Mathematik für Geowissenschaftler II

Christian-Albrechts-Universität zu Kiel

Marco Kühnel, Richard Weidmann

Inhaltsverzeichnis

KAPITEL 1

Kombinatorik

1. Mengen

Im Folgenden betrachten wir Mengen mit endlich vielen Elementen, so ist z.B.

$$A := \{a, b, c, \ldots, x, y, z\}$$

eine Menge mit 26 Elementen. Der Einfachheit halber werden wir meist Mengen betrachten, die nur aus natürlichen Zahlen bestehen, z.B.

$$\{1, 2, 3, \ldots, 24, 25, 26\}.$$

Jede endliche Menge kann so dargestellt werden, wenn wir die Elemente in natürliche Zahlen umbenennen, in dem Beispiel durch $a \to 1$, $b \to 2$ usw.

Wir wollen zunächst zählen wie viele Teilmengen die Menge $X_n := \{1, \ldots, n\}$ hat. Hier ist eine Teilmenge von X_n eine Menge Y, so dass jedes Element aus Y auch eine Element aus Y ist.

Im Fall $n = 4$ hat $X_4 = \{1, 2, 3, 4\}$ offensichtlich die Teilmengen $\emptyset$ (die leere Menge, die kein Element enthält), $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, $\{3, 4\}$, $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$, $\{2, 3, 4\}$, $\{1, 2, 3, 4\}$, also einmal die leere Menge, 4 einelementige Mengen, 6 zweielementige Mengen, 4 dreielementige Mengen und eine vierelementige Menge.

Insgesamt hat X_4 also $1 + 4 + 6 + 4 + 1 = 16 = 2^4$ Teilmengen.

Das hätte man sich auch schneller überlegen können: Um eine Teilmenge von X_n zu beschreiben, muss man für jedes Element aus X_n festlegen, ob es in der Teilmenge liegt oder nicht. Da es n Element in X_n gibt, gibt es also insgesamt 2^n Möglichkeiten, diese Festlegungen zu treffen und somit 2^n Teilmengen. Wir haben also den folgenden Satz:

SATZ 1.1. *Eine n-elementige Menge hat genau 2^n (verschiedene) Teilmengen.*

Mengen sind nicht mit einer Ordnung versehen, d.h. es gilt z.B. $\{1, 2\} = \{2, 1\}$. Eine Menge kann jedoch (an)geordnet werden. Im folgenden zählen wir die verschiedenen Anordnungen einer endlichen Menge. So kann die Menge $X_3 = \{1, 2, 3\}$ in den folgenden Weisen (als n-Tupel) angeordnet werden: $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$ und $(3, 2, 1)$. Insgesamt gibt es also $6 = 1 \cdot 2 \cdot 3$ verschiedene Anordnungen.

Das Produkt der Zahlen von 1 bis n hat einen Namen:

DEFINITION 1.2 (Fakultät). *Für jede natürliche Zahl $n \in \mathbb{N}$ heißt die Zahl*

$$n! := 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n = \prod_{i=1}^{n} i$$

n Fakultät. Ferner sei $0! = 0$.

Die Anzahl der Anordnungen einer n-elementigen Menge ist nun leicht bestimmt. Es gibt n Möglichkeiten das erste Element zu wählen. Wenn das erste gewählt ist, gibt es noch $(n-1)$ für das zweite usw. Insgesamt gibt es also $n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$ verschiedene Anordnungen. Wir erhalten:

SATZ 1.3. *Eine n-elementige Menge kann auf genau $n!$ verschiedene Arten angeordnet werden.*

In der Mathematik heißen Anordnungen auch Permutationen. Eigentlich meint man mit Permutationen hier jedoch die Abbildungen, die die Menge umordnen.

Manchmal interessiert man sich auch für zyklische Anordnungen der Elemente einer endlichen Menge. Damit sind ringförmige Anordnungen gemeint, bei denen keine Umlaufrichtung vorgegeben ist. Jedes Element hat also genau zwei benachbarte Elemente und zwei solche Anordnungen sollen für uns identisch sein, wenn die eine nur durch Vertauschen aller rechten mit den entsprechenden linken Nachbarn aus der anderen hervorgeht. Ausgehend von einer gewöhnlichen Anordnung erhalten wir eine zyklische, indem wir das letzte Element der Anordnung an das erste "knüpfen". Wir erhalten so alle zyklischen Anordnungen, aber jede zyklische Anordnung kann auf $2n$ verschiedene Weisen erhalten werden: Indem wir einen Startpunkt und eine Umlaufrichtung einer zyklischen Anordnung festlegen, erhalten wir eine gewöhnliche Anordnung, die durch "Zusammenknüpfen" wieder die zyklische Anordnung ergibt, mit der wir begonnen haben; es gibt n mögliche Startpunkte und 2 Umlaufrichtungen. Also haben wir gezeigt

SATZ 1.4. *Eine n-elementige Menge (mit $n \geq 3$) kann auf $\frac{(n-1)!}{2}$ verschiedene Arten zyklisch angeordnet werden.*

Dies ist ein Spezialfall geschlossener Wege in Graphen.

DEFINITION 1.5 ((vollständiger)Graph/geschlossener Weg).
(1) *Ein Graph $G = (V, E)$ besteht aus einer endlichen Menge V von sogenannten "Knoten" und einer Menge E von "Kanten", die je zwei Knoten verbinden. Zwischen zwei Knoten darf es mehrere Kanten geben.*
(2) *Ein Graph $G = (V, E)$ heißt vollständiger Graph, wenn zwischen je zwei Knoten genau eine Kante besteht.*
(3) *Ein geschlossener Weg in einem Graph $G = (V, E)$ ist eine Teilmenge $F \subseteq E$ von Kanten so, dass jeder Knoten in V in genau zwei Kanten in F vorkommt und man von jedem Knoten in V zu jedem anderen Knoten in V über Kanten in F gelangen kann.*

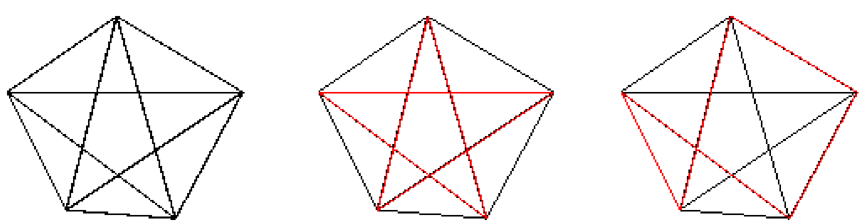


ABBILDUNG 1.1. Der vollständige Graph mit 5 Knoten. In den farbigen Abbildungen bilden sowohl die roten als auch die schwarzen Kanten geschlossene Wege. Durch Rotation der geschlossenen Wege im rechten Bild erhält man schließlich alle 12 geschlossenen Wege.

Nun ist das folgende Resultat nicht schwer einzusehen.

SATZ 1.6. *Die Anzahl geschlossener Wege im vollständigen Graphen mit n Knoten ist die Anzahl zyklischer Anordnungen einer n-elementigen Mengen, also $\frac{(n-1)!}{2}$.*

Da jeder Graph mit n Knoten, der zwischen zwei Knoten keine mehrfachen Kante besitzt, zum vollständigen Graphen mit n Knoten komplettiert werden kann, gibt es in einem solchen Graphen maximal $\frac{(n-1)!}{2}$ geschlossene Wege. Lässt man mehrfache Kanten zu, so kann es auch mehr als $\frac{(n-1)!}{2}$ geschlossene Wege geben.

Als nächstes wollen wir die für $m \leq n$ die m-elementigen Teilmengen einer n-elementigen Menge zählen. Wir zählen die Möglichkeiten, m Elemente aus X_n auszuwählen. Wenn wir die Reihenfolge berücksichtigen, so gibt es hier offensichtlich

$$n \cdot (n-1) \cdot \ldots \cdot (n-m+2) \cdot (n-m+1) = \frac{n!}{(n-m)!}$$

Möglichkeiten. Da für Teilmengen die Reihenfolge keine Rolle spielt und jede Teilmenge nach Satz 1.3 $m!$ oft ausgewählt wurde, erhalten wir

$$\frac{n!}{m!(n-m)!}$$

verschieden Teilmengen. Diese Zahl hat einen Namen:

DEFINITION 1.7 (Binomialkoeffizient). *Der Binomialkoeffizient $\binom{n}{m}$ (sprich n* über *m) ist definiert als*

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

Wir haben also den folgenden Satz:

SATZ 1.8. *Eine n-elementige Menge hat $\binom{n}{m}$ m-elementige Teilmengen.*

Es ist klar, dass $\binom{n}{m} = \binom{n}{m-n}$. Diese Symmetrie hatten wir bereits beim Zählen der Teilmengen von X_4 gesehen.

Da jede Teilmenge einer n-elementigen Menge zwischen 0 und n Elemente hat erhalten wir aus Satz 1.3 und Satz 1.8, dass

$$2^n = \sum_{m=0}^{n} \binom{n}{m}.$$

Im Falle der Anzahl der Teilmengen von X_4 stimmt dies wie erwartet mit unserem expliziten Zählen überein:

$$16 = 2^4 = \sum_{m=0}^{4} \binom{4}{m} = 1 + 4 + 6 + 4 + 1.$$

Allgemeiner erhalten wir den folgenden Satz. Man erhält ihn durch das Ausmultiplizieren von $(a+b)^n$ und zählen der einzelnen Terme. So kommt der Term $a^m b^{n-m}$ genau $\binom{n}{m}$ mal vor, da es genau $\binom{n}{m}$ Möglichkeiten gibt, die m Vorkommnisse von a auf die n Faktoren von $(a+b)^n$ zu verteilen.

SATZ 1.9 (Binomischer Lehrsatz). *Es gilt*

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

Im Fall $a = b = 1$ ergibt der binomische Lehrsatz die vorige Gleichung.

2. Urnenmodelle

Urnenmodelle sind oft nützlich um Sachen abzuzählen. Es geht hier um eine Urne mit n Kugeln, aus der m Kugeln gezogen werden. Wir betrachten sowohl den Fall, dass wir die gezogenen Kugel sofort wieder zurücklegen als auch den Fall, dass die gezogenen Kugeln nicht zurück gelegt werden. Im ersten Fall können wir beliebig oft ziehen, im letzten höchstens n mal.

Beachten wir die Reihenfolge, in der wir die Kugeln ziehen, erhalten wir leicht den folgenden Satz:

SATZ 1.10. *In einer Urne liegen n Kugeln. Wir ziehen m Kugeln und beachten (anders als beim Lotto) die Reihenfolge der Kugeln.*

(1) *Beim Ziehen ohne Zurücklegen gibt es* $n \cdot (n-1) \cdot \ldots \cdot (n-m+1) = \frac{n!}{(n-m)!}$ *Möglichkeiten, m Kugeln zu ziehen. Hier ist* $m \leq n$.
(2) *Beim Ziehen mit Zurücklegen gibt es* n^m *Möglichkeiten, m Kugeln zu ziehen. Hier ist* $m \in \mathbb{N}$ *beliebig.*

Interessieren wir uns (wie beim Lotto) nur für die gezogenen Kugeln, nicht aber für die Reihenfolge in der sie gezogen werden, so erhalten wir den folgenden Satz:

SATZ 1.11. *In einer Urne liegen n Kugeln. Wir ziehen m Kugeln und beachten (wie beim Lotto) die Reihenfolge der Kugeln nicht.*

(1) *Beim Ziehen ohne Zurücklegen gibt es* $\binom{n}{m}$ *Möglichkeiten, m Kugeln zu ziehen. Hier ist* $m \leq n$.
(2) *Beim Ziehen mit Zurücklegen gibt es* $\binom{n+m-1}{m}$ *Möglichkeiten, m Kugeln zu ziehen. Hier ist* $m \in \mathbb{N}$ *beliebig.*

Die erste Aussage ist klar, da das Ziehen ohne Zurücklegen genau der Auswahl einer m-elementigen Teilmenge entspricht. Die Behauptung folgt somit aus Satz 1.8. Zum Beweis der zweiten Aussage betrachten wir neben den n Kugeln weitere m "Hilfskugeln". Nachdem wir m Kugeln aus den n gezogen, zählen wir wie oft Kugel 1 gezogen wurde und nennen diese Anzahl a_1. Ebenso für jede andere der n Kugeln. Wir erhalten also nichtnegative Zahlen $a_1, \ldots, a_n$ mit der Eigenschaft $a_1 + \cdots + a_n = m$. Seien davon $a_{i_1}, \ldots, a_{i_k}$ verschieden von 0, die verschiedenen gezogenen Kugeln also Nummer $i_1, i_2, \ldots, i_k$. Von den m Hilfskugeln ziehen wir nun wie folgt: die erste lassen wir weg, nehmen die $a_1 - 1$ nächsten, lassen wieder eine aus, nehmen die nächsten $a_2 - 1$ und so fort. Auf diese Weise haben wir nun wirklich m verschiedene Kugeln gezogen und zwar aus einer Gesamtmenge von $n+m-1$ Kugeln: die erste Hilfskugel wird nie gezogen. Umgekehrt kann man sich überlegen, dass jede Wahl von m verschiedenen Kugeln aus den $n+m-1$ Kugeln (niemals die erste Hilfskugel wählen) eine Wahl von m Kugeln aus n mit Zurücklegen auf die angegebene Weise codiert. Wir erhalten also $\binom{n+m-1}{m}$ Möglichkeiten, m Kugeln aus n mit Zurücklegen zu ziehen.

3. Beispiele

BEISPIEL 1: Der Achter rudert nicht schnell genug. Der Trainer entscheidet sich, alle möglichen Sitzordnung auszuprobieren. Wie viele sind dies?

Antwort: Nach Satz 1.3 gibt es $8! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot 8 = 40320$ Sitzordnungen. Alternativ benutzen wir Satz 1.10(1).

BEISPIEL 2: An dem Marathonlauf nehmen 1000 Läufer teil. Die ersten drei werden geehrt. Wie viele mögliche Siegerehrungen sind vorstellbar?

Antwort: Die Reihenfolge spielt offensichtlich eine Rolle und kein Läufer wird mehr als einmal geehrt. Es gibt also $1000 \cdot 999 \cdot 998 = 997.002.000$ mögliche Ehrungen. Dies entspricht dem Ziehen ohne Zurücksetzen mit Beachtung der Ordnung, also Satz 1.10(1)

BEISPIEL 3: Eine Klasse hat 30 Schüler. Meistens sind einige krank. Wie viele mögliche Anwesenheiten gibt es?

Antwort: Nach Satz 1.8 gibt es 2^{30} Teilmengen der Menge der Schüler also $2^{30} = 1.073.741.824$ Besetzungen. Alternativ kann man hier mit Satz 1.10(2) argumentieren.

BEISPIEL 4: Wie viele verschiedene Würfe mit zwei Würfeln gibt es?

Antwort: Es ist das Urnenmodell mit $n = 6$ und $m = 2$ mit Zurücklegen und ohne Beachtung der Reihenfolge anwendbar. Ohne Reihenfolge deshalb, weil nicht wesentlich ist, welcher Würfel welche Zahl liefert. Satz 1.11(2) liefert also $\binom{n+m-1}{m} = \binom{6+2-1}{2} = \binom{7}{2} = 21$ mögliche Würfe.

BEISPIEL 5: Wie viele Möglichkeiten gibt es, 8 Türme so auf ein Schachbrett zu stellen, dass sie sich (paarweise) nicht bedrohen?

Antwort: In jeder Spalte (Vertikalen) und Reihe (Horizontalen) muss genau ein Turm stehen. Den Turm in der 1. Spalte können wir in eine beliebige Zeile stellen, haben also 8 Möglichkeiten. Für den Turm der 2. Spalte haben wir dann noch 7 Möglichkeiten usw. Insgesamt haben wir also $8 \cdot 7 \cdot 6 \cdot \ldots \cdot 1 = 8! = 40320$ Möglichkeiten.

BEISPIEL 6: Wie viele Möglichkeiten gibt es, 8 Türme auf ein Schachbrett zu stellen?

Antwort: Nach Satz 1.11(1) haben wir hier $\binom{64}{8}$.

BEISPIEL 7: Beim Lotto werden 6 von 49 Kugeln gezogen. Wie viele mögliche Ziehungen bzw. Tipps gibt es?

Antwort: Es handelt sich um ein Ziehen ohne Zurücklegen und ohne Beachtung der Ordnung. Nach Satz 1.11(1) gibt es also $\binom{49}{6} = 13.983.816$ verschiedene Ziehungen.

BEISPIEL 8: Bei wie vielen Tipps hat man genau 4 Richtige?

Antwort: Bei 4 Richtigen hat man 2 Falsche. Es gibt $\binom{6}{4}$ Möglichkeiten für die 4 Richtigen und $\binom{43}{2}$ für die 2 Falschen. Insgesamt gibt es also $\binom{6}{4}\binom{43}{2}$ Tipps mit 4 Richtigen.

BEISPIEL 9: Wie groß ist die Wahrscheinlichkeit, beim Lotto genau 4 Richtige zu haben?

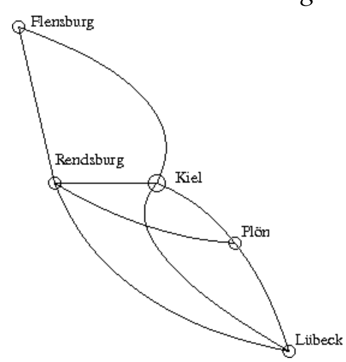
Antwort: Es gibt $\binom{49}{6}$ mögliche Tipps nach Beispiel 7. Davon führen genau $\binom{6}{4}\binom{43}{2}$ zu 4 Richtigen. Also ist die Wahrscheinlichkeit 4 Richtige zu tippen

$$\frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}}.$$

4. Aufgaben

Aufgabe 1-1 Ein Radfahrer plant eine Fahrradtour durch Kiel, Flensburg, Plön, Rendsburg und Lübeck. Er will jede Stadt nur einmal aufsuchen.

(1) Wieviele verschiedene Reihenfolgen der Städte sind denkbar?
(2) Wieviele verschiedene Rundwege sind denkbar (ohne fixierten Start-/Endpunkt, ohne Umlaufrichtung), wenn es zwischen je zwei der Städte genau einen Weg gibt?
(3) Nun hat er folgende Strassenkarte zur Verfügung. Wieviele Rundwege gibt es nun (ohne definierten Start-/Endpunkt, ohne Umlaufrichtung, an Kreuzungen darf abgebogen werden)?
Hinweis: Bestimmen Sie zunächst den dem Problem zugrundeliegenden Graphen.



Aufgabe 1-2 Es gibt in Kiel-Holtenau 23 Eistage pro Jahr (ganztägig unter 0 Grad Celsius), die alle in das Winterhalbjahr Oktober bis März fallen (insgesamt 182 Tage).

(1) Wie viele verschiedene Verteilungen der Eistage auf das Winterhalbjahr gibt es?
(2) Wie viele verschiedene Verteilungen der Eistage auf das Winterhalbjahr gibt es, bei denen alle Tage von 24.-26. Dezember oder alle Tage vom 31.12-1.1. Eistage sind? Wie hoch ist deren Anteil an allen möglichen Verteilungen (aus Teil (1))?

Aufgabe 1-3 An fünf Orten soll je ein Präzisionsthermometer aufgestellt werden. Die zu erwartenden Temperaturspannen in Grad Celsius an jedem Ort sind

Ort	Temperaturbereich
1	-20 – 37
2	-5 – 25
3	12 – 32
4	-65 – -7
5	0 – 18

Zur Verfügung stehen Thermometer der Firma A, die von -10 bis 35 Grad Celsius genau messen, von der Firma B (-70 bis 20 Grad Celsius) und der Firma C (-25 bis 45 Grad Celsius). Wie viele verschiedene Thermometerverteilungen auf die Messstationen gibt es?

Aufgabe 1-4 Auf 7 verschiedene Messstationen soll je ein Thermometer gestellt werden. Wir haben 10 verschiedene Thermometer zu Auswahl. Wie viele verschiedene Möglichkeiten gibt es, die Messstationen mit je einem Thermometer zu bestücken? Wie viele unterschiedliche Mengen von unverteilten Thermometern gibt es?

KAPITEL 2

Elementare Wahrscheinlichkeitstheorie

Im Folgenden wollen wir ein Zufallsexperiment durch ein mathematisches Modell beschreiben. Wir diskutieren einige einfache wahrscheinlichkeitstheoretische Begriffe.

1. Elementarereignisse und Wahrscheinlichkeit

Wir bezeichnen die möglichen Ausgänge eines Experiments als die Elementarereignisse, wir bezeichnen diese Menge mit Ω.

BEISPIELE:

(1) Beim 5-maligen Würfeln besteht die Menge der Elementarereignisse aus der Menge der 5-Tupel mit Einträgen zwischen 1 und 6, d.h.
$$\Omega = \{(x_1, \ldots, x_5) \mid x_i \in \{1, \ldots, 6\} \text{ für } 1 \leq i \leq 5\}.$$
Insbesondere hat Ω genau 6^5 Elemente.

(2) Aus einer Menge von n Messstationen (davon m auf der Nordhalbkugel und $n - m$ auf der Südhalbkugel) werden "zufällig" 5 ausgewählt. Hier entspricht Ω der Menge der 5-Elementige Teilmengen von $\{1, \ldots, n\}$. Insbesondere hat Ω genau $\binom{n}{5}$ Elemente.

Eine Teilmenge von Ω nennt man ein Ereignis, ein Ereignis besteht also aus einer Menge von Elementarereignissen. Insbesondere sind die leere Menge $\emptyset$ und Ω Ereignisse. So ist im ersten Beispiel die Menge der Elementarereignisse, bei denen 5 mal die gleiche Zahl gewürfelt wird, ein Ereignis.

Der Begriff der Wahrscheinlichkeit ordnet nun jedem Ereignis $A \subseteq \Omega$ eine Zahl $P(A)$ zu. Ist diese Zahl 1, so erwarten wir, dass das Ereignis immer eintritt, ist sie 0, so erwarten wir, dass es nie eintritt. Wir beschränken uns im folgenden auf endliches Ω.

DEFINITION 2.1. *Sei Ω ein endlicher[1] Ereignisraum und für jedes $A \subseteq \Omega$ sei $P(A) \in \mathbb{R}$ definiert.*

Wir sagen, dass dies ein Wahrscheinlichkeitsmaß, bzw. eine Wahrscheinlichkeitsverteilung auf Ω definiert, wenn folgende Bedingungen erfüllt sind:

(1) $P(A) \geq 0$ *für alle* $A \subseteq \Omega$.
(2) $P(\emptyset) = 0$ *und* $P(\Omega) = 1$.
(3) *Sind* $A_1, \ldots, A_n \subseteq \Omega$ *disjunkt, so gilt* $P(A_1 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_n)$.

Ist also $A = \{a_1, \ldots, a_n\} \subseteq \Omega$, so gilt $P(A) = P(a_1) + \ldots + P(a_n)$, man kann also die Wahrscheinlichkeiten beliebiger Ereignisse aus den Wahrscheinlichkeiten der Elementarereignisse bestimmen.

Ist A^C das Komplement von A in Ω, gilt also $A \cup A^C = \Omega$ und $A \cap A^C = \emptyset$, so gilt $P(A^C) = 1 - P(A)$. Man nennt A^C auch das Gegenereignis von A.

[1] Im Fall von unendlichem Ω muss man in (3) auch unendliche (abzählbare) Vereinigungen zulassen.

Besteht Ω aus n Elementen und ist jedes Elementarereignis gleich wahrscheinlich, also $P(a) = \frac{1}{n}$ für alle $a \in \Omega$, so sprechen wir von einem *Laplace-Wahrscheinlichkeitsraum*.

BEISPIELE:

(1) Bei 5-maligem Würfeln ist es sinnvoll, davon auszugehen, dass jedes Elementarereignis gleich wahrscheinlich ist und somit gilt $P(a) = \frac{1}{6^5}$ für alle $a \in A$. Für das Ereignis A, dass alle 5 Zahlen gleich sind, gilt also $P(A) = P(1,1,1,1,1) + P(2,2,2,2,2) + \ldots + P(6,6,6,6,6) = \frac{1}{6^5} + \ldots + \frac{1}{6^5} = 6 \cdot \frac{1}{6^5} = \frac{1}{6^4}$.

(2) Bei dem Beispiel mit den Messstationen ist es vernünftig, davon auszugehen, dass alle Elementarereignisse gleich wahrscheinlich sind, also gilt $P(a) = \frac{1}{\binom{n}{5}}$ für alle $a \in \Omega$. Sei A das Ereignis, dass 2 Stationen auf der Nordhalbkugel und 3 auf der Südhalbkugel ausgewählt werden. Es gibt $\binom{m}{2}$ mögliche Auswahlen für die Stationen der Nordhalbkugel und $\binom{n-m}{3}$ für die auf der Südhalbkugel. Insgesamt besteht A aus $\binom{m}{2} \cdot \binom{n-m}{3}$ Elementarereignissen. Also gilt $P(A) = \frac{\binom{m}{2} \cdot \binom{n-m}{3}}{\binom{n}{5}}$.

2. Bedingte Wahrscheinlichkeit

2.1. Definition und Unabhängigkeit. Seinen A und B Ereignisse. Man interessiert sich häufig für die Wahrscheinlichkeit, dass das Ereignis B eintritt unter der Bedingung, dass A eingetroffen ist. Dies führt uns zum Begriff der bedingten Wahrscheinlichkeit.

DEFINITION 2.2. *Sind A und B Ereignisse, so bezeichnen wir mit*

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

die bedingte Wahrscheinlichkeit von B unter der Hypothese A. Gilt $P(B|A) = P(B)$, so nennt man A und B unabhängig.

Sind A und B unabhängig, gilt also $P(B) = P(B|A) = \frac{P(A \cap B)}{P(A)}$, so gilt $P(A \cap B) = P(A) \cdot P(B)$, d.h. die Wahrscheinlichkeit, dass zwei unabhängige Ereignisse eintreten, ist das Produkt der Wahrscheinlichkeiten der Ereignisse. In der Tat sind äquivalent

(1) A und B sind unabhängig,
(2) $P(B|A) = P(B)$,
(3) $P(A|B) = P(A)$,
(4) $P(A \cap B) = P(A) \cdot P(B)$.

Es ist leicht einzusehen, dass mit A und B auch B und A, A und B^C, A^C und B^C sowie A^C und B unabhängig sind.

BEISPIEL 2.3. (1) Man würfelt einmal, wir haben also $\Omega = \{1,2,3,4,5,6\}$ und $P(a) = \frac{1}{6}$ für alle $a \in \Omega$. Die Wahrscheinlichkeit, eine 2 zu würfeln, ist also $P(\{2\}) = \frac{1}{6}$ und die Wahrscheinlichkeit, eine gerade Zahl zu würfeln, ist $P(\{2,4,6\}) = 3 \cdot \frac{1}{6} = \frac{1}{2}$. Es gilt $P(\{2\}|\{2,4,6\}) = \frac{P(\{2\})}{P(\{2,4,6\})} = \frac{1/6}{1/2} = \frac{1}{3}$; das Ereignis, eine 2 zu würfeln und das Ereignis, eine gerade Zahl zu würfeln, sind also nicht unabhängig. Das war natürlich zu erwarten.

(2) Von 30 Schülern sind 12 Jungs und 18 Mädchen. 5 Schüler stehen in Mathe auf einer 5. Von diesen 5 Schülern sind 2 Jungs. Wir wählen nun zufällig einen Schüler aus. Sei A das Ereignis, dass der Schüler ein Junge ist und B das Ereignis, dass der Schüler eine 5 hat. Es

gilt offensichtlich, dass $P(A) = \frac{2}{5}$, $P(B) = \frac{1}{6}$ und $P(A \cap B) = \frac{2}{30} = \frac{1}{15}$. Es folgt, dass

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{15}}{\frac{1}{6}} = \frac{6}{15} = \frac{2}{5} = P(A).$$

Die Ereignisse A und B sind also unabhängig.

$\triangle$

2.2. Relative Häufigkeiten vs. bedingte Wahrscheinlichkeiten. Wir betrachten eine endliche Menge Ω mit N Elementen und die Teilmengen (Ereignisse) $A, B \subseteq \Omega$. Die Teilmenge A habe a Elemente, B habe b Elemente und $A \cap B$ habe c. Wir definieren dann

DEFINITION 2.4. (1) *die relative Häufigkeit* $H(A) := \frac{a}{N}$ *und*
(2) *die bedingte relative Häufigkeit* $H(A|B) := \frac{c}{b}$.

Durch Vertauschung der Rollen von A und B sind dann natürlich auch $H(B)$ und $H(B|A)$ definiert.

Die relativen Häufigkeiten sind die (bedingten) Wahrscheinlichkeiten der Ereignisse, wenn Ω als Laplace-Wahrscheinlichkeitsraum aufgefasst wird:

PROPOSITION 2.5. *Ist* (Ω, P) *eine Laplace-Wahrscheinlichkeitsraum, so gilt* $H(A) = P(A)$ *und* $H(A|B) = P(A|B)$ *für alle Ereignisse* $A, B \subseteq \Omega$.

BEISPIEL 2.6. Folgende Daten über den Seeverkehr deutscher Häfen veröffentlichte das Statistische Bundesamt 2010:

	Ein- und ausgestiegene Fahrgäste 2010					
		Ein- und Ausstiegsgebiete				
	insgesamt	Schleswig-Holstein	Hamburg	Niedersachsen	Bremen/ Bremerhaven	Mecklenburg-Vorpommern
Fahrtgebiete	In 1000					
Verkehr innerhalb Deutschlands	**16809**	**5694**	**72**	**11042**	**1**	**–**
darunter:						
Niedersachsen	11324	293	–	11031	–	–
Schleswig-Holstein, Nordsee	5397	5330	56	11	1	–
Verkehr mit Häfen außerhalb Deutschlands	**11972**	**8848**	**200**	**–**	**47**	**2876**
Europa	11959	8848	190	–	44	2876
Europäische Union	10686	7689	126	–	24	2844
Sonstiges Europa	1273	1155	65	–	20	33
Ostseegebiet	10191	7309	12	–	10	2860
darunter:						
Dänemark, Ostsee	7879	6362	2	–	2	1513
Schweden	1772	722	6	–	1	1043
Polen	139	7	2	–	3	126
Nordeuropa	1644	1496	102	–	29	17
darunter:						
Norwegen	1244	1148	63	–	17	17
Dänemark, Nordsee	328	326	2	–	–	–
Vereinigtes Königreich	65	22	34	–	9	–
Insgesamt	**28780**	**14542**	**272**	**11042**	**48**	**2876**

Wir wählen einen Passagier, der 2010 eine Seefahrt von oder zu einem deutschen Hafen unternommen hat, zufällig (im Sinn eines Laplace-Wahrscheinlichkeitsraums) aus.

(1) Wie hoch ist die Wahrscheinlichkeit, dass der Passagier zwischen Deutschland und Norwegen gereist ist? Antwort: $\frac{1244}{28780} = 4.3\%$
(2) Wie hoch ist die Wahrscheinlichkeit, dass der Passagier seine Reise an einem dänischen Ostseehafen begann oder endete, unter der Annahme, dass er in Schleswig-Holstein ein- oder ausgestiegen ist? Antwort: $\frac{6362}{14542} = 43.7\%$
(3) Wie hoch ist die Wahrscheinlichkeit, dass der Passagier in Hamburg ein- oder ausgestiegen ist, unter der Annahme, dass die Fahrt im Vereinigten Köngireich begann oder endete? Antwort: $\frac{34}{65} = 52.3\%$
(4) Wie hoch ist die Wahrscheinlichkeit, dass die Fahrt des Passagiers in Niedersachsen begann und endete? Antwort: $\frac{11031}{28780} = 38.3\%$
(5) Wie hoch ist die Wahrscheinlichkeit, dass der Passagier in Bremen/Bremerhaven ein- oder ausgestiegen ist, unter der Annahme, dass die Fahrt außerhalb Europas begann oder endete? Antwort: $\frac{47-44}{11972-11959} = \frac{3}{13} = 23.1\%$

△

BEISPIEL 2.7. Beim gleichzeitigen Wurf zweier unabhängiger Laplace-Würfel (also ohne Beachtung der Reihenfolge) gibt es 21 mögliche Wurfpaare. Die relative Häufigkeit eines Pasches unter allen möglichen Paaren ist also $\frac{6}{21} = \frac{2}{7} \cong 29\%$. Die Wahrscheinlichkeit, einen Pasch zu würfeln ist aber $\frac{1}{6} \cong 17\%$. Dies sehen wir wie folgt: Nach Annahme sind beide Würfel Laplace-Würfel und unabhängig voneinander. Werfen wir sie nacheinander, so ist damit der Ereignisraum

$$\Omega := \{(a,b)|a,b \in \{1,\dots,6\}\}$$

ein Laplace-Wahrscheinlichkeitsraum, wobei a das Ergebnis des ersten Würfels und b das des zweiten Würfels ist. Das heißt,

$$P((a,b)) = \frac{1}{36}$$

für alle a, b. Vergessen wir nun die Reihenfolge, das heißt, anstelle des Tupels (a,b) betrachten wir die Menge $\{a,b\}$, so erhalten wir

$$P(\{a,b\}) = \begin{cases} P(\{(a,b),(b,a)\} = \frac{1}{18} & \text{, falls } a \neq b \\ P((a,a)) = \frac{1}{36} & \text{, falls } a = b \end{cases}$$

Damit ist die Wahrscheinlichkeit für einen Pasch gleich $6 \cdot \frac{1}{36} = \frac{1}{6}$.

Bevor man also relative Häufigkeiten mit Wahrscheinlichkeiten gleichsetzen kann, muss man prüfen, ob wirklich alle Elementarereignisse gleich wahrscheinlich sind. △

3. Zufallsvariablen

Oft interessiert man sich nicht für die einzelnen Elementarereignisse, sondern für eine Messgröße, die den Elementarereignissen zugeordnet wird, beim Würfeln z.B. die Würfelsumme. Ähnlich interessieren wir uns bei Gesteinsproben vielleicht nicht für die einzelne Gesteinsprobe, sondern nur für seinen Goldgehalt. Dies führt uns zum Begriff der Zufallsvariablen.

DEFINITION 2.8 (Zufallsvariable). *Sei Ω die endliche Menge der Elementarereignisse. Eine Zufallsvariable ist eine Abbildung $X : \Omega \to \mathbb{R}$.*

Die Annahme, dass Zufallsvariablen reellwertig sind, ist vernünftig, da beobachtete Größen gewöhnlich durch reelle Zahlen (oder Tupel reeller Zahlen) beschrieben werden.

Für eine Zufallsvariable $X : \Omega \to \mathbb{R}$ ist die Wahrscheinlichkeit, dass X einen Wert in einer Teilmenge $S \subseteq \mathbb{R}$ annimmt gegeben durch

$$P(X \in S) := P(\{a \in \Omega \mid X(a) \in S\}).$$

Analog ist die Wahrscheinlichkeit, dass X den Wert $r \in \mathbb{R}$ annimmt gegeben durch

$$P(X = r) := P(\{a \in \Omega \mid X(a) = r\}).$$

Im eingangs angeführten Beispiel wäre Ω die Menge aller Gesteinsproben, $X(a)$ der Goldgehalt (z.B. in Milligramm) der Gesteinsprobe a und S der Milligrammbereich, in dem uns der Goldgehalt interessiert.

Die Unterscheidung zwischen dem Wahrscheinlichkeitsraum und der Zufallsvariablen ist etwas künstlich. Dasselbe Experiment kann verschieden modelliert werden.

BEISPIEL 2.9. Ein Würfel wird zweimal nacheinander geworfen. Wir interessieren uns a) für die Augensumme der beiden Würfe und b) für die Differenz 1. Wurf – 2. Wurf. Konkret wollen wir dann die Wahrscheinlichkeit errechnen, dass die Augensumme 4 beträgt. Folgende zwei Modelle betrachten wir:

(1) $\Omega = \{(a_1, a_2) | a_1, a_2 \in \{1, \ldots, 6\}\}$ mit $P((a_1, a_2)) = \frac{1}{36}$. Der Wahrscheinlichkeitsraum Ω besteht aus 36 gleich wahrscheinlichen Elementarereignissen. Im Fall a) betrachten wir $X_a((a_1, a_2)) := a_1 + a_2$. Im Fall b) ist die Zufallsvariable $X_b((a_1, a_2)) := a_1 - a_2$. Die gesuchte Wahrscheinlichkeit ist $P(X_a = 4) = P(\{(1,3), (2,2), (3,1)\}) = \frac{3}{36} = \frac{1}{12}$.
(2) $\tilde{\Omega} := \{\{a_1, a_2\} | a_1, a_2 \in \{1, \ldots, 6\}\}$ mit

$$P(\{a_1, a_2\}) = \begin{cases} \frac{1}{18} & \text{, falls } a_1 \neq a_2 \\ \frac{1}{36} & \text{, falls } a_1 = a_2. \end{cases}$$

Die Zufallsvariable im Fall a) ist

$$\tilde{X}_a(\{a_1, a_2\}) := \begin{cases} a_1 + a_2 & \text{, falls } a_1 \neq a_2 \\ 2a_1 & \text{, falls } a_1 = a_2. \end{cases}$$

Der Fall b) lässt sich hier nicht modellieren: Wir haben bei der Wahl des Ereignisraumes $\tilde{\Omega}$ die Reihenfolge der Würfe vergessen, also können wir nicht mehr feststellen, ob wir $a_1 - a_2$ oder $a_2 - a_1$ wissen wollen, wenn $a_1 \neq a_2$.

Die gesuchte Wahrscheinlichkeit ist in diesem Setting

$$P(\tilde{X}_a = 4) = P(\{1,3\}, \{2\}) = \frac{1}{18} + \frac{1}{36} = \frac{1}{12},$$

natürlich dieselbe wie im Fall zuvor.

Wir lernen daraus: Man muss darauf achten, dass das Modell auch noch alle Informationen beinhaltet, die wir später benötigen. Die Wahl des Modells sollte außerdem so erfolgen, dass einerseits die Wahrscheinlichkeiten der Elementarereignisse leicht berechenbar sind (spricht hier für das erste Modell), andererseits die Abzählung der für uns günstigen Elementarereignisse einfach bleibt (spricht etwas für das zweite Modell). Am Ende sollte man ein Modell wählen, mit dem man sich am sichersten fühlt. △

BEISPIEL 2.10. Wir betrachten erneut das 5-malige Würfeln. Es sei $X : \Omega \to \mathbb{R}$ die Zufallsvariable, die einem Elementarereignis $a = (a_1, \ldots, a_5)$ die Summe $a_1 + \ldots + a_5$ zuordnet. Die möglichen Werte der Zufallsvariable sind $5, 6, 7, \ldots, 30$.

Wir bestimmen die Wahrscheinlichkeit $P(X = 8)$. Sei also A das Ereignis, das aus den Elementarereignissen besteht, für die die Zufallsvariable den Wert 8 annimmt. Wir suchen also alle Tupel

$(a_1, \ldots, a_5)$ mit $1 \leq a_i \leq 6$, so dass $a_1 + \ldots + a_5 = 8$. Unter diesen Elementarereignissen gibt es drei Typen:

(1) Es werden 3 Zweier und 2 Einer gewürfelt. Davon gibt es $\binom{5}{2} = 10$ Elementarereignisse.
(2) Es werden 1 Dreier, 1 Zweier und 3 Einer gewürfelt. Davon gibt es $5 \cdot 4 = 20$ Elementarereignisse.
(3) Es werden 1 Vierer und 4 Einer gewürfelt. Dazu gibt es 5 Elementarereignisse.

Insgesamt besteht das Ereignis A also aus 35 Elementarereignissen. Es folgt, dass

$$P(X = 8) = P(A) = 35 \cdot \frac{1}{6^5} \approx 0,004501029 = 0,4501029\%.$$

$\triangle$

4. Binomialverteilung

Hier beschäftigen wir uns mit einem Experiment, das zwei mögliche Ausgänge hat, und mehrfach hintereinander ausgeführt wird. Ferner machen wir die Annahme, dass die Ausgänge der Einzelexperimente unabhängig voneinander sind.

Wir haben also für das Einzelexperiment einen Ereignisraum $\bar{\Omega} = \{0, 1\}$ und eine Wahrscheinlichkeitsverteilung P mit $P(1) = q$ und $P(0) = 1 - q$ wobei $q \in [0, 1]$. Führen wir das Experiment n-fach aus, so erhalten wir einen Ereignisraum $\Omega = \{(a_1, \ldots, a_n) \,|\, a_i \in \{0, 1\} \text{ für } 1 \leq i \leq n\}$. Ferner folgt aus der Unabhängigkeit der Teilexperimente, dass

$$P(a_1, \ldots, a_n) = P(a_1) \cdot \ldots \cdot P(a_n) = q^k (1 - q)^{n-k}$$

wobei k die Anzahl der a_i ist, für die $a_i = 1$ gilt.

Wir betrachten nun die Zufallsvariable X, die einem gegebenem Elementarereignis $(a_1, \ldots, a_n)$ die Anzahl der a_i zuordnet, für die $a_i = 1$ gilt. Es gilt also $X(a_1, \ldots, a_n) = a_1 + \ldots + a_n$.

Wir bestimmen $P(X = k)$. Sei A die Menge der Elementarereignisse $a \in \Omega$, für die gilt $X(a) = k$. Wir haben bereits gesehen, dass die Wahrscheinlichkeit eines solchen Ereignisses $q^k(1 - q)^{n-k}$ ist. Die Anzahl dieser Ereignisse ist $\binom{n}{k}$, da es $\binom{n}{k}$ Möglichkeiten gibt, wie die k 1-Einträge auftreten können. Wir haben also den folgenden Satz:

SATZ 2.11. *(Binomialverteilung) In obiger Situation gilt* $P(X = k) = \binom{n}{k} \cdot q^k (1 - q)^{n-k}$.

BEISPIEL 2.12. Wir bestimmen die Wahrscheinlichkeit, beim 10-maligen Würfeln genau 8 mal eine 5 oder 6 zu würfeln. Die Wahrscheinlichkeit, beim einmaligen Würfeln eine 5 oder 6 zu würfeln ist $q = \frac{1}{3}$. Nach Satz 2.11 ist die Wahrscheinlichkeit also $\binom{10}{8} \cdot \left(\frac{1}{3}\right)^8 \cdot \left(1 - \frac{1}{3}\right)^2 = 45 \cdot \frac{1}{3^8} \cdot \frac{4}{9} = 0,003048316 = 0,3048316\%$. $\triangle$

5. Beispiele

BEISPIEL 1: 20% aller Gesteinsproben enthalten Gold. Wie groß ist die Wahrscheinlichkeit, dass bei einer Stichprobe von 10 Gesteinsproben 8 oder mehr Gold enthalten?

Antwort: Wir betrachten die Binomialverteilung mit $n = 10$, $q = 0,2$ und bestimmen $P(X \geq 8) = P(X \in \{8, 9, 10\}) = P(X = 8) + P(X = 9) + P(X = 10)$. Es gilt nach Satz 2.11, dass $P(X = 8) = \binom{10}{8} \cdot 0,2^8 \cdot 0,8^2 = 0,000073728$, $P(X = 9) = \binom{10}{9} \cdot 0,2^9 \cdot 0,8^1 = 0,000004096$ und $P(X = 10) = \binom{10}{10} \cdot 0,2^{10} \cdot 0,8^0 = 0,2^{10} = 0,000000102$ und somit $P(X \geq 8) = 0,00007373 + 0,000004096 + 0,000000102 = 0,000077928 = 0,0077928\%$.

BEISPIEL 2: Betrachte eine Menge verheirateter Paare. Sei A das Ereignis, dass ein verheirateter Mann nach 15 Jahren noch lebt, B das Ereignis, dass seine Frau nach 15 Jahren noch lebt. Es seien A und B unabhängige Ereignisse mit $P(A) = \frac{2}{5}$ und $P(B) = \frac{3}{7}$. Wie groß ist die Wahrscheinlichkeit, dass nach 15 Jahren

(1) beide noch leben? Wir suchen also $P(A \cap B)$.
(2) mindestens einer von beiden noch lebt? Wir suchen also $P(A \cup B)$.
(3) nur noch die Frau lebt?

Antwort: Es gibt eine Ereignismenge $\Omega = \{(0,0), (1,0), (0,1), (1,1)\}$ wobei $(0,0)$ bedeutet, dass beide sterben, $(1,1)$, dass beide überleben, $(1,0)$, dass nur der Mann überlebt und $(0,1)$, dass nur die Frau überlebt. Es gilt $P(1,1) = \frac{2}{5} \cdot \frac{3}{7} = \frac{6}{35}$, da die Ereignisse unabhängig sind. Analog gilt $P(0,0) = \frac{3}{5} \cdot \frac{4}{7} = \frac{12}{35}$, $P(1,0) = \frac{2}{5} \cdot \frac{4}{7} = \frac{8}{35}$ und $P(0,1) = \frac{3}{5} \cdot \frac{3}{7} = \frac{9}{35}$. Wir erhalten folgende Antworten:

(1) Es gilt $P(1,1) = \frac{6}{35}$.
(2) Es gilt $P(A \cup B) = P(\{(1,1), (0,1), (1,0)\}) = \frac{6}{35} + \frac{8}{35} + \frac{9}{35} = \frac{23}{35}$. Alternativ kann man hier sagen, dass $A \cup B$ das Gegenereignis von $\{(0,0)\}$ ist und somit gilt $P(A \cup B) = 1 - P(0,0) = 1 - \frac{12}{35} = \frac{23}{35}$.
(3) Es gilt $P(0,1) = \frac{9}{35}$.

6. Aufgaben

Aufgabe 2-1 Eine Münze wird 3 mal geworfen und liefert jedes mal Kopf (K) oder Zahl (Z). Die Menge der Elementarereignis Ω besteht also aus Tripeln mit Einträgen aus $\{K, Z\}$. Sei A das Ereignis, dass beim ersten Wurf Z auftritt, B das Ereignis, dass beim zweiten Wurf Z auftritt und C das Ereignis, dass beim dritten Wurf Z auftritt.

(1) Man schreibe die Ereignisse A, B, C, $A \cap B$, $A \cap B \cap C$ als Teilmengen von Ω und gebe ihre Wahrscheinlichkeiten an (wir machen die Annahme, dass alle Elementarereignisse gleich wahrscheinlich sind).
(2) Sind A und B unabhängig? Gleiches für A und C bzw. für B und C.
(3) Sind A, B und C unabhängig, gilt also $P(A \cap B \cap V) = P(A) \cdot P(B) \cdot P(C)$?

Aufgabe 2-2 Wir würfeln mit 2 Würfeln. Sei X die Zufallsvariable, die jedem Ergebnis die Augensumme zuordnet. Bestimmen Sie für jede mögliche Augensumme die Wahrscheinlichkeit, mit der sie auftritt. Sie dürfen, wenn Ihnen das hilft, eine Symmetrieüberlegung in Ihre Lösung einfließen lassen.

Aufgabe 2-3 Bei einer Wahl stimmen 55% für Partei A und 45% für Partei B. Wie groß ist die Wahrscheinlichkeit, dass Sie nach einer Befragung von 13 Wählern vor der Wahl, das Wahlergebnis falsch vorausgesagt hätten?

Aufgabe 2-4 Ein Medikament verursacht bei 2% aller Patienten gravierende Nebenwirkungen. Wie groß ist die Wahrscheinlichkeit, dass bei einem Versuch mit 50 Patienten bei keinem Patienten Nebenwirkungen auftreten?

Aufgabe 2-5 Die Wahlergebnisse der vorletzten Landtagswahl in Schleswig-Holstein 2005[2], aufgeschlüsselt nach einigen Bevölkerungsgruppen[3], waren

[2]Bei der Landtagswahl 2009 wurden keine repräsentativen Wahlumfragen durchgeführt

[3]Quelle: Bundeswahlleiter

Alter von ... bis unter ... Jahren	Stimmabgabe in %																	
	Insgesamt						Männer						Frauen					
	SPD	CDU[1)]	FDP	GRÜNE	DIE LINKE	Sonstige	SPD	CDU[1)]	FDP	GRÜNE	DIE LINKE	Sonstige	SPD	CDU[1)]	FDP	GRÜNE	DIE LINKE	Sonstige
	Schleswig-Holstein 20.02.2005[2)]																	
18 - 25	37,7	28,9	8,3	11,3		13,8	35,0	29,5	9,1	10,7		15,6	40,7	28,2	7,3	11,9		11,8
25 - 35	32,9	36,7	8,8	8,1		13,5	29,7	38,4	10,9	7,3		13,6	36,0	35,0	6,8	8,9		13,3
35 - 45	38,8	33,9	6,8	9,5	3)	11,1	35,8	36,0	7,5	8,6	3)	12,1	41,8	31,9	6,0	10,3	3)	10,0
45 - 60	43,3	34,2	6,4	7,6		8,6	40,6	35,4	6,6	7,6		9,8	45,9	32,9	6,2	7,7		7,2
60 und mehr	37,8	48,5	5,5	2,2		6,0	34,8	49,8	6,0	2,3		7,0	40,5	47,3	5,1	2,1		5,1
Insgesamt	39,0	39,5	6,5	6,2		8,9	36,2	40,6	7,1	6,1		10,0	41,8	38,4	5,8	6,3		7,7

1) In Bayern CSU. – 2) Bei der Landtagswahl 2009 wurde keine repräsentative Wahlstatistik durchgeführt. – 3) In den Sonstigen enthalten.

Es sei (alles zum Stichtag der Landtagswahl 2005)

$$\Omega := \{\text{Wahlberechtigte in Schleswig-Holstein}\}$$

unser Ereignisraum. Die relativen Häufigkeiten der Tabelle interpretieren wir als Wahrscheinlichkeiten. Weiter betrachten wir die Ereignisse[4]

$$A := \{\text{Frauen in } \Omega\},\ B := \{\text{18–25-jährige in } \Omega\},$$
$$C := \{\text{Wähler der Partei } X \text{ in } \Omega\},\ D := \{\text{Wähler in } \Omega\}$$

(1) Stellen Sie die folgenden Größen als bedingte Wahrscheinlichkeiten dar (z.B. wäre die Wahlbeteiligung unter den 18–25-jährigen Wahlberechtigten $P(D|B)$):
   (a) Anteil der der 18–25-jährigen unter den wahlberechtigten Frauen
   Lösung: $P(B|A)$
   (b) Anteil der Wähler der Partei X unter den 18–25-jährigen wahlberechtigten Frauen
   Lösung: $P(C|A \cap B)$
   (c) Wahlbeteiligung unter den 18–25-jährigen wahlberechtigten Männern
   Lösung: $P(D|A^c \cap B)$
   (d) Anteil der Frauen unter allen Wählern
   Lösung: $P(A|D)$
(2) Zeigen Sie für beliebige Ereignisse $A, B, C \subseteq \Omega$ die Gültigkeit der Gleichung

$$P(A|B) \cdot P(C|A \cap B) + (1 - P(A|B)) \cdot P(C|A^c \cap B) = P(C|B).$$

Lösung: $P(A|B) \cdot P(C|A \cap B) + (1 - P(A|B)) \cdot P(C|A^c \cap B) = \frac{P(A\cap B)}{P(B)} \cdot \frac{P(A\cap B\cap C)}{P(A\cap B)} + \frac{P(A^c\cap B)}{P(B)} \cdot \frac{P(A^c\cap B\cap C)}{P(A^c\cap B)} = \frac{P(A\cap B\cap C)+P(A^c\cap B\cap C)}{P(B)} = \frac{P(B\cap C)}{P(B)} = P(C|B)$.
(3) Geben Sie die fünf Werte für den Anteil der Frauen unter den 18–25-jährigen Wählern an, die sie mit Hilfe der Gleichung aus (2) und dem sukzessiven Einsetzen von SPD, CDU, FDP, Grüne und Sonstige anstelle von X erhalten. Verwenden Sie dafür $\Omega' := D, A' := A \cap D, B' := B \cap D, C' := C$ und die Gleichung aus (2) für Ω', A', B', C' anstelle von Ω, A, B, C. (Der Unterschied der Zahlen kommt von der für diesen Zweck relativ großen Ungenauigkeit der Wahlergebnisse.)
Lösung: Gesucht ist offenbar $P(A'|B')$ *und durch die Tabelle sind gegeben* $P(C'|B'), P(C'|A' \cap B')$ *und* $P(C'|(A')^c \cap B')$. *Wir lösen also die Gleichung aus (2) nach* $P(A'|B')$ *auf und erhalten*

$$P(A'|B') = \frac{P(C'|B') - P(C'|(A')^c \cap B')}{P(C'|A' \cap B') - P(C'|(A')^c \cap B')}.$$

[4] "Wähler" ist hier geschlechtsneutral zu verstehen.

Nun setzen wir für X in C' die verschiedenen Parteien und Sonstige ein und erhalten beispielsweise für $X = SPD$

$$P(A'|B') = \frac{37{,}7 - 35{,}0}{40{,}7 - 35{,}0} \sim 47{,}4\%.$$

Der errechnete Wert ist um so genauer, je unterschiedlicher das Wahlverhalten zwischen Männern und Frauen bei Partei X war. Aller Resultate sind[5]

X =	*geschätztes $P(A'\|B')$*	*$P(A'\|B')$ sicher im Intervall*
SPD	*47,4 %*	*44,8–50,0 %*
CDU	*46,2 %*	*35,7–58,3 %*
FDP	*44,4 %*	*36,8–52,9 %*
Grüne	*50,0%*	*38,5–63,6%*
Sonstige	*47,4%*	*43,6–51,4%*

Der wahre Wert muss im Schnitt aller angegebenen Intervalle liegen, was mit dem erstgenannten, 44,8–50,0% übereinstimmt. Unsere Fehlerabschätzung ist aber sehr vorsichtig. Wir würden also erwarten, dass in der Altersgruppe 18–25 Jahre weniger Frauen als Männer gewählt haben. In der Tat war nach weiteren Daten des Bundeswahlleiters die Wahlbeteiligung bei Männern dieser Altersgruppe bei etwa 46%, bei den Frauen gleichen Alters etwa 42%. Dies würde einen Wert für $P(A'|B')$ von etwa 47–48% bedeuten.

(4) Welche der folgenden Ereignispaare sind (im Rahmen der Tabellengenauigkeit) in den angegebenen Ereignisräumen stochastisch unabhängig?
   (a) In {weibliche Wähler}: {SPD-Wählerinnen} und {35–45-jährige}
   Lösung: $P(SPD|weiblich) = 0.418 = P(SPD|weiblich\ und\ 35\text{–}45)$, also ja, im angegebenen Ereignisraum sind die beiden Ereignisse stochastisch unabhängig.
   (b) In {25–35-jährige Wähler}: {CDU-Wähler} und {männliche Wähler}
   Lösung: $P(CDU|25\text{–}35) = 0.367 \neq 0.384 = P(CDU|25\text{–}35\ und\ männlich)$, also nicht stochastisch unabhängig.
   (c) In {Wähler}: {60-jährige und ältere Wähler} und {FDP-Wähler}
   Lösung: $P(FDP) = 0.065 \neq 0.055 = P(FDP|60\ und\ älter)$, also nicht stochastisch unabhängig.
   (d) In {45–60-jährige Wähler}: {Grünen-Wähler} und {männliche Wähler}
   Lösung: $P(Grüne|45\text{–}60) = 0.076 = P(Grüne|45\text{–}60\ und\ männlich)$, also ja, stochastisch unabhängig.

Aufgabe 2-6 Im Jahr 2011 wurden von deutschen Häfen aus etwa 565 Millionen Passagiere per Schiff befördert. Davon stiegen 51 Millionen in Kiel ein.[6] Wie hoch ist die Wahrscheinlichkeit, dass unter 10 aus den 565 Millionen zufällig ausgewählten Passagieren mindestens zwei in Kiel an Bord gegangen sind?

Aufgabe 2-7 Im Jahr 2011 wurden in Deutschland folgende Weinmengen erzeugt (in Millionen Liter)[1]:

	Weißwein	Rotwein	insgesamt
Tafel-/Landwein	26,5	4,5	31,0
Qualitätswein	514,8	367,5	882,3
insgesamt	541,3	372,0	913,3

Eine Weinflasche wird zufällig ausgewählt.

(1) Wie hoch ist die Wahrscheinlichkeit, dass die Weinflasche Tafel- oder Landwein enthält?

[5] Jeder Häufigkeitswert ist auf 0.05 genau, so dass Zähler und Nenner der obigen Formel jeweils auf 0.1 genau sind.
[6] Quelle: Destatis

(2) Wie hoch ist diese Wahrscheinlichkeit unter der Bedingung, dass es sich um Weißwein handelt?
(3) Wie hoch ist diese Wahrscheinlichkeit unter der Bedingung, dass es sich um Rotwein handelt?
(4) Wie hoch ist die Wahrscheinlichkeit, dass man Weißwein in der Hand hält, wenn man nur aus Qualitätsweinen gewählt hat?
(5) Wie hoch ist die Wahrscheinlichkeit, dass man Weißwein in der Hand hält, wenn man nur aus Tafel-/Landweinen gewählt hat?
(6) Wie hoch ist die Wahrscheinlichkeit, dass man nach zufälliger Auswahl von 11 Flaschen Qualitätswein den Anteil von Rotwein daran höher einschätzt als den von Weißwein?

Aufgabe 2-8 Im Jahr 2010 wurden an deutschen Hochschulen folgende Anzahlen an Abschlussprüfungen abgehalten[7]:

	Bestanden	Nicht bestanden	insgesamt
männlich	175928	6826	182754
weiblich	185769	3260	189029
insgesamt	361697	10086	371783

Eine Person, die 2010 eine Abschlussprüfung an einer deutschen Hochschule hatte, wird zufällig ausgewählt.

(1) Wie hoch ist die Wahrscheinlichkeit, dass die Person ihre Abschlussprüfung nicht bestanden hat?
(2) Wie hoch ist diese Wahrscheinlichkeit unter der Bedingung, dass die Person ein Mann ist?
(3) Wie hoch ist diese Wahrscheinlichkeit unter der Bedingung, dass die Person eine Frau ist?
(4) Wie hoch ist die Wahrscheinlichkeit, dass die Person eine Frau ist, wenn man weiß, dass sie bestanden hat?
(5) Wie hoch ist die Wahrscheinlichkeit, dass die Person eine Frau ist, wenn man weiß, dass sie nicht bestanden hat?
(6) Wie hoch ist die Wahrscheinlichkeit, dass man nach zufälliger Auswahl von 11 Personen, die bestanden haben, den Männeranteil an den Personen mit bestandener Prüfung höher einschätzt als den Frauenanteil?

[7] Quelle: Destatis

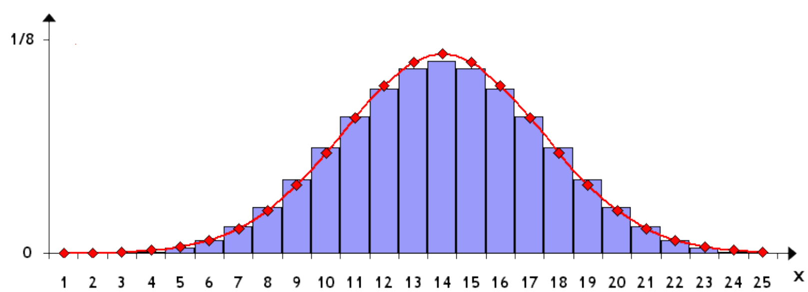
KAPITEL 3

Die Normalverteilung

Die Verteilungen vieler in der Natur vorkommenden Zufallsvariablen sind normalverteilt, oder annähernd normalverteilt. Dies liegt nicht zuletzt am zentralen Grenzwertsatz.

1. Die Normalverteilung mit Parametern μ und σ

Betrachtet man eine graphische Darstellung einer Binomialverteilung mit hinreichend großem n, so sieht man eine charakteristische Form. Es entsteht eine Approximation einer symmetrischen Glockenkurve. Hier für n=27 und $p = \frac{1}{2}$.



Diese Form hängt weder von n noch von p ab, lediglich die Höhe und Breite variiert. Obwohl die Symmetrie der Kurve in dem Beispiel daher kommt, dass $p = \frac{1}{2}$ ist, ist der relevante Teil der Kurve für hinreichend großes n stets fast symmetrisch, hier der Falle $n = 48$ und $p = 0,25$

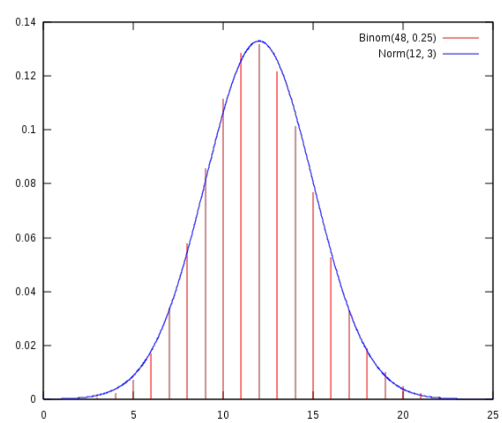


ABBILDUNG 3.1. GNUplot von Vasiliev Mihail, Lizenz: cc-by-sa-3.0-migrated

Man sieht an diesen Beispielen auch, dass der wahrscheinlichste Wert der Binomialverteilung bei $p \cdot n$ liegt, diese Größe werden wir noch als den Erwartungswert kennenlernen.

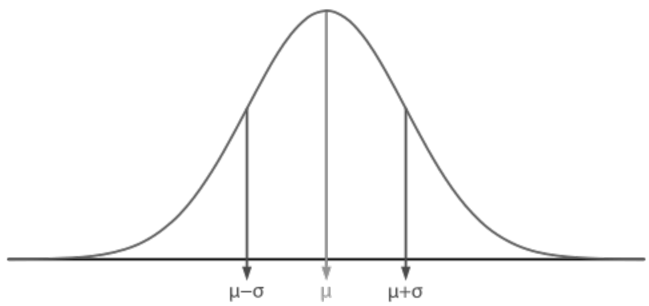
Es stellt sich heraus, dass die Kurven, die von den Binomialverteilungen approximiert werden, explizit angegeben werden können, sie hängen von zwei Parametern μ und σ ab.

DEFINITION 3.1. *Sei $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_{>0}$. Dann heißt die Funktion (bzw. ihr Graph)*

$$f_{\mu,\sigma} : \mathbb{R} \to \mathbb{R},\ f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

die Gaußsche Glockenkurve zu den Parametern μ und σ.

Die Kurve hat die folgende Form:



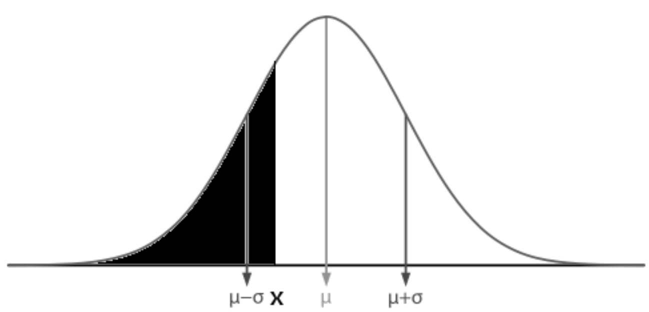
Der Flächeninhalt unter der Kurve ist 1; dies wird uns im folgenden erlauben, diese Kurve mit der Wahrscheinlichkeitsverteilung einer Zufallsvariablen in Verbindung zu bringen. Dazu führen wir zunächst die folgende Funktion ein:

DEFINITION 3.2. *Sei $f_{\mu,\sigma}$ wie oben. Dann Definieren wir*

$$F_{\mu,\sigma} : \mathbb{R} \to \mathbb{R},\ F_{\mu,\sigma}(x) := \int_{-\infty}^{x} f_{\mu,\sigma}(t)dt.$$

Wir bezeichnen die Funktion $F_{0,1}$ schlicht mit ϕ.

Der Wert $F_{\mu,\sigma}(x)$ misst also den Flächeninhalt unter dem Graph von $f_{\mu,\sigma}$ zwischen $-\infty$ und x:



Insbesondere ist also der Flächeninhalt unterhalb der Kurve zwischen x_1 und x_2 nichts anderes als $\int_{x_1}^{x_2} f_{\mu,\sigma}(t)dt = F_{\mu,\sigma}(x_2) - F_{\mu,\sigma}(x_1)$. Wir müssen die Funktion $F_{\mu,\sigma}$ nicht für alle μ und σ kennen:

SATZ 3.3. *Für alle $x \in \mathbb{R}$ gilt $F_{\mu,\sigma}(x) = \phi\left(\frac{x-\mu}{\sigma}\right)$.*

Die Bedeutung dieser Funktion folgt aus der Tatsache, dass viele Zufallsvariablen normalverteilt sind im folgende Sinne. Anders als zuvor ist hier der Ereignisraum Ω nicht endlich.

DEFINITION 3.4. *Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Wir sagen, dass X eine (μ, σ)-normalverteilte Zufallsvariable ist, wenn für alle $x_1, x_2 \in \mathbb{R} \cup \{\infty, -\infty\}$ mit $x_1 \le x_2$ gilt, dass*

$$P(X \in [x_1, x_2]) = \int_{x_1}^{x_2} f_{\mu,\sigma}(t)dt.$$

Insbesondere folgt also aus der obigen Diskussion und Satz 3.3, dass für eine (μ, σ)-normalverteilte Zufallsvariable X gilt, dass

$$P(X \in [x_1, x_2]) = \phi\left(\frac{x_2 - \mu}{\sigma}\right) - \phi\left(\frac{x_1 - \mu}{\sigma}\right).$$

Die Funktion ϕ ist nicht leicht zu berechnen. Früher hat man dazu meist Tabellen genommen, heute können viele Taschenrechner die Funktion ϕ berechnen.

Ist nun eine Zufallsvariable X (μ, σ)-normalverteilt, so gilt

(1) $P(X \in [\mu - \sigma, \mu + \sigma]) = \phi(1) - \phi(-1) = 0,8413 - 0,1587 = 0,6826.$
(2) $P(X \in [\mu - 2\sigma, \mu + 2\sigma]) = \phi(2) - \phi(-2) = 0,9772 - 0,0228 = 0,9544.$
(3) $P(X \in [\mu - 3\sigma, \mu + 3\sigma]) = \phi(3) - \phi(-3) = 0,9987 - 0,0013 = 0,9974.$

2. Beispiele

BEISPIEL 1: Die Lufttemperatur (gemessen um 12.00 Uhr am 15. Juni) ist von Jahr zu Jahr verschieden. Wir nehmen an, sie ist (μ, σ)-normalverteilt mit $\mu = 20°$ und $\sigma = 3°$.

(1) Wie groß ist die Wahrscheinlichkeit, dass sie am 15. Juni eines Tages zwischen 11° und 17° liegt?
(2) Wie groß ist die Wahrscheinlichkeit, dass die Temperatur höher als 25° ist?

Antwort: (1) Es gilt $P(X \in [11, 17]) = \phi(\frac{17-20}{3}) - \phi(\frac{11-20}{3}) = \phi(-1) - \phi(-3) = 0,1587 - 0,0013 = 0,1574$.

(2) Es gilt $P(X \ge 25) = 1 - P(X \le 25) = 1 - \phi(\frac{25-20}{3}) = 1 - \phi(1,6666) = 1 - 0,9521 = 0,0479$.

BEISPIEL 2: In einem Wald stehen 10000 Bäume. Ihre Größe ist (μ, σ)-normalverteilt mit $\mu = 30m$ und $\sigma = 5m$. Es sollen die 2000 kleinsten Bäume gefällt werden. Bis zu welcher Größe müssen die Bäume gefällt werden?

Antwort: Sei y die gesuchte Größe. Es gilt also $0,2 = P(X \le y) = \phi(\frac{y-30}{5})$. Es gilt also $\frac{y-30}{5} = -0,84$ und somit $y - 30 = -4,2$, also $y = 25,8$.

3. Aufgaben

Aufgabe 3-1 Die Verteilung der Tablettenmassen einer Charge sei (μ, σ)-normalverteilt mit $\mu = 250mg$ und $\sigma = 3mg$. Wie viel Prozent der Tabletten besitzen eine Masse zwischen $245mg$ und $252mg$?

Aufgabe 3-2 Die Größenverteilung 4-jähriger Kinder wird als normalverteilt vorausgesetzt. Es sind 3% der Kinder größer als $111cm$ und 3% kleiner als $96cm$.

(1) Bestimmen Sie die Parameter μ und σ der Normalverteilung.
(2) Wieviel Prozent der Vierjährigen sind größer als $99cm$?
(3) Wieviel Prozent der Vierjährigen sind zwischen $99cm$ und $102cm$ groß?

Aufgabe 3-3 Der Niederschlag im Monat Juli ist von Jahr zu Jahr verschieden. Wir nehmen an, er ist normalverteilt mit $\mu = 83mm$ und $\sigma = 30mm$. Berechnen Sie die Wahrscheinlichkeit, dass mehr als $150mm$ Regen fällt.

Aufgabe 3-4 Der tägliche Wasserverbrauch (ohne indirekten durch Konsum) betrug durchschnittlich 121 Liter pro Person in Deutschland im Jahr 2010[1]. Wir nehmen daher an, dass der Wasserverbrauch (μ, σ)-normalverteilt ist mit $\mu = 121l$. Schätzungsweise 14% der Bevölkerung verbrauchten durchschnittlich weniger als 100 Liter pro Tag.

(1) Bestimmen Sie σ.
(2) Bestimmen Sie die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person durchschnittlich mehr als 135 Liter pro Tag verbraucht.

4. Quellen

Die gemeinfreien Grafiken sind den folgenden Webseiten entnommen:

(1) http://sites.google.com/site/approxbvnv/
(2) http://commons.wikimedia.org/wiki/File:Gaussian_distribution.svg

[1] Quelle: spiegel.de

KAPITEL 4

Erwartungswert, Varianz und zentraler Grenzwertsatz

1. Erwartungswert und Varianz

Im Folgenden sei X eine Zufallsvariable. Unser erstes Ziel ist es, den Erwartungswert zu definieren. Wie das Wort bereits sagt, soll der Erwartungswert eine Größe sein, die uns sagt, welchen Wert einer Zufallsvariable wir erwarten, also eine Art Durchschnittswert. Werfen wir z.B. eine Münze mit Kopf=0 und Zahl=1, so erwarten wir mit gleicher Wahrscheinlich keit (nämlich 0,5) 0 und 1, im Mittel erwarten wir also 0,5. Dieses Beispiel zeigt bereits, dass der Erwartungswert nicht immer ein Wert ist, der von der Zufallsvariablen angenommen wird.

Wir geben zunächst die Definition des Erwartungswertes im Falle einer Zufallsvariable X, die nur endlich viele Werte annimmt.

DEFINITION 4.1 (Erwartungswert). *Sei X eine Zufallsvariable, die die Werte $x_1, \dots, x_n$ annimmt. Sei $p_i := P(X = x_i)$ für $1 \leq i \leq k$. Dann ist der Erwartungswert $E(X)$ von X definiert als*

$$E(X) := \sum_{i=1}^{k} p_i \cdot x_i.$$

Ist der zugrunde liegende Ereignisraum endlich, so kann man $E(X)$ auch mittels

$$E(X) = \sum_{a \in \Omega} P(a) \cdot X(a)$$

errechnen und am allgemeinsten

$$E(X) = \sum_{i=1}^{k} P(A_i) \cdot y_i,$$

wobei $X = \bigcup_{i=1}^{k} A_i$, $A_i \cap A_j = \emptyset$ für alle $i \neq j$ und $X(a) = y_i$ für alle $a \in A_i$ gelten muss.

Nimmt die Zufallsvariable unendlich viele verschiedene Werte an, so muss die Summe durch ein Integral ersetzt werden. Die entsprechende Integrationstheorie haben wir nicht zur Verfügung. Ist das Wahrscheinlichkeitsmaß jedoch wie im Falle der Normalverteilung durch eine Dichtefunktion f gegeben, so erhalten wir

$$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x \, dx.$$

Neben dem Erwartungswert spielt eine weitere Kenngröße einer Zufallsvariable X eine herausragende Rolle, nämlich die Varianz $\mathrm{Var}(X)$. Die Varianz misst die Streuung einer Zufallsvariable, d.h. wie weit die Werte um den Erwartungswert verteilt sind.

DEFINITION 4.2 (Varianz). *Sei X eine Zufallsvariable, die die Werten $x_1, \dots, x_n$ annimmt. Sei $p_i := P(X = x_i)$ für $1 \leq i \leq k$. Dann ist die Varianz $\mathrm{Var}(X)$ von X definiert als*

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

Dies können wir wie oben ausdrücken (wenn der Ereignisraum endlich ist) also

$$\mathrm{Var}(X) = \sum_{i=1}^{k} p_i \cdot (x_i - E(X))^2 = \sum_{a\in\Omega} P(a)(X(a) - E(X))^2 = \sum_{i=1}^{k} P(A_i)(y_i - E(X))^2.$$

Man nennt $\sigma(X) := \sqrt{\mathrm{Var}(X)}$ *auch die Standardabweichung von* X*.*

Bemerkung Die Varianz misst nicht genau die durchschnittliche Abweichung vom Erwartungswert; wollte man diese messen, so müsste man $(X - E(X))^2$ durch $|X - E(X))|$ ersetzten.

BEISPIELE:

(1) Es wird einmal (ein Laplace-Würfel) gewürfelt. Die Zufallsvariable X sei die Augenzahl des Wurfs. Dann ist

$$E(X) = \frac{1}{6}\sum_{a=1}^{6} a = \frac{21}{6} = 3.5$$

und

$$\mathrm{Var}(X) = \frac{1}{6}\sum_{a=1}^{6}(a-3.5)^2 = \frac{1}{24}(25+9+1+1+9+25) = \frac{70}{24} = \frac{35}{12}.$$

(2) Es wird einmal (Laplace-)gewürfelt. Die Zufallsvariable ist das Quadrat der erzielten Augenzahl. Wir berechnen den Erwartungswert und die Varianz. Es gilt

$$E(X) = \frac{1}{6}\cdot 1^2 + \frac{1}{6}\cdot 2^2 + \frac{1}{6}\cdot 3^2 + \frac{1}{6}\cdot 4^2 + \frac{1}{6}\cdot 5^2 + \frac{1}{6}\cdot 6^2 = \frac{1}{6}\cdot 91 = 15,167$$

Die Varianz ist

$$\mathrm{Var}(X) = \sum_{n=1}^{6}\frac{1}{6}\cdot(n^2 - 15,167)^2 = \frac{1}{6}\cdot 894,833 = 149,1388$$

(3) Ist X eine (μ, σ)-normalverteilte Zufallsvariable, so gilt $E(X) = \mu$ und $\mathrm{Var}(X) = \sigma^2$. Es ist also σ die Standardabweichung von X.

(4) Im Falle der Binomialverteilung X mit Parametern n und q gilt nach Satz 2.11, dass $P(X = i) = \binom{n}{i} q^i (1-q)^{n-i}$ und wir erhalten

$$E(X) = \sum_{i=0}^{n} i \cdot \binom{n}{i} q^i (1-q)^{n-i} = nq.$$

Ferner erhalten wir

$$\mathrm{Var}(X) = nq(1-q)$$

und somit die Standardabweichung von $\sqrt{nq(1-q)}$.

BEISPIEL 4.3. **(Eine Strategie braucht mehr als nur den Erwartungswert)** Wir wetten gegen einen reichen Freund bei einem Münzwurf auf Kopf oder Zahl. Der Freund ist großzügig und gibt uns den dreifachen Einsatz zurück, wenn wir gewinnen. Wir suchen nach einer guten Spielstrategie: wie oft spielen wir und was setzen wir? Der Erwartungswert des Nettogewinns pro Einsatz eines solchen Spieles ($=: X$) ist

$$E(X) = \frac{1}{2}\cdot(-1) + \frac{1}{2}\cdot 2 = \frac{1}{2} > 0.$$

Würde in unsere Strategie nur der Erwartungswert des Spiels einfließen, so wäre damit klar: wir spielen so oft, wie es geht und setzen jeweils alles. Damit endet das Spiel aber immer mit Totalverlust, wenn es endet. Dass es nach n Würfen zu Ende ist, passiert mit der Wahrscheinlichkeit $1 - (\frac{1}{2})^n$, was gegen 1 konvergiert für $n \longrightarrow \infty$. Das heißt, wir erleiden mit Wahrscheinlichkeit 1 Totalverlust. Dies stellt die Strategie zumindest in Frage. △

Wir benötigen im Folgenden den Begriff unabhängiger Zufallsgrößen:

DEFINITION 4.4. *Zwei Zufallsgrößen X, Y auf dem Wahrscheinlichkeitsraum Ω heißen unabhängig, wenn für alle $a, b \in \mathbb{R}$ die Ereignisse $X = a$ und $Y = b$ unabhängig sind, d.h., wenn*

$$P(X = a, Y = b) = P(X = a) \cdot P(Y = b).$$

Allgemeiner heißen Zufallsgrößen $X_1, \ldots, X_n$ unabhängig, wenn

$$P(X_1 = a_1, \ldots, X_n = a_n) = P(X_1 = a_1) \cdot \cdots \cdot P(X_n = a_n)$$

für alle $a_1, \ldots, a_n \in \mathbb{R}$ gilt.

Damit können wir nun Rechenregeln für Erwartungswert und Varianz notieren.

SATZ 4.5. *Sei Ω ein endlicher Wahrscheinlichkeitsraum und $X, Y, X_1, \ldots, X_n$ Zufallsgrößen auf Ω und $r \in \mathbb{R}$. Dann gilt*

(1) $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$,
(2) $E(rX) = rE(X)$,
(3) $\mathrm{Var}(rX) = r^2 \mathrm{Var}(X)$.
(4) $\mathrm{Var}(X_1 + \cdots + X_n) \leq n(\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n))$.

Sind $X_1, \ldots, X_n$ unabhängig, so gilt weiter

(1) $E(X_1 \cdot \cdots \cdot X_n) = E(X_1) \cdot \cdots \cdot E(X_n)$,
(2) $\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$.

Wenn wir von "n unabhängigen Kopien von X" sprechen, sind damit die Zufallsvariablen $X_i(a_1, \ldots, a_n) := X(a_i)$ auf dem Wahrscheinlichkeitsraum Ω^n mit $P(a_1, \ldots, a_n) := P(a_1) \cdot \cdots \cdot P(a_n)$ gemeint. Diese sind automatisch unabhängig.

BEISPIEL 4.6. Wir würfeln zweimal (wie immer einen Laplace-Würfel und unter Annahme der Unabhängigkeit der Würfe). und betrachten die Zufallsgrößen

$$X := \begin{cases} 0 & \text{, falls die Summe der Augenzahlen gerade ist} \\ 1 & \text{, falls die Summe der Augenzahlen ungerade ist} \end{cases}$$

und

$$Y := \text{Augenzahl des zweiten Wurfs.}$$

Es ist

$$P(X = 0) = \frac{1}{2}, P(X = 1) = \frac{1}{2},$$

da es bei jedem einzelnen Wurf genauso viele gerade wie ungerade Möglichkeiten gibt und die Summe genau dann ungerade ist, wenn genau ein Summand ungerade war. Der Erwartungswert von X ist also

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

und die Varianz ist

$$\mathrm{Var}(X) = \frac{1}{2}(0 - \frac{1}{2})^2 + \frac{1}{2}(1 - \frac{1}{2})^2 = \frac{1}{4}.$$

für Y kennen wir die Werte bereits: $E(Y) = 3{,}5, \mathrm{Var}(Y) = \frac{35}{12}$. Sind X und Y unabhängig? Für jeden beliebigen zweiten Wurf gibt es genau drei Möglichkeiten für den ersten Wurf so, dass die Summe gerade wird. Ebenso für ungerade Summe. Das heißt

$$P(X = 1, Y = a) = \frac{3}{6} P(Y = a) = P(X = 1) \cdot P(Y = a)$$

und

$$P(X = 0, Y = a) = \frac{3}{6} P(Y = a) = P(X = 0) \cdot P(Y = a).$$

Also sind in der Tat X und Y unabhängig und ohne weitere Rechnung wissen wir

$$E(XY) = \frac{7}{4}, \mathrm{Var}(X + Y) = \frac{19}{6}.$$

$\triangle$

2. Kovarianz und Korrelation

Die "binäre" Unterscheidung zwischen unabhängigen und abhängigen Zufallsgrößen ist etwas grob. In der Praxis hat man sowieso nur fehlerbehaftete Messwerte, so dass eine exakte Gültigkeit der Gleichung, die Unabhängigkeit beschreibt, nicht möglich ist. Wir benötigen also eine quantitative Version der Unabhängigkeit. Diese liefert der Korrelationskoeffizient.

DEFINITION 4.7. *Seien X und Y Zufallsgrößen auf dem endlichen Wahrscheinlichkeitsraum Ω. Dann heißt*

$$\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Kovarianz von X und Y. Es ist $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$. Um den Wert dieser Größe beurteilen zu können, muss man ihn ins Verhältnis zu den Standardabweichungen von X und Y setzen:

$$\rho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}$$

wird Korrelationskoeffizient von X und Y genannt.

Es ist stets $|\rho(X, Y)| \leq 1$. Ein positiver Korrelationskoeffizient bedeutet, dass X tendenziell größer wird, wenn Y anwächst. Entsprechend bedeutet ein negativer Korrelationskoeffizient, dass der Wert von X tendenziell schrumpft, wenn der von Y wächst. Es gilt nun in der Tat

SATZ 4.8. *Sind X und Y unabhängig, so ist $\rho(X, Y) = 0$.*

Die Umkehrung ist aber im Allgemeinen falsch! Der Korrelationskoeffizient erfasst nicht die ganze Komplexität des Unabhängigkeitskonzepts.

BEISPIEL 4.9. Wir betrachten wie in Beispiel 4.6 das zweifache Würfeln und X, Y aus Beispiel 4.6 zusammen mit der neuen Zufallsvariable $Z :=$ Summe der beiden Augenzahlen.

(1) Da wir aus dem Wert von Z sofort den von X berechnen können, erwarten wir hier eine Abhängigkeit. In der Tat ist zum Beispiel

$$P(X = 0, Z = 3) = 0 \neq \frac{1}{2} \cdot \frac{1}{18} = P(X = 0) \cdot P(Z = 3),$$

also sind X und Z nicht unabhängig. Drückt sich das auch in Korrelation aus? Man kann nicht sagen, dass X mit wachsendem Z steigen oder fallen würde. Und genau das wird der Grund für fehlende Korrelation sein. Zunächst berechnen wir $E(Z) = 2\mathrm{Var}(Y) = 7$ und $\mathrm{Var}(Z) = 2\mathrm{Var}Y = \frac{35}{6}$. Die Zufallsvariable XZ wird beschrieben durch

$$XZ(a_1, a_2) = \begin{cases} a_1 + a_2 & \text{, falls } a_1 + a_2 \text{ ungerade ist} \\ 0 & \text{, sonst} \end{cases}.$$

Damit ist

$$E(XZ) = \frac{1}{36} \sum_{a_1, a_2 \in \{1, \ldots, 6\} |\ a_1 + a_2 \text{ ungerade}} a_1 + a_2 = \frac{1}{36} (2 \cdot 3 + 4 \cdot 5 + 6 \cdot 7 + 4 \cdot 9 + 2 \cdot 11).$$

Hierbei stellt der zweite Faktor jedes Summanden die möglichen Werte von $a_1 + a_2$ dar und der erste Faktor, durch wieviele Würfe er realisierbar ist. Als Ergebnis erhält man schließlich

$$E(XZ) = 3{,}5 = \frac{1}{2} \cdot 7 = E(X) \cdot E(Z),$$

also in der Tat $\mathrm{Cov}(X, Z) = E(XZ) - E(X)E(Z) = 0$ und damit $\rho(X, Z) = 0$.

(2) Ist hingegen der zweite Wurf recht groß, so wird auch die Summe der beiden Würfe mit höherer Wahrscheinlichkeit einen großen Wert haben. Wir erwarten also eine positive Korrelation zwischen Y und Z. Dazu berechnen wir wieder

$$YZ(a_1, a_2) = a_2(a_1 + a_2)$$

und damit

$$\begin{aligned} E(YZ) &= \frac{1}{36} \sum_{a_1=1}^{6} \sum_{a_2=1}^{6} (a_2 a_1 + a_2^2) = \frac{1}{36} \left((\sum_{a=1}^{6} a)^2 + 6 \sum_{a=1}^{6} a^2 \right) \\ &= \frac{1}{36}(441 + 546) = \frac{329}{12} = 27{,}417. \end{aligned}$$

Also haben wir $\mathrm{Cov}(Y, Z) = \frac{329}{12} - 3{,}5 \cdot 7 = \frac{35}{12}$ und daher

$$\rho(Y, Z) = \frac{\mathrm{Cov}(Y, Z)}{\sigma(Y)\sigma(Z)} = \frac{\frac{35}{12}}{\sqrt{\frac{35}{12} \cdot \frac{35}{6}}} = \frac{1}{2}\sqrt{2} = 0{,}7071.$$

△

3. Die Tschebyscheff-Ungleichung

Erwartungswert und Varianz bilden nur zwei von unendlich vielen Informationen, die in einer Wahrscheinlichkeitsverteilung stecken. Sie sind aber immerhin so wichtig, dass man ohne darüber hinausgehendes Wissen über die Wahrscheinlichkeitsverteilung die Wahrscheinlichkeiten für große Abweichungen vom Erwartungswert nach oben abschätzen kann.

SATZ 4.10 (Tschebyscheff-Ungleichung).

$$P(|X - E(X)| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}$$

oder äquivalent

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

BEWEIS.

$$\begin{aligned} P(|X - E(X)| \geq k\sigma(X)) &= \sum_{a \in X(\Omega), |a - E(X)| \geq k\sigma(X)} P(X = a) \\ &\leq \sum_{a \in X(\Omega), |a - E(X)| \geq k\sigma(X)} \frac{(a - E(X))^2}{k^2 \mathrm{Var}(X)} P(X = a) \\ &\leq \frac{1}{k^2 \mathrm{Var}(X)} \sum_{a \in X(\Omega)} (a - E(X))^2 P(X = a) = \frac{1}{k^2}. \end{aligned}$$

Die erste Form der Tschebyscheff-Ungleichung erhält man, indem man $k := \frac{c}{\sigma(X)}$ setzt. □

Es ist zu bemerken, dass Messgrößen normalerweise eine physikalische Einheit haben wie $m, s, kg \ldots$. In der Tat haben $X, E(X), \sigma(X)$ dieselbe Einheit. Die zweite Version der Tschebyscheff-Ungleichung ist also skaleninvariant (d.h. wenn man z.B. Fuß statt Meter nimmt, bleibt das k in der Tschebyscheff-Ungleichung gleich).

Die Tschebyscheff-Ungleichung ist für alle Wahrscheinlichkeitsverteilungen richtig und liefert daher eine in der Praxis sehr vorsichtige Abschätzung für die Wahrscheinlichkeiten extremer Ereignisse. Folgende Tabelle vergleicht die Abweichungswahrscheinlichkeit wie von der Tschebyscheff-Ungleichung geschätzt mit der, wenn bekannt ist, dass X normalverteilt ist.

$$P(|X - E(X)| \geq k\sigma(X))$$

k	Tschebyscheff	X normalverteilt
1	≤ 1 (keine Information)	$\approx 32\%$
2	$\leq 25\%$	$\approx 5\%$
3	$\leq 11\%$	$\approx 0,25\%$

Die puren Wahrscheinlichkeiten für extreme Ereignisse wie in der Tschebyscheff-Ungleichung beschrieben, sagen aber noch nichts über die langfristigen Auswirkungen solcher Ereignisse aus. Dies liegt daran, dass die tatsächliche Abweichung vom Erwartungswert solcher Ereignisse in der Ungleichung gar nicht vorkommt. Folgendes Beispiel, das in der Finanzdienstleistungsbranche gängig ist, soll ein Licht darauf werfen.

BEISPIEL 4.11. **(Value at Risk/ Expected Shortfall)** Den momentanen Wert einer Geldanlage nennen wir $100\% = 1$. Uns interessiert $X :=$ Wert der Geldanlage nach einem Jahr. Um verschiedene Anlageoptionen vergleichen zu können, gibt der Finanzdienstleister den sogenannten Value at Risk (VaR) der Anlage an: Zu einem Risikoniveau α (meistens ist $\alpha = 5\%$) wird der minimale Prozentsatz r bestimmt so, dass ein Verlust von mehr als r des aktuellen Wertes höchstens mit Wahrscheinlichkeit α eintritt, also

$$\mathrm{VaR}_\alpha(X) := \inf\{r \in [0,1] |\ P(X \leq (1-r)) \leq \alpha\}.$$

Die Zufallsvariable kann aus vergangenen Daten oder aus theoretischen Annahmen heraus modelliert werden. In der Praxis tritt meist der letzte Fall ein, da die meisten Anlagefonds noch nicht hinreichend lange existieren, um aus den Daten ein realistisches Modell zu basteln. Das Hauptproblem des VaR besteht aber darin, dass er nichts über die Höhe des tatsächlichen Verlustes in dem α-Anteil der Fälle hohen Verlustes aussagt.

Es könnte zum Beispiel folgende Situation vorliegen. Die Anlage geschieht in Aktien einer wachsenden, aber stark von der Weltkonjunktur abhängigen Branche. Die mittlere Rendite liegt bei 5.8%, der $\mathrm{VaR}_{5\%}$ ist 0.0%. Ist die Anlage zum Sparen auf eine Rente die richtige? Es mag auf den ersten Blick so scheinen, in Wirklichkeit sind die Daten aber recht aussagelos. Es könnte beispielsweise sein, dass hinter diesen Daten folgende Verteilung steht

$$P(X = 1.1) = 0.96,\ P(X = 0.1) = 0.04.$$

Meistens ist also eine 10%-ige Steigerung zu erwarten, etwa alle 25 Jahre aber ein totaler Crash. Unter der Annahme, dass die Entwicklung in aufeinander folgenden Jahren unabhängig ist, würde dies zur angegebenen mittleren Rendite von $5.8\% = E(X) - 1$ führen. Sei Y_n der Wert der Anlage nach n Jahren. Nach n Jahren mit k Gewinnjahren tritt genau dann ein Gesamtverlust ein, wenn

$$k \leq k_0 := \max\{m \in \{0, \ldots, n\} |\ 1.1^m \cdot 0.1^{n-m} < 1\}.$$

Die Wahrscheinlichkeit für einen Verlust nach n Jahren beträgt also

$$P(Y_n < 1) = \sum_{k=0}^{k_0} \binom{n}{k} (0.96)^k \cdot (0.04)^{n-k}.$$

Damit errechnet man etwa

$$P(Y_{25} < 1) \approx 64\%, \; P(Y_{50} < 1) \approx 60\%.$$

Für große n übersetzt sich die Bedingung $k \le k_0$ in

$$k \le n \frac{\ln 10}{\ln 11}.$$

Die Binomialverteilung zu den Parametern n und $q = 0.96$ hat Erwartungswert $0.96n$ und Varianz $0.96 \cdot 0.04n$. Im Limes gilt daher nach dem zentralen Grenzwertsatz (siehe nächsten Abschnitt) sogar

$$\lim_{n \longrightarrow \infty} P(Y_n < 1) = \lim_{n \longrightarrow \infty} \Phi\left(\frac{n(\frac{\ln 10}{\ln 11} - 0.96)}{\sqrt{0.96 \cdot 0.04n}}\right) = \lim_{n \longrightarrow \infty} \Phi(0.0013\sqrt{n}) = 1,$$

allerdings ist die Konvergenz sehr langsam. Für ein Rentensparen ist eine Anlage mit solch hoher langfristiger Verlustwahrscheinlichkeit sicherlich nicht empfehlenswert.

Eine weniger vergessliche und daher aussagekräftigere Kenngröße einer Geldanlage ist der Expected Shortfall (ES)

$$\mathrm{ES}_\alpha(X) := \frac{1}{\alpha} \int_0^\alpha \mathrm{VaR}_\gamma(X) d\gamma.$$

Er misst so etwas wie den zu erwartenden Verlust im schlimmsten α-Anteil aller Fälle. In der hier vorliegenden Situation ist

$$\mathrm{VaR}_\gamma(X) = \begin{cases} 0.9 & \text{, falls } \gamma \le 0.04 \\ 0.0 & \text{, falls } 0.04 < \gamma \end{cases}$$

Damit ist

$$\mathrm{ES}_{5\%}(X) = 20 \cdot (0.04 \cdot 0.9) = 0.72 = 72\%,$$

ein Wert, der schon eher einer adäquaten Warnung vor diesem Finanzprodukt (zumindest zum Zweck des Rentensparens) gleicht. $\triangle$

Wenden wir zum Schluss dieses Abschnitts noch die Tschebyscheff-Ungleichung auf die Mittelung von n unabhängigen Kopien von X an. Seien $X_1, \ldots, X_n$ diese Kopien und

$$Y_n := \frac{1}{n} \sum_{i=1}^{n} X_i : \Omega^n \longrightarrow \mathbb{R}, (a_1, \ldots, a_n) \mapsto \frac{1}{n}(X(a_1) + \cdots + X(a_n)).$$

Wir haben in Satz 4.5 gelernt, dass dann $E(Y_n) = E(X)$ und $\mathrm{Var}(Y_n) = \frac{1}{n}\mathrm{Var}(X)$ gilt. Setzen wir das in die Tschebyscheff-Ungleichung ein, so folgt

$$P(|Y_n - E(X)| \ge k\sigma(X)) \le \frac{1}{nk^2},$$

insbesondere folgt das Gesetz der großen Zahlen

SATZ 4.12 (Gesetz der großen Zahlen). *Für alle $\varepsilon > 0$ ist*

$$\lim_{n \longrightarrow \infty} P(|Y_n - E(X)| \ge \varepsilon) = 0.$$

Es besagt, dass wir den Erwartungswert einer Zufallsvariablen durch Mitteln häufiger Messungen approximieren können. Das ist aber noch eine recht schwache Aussage über die Verteilung nach vielen Wiederholungen. Eine stärkere Aussage liefert der zentrale Grenzwertsatz.

4. Zentraler Grenzwertsatz

Wir haben bereits beobachtet, dass die Binomialverteilung für großes n der Normalverteilung sehr ähnelt. Dies gilt nicht nur für die Binomialverteilung, sondern für alle Verteilungen, die durch unabhängiges Wiederholen eines Experimentes und Aufsummieren der Werte der Zufallsvariablen entstehen.

Sei also ein durch die Zufallsvariable X beschriebenes Zufallsexperiment gegeben, z.B. das Würfeln oder das Werfen einer Münze. Sei $\mu = E(X)$ und $\sigma^2 = \mathrm{Var}(X)$.

Das Experiment werde nun r-mal wiederholt. Die Annahme ist hier natürlich, dass die verschiedenen Wiederholungen voneinander unabhängig sind. Wir ordnen der i-ten Wiederholung die Zufallsvariable X_i zu. X_i ist also eine Kopie von X mit der gleichen Verteilung. Man bilde nun die neue Zufallsvariable

$$\bar{Y}_r := X_1 + \ldots + X_r.$$

Es ist leicht einzusehen, dass der Erwartungswert von $\bar{Y}_r$ genau $r \cdot \mu$ ist. Man kann ebenso berechnen, dass die Varianz genau $r \cdot \sigma^2$ ist, also die Standardabweichung $\sqrt{r} \cdot \sigma$ ist. Wir betrachten die neue Zufallsvariable

$$Y_r := \frac{\bar{Y}_r - r\mu}{\sigma\sqrt{r}} = \frac{X_1 + \ldots + X_r - r\mu}{\sigma\sqrt{r}}.$$

Man rechnet nach, dass $E(Y_r) = 0$ und $\mathrm{Var}(X) = 1$. Die Kenngrößen Erwartungswert und Varianz von Y_r sind also die gleichen, wie die einer $(0,1)$-normalverteilten Zufallsvariablen.

Eine viel stärkere Aussage ist jedoch richtig: Die Verteilung von Y_r nähert sich für größer werdendes r immer weiter an die normierten Normalverteilung ($\mu = 0$ und $\sigma = 1$) an:

SATZ 4.13 (Zentraler Grenzwertsatz). *Für alle $a, b \in \mathbb{R}$ und $r \in \mathbb{N}$ hinreichend groß ist die Wahrscheinlichkeit, dass Y_r einen Wert aus $[a,b]$ annimmt, annähernd gleich der Wahrscheinlichkeit, dass eine $(0,1)$-normalverteilte Zufallsvariable einen Wert in $[a,b]$ annimmt. Mathematischer ausgedrückt:*

$$\lim_{r\to\infty} P(Y_r \in [a,b]) = \phi(b) - \phi(a).$$

BEISPIEL: Eine Münze wird 500 mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass mindestens 240 und höchstens 270 mal Zahl geworfen wird?

Antwort: Zunächst können wir dies mit Binomialverteilung berechnen: Wir wissen, dass die Wahrscheinlichkeit, dass k mal Kopf kommt, genau

$$\binom{500}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{500-k} = \binom{500}{k}\left(\frac{1}{2}\right)^{500}$$

ist, die gesuchte Wahrscheinlichkeit ist also

$$\sum_{k=240}^{270}\binom{500}{k}\left(\frac{1}{2}\right)^{500} \approx 0,793.$$

Alternativ können wir dies berechnen, indem wir die Binomialverteilung durch die Normalverteilung approximieren. Sei also X die Zufallsvariable mit Wert 0 für Kopf und Wert 1 für Zahl und $P(X=0) = P(X=1) = \frac{1}{2}$. Der Erwartungswert ist also $\mu = \frac{1}{2}$, die Varianz $\sigma^2 = \frac{1}{2}^2 = \frac{1}{4}$ und die Standardabweichung $\sigma = \frac{1}{2}$.

Für $i = 1, \ldots, 500 = r$ sei X_i ein Kopie von X und $\bar{Y}_{500} = X_1 + \ldots + X_{500}$. Uns interessiert $P(\bar{Y}_{500} \in [240, 270])$. Es gilt $E(\bar{Y}_{500}) = 500\mu = 500 \cdot \frac{1}{2} = 250$ und $\mathrm{Var}(\bar{Y}_{500}) = 500\sigma^2 = 500 \cdot \frac{1}{4} = 125$. Es gilt ferner, dass $\sigma\sqrt{r} = \frac{1}{2}\sqrt{500} = \sqrt{125} = 11,18$.

Sei nun $Y_{500} = \frac{\bar{Y}_{500}-250}{\sqrt{125}}$ wie oben. Es gilt

$$\begin{aligned} P(\bar{Y}_{500} \in [240,270]) &= P\left(Y_{500} \in \left[\frac{240-250}{\sqrt{125}}, \frac{270-250}{\sqrt{125}}\right]\right) = P(Y_{500} \in [-0,894428, 1,788854]) \\ &\approx \phi(1,788854) - \phi(-0,894428) = 0,963 - 0,186 = 0,777. \end{aligned}$$

Bei kleinem r bzw. einem kleinen Intervall für $\bar{Y}_r$ und ganzzahligen Werten von $\bar{Y}_r$ ist es ratsam, das ursprüngliche ganzzahlige Intervall (hier $[240, 270]$) an beiden Grenzen um $0,5$ zu vergrößern, da dies eine genauere Annäherung an die tatsächliche Wahrscheinlichkeit gibt. Hier sollten wir daher Werte von $\bar{Y}_{500}$ im Intervall $[239,5\,,\, 270,5]$ betrachten. Wir rechnen also in obigem Beispiel besser

$$\begin{aligned} P(\bar{Y}_{500} \in [240,270]) &= \left(P(Y_{500} \in \left[\frac{240-250}{\sqrt{125}}, \frac{270-250}{\sqrt{125}}\right]\right) \approx \phi\left(\frac{270,5-250}{\sqrt{125}}\right) - \phi\left(\frac{239,5-250}{\sqrt{125}}\right) \\ &= \phi(1,8336) - \phi(0,93915) = 0,967 - 0,176 = 0,791. \end{aligned}$$

Bei sehr großem r wird dieser Fehler vernachlässigbar sein.

5. Aufgaben

Aufgabe 4-1 Ein (Laplace)-Würfel werde dreimal (unabhängig voneinander) nacheinander geworfen. Es bezeichne X_i das Ergebnis des i-ten Wurfes und X die Summe aus dem Produkt aus den ersten beiden Würfen mit dem dritten, also $X := X_1 \cdot X_2 + X_3$. Weiter sei Y die Differenz von erstem und zweitem Wurf, also $Y := X_1 - X_2$.

(1) Berechnen Sie den Erwartungswert von X.
(2) Berechnen Sie die Varianz von Y.

Aufgabe 4-2 Es wird mit 2 Würfeln gewürfelt. Die Zufallsvariable X misst das Produkt der Augenzahlen. Berechnen Sie $E(X)$ und $\mathrm{Var}(X)$.

Aufgabe 4-3 Bestimmen Sie die Wahrscheinlichkeit, dass beim 1000-maligen Würfeln mindestens 160 mal eine 6 gewürfelt wird. Benutzen Sie hierbei die Normalverteilung.

Aufgabe 4-4 (1+1+1 Punkte) Eine Münze wird 300 mal geworfen. Berechnen Sie die Wahrscheinlichkeit, dass mindestens 149 und höchstens 151 mal Kopf geworfen wird. Berechnen Sie die Wahrscheinlichkeit auf die folgenden drei Arten:

(1) Mit Hilfe der Binomialverteilung (ohne Normalverteilung).
(2) Mit Hilfe der Normalverteilung ohne Verbesserung der Genauigkeit (d.h. ohne $\pm 0,5$).
(3) Mit Hilfe der Normalverteilung mit Verbesserung der Genauigkeit (d.h. mit $\pm 0,5$).

Aufgabe 4-5 (1+1+1 Punkte) Eine Münze wird 200 mal geworfen. Berechnen Sie die Wahrscheinlichkeit, dass mindestens 149 und höchstens 151 mal Kopf geworfen wird. Berechnen Sie die Wahrscheinlichkeit auf die folgenden drei Arten:

(1) Mit Hilfe der Binomialverteilung (ohne Normalverteilung).
(2) Mit Hilfe der Normalverteilung ohne "Verbesserung der Genauigkeit" (d.h. ohne $\pm 0,5$).
(3) Mit Hilfe der Normalverteilung mit "Verbesserung der Genauigkeit" (d.h. mit $\pm 0,5$).

Die "Verbesserung der Genauigkeit" hat hier zu einer wesentlichen Verschlechterung der Genauigkeit geführt: Man schätzt die Wahrscheinlichkeit fast doppelt so gross wie sie wirklich ist. Das liegt daran, dass die Normalverteilung die Binomialverteilung vor allem um den Erwartungswert herum schnell annähert, in den Randbereichen dauert das wesentlich länger. Daher benutzt man für seltene Ereignisse eine andere Verteilungsnäherung: Die sogenannte Poissonverteilung wird noch Gegenstand der Vorlesung werden.

Aufgabe 4-6 Beim Lotto 6 aus 49 werden 50% der Einsätze an den Staat abgeführt, die andere Hälfte wird nach einem bestimmten Schlüssel auf die Gewinnklassen und den Jackpot verteilt. Wir betrachten einen einzelnen Tipp und die Zufallsvariable X, die den Nettogewinn pro Einsatz angibt, also $\frac{\text{Gewinn}-\text{Einsatz}}{\text{Einsatz}}$. Geben Sie $E(X)$ an und begründen Sie ihre Antwort kurz.

Aufgabe 4-7 Beim Roulette wird eine Zahl zwischen 0 und 36 mittels einer Kugel bestimmt, 18 davon sind rot, 18 sind schwarz und die 0 ist grün. Setzt man auf eine Zahl und gewinnt, so bekommt man den 36fachen Einsatz zurück. Setzt man auf Farbe (rot oder schwarz) und gewinnt, so erhält man den doppelten Einsatz zurück. Zu jedem der beiden Spiele betrachten wir die Zufallsvariable Nettogewinn pro Einsatz wie in der Aufgabe zuvor. Berechnen Sie Erwartungswert und Varianz der beiden Zufallsvariablen.

Aufgabe 4-8 In der 20-Uhr-Tagesschau vom 4. März 2012 äußerte sich die Korrespondentin aus Moskau dahingehend, dass auf Putin momentan ein Stimmenanteil von etwa 64% entfalle, es seien aber erst 30% der Wahlbezirke ausgezählt.

Sei also $\Omega := \{\text{Wahlbezirke}\}$ ein Laplace-Wahrscheinlichkeitsraum, d.h. $P(a) = \frac{1}{|\Omega|}$ für alle Wahlbezirke a, und $X(a) =$ Stimmenanteil für Putin im Bezirk a. Die Zufallsvariable X entspricht also dem Experiment der zufälligen Auswahl eines Wahlbezirks und ordnet als Wert das dortige Wahlergebnis zu. Anders interpretiert: ein Wahlergebnis wird gemeldet. Wir nehmen vereinfachend an, dass die Wahlbezirke alle gleich viele Einwohner enthalten, das Gesamtergebnis der Wahl also mit $E(X)$ übereinstimmt. Nach n Meldungen von Wahlergebnissen wollen wir eine Hochrechnung machen. Wir nehmen an, dass die Meldungen voneinander unabhängig geschehen (was nicht ganz realistisch ist), wir also n unabhängige Kopien $X_1, \ldots, X_n$ des Experiments X haben. Unsere Hochrechnung soll

$$Y_n(a_1, \ldots, a_n) := \frac{1}{n} \sum_{i=1}^{n} X_i(a_i)$$

sein auf dem Wahrscheinlichkeitsraum Ω^n mit $P(a_1, \ldots, a_n) = \frac{1}{|\Omega|^n}$.

(1) Zeigen Sie $0 \le E(X) \le 1$ und $Var(X) < 1$. Geben Sie $E(Y_n)$ und $Var(Y_n)$ an.
(2) Geben Sie mit Hilfe der Tschebyscheff-Ungleichung eine obere Schranke für $P(|Y_n - E(X)| \ge 0.02)$ an. Diese soll nur noch von n abhängen.
(3) Welche wichtige Information fehlte also in der Tagesschau? Welche Information war unnötig und irreführend?
(4) Es gab 95000 Wahlbezirke[1]. Mit welcher Mindestwahrscheinlichkeit differierte zum Zeitpunkt der Tagesschau das Wahlergebnis um höchstens 2% bzw. 5%? Mit welcher Mindestwahrscheinlichkeit wurde der Wahlsieg Putins richtig prognostiziert?
(5) Nehmen Sie nun an, dass X normalverteilt ist (mit unbekanntem $\mu \in [0,1], \sigma \in [0,1]$) und schätzen Sie mit dieser Zusatzinformation die Wahrscheinlichkeit, dass die Hochrechnung maximal 1.5% vom Endergebnis abweicht.
(6) Halten Sie die Annahme einer Normalverteilung für gerechtfertigt?

Aufgabe 4-9 Der Verlauf des Erzeugerpreisindex landwirtschaftlicher Produkte zwischen Januar 2011 und März 2012 war der folgende[2]:

[1] Quelle: Wikipedia
[2] Quelle: Destatis

Monat	pflanzlich	tierisch
2012 Mär	140,8	125,0
Feb	138,6	124,0
Jan	135,6	121,9
2011 Dez	132,8	124,4
Nov	133,6	126,6
Okt	133,3	125,0
Sep	138,2	123,7
Aug	139,1	123,1
Jul	142,9	122,7
Jun	150,4	122,7
Mai	153,7	122,5
Apr	153,5	121,5
Mär	151,8	118,7
Feb	158,5	114,7
Jan	155,6	110,8

(Der Index ist so gewählt, dass die Preise im jeweiligen Monat des Jahres 2005 stets 100 sind. Damit sind jahreszeitliche Schwankungen bereits heraus gerechnet.)

Sei $\Omega := \{\text{Monate von Jan 2011 bis Mär 2012}\}$ gegeben als Laplace-Wahrscheinlichkeitsraum sowie

$$X(a) := \text{Index pflanzlicher Produkte im Monat } a$$

und

$$Y(a) := \text{Index tierischer Produkte im Monat } a.$$

Berechnen Sie den Korrelationskoeffizienten $\rho(X, Y)$.

Aufgabe 4-10 Diskutieren Sie für folgende Spiele, ob Sie bei großem N den zentralen Grenzwertsatz anwenden können:

(1) Sie würfeln und interessieren sich für die Augensumme aller N Würfe, die Sie durchgeführt haben.
(2) Sie werfen eine Münze und wetten bei Wurf Nummer n einen Einsatz von n Euro auf "Kopf". Sie interessieren sich für den Gesamtgewinn nach N Durchgängen.
(3) Ihre Geldanlage wird jährlich mit einem variablen Zinssatz X verzinst, den wir mit einer Zufallsvariablen modellieren. Sie interessieren sich für den Wert ihrer Anlage nach N Jahren.

Aufgabe 4-11 Es wird 1000 mal gewürfelt. Die Zufallsvariable X sei die Augensumme der 1000 Würfe. Bestimmen sie den Erwartungswert $E(X)$ von X. Bestimmen Sie ferner s näherungsweise so, dass $P(X \in [E(X) - s, E(X) + s]) = 0,99$.

Aufgabe 4-12 Die durchschnittliche Windgeschwindigkeit im Mai am Kieler Leuchtturm beträgt 13 Knoten. Die Wahrscheinlichkeit, dass der Wind zu einem zufälligen Zeitpunkt im Mai eines Jahres stärker als mit 10 Knoten weht, beträgt 58%.[3] Es ist experimentell bekannt, dass Windgeschwindigkeiten log-normalverteilt sind, d.h. für

$$X(t) := \text{Windgeschwindigkeit in Knoten zum Zeitpunkt } t \text{ im Mai}$$

ist $\ln X$ normalverteilt. Man kann aus obigen Daten errechnen, dass $\ln X$ $(2.41, 0.55)$-normalverteilt ist. Berechnen Sie die Wahrscheinlichkeit, dass an einem zufällig gewählten Zeitpunkt im Mai 2012 mindestens Beaufort 6 (=20 Knoten) erreicht wird.

[3] Quelle: windfinder.com

KAPITEL 5

Schätzen aus Stichproben

1. Schätzung von Erwartungswert und Kovarianz

Auch wenn man oft davon ausgehen kann, dass in der Natur gemessene Größen näherungsweise normalverteilt sind, sind Erwartungswert und Varianz in der Regel nicht bekannt. Diese werden dann experimentell durch eine (große) Stichprobe ermittelt. In diesem Abschnitt beschäftigen wir uns damit, wie man sie schätzt. Später werden wir bestimmen, wie zuverlässig diese Schätzungen sind.

Bei der Wahl der Stichprobe ist darauf zu achten, dass sie möglichst zufällig ist, d.h. der Auswahlprozess nicht bestimmte Messungen bevorzugt. Will ich beispielsweise den Sauerstoffgehalt in 100m Meerestiefe an einem bestimmten Meerespunkt regelmässig messen, so ist durch den Umstand, dass ich bei Sturm die Messung nicht durchführen kann, bereits mit einer strukturellen Abweichung der gemessenen Mittelwerte und Varianzen von den tatsächlichen zu rechnen. Rufe ich zur freiwilligen Teilnahme an einer psychologischen Studie auf, so muss man sicherstellen, dass der Umstand der Freiwilligkeit, also der Offenheit für solche Studien, nicht das Messergebnis der Studie beeinflusst, damit das Ergebnis als repräsentativ gedeutet werden kann.

Sind solche Einflüsse durch den Auswahlprozess ausgeschlossen, spricht man auch von einer Zufallsstichprobe. Im Folgenden sollen alle unsere Stichproben Zufallsstichproben sein.

Oft ist für eine Meßgröße die Verteilung nicht bekannt, sondern man hat lediglich eine Stichprobe. Man kann in diesen Fällen dann den Erwartungswert und die Varianz schätzen. Eine Stichprobe ist mathematisch nichts anderes als ein Tupel $(x_1, \ldots, x_n)$ reeller Zahlen, in der Regel eine Meßreihe. Wir können dann definieren:

DEFINITION 5.1. *Sei eine Stichprobe $(x_1, \ldots, x_n)$ gegeben.*

(1) *Es heißt $\bar{x} := \frac{1}{n}(x_1 + \ldots + x_n)$ der (empirische) Mittelwert der Stichprobe.*
(2) *Es heißt $s^2 := \frac{1}{n-1}\left((x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2\right)$ die (empirische) Varianz der Stichprobe.*

Der Faktor $\frac{1}{n-1}$ bei der Schätzung der Varianz kommt daher, dass unsere n Messungen nicht den ganzen Ereignisraum bilden, sondern nur einen kleinen Ausschnitt. Daher ist auch $\bar{x}$ nicht der exakte Erwartungswert, sondern fehlerbehaftet, d.h. er variiert mit wechselnder Stichprobe. Nutzen wir also $\bar{x}$ anstelle des unbekannten Erwartungswertes, um die Varianz zu schätzen, so muss die Variabilität von $\bar{x}$ selbst in die Rechnung eingehen. Dies führt zu dem Vorfaktor $\frac{1}{n-1}$.

Das ist auch recht intuitiv: Aus den Daten $\bar{x}, x_2, \ldots, x_n$ können wir x_1 errechnen, also enthalten diese n Daten die gleiche Information wie unsere ursprüngliche Stichprobe. Da der Erwartungswert uns nichts über die Varianz mitteilen kann, stehen zu deren Schätzung also in Wirklichkeit nur noch $n-1$ Daten, $x_2, \ldots, x_n$ zur Verfügung.

Mathematisch genauer ist der Grund: Sind $X_1, \ldots, X_n$ unabhängige Kopien von X und

$$\bar{X} := \frac{1}{n}(X_1 + \ldots + X_n), \; S^2 := \frac{1}{n-1}\left((X_1 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2\right),$$

so ist $E(\bar{X}) = E(X)$ und $E(S^2) = \mathrm{Var}(X)$. Mit der Tschebyscheff-Ungleichung folgt also, dass die zu erwartende Abweichung unserer Schätzung von den zu schätzenden Werten mit wachsendem n gegen 0 geht.

Hätten wir naiv $\frac{1}{n}$ anstatt dem Faktor $\frac{1}{n-1}$ gewählt, so würde unsere geschätzte Varianz zu klein ausfallen. Ähnliches gilt für die Schätzung von Kovarianzen:

DEFINITION 5.2. *Sind Stichproben $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$ zweier Messgrößen gegeben, so heißt $c := \frac{1}{n-1}\left((x_1 - \bar{x})(y_1 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})\right)$ die (empirische) Kovarianz der beiden Stichproben.*

Das Entnehmen einer Stichprobe $(x_1, \ldots, x_n)$ einer durch die Zufallsvariable X beschriebenen Größe und das Bilden des Mittelwertes, kann, wie bereits erwähnt, selbst als Zufallsvariable aufgefasst werden, nämlich als die Zufallsvariable $X_M^n = \frac{1}{n}(X_1 + \ldots + X_n)$, wobei X_i unabhängige Kopien von X sind für $1 \leq i \leq n$.

Zumindest der erste Teil des folgenden Satzes folgt bereits aus unserer vorherigen Diskussion.

SATZ 5.3. *Hat X den Erwartungswert μ und die Varianz σ^2, so hat X_M^n den Erwartungswert μ und die Varianz $\frac{\sigma^2}{n}$.*

Ist X eine (μ, σ)-normalverteilte Zufallsvariable, so ist X_M^n eine $\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$-normalverteilte Zufallsvariable. Insbesondere ist $\frac{(X_M^n - \mu)\sqrt{n}}{\sigma}$ eine $(0,1)$-normalverteilte Zufallsvariable.

BEISPIEL 5.4. Das 10-malige Wiederholen eines Experimentes ergibt die Stichprobe

$$x = (7, 5, 6, 11, 6, 7, 8, 7, 10, 9).$$

Wir bestimmen den empirischen Mittelwert und die empirische Varianz der Stichprobe. Für $\bar{x}$ rechnen wir

$$\bar{x} = \frac{1}{10}\left(7 + 5 + 6 + 11 + 6 + 7 + 8 + 7 + 10 + 9\right) = 7{,}6.$$

Die empirische Varianz ist damit

$$\begin{aligned} s^2 &= \frac{1}{9}\left((7-7{,}6)^2 + (5-7{,}6)^2 + (6-7{,}6)^2 + (11-7{,}6)^2 + (6-7{,}6)^2 + (7-7{,}6)^2 + (8-7{,}6)^2 \right. \\ &\quad \left. + (7-7{,}6)^2 + (10-7{,}6)^2 + (9-7{,}6)^2\right) = 3{,}6. \end{aligned}$$

△

2. Aufgaben

Aufgabe 5-1 Sei $\Omega = \{$volle Minuten zwischen 16:09 und 18:09 am 23. Mai 2012$\}$ als Laplace-Wahrscheinlichkeitsraum gegeben. Wir bezeichnen mit $X(t)$ die Windrichtung am Kieler Leuchtturm für $t \in \Omega$. Folgende Messungen liegen uns vor[1]:

[1]Quelle: GEOMAR

Uhrzeit	Windrichtung
18:05	78 Grad
17:57	77 Grad
17:49	76 Grad
17:41	74 Grad
17:33	76 Grad
17:25	75 Grad
17:17	74 Grad
17:09	72 Grad
17:01	70 Grad
16:53	70 Grad
16:45	73 Grad
16:37	73 Grad
16:29	71 Grad
16:21	70 Grad
16:13	68 Grad

(1) Schätzen Sie den Erwartungswert und die Standardabweichung von X.

(2) Wir nehmen nun an, dass die Standardabweichung von X mit der geschätzten übereinstimmt. Geben Sie mit Hilfe der Tschebyscheff-Ungleichung eine Mindestwahrscheinlichkeit an, mit der die Windgeschwindigkeit zu einem zufälligen Zeitpunkt $t \in \Omega$ höchstens 3,5 Grad vom Erwartungswert von X abweicht.

(3) Wie hoch ist die relative Häufigkeit der Messwerte, die höchstens 3,5 Grad vom Mittelwert der Messwerte abweichen?

KAPITEL 6

Lineare Regression

1. Berechnung der Regressionsgerade

Bisher haben wir uns auf den Standpunkt gestellt, dass hinter allen Messwerten eine zeitlich unabhängige Zufallsvariable steht. Oftmals unterliegt diese aber auch einem zeitlichen Trend. Die Punkte (x_i, t_i) der Messwerte x_i zu den Zeiten t_i gruppieren sich dann um den Graph einer mehr oder weniger klar erkennbaren Funktion. Würden wir zum Beispiel die Windgeschwindigkeit am Kieler Leuchtturm im Monatsdurchschnitt betrachten, so wäre es klar, dass wir in verschiedenen Jahren die Messwerte zum gleichen Monat vergleichen sollten, um eine zeitlich unabhängige Zufallsvariable zugrunde legen zu können. (Und selbst dann könnte man immer noch langfristige klimatische Trends ausmachen.) In monatlicher Abfolge betrachtet würden wir erwarten, dass die Messungen um eine periodische Funktion streuen mit Periodenlänge ein Jahr. Dies formalisiert man wie folgt:

Sei Ω ein Ereignisraum mit Wahrscheinlichkeitsverteilung P und $Y : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable. Jedem Ereignis $a \in \Omega$ sei ein Zeitpunkt $T(a)$ zugeordnet, d.h. wir haben eine weitere Zufallsvariable $T : \Omega \longrightarrow \mathbb{R}$. Diese parametrisiert Y, indem wir

$$\Omega_t := \{a \in \Omega |\ T(a) = t\}, P_t(A) := \frac{P(A)}{P(\Omega_t)} \text{ und } Y_t := Y|_{\Omega_t}$$

(für alle $A \subseteq \Omega_t$) betrachten.

Die Frage ist: Wie hängt Y_t von t ab, d.h. wie sind Y und T korreliert? Unsere erste Beobachtung ist, dass wir anstatt der Zeit T eine beliebige Zufallsvariable X wählen können (der Parameter t wird in x umbenannt). Die Fragestellung impliziert dann bereits, dass die Antwort eine geometrische Interpretation des Korrelationskoeffizienten sein wird.

Der Ansatz ist, dass wir uns auf eine lineare Abhängigkeit zwischen X und Y beschränken. Sicherlich können wir $a, b \in \mathbb{R}$ im Allgemeinen nicht so wählen, dass die Gleichung $Y = aX + b$ erfüllt werden kann, denn das hieße $Y_x = ax + b$, also nähme Y mit Sicherheit den Wert $ax + b$ an, wenn X den Wert x annimmt. Aber wir können a, b so wählen, dass

$$E(Y) = E(aX + b) \text{ und } \mathrm{Var}(Y - aX - b) \text{ minimal}$$

wird und damit in gewissem Sinne eine beste Approximation von Y durch eine affin-lineare Funktion von X erhalten. Bei großem Korrelationskoeffizient zwischen X und Y ist eine solche lineare Approximation sicherlich gerechtfertigt. Beachte, dass die Varianz invariant unter Addition von Konstanten ist, also in Wirklichkeit nur $\mathrm{Var}(Y - aX)$ zu minimieren ist.

Aus $E(Y) = E(aX + b)$ folgt, dass b durch $b = E(Y) - aE(X)$ bestimmt ist, wenn wir a kennen. Die zu minimierende Größe ist

$$\mathrm{Var}(Y - aX) = E(((Y - E(Y)) - a(X - E(X)))^2) = \mathrm{Var}(Y) - 2a\mathrm{Cov}(X, Y) + a^2\mathrm{Var}(X).$$

Um dies zu minimieren, leiten wir nach a ab, setzen das Ergebnis gleich 0 und formen das um zu

$$a = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} = \frac{\sigma(Y)}{\sigma(X)}\rho(X, Y).$$

Hier ist $\rho(X,Y)$ der Korrelationskoeffizient zwischen X und Y.

Die (im Sinne "kleinster Quadrate") beste affin-lineare Approximation von Y durch X ist also gegeben durch

$$Y - E(Y) \approx \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}(X - E(X))$$

oder äquivalent

$$\text{(*)} \qquad \frac{Y - E(Y)}{\sigma(Y)} \approx \rho(X,Y)\frac{X - E(X)}{\sigma(X)}.$$

DEFINITION 6.1. *Gegeben die Zufallsvariablen X und Y, heißt die Gerade*

$$\{(x,y) \in \mathbb{R}^2 \mid y - E(Y) = \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}(x - E(X))\}$$

Regressionsgerade von Y bezüglich X.

Es ist zu bemerken, dass die Regressionsgerade von X bezüglich Y eine andere als die von Y bezüglich X ist, wenn $|\rho(X,Y)| < 1$, denn letztere wird durch

$$\frac{y - E(Y)}{\sigma(Y)} = \rho(X,Y)\frac{x - E(X)}{\sigma(X)}$$

beschrieben, erstere aber durch

$$\frac{y - E(Y)}{\sigma(Y)} = \frac{1}{\rho(X,Y)}\frac{x - E(X)}{\sigma(X)}.$$

Die Regressionsgerade wird noch anschaulicher, wenn wir das Problem auf dem Niveau von Stichproben betrachten. Gegeben Zufallsstichproben $x := (x_1, \ldots, x_n)$ von X und $y := (y_1, \ldots, y_n)$ von Y, so dass die Messungen x_i und y_i in Beziehung stehen (z.B. am gleichen Ort oder zur gleichen Zeit oder unter anderen ähnlichen Umständen gewonnen), wollen wir die Regressionsgerade schätzen. Naheliegend und sinnvoll ist, einfach die uns bekannten Schätzungen $c_{x,y}$ für die Kovarianz zwischen X und Y sowie $s_x^2, \bar{x}, \bar{y}$ für die Varianz von X, den Erwartungswert von X und den von Y zu verwenden.

DEFINITION 6.2. *Die Gerade*

$$\{(u,v) \in \mathbb{R}^2 \mid v - \bar{y} = \frac{c_{x,y}}{s_x^2}(u - \bar{x})\}$$

heißt empirische Regressionsgerade der Stichprobe y bezüglich der Stichprobe x.

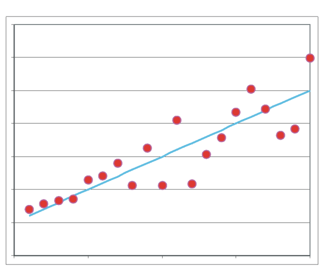


ABBILDUNG 6.1. Darstellung einer (empirischen) Regressionsgerade. Quelle: Wikibooks, Autor: Philipendula

Sie hat die Eigenschaft, dass sie die Summe der Quadrate der vertikalen Abstände zwischen (x_i, y_i) und einer festen Gerade unter allen Geraden minimiert. Die Rechnung dazu ist ganz ähnlich zu unserer obigen Argumentation. Und nun sehen wir auch anschaulich, warum die Regressionsgerade von y bezüglich x eine andere ist als die von x bezüglich y: erstere minimiert die Quadrate der vertikalen Abstände, letztere die Quadrate der horizontalen Abstände zu Geraden. Bemerke weiter, dass beide Geraden durch den Punkt $(\bar{x}, \bar{y})$ gehen. Skaliert man die Achsen in Einheiten von s_x und s_y, so sind nach dem empirischen Analogon zu Gleichung (*) die Steigungen der beiden Geraden invers zueinander, d.h. die beiden Geraden gehen dann durch Spiegelung an der Geraden mit Steigung 1 durch $(\bar{x}, \bar{y})$ auseinander hervor.

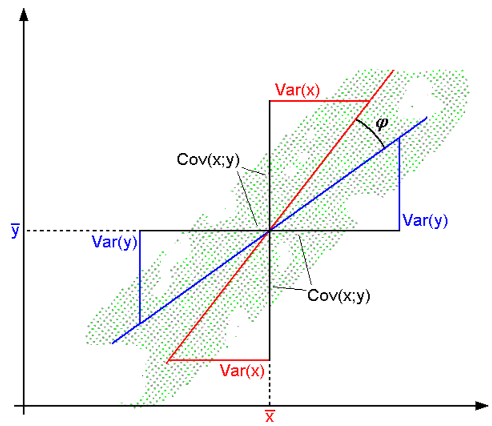


ABBILDUNG 6.2. In dieser schematischen Darstellung (basierend auf einer Graphik von Qniemic auf Wikibooks) wurde eine leicht andere, aber selbsterklärende Notation verwendet. Die rote Gerade ist die Regressionsgerade von y bezüglich x, die blaue stellt die Regressionsgerade von x bezüglich y dar.

BEISPIEL 6.3. Die Energieproduktivität bezeichnet den Quotienten aus inflationsbereinigtem Bruttoinlandsprodukt und Primärenergieverbrauch relativ zu einem gewählten Basisjahr (Wert im Basisjahr:=100). Treibhausgasemissionen in CO_2-Äquivalent bezeichnet die gesamten Emissionen von Treibhausgasen gewichtet nach ihrer Klimaaktivität relativ zu einem Basisjahr (Wert im Basisjahr:=100). Sei $X :=$ Energieproduktivität in Deutschland und $Y :=$ Treibhausgasemissionen in Deutschland. Die Daten[1] aus den Jahren 1990 bis 2010 erlauben folgende Schätzungen:

$$E(X) = 119.84, E(Y) = 86.19, \sigma(X) = 11.43, \sigma(Y) = 6.95, \mathrm{Cov}(X, Y) = -76.86.$$

Daraus errechnet sich die Schätzung für den Korrelationskoeffizienten $\rho(X, Y) = -0.97$ und die Regressionsgerade von Y bezüglich X als

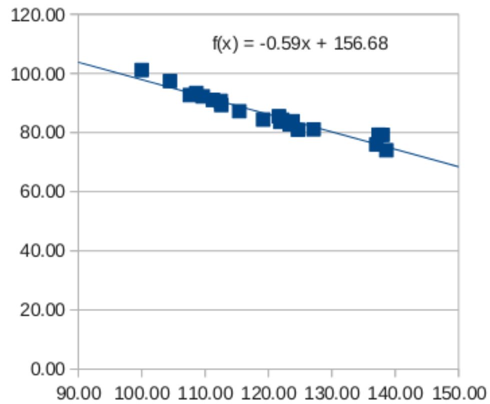
$$y = -0.59x + 156.68.$$

Die andere Regressionsgerade (von x bezüglich y) wäre umgerechnet durch

$$y = -0.63x + 161.55$$

gegeben. Das folgende Diagramm zeigt die verwendeten Datenpaare zusammen mit der Regressionsgerade von Y bezüglich X.

[1] Quelle: Destatis



Die Regressionsgerade kann so interpretiert werden, dass jeder Prozentpunkt mehr an Energieproduktivität (bzgl. des Basisjahrs) die Treibhausgasemissionen um etwa 0.6 Prozentpunkte (bzgl. des Basisjahrs) senkt, statistisch betrachtet. Es ist damit keine Aussage über einen ursächlichen Zusammenhang gemacht! Man würde hier aber sicherlich einen solchen vermuten. △

2. Aufgaben

Aufgabe 6-1 Auf Fehmarn liegen folgende jeweils im gleichen Monat gemessenen Messpaare über die monatliche Sonnenscheindauer (in h) und die durchschnittliche Windstärke (in Beaufort) vor.[2]

Sonnenschein	189.3	237.3	45.0	58.5	220.2	241.2	156.2
Windstärke	2.9	3.5	4.2	4.7	3.1	2.7	3.7

Bestimmen Sie die Regressionsgerade der Sonnenscheindauer bezüglich der Windgeschwindigkeit.

Aufgabe 6-2 Ein Laplace-Würfel werde zweimal unabhängig von einander geworfen, X_1 bezeichne die Augenzahl des ersten Wurfs, X_2 die des zweiten. Weiter sei $X_3 := X_1 + X_2$ die Summe der Augenzahlen.

(1) Berechnen Sie die Regressionsgerade von X_1 bezüglich X_2.
(2) Berechnen Sie die Regressionsgerade von X_2 bezüglich X_1.
(3) Berechnen Sie die Regressionsgerade von X_3 bezüglich X_2.
(4) Berechnen Sie die Regressionsgerade von X_2 bezüglich X_3.
(5) Erläutern Sie, warum die Ergebnisse der vorhergehenden Teilaufgaben auch ohne genauere Rechnung plausibel sind.

[2]Quelle: Deutscher Wetterdienst

KAPITEL 7

Kriging: Interpolation von Messwerten

1. Die Methode

Die von Danie Krige behandelte Fragestellung ist: Gegeben Messwerte an den Orten $x_1, \ldots, x_n$, wie bestimme ich den Wert der Messgröße an einem anderen Ort x möglichst präzise, wenn ich dort nicht messen konnte? Mit dieser Informationslage gibt es keine sinnvolle Antwort. Wird die Messgröße jedoch unter ähnlichen Umständen viele Male gemessen und liegen auch über den Ort x viele andere Messungen vor, so liefert Krige eine Antwort, die sich statistischer Methoden bedient.

Die Grundannahme ist, dass die Messungen an einem Ort y Stichproben einer dahinter stehenden Zufallsvariablen $Z(y)$ sind. Ist $\Omega \subseteq \mathbb{R}^n$ (in der Praxis: $n \in \{1,2,3\}$) das Messgebiet, so nennen wir die Gesamtheit $\{Z(y) |\ y \in \Omega\}$ ein Zufallsfeld auf Ω. Weiter nehmen wir an, dass wir so viele Messungen an den Orten $x, x_1, \ldots, x_n \in \Omega$ haben, dass wir mit hinreichender Genauigkeit

$$m_i := E(Z(x_i)), \mu := E(Z(x)), c_{ij} := \operatorname{Cov}(Z(x_i), Z(x_j)), c_i := \operatorname{Cov}(Z(x_i), Z(x))$$

kennen. Diese Größen fassen wir zusammen in den Vektoren und der Matrix

$$\vec{m} := \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}, \ \vec{c} := \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \ C := \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}.$$

Beachte, dass C eine symmetrische Matrix ist. Mehr Information über die Zufallsvariablen $Z(x_i), Z(x)$ haben wir nicht.

Der Ansatz besteht nun darin, dass wir die unbekannte Zufallsvariable $Z(x)$ durch Linearkombinationen der $Z(x_i)$ zu approximieren versuchen:

$$\hat{Z}(x) := \sum_{i=1}^{n} w_i Z(x_i)$$

und dabei die w_i optimal wählen. Auch hier fassen wir die noch zu bestimmenden Koeffizienten w_i zu dem Vektor $\vec{w} := \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$ zusammen. Doch was bedeutet hier Optimalität? Zuallererst muss natürlich der Erwartungswert auf diese Weise wiedergewonnen werden, das heißt wir fordern

$$\mu = E(\hat{Z}(x)) = \sum_{i=1}^{n} w_i m_i = \langle \vec{w}, \vec{m} \rangle.$$

Wie bei der Bestimmung der Regressionsgerade wollen wir zusätzlich $\operatorname{Var}(\hat{Z}(x) - Z(x))$ minimieren. Die Aufgabenstellung lautet damit:

Minimiere

$$\begin{aligned}
\mathrm{Var}(\hat{Z}(x)-Z(x)) &= E(((\hat{Z}(x)-\mu)-(Z(x)-\mu))^2) \\
&= \mathrm{Var}(Z(x)) + \mathrm{Var}(\hat{Z}(x)) - 2\mathrm{Cov}(\hat{Z}(x), Z(x)) \\
&= \mathrm{Var}(Z(x)) + \sum_{i,j=1}^{n} c_{ij} w_i w_j - 2\sum_{i=1}^{n} c_i w_i
\end{aligned}$$

unter der Nebenbedingung

$$\mu = \sum_{i=1}^{n} m_i w_i.$$

Dies löst man durch Einführung des Lagrange-Multiplikators λ. Es muss in den minimierenden w_i gelten

$$\frac{\partial}{\partial w_k}\left(\mathrm{Var}(Z(x)) + \sum_{i,j=1}^{n} c_{ij} w_i w_j - 2\sum_{i=1}^{n} c_i w_i + \lambda(\mu - \sum_{i=1}^{n} m_i w_i)\right) = 0$$

für alle k sowie die Nebenbedingung (die durch Ableiten dieser Funktion nach λ wiedergewonnen werden kann). Führt man diese Ableitungen aus, so erhält man in Matrix-/Vektorenschreibweise

$$2C\cdot\vec{w} - 2\vec{c} - \lambda\vec{m} = 0,\ \ \mu = \langle\vec{w},\vec{m}\rangle.$$

Die linke Gleichung liefert

$$\vec{w} = C^{-1}\cdot(\vec{c} + \frac{\lambda}{2}\vec{m}),$$

eingesetzt in die rechte Gleichung erhalten wir

$$\frac{\lambda}{2} = \frac{\mu - \langle\vec{m}, C^{-1}\cdot\vec{c}\rangle}{\langle\vec{m}, C^{-1}\cdot\vec{m}\rangle}$$

und damit (unter Verwendung der Symmetrie von C)

$$\vec{w} = C^{-1}\cdot\vec{c} + \frac{\mu - \langle C^{-1}\cdot\vec{m}, \vec{c}\rangle}{\langle C^{-1}\cdot\vec{m}, \vec{m}\rangle} C^{-1}\cdot\vec{m}.$$

Haben wir nun Messwerte z_i an den Orten x_i, aber keinen in x, so schätzen wir gemäß unseres Ansatzes den Wert in x durch

$$\langle\vec{w},\vec{z}\rangle = \sum_{i=1}^{n} w_i z_i,$$

wobei hier $\vec{z} := \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$ definiert wurde.

BEISPIEL 7.1. An den Orten D, E und F werden seit 100 Jahren Niederschlagsmengen gemessen. In D fallen im Juli durchschnittlich 20 mm Regen, in E 25 mm und in F 30 mm. Die Kovarianzmatrix zwischen den Orten ist (in Quadratmillimetern)

	D	E	F
D	4	5	1
E	5	9	2
F	1	2	1

Aktuell (2012) gemessen wurden in E 18 mm Niederschlag und in F 26 mm. In D war der Wasserbehälter undicht und es liegt kein Messwert vor. Gesucht ist eine gute Schätzung der Niederschlagsmenge in D im Juli 2012.

Eine Bemerkung zuvor: Aus der Kovarianzmatrix lesen wir auch die Korrelationskoeffizienten ab:

$$\rho(D,E) = \frac{\mathrm{Cov}(D,E)}{\sqrt{\mathrm{Cov}(D,D)\cdot \mathrm{Cov}(E,E)}} = \frac{5}{6}, \rho(D,F) = \frac{1}{2}, \rho(E,F) = \frac{2}{3}.$$

Wir erwarten also eine stärkere Orientierung unserer Schätzung an dem Messwert in E als an dem in F.

In der Kovarianzmatrix erkennen wir als Vektor $\vec{c}$ die unteren beiden Einträge der ersten Spalte und als Matrix C die untere rechte 2×2-Matrix, also

$$\vec{c} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \; C = \begin{pmatrix} 9 & 2 \\ 2 & 1 \end{pmatrix}.$$

Weiter ist $\mu = 20$ und $\vec{m} = \begin{pmatrix} 25 \\ 30 \end{pmatrix}$. Um weiter rechnen zu können, müssen wir zunächst C invertieren. Allgemein gilt für eine 2×2-Matrix $A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

In unserem Fall also

$$C^{-1} = \frac{1}{5}\begin{pmatrix} 1 & -2 \\ -2 & 9 \end{pmatrix}.$$

Wir berechnen die Ingredienzen der Kriging-Formel schrittweise:

$$C^{-1}\cdot\vec{c} = \begin{pmatrix} \frac{3}{5} \\ -\frac{1}{5} \end{pmatrix}, \; C^{-1}\cdot\vec{m} = \begin{pmatrix} -7 \\ 44 \end{pmatrix}$$

und

$$\langle C^{-1}\cdot\vec{m}, \vec{c}\rangle = 9, \; \langle C^{-1}\cdot\vec{m}, \vec{m}\rangle = 1145.$$

Dies setzen wir nun in die Formel ein und erhalten

$$\vec{w} = \begin{pmatrix} \frac{3}{5} + \frac{20-9}{1145}\cdot(-7) \\ -\frac{1}{5} + \frac{20-9}{1145}\cdot 44 \end{pmatrix} \approx \begin{pmatrix} 0.53 \\ 0.22 \end{pmatrix}.$$

Der durch Kriging geschätzte Wert für D ist also $(0.53\cdot 18 + 0.22\cdot 26)mm = 15.3mm$. △

Kriging hat gegenüber anderen, herkömmlichen Interpolationsmethoden den entscheidenden Vorteil, dass plötzliche Änderungen in der geophysikalischen Struktur des Gebiets adäquat erfasst werden: Orte, deren Messwerte völlig unkorreliert zu denen in x sind, werden von der Methode automatisch außer Acht gelassen.

Daher sehen durch Kriging ermittelte graphische Darstellungen geophysikalischer Meßgrößen den tatsächlich vorliegenden Werten auch in der Feinstruktur noch sehr ähnlich.

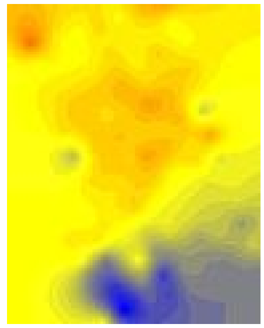


ABBILDUNG 7.1. Die Darstellung (Quelle: Wikipedia, gemeinfreie Graphik erzeugt von Stefoster) zeigt eine Kriging-Interpolation der Ertragswerte eines Ackers.

2. Aufgaben

Aufgabe 7-1 In Kiel, Kassel und auf dem Feldberg im Schwarzwald wurden folgende Monatsmittelwerte der Lufttemperatur für den Monat Mai in Grad Celsius gemessen: [1]

	Kiel-Holtenau	Feldberg	Kassel
2007	12.6	7.9	14.4
2008	13.1	8.6	15.3
2009	12.5	8.6	14.0
2010	9.6	3.9	10.5
2011	12.8	8.7	14.2

(1) Schätzen Sie die Erwartungswerte der Lufttemperatur im Mai in Kiel-Holtenau, in Kassel und auf dem Feldberg anhand der angegebenen Stichproben.
(2) Schätzen Sie die Kovarianzmatrix zwischen den Lufttemperaturen im Mai in Kiel-Holtenau, in Kassel und auf dem Feldberg anhand der angegebenen Stichproben.
(3) Im Jahr 2006 wurde als Mittelwert der Lufttemperatur im Mai in Kassel 13.3 und auf dem Feldberg 6.2 Grad Celsius gemessen. Schätzen Sie mittels Kriging den entsprechenden Wert für Kiel-Holtenau im Mai 2006. Benutzen Sie dabei die geschätzten Erwartungswerte und Kovarianzen aus den vorigen Teilaufgaben.

    Im Jahr 2000 wurde als Mittelwert der Lufttemperatur im Mai in Kassel 15.0 und auf dem Feldberg 8.2 Grad Celsius gemessen. Schätzen Sie wie zuvor den entsprechenden Wert für Kiel-Holtenau im Mai 2000.

    Zum Vergleich: Der Wert für 2006 in Kiel-Holtenau betrug 12.0 Grad und im Jahr 2000 waren es 13.2 Grad.

Aufgabe 7-2 In den Städten U, V, W werden seit langem die Sommertage pro Jahr gezählt (Tageshöchsttemperatur über 25 Grad Celsius). Die gemessenen Erwartungswerte sind

$$\mu_U = 30, \mu_V = 5, \mu_W = 50.$$

Die Kovarianzmatrix ist

	U	V	W
U	9	5	9
V	5	4	6
W	9	6	16

[1] Quelle: Deutscher Wetterdienst

Im Jahr 2011 wurden in U 22 und in V 2 Sommertage gezählt. Der Wert von W liegt nicht vor. Bestimmen Sie die aus den vorliegenden Daten zu erwartende Anzahl von Sommertagen in W.

KAPITEL 8

Einige weitere wichtige Verteilungen

1. Log-Normalverteilung

Wie im zentralen Grenzwertsatz erklärt, ist die Normalverteilung bei der Addition verschiedener Einflüsse anwendbar. Oftmals sind Messgrößen X aber a priori positiv, d.h. die Normalverteilung ist für die negativen Werte eine sehr schlechte Näherung. Hinzu kommt, dass bei prinzipiell positiven Messgrößen die Einflüsse oft multiplikativ wirken. Da $\ln(ab) = \ln(a) + \ln(b)$, wirken die Einflüsse dann auf $\ln X$ additiv.

DEFINITION 8.1. *Eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}^+$ heißt log-normalverteilt mit Parametern $\tilde{\mu}, \tilde{\sigma}$, wenn $\ln X$ $(\tilde{\mu}, \tilde{\sigma})$-normalverteilt ist, d.h. wenn*

$$P(X \in [a,b])(= P(\ln X \in [\ln a, \ln b])) = \Phi(\frac{\ln b - \tilde{\mu}}{\tilde{\sigma}}) - \Phi(\frac{\ln a - \tilde{\mu}}{\tilde{\sigma}}).$$

Dies drückt sich in der Praxis in einer gewissen Skaleninvarianz ist: Ist das Gewicht einer Einflussgröße unabhängig vom aktuellen Wert, so ist die Annahme einer log-Normalverteilung oft gerechtfertigt. Die Log-Normalverteilung beschreibt in der Tat viele Größen wie zum Beispiel den Stammdurchmesser von Bäumen, Durchmesser von Bakterien, Körpergröße des Menschen, Verteilung der Galaxien, Größe von Ölfeldern, Verteilung chemischer Elemente in Sedimenten, Größe von Städten, Partikelgröße in Wolken, Wasserdurchfluss von Flüssen, Windgeschwindigkeiten, ...

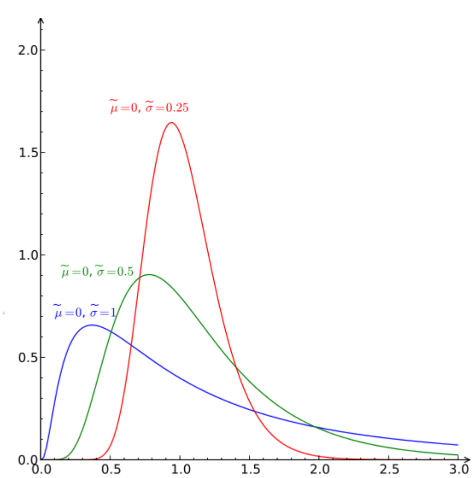


ABBILDUNG 8.1. Das Diagramm zeigt die Wahrscheinlichkeitsdichten von ausgewählten Log-Normalverteilungen mit Parameter $\tilde{\mu} = 0$. Wie bei der Normalverteilung erhält man die Wahrscheinlichekeit $P(X \leq a)$ als Fläche unter der Kurve bis zum Wert a auf der x-Achse. Quelle: Wikipedia, Autor: Mikael Häggström

Ist X log-normalverteilt mit den Parametern $\tilde{\mu}, \tilde{\sigma}$, so ist offenbar $\tilde{\mu} = E(\ln X), \tilde{\sigma} = \sigma(\ln X)$. Gegeben sind aber meistens $\mu := E(X)$ und $\sigma := \sigma(X)$. Wie kann man das ineinander umrechnen?

SATZ 8.2. *Ist X log-normalverteilt mit den Parametern $\tilde{\mu}, \tilde{\sigma}$ und $\mu := E(X), \sigma := \sigma(X)$, so gilt*

$$\mu = e^{\tilde{\mu}+\frac{1}{2}\tilde{\sigma}^2}, \ \sigma^2 = e^{2\tilde{\mu}+\tilde{\sigma}^2}(e^{\tilde{\sigma}^2} - 1).$$

Umgekehrt hat man

$$\tilde{\mu} = \ln \mu - \frac{1}{2}\ln(\frac{\sigma^2}{\mu^2} + 1), \ \tilde{\sigma}^2 = \ln(\frac{\sigma^2}{\mu^2} + 1).$$

BEISPIEL 8.3. Im Wald stehen 1000 Bäume mit durchschnittlichem Stammdurchmesser 80cm und Standardabweichung 30cm. Man weiß, dass Stammdurchmesser log-normalverteilt sind. Schätzen Sie die Anzahl von Bäumen mit Durchmesser mindestens 120cm.

Lösung: Sei X der Stammdurchmesser eines Baumes. Es ist $\mu = 0.8, \sigma = 0.3$ (in Metern). Damit bestimmen wir die Parameter $\tilde{\mu}, \tilde{\sigma}$ als

$$\tilde{\mu} = \ln 0.8 - \frac{1}{2}\ln(\frac{0.09}{0.64} + 1) \approx -0.29, \tilde{\sigma} = \sqrt{\ln(\frac{0.09}{0.64} + 1)} \approx 0.36.$$

Also ist

$$P(X \geq 1.2) = P(\ln X \geq \ln 1.2) = 1 - \Phi(\frac{\ln 1.2 + 0.29}{0.36}) \approx 1 - \Phi(1.31) \approx 0.095.$$

Wir erwarten also $1000 \cdot 0.095 = 95$ Bäume mit Stammdurchmesser mindestens 120cm.

Vergleichen wir das mit der (naiveren) Annahme, dass Stammdurchmesser normalverteilt sind. Dann wäre

$$P(X \geq 1.2) = 1 - \Phi(\frac{1.2 - 0.8}{0.3}) \approx 0.092,$$

d.h. wir würden 92 Bäume mit einem Stammdurchmesser von mehr als 120 cm erwarten. Die Abweichung ist hier klein, kann aber durchaus beträchtlicher ausfallen. △

BEISPIEL 8.4. Die durchschnittliche Windgeschwindigkeiten im Mai ist am Kieler Leuchtturm 13 Knoten. Die Windwahrscheinlichkeit (mindestens 10 Knoten) beträgt im Mai 58%.[1] Die Windgeschwindigkeit ist log-normalverteilt. Berechnen Sie die Parameter $\tilde{\mu}, \tilde{\sigma}$.

Lösung: Sei X die Windgeschwindigkeit. Wir haben $\mu = 13$ (in Knoten) und $P(X \geq 10) = 0.58$ gegeben. Die letzte Gleichung sagt uns

$$0.58 = P(\ln X \geq \ln 10) = 1 - \Phi(\frac{\ln 10 - \tilde{\mu}}{\tilde{\sigma}}) = \Phi(\frac{\tilde{\mu} - \ln 10}{\tilde{\sigma}}).$$

Aus der Tabelle der Normalverteilung lesen wir ab

$$\frac{\tilde{\mu} - \ln 10}{\tilde{\sigma}} = 0.20.$$

Logarithmieren wir die Gleichung $13 = \mu = e^{\tilde{\mu}+\frac{1}{2}\tilde{\sigma}^2}$, so erhalten wir damit zwei Gleichungen für $\tilde{\mu}$

$$\tilde{\mu} = 0.2\tilde{\sigma} + \ln 10 = \ln 13 - \frac{1}{2}\tilde{\sigma}^2.$$

Die daraus resultierende quadratische Gleichung

$$\frac{1}{2}\tilde{\sigma}^2 + 0.2\tilde{\sigma} + \ln\frac{10}{13} = 0$$

[1] Quelle: windfinder.com

hat als einzige positive Lösung

$$\tilde{\sigma} = -0.2 + \sqrt{0.04 + 2\ln\frac{13}{10}} \approx 0.55.$$

Dies eingesetzt in eine der beiden Gleichungen für $\tilde{\mu}$ liefert

$$\tilde{\mu} \approx 2.41.$$

△

2. Poissonverteilung

DEFINITION 8.5. *Eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{N}$ heißt Poisson-verteilt mit Parameter $\lambda > 0$, wenn*

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

gilt.

Anwendung findet die Poissonverteilung bei der Statistik seltener Ereignisse, die einer Binomialverteilung gehorchen. Bei kleinem q und großem n ist der Rechenaufwand für die Binomialverteilung sehr hoch und man benötigt eine effektivere Berechnungsmethode. Diese liefert die Poissonverteilung. Exakt gilt

SATZ 8.6. *Ist $X_n : \Omega \longrightarrow \mathbb{N}$ eine Folge binomialverteilter Zufallsvariablen mit Parametern n und q_n so, dass*

$$\lim_{n\longrightarrow\infty} nq_n = \lambda,$$

so ist

$$\lim_{n\longrightarrow\infty} P(X_n = k) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

In der Praxis wendet man die Poissonverteilung anstatt der Binomialverteilung an, wenn $0 < nq \leq 10$ erfüllt ist, manchmal auch rigoroser erst dann, wenn zusätzlich $p \leq 0.05$ und $n \geq 50$ erfüllt sind.

Ist X Poisson-verteilt mit Parameter $\lambda > 0$, so ist $E(X) = \lambda = \mathrm{Var}(X)$.

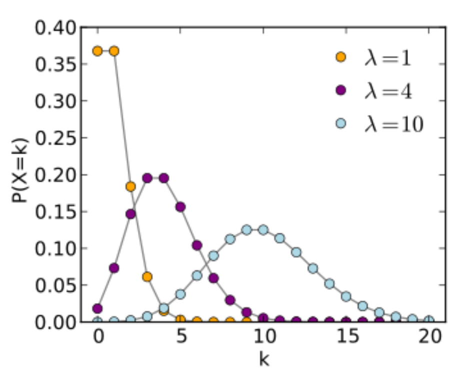


ABBILDUNG 8.2. Das Diagramm zeigt verschiedene Poissonverteilungen. Quelle: Wikipedia, Autor: Skbkekas

BEISPIEL 8.7. Bei einer seriellen Datenübertragung ist die Fehlerwahrscheinlichkeit 10^{-6}. Wie hoch ist die Wahrscheinlichkeit, dass bei der Übertragung eines 2.4 MB-Files höchstens drei Bytes falsch übertragen werden?

Lösung: Die Anzahl X der Fehler ist binomialverteilt und es ist $n = 2.4 \cdot 10^6, q = 10^{-6}$, also $\mu = nq = 2.4 < 10$. Daher wenden wir die Poissonverteilung an und erhalten

$$P(X \leq 3) = e^{-2.4}(1 + 2.4 + \frac{1}{2}2.4^2 + \frac{1}{6}2.4^3) \approx 77,9\%.$$

△

3. Die hypergeometrische Verteilung

Gegeben eine Menge mit N Elementen, haben M davon eine gewisse Eigenschaft A, die restlichen $N - M$ Elemente haben diese Eigenschaft nicht. Wir ziehen aus den N Elementen n zufällig (ohne Zurücklegen). X sei die Anzahl der gezogenen Elemente mit der Eigenschaft A. Dann gilt

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

für $k \in \{0, \dots, n\}$.

DEFINITION 8.8. *Eine Zufallsgröße $X : \Omega \longrightarrow \{0, \dots, n\}$ heißt hypergeometrisch verteilt mit den Parametern n, N, M, wenn*

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

für $k \in \{0, \dots, n\}$ gilt.

BEISPIEL 8.9. In einer Lostrommel befinden sich 12 Lose, davon 4 Gewinnlose. Wie hoch ist die Wahrscheinlichkeit, beim Kauf von 3 Losen 2 Gewinnlose zu haben?

Lösung: Es liegt offenbar eine hypergeometrische Verteilung vor für $X :=$ Anzahl der Gewinnlose nach dem Kauf dreier Lose. Die Eigenschaft A ist Gewinnlos zu sein, $N = 12, M = 4, n = 3, k = 2$. Damit berechnen wir

$$P(X = 2) = \frac{\binom{4}{2}\binom{8}{1}}{\binom{12}{3}} = \frac{12}{55} \approx 21,8\%.$$

△

4. χ^2-Verteilung

Die folgende Verteilung findet Anwendung, wenn man die Genauigkeit einer geschätzten Varianz schätzen will. Dies werden wir in einem späteren Kapitel tun.

Seien $X_1, \dots, X_n$ unabhängige $(0, 1)$-normalverteilte Zufallsvariablen. Die Zufallsvariable

$$\chi_n^2 := X_1^2 + \dots + X_n^2$$

erfüllt

$$E(\chi_n^2) = n,$$

denn wegen den Annahmen $E(X_i) = 0, \mathrm{Var}(X_i) = 1$ ist

$$E(\chi_n^2) = nE(X_1^2) = n(E(X_1^2) - E(X_1)^2) = n\mathrm{Var}(X_1) = n.$$

Die Varianz von χ_n^2 ist komplizierter zu berechnen. Das Ergebnis ist

$$\mathrm{Var}(\chi_n^2) = 2n.$$

Genauer ist

$$P(\chi_n^2 \leq x) = c_n \int_0^x t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt,$$

wobei c_n so ist, dass $c_n \int_0^\infty t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt = 1$.

DEFINITION 8.10. *Eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}^+$ heißt χ^2-verteilt mit n Freiheitsgraden, wenn*

$$P(X \leq x) = c_n \int_0^x t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt$$

(mit dem oben angegebenen c_n).

Die χ^2-Verteilung kann nicht expliziter angegeben werden. Die Werte sind wie für die Normalverteilung tabelliert.

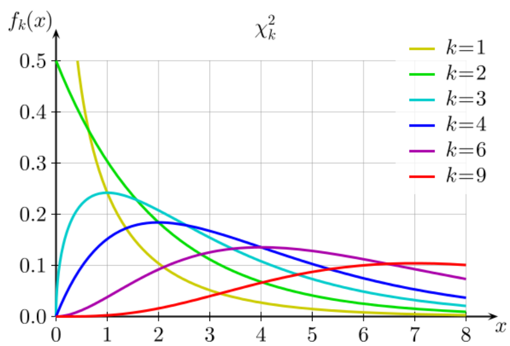


ABBILDUNG 8.3. Das Diagramm zeigt verschiedene χ^2-Wahrscheinlichkeitsdichten. Die Anzahl der Freiheitsgrade heißen im Diagramm k. Quelle: Wikipedia, Autor: Geek3

5. Student-t-Verteilung

Die Student-t-Verteilung wird angewandt, wenn man die Genauigkeit eines geschätzten Erwartungswertes bei unbekannter Varianz schätzen will. Auch dies wird in einem späteren Kapitel erklärt.

DEFINITION 8.11. *Sei Z eine $(0,1)$-normalverteilte Zufallsvariable. Eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$ heißt (Student)-t-verteilt mit n Freiheitsgraden, wenn*

$$P(X \leq x) = P\left(\frac{\sqrt{n}Z}{\sqrt{\chi_n^2}} \leq x\right) = \tilde{c}_n \int_{-\infty}^x \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt,$$

wobei $\tilde{c}_n$ so ist, dass $\tilde{c}_n \int_{-\infty}^x (1 + \frac{t^2}{n})^{-\frac{n+1}{2}} dt = 1$.

Auch die Werte der Student-t-Verteilung sind tabelliert. Die Student-t-Verteilung hat als Erwartungswert und Varianz $E(X) = 0$ für $n > 1$ (für $n = 1$ existiert der Erwartungswert nicht) und $\mathrm{Var}(X) = \frac{n}{n-2}$ für $n > 2$ (für $n \leq 2$ existiert die Varianz nicht).

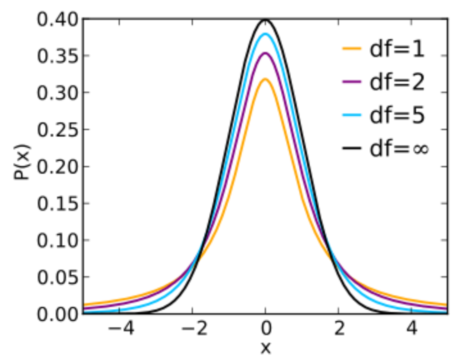


ABBILDUNG 8.4. Das Diagramm zeigt Wahrscheinlichkeitsdichten verschiedener Student-t-Verteilungen. df steht dabei für die Anzahl der Freiheitsgrade ("degrees of freedom"), was wir hier mit n bezeichnet haben. Quelle: Wikipedia, Autor: Skbkekas

Die starke Ähnlichkeit der Student-t-Verteilungen zur Normalverteilung (der Fall genau eines Freiheitsgrades) hat zur Folge, dass man die Varianz nicht allzu genau kennen muss, um die Genauigkeit der Schätzung des Erwartungswertes bestimmen zu können. Genaueres dazu im nächsten Kapitel.

6. Aufgaben

Aufgabe 8-1 Die Wahrscheinlichkeit, innerhalb eines Jahres bei einem (polizeilich erfassten) Verkehrsunfall verletzt zu werden, beträgt etwa 0,5% aktuell[2]. Schätzen Sie die Wahrscheinlichkeit, dass sich unter 1000 zufällig gewählten Personen mindestens 5 befinden, die im vergangenen Jahr bei einem Verkehrsunfall verletzt wurden.

Aufgabe 8-2 Die Einwohnerzahl von Gemeinden in Deutschland ist in etwa log-normalverteilt. Es leben 81,859 Millionen Menschen in 12316 Gemeinden. Von diesen haben 76,6% weniger als 5000 Einwohner.[3] Schätzen Sie mit diesen Informationen die Anzahl der Großstädte (mehr als 100 000 Einwohner).

Aufgabe 8-3 Bei einer täglichen Vogelzählung in einem Gebiet im Wattenmeer werden folgende Anzahlen von Sandregenpfeifern an 5 zufällig ausgewählten Tagen gezählt:

Tag 1	Tag 2	Tag 3	Tag 4	Tag 5
27	35	15	28	45

(1) Schätzen Sie den Erwartungswert und die Standardabweichung der (langfristig erhobenen) Anzahl täglich gezählter Sandregenpfeifer. Geben Sie die verwendeten Formeln an.
(2) Berechnen Sie eine Mindestwahrscheinlichkeit, mit der an einem beliebig gewählten Tag zwischen 15 und 45 Sandregenpfeifer gezählt werden. (Falls Sie Teilaufgabe (1) nicht lösen konnten, verwenden Sie $\bar{x} = 30$ und $s = 10$.)
(3) Schätzen Sie diese Wahrscheinlichkeit unter der Annahme, dass die Anzahlen gezählter Sandregenpfeifer pro Tag log-normalverteilt ist.

[2]Quelle: Destatis

[3]Quelle: Junkernheinrich/Micosatt: Kommunalstrukturen in Deutschland Eine Analyse zur länderübergreifenden Vergleichbarkeit kommunaler Finanzkennzahlen, Bertelsmann-Stiftung 2009

KAPITEL 9

Tests

Unser erstes Ziel ist die Schätzung der Genauigkeit einer Mittelwertschätzung aus n Stichproben. Dabei gehen wir wie folgt vor: Zuerst schätzen wir die Genauigkeit unserer Varianzschätzung. Sind wir mit der Genauigkeit zufrieden, nehmen wir an, die geschätzte Varianz ist die tatsächliche und schätzen damit die Genauigkeit des Erwartungswerts. Sind wir damit nicht zufrieden, so wenden wir eine leicht aufwändigere Methode an, um die Genauigkeit des Erwartungswerts bei unbekannter Varianz zu schätzen.

1. Quantile

Gegeben einen Wahrscheinlichkeitsraum (Ω, P), eine Zufallsvariable X auf Ω und $\beta \in [0,1]$, nennen wir x_β das β-Quantil der Wahrscheinlichkeitsverteilung von X unter P, wenn

$$P(X \leq x_\beta) = \beta.$$

Im Folgenden benötigen wir die Quantile der Standardnormalverteilung, der χ^2- und der t-Verteilungen. Daher bezeichnen wir sie separat:

- z_β heiße das β-Quantil, wenn X $(0,1)$-normalverteilt ist,
- $\chi^2_{n,\beta}$ heiße das β-Quantil, wenn X χ^2-verteilt mit n Freiheitsgraden ist,
- $t_{n,\beta}$ heiße das β-Quantil, wenn X Student-t-verteilt mit n Freiheitsgraden ist.

Im Anhang finden Sie Tabellen einiger wichtiger Quantile.

2. Konfidenzintervalle und -schranken

Gegeben eine Zufallsvariable X, eine Messung x und ein "Konfidenzniveau" $\gamma \in [0,1]$, wollen wir a so bestimmen, dass die Wahrscheinlichkeit, dass $E(X)$ sich im Intervall $[x-a, x+a]$ befindet, gleich γ ist. Unter leichtem Missbrauch der Notation schreiben wir dafür

$$P(E(X) \in [x-a, x+a]) = \gamma.$$

In der Praxis wird γ oft 95% oder 99% gewählt. Das Intervall $[x-a, x+a]$ heißt Konfidenzintervall für $E(X)$ zum Konfidenzniveau γ. Manchmal interessiert man sich nur für einseitige "Konfidenzschranken": a_+ bzw. a_- so, dass

$$P(E(X) \leq x + a^+) = \gamma \text{ bzw. } P(E(X) \geq x - a_-) = \gamma.$$

3. Verteilung der Schätzer

Erfolg können wir bei der Bestimmung des Konfidenzintervalls oder der -schranken nur haben, wenn wir die zugrunde liegende Form der Wahrscheinlichkeitsverteilung kennen. Der zentrale Grenzwertsatz sagt aus, dass wir diese für die Mittelung unabhängiger Kopien $X_1, \ldots, X_n$ von X für große n asymptotisch kennen: sie wird immer normalverteilter. Ähnliches gilt für die Schätzer der Varianz und des Erwartungswertes bei unbekannter Varianz.

SATZ 9.1. *Seine $X_1, \ldots, X_n$ unabhängige Kopien einer* normalverteilten *Zufallsvariablen X mit Erwartungswert μ und Varianz σ^2. Weiter seien $\bar{X}$ sowie S^2 die Schätzungen für den Erwartungswert und die Varianz X wie in Kapitel 5. Dann ist*

(1) $T_\mu := \sqrt{n}\frac{\bar{X}-\mu}{S}$ *Student-t-verteilt mit $n-1$ Freiheitsgraden und*
(2) $Z_\sigma := \frac{n-1}{\sigma^2}S^2$ *ist χ^2-verteilt mit $n-1$ Freiheitsgraden.*

Dasselbe gilt asymptotisch für große n für beliebig verteiltes X.

Im Rest des Kapitels setzen wir folgende Situation voraus: Wir haben eine Zufallsstichprobe $x_1, \ldots, x_n$ einer Zufallsvariablen X und wollen die Genauigkeit der empirischen Varianz s^2 und des empirische Mittelwerts $\bar{x}$ schätzen. Dabei nehmen wir stets an, dass X normalverteilt ist oder, dass n groß ist; so, dass Satz 9.1 anwendbar ist.

4. Konfidenzintervall und -schranken für die Varianz

Um ein möglichst einfaches Konfidenzintervall angeben zu können, modifizieren wir hier das Problem leicht und erlauben auch unsymmetrische Intervalle um s^2 der Form $[c \cdot s^2, C \cdot s^2]$ mit $c < 1, C > 1$. Wir wollen c, C so haben, dass

$$\begin{aligned}\gamma &= P(\sigma^2 \in [cs^2, Cs^2]) = P(s^2 \in [\frac{1}{C}\sigma^2, \frac{1}{c}\sigma^2]) = P(S^2 \in [\frac{1}{C}\sigma^2, \frac{1}{c}\sigma^2]) \\ &= P(S^2 \leq \frac{1}{c}\sigma^2) - P(S^2 \leq \frac{1}{C}\sigma^2).\end{aligned}$$

Dies ist offenbar insbesondere dann erfüllt, wenn $P(S^2 \leq \frac{1}{c}\sigma^2) = \frac{1+\gamma}{2}$ und $P(S^2 \leq \frac{1}{C}\sigma^2) = \frac{1-\gamma}{2}$, d.h. wenn

$$P(Z_\sigma \leq \frac{n-1}{c}) = \frac{1+\gamma}{2},\ P(Z_\sigma \leq \frac{n-1}{C}) = \frac{1-\gamma}{2}.$$

Nach Satz 9.1 bedeutet dies wiederum

$$c = \frac{n-1}{\chi^2_{n-1,\frac{1+\gamma}{2}}},\ C = \frac{n-1}{\chi^2_{n-1,\frac{1-\gamma}{2}}}.$$

Mit wachsendem n wird das so bestimmte Konfidenzintervall immer symmetrischer um den Erwartungswert $n-1$.

Wir erhalten damit und mit ähnlicher Rechnung

SATZ 9.2. *Seien $x_1, \ldots, x_n$ eine Zufallsstichprobe der Zufallsvariable X. Es sei X normalverteilt oder n groß. Zum Konfidenzniveau γ ist $[\frac{n-1}{\chi^2_{n-1,\frac{1+\gamma}{2}}}s^2, \frac{n-1}{\chi^2_{n-1,\frac{1-\gamma}{2}}}s^2]$ ein (unsymmetrisches) Konfidenzintervall, d.h.*

$$P(\sigma^2 \in [\frac{n-1}{\chi^2_{n-1,\frac{1+\gamma}{2}}}s^2, \frac{n-1}{\chi^2_{n-1,\frac{1-\gamma}{2}}}s^2]) = \gamma.$$

Die obere Konfidenzschranke ist $\frac{n-1}{\chi^2_{n-1,1-\gamma}}s^2$ und die untere Konfidenzschranke ist $\frac{n-1}{\chi^2_{n-1,\gamma}}s^2$, d.h.

$$P(\sigma^2 \leq \frac{n-1}{\chi^2_{n-1,1-\gamma}}s^2) = \gamma = P(\sigma^2 \geq \frac{n-1}{\chi^2_{n-1,\gamma}}s^2).$$

BEISPIEL 9.3. **(Zwei Stichproben I)** Zwei Stichproben (der Zufallsvariablen X und Y) mit $n = 20$ liefern Mittelwerte von $\bar{x} = 60$ und $\bar{y} = 50$. Die empirischen Standardabweichungen seien $s_x = 30$

bzw $s_y = 40$. Wir bestimmen die Konfidenzintervalle von s_x und s_y zu den Konfidenzniveaus 95% und 99%. Wir benötigen also die Tabellenwerte $\chi^2_{19,0.005}, \chi^2_{19,0.025}, \chi^2_{19,0.975}$ und $\chi^2_{19,0.995}$. Diese sind

$$\chi^2_{19,0.005} = 6.84,\ \chi^2_{19,0.025} = 8.91,\ \chi^2_{19,0.975} = 32.85,\ \chi^2_{19,0.995} = 38.58.$$

Damit sind die Konfidenzintervalle $[\sqrt{\frac{19}{\chi^2_{19,\frac{1+\gamma}{2}}}}s, \sqrt{\frac{19}{\chi^2_{19,\frac{1-\gamma}{2}}}}s]$ der Standardabweichungen für $\gamma = 0.95, 0.99$ und X und Y

σ	X	Y
$\gamma = 0.95$	[22.82,43.81]	[30.43,58.41]
$\gamma = 0.99$	[21.05,50.00]	[28.07,66.67]

$\triangle$

5. Konfidenzintervall und -schranken für den Erwartungswert bei unbekannter Varianz

Ist die Varianz von X unbekannt, so müssen wir s^2 stattdessen verwenden und damit die Zufallsvariable T_μ betrachten. Ähnlich wie zuvor argumentieren wir

$$\begin{aligned}\gamma &= P(\mu \in [\bar{x} - a, \bar{x} + a]) = P(\bar{x} \in [\mu - a, \mu + a]) = P(\bar{X} \in [\mu - a, \mu + a]) \\ &= P(T_\mu \in [-\frac{\sqrt{n}a}{s}, \frac{\sqrt{n}a}{s}]) = 2P(T_\mu \le \frac{\sqrt{n}a}{s}) - 1,\end{aligned}$$

letzteres wegen der Symmetrie der Student-t-Verteilung um 0 (wie bei der Standardnormalverteilung). Das löst man wiederum auf zu

$$a = \frac{s}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}.$$

Damit und mit ähnlicher Rechnung bekommen wir

SATZ 9.4. *Seien $x_1, \ldots, x_n$ eine Zufallsstichprobe der Zufallsvariable X. Es sei X normalverteilt oder n groß. Zum Konfidenzniveau γ ist das Konfidenzintervall für den Erwartungswert bei unbekannter Varianz $[\bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}, \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}]$, d.h.*

$$P(\mu \in [\bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}, \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}]) = \gamma.$$

Weiter gilt

$$P(\mu \le \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\gamma}) = \gamma = P(\mu \ge \bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\gamma}),$$

d.h. die obere Konfidenzschranke für den Erwartungswert bei unbekannter Varianz von X ist $\bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\gamma}$ und die untere Konfidenzschranke ist $\bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\gamma}$, jeweils zum Konfidenzniveau γ.

BEISPIEL 9.5. **(Zwei Stichproben II)** Wir führen das Beispiel aus dem vorigen Abschnitt fort ($n = 20, \bar{x} = 60, \bar{y} = 50, s_x = 30, s_y = 40$). Angesichts der großen Konfidenzintervalle für die Varianz entscheiden wir uns, die Konfidenzintervalle für die Erwartungswerte mit unbekannter Varianz zu berechnen. Wir benötigen also die Tabellenwerte

$$t_{19,0.975} = 2.093 \text{ und } t_{19,0.995} = 2.861.$$

Damit erhalten wir als Konfidenzintervalle für die Erwartungswerte

μ	X	Y
$\gamma = 0.95$	[45.96,74.04]	[31.28,68.72]
$\gamma = 0.99$	[40.81,79.19]]	[24.41,75.59]

△

6. Konfidenzintervall und -schranken für den Erwartungswert bei bekannter Varianz

Ist nun zusätzlich die Varianz σ^2 von X bekannt, d.h. in der Praxis, dass das Konfidenzintervall für die Varianz uns klein genug erscheint, dann können wir statt T_μ die Zufallsvariable

$$N_\mu := \sqrt{n}\frac{\overline{X} - \mu}{\sigma}$$

betrachten, von der wir aus dem zentralen Grenzwertsatz wissen, dass sie einer $(0, 1)$-Normalverteilung zustrebt bzw. $(0, 1)$-normalverteilt ist, wenn X normalverteilt ist. Wir wiederholen also Argumente aus dem vorigen Abschnitt, nur unter Verwendung der (von n unabhängigen und damit leichter handhabbaren) Standardnormalverteilung anstatt der Student-t-Verteilung und von σ anstatt von s. Das Ergebnis ist also

SATZ 9.6. *Seien $x_1, \dots, x_n$ eine Zufallsstichprobe der Zufallsvariable X. Es sei X normalverteilt oder n groß. Zum Konfidenzniveau γ ist das Konfidenzintervall für den Erwartungswert bei bekannter Varianz σ^2 von X $[\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}]$, d.h.*

$$P(\mu \in [\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}]) = \gamma.$$

Weiter gilt

$$P(\mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_\gamma) = \gamma = P(\mu \geq \bar{x} - \frac{\sigma}{\sqrt{n}} z_\gamma),$$

d.h. die obere Konfidenzschranke für den Erwartungswert bei unbekannter Varianz von X ist $\bar{x} + \frac{\sigma}{\sqrt{n}} z_\gamma$ und die untere Konfidenzschranke ist $\bar{x} - \frac{\sigma}{\sqrt{n}} z_\gamma$, jeweils zum Konfidenzniveau γ.

Meist wird diese vereinfachte Version angewandt, wenn $n \geq 30$.

BEISPIEL 9.7. **(Zwei Stichproben III)** Zu Demonstrationszwecken wenden wir diese vereinfachte Methode auf unser Beispiel der zwei Stichproben bereits bei $n = 20$ an. Die benötigten Quantile sind

$$z_{0.975} = 1.96, \; z_{0.995} = 2.57$$

und die daraus errechneten Konfidenzintervalle

μ	X	Y
$\gamma = 0.95$	[46.85,73.15]	[32.47,67.53]
$\gamma = 0.99$	[42.76,77.24]]	[27.01,72.99]

Diese Intervalle unterscheiden sich nicht mehr stark, aber spürbar von den exakteren aus der Schätzung mit unbekannter Varianz. Die Verwendung dieser Vereinfachung ab $n = 30$ scheint sinnvoll zu sein. △

7. Hypothesentests für Erwartungswerte

Wir haben eine Zufallsvariable X gegeben und Stichproben $x_1, \dots, x_n$. Aus theoretischen Erwägungen haben wir eine Hypothese über den Erwartungswert μ von X, die wir anhand der Stichprobe testen wollen. Unsere Hypothese hat eine der drei Formen

(1) $\mu = \mu_0$,
(2) $\mu \geq \mu_0$, oder
(3) $\mu \leq \mu_0$.

für ein $\mu_0 \in \mathbb{R}$.

Ziel des Tests ist eine Widerlegung der Hypothese. Um ein solches Ergebnis verkünden zu können, wollen wir uns möglichst sicher sein. Zunächst müssen wir also eine Irrtumswahrscheinlichkeit α vorgeben, auch Signifikanzniveau genannt. α soll die Wahrscheinlichkeit sein, dass wir die Hypothese verwerfen, obwohl sie richtig ist. α soll also möglichst klein sein.

Betrachten wir die Hypothese $\mu = \mu_0$. Wir stellen die Regel auf, dass wir sie nicht verwerfen, wenn $\bar{x}$ hinreichend nahe an μ_0 liegt. Wenn sie wahr ist, wollen wir sie mit Wahrscheinlichkeit $1 - \alpha$ akzeptieren. Wir suchen also ein a so, dass

$$P(\mu \in [\bar{x} - a, \bar{x} + a]) = 1 - \alpha.$$

Dieses Problem haben wir aber bereits oben gelöst. Analoge Überlegungen sind auf die beiden anderen Hypothesen anwendbar. Unsere Regel lautet damit

SATZ 9.8.
(1) *Wir verwerfen die Hypothese $\mu = \mu_0$ zum Signifikanzniveau α nicht, wenn μ_0 im Konfidenzintervall des Erwartungswertes zum Konfidenzniveau $1 - \alpha$ liegt. Andernfalls verwerfen wir sie.*
(2) *Wir verwerfen die Hypothese $\mu \geq \mu_0$ zum Signifikanzniveau α nicht, wenn μ_0 unterhalb der oberen Konfidenzschranke des Erwartungswertes zum Konfidenzniveau $1 - \alpha$ liegt. Andernfalls verwerfen wir sie.*
(3) *Wir verwerfen die Hypothese $\mu \leq \mu_0$ zum Signifikanzniveau α nicht, wenn μ_0 oberhalb der unteren Konfidenzschranke des Erwartungswertes zum Konfidenzniveau $1 - \alpha$ liegt. Andernfalls verwerfen wir sie.*

Solche Hypothesentests haben natürlich nur begrenzten Aussagewert. Außer der Irrtumswahrscheinlichkeit α ist eine andere Fehlerquelle, nicht verworfene Hypothesen für wahr zu halten. Zu einem gegebene Signifikanzniveau gibt es ein ganzes Intervall von μ_0's, für die die entsprechende Hypothese nicht verworfen werden kann. Natürlich kann nur eine davon wahr sein. Daher ist es wichtig, andere (theoretische) Gründe für das Aufstellen der Hypothese anführen zu können. Und schließlich ist mit der Festlegung von α noch keine Aussage über die Wahrscheinlichkeit der irrtümlichen Akzeptanz der Hypothese gemacht. Dies zeigt, dass man sich sehr wohl überlegen muss, ob man die Hypothese oder ihre Negation testen will. Der Test ist so konzipiert, dass man möglichst selten eine wahre Hypothese verwirft. Die zu testende Hypothese muss also so gewählt werden, dass die Konsequenzen einer fälschlichen Verwerfung schlimmer sind als die einer irrtümlichen Akzeptanz.

In der Realität z.B. medizinischer Studien sind Signifikanzniveaus von 5% und 1% gängig. Das heißt aber auch, dass man nach 20 Studien zum Signifikanzniveau 5% mit hoher Wahrscheinlichkeit mindestens eine findet, die die bezweifelte Hypothese tatsächlich verwirft, auch, wenn sie wahr ist. In der Physik werden einer sehr gut bestätigten Theorie widersprechende Messungen erst dann als Widerlegung akzeptiert, wenn die Wahrscheinlichkeit einer irrtümlichen Verwerfung (resultierend aus statistisch modellierten Messunsicherheiten) unterhalb von etwa $\Phi(-5) \approx 10^{-7}$ liegt.

BEISPIEL 9.9. **(Zwei Stichproben IV)** Nehmen wir an, aus theoretischen Erwägungen heraus würden wir sowohl für X als auch Y einen Erwartungswert von 75 erwarten. Die zu testenden Hypothesen lauten $\mu_x = 75$ und $\mu_y = 75$. Das ist auch sinnvoll: Eine Verwerfung der Hypothesen würde wohl eine ganze wissenschaftliche Theorie entwerten. Wenn wir das behaupten wollen, sollten wir unserer Sache sehr sicher sein. Wir testen sie zu den Signifikanzniveaus $\alpha = 5\%$ und $\alpha = 1\%$. Aus den in Abschnitt 5 errechneten Konfidenzintervallen ergibt sich folgendes Akzeptanz-/Verwerfungsschema

$\mu = 75?$	X	Y
$\alpha = 0.05$	verworfen	verworfen
$\alpha = 0.01$	akzeptiert	akzeptiert

Die unterschiedlichen Entscheidungen bei unterschiedlichem Signifikanzniveau spiegeln wieder, dass wir bei geringerer Irrtumswahrscheinlichkeit α die Hypothese eher akzeptieren, um sie nicht fälschlicherweise zu verwerfen. Bei vereinfachter Berechnung der Konfidenzintervalle wie im Beispiel in Abschnitt 6 ergäbe sich das Schema

$\mu = 75?$	X	Y
$\alpha = 0.05$	verworfen	verworfen
$\alpha = 0.01$	akzeptiert	verworfen

Eine möglichst genaue Berechnung der Konfidenzintervalle kann also von Bedeutung sein. △

BEISPIEL 9.10. Das Lottospiel "6 aus 49" wird im allgemeinen so modelliert, dass bei jeder Ziehung jede (verbleibende) Zahl gleich wahrscheinlich gezogen wird. Der Ereignisraum $\Omega := \{A \subseteq \{1, \ldots, 49\} |\ A \text{ hat sechs Elemente }\}$ wird daher als Laplace-Wahrscheinlichkeitsraum angesehen. Die empirischen relativen Häufigkeiten einzelner Zahlen sind aber nicht gleich, zum Beispiel[1] kommen beim Samstagslotto auf drei Ziehungen der 13 mehr als vier Ziehungen der 49. Wir wollen testen, ob das mit der Annahme einer Gleichverteilung noch kompatibel ist.

Genauer heißt das: Sei für eine Ziehung $A \subseteq \{1, \ldots, 49\}$

$$X(A) := \begin{cases} 1 & \text{, falls } 13 \in A \\ 0 & \text{sonst.} \end{cases}$$

Bis zum Zeitpunkt der Datenerhebung gab es $N := 3005$ Ziehungen, davon 303 mit einer 13. Wie bei der Binomialverteilung kann die empirische Varianz einer 0/1-wertigen Zufallsgröße aus dem empirischem Erwartungswert berechnet werden:

$$s^2 = \frac{N}{N-1}\bar{x}(1-\bar{x}),$$

wobei hier $\bar{x} = \frac{303}{3005} \cong 0.101$. Also ist $s \cong 0.300$.

Wären die Ziehungen gleichverteilt, so würden wir als Erwartungswert von X

$$\mu = \frac{6}{49} \cong 0.122$$

erwarten.

Wir wollen nun die Hypothese $\mu = 0.122$ zum Signifikanzniveau $\alpha = 1\%$ testen. Da $N > 30$, verwenden wir die vereinfache Version mit Hilfe der Normalverteilung. Das Konfidenzintervall ist also in unserem Fall

$$[0.101 - \frac{0.300}{\sqrt{3005}} z_{0.995}, 0.101 + \frac{0.300}{\sqrt{3005}} z_{0.995}] \cong [0.087, 0.115]$$

mit $z_{0.995} = 2.58$. Der Wert 0.122 liegt nicht im Konfidenzintervall, also kann die Hypothese mit einer Irrtumswahrscheinlichkeit von weniger als 1% verworfen werden.

Nun haben wir aber die 13 genau deshalb ausgewählt, weil sie die kleinste relative Häufigkeit aufwies. Wir wollen also wissen, wie wahrscheinlich es ist, dass mindestens eine der 49 Zahlen eine relative Häufigkeit hat, so dass der Gleichverteilungserwartungswert 0.122 außerhalb des Konfidenzintervalls zum Konfidenzniveau 1% liegt. Dies Wahrscheinlichkeit ist $1 - 0.99^{49} \approx 38.9\%$. Ein

[1] Quelle: lottodatenbank.de

Fall wie die Seltenheit der 13 hier tritt also mit Wahrscheinlichkeit von 38.9% nach 3005 Ziehungen ein. △

8. Hypothesentest für Differenzen von Mittelwerten

Sind X und Y zwei Messgrößen, die an n (Zeit-)Punkten die jeweiligen Werte $x_1, \ldots, x_n$ bzw. $y_1, \ldots, y_n$ ergeben, und wollen wir die Hypothesen $E(X) = E(Y), E(X) \geq E(Y), E(X) \leq E(Y)$ testen, so bilden wir die Zufallsgröße $Z := X - Y$, errechnen aus den Messwerten $z_i := x_i - y_i$ die empirischen Schätzungen $\bar{z}$ und s_z und testen die Hypothesen $E(Z) = 0, E(Z) \geq 0, E(Z) \leq 0$ wie oben beschrieben. Dabei kommt die empirische Kovarianz zwischen X und Y zum Tragen. Genauer gilt

$$\bar{z} = \bar{x} - \bar{y},\ s_z^2 = s_x^2 + s_y^2 - 2c_{x,y}.$$

Eine andere Situation liegt aber vor, wenn zwei Messreihen $x_1, \ldots, x_n$ von X und $y_1, \ldots, y_m$ von Y (es darf durchaus $n \neq m$ sein) gegeben sind, die nicht in vergleichbaren Umständen gemessen wurden (x_i und y_i wurden nicht am gleichen (Zeit-)Punkt gemessen). Zum Beispiel könnten die zwei Stichproben Messungen der gleichen physikalischen Größe von verschiedenen Forschergruppen sein. Hier werden wie oben die Messungen $x_1, \ldots, x_n$ durch unabhängige Kopien $X_1, \ldots, X_n$ und $y_1, \ldots, y_m$ durch unabhängige Kopien $Y_1, \ldots, Y_m$ von Y modelliert, jedoch dürfen wir nun annehmen, dass auch $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ unabhängig sind. Die Zufallsgrößen $\bar{X}$ und $\bar{Y}$ sind dann ebenfalls unabhängig und die Rechenregeln für Erwartungswert und Varianz implizieren

$$\delta := E(\bar{X} - \bar{Y}) = E(X) - E(Y), \tau^2 := \mathrm{Var}(\bar{X} - \bar{Y}) = \frac{1}{n}\mathrm{Var}(X) + \frac{1}{m}\mathrm{Var}(Y).$$

Sind X und Y normalverteilt, so kann man zeigen, dass dann in der Tat $\bar{X} - \bar{Y}$ (δ, τ)-normalverteilt ist. Für beliebige X und Y gilt dies nach dem zentralen Grenzwertsatz immer noch asymptotisch für große n und m.

Gegeben ein Signifikanzniveau α, suchen wir zunächst das Konfidenzintervall zum Konfidenzniveau $1 - \alpha$, so dass gilt

$$1 - \alpha = P(\delta \in [\bar{x} - \bar{y} - a, \bar{x} - \bar{y} + a]) = P(\bar{X} - \bar{Y} \in [\delta - a, \delta + a]).$$

8.1. Test bei bekannten Varianzen $\sigma(X)^2$ und $\sigma(Y)^2$. Bei bekannten Varianzen kann man τ als bekannt voraussetzen und weiter rechnen

$$1 - \alpha = 2\Phi(\frac{a}{\tau}) - 1,$$

also ist $a = z_{1-\frac{\alpha}{2}}\tau$ und das Konfidenzintervall

$$[\bar{x} - \bar{y} - \tau z_{1-\frac{\alpha}{2}}, \bar{x} - \bar{y} + \tau z_{1-\frac{\alpha}{2}}].$$

Ähnlich berechnet man die obere Konfidenzschranke als $\bar{x} - \bar{y} + \tau z_{1-\alpha}$ und die untere als $\bar{x} - \bar{y} - \tau z_{1-\alpha}$. Nun testen wir die Hypothesen $\delta = 0, \delta \geq 0, \delta \leq 0$ wie in Abschnitt 7 beschrieben.

BEISPIEL 9.11. **(Zwei Stichproben V)** Die Hypothese ist, dass X und Y den gleichen Erwartungswert haben. Wir setzen hier also $\tau := \sqrt{\frac{1}{20}(30^2 + 40^2)} = 11.18$ und benötigen wieder nur $z_{0.975} = 1.96$ und $z_{0.995} = 2.57$. Die entsprechenden Konfidenzintervalle sind

$\alpha = 0.05$	[-11.91,31.91]
$\alpha = 0.01$	[-18.73,38.73]

Beide enthalten 0 und damit kann die Hypothese zu keinem der beiden Signifikanzniveaus verworfen werden. Dafür hätte es freilich gereicht, zuerst nur das Konfidenzintervall zum Signifikanzniveau 5% zu bestimmen. Wenn das schon 0 enthält, dann erst recht das zum Signifikanzniveau 1%. Erst beim Signifikanzniveau 40% erhalten wir (mit $z_{0.8} = 0.84$) ein Konfidenzintervall, das die 0 gerade so nicht enthält, nämlich $[1.06, 18.94]$.

Um die beiden anderen Hypothesen $E(X) \geq E(Y)$ und $E(X) \leq E(Y)$ zu testen, bestimmen wir $z_{0.95} = 1.64, z_{0.99} = 2.33$ und errechnen die Konfidenzschranken

Konfidenzschranken	obere	untere
$\alpha = 0.05$	28.34	-8.34
$\alpha = 0.01$	36.05	-16.05

Wenig überraschend kann keine der beiden Hypothesen zum Signifikanzniveau 5% verworfen werden; erst recht nicht zum Signifikanzniveau 1%. Erst zum Signifikanzniveau 20% kann die Hypothese $E(X) \leq E(Y)$ verworfen werden (mit gleicher Rechnung wie zum Konfidenzintervall beim Signifikanzniveau 40%). △

8.2. Test bei unbekannten Varianzen. Hier nehmen wir zusätzlich $n = m$ an. Seien s_x, s_y die empirischen Varianzen der Stichproben und S_X, S_Y die schätzenden Zufallsgrößen wie zuvor. Dann betrachten wir hilfsweise die Differenzen $z_i := x_i - y_i$. Diese werden modelliert durch unabhängige Kopien $Z_i := X_i - Y_i$ von $Z := X - Y$, wobei jedoch $X_1, \dots X_n, Y_1, \dots, Y_n$ als unabhängig angenommen werden dürfen. Daher ist

$$E(S_Z^2) = E(S_X^2) + E(S_Y^2),$$

wobei S_Z die Varianz von Z schätzt. Wir dürfen also das Konfidenzintervall für den Erwartungswert von Z aus Abschnitt 5 mit den Parametern $\tilde{s}_z := \sqrt{s_x^2 + s_y^2}$ und dem Mittelwert $\bar{z} = \bar{x} - \bar{y}$ verwenden, also

$$P(E(X) - E(Y) \in [\bar{z} - \frac{\tilde{s}_z}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}, \bar{z} + \frac{\tilde{s}_z}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}]) = \gamma.$$

Beachte, dass der Term $\frac{\tilde{s}_z}{\sqrt{n}}$ tatsächlich die empirische Version des Parameters τ aus Abschnitt 8.1 für $n = m$ ist. Entsprechend hat man die Konfidenzschranken

$$P(E(X) - E(Y) \leq \bar{z} + \frac{\tilde{s}_z}{\sqrt{n}} t_{n-1,\gamma}) = \gamma = P(E(X) - E(Y) \geq \bar{z} - \frac{\tilde{s}_z}{\sqrt{n}} t_{n-1,\gamma}).$$

Für die Hypothesen $E(X) = E(Y), E(X) \leq E(Y), E(X) \geq E(Y)$ testet man wieder entsprechend, ob 0 im Konfidenzintervall oder diesseits der Schranken liegt.

BEISPIEL 9.12. **(Zwei Stichproben VI)** Wir erinnern uns an die Werte bzw. bestimmen

$$t_{19,0.975} = 2.093,\ t_{19,0.995} = 2.861 \text{ und } t_{19,0.8} = 0.86$$

ermitteln $\tilde{s}_z = \sqrt{30^2 + 40^2} = 50$ und berechnen daraus die Konfidenzintervalle

$\alpha = 0.05$	[-13.40,33.40]
$\alpha = 0.01$	[-21.99,41.99]
$\alpha = 0.4$	[0.38,19.62]

In der Tat kann also auch noch nach dem exakteren Test die Hypothese $E(X) = E(Y)$ zum Signifikanzniveau 40% noch knapp verworfen werden, aber natürlich nicht zu den seriöseren Signifikanzniveaus 5% und 1%. △

9. Aufgaben

Aufgabe 9-1 Im September der Jahre 2000 bis 2010 sind folgende Mengen an Niederschlag (in mm) in Kiel-Holtenau gefallen: [2]

Jahr	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Niederschlag	63.2	144.2	16.7	37.8	80.1	30.5	29.3	65.0	57.9	33.1	132.4

Da die Annahme, dass die betrachteten Niederschlagsmengen etwa log-normalverteilt sind vernünftiger ist als die Annahme einer Normalverteilung, betrachten wir als Zufallsgröße

$$X(a) := \ln(\text{Niederschlagsmenge in Kiel im September des Jahres } a)$$

und nehmen an, dass X über lange Zeiträume betrachtet etwa normalverteilt ist und die angegebenen Messwerte eine Zufallsstichprobe von X darstellen.

(1) Schätzen Sie den (langjährigen) Erwartungswert und die (langjährige) Varianz von X.
(2) Geben Sie die Konfidenzintervalle für den Erwartungswert und die Varianz von X für die Konfidenzniveaus 90% und 95% an.
(3) Geben Sie die Konfidenzintervalle für den Erwartungswert für die Konfidenzniveaus 90% und 95% unter der Annahme an, dass die Varianz mit der geschätzten übereinstimmt.

Aufgabe 9-2 In den Städten D und E wird seit langem die jährliche Sonnenscheindauer gemessen. Folgende Messreihen liegen vor (Angaben in Stunden):

	1986	1990	1993	1996	2000	2001	2002	2003	2008
D	1800	1500	k.A.	2100	k.A.	1900	1700	2500	1800
E	1100	k.A.	1300	1900	1400	2300	k.A.	1600	1500

Testen Sie die Hypothese "In D und E scheint die Sonne im Durchschnitt gleich lange" auf dem Signifikanzniveau 5%

(1) unter Verwendung nur der Jahresdaten, die in beiden Städten vorliegen.
(2) unter Verwendung aller Daten von D und E separat, d.h. unter Vernachlässigung des Wissens um die Gleichzeitigkeit von Messungen in D und E.

[2] Quelle: Deutscher Wetterdienst

ANHANG A

Weitere Kennzahlen

1. Spannweite

Gegeben eine Stichprobe $x_1, \ldots, x_n$, heißt $R := \max_{i=1,\ldots,n} x_i - \min_{i=1,\ldots,n} x_i$ die Spannweite der Stichprobe. Sie wird auch benutzt, um die Standardabweichung zu schätzen. Allerdings funktioniert das nur für kleine n, denn bei großen n werden große Ausreißer in beiden Richtungen auftreten und deren Größe wird immer schlechter kontrollierbar.

2. Schiefe

Nach der Betrachtung des Erwartungswertes $E(X)$ als linearer Größe in X und der Varianz $\mathrm{Var}(X)$ als quadratischer Größe, betrachtet man zunächst die kubische Größe $E((X - E(X))^3)$. Ist diese Größe positiv, so ist der Graph der Dichte der Wahrscheinlichkeitsverteilung nach links geneigt, ist sie negativ, so ist er nach rechts geneigt. Um ein von Einheiten unabhängiges Maß zu erhalten, definiert man

$$v(X) := \frac{E((X - E(X))^3)}{\sigma(X)^3}$$

und nennt das die Schiefe von X. Ist der Graph der Dichte symmetrisch um den Erwartungswert, d.h. ist $P(X - E(X) = k) = P(X - E(X) = -k)$ für alle k bei endlich vielen Werten von X, so ist $v(X) = 0$.

Für eine Stichprobe $x_1, \ldots, x_n$ mit $n \geq 3$ ist

$$v := \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (\frac{x_i - \bar{x}}{s})^3$$

die empirische Schiefe der Stichprobe. Der Vorfaktor $\frac{n}{(n-1)(n-2)}$ ist wieder dafür notwendig, um damit die tatsächliche Schiefe von X schätzen zu können. Der Grund ist ähnlich wie bei der Varianz: Erst mit 3 Messwerten kann eine Aussage in dritter Ordnung gemacht werden: Mit zwei Messwerten ist eine Neigung der Verteilung nicht erkennbar, eine naiv mit dem Vorfaktor $\frac{1}{n}$ definierte Schiefe wäre 0.

3. Median

Der Median $m(X)$ einer Zufallsvariable X ist das Quantil $x_{0.5}$. Er ist dadurch gegeben, dass die Wahrscheinlichkeit für einen Wert größer als der Median gleich der Wahrscheinlichkeit für einen Wert kleiner als der Median ist. Für eine Stichprobe $x_1, \ldots, x_n$ sortiert man die Werte x_i in aufsteigender Reihenfolge. Ist n ungerade, so definiert man dann den Median der Stichprobe als den Wert an mittlerer Stelle der Reihenfolge. Ist n gerade, so nimmt man den Mittelwert der beiden mittleren Werte.

Der Median einer Zufallsvariable kann nicht allzu weit vom Erwartungswert abweichen: Er erfüllt

$$|m(X) - E(X)| \leq \sigma(X).$$

Der Beweis funktioniert nach einem ähnlichen Prinzip wie bei der Tschebyscheff-Ungleichung. Für ein beliebiges $k \geq 0$ beobachtet man zunächst, dass

$$P(|X - E(X)| \geq k) = P(X - E(X) \geq k) + P(X - E(X) \leq -k),$$

also muss einer der Summanden kleiner oder gleich $\frac{1}{2}P(|X - E(X)| \geq k$ sein. Wir nehmen ohne Einschränkung an, dass es der erste Summand. (Falls nicht, betrachten wir $-X$, dann ist es der erste und alles folgende bleibt gleich.) Für $k := m(X) - E(X)$ rechnet man nun

$$\begin{aligned}\frac{1}{2} &= P(X \geq m(X)) = P(X - E(X) \geq k) \leq \frac{1}{2}P(|X - E(X)| \geq k) \leq \frac{1}{2k^2} \sum_{a - E(X) \geq k} (a - E(X))^2 \cdot P(a) \\ &\leq \frac{1}{2k^2} \sum_{a \in X(\Omega)} (a - E(X))^2 \cdot P(a) = \frac{1}{2k^2} E((X - E(X))^2) = \frac{\sigma(X)^2}{2k^2},\end{aligned}$$

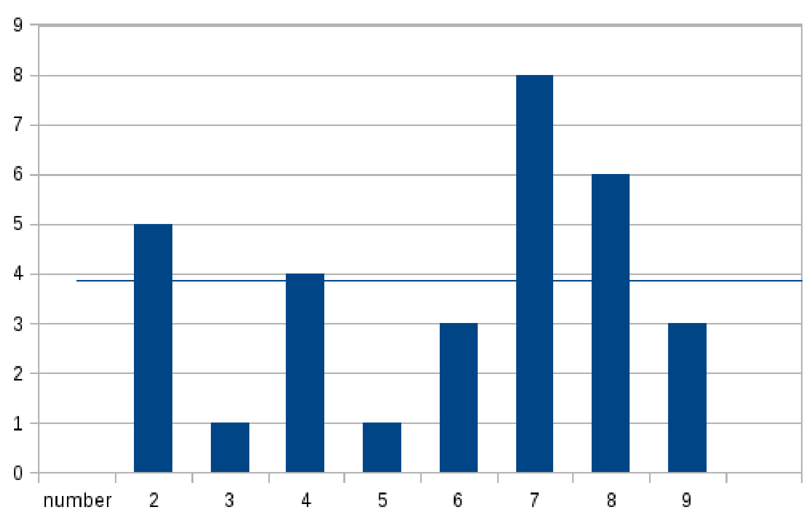
was man umformt zu

$$|m(X) - E(X)| = |k| \leq \sigma(X).$$

Ist die Verteilung von X symmetrisch um $E(X)$ (wie z.B. bei der Normalverteilung oder der t-Verteilung), so ist $m(X) = E(X)$.

4. Aufgaben

Aufgabe 10-1 Die Teilnehmer der Vorlesung wurden gebeten, eine ganze Zahl zwischen 1 und 10 zu nennen. Es gab 31 Antworten. Das folgende Diagramm zeigt die Häufigkeit der einzelnen Zahlen. Nach rechts ist die jeweils genannte Zahl aufgetragen, nach oben die Anzahl der Nennungen der jeweiligen Zahl.



(1) Bestimmen Sie den Erwartungswert, Standardabweichung, Median, Spannweite und Schiefe der Stichprobe. Was ist die Bedeutung des blauen waagrechten Strichs?
(2) Wenn man eine Gleichverteilung der genannten Zahlen unterstellt, würde $\mu = 5{,}5$ folgen. Ab welchem Signifikanzniveau können wir die Hypothese $\mu = 5{,}5$ verwerfen?

ANHANG K

Tests bei partiell korrelierten Stichproben

Gegeben seien zwei Stichproben $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$ so, dass x_i mit y_i in Beziehung steht (z.B. zum gleichen Zeitpunkt gemessen wurde) für $i = 1, \ldots, l$. Die Messwerte x_j, y_j für $j \geq l$ hingegen haben nichts miteinander zu tun. Wir wollen hier beispielhaft den Hypothesentest für $E(X) = E(Y)$ beschreiben. Um das Konfidenzintervall zu berechnen, benötigen wir hier die (wie auch sonst definierten) Größen $\bar{x}, \bar{y}, s_x^2, s_y^2$ sowie $\bar{z} := \bar{x} - \bar{y}$ und

$$c_{x,y}^{(l)} := \frac{1}{n-1} \sum_{i=1}^{l} (x_i - \bar{x})(y_i - \bar{y}).$$

Beachte, dass die Anzahl der Summanden auf der rechten Seite die einzige Ingredienz ist, die l verwendet! Der Vorfaktor $\frac{1}{n-1}$ ist korrekt. Nun definiert man

$$s_z^{(l)} := \sqrt{s_x^2 + s_y^2 - 2c_{x,y}^{(l)}}.$$

Für $l = 0$ und $l = n$ ist ergibt dies genau die in Kapitel 9 Abschnitt 8 verwendeten Standardabweichungen für vollkommen unkorrelierte und vollkommen korrelierte Daten. Man verwendet $s_z^{(l)}$ anstelle von s_z bzw. $\tilde{s}_z$ zur Bestimmung der jeweiligen Konfidenzintervalle oder -schranken. Das Konfidenzintervall zum Konfidenzniveau γ bei unbekannten Varianzen ist dann zum Beispiel

$$[\bar{z} - \frac{s_z^{(l)}}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}, \bar{z} + \frac{s_z^{(l)}}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}].$$

Diese Formeln können aus ähnlichen Überlegungen wie in Kapitel 9 angestellt berechnet werden.

BEISPIEL K.1. In den Städten D und E wird seit langem die jährliche Sonnenscheindauer gemessen. Folgende Messreihen liegen vor (Angaben in Stunden):

	1986	1990	1993	1996	2000	2001	2002	2003	2008
D	1950	1650	k.A.	2250	k.A.	2050	1850	2650	1950
E	1100	k.A.	1300	1900	1400	2300	k.A.	1600	1500

Wir testen die Hypothese "In D und E scheint die Sonne im Durchschnitt gleich lange" auf dem Signifikanzniveau 5%. Sortieren wir die Jahre so, dass Jahre mit beiden Werten zuerst kommen, so sind die Stichproben

	1986	1996	2001	2003	2008	1990/93	2000/2002
D	1950	2250	2050	2650	1950	1650	1850
E	1100	1900	2300	1600	1500	1300	1400

Die letzten beiden Jahrespaarungen sind willkürlich. Wir hätten auch 1990/2000 und 1993/2002 nehmen können. Die Hauptsache ist, dass wir alle gleichzeitig gemessenen Werte nach vorne stellen. Seien die Werte von D mit $x_1, \ldots, x_7$ notiert und die von E mit $y_1, \ldots, y_7$. Offenbar ist also $n = 7$ und $l = 5$. Es ist

$$\bar{x} = 2050, s_x \approx 321, \bar{y} \approx 1586, s_y \approx 402$$

und damit auch $\bar{z} = 464$. Nun benötigen wir

$$c_{x,y}^{(5)} = \frac{1}{6}\sum_{i=1}^{5}(x_i - 2050)\cdot(y_i - 1586) = 21400.$$

Damit haben wir $s_z^{(5)} = \sqrt{s_x^2 + s_y^2 - 2c_{x,y}^{(5)}} \approx 471$. Das Konfidenzintervall zum Konfidenzniveau 95% ist unter Verwendung von $t_{6,0.975} = 2.45$

$$[\bar{z} - \frac{s_z^{(5)}}{\sqrt{7}}t_{6,0.975}, \bar{z} + \frac{s_z^{(5)}}{\sqrt{7}}t_{6,0.975}] = [28, 900].$$

Dies enthält nicht 0, also können wir die Hypothese "In D und E scheint die Sonne im Durchschnitt gleich lange" zum Signifikanzniveau 5% verwerfen.

Vergleichen wir das mit den beiden in Kapitel 8 zur Verfügung stehenden Methoden:

(1) Unter Verwendung nur der Jahresdaten, die in beiden Städten vorliegen:
Dies sind die Daten der Jahre 1986,1996,2001,2003 und 2008. Für diese bilden wir die Stichprobendifferenz

$$z_i := x_i - y_i$$

der Zufallsgrße $Z := X - Y$ und testen die Hypothese $E(Z) = 0$. Es ist

$$z_1 = 850, z_2 = 350, z_3 = -250, z_4 = 1050, z_5 = 450.$$

Damit errechnet man

$$\bar{z} = 490, s_z \approx 503.$$

Mit diesen Daten verwenden den Hypothesentest mit unbekannter Varianz (wegen Stichprobenumfang $= 5 < 30$) und erhalten mit $t_{4,0.975} = 2.78$ das Konfidenzintervall entsprechend Abschnitt 8.5 der Vorlesung

$$[\bar{z} - \frac{s}{\sqrt{5}}t_{4,0.975}, \bar{z} + \frac{s}{\sqrt{5}}t_{4,0.975}] = [490 - \frac{503}{\sqrt{5}}\cdot 2.78, 490 - \frac{503}{\sqrt{5}}\cdot 2.78] = [-135, 1115].$$

Dies enthält die 0, also können wir die Hypothese "In D und E scheint die Sonne im Durchschnitt gleich lange" auf dem Signifikanzniveau 5% auf Grundlage dieser fünf Daten nicht verwerfen.

(2) Unter Verwendung aller Daten von D und E separat, d.h. unter Vernachlässigung des Wissens um die Gleichzeitigkeit von Messungen in D und E.
Hier vergessen wir die zeitliche Information vollkommen und benutzen wie oben

$$\bar{x} = 2050, s_x \approx 321, \bar{y} \approx 1586, s_y \approx 402.$$

Wir verwenden wieder die Methode mit unbekannter Varianz, aber diesmal für unkorrelierte Daten nach Abschnitt 8.8.2 (wegen $n = m = 7 < 30$) und berechnen daher

$$\bar{z} = \bar{x} - \bar{y} = 464, \tilde{s}_z = \sqrt{s_x^2 + s_y^2} \approx 514.$$

Diesmal benötigen wir $t_{6,0.975} = 2.45$ und errechnen damit das Konfidenzintervall

$$[\bar{z} - \frac{\tilde{s}_z}{\sqrt{7}}t_{6,0.975}, \bar{z} + \frac{\tilde{s}_z}{\sqrt{7}}t_{6,0.975}] = [-12, 940],$$

das heißt, wir können auch mit diesen Daten die Hypothese "In D und E scheint die Sonne im Durchschnitt gleich lange" zum Signifikanzniveau 5% nicht verwerfen.

Die Nutzung aller Messwerte sowie der partiellen Relation der Messwerte ergibt also in der Tat eine Verbesserung des Konfidenzintervalls. △

ANHANG L

Tabellen

Werte der Normalverteilung

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	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	0.7422	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	0.8888	0.8907	0.8925	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	0.9484	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	0.9854	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	0.9989	0.9989	0.9989	0.9990	0.9990

Werte erzeugt von William Knight

Quantile der Chi-Quadrat-Verteilung nach ausgewählten Wahrscheinlichkeiten p und Freiheitsgraden

Wahrscheinlichkeit p

Freiheitsgrade	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,5**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
1	0,00	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19
p →	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,5**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
11	2,60	3,05	3,82	4,57	5,58	10,34	17,28	19,68	21,92	24,73	26,76
12	3,07	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	14,34	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	15,34	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	18,34	27,20	30,14	32,85	36,19	38,58

p →	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995
21	8,03	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,34	14,04	21,34	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	24,34	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,65
28	12,46	13,56	15,31	16,93	18,94	27,34	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67
p →	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,5**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
31	14,46	15,66	17,54	19,28	21,43	30,34	41,42	44,99	48,23	52,19	55,00
32	15,13	16,36	18,29	20,07	22,27	31,34	42,59	46,19	49,48	53,49	56,33
33	15,82	17,07	19,05	20,87	23,11	32,34	43,75	47,40	50,73	54,78	57,65
34	16,50	17,79	19,81	21,66	23,95	33,34	44,90	48,60	51,97	56,06	58,96
35	17,19	18,51	20,57	22,47	24,80	34,34	46,06	49,80	53,20	57,34	60,28
36	17,89	19,23	21,34	23,27	25,64	35,34	47,21	51,00	54,44	58,62	61,58
37	18,59	19,96	22,11	24,08	26,49	36,34	48,36	52,19	55,67	59,89	62,88
38	19,29	20,69	22,88	24,88	27,34	37,34	49,51	53,38	56,90	61,16	64,18
39	20,00	21,43	23,65	25,70	28,20	38,34	50,66	54,57	58,12	62,43	65,48
40	20,71	22,16	24,43	26,51	29,05	39,34	51,80	55,76	59,34	63,69	66,77

Quantile der t-Verteilung nach ausgewählten Wahrscheinlichkeiten p und Freiheitsgraden

	Wahrscheinlichkeit p				
Freiheitsgrade	**0.9**	**0.95**	**0.975**	**0.99**	**0.995**
1	3,078	6,314	12,706	31,821	63,656
2	1,886	2,92	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,44	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,86	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,25
10	1,372	1,812	2,228	2,764	3,169
p →	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,35	1,771	2,16	2,65	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,12	2,583	2,921
17	1,333	1,74	2,11	2,567	2,898
18	1,33	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
p →	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
21	1,323	1,721	2,08	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,5	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,06	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,31	1,697	2,042	2,457	2,75
1000	1,282	1,646	1,962	2,33	2,581