(2t)$$

eine Zeitreihe $\boldsymbol{X}$, wobei die zufällige Störung e_k mit den Parametern $\mu := 0$ und $\sigma := 2$ normalverteilt ist. Die zugehörige Punktwolke wird in Abbildung 38 durch die roten Punkte visualisiert. Da $f(t)$ ein trigonometrisches Polynom ist, verwundert das Oszillationsverhalten nicht, wobei der Cosinusanteil mit dem Gewicht 20 dominiert. Mittels der Rekursion

$$z_{k+4} := w_0 z_k + w_1 z_{k+1} + w_2 z_{k+2} + w_3 z_{k+3}, \quad z_i := \hat{x}_i,\ i = 0, 1, 2, 3 \tag{74}$$

soll die Zeitreihe fortgeschrieben werden. Zur Berechnung geeigneter Gewichte w_i glätten wir $\boldsymbol{X}$ durch einen gleitenden Durchschnitt von der Ordnung 10, was die Zahlenfolge $(\hat{x}_k)_{k=0}^{91}$ liefert. Den Gewichtsvektor w bestimmen wir so, daß die Differenz

$$\begin{bmatrix} \hat{x}_4 - z_4 \\ \vdots \\ \hat{x}_{91} - z_{91} \end{bmatrix}$$

kurz im Euklidischen Sinn wird, d.h. die Summe

$$\sum_{k=4}^{91} (\hat{x}_k - z_k)^2$$

soll klein werden. Falls die geglättete Zahlenfolge $(\hat{x}_k)_{k=0}^{91}$ der Rekursion (74) gehorcht, erfüllt der Vektor w die Gleichung

$$\begin{bmatrix} \hat{x}_4 \\ \hat{x}_5 \\ \vdots \\ \hat{x}_{91} \end{bmatrix} = \underbrace{\begin{bmatrix} \hat{x}_0 & \hat{x}_1 & \hat{x}_2 & \hat{x}_3 \\ \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & \hat{x}_4 \\ \vdots & \vdots & \vdots & \vdots \\ \hat{x}_{87} & \hat{x}_{88} & \hat{x}_{89} & \hat{x}_{90} \end{bmatrix}}_{=:\, H} w$$

Daher kann man

$$w := \underbrace{(H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}}}_{=:\, H^{\dagger}} \begin{bmatrix} \hat{x}_4 \\ \hat{x}_5 \\ \vdots \\ \hat{x}_{91} \end{bmatrix}$$

wählen. Offenbar ist H entlang aller Nebendiagonalen konstant. Solch eine Matrix heißt **Hankel-Matrix**. Bei unserem Beispiel bekommen wir

$$w = [-0.355, 0.078, -0.135, 1.342]^{\mathrm{T}}.$$

Benutzt man diese Gewichte in der Rekursion (74) und die Anfangsbedingung

$$z_i = x_i, \quad i = 0, 1, 2, 3$$

so erhält man die grüne Kurve in Abbildung 38 als Fortschreibung von $\boldsymbol{X}$. ◇

Der Ansatz (74) ist durch folgenden Zusammenhang gerechtfertigt. Die Zeitreihe

$$\boldsymbol{X} := (x_k)_{k=0}^{N}$$

sei durch die Werte eines trigonometrischen Polynoms zustandegekommen, d.h.

$$x_k := f(k), \quad f(t) := \sum_{k=1}^{n} a_k \cos(\alpha_k t) + b_k \sin(\beta_k t), \quad \alpha_k, \beta_k \in \mathbb{R}$$

wobei in der Menge $\{\alpha_i, \beta_i : i = 1, \ldots, n\}$ keine zwei Elemente übereinstimmen. Bildet man dann für $m \geq 2n$ das Polynom

$$p(z) := z^{2n} - \sum_{k=0}^{2n-1} v_k z^k, \quad v := [v_k]_{k=0}^{2n-1} := \underbrace{\begin{bmatrix} x_0 & \ldots & x_{2n-1} \\ \vdots & & \vdots \\ x_{m-1} & \ldots & x_{m+2n} \end{bmatrix}^{\dagger}}_{=:\, H_{m,2n}(\boldsymbol{X})} \begin{bmatrix} x_{2n} \\ \vdots \\ x_{m+2n+1} \end{bmatrix}$$

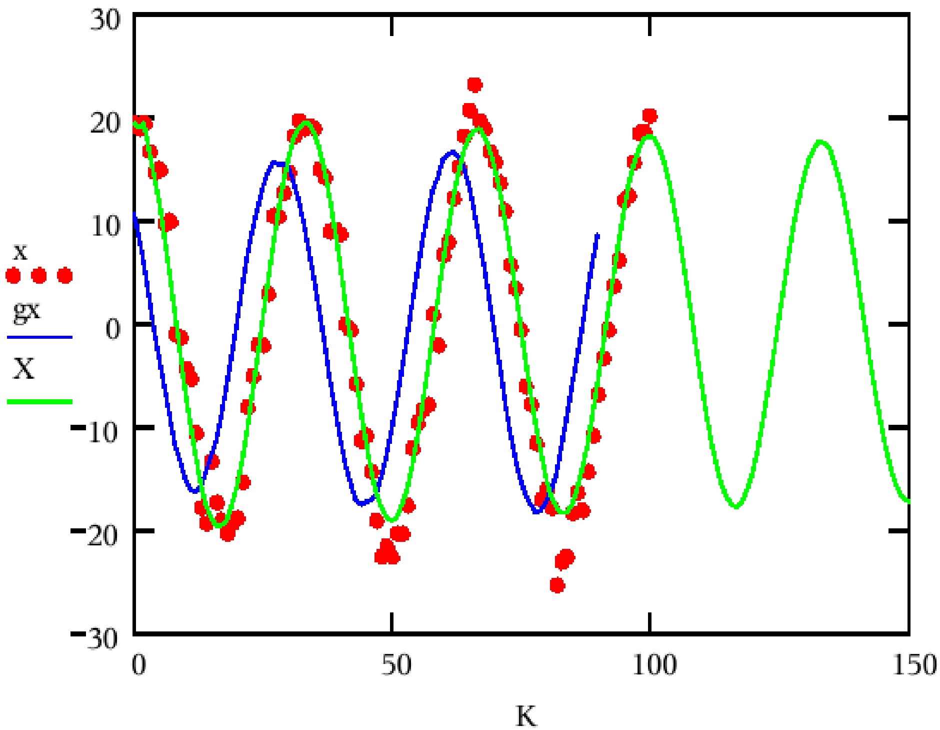


Abbildung 38: Die rote Punktwolke $\{(k, x_k) : k = 0, \ldots, 100\}$ stellt die Zeitreihe $\boldsymbol{X}$ dar, die blaue Kurve verbindet alle Punkte, die man aus $\boldsymbol{X}$ durch den gleitenden Durchschnitt der Ordnung 10 bekommt und die grüne Kurve entsteht durch Verbinden der Punkte (k, z_k), wobei die Folge $(z_k)_{k=0}^{150}$ mittels der Rekursion (74) erzeugt wird, allerdings mit der durch die ungeglättete Zeitreihe gegebenen Anfangsbedingung $z_i := x_i$, $i = 0, \ldots, 3$.

so ist p unabhängig von m und die Nullstellen von p lauten

$$e^{\mathrm{i}\alpha_k},\ e^{\mathrm{i}\beta_k}, \quad k = 1, \dots, n.$$

Daher ist der Vektor v geeignet, um die $\boldsymbol{X}$ zugrundeliegenden Frequenzen α_k und β_k zu schätzen.

Hier bedeutet $H^\dagger$ die Moore-Penrose Inverse von H. Falls $m = 2n$, so wird $H_{m,2n}$ quadratisch und man kann zur gewöhnlichen Inversen übergehen.

Die Koeffizienten a_k und b_k können durch das Gleichungssystem

$$x_i = \sum_{k=1}^{n} a_k \cos(\alpha_k i) + b_k \sin(\beta_k i), \quad i = 0, \dots, 2n-1$$

bestimmt werden.

4.9 Nichlineare Regressionsrechnung

Die lineare Regressionsrechnung versucht ein Polynom ersten, zweiten oder dritten Grades optimal im Sinne der kleinsten Quadrate durch eine gegebene Punktwolke

$$W := \{(x_k, y_k) : k = 0, \dots, n-1\} \subset \mathbb{R}^2$$

hindurch zu legen. Anstatt von Polynomen betrachten wir jetzt die Funktionenklasse

$$y(x) = ae^{bx} + c, \quad a, b, c \in \mathbb{R} \tag{75}$$

der modifizierten Exponentialfunktionen ([17], S.199) und versuchen, die Parameter a, b, c so zu bestimmen, daß die Summe

$$f(a, b, c) := \sum_{k=0}^{n-1} (y_k - ae^{bx_k} - c)^2 \tag{76}$$

möglichst klein wird. Solange man sich in der Klasse der Polynome bewegt, führt solch ein Minimierungsproblem auf ein lineares Gleichungssystem, dessen Lösung den gewünschten Satz an Parametern liefert. Bei Zugrundelegung der Klasse (75) läßt sich das resultierende Minimierungsproblem nicht mehr auf ein lineares Gleichungssystem zurückführen, deshalb sprechen wir hier von nichtlinearer Regressionsrechnung.

Für $n = 3$ erwartet man, aufgrund dreier frei verfügbarer Parameter, die Möglichkeit der Interpolation von drei Punkten:

$$y_k = ae^{bx_k} + c,\ k = 0, 1, 2.$$

Man bekommt unmittelbar

$$\frac{y_0 - c}{y_1 - c} = e^{b(x_0 - x_1)}, \quad \frac{y_1 - c}{y_2 - c} = e^{b(x_1 - x_2)}$$

was auf

$$(y_0 - c)^{x_1 - x_2}(y_2 - c)^{x_0 - x_1} = (y_1 - c)^{x_0 - x_2}$$

führt. Nimmt man äquidistante Abstände zwischen den x_i an, d.h.

$$\Delta := x_1 - x_0 = x_2 - x_1$$

so bekommt man

$$(y_0 - c)^{\Delta}(y_2 - c)^{\Delta} = (y_1 - c)^{2\Delta}.$$

Demnach gilt unter der Voraussetzung $y_0 + y_2 \neq 2y_1$ für den Parameter c die Gleichung

$$c := \frac{y_1^2 - y_0 y_2}{(y_1 - y_0) - (y_2 - y_1)}.$$

Unter Benutzung dieser Darstellung von c bekommt man

$$b := \frac{1}{\Delta} \ln\left(\frac{y_2 - y_1}{y_1 - y_0}\right), \quad a := \frac{y_k - c}{\left(\frac{y_2 - y_1}{y_1 - y_0}\right)^{\frac{x_k}{\Delta}}}$$

wobei bei der Bestimmung von a der Index $k \in \{0, 1, 2\}$ beliebig gewählt werden kann.

In **mathcad** liefert der Befehl

$$\mathrm{expanp}(\boldsymbol{x}, \boldsymbol{y})$$

den Vektor $[a, b, c]^{\mathrm{T}}$. Allerdings ist man bei seiner Anwendung nicht auf die Restriktion $n = 3$ angewiesen. Mit seiner Hilfe bestimmt man das Minimum von (76), was auf die Lösung des nichtlinearen Gleichungssystem

$$\frac{\partial}{\partial a} f(a, b, c) = 0, \quad \frac{\partial}{\partial b} f(a, b, c) = 0, \quad \frac{\partial}{\partial c} f(a, b, c) = 0$$

hinausläuft. Da a und c nur als Linearfaktoren vorkommen, kann man diese beiden Parameter eliminieren, was für b die Gleichung

$$F(b) = 0$$

bedeutet, wobei

$$F := (f_1 - \overline{\boldsymbol{y}} f_2)\left(\frac{1}{n} f_5^2 - f_3\right) - (f_4 - \overline{\boldsymbol{y}} f_5)\left(\frac{1}{n} f_5 f_2 - f_6\right)$$

und

$$\begin{aligned} f_1(b) &:= \boldsymbol{y}^{\mathrm{T}}[x_k e^{b x_k}]_{k=0}^{n-1} & f_2(b) &:= \boldsymbol{x}^{\mathrm{T}} v(b) & f_3(b) &:= v(0)^{\mathrm{T}} v(2b) \\ f_4(b) &:= \boldsymbol{y}^{\mathrm{T}} v(b) & f_5(b) &:= v(0)^{\mathrm{T}} v(b) & f_6(b) &:= \boldsymbol{x}^{\mathrm{T}} v(2b) \end{aligned}$$

$$v(b) := [e^{b x_k}]_{k=0}^{n-1}.$$

Die Nullstellen von F stellen Kandidaten für den Parameter b dar, die Parameter a und c ergeben sich nach Fixierung von b aus dem linearen (2×2)-Gleichungssystem

$$\frac{\partial}{\partial a} f(a, b, c) = 0, \qquad \frac{\partial}{\partial c} f(a, b, c) = 0.$$

Beispiel 69 Man bestimme die Parameter a und b so, daß die beiden Punkte (x_1, y_1) und (x_2, y_2) auf dem Graphen von $y(x) := ae^{bx}$ liegen.

Offenbar müssen a und b das Gleichungssystem

$$y_1 = ae^{bx_1}, \quad y_2 = ae^{bx_2}$$

erfüllen. Vorausgesetzt, daß $y_2 \neq 0$, folgt

$$\frac{y_1}{y_2} = e^{b(x_1 - x_2)}.$$

Sind daher die Vorzeichen von y_1 und y_2 gleich, so gilt

$$b = \frac{\ln\left(\dfrac{y_1}{y_2}\right)}{(x_1 - x_2)} = \frac{\ln(y_1) - \ln(y_2)}{(x_1 - x_2)}.$$

Den Parameter a bekommt man entsprechend

$$a = y_1 e^{-bx_1} = y_2 e^{-bx_2}.$$

Die Berechnung zeigt, daß nicht für jedes Punktepaar diese Aufgabe eine Lösung besitzt, sobald y_1 und y_2 ungleiches Vorzeichen besitzt oder genau einer von beiden verschwindet, besitzt das Problem keine Lösung.

Man bestimme eine kleinste Quadratelösung.

Zu diesem Zweck hat man die Funktion

$$f(a, b) := (y_1 - ae^{bx_1})^2 + (y_2 - ae^{bx_2})^2$$

zu minimieren. Die partiellen Ableitungen lauten

$$\begin{aligned} f_a(a, b) &= -2(y_1 - ae^{bx_1})e^{bx_1} - 2(y_2 - ae^{bx_2})e^{bx_2}, \\ f_b(a, b) &= -2(y_1 - ae^{bx_1})e^{bx_1}ax_1 - 2(y_2 - ae^{bx_2})e^{bx_2}ax_2. \end{aligned}$$

Daher liefern die Gleichungen $f_a(a, b) = f_b(a, b) = 0$ das Gleichungssystem

$$\begin{aligned} (y_1 - ae^{bx_1})e^{bx_1} + (y_2 - ae^{bx_2})e^{bx_2} &= 0 \\ (y_1 - ae^{bx_1})e^{bx_1}x_1 + (y_2 - ae^{bx_2})e^{bx_2}x_2 &= 0 \end{aligned}$$

Auflösung der ersten Gleichung nach a liefert

$$a = \frac{y_1 e^{bx_1} + y_2 e^{bx_2}}{e^{2bx_1} + e^{2bx_2}}.$$

Einsetzen in die zweite Gleichung liefert

$$\begin{aligned} 0 &= y_1 e^{bx_1} x_1 + y_2 e^{bx_2} x_2 - a\left(e^{2bx_1}x_1 + e^{2bx_2}x_2\right) \\ &= y_1 e^{bx_1} x_1 + y_2 e^{bx_2} x_2 - \left(\frac{y_1 e^{bx_1} + y_2 e^{bx_2}}{e^{2bx_1} + e^{2bx_2}}\right)\left(e^{2bx_1}x_1 + e^{2bx_2}x_2\right). \end{aligned}$$

Multiplikation mit $e^{2bx_1} + e^{2bx_2}$ liefert

$$\begin{aligned} 0 &= \left(y_1 e^{bx_1} x_1 + y_2 e^{bx_2} x_2\right)\left(e^{2bx_1} + e^{2bx_2}\right) - \left(y_1 e^{bx_1} + y_2 e^{bx_2}\right)\left(e^{2bx_1} x_1 + e^{2bx_2} x_2\right) \\ &= (x_1 - x_2)(y_1 e^{(x_1+2x_2)b} - y_2 e^{(x_2+2x_1)b}) \end{aligned}$$

Da $x_1 \neq x_2$ folgt aus dieser Gleichung

$$y_1 e^{(x_1+2x_2)b} = y_2 e^{(x_2+2x_1)b}$$

d.h.

$$y_1 = y_2 e^{(x_1-x_2)b}$$

sodaß das kleinste Quadrate Problem auch nur genau dann eine Lösung besitzt, wenn das exakte Problem lösbar ist. Insbesondere stimmen beide Lösungen überein.

Man bestimme für drei Punkte (x_k, y_k), $k = 1, 2, 3$ eine kleinste Quadrate Lösung.

Wie oben setzt man

$$f(a, b) := (y_1 - ae^{bx_1})^2 + (y_2 - ae^{bx_2})^2 + (y_3 - ae^{bx_3})^2.$$

Die partiellen Ableitungen lauten

$$\begin{aligned} f_a(a,b) &= -2(y_1 - ae^{bx_1})e^{bx_1} - 2(y_2 - ae^{bx_2})e^{bx_2} - 2(y_3 - ae^{bx_3})e^{bx_3} \\ f_b(a,b) &= -2(y_1 - ae^{bx_1})e^{bx_1} ax_1 - 2(y_2 - ae^{bx_2})e^{bx_2} ax_2 - 2(y_3 - ae^{bx_3})e^{bx_3} ax_3 \end{aligned}$$

Daher liefern die Gleichungen $f_a(a,b) = f_b(a,b) = 0$ das Gleichungssystem

$$\begin{aligned} (y_1 - ae^{bx_1})e^{bx_1} + (y_2 - ae^{bx_2})e^{bx_2} + (y_3 - ae^{bx_3})e^{bx_3} &= 0, \\ (y_1 - ae^{bx_1})e^{bx_1} x_1 + (y_2 - ae^{bx_2})e^{bx_2} x_2 + (y_3 - ae^{bx_3})e^{bx_3} x_3 &= 0. \end{aligned}$$

Auflösung der ersten Gleichung nach a liefert

$$a = \frac{y_1 e^{bx_1} + y_2 e^{bx_2} + y_3 e^{bx_3}}{e^{2bx_1} + e^{2bx_2} + e^{3bx_3}}.$$

Einsetzen in die zweite Gleichung und Multiplikation mit $e^{2bx_1} + e^{2bx_2} + e^{2bx_3}$ liefert

$$\begin{aligned} 0 = &\ (y_1 e^{bx_1} x_1 + y_2 e^{bx_2} x_2 + y_3 e^{bx_3} x_3)(e^{2bx_1} + e^{2bx_2} + e^{2bx_3}) \\ - &\ (y_1 e^{bx_1} + y_2 e^{bx_2} + y_3 e^{bx_3})(e^{2bx_1} x_1 + e^{2bx_2} x_2 + e^{2bx_3} x_3) \end{aligned}$$

was auf

$$\begin{aligned} 0 = &\ y_1 e^{bx_1}((x_1 - x_2)e^{2bx_2} + (x_1 - x_3)e^{2bx_3}) \\ + &\ y_2 e^{bx_2}((x_2 - x_1)e^{2bx_1} + (x_2 - x_3)e^{2bx_3}) \\ + &\ y_3 e^{bx_3}((x_3 - x_1)e^{2bx_1} + (x_3 - x_2)e^{2bx_2}) \end{aligned}$$

führt. Setzen wir $z = e^{b\Delta}$ und nehmen wir äquidistante Abstände zwischen den x_k, d.h. $x_k = k\Delta$, so bekommen wir folgendes Polynom

$$\begin{aligned} p(z) = &\ y_1 z(-\Delta z^4 - 2\Delta z^6) + y_2 z^2(\Delta z^2 - \Delta z^6) + y_3 z^3(2\Delta z^2 + \Delta z^4) \\ = &\ \Delta(y_1 z(-z^4 - 2z^6) + y_2 z^2(z^2 - z^6) + y_3 z^3(2z^2 + z^4)) \\ = &\ \Delta z^4(y_1 z(-1 - 2z^2) + y_2(1 - z^2) + y_3(2z + z^3)) \end{aligned}$$

dessen positive Nullstellen Kandidaten für b liefern. ◇

5 Übungsaufgaben mit Lösungen

1. Aufgabe: Es gelte $\Omega := \{\omega_0, \ldots, \omega_4\}$, wobei

$$P(\{\omega_0\}) = 1/8,\ P(\{\omega_1\}) = 1/4,\ P(\{\omega_2\}) = 1/8,\ P(\{\omega_3\}) = 1/4,\ P(\{\omega_4\}) = 1/4.$$

Das Ereignis A sei erklärt durch

$$A = \{\omega_0, \omega_1, \omega_2\}.$$

Man erkläre ein zu A unabhängiges Ereignis B und begründe die Unabhängigkeit.

Antwort: Eine Möglichkeit besteht in

$$B := \{\omega_1, \omega_4\}.$$

Dann gilt

$$A \cap B = \{\omega_1\} \Rightarrow P(A \cap B) = 1/4$$
$$P(A) = P(\{\omega_0\}) + P(\{\omega_1\}) + P(\{\omega_2\}) = 1/2$$

und

$$P(B) = P(\{\omega_1\}) + P(\{\omega_4\}) = 1/2$$

sodaß

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

Aufgabe 1a: Der Ereignisraum Ω bestehe aus 9 Elementarereignissen $\{\omega_1, \ldots, \omega_9\}$, wobei für das Eintreten von ω_k die Wahrscheinlichkeit

$$P(\{\omega_k\}) = \frac{k}{45}$$

gelte.

- Überprüfen Sie $P(\Omega) = 1$.
- Die Ereignisse A und B seien durch $A = \{\omega_6, \omega_7, \omega_8\}$ und $B := \{\omega_3, \omega_5, \omega_7\}$ erklärt. Sind A und B unabhängig voneinander oder nicht? Begründen Sie Ihre Entscheidung.

Antwort: Es liegt Unabhängigkeit vor, denn es gilt

$$A \cap B = \{\omega_7\} \Rightarrow P(A \cap B) = \frac{7}{45}$$
$$P(A) = P(\{\omega_6\}) + P(\{\omega_7\}) + P(\{\omega_8\}) = \frac{6+7+8}{45} = \frac{21}{45}$$

und

$$P(B) = P(\{\omega_3\}) + P(\{\omega_5\}) + P(\{\omega_7\}) = \frac{3+5+7}{45} = \frac{15}{45}$$

sodaß

$$P(A \cap B) = \frac{7}{45} = \frac{21}{45} \times \frac{15}{45} = P(A)P(B).$$

Aufgabe 1b: Es werde mit zwei Hexaedern gewürfelt, die Zg X sei durch die Summe der oben aufliegenden Augenanzahlen erklärt. Das Ereignis A tritt ein, falls X einen durch 3 teilbaren Wert annimmt, und das Ereignis B, falls X einen Wert von nicht mehr als 7 annimmt. Man überprüfe, ob A und B unabhängig sind.

2. Aufgabe: Der Ereignisraum Ω sei gegeben durch

$$\Omega := \{(i,j) : i,j = 1,2,3\}$$

und die Zufallsgröße $X : \Omega \mapsto \mathbb{N}$ durch

$$X((i,j)) := 2 - \frac{ij}{4}(9 - 3(i+j) + ij).$$

Alle Elementarereignisse treten mit der gleichen Wahrscheinlichkeit $\frac{1}{\#\Omega}$ auf.

- Man bestimme die Mengen

$$A_k := \{\omega \in \Omega : X(\omega) = k\}$$

  wobei k den Wertevorrat von X durchläuft.

- Man bestimme die Wahrscheinlichkeitsfunktion f von X.

- Man bestimme die Verteilungsfunktion F von X.

Antwort: Zur Bestimmung der Mengen A_k benötigt man den Werteverlauf von X, der

X	1	2	3
1	1	1	2
2	1	1	2
3	2	2	2

lautet. Demnach besteht der Wertevorrat von X nur aus 1 und 2 und es verbleibt, die Mengen A_1 und A_2 zu berechnen:

$$A_1 = \{(1,1),(1,2),(2,1),(2,2)\}, \quad A_2 = \{(1,3),(2,3),(3,3),(3,2),(3,1)\}.$$

Die Wahrscheinlichkeitsfunktion f von X ist definiert durch

$$f : X(\Omega) \mapsto [0,1], \quad f(k) := P(X = k)$$

wobei wir, wegen der Voraussetzung, daß alle Elementarereignisse gleichwahrscheinlich sind, zu

$$f(k) = P(X = k) = \frac{\#A_k}{\#\Omega} = \begin{cases} \frac{4}{9} & \text{falls } k = 1 \\ \frac{5}{9} & \text{falls } k = 2 \end{cases}$$

gelangen.

Die Verteilungsfunktion F ist definiert durch

$$F : \mathbb{R} \mapsto [0,1], \quad F(x) := P(X \leq x).$$

Da unser X keine Werte kleiner als 1 annehmen kann, gilt für $x < 1$ die Gleichung $F(x) = 0$.

Da für $x \in [1,2)$ aus der Ungleichung $X \leq x$ die Gleichung $X = 1$ folgt, bekommen wir

$$F(x) = P(X \leq x) = P(X = 1) = f(1) = \frac{4}{9} \quad \text{falls } x \in [1,2).$$

Da für $2 \leq x$ aus der Ungleichung $X \leq x$ entweder $X = 1$ oder $X = 2$ folgt, bekommen wir

$$F(x) = P(X \leq x) = P(X = 1) + P(X = 2) = f(1) + f(2) = 1, \quad \text{falls } x \in [2,+\infty).$$

Zusammenfassend können wir

$$F(x) = \begin{cases} 0 & \text{falls } x < 1 \\ 4/9 & \text{falls } 1 \leq x < 2 \\ 1 & \text{falls } 2 \leq x \end{cases}$$

schreiben.

Aufgabe 2a: Der Ereignisraum Ω sei gegeben durch $\Omega := \{(i,j) : i,j = 0,1,2\}$ und die Zufallsgröße $X : \Omega \mapsto \mathbb{N}$ durch

$$X((i,j)) := 1 + ij(4 - 2(i+j) + ij).$$

Alle Elementarereignisse treten mit der gleichen Wahrscheinlichkeit $\frac{1}{\#\Omega}$ auf.

- Man bestimme den Wertevorrat von X.
- Man bestimme die Mengen $A_k := \{\omega \in \Omega : X(\omega) = k\}$, wobei k den Wertevorrat von X durchläuft.
- Man bestimme die Wahrscheinlichkeitsfunktion $f(k) := P(A_k)$ von X.
- Man bestimme die Verteilungsfunktion $F(x) := P(X \leq x)$ von X.

Antwort: Der Werteverlauf von X lautet

X	0	1	2
0	1	1	1
1	1	2	1
2	1	1	1

Demnach besteht der Wertevorrat von X nur aus 1 und 2 und es verbleibt, die Mengen A_1 und A_2 zu berechnen:

$$A_1 = \Omega \setminus \{(1,1)\} \quad A_2 = \{(1,1)\}.$$

Die Wahrscheinlichkeitsfunktion f von X ist definiert durch

$$f : X(\Omega) \mapsto [0,1], \quad f(k) := P(X = k)$$

wobei wir, wegen der Voraussetzung, daß alle Elementarereignisse gleichwahrscheinlich sind, zu

$$f(k) = P(X = k) = \frac{\#A_k}{\#\Omega} = \begin{cases} \frac{8}{9} & \text{falls } k = 1 \\ \frac{1}{9} & \text{falls } k = 2 \end{cases}$$

gelangen.

Die Verteilungsfunktion F ist definiert durch

$$F : \mathbb{R} \mapsto [0,1], \quad F(x) := P(X \leq x).$$

Da unser X keine Werte kleiner als 1 annehmen kann, gilt für $x < 1$ die Gleichung $F(x) = 0$.

Da für $x \in [1,2)$ aus der Ungleichung $X \leq x$ die Gleichung $X = 1$ folgt, bekommen wir

$$F(x) = P(X \leq x) = P(X = 1) = f(1) = \frac{8}{9} \quad \text{falls } x \in [1,2).$$

Da für $2 \leq x$ aus der Ungleichung $X \leq x$ entweder $X = 1$ oder $X = 2$ folgt, bekommen wir

$$F(x) = P(X \leq x) = P(X = 1) + P(X = 2) = f(1) + f(2) = 1, \quad \text{falls } x \in [2, +\infty).$$

Zusammenfassend können wir

$$F(x) = \begin{cases} 0 & \text{falls } x < 1 \\ 8/9 & \text{falls } 1 \leq x < 2 \\ 1 & \text{falls } 2 \leq x \end{cases}$$

schreiben.

3. Aufgabe: Bei einem Spielautomaten fangen nach Einwurf eines Euros drei Symbolscheiben an zu rotieren, die nach einer gewissen Zeit zum Stehen kommen und in drei Sichtfeldern jeweils eines der Symbole 'Apfel', 'Birne' oder 'Pflaume' zeigen. Jede Scheibe zeigt mit gleicher Wahrscheinlichkeit eines dieser Symbole an, alle Scheiben drehen sich unabhängig voneinander. Falls die ersten beiden Symbole übereinstimmen und das dritte verschieden ist, bekommt der Spieler als Gewinn zwei Euro, falls alle drei Symbole übereinstimmen, bekommt er drei Euro. Bei allen anderen Konstellationen gewinnt der Spieler nichts.

- Beschreiben Sie den Ereignisraum Ω mengenmäßig. Wieviele Elementarereignisse enthält Ω?
- Beschreiben Sie die Ereignisse $A :=$ 'Erste beiden Symbole stimmen überein, drittes ist verschieden' und $B :=$ 'Alle drei Symbole stimmen überein'.
- Erklären Sie eine Zufallsgröße X, die die Gewinnsituation widergibt.
- Berechnen Sie den Erwartungswert von X und entscheiden Sie, ob es sich lohnt, längere Zeit mit dem Automaten zu spielen. Wie müßte man den Gewinn beim Eintreten des Ereignisses A abändern, damit man bei längerem Spielen weder verliert noch gewinnt?
- Geben Sie die Verteilungsfunktion von X an.

Antwort:

- Zur Beschreibung von Ω kann man die Symbole a, b und p benutzen. Dann gilt

$$\Omega := \left\{ \begin{array}{ccccccccc} aaa & aab & aap & aba & abb & abp & apa & apb & app \\ baa & bab & bap & bba & bbb & bbp & bpa & bpb & bpp \\ paa & pab & pap & pba & pbb & pbp & ppa & ppb & ppp \end{array} \right\}$$

  Offenbar besteht Ω aus 27 Elementarereignissen.

- $$A = \{aab, aap, bba, bbp, ppa, ppb\}\,, \quad B := \{aaa, bbb, ppp\}.$$

- $$X(\omega) := \begin{cases} 0 & \text{falls } \omega \in \Omega \setminus (A \cup B) \\ 2 & \text{falls } \omega \in A, \\ 3 & \text{falls } \omega \in B. \end{cases}$$

- $$EX = 2P(A) + 3P(B) = 2 \times \frac{6}{27} + 3 \times \frac{3}{27} = \frac{12}{27} + \frac{9}{27} = \frac{21}{27}.$$

  Da $21/27 < 1$ lohnt sich das Spiel nicht.

  Damit das Spiel ausgeglichen wird, muß der Erwartungswert mit dem Einsatz übereinstimmen, d.h.

$$EX = 1.$$

  Demnach hat man die Gleichung

$$x \times \frac{6}{27} + 3 \times \frac{3}{27} = 1$$

  zu lösen, was $x = 3$ ergibt.

- $$F(x) = P(X \le x) = \begin{cases} 0 & \text{falls } x < 0 \\ 18/27 & \text{falls } 0 \le x < 2 \\ 18/27 + 6/27 & \text{falls } 2 \le x < 3 \\ 1 & \text{falls } 3 \le x \end{cases}$$

Aufgabe 3a: Bei einem Spielautomaten fangen nach Einwurf eines Euros, drei Symbolscheiben an zu rotieren, die nach einer gewissen Zeit zum Stehen kommen und in drei Sichtfeldern jeweils eine Eins, eine Zwei oder eine Drei zeigen. Jede Scheibe zeigt mit gleicher Wahrscheinlichkeit eine dieser Zahlen an und alle drei Scheiben drehen sich unabhängig voneinander. Falls die Summe der drei Zahlen gleich

- 6 ist, bekommt der Spieler 1.5 Euro
- 7 oder 8 ist, bekommt der Spieler 2.0 Euro
- 9 ist, bekommt der Spieler 2.5 Euro.

Bei allen anderen Konstellationen bekommt der Spieler nichts.

- Wieviele Elementarereignisse enthält der Ereignisraum Ω?
- Beschreiben Sie die Ereignisse

$$\begin{aligned} A &:= \text{Die Summe ist gleich 6} \\ B &:= \text{Die Summe ist gleich 7 oder 8} \\ C &:= \text{Die Summe ist gleich 9} \end{aligned}$$

  mengenmäßig und geben Sie $P(A)$, $P(B)$, $P(C)$ und $P(\Omega \setminus (A \cup B \cup C))$ an.

- Erklären Sie eine Zufallsgröße X, die das Ausgabeverhalten des Automaten widergibt.
- Berechnen Sie den Erwartungswert von X und entscheiden Sie, ob es sich lohnt, mit dem Automaten längere Zeit zu spielen. Wie müßte man den Ausgabewert beim Eintreten des Ereignisses B abändern, damit man bei längerem Spielen weder verliert noch gewinnt?
- Geben Sie die Verteilungsfunktion $F(x) := P(X \le x)$ von X an.

Antwort:

- $$\Omega := \left\{ \begin{array}{ccccccccc} 111 & 112 & 113 & 121 & 122 & 123 & 131 & 132 & 133 \\ 211 & 212 & 213 & 221 & 222 & 223 & 231 & 232 & 233 \\ 311 & 312 & 313 & 321 & 322 & 323 & 331 & 332 & 333 \end{array} \right\}$$

Offenbar besteht Ω aus 27 Elementarereignissen. Die zugehörigen Summen lauten

$$\left\{\begin{array}{ccccccccc} 3 & 4 & 5 & 4 & 5 & 6 & 5 & 6 & 7 \\ 4 & 5 & 6 & 5 & 6 & 7 & 6 & 7 & 8 \\ 5 & 6 & 7 & 6 & 7 & 8 & 7 & 8 & 9 \end{array}\right\}$$

-
$$A := \{123, 132, 213, 222, 231, 312, 321\}$$
$$B := \{133, 223, 232, 313, 322, 331, 233, 323, 332\}$$
$$C := \{333\}$$

-
$$X(\omega) := \begin{cases} 0.0 & \text{falls } \omega \in \Omega \setminus (A \cup B \cup C) \\ 1.5 & \text{falls } \omega \in A, \\ 2.0 & \text{falls } \omega \in B, \\ 2.5 & \text{falls } \omega \in C. \end{cases}$$

-
$$EX = 1.5P(A) + 2P(B) + 2.5P(C) = \frac{3}{2} \times \frac{7}{27} + 2 \times \frac{9}{27} + \frac{5}{2} \times \frac{1}{27} = \frac{31}{27} = 1.148.$$

  Da $1 < 31/27$ lohnt sich das Spiel.

  Damit das Spiel ausgeglichen wird, muß der Erwartungswert mit dem Einsatz übereinstimmen, d.h.
$$EX = 1.$$

  Demnach hat man die Gleichung
$$\frac{3}{2} \times \frac{7}{27} + x \times \frac{9}{27} + \frac{5}{2} \times \frac{1}{27} = 1$$

  zu lösen, was $x = \dfrac{14}{9}$ ergibt.

-
$$F(x) = P(X \leq x) = \frac{1}{27} \begin{cases} 0 & \text{falls } x < 0 \\ 10 & \text{falls } 0 \leq x < 1.5 \\ 17 & \text{falls } 1.5 \leq x < 2.0 \\ 26 & \text{falls } 2.0 \leq x < 2.5 \\ 27 & \text{falls } 2.5 \leq x \end{cases}$$

4. Aufgabe: Ein Hotelbesitzer geht davon aus, daß 10 % der Buchungen nicht zustande kommen. Deshalb hat er für seine 10 Betten 11 Buchungen angenommen. Wie groß ist die Wahrscheinlichkeit, daß zu viele Gäste kommen? (Es darf Unabhängigkeit unter den Gästen angenommen werden. Binomialverteilung)

Antwort: Die Annahme, daß 10 % der Buchungen nicht zustande kommen, bedeutet, daß mit der Wahrscheinlichkeit $p = 0.9$ der Gast tatsächlich kommt. Zuviele Gäste kommen, falls alle 11 Gäste kommen. Die Wahrscheinlichkeit dafür ist

$$p^{11} = 0.9^{11} = 0.3138.$$

Aufgabe 4a: Eine Fluggesellschaft geht davon aus, daß 15 % der Buchungen nicht zustande kommen. Deshalb hat sie für die 20 Plätze in einer ihrer Maschinen 22 Buchungen angenommen. Wie groß ist die Wahrscheinlichkeit, daß zu viele Fluggäste kommen? Es darf Unabhängigkeit unter den Gästen angenommen werden. (Binomialverteilung)

Antwort: Die Annahme, daß 15 % der Buchungen nicht zustande kommen, bedeutet, daß mit der Wahrscheinlichkeit $p = 0.85$ der Fluggast erscheint. Zuviele Gäste kommen, falls 21 oder 22 Gäste erscheinen. Die Wahrscheinlichkeit dafür ist

$$\binom{22}{21} p^{21}(1-p)^1 + \binom{22}{22} p^{22}(1-p)^0 = 22(0.85)^{21} 0.15 + (0.85)^{22} = 0.137.$$

5. Aufgabe: 20 % der Fahrgäste eines Verkehrsunternehmens sind Schwarzfahrer. Es werden 100 Personen zufällig ausgewählt und überprüft. Die Zg X sei die 'Anzahl der Schwarzfahrer unter den 100 ausgewählten Personen'.

a) Falls der Grenzwertsatz von de-Moivre-Laplace sich anwenden läßt, berechnen Sie folgende Wahrscheinlichkeiten:

- die Wahrscheinlichkeit, daß sich unter den 100 ausgewählten Personen höchstens 20
- von 10 bis 30
- mehr als 25 Schwarzfahrer befinden.

(Binomialverteilung, Normalverteilung)

b) Bestimmen Sie mit dem Grenzwertsatz ein zu 20 symmetrisch gelegenes möglichst kleines Intervall mit der Eigenschaft, daß X mit einer Wahrscheinlichkeit von 95 % in diesem Intervall liegt.

Antwort:

a) Die Aufgabe kann mittels eines Bernouilliexperiments vom Umfang $n := 100$ behandelt werden, wobei das Experiment mit der Wahrscheinlichkeit $p = 0.2$ gelingt. Im Einzelnen hat man

- $P(X \leq 20) = \sum_{k=0}^{20} \binom{100}{k} p^k (1-p)^{100-k}$
- $P(10 \leq X \leq 30) = \sum_{k=10}^{30} \binom{100}{k} p^k (1-p)^{100-k}$

- $P(25 < X) = 1 - P(X \leq 25) = 1 - \sum_{k=0}^{25} \binom{100}{k} p^k (1-p)^{100-k}$

zu berechnen. Für diese Zwecke darf der Grenzwertsatz von de-Moivre-Laplace verwendet werden, falls $np > 10$ und $n(1-p) > 10$. Wegen $np = 20$ und $n(1-p) = 80$ ist diese Bedingung erfüllt, so daß

$$P(a \leq X \leq b) \approx \Phi_{\mu,\sigma}(b) - \Phi_{\mu,\sigma}(a), \quad \mu = np = 20, \quad \sigma = \sqrt{np(1-p)} = \sqrt{16} = 4.$$

Demzufolge bekommen wir

- $P(X \leq 20) = \Phi_{20,4}(20) = \Phi_{0,1}(0) = 0.5$
- $$\begin{aligned} P(10 \leq X \leq 30) &= \Phi_{20,4}(30) - \Phi_{20,4}(10) \\ &= \Phi_{0,1}((30-20)/4) - \Phi_{0,1}((10-20)/4) \\ &= \Phi_{0,1}(2.5) - \Phi_{0,1}(-2.5) = 2\Phi_{0,1}(2.5) - 1 \\ &= 2 \times 0.9938 - 1 = 0.9876 \end{aligned}$$
- $$\begin{aligned} P(25 < X) &= 1 - \Phi_{20,4}(25) \\ &= 1 - \Phi_{0,1}((25-20)/4) \\ &= 1 - \Phi_{0,1}(1.25) \\ &= 1 - 0.8944 = 0.1056. \end{aligned}$$

b) Es wird verlangt, eine Konstante c so zu bestimmen, daß

$$0.95 = P(20 - c \leq X \leq 20 + c).$$

Da der Grenzwertsatz benutzt werden soll, bekommt man

$$0.95 = \Phi_{20,4}(20+c) - \Phi_{20,4}(20-c) = \Phi_{0,1}(c/4) - \Phi_{0,1}(-c/4) = 2\Phi_{0,1}(c/4) - 1$$

d.h.

$$1.95/2 = 0.975 = \Phi_{0,1}(c/4).$$

Demnach gilt

$$1.96 = c/4$$

d.h. $c = 7.84$. Da X nur natürliche Werte annehmen kann, verwenden wir $c = 8$, d.h. das gewünschte Intervall lautet $[12, 28]$.

Aufgabe 5a: Es zeigt sich, daß bei einer Getränkeabfüllanlage 6 % der Flaschen nicht ausreichend gefüllt werden (halbvoll). Es werden 200 Flaschen zufällig ausgewählt und die Anzahl halbvoller Flaschen gezählt. Die Zg X lautet daher

Anzahl der halbvollen Flaschen unter 200 zufällig ausgewählten Flaschen.

- Falls der Grenzwertsatz von de-Moivre-Laplace sich anwenden läßt, berechnen Sie die Wahrscheinlichkeit, daß sich unter 200 zufällig ausgewählten Flaschen genau 12 halbvolle befinden. (Binomialverteilung, Dichte der Normalverteilung)

- Bestimmen Sie mit dem Grenzwertsatz ein zu 12 symmetrisch gelegenes Intervall mit der Eigenschaft, daß X mit einer Wahrscheinlichkeit von 90 % in diesem Intervall liegt.

Antwort: a) Die Aufgabe kann mittels eines Bernouilliexperiments vom Umfang $n := 200$ behandelt werden, wobei das Experiment mit der Wahrscheinlichkeit $p = 0.06$ gelingt. Im Einzelnen hat man

$$P(X = 12) = \binom{200}{12} p^{12}(1-p)^{200-12}$$

zu berechnen. Für diese Zwecke darf der (lokale) Grenzwertsatz von de-Moivre-Laplace verwendet werden, falls $np > 10$ und $n(1-p) > 10$. Wegen $np = 12$ und $n(1-p) = 188$, ist diese Bedingung erfüllt, so daß ([20], S.1054)

$$P(X = k) \approx \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(k-\mu)^2}{2\sigma^2}}, \quad \mu := np = 12, \quad \sigma := \sqrt{np(1-p)} = 3.359.$$

Demzufolge bekommen wir $P(X = 12) \approx 0.119$.

b) Es wird verlangt, eine Konstante c so zu bestimmen, daß

$$0.90 = P(12 - c \leq X \leq 12 + c).$$

Da der Grenzwertsatz benutzt werden soll, bekommt man

$$0.90 = \Phi_{\mu,\sigma}(\mu + c) - \Phi_{\mu,\sigma}(\mu - c) = \Phi_{0,1}(c/\sigma) - \Phi_{0,1}(-c/\sigma) = 2\Phi_{0,1}(c/\sigma) - 1$$

d.h.

$$1.90/2 = 0.95 = \Phi_{0,1}(c/\sigma).$$

Durch Aufsuchen von 0.95 im Inneren der Normalverteilungstabelle findet man $x = 1.645$. Demnach gilt

$$1.645 = c/\sigma$$

d.h.

$$c = 1.645 \times 3.359 = 5.52.$$

6. Aufgabe: Eine Maschine füllt Bonbons in Tüten ab. Das Gewicht der Tüten wird als normalverteilt angesehen mit den Parametern $\mu = 100$ g und $\sigma = 2$ g. Eine Tüte wird als untergewichtig bezeichnet, falls sie weniger als 96 g wiegt. Für den Großversand werden jeweils 200 Tüten in einen Karton verpackt. Wieviele untergewichtige Tüten sind pro Karton zu erwarten? (Normalverteilung, Binomialverteilung)

Antwort: X sei das Tütengewicht. Dann gilt $X \sim \mathscr{N}(\mu, \sigma^2)$, $\mu = 100$, $\sigma = 2$. Die Wahrscheinlichkeit, daß eine Tüte untergewichtig ist ergibt sich dann aus

$$P(X \leq 96) = \Phi_{0,1}\left(\frac{96 - 100}{2}\right) = \Phi_{0,1}(-2) = 1 - \Phi_{0,1}(2) = 1 - 0.9778 = 0.0228.$$

Das Abpacken in die Kartons kann als ein Bernoulliexperiment im Umfang von $n = 200$ aufgefaßt werden, wobei der Grundversuch positiv ausgeht, wenn man eine untergewichtige Tüte in den Karton hineinlegt. Die Grundwahrscheinlichkeit dafür lautet $p := 0.0228$. Da der Erwartungswert np lautet hat man mit

$$np = 200 \times 0.0228 = 4.56$$

untergewichtigen Tüten pro Karton zu rechnen.

Aufgabe 6a: Eine Fluggesellschaft möchte ihren Gewinn steigern. Zu diesem Zweck sollen übergewichtige Passagiere außer dem Grundpreis noch einen Zuschlag bezahlen. Das Gewicht der Passagiere wird als normalverteilt angesehen mit den Parametern $\mu = 80$ kg und $\sigma = 8$ kg. Die Gesellschaft hat sich entschieden, einen Passagier als übergewichtig zu klassifizieren, falls er mehr als 90 kg wiegt. Pro Flug werden jeweils 240 Passagiere transportiert.

- Zur Schätzung der Mehreinnahmen möchte man nun wissen, wieviele übergewichtige Passagiere pro Flug zu erwarten sind? Beantworten Sie diese Frage. (Normalverteilung, Binomialverteilung)
- Nachdem man die Anzahl der zu erwartenden übergewichtigen Passagiere kennt, macht sich bei der Finanzverwaltung Unzufriedenheit breit. Wie muß man die Grenze, ab der eine Person als übergewichtig gilt, wenigstens absenken, um mit 30 übergewichtigen Fluggästen pro Flug rechnen zu können?

Antwort: X sei das Gewicht eines einzelnen Passagiers. Dann gilt

$$X \sim \mathscr{N}(\mu, \sigma^2), \quad \mu = 80, \quad \sigma = 8.$$

Die Wahrscheinlichkeit, daß ein Passagier übergewichtig ist, ergibt sich dann aus

$$\begin{aligned} P(90 \leq X) = 1 - P(X < 90) &= 1 - \Phi_{0,1}\left(\frac{90-80}{8}\right) \\ &= 1 - \Phi_{0,1}(1.25) \\ &= 1 - 0.8944 = 0.1056. \end{aligned}$$

Eine Ansammlung von 240 Passagieren kann als ein Bernoulliexperiment im Umfang von $n = 240$ aufgefaßt werden, wobei der Grundversuch positiv ausgeht, wenn man auf einen übergewichtigen Passagier trifft. Die Grundwahrscheinlichkeit dafür lautet

$$p := 0.1056.$$

Da der Erwartungswert np lautet, hat man mit

$$np = 240 \times 0.1056 = 25.344$$

übergewichtigen Passagieren pro Flug zu rechnen.

Wenn man wenigstens 30 übergewichtige Passagiere unter 240 Passagieren erwarten möchte, so muß

$$240p \geq 30$$

gelten, d.h.

$$p \geq 30/240 = 1/8.$$

Für die unbekannte Übergewichtsgrenze x muß daher die Ungleichung

$$1 - \Phi_{0,1}\left(\frac{x-80}{8}\right) \geq \frac{1}{8}$$

erfüllt sein. Da x so groß wie möglich sein soll, hat man

$$0.875 = \frac{7}{8} = \Phi_{0,1}\left(\frac{x-80}{8}\right)$$

zu lösen, was auf

$$1.15 = \frac{x-80}{8}$$

führt. Demnach muß man einen Passagier als übergewichtig klassifizieren, falls er mehr als

$$x = 89.2 \text{ kg}$$

wiegt.

7. Aufgabe: Man berechne alle reellen Zahlen a und b, für die die Funktion

$$f(x,a,b) := \begin{cases} 0 & \text{falls} \quad x < -1 \\ x^2 + bx + a & \text{falls} \ -1 \leq x < 1 \\ 0 & \text{falls} \quad 1 \leq x \end{cases}$$

als Dichtefunktion verwendet werden kann.

Unter Benutzung dieser Werte löse man folgende Teilaufgaben.

- Berechnen Sie EX und $\mathrm{Var}(X)$, wobei X entsprechend $F_{a,b}(x) := \int_{-\infty}^{x} f(t,a,b)\mathrm{d}t$ verteilt ist. Für welche b streut die Zg X am wenigsten um ihren Erwartungswert?
- Es sei $\boldsymbol{x}$ eine Stichprobe aus einer entsprechend $F_{a,b}$ verteilten Grundgesamtheit. Schätzen Sie ab, welche Werte die Spannweite S und das Stichprobenmittel $\overline{\boldsymbol{x}}$ annehmen werden.
- Für welche Parameterwerte b wird $f(x,a,b)$ symmetrisch.
- Für welche Parameterwerte b wird $f(x,a,b)$ linksschief und für welche Parameterwerte rechtsschief?

- Überprüfen Sie die Resultate aus der vorhergehenden Teilaufgabe, indem Sie in Abhängigkeit von b die Anteile der statistischen Masse rechts und links vom oben geschätzten Zentralwert berechnen. Der Anteil der statistischen Masse links von z ist durch $F_{a,b}(z)$ und der Anteil rechts von z durch $1 - F_{a,b}(z)$ erklärt.

Antwort: Es müssen die beiden Eigenschaften

$$\int_{-\infty}^{+\infty} f(x,a,b)\mathrm{d}x = 1, \quad 0 \leq f(x,a,b)$$

gewährleistet sein. Hinsichtlich der ersten bekommt man die Bedingung

$$1 = \int_{-1}^{+1} x^2 + bx + a\mathrm{d}x = \frac{x^3}{3} + b\frac{x^2}{2} + ax\Big|_{-1}^{+1} = \frac{2}{3} + 2a$$

d.h. für jedes b muß man $a = 1/6$ wählen.

Für b bleibt die Bedingung

$$p(x) := x^2 + bx + \frac{1}{6} \geq 0$$

für alle $x \in [-1, 1]$. Das Polynom p nimmt sein Minimum in

$$-b/2$$

an, wobei

$$p(-b/2) = -b^2/4 + 1/6.$$

Da p in $[-1, 1]$ nichtnegativ sein soll, verlangen wir $p(-b/2) = 0$, d.h.

$$b = \pm\sqrt{2/3}.$$

Daraus folgt, daß $p(-b/2)$ in

$$b \in [-\sqrt{2/3}, +\sqrt{2/3}]$$

nichtnegativ ist und außerhalb von $[-\sqrt{2/3}, +\sqrt{2/3}]$ negativ. Da schließlich für $b \in [-\sqrt{2/3}, +\sqrt{2/3}]$ die Zahl $-b/2$ zu $[-1, +1]$ gehört, darf b nur in $[-\sqrt{2/3}, +\sqrt{2/3}]$ variieren. Zusammenfassend stellen wir fest: $f(x, a, b)$ kann als Dichtefunktion verwendet werden, falls $a = 1/6$ und $b \in [-\sqrt{2/3}, +\sqrt{2/3}]$. Abbildung 31 zeigt für verschiedene zulässige b die resultierende Dichtfunktion und die dazugehörigen Verteilungsfunktionen.

Sind das alle Möglichkeiten für b? Es stellt sich heraus, daß für $b \in \mathbb{R} \setminus [-\sqrt{2/3}, +\sqrt{2/3}]$ die Funktion $f(x, 1/6, b)$ innerhalb $[-1, 1]$ negative Werte annimmt.

Für den Erwartungswert ergibt sich

$$EX = \int_{-\infty}^{+\infty} xf(x, 1/6, b)\mathrm{d}x = \int_{-1}^{+1} x\left(x^2 + bx + \frac{1}{6}\right)\mathrm{d}x = \frac{x^4}{4} + b\frac{x^3}{3} + \frac{1}{6}\frac{x^2}{2}\Big|_{-1}^{+1} = \frac{2}{3}b.$$

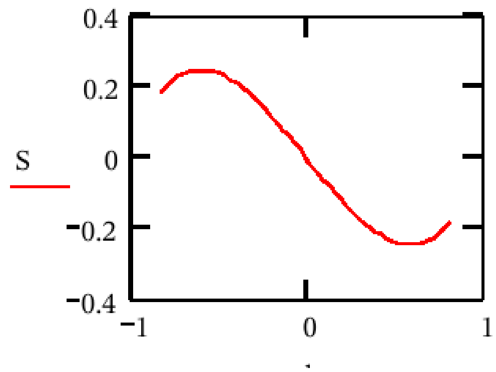


Abbildung 39: Es wird der Verlauf von $m_3(2b/3)$ für $b \in [-\sqrt{2/3}, +\sqrt{2/3}]$ gezeigt.

Für die Varianz von X berechnen wir

$$EX^2 = \int_{-1}^{+1} x^2(x^2 + bx + 1/6)\mathrm{d}x = \frac{x^5}{5} + b\frac{x^4}{4} + \frac{1}{6}\frac{x^3}{3}\bigg|_{-1}^{+1} = \frac{2}{5} + \frac{1}{9} = \frac{23}{45}.$$

Daher gilt

$$\mathrm{Var}(X) = EX^2 - (EX)^2 = \frac{23}{45} - \frac{4}{9}b^2.$$

Offensichtlich wird $\mathrm{Var}(X)$ minimal für $b = \pm\sqrt{2/3}$.

Hinsichtlich Symmetrie und Schiefe kann man das dritte Moment bezüglich $(2/3)b$ berechnen:

$$m_3\left(\frac{2}{3}b\right) = \int_{-1}^{1}\left(x - \frac{2}{3}b\right)^3\left(x^2 + bx + \frac{1}{6}\right)\mathrm{d}x = \frac{16}{27}b^3 - \frac{28}{45}b.$$

Abbildung 39 zeigt den Werteverlauf von $m_3(2b/3)$ für $b \in [-\sqrt{2/3}, +\sqrt{2/3}]$. Offensichtlich wird für positives b das dritte Moment bzgl. $\mu := 2b/3$ negativ und für negatives b positiv, sodaß positives b rechtsschiefes f und negatives b linksschiefes f nachsichzieht. Hinsichtlich der Symmetrie haben wir zu überprüfen, für welche b die Funktionalgleichung

$$f\left(\frac{2}{3}b + x, \frac{1}{6}, b\right) = f\left(\frac{2}{3}b - x, \frac{1}{6}, b\right)$$

erfüllt ist. Das ist genau dann der Fall, wenn $m_3\left(\frac{2}{3}b\right) = 0$, d.h. wenn $b = 0$.

Um Schiefheitsaussagen zu überprüfen visualisieren wir f in $[-1, 1]$ für $b \in \left[-\sqrt{\frac{2}{3}}, +\sqrt{\frac{2}{3}}\right]$ (Abbildung 40).

Hinsichtlich der statistischen Masse links und recht von Null bekommt man

$$M_L := \int_{-\infty}^{0} f(t, 1/6, b)\mathrm{d}t = \int_{-1}^{0}(t^2 + bt + 1/6)\mathrm{d}t = 1/2 - b/2$$

$$M_R := 1 - M_L = 1/2 + b/2.$$

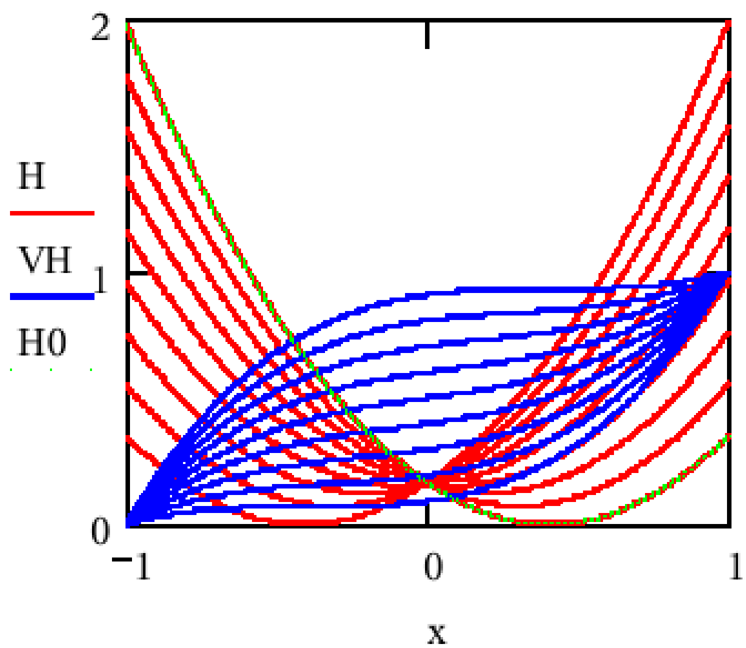


Abbildung 40: Es werden die Graphen (H, rot) der Dichtefunktion f über $[-1,1]$ für $b_k := -\sqrt{2/3} + \frac{2\sqrt{2/3}}{8}k$, $k = 0, \ldots, 8$, gezeigt. Die Parabel H0 gehört dabei zum Parameter $b_0 := -\sqrt{2/3}$. Sie besitzt offensichtlich in $-b_0 = \sqrt{2/3}$ ihr Minimum. Das Stichprobenmittel $\overline{\boldsymbol{x}}$ einer Stichprobe $\boldsymbol{x}$ aus einer entsprechend $F_{a,b}$ verteilten Grundgesamtheit liegt ungefähr bei bei $EX = 2b/3$. Für positives b überwiegt der Anteil der statistischen Masse rechts von Null seinen Anteil links von Null. Dieses Verhältnis wird durch die Eigenschaft 'rechtsschief' charakterisiert. Eine numerische Charakterisierung ist durch $m_3(\mu) < 0$ gegeben. Für $b = 0$ sind beide Anteile gleich, was Symmetrie von f bedeutet. Für negatives b überwiegt schließlich der Anteil der statistischen Masse links den Anteil rechts. Dieses Verhältnis wird durch die Eigenschaft 'linksschief' charakterisiert. Eine numerische Charakterisierung ist durch $m_3(\mu) > 0$ gegeben. Die blauen Kurven VH stellen die zu den Dichtefunktionen $f(., 1/6, b_k)$, $k = 0, \ldots, 8$, gehörenden Verteilungsfunktionen dar.

Demnach gilt für negatives b die Ungleichung $M_R < M_L$ (linksschiefes $f(x, 1/6, b)$) und für positives b die Ungleichung $M_L < M_R$ (rechtssschiefes $f(x, 1/6, b)$).

8. Aufgabe: Gegeben seien die Meßwertpaare

k	1	2	3	4	5	6	7	8
x_k	1.20	1.50	1.80	1.90	2.20	2.40	2.6	3.0
y_k	3.41	4.01	4.59	4.79	5.41	5.79	?	?

Es wird ein linearer Zusammenhang unterstellt. Welche Werte sind nach der Methode der kleinsten Quadrate ungefähr für y_7 und y_8 zu erwarten, wenn man $x_7 := 2.6$ und $x_8 := 3.0$ misst?

Antwort: Da mit der Methode der kleinsten Quadrate gearbeitet werden soll, hat man die Regressionsgerade

$$y(x) = ax + b$$

zu bestimmen und dann $y(x_7)$ und $y(x_8)$ als Schätzung zu benutzen. Aus

$$a := \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - n\overline{\boldsymbol{x}}^2}, \quad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 21.14,\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 53.268,\ \overline{\boldsymbol{x}} = 1.833,\ \overline{\boldsymbol{y}} = 4.667,\ n = 6$$

folgt $a = 1.988$ und aus

$$b = \overline{\boldsymbol{y}} - a\overline{\boldsymbol{x}}$$

bekommt man $b = 1.023$. Daher gilt

$$y_7 \approx ax_7 + b = 1.988 \times 2.6 + 1.023 = 6.191, \quad y_8 \approx ax_8 + b = 1.988 \times 3.0 + 1.023 = 6.986.$$

9. Aufgabe: Der Stichprobenvektor

$$\boldsymbol{x} := [2, -1, 3, 2, 1, 6, -2, 3]^{\mathrm{T}}$$

stammt aus einer Grundgesamtheit, die entsprechend $F(x) := \int_{-\infty}^{x} f(t)\mathrm{d}t$ verteilt ist.

- Bestimmen Sie die empirische Verteilungsfunktion F_n.
- Bestimmen Sie den Mittelwert $\overline{\boldsymbol{x}}$, den Zentralwert z und die Spannweite S von $\boldsymbol{x}$.
- Stellen Sie aufgrund vorliegender Stichprobe eine Vermutung über die Symmetrieeigenschaften der Dichtefunktion f an.

Antwort: Nach Definition gilt

$$F_n(x) := \frac{\text{Anzahl der Komponenten von } \boldsymbol{x}\text{, die kleiner oder gleich } x \text{ sind}}{n}.$$

Offenbar gilt $n = 8$ und der sortierte Vektor $\boldsymbol{x}$ lautet

$$\mathrm{sort}(\boldsymbol{x}) = [-2, -1, 1, 2, 2, 3, 3, 6]^{\mathrm{T}} =: [x_1, \ldots, x_8].$$

Daher gilt

$$F_8(x) := \frac{1}{8}\begin{cases} 0 & \text{falls} \quad x < -2 \\ 1 & \text{falls} \; -2 \le x < -1 \\ 2 & \text{falls} \; -1 \le x < 1 \\ 3 & \text{falls} \quad 1 \le x < 2 \\ 5 & \text{falls} \quad 2 \le x < 3 \\ 7 & \text{falls} \quad 3 \le x < 6 \\ 8 & \text{falls} \quad 6 \le x \end{cases}$$

Für die Lageparameter bekommen wir

$$\overline{\boldsymbol{x}} = \frac{7}{4}, \quad z = \frac{x_4 + x_5}{2} = 2, \quad S = \max(\boldsymbol{x}) - \min(\boldsymbol{x}) = 8.$$

Um etwas über die Symmetrieeigenschaften von f aussagen zu können, approximieren wir das dritte Moment $m_3(\mu)$ bezüglich μ entsprechend

$$m_3(\mu) \approx \frac{1}{8}\sum_{k=1}^{8}(x_k - \overline{\boldsymbol{x}})^3 = \frac{27}{32}.$$

Demnach wird $m_3(\mu)$ positiv sein, sodaß man es wahrscheinlich mit einem linksschiefen f zu tun hat.

10. Aufgabe: Testen Sie mit einer Irrtumswahrscheinlichkeit von 10 %, ob die $n := 100$ spaltenweise geordneten Meßwerte

0.637	$\underline{1.381}$	1.598	1.794	2.034	2.157	2.238	2.312	2.435	2.732
0.852	1.473	1.618	$\underline{1.807}$	2.044	2.173	2.248	2.331	2.437	2.737
1.041	1.475	1.631	1.877	$\underline{2.067}$	2.178	2.254	2.332	$\underline{2.448}$	2.739
1.089	1.488	1.650	1.894	2.072	2.193	2.259	2.342	2.465	$\underline{2.773}$
1.153	1.514	$\underline{1.653}$	1.914	2.075	$\underline{2.206}$	2.260	$\underline{2.358}$	2.521	2.880
$\underline{1.212}$	1.532	1.676	1.922	2.093	2.212	2.270	2.395	2.542	2.945
1.285	$\underline{1.545}$	$\underline{1.715}$	$\underline{1.924}$	2.098	2.218	2.275	2.396	$\underline{2.548}$	2.950
1.374	1.562	1.763	1.938	2.124	2.229	$\underline{2.278}$	2.405	2.582	2.979
1.378	1.566	1.773	1.957	$\underline{2.130}$	2.231	2.282	2.411	2.587	3.057
1.382	1.568	1.779	$\underline{1.974}$	2.146	2.232	2.308	2.431	$\underline{2.653}$	3.089

aus einer normalverteilten Grundgesamtheit stammen. Benutzen Sie dazu den χ^2-Anpassungstest unter Verwendung folgender Einteilung in $k := 18$ Klassen:

$$[A_1, \ldots, A_{19}] :=$$

$$[0.50, 1.25, 1.44, 1.56, 1.66, 1.75, 1.83, 1.93, 2.00, 2.07$$
$$2.14, 2.21, 2.28, 2.36, 2.45, 2.55, 2.67, 2.85, 3.40].$$

d.h.

$$K_1 := [0.50, 1.25), \; K_2 := [1.25, 1.44), \ldots, K_{18} := [2.85, 3.40).$$

Außerdem gilt

$$\boldsymbol{\overline{x}} = 2.058, \quad s = 0.499.$$

Antwort: Man hat zuerst die absoluten Häufigkeiten h_r, $r = 1, \ldots, 18$ zu bestimmen, d.h. es ist auszuzählen, wieviele Meßwerte in jeder Klasse liegen:

$$[h_1, \ldots, h_{18}] = [6, 5, 6, 8, 2, 5, 5, 3, 3, 6, 6, 13, 7, 8, 4, 3, 4, 6].$$

Nun sind die Zahlen

$$p_r := \Phi_{0,1}\left(\frac{A_{r+1} - \boldsymbol{\overline{x}}}{s}\right) - \Phi_{0,1}\left(\frac{A_r - \boldsymbol{\overline{x}}}{s}\right), \quad r = 1, \ldots, 18$$

zu berechnen. Man erhält

$$100 \times [p_1, \ldots, p_{18}] =$$

$$[5.17, 5.50, 5.14, 5.34, 5.60, 5.53, 7.49, 5.50, 5.59,$$

$$5.57, 5.44, 5.22, 5.57, 5.65, 5.40, 5.21, 5.38, 5.26]$$

d.h. die Forderung

$$np_r \geq 5$$

ist erfüllt. Schließlich hat man die Testgröße

$$c^2 := \sum_{r=1}^{18} \frac{(h_r - np_r)^2}{np_r}$$

zu berechnen:

$$c^2 = 21.987$$

und mit der Lösung der Gleichung

$$F_m(x) = 1 - \alpha, \quad \alpha := 0.10$$

zu vergleichen, wobei F_m die χ^2-Verteilungsfunktion mit m Freiheitsgraden darstellt und

$$m := k - 3 = 15$$

zu wählen ist. Blick in Tabelle (48) in Zeile 15 und Spalte 3 liefert

$$F_{15}^{-1}(1 - 0.1) = F_{15}^{-1}(0.9) = 22.31.$$

Da schließlich

$$c^2 = 21.987 < 22.31 = F_{15}^{-1}(1 - \alpha)$$

können die in Frage stehenden Meßwerte mit einer Irrtumswahrscheinlichkeit von 10 % als normalverteilt angesehen werden.

Aufgabe 10a: Testen Sie mit einer Irrtumswahrscheinlichkeit von 1 %, ob die $n := 30$ Meßwerte

$$\begin{array}{rrrrrrrrrr} -7.58 & -6.46 & -5.14 & -4.21 & -4.17 & -4.14 & -3.26 & -3.24 & -2.31 & -1.46 \\ -0.05 & 0.07 & 1.24 & 1.38 & 2.93 & 2.98 & 3.30 & 3.93 & 4.06 & 4.12 \\ 4.29 & 4.32 & 4.38 & 4.60 & 6.11 & 6.48 & 6.65 & 8.53 & 9.32 & 12.57 \end{array}$$

für die

$$\overline{\boldsymbol{x}} = 1.64, \quad s = 5.00$$

gilt, einer normalverteilten Grundgesamtheit entstammen. Benutzen Sie dazu den χ^2-Anpassungstest unter Verwendung der Klasseneinteilung

$$K_1 := [-8, -2.5), \quad K_2 := [-2.5, 0), \quad K_3 := [0, 2.4), \quad K_4 = [2.4, 5), \quad K_5 := [5, 13).$$

Es zeigt sich, daß der Test positiv ausfällt. Konstruieren Sie daher unter Verwendung von $\overline{\boldsymbol{x}}$ und s ein Vertrauensintervall für μ zum Vertrauensniveau 99 % (Alg. zur Berechnung eines Vertrauensintervalls für μ von normalverteilter Zg bei unbekannter Standardabweichung, t-Verteilung).

Antwort: Man hat zuerst die absoluten Häufigkeiten h_r, $r = 1, \ldots, k := 5$, zu bestimmen, d.h. es ist auszuzählen, wieviele Meßwerte in jeder Klasse liegen:

$$[h_1, \ldots, h_5] = [8, 3, 3, 10, 6].$$

Nun sind die Zahlen

$$\phi_r := \Phi_{0,1}\left(\frac{A_r - \overline{\boldsymbol{x}}}{s}\right), \quad r = 1, \ldots, 6$$

zu berechnen, wobei

$$A := [-8, -2.5, 0, 2.4, 5, 13]^{\mathrm{T}}.$$

Wegen

$$\frac{A_1 - \overline{\boldsymbol{x}}}{s} = \frac{-9.64}{5} = -1.928 \quad \frac{A_2 - \overline{\boldsymbol{x}}}{s} = \frac{-4.14}{5} = -0.828 \quad \frac{A_3 - \overline{\boldsymbol{x}}}{s} = \frac{-1.64}{5} = -0.328$$

$$\frac{A_4 - \overline{\boldsymbol{x}}}{s} = \frac{0.76}{5} = 0.152 \qquad \frac{A_5 - \overline{\boldsymbol{x}}}{s} = \frac{3.36}{5} = 0.672 \qquad \frac{A_6 - \overline{\boldsymbol{x}}}{s} = \frac{11.36}{5} = 2.272$$

erhält man

$$\begin{aligned} \phi_1 &= \Phi_{0,1}(-1.928) = 1 - \Phi_{0,1}(1.928) = 1 - 0.9732 = 0.027 \\ \phi_2 &= \Phi_{0,1}(-0.828) = 1 - \Phi_{0,1}(0.828) = 1 - 0.7967 = 0.203 \\ \phi_3 &= \Phi_{0,1}(-0.328) = 1 - \Phi_{0,1}(0.328) = 1 - 0.6293 = 0.371 \\ \phi_4 &= \Phi_{0,1}(0.152) = 0.560 \\ \phi_5 &= \Phi_{0,1}(0.672) = 0.749 \\ \phi_6 &= \Phi_{0,1}(2.272) = 0.988 \end{aligned}$$

sodaß

$p_i := \phi_{i+1} - \phi_i$	theoretische Besetzungszahl
$p_1 = 0.177$	$np_1 := n(\phi_2 - \phi_1) = 5.31$
$p_2 = 0.168$	$np_2 := n(\phi_3 - \phi_2) = 5.02$
$p_3 = 0.189$	$np_3 := n(\phi_4 - \phi_3) = 5.67$
$p_4 = 0.189$	$np_4 := n(\phi_5 - \phi_4) = 5.66$
$p_5 = 0.239$	$np_5 := n(\phi_6 - \phi_5) = 7.18$

d.h. die Forderung

$$np_r \geq 5, \quad r = 1, \ldots, 5$$

ist erfüllt. Schließlich hat man die Testgröße

$$c^2 := \sum_{r=1}^{5} \frac{(h_r - np_r)^2}{np_r} = 6.976$$

zu berechnen und mit der Lösung der Gleichung

$$F_m(x) = 1 - \alpha, \quad \alpha := 0.01$$

zu vergleichen, wobei F_m die χ^2-Verteilungsfunktion mit m Freiheitsgraden darstellt und

$$m := k - 3 = 2$$

zu wählen ist. Blick in Tabelle (48) in Zeile 2 und Spalte 1 liefert

$$F_2^{-1}(1 - 0.01) = F_2^{-1}(0.99) = 9.21.$$

Da schließlich

$$c^2 < F_2^{-1}(1 - \alpha)$$

können die in Frage stehenden Meßswerte mit einer Irrtumswahrscheinlichkeit von 1 % als normalverteilt angesehen werden.

Hinsichtlich des Vertrauensintervalls für μ hat man

$$c := \frac{sz}{\sqrt{n}} = \frac{5 \times z}{\sqrt{30}}$$

zu berechnen, wobei sich z aus der Gleichung

$$F_{n-1}(z) = 1 - \frac{\alpha}{2} = 1 - \frac{0.01}{2} = 0.995$$

ergibt und F_n die t-Verteilung mit n-Freiheitsgraden darstellt. Blick in Zeile 29 und Spalte 5 von Tabelle (79) liefert

$$z = 2.756$$

sodaß

$$c = \frac{5 \times 2.756}{\sqrt{30}} = 2.516.$$

Demnach lautet das gesuchte Vertrauensintervall

$$[\overline{\boldsymbol{x}} - c, \overline{\boldsymbol{x}} + c] = [1.64 - 2.516, 1.64 + 2.516] = [-0.876, 4.156].$$

11. Aufgabe: Für welche $c \in [0, 1]$ wird die Dichtefunktion

$$f(x,c) := \begin{cases} 0 & \text{falls} \qquad x < -1 \\ c & \text{falls } -1 \leq x < 0 \\ 1-c & \text{falls} \quad 0 \leq x < 1 \\ 0 & \text{falls} \quad 1 \leq x \end{cases}$$

symmetrisch?

Antwort: Offensichtlich wird $f(x,c)$ für $c = 1/2$ symmetrisch. Aber auch für $c = 0$ und $c = 1$ stellt sich Symmetrie ein. Zur Überprüfung kann das dritte Moment

$$m_3(\mu) := \int_{-\infty}^{+\infty} (x-\mu)^3 f(x,c)\mathrm{d}x, \quad \mu := \frac{1}{2} - c = \int_{-\infty}^{+\infty} x f(x,c)\mathrm{d}x$$

bezüglich des Erwartungswerts von $f(x,c)$ benutzt werden, welches in Abhängigkeit von c mit den Werten des Polynoms

$$p(c) := -2c^3 + 3c^2 - c = -2c(c-0.5)(c-1)$$

übereinstimmt.

12. Aufgabe: Ein Briefeschreiber verfaßt 5 Briefe verschiedenen Inhalts, addressiert die zugehörigen Umschläge und überläßt es dem Zufall, welcher Brief in welchen Umschlag gelangt. Wie groß ist die Wahrscheinlichkeit, daß

- genau ein Brief
- genau zwei Briefe
- genau drei Briefe
- wenigstens zwei Briefe

den richtigen Adressaten erreichen?

Antwort: Es gibt $5! = 120$ Möglichkeiten die 5 Briefe den 5 Umschlägen zu zuordnen. Die Anzahl der Möglichkeiten, daß genau ein Brief in den richtigen Umschlag kommt, ergibt sich aus dem Produkt der 5 Möglichkeiten einen Brief auszuwählen, der in den richtigen Umschlag kommt, und der Anzahl der Möglichkeiten, die restlichen vier Briefe alle in falsche Umschläge zu verteilen. Diese Anzahl stimmt mit 9 überein, da es genau 9 Permutationen im Umfang von 4 gibt, bei denen kein Element auf seinem Platz bleibt:

$$\begin{array}{cccccc} 1234 & 1243 & 1324 & 1342 & 1423 & 1432 \\ 2134 & \boxed{2143} & 2314 & \boxed{2341} & \boxed{2413} & 2431 \\ 3124 & \boxed{3142} & 3214 & 3241 & \boxed{3412} & \boxed{3421} \\ \boxed{4123} & 4132 & 4213 & 4231 & \boxed{4312} & \boxed{4321} \end{array}$$

Daher gibt es 45 Möglichkeiten einen Brief richtig und die restlichen vier falsch zu verteilen. Demnach lautet die erste gefragte Wahrscheinlichkeit $\frac{45}{120}$.

Die Anzahl der Möglichkeiten, genau zwei Briefe richtig zu verteilen und die restlichen drei falsch, ist gleich dem Produkt aus der Anzahl zwei aus fünf Briefen auszuwählen und der Anzahl der Permutationen vom Umfang 3, bei denen kein Element auf seiner Position bleibt:

$$123 \quad 132 \quad 213 \quad \boxed{231} \quad \boxed{312} \quad 321$$

Daher hat man

$$\binom{5}{2} \times 2 = \frac{5!}{2!3!} \times 2 = 20$$

Möglichkeiten genau zwei Briefe in richtige Umschläge zu stecken, sodaß die zweite gefragte Wahrscheinlichkeit $\frac{1}{6}$ beträgt.

Die Anzahl der Möglichkeiten, genau drei Briefe richtig zu verteilen und die restlichen zwei falsch, ist gleich dem Produkt aus der Anzahl drei aus fünf Briefen auszuwählen und der Anzahl der Permutationen vom Umfang 2, bei denen kein Element auf seiner Position bleibt:

$$12 \quad \boxed{21}$$

Daher hat man

$$\binom{5}{3} = \frac{5!}{2!3!} = 10$$

Möglichkeiten genau drei Briefe in richtige Umschläge zu stecken, sodaß die dritte gefragte Wahrscheinlichkeit $\frac{1}{12}$ beträgt.

Die Wahrscheinlichkeit, daß wenigstens zwei Briefe den richtigen Adressaten erreichen ist gleich der Summe der Wahrscheinlichkeiten, daß genau zwei, drei, vier und fünf Briefe richtig ankommen. Die Wahrscheinlichkeit, daß genau vier Briefe richtig ankommen ist Null, da diese Ereignis nicht eintreten kann. Die Wahrscheinlichkeit, daß alle fünf Briefe richtig ankommen, beträgt $1/120$. Daher lautet die in Frage stehende Wahrscheinlichkeit

$$\frac{1}{6} + \frac{1}{12} + \frac{1}{120} = \frac{20 + 10 + 1}{120} = \frac{31}{120}.$$

Wir haben die Aufgabe im wesentlichen durch Auszählen gelöst, wobei wir die Permutationen gezählt haben, bei denen kein Element an seinem Platz bleibt. Diese Anzahl ist bei Permutionen vom Umfang n durch

$$f(n) := n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

gegeben. So gilt tatsächlich

$$f(1) = 0, \quad f(2) = 1, \quad f(3) = 2, \quad f(4) = 9$$

in Übereinstimmung mit unseren Zählergebnissen von oben. ([4] S.92)

Geht man zu Wahrscheinlichkeiten über, so liefert der Quotient

$$\frac{f(n)}{n!}$$

die Wahrscheinlichkeit dafür, daß bei der Verteilung von n Briefen keiner in den richtigen Umschlag kommt. Dieser Quotient stimmt mit der Summe

$$\sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

überein, die offensichtlich den Anfangsabschnitt der Taylorreihe von $1/e$ darstellt. Daher kann man für große n die Wahrscheinlichkeit, daß kein Brief im richtigen Umschlag steckt, durch $1/e$ approximieren. Für unendlich viele Briefe ist dieser Wert exakt. Schließlich bekommt man für die Wahrscheinlichkeit, daß bei der Verteilung von unendlich vielen Briefen in unendlich viele Umschläge wenigstens ein Brief in den richtigen Umschlag kommt, den Wert

$$1 - \frac{1}{e} \approx 0.632.$$

13. Aufgabe: Eine Zufallsgröße X sei normalverteilt.

1. Berechnen Sie für

   $$a := 8.7, \quad b := 9.2, \quad \mu := 9, \quad \sigma := 0.5$$

   die Wahrscheinlichkeit, daß X Werte zwischen a und b annimmt.

2. Gegeben sind

   $$a := 11.8, \quad \mu := 12, \quad \sigma := 0.5.$$

   Wie muß man die Konstante b wählen, damit die Wahrscheinlichkeit, daß X Werte zwischen a und b annimmt, gleich 0.4435 wird.

3. Gegeben sind

   $$a := 4, \quad b := +\infty, \quad \sigma := 3.$$

   Außerdem ist die Wahrscheinlichkeit dafür, daß X Werte größer als a annimmt, gleich 0.1587. Berechnen Sie den Erwartungswert μ.

4. Gegeben sind

   $$a := 0.6, \quad \mu := 13.5, \quad \sigma := 1.2.$$

   Berechnen Sie

   $$P(|X - \mu| \leq a).$$

5. Gegeben sind

   $$\mu := 0.6, \quad \sigma := 0.02, \quad P(|X - \mu| \leq a) = 0.9876.$$

   Berechnen Sie die Schranke a.

6. Gegeben sind

$$a := 0.8, \quad P(|X-\mu| \leq a) = 0.97.$$

Berechnen Sie σ.

Antwort:

1. Es gilt

$$\begin{aligned}
P(a \leq X \leq b) &= \Phi_{\mu,\sigma}(b) - \Phi_{\mu,\sigma}(a) \\
&= \Phi_{0,1}\left(\frac{b-\mu}{\sigma}\right) - \Phi_{0,1}\left(\frac{a-\mu}{\sigma}\right) = \Phi_{0,1}\left(\frac{9.2-9}{0.5}\right) - \Phi_{0,1}\left(\frac{8.7-9}{0.5}\right) \\
&= \Phi_{0,1}\left(\frac{2}{5}\right) - \Phi_{0,1}\left(-\frac{3}{5}\right) = 0.6554 - \left(1 - \Phi_{0,1}\left(\frac{3}{5}\right)\right) \\
&= 0.6554 - 1 + 0.7257 = 0.3811.
\end{aligned}$$

2. Es gilt

$$\begin{aligned}
0.4435 &= \Phi_{0,1}\left(\frac{b-12}{0.5}\right) - \Phi_{0,1}\left(\frac{11.8-12}{0.5}\right) \\
&= \Phi_{0,1}\left(\frac{b-12}{0.5}\right) - \Phi_{0,1}\left(-\frac{2}{5}\right) \\
&= \Phi_{0,1}\left(\frac{b-12}{0.5}\right) - \left(1 - \Phi_{0,1}\left(\frac{2}{5}\right)\right) \\
&= \Phi_{0,1}\left(\frac{b-12}{0.5}\right) - 1 + 0.6554
\end{aligned}$$

d.h. es gilt die Gleichung

$$\Phi_{0,1}\left(\frac{b-12}{0.5}\right) = 0.7881.$$

Da $\Phi_{0,1}$ streng monoton wachsend ist, folgt

$$\frac{b-12}{0.5} = 0.8.$$

Auflösung nach b liefert $b = 12.4$.

3. Nach Voraussetzung gilt

$$P(4 \leq X < +\infty) = 0.1587.$$

Demnach gilt

$$0.1587 = \Phi_{0,1}(+\infty) - \Phi_{0,1}\left(\frac{4-\mu}{3}\right)$$

d.h.
$$\Phi_{0,1}\left(\frac{4-\mu}{3}\right) = 1 - 0.1587 = 0.8413.$$
Da $\Phi_{0,1}$ streng monoton wachsend ist, folgt
$$\frac{4-\mu}{3} = 1$$
d.h. $\mu = 1$.

4. Die Ungleichung
$$|X-\mu| \leq a$$
bedeutet, daß der Abstand zwischen X und μ nicht größer sein soll als a, d.h. $X \in [\mu - a, \mu + a]$, d.h.
$$\mu - a \leq X \leq \mu + a.$$
Demnach gilt
$$\begin{aligned} P(|X-\mu| \leq a) = \;& P(\mu - a \leq X \leq \mu + a) = \Phi_{\mu,\sigma}(\mu + a) - \Phi_{\mu,\sigma}(\mu - a) \\ = \;& \Phi_{0,1}\left(\frac{a}{\sigma}\right) - \Phi_{0.1}\left(-\frac{a}{\sigma}\right) = \Phi_{0,1}\left(\frac{a}{\sigma}\right) - \left(1 - \Phi_{0.1}\left(\frac{a}{\sigma}\right)\right) \\ = \;& 2\Phi_{0,1}\left(\frac{a}{\sigma}\right) - 1 = 2\Phi_{0,1}\left(\frac{0.6}{1.2}\right) - 1 = 2 \times 0.6915 - 1 = 0.383. \end{aligned}$$
Bemerkung: Die Kenntnis von μ war nicht notwendig.

5. Nach Voraussetzung gilt
$$P(|X-\mu| \leq a) = 0.9876.$$
Entsprechend der vorhergehenden Aufgabe gilt
$$P(|X-\mu| \leq a) = 2\Phi_{0,1}\left(\frac{a}{\sigma}\right) - 1.$$
Kombination beider Gleichungen liefert
$$0.9876 = 2\Phi_{0,1}\left(\frac{a}{\sigma}\right) - 1$$
d.h.
$$\frac{1.9876}{2} = 0.9938 = \Phi_{0,1}\left(\frac{a}{\sigma}\right).$$
Aus der strengen Monotonie von $\Phi_{0,1}$ folgt $\frac{a}{\sigma} = 2.5$, woraus mit $\sigma = 0.02$ für a die Gleichung $a = 0.05$ folgt.

6. Nach Voraussetzung gilt
$$P(|X-\mu| \le a) = 0.97.$$
Entsprechend der vorhergehenden Aufgabe gilt
$$P(|X-\mu| \le a) = 2\Phi_{0,1}\left(\frac{a}{\sigma}\right) - 1.$$
Kombination beider Gleichungen liefert
$$0.97 = 2\Phi_{0,1}\left(\frac{a}{\sigma}\right) - 1$$
d.h.
$$\frac{1.97}{2} = 0.985 = \Phi_{0,1}\left(\frac{a}{\sigma}\right).$$
Aus der strengen Monotonie von $\Phi_{0,1}$ erhält man
$$\frac{a}{\sigma} = 2.175$$
woraus zusammen mit $a = 0.8$ die Gleichung $\sigma = 0.3687$ folgt.

14. Aufgabe: Es wird mit zwei Würfeln (Hexaedern) gewürfelt. Die Zufallsgröße X sei durch das Produkt der oben aufliegenden Augenzahlen erklärt. Bestimmen Sie für X

- den Wertevorrat $X(\Omega)$
- die Mengen A_k
- die Wahrscheinlichkeitsfunktion $f(x)$
- den Erwartungswert EX
- die Varianz $\mathrm{Var}(X)$
- die Verteilungsfunktion $F(x)$.

Berechnen Sie die Wahrscheinlichkeit $P(2.3 \le X \le 16.7)$ mit Hilfe der Verteilungsfunktion.

Antwort: Der Ereignisraum kann durch
$$\Omega := \{(i,j) : i,j = 1,\dots,6\}$$
modelliert werden. Der Wertverlauf von X auf Ω ist dann durch

X	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

gegeben, d.h. der Wertevorrat von X lautet

$$X(\Omega) = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36\}$$

was in

$$\begin{array}{lll}
A_1 = \{(1,1)\} & A_2 = \{(1,2),(2,1)\} & A_3 = \{(1,3),(3,1)\} \\
A_4 = \{(1,4),(2,2),(4,1)\} & A_5 = \{(1,5),(5,1)\} & A_6 = \{(1,6),(2,3),(3,2),(6,1)\} \\
A_8 = \{(2,4),(4,2)\} & A_9 = \{(3,3)\} & A_{10} = \{(2,5),(5,2)\} \\
A_{12} = \{(2,6),(3,4),(4,3),(6,2)\} & A_{15} = \{(3,5),(5,3)\} & A_{16} = \{(4,4)\} \\
A_{18} = \{(3,6),(6,3)\} & A_{20} = \{(4,5),(5,4)\} & A_{24} = \{(4,6),(6,4)\} \\
A_{25} = \{(5,5)\} & A_{30} = \{(5,6),(6,5)\} & A_{36} = \{(6,6)\}
\end{array}$$

resultiert. Da das Eintreten eines jeden Elementarereigniss gleichwahrscheinlich ist, setzen wir $P((i,j)) := \frac{1}{36}$. Daraus folgt für die Wahrscheinlichkeitsfunktion f auf Ω der Werteverlauf

$$\begin{array}{rrr}
f(1) := P(A_1) = \dfrac{1}{36} & f(2) := P(A_2) = \dfrac{2}{36} & f(3) := P(A_3) = \dfrac{2}{36} \\
f(4) := P(A_4) = \dfrac{3}{36} & f(5) := P(A_5) = \dfrac{2}{36} & f(6) := P(A_6) = \dfrac{4}{36} \\
f(8) := P(A_8) = \dfrac{2}{36} & f(9) := P(A_9) = \dfrac{1}{36} & f(10) := P(A_{10}) = \dfrac{2}{36} \\
f(12) := P(A_{12}) = \dfrac{4}{36} & f(15) := P(A_{15}) = \dfrac{2}{36} & f(16) := P(A_{16}) = \dfrac{1}{36} \\
f(18) := P(A_{18}) = \dfrac{2}{36} & f(20) := P(A_{20}) = \dfrac{2}{36} & f(24) := P(A_{24}) = \dfrac{2}{36} \\
f(25) := P(A_{25}) = \dfrac{1}{36} & f(30) := P(A_{30}) = \dfrac{2}{36} & f(36) := P(A_{36}) = \dfrac{1}{36}
\end{array}$$

Für EX folgt daraus

$$EX =$$

$$\frac{1}{36}(1+4+6+12+10+24+16+9+20+48+30+16+36+40+48+25+60+36)$$

d.h.

$$EX = \frac{441}{36} = \frac{49}{4} = 12.25.$$

Diese Rechnung kann man sich vereinfachen, wenn man bedenkt, daß das Werfen von Würfel 1 das Werfen von Würfel 2 nicht beeinflußt. Definiert man X_1 durch die obenaufliegende Augenanzahl von Würfel 1 und X_2 durch die obenaufliegende Augenanzahl von Würfel 2, so gilt

$$X = X_1 X_2$$

und wegen der Unabhängigkeit von X_1 und X_2 die Gleichung

$$EX = EX_1 EX_2.$$

Wegen $EX_1 = EX_2 = \frac{7}{2}$ ergibt sich sofort $EX = \frac{49}{4}$.

Zur Berechnung von $\mathrm{Var}(X)$ benutzen wir die Gleichung

$$\mathrm{Var}(X) = EX^2 - (EX)^2.$$

Für EX^2 bekommt man

$$EX^2 = \frac{1}{36}\left(\begin{array}{l} (1\times 1^2)+(2\times 2^2)+(2\times 3^2)+(3\times 4^2)+(2\times 5^2)+(4\times 6^2)+ \\ (2\times 8^2)+(1\times 9^2)+(2\times 10^2)+(4\times 12^2)+(2\times 15^2)+(1\times 16^2)+ \\ (2\times 18^2)+(2\times 20^2)+(2\times 24^2)+(1\times 25^2)+(2\times 30^2)+(1\times 36^2) \end{array}\right)$$

$$= \frac{8281}{36}$$

sodaß

$$\mathrm{Var}(X) = EX^2 - (EX)^2 = \frac{8281}{36} - \left(\frac{49}{4}\right)^2 = \frac{11515}{144} = 79.965.$$

Bei der Berechnung von EX^2 kann auch wieder die Unabhängigkeit benutzt werden. Da unsere beiden Zg nur positive Werte annehmen können, wird die Unabhängigkeit durch Quadrieren nicht gestört, sodaß

$$EX^2 = E(X_1^2 X_2^2) = EX_1^2 EX_2^2.$$

Nach Definition gilt

$$EX_1^2 = \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6} = EX_2^2$$

sodaß

$$EX^2 = \left(\frac{91}{6}\right)^2 = \frac{8291}{36}.$$

Für die Verteilungsfunktion bekommt man schließlich

$$F(x) := \frac{1}{36}\begin{cases} 0, & \text{falls } x < 1 \\ 1, & \text{falls } 1 \le x < 2 \\ 3, & \text{falls } 2 \le x < 3 \\ 5, & \text{falls } 3 \le x < 4 \\ 8, & \text{falls } 4 \le x < 5 \\ 10, & \text{falls } 5 \le x < 6 \\ 14, & \text{falls } 6 \le x < 8 \\ 16, & \text{falls } 8 \le x < 9 \\ 17, & \text{falls } 9 \le x < 10 \\ 19, & \text{falls } 10 \le x < 12 \\ 23, & \text{falls } 12 \le x < 15 \\ 25, & \text{falls } 15 \le x < 16 \\ 26, & \text{falls } 16 \le x < 18 \\ 28, & \text{falls } 18 \le x < 20 \\ 30, & \text{falls } 20 \le x < 24 \\ 32, & \text{falls } 24 \le x < 25 \\ 33, & \text{falls } 25 \le x < 30 \\ 35, & \text{falls } 30 \le x < 36 \\ 36, & \text{falls } 36 \le x \end{cases}$$

sodaß

$$P(2.3 \le X \le 16.7) = F(16.7) - F(2.3) = \frac{26}{36} - \frac{3}{36} = \frac{23}{36}.$$

15. Aufgabe: Ein Verpackungsautomat füllt Zündhölzer in dafür geeignete Schachteln ab. Dabei stellt sich heraus, daß mit gewissen Wahrscheinlichkeiten p_k immer nur bestimmte Anzahlen x_k in den Schachteln vorkommen:

Anzahl x_k	56	58	60	61	62	63	64
Wahrscheinlichkeit p_k	0.02	0.11	0.42	0.16	0.13	0.12	0.04

Man berechne die Wahrscheinlichkeit dafür, daß sich in 5 beliebig ausgewählten Schachteln insgesamt zwischen einschließlich 295 und 305 Zündhölzer befinden. Es darf Unabhängigkeit unter den Stückzahlen in den einzelnen Streichholzschachteln vorausgesetzt werden. Mit Hilfe des Zentralen Grenzwertsatzes approximiere man das Ergebnis durch die Normalverteilung.

Antwort: Die Wahrscheinlichkeitsfunktion für die Stückzahl in einer Schachtel ergibt sich entsprechend

$$f(x) := \sum_{k=1}^{7} p_k \delta(x, x_k), \quad \delta(x,y) := \begin{cases} 1, & \text{falls } x = y, \\ 0, & \text{falls } x \neq y. \end{cases}$$

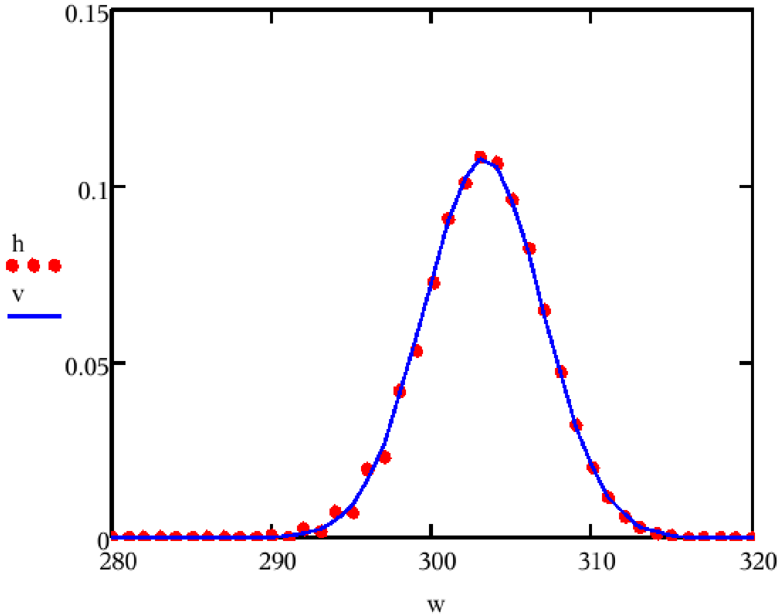


Abbildung 41: Die roten Punkte stellen die exakten Wahrscheinlichkeiten $f_4(k)$ dar und die blaue Kurve den Werterlauf von $\varphi_{5\mu,\sqrt{5}\sigma}$.

Da Unabhängigkeit vorliegt, liefert rekursive Faltung die Wahrscheinlichkeitsfunktion f_4 für die Summe X der Zündhölzeranzahlen in 5 beliebig ausgewählten Schachteln;

$$f_i(x) := \sum_{k=56}^{64} f_{i-1}(x-k)f(k), \quad f_0(x) := f(x), \quad i = 1, \ldots, 4.$$

Die exakte Lösung lautet

$$P(295 \leq X \leq 305) = \sum_{k=295}^{305} f_4(k) = 0.7198.$$

Entsprechend des zentralen Grenzwertsatzes erhält man als Näherungslösung

$$P(295 \leq X \leq 305) \approx \sum_{k=295}^{305} \varphi_{5\mu,\sqrt{5}\sigma}(k) = 0.7238$$

wobei

$$\mu := \sum_{k=1}^{7} x_k p_k = 60.64, \quad \sigma := \sqrt{\sum_{k=1}^{7} x_k^2 p_k - \mu^2} = \sqrt{2.75}.$$

Offenbar handelt es sich um eine gute Approximation des exakten Ergebnisses. Die roten Punkte in Abbildung 41 zeigen die Werte von $f_4(k)$ für $k = 280, \ldots, 320$ und die blaue Kurve den Werteverlauf von $\varphi_{5\mu,\sqrt{5}\sigma}(x)$ in diesem Bereich.

16. Aufgabe: Es werden 20 Personen nach ihrem Alter und Körpergewicht befragt. Das Befragungsresultat ist in folgender Tabelle zusammengefaßt:

Befragter	1	2	3	4	5	6	7	8	9	10
Alter	27	27	28	29	30	30	32	33	35	38
Gewicht	80	70	72	105	95	80	55	60	93	113
Befragter	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Alter	38	39	43	45	45	51	53	54	54	54
Gewicht	56	78	95	94	78	62	63	79	82	83

Man bilde die Kreuztabelle für die Einteilung

$$\begin{array}{rlll} \text{Altersklassen} & [0,30), & [30,45), & [45,60), \\ \text{Gewichtsklassen} & [0,70), & [70,90), & [90,120). \end{array}$$

Man bestimme die Spannweite, den Zentralwert und den häufigsten Wert für die Zahlenreihe, die sich aus den Gewichten ergibt.

Antwort: Man hat auszuzählen, wieviele der befragten Personen in jeder Kombination aus einer Alters- und einer Gewichtsklasse liegen:

	0 – 29	30 – 44	45 – 60
0 – 69	0	3	2
70 – 89	3	2	4
90 – 120	1	4	1

Die Summe der Einträge muß 20 betragen.

Wir sortieren die Gewichte in aufsteigender Reihenfolge in einen Gewichtsvektor

$$g := [55,\ 56,\ 60,\ 62,\ 63,\ 70,\ 72,\ 78,\ 78,\ 79,\ 80,\ 80,\ 82,\ 83,\ 93,\ 94,\ 95,\ 95,\ 105,\ 113]^{\mathrm{T}}$$

- Die Spannweite S beträgt $\max(g) - \min(g) = 113 - 55 = 58$.
- Der Zentralwert beträgt
$$\frac{g_{10} + g_{11}}{2} = 79.5.$$
- Die häufigsten Werte sind 78, 80 und 95.

17. Aufgabe: Diese Aufgabe zielt auf die Konzentration eines bestimmten Merkmals innerhalb einer Individuengruppe ab. Einzelheiten dazu finden Sie im Skript auf den Seiten 171 bis 174, die Sie auf folgende Daten anwenden sollen.

Der Markt für eine Produktgruppe wird von 7 Herstellerfirmen A bis G dominiert. Dabei wurden im Jahr 2010 folgende Umsätze in Mio EUR bezüglich dieser Produktgruppe erreicht:

Hersteller	A	B	C	D	E	F	G	sonstige Hersteller
Umsatz	250	500	350	750	200	150	180	120

Skizzieren Sie die resultierende Lorenzkurve und berechnen Sie den Gini-Koeffizienten.

Antwort: Wir sortieren den Umsatzvektor in aufsteigender Reihenfolge

$$\boldsymbol{x} := [120, 150, 180, 200, 250, 350, 500, 750]^{\mathrm{T}}$$

Offenbar besteht er aus 8 Komponenten, sodaß wir

$$u_0 := 0,\ u_1 := \frac{1}{8},\ u_1 := \frac{2}{8}, \ldots, u_8 = 1$$

erhalten. Mit der Definition

$$v_0 := 0, \quad v_k := \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{8} x_i}, \quad k = 1, \ldots 8$$

bekommt man

$$v_1 = 0.048,\ v_2 = 0.108,\ v_3 = 0.18,\ v_4 = 0.26,\ v_5 = 0.36,\ v_6 = 0.5,\ v_7 = 0.7,\ v_8 = 1.$$

Der Polygonzug, der die Punkte $(u_0, v_0), \ldots, (u_8, v_8)$ der Reihe nach miteinander verbindet, liefert die gewünschte Lorenzkurve L. Der Ginikoeffizient ist entsprechend S.174 im Skript durch

$$1 - \frac{1}{8} \sum_{k=1}^{8} (v_{k-1} + v_k)$$

gegeben. Die Auswertung führt auf 0.336, sodaß man noch keine extreme Konzentration vorliegen hat.

18. Aufgabe: Ein Schütze schießt 20 mal auf eine Zielscheibe. Welche Trefferwahrscheinlichkeit p muß er besitzen, damit er mit 10 %-ger Wahrscheinlichkeit genau 4 mal trifft?

Antwort: Die gesuchte Trefferwahrscheinlichkeit p ergibt sich aus der Gleichung

$$0.1 = \binom{20}{4} p^4 (1-p)^{16}.$$

Demnach interessiert man sich für die Nullstellen des Polynoms

$$f(p) := \binom{20}{4} p^4 (1-p)^{16} - 0.1$$

die im Intervall $[0, 1]$ liegen. Abbildung 42 zeigt den Werteverlauf. Offenbar gibt es zwei Nullstellen, die eine liegt in der Nähe von 0.1 und die andere bei 0.3. Mit dem Newtonverfahren

$$p_{k+1} := p_k - \frac{f(p_k)}{f'(p_k)}$$

und den Startwerten $p_0 := 0.1$ und $p_0 := 0.3$ bekommt man exakter

$$p = 0.1051 \quad \text{oder} \quad p = 0.3253.$$

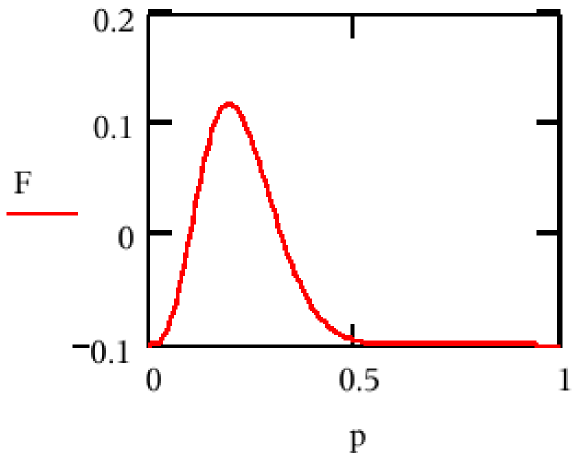


Abbildung 42: Werteverlauf von $f(p)$ aus Aufgabe 18. Offensichtlich existieren zwei Nullstellen, d.h. die gesuchte Trefferwahrscheinlichkeit ist nicht eindeutig bestimmt.

Daher ist die Fragestellung nicht eindeutig zu beantworten. Sowohl eine Trefferwahrscheinlichkeit von 0.1051 als auch von 0.3253 bewirkt, daß mit 10 % Wahrscheinlichkeit bei 20 maligem Schießen genau 4 mal getroffen wird.

Ergänzung: Um das Problem eindeutig lösbar zu machen, kann man fragen, wie groß die Treffferwahrscheinlichkeit sein muß, wenn mit 10 % Wahrscheinlichkeit wenigstens 4 mal getroffen werden soll bei 20 Schießversuchen.

Unser Polynom $f(p)$ geht dann über in

$$f(p) = 1 - \sum_{k=0}^{3} \binom{20}{k} p^k (1-p)^{20-k} - 0.1 = 0.9 - \sum_{k=0}^{3} \binom{20}{k} p^k (1-p)^{20-k}. \qquad (77)$$

Abbildung 43 versichert die Existenz einer Nullstelle in der Nähe von 0.1. Das Newtonverfahren liefert mit dem Startwert $p_0 := 0.1$ nach drei Iterationen den Wert 0.09021, der sich bei weitere Iteration nicht mehr ändert. Daher hat man auf die modifizierte Fragestellung mit $p := 0.09021$ zu antworten.

19. Aufgabe: Wieviele Möglichkeiten gibt es, die Zahl 22 in eine Summe von 5 Summanden zu zerlegen, wobei jeder Summand aus der Menge $\{1, \ldots, 6\}$ stammt? Berechnen Sie einmal das Ergebnis exakt und einmal näherungsweise mit Hilfe des zentralen Grenzwertsatzes.

Antwort: Die gesuchte Anzahl ist gleich der Wahrscheinlichkeit in der Summe eine 22 zu würfeln, wenn man mit fünf unabhängigen Hexaedern würfelt, multipliziert mit 6^5. Setzen wir

$$f(x) := \begin{cases} \frac{1}{6} & \text{falls } x \in \{1, \ldots, 6\} \\ 0 & \text{falls } x \notin \{1, \ldots, 6\} \end{cases}$$

so hat man f viermal mit sich selbst zu falten und den Funktionswert des Faltungsresul-

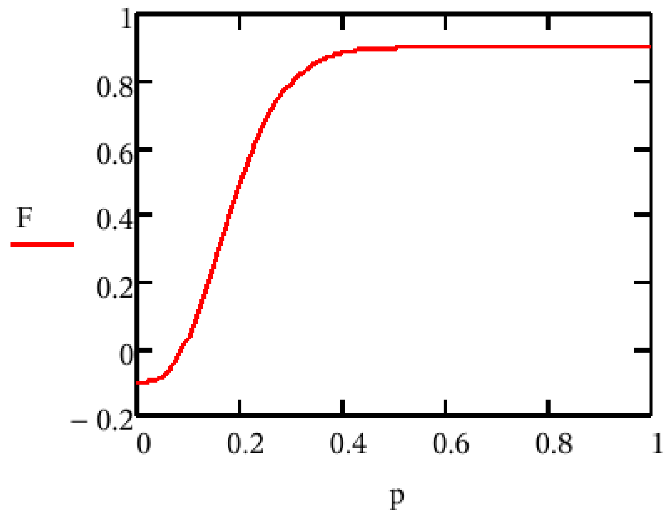


Abbildung 43: Werteverlauf von $f(p)$ definiert wie in (77). Offensichtlich existiert in $[0, 1]$ genau eine Nullstelle, welche in der Nähe von 0.1 liegt.

tates an der Stelle 22 mit 6^5 zu multiplizieren. Wir setzen $h_0(x) := f(x)$ und

$$\begin{aligned} h_1(z) &:= \textstyle\sum_{k=1}^{6} h_0(z-k)f(k), \\ h_2(z) &:= \textstyle\sum_{k=1}^{6} h_1(z-k)f(k), \\ h_3(z) &:= \textstyle\sum_{k=1}^{6} h_2(z-k)f(k), \\ h_4(z) &:= \textstyle\sum_{k=1}^{6} h_3(z-k)f(k). \end{aligned}$$

Dann haben wir

$$6^5 h_4(22)$$

zu berechnen, was in

$$420$$

resultiert. Demnach gibt es 420 Möglichkeiten die Zahl 22 in eine Summe von fünf Summanden zu zerlegen, wobei jeder Summand die Werte $1, 2, 3, 4, 5, 6$ annehmen kann. Abbildung 44 zeigt die Approximation des Werteverlaufs von $h_4(k)$ durch die den Werteverlauf der Dichtefunktion

$$\varphi_{\mu,\sigma}(k)$$

für die Parameter

$$\mu := EX_1 + \ldots + EX_5 = 5 \times 3.5 = 17.5$$

$$\sigma := \sqrt{\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_5)} = \sqrt{5 \times \tfrac{35}{12}} = 3.819.$$

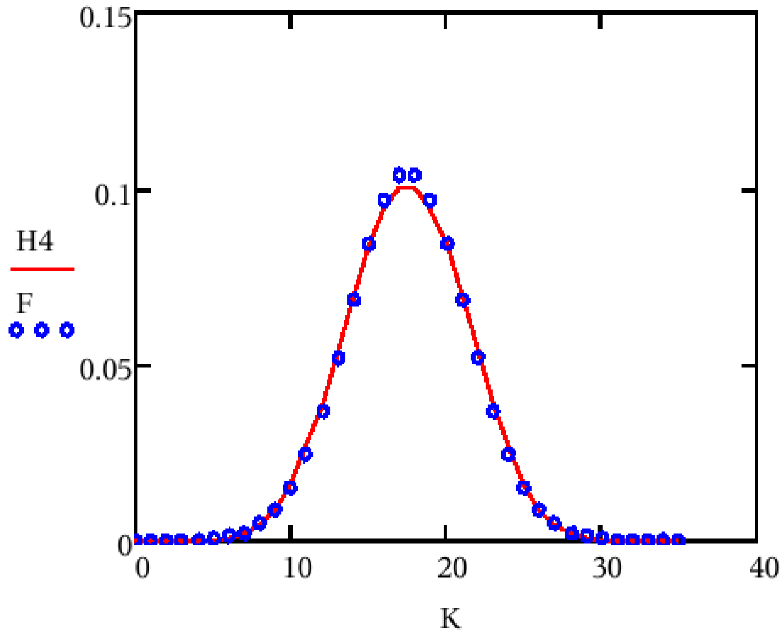


Abbildung 44:

Hierbei bezeichnet X_i die Zg, die den Zufallsprozess der oben aufliegenden Augenanzahl von Würfel i beschreibt. In Beispiel 8 hatten wir

$$EX_i = 3.5, \quad \mathrm{Var}(X_i) = \frac{35}{12}$$

festgestellt. Aufgrund der guten Approximationseigenschaften, kann man nun zur näherungsweisen Berechnung den Wert von

$$\frac{6^5}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right)$$

an der Stelle $k = 22$ heranziehen, der 405.708 lautet.

20. Aufgabe: Es soll der Parameter $a \in \mathbb{R}_+$ so bestimmt werden, damit

$$f_a(x) := \begin{cases} 0, & \text{falls} \quad x \leq -a \\ p_a(x), & \text{falls} \quad -a \leq x \leq a \\ 0, & \text{falls} \quad a \leq x \end{cases}$$

als Dichtefunktion dienen kann, wobei das quadratische Polynom $p_a(x) := p_2x^2 + p_1x + p_0$ durch die Punkte $(-a, 0)$, $(0, 1)$ und $(a, 0)$ hindurchgehen soll.

Antwort: Offenbar besitzt das Polynom

$$p(x) := (x-a)(x+a)b$$

für jedes $b \in \mathbb{R}$ die Nullstellen $\pm a$. Es reicht daher aus, b so zu bestimmen, daß $p(0) = 1$ gilt, d.h.

$$-a^2 b = 1.$$

Demnach hat man $b := -\frac{1}{a^2}$ zu wählen, sodaß

$$p_a(x) = -\frac{1}{a^2}(x-a)(x+a)$$

gilt. Es verbleibt, a so zu bestimmen, daß

$$\int_{-a}^{a} p_a(x)\mathrm{d}x = 1$$

gilt. Offenbar gilt

$$\int_{-a}^{a} p_a(x)\mathrm{d}x = -\frac{1}{a^2}\int_{-a}^{a}(x^2-a^2)\mathrm{d}x = -\frac{1}{a^2}\left(\left.\frac{x^3}{3}\right|_{-a}^{a} - a^2\, x|_{-a}^{a}\right)$$

$$= -\frac{1}{a^2}\left(\frac{2a^3}{3} - a^2(2a)\right) = \frac{4}{3}a.$$

Daher führt die Normierungsbedingung $\int_{-a}^{a} p_a(x)\mathrm{d}x = 1$ auf $a = \frac{3}{4}$.

6 Tabellen

6.1 Werteverlauf der Standard-Normalverteilung

x	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5318	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7421	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8887	0.8907	0.8926	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9853	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9988	0.9989	0.9989	0.9990	0.9990

(78)

Tabelle (78) zeigt den Werteverlauf von $\Phi_{0,1}(x)$ für $x = 0.00, \ldots, 3.09$. Ablesebeispiel: Man möchte $\Phi_{0,1}(1.84)$ wissen. Zu diesem Zweck geht man in Zeile 19 und Spalte 5 und entnimmt dort den Wert 0.9671. Falls man die Gleichung $\Phi_{0,1}(x) = z$ lösen möchte, wobei $z \in [0.5, 1]$ sucht man innerhalb der Tabelle die Stelle, die am nächsten zu z liegt und liest auf dem Rand das gesuchte x ab. Z.B. soll $\Phi_{0,1}(x) = 0.8508$ gelöst werden. Offenbar befindet sich 0.8508 in Zeile 11 und Spalte 5, daher gilt $x = 1.04$.

6.2 Werteverlauf der Umkehrfunktion zur STUDENTschen Verteilung

Mittels des **mathcad**-Befehls qt(x, n) bekommt man folgenden Werteverlauf von $F_n^{-1}(x)$ für

$$(n, x) \in \{1, \ldots, 30\} \times \{0.900,\ 0.950,\ 0.975,\ 0.990,\ 0.995,\ 0.999\}, \quad x := 1 - \frac{\alpha}{2}$$

wobei F_n die Verteilungsfunktion der t-Verteilung für n Freiheitsgrade darstellt:

$n \diagdown x$	0.900	0.950	0.975	0.990	0.995	0.999
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.885	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.724	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.410	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.313	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385

(79)

Vgl. Tabelle A.2 [15] S.180.

Ablesebeispiel: Wir wollen den Wert z von $F_6^{-1}(0.990)$ wissen. Zu diesem Zweck liest man in Zeile 6 und Spalte 4 den Wert 3.143 ab. Die für uns in diesem Zusammenhang wichtigen Werte von z befinden sich im Inneren der Tabelle.

6.3 Werteverlauf der Umkehrfunktion zur χ^2-Verteilungsfunktion

Im Inneren von Tabelle (80) ist der Werteverlauf von

$$F_m^{-1}(1-\alpha)$$

für $\alpha = 0.01, 0.05, 0.10$ und $m = 1, \ldots, 30$ zu finden, wobei F_m die χ^2-Verteilungsfunktion mit m Freiheitsgraden darstellt. Die Tabelle wurde mit dem mathcad-Befehl qchisq(x, m) erzeugt. Vgl. Tabelle 3 [13] S.738, Tabelle A.3 [15] S.181, Tabelle 0.4.6.4 [20] S.93.

m \ α	0.01	0.05	0.10
1	6.63	3.84	2.71
2	9.21	5.99	4.61
3	11.35	7.81	6.25
4	13.28	9.49	7.78
5	15.09	11.07	9.24
6	16.81	12.59	10.64
7	18.48	14.07	12.02
8	20.09	15.51	13.36
9	21.67	16.92	14.68
10	23.21	18.31	15.99
11	24.73	19.68	17.28
12	26.22	21.03	18.55
13	27.69	22.36	19.81
14	29.14	23.68	21.06
15	30.58	25.00	22.31
16	32.00	26.30	23.54
17	33.41	27.59	24.77
18	34.81	28.87	25.99
19	36.19	30.14	27.20
20	37.57	31.41	28.41
21	38.93	32.67	29.62
22	40.29	33.92	30.81
23	41.64	35.17	32.01
24	42.98	36.42	33.20
25	44.31	37.65	34.38
26	45.64	38.89	35.56
27	46.96	40.11	36.74
28	48.28	41.34	37.92
29	49.59	42.56	39.09
30	50.89	43.77	40.26

$F_m(x) = 1 - \alpha$
x steht im Inneren der Tabelle
$F_m(x)$ ist die χ^2-Verteilungsfunktion
mit m Freiheitsgraden.

Wegen $F_{2k}(x) = 1 - e^{-x/2} \sum_{i=0}^{k-1} \frac{(x/2)^i}{i!}$
entstehen die Einträge der Zeile $2k$
durch Lösen der Gleichung (80)

$$\alpha e^{x/2} = \sum_{i=0}^{k-1} \frac{(x/2)^i}{i!}$$

für $\alpha = 0.01, 0.05, 0.10$.

7 Literatur

Literatur

[1] H. Bauer, *Wahrscheinlichkeitstheorie*, de Gruyter, Berlin, 1991

[2] Bronstein, Semandjajew, Musiol, Mühlig, *Taschenbuch der Mathematik*, Harri Deutsch, Frankfurt, 1997

[3] A. Caputo, L. Fahrmeir, R. Künstler, St.Lang, I. Pigeot, G. Tutz, *Arbeitsbuch Statistik*, Springer, Berlin, 2009

[4] R. Courant, H. Robbins, *Was ist Mathematik*, Springer, Berlin, 2001

[5] L. Fahrmeir, R. Künstler, I. Pigeot, G. Tutz, *Statistik*, Springer, Berlin, 2010

[6] M. Fisz, *Wahrscheinlichkeitsrechnung und mathematische Statistik*, Deutscher Verlag der Wissenschaften, Berlin, 1970

[7] F. R. Gantmacher, *Matrizenrechnung*, Deutscher Verlag der Wissenschaften, Berlin, 1958

[8] B.W. Gnedenko, A.J. Chintschin, *Elementare Einführung in die Wahrscheinlichkeitsrechnung*, Deutscher Veralg der Wissenschaften, Berlin, 1979

[9] B.W. Gnedenko, *Lehrbuch der Wahrscheinlichkeitsrechnung*, Akademie-Verlag, Berlin, 1970

[10] B.W. Gnedenko, I.O. Kowalenko, *Einführung in die Bedienungstheorie*, Akademieverlag, Berlin, 1971

[11] O. Häggström, *Streifzüge durch die Wahrswcheinlichkeitsrechnung*, Springer, Berlin, 2006

[12] G. Moschytz, M. Hofbauer, *Adaptive Filter*, Springer, Berlin, 2000

[13] L. Papula, *Mathematik für Ingenieure und Naturwissenschaftler*, Band 3, Viehwegs Fachbücher der Technik, Braunschweig, Wiesbaden, 1997

[14] W. Rudin, *Reelle und Komplex Analysis*, Oldenbourg, München,1999

[15] M. Sachs, *Wahrscheinlichkeitsrechnung und Statistik*, Fachbuchverlag Leipzig, 2009

[16] D. Struik, *Abriss der Geschichte der Mathematik*, Deutscher Verlag der Wissenschaften, Berlin, 1967

[17] H. Toutenburg, C. Heumann, *Deskriptive Statistik*, Springer, Berlin, 2008

[18] W. Voss Hrsg., *Taschenbuch der Statistik*, Fachbuchverlag Leipzig, 2004

[19] H. G. Zachmann, *Mathematik für Chemiker*, Wiley-VCH, 2004

[20] E. Zeidler, *Teubner-Taschenbuch der Mathematik*, B.G.Teubner, Stuttgart, Leipzig, 1996