

<https://matse.paddel.xyz/spicker>

Stochastik

Patrick Gustav Blaneck

Letzte Änderung: 2. Februar 2022

Inhaltsverzeichnis

1 Wahrscheinlichkeitsrechnung	2
1.1 Einführung in die Kombinatorik	2
1.2 Grundbegriffe	4
1.3 Wahrscheinlichkeit	5
1.4 Wahrscheinlichkeitsverteilung einer Zufallsvariablen	11
1.5 Spezielle Wahrscheinlichkeitsverteilungen	19
1.6 Mehrdimensionale Zufallsvariablen	26
1.7 Kovarianz und Korrelation	32
1.8 Gesetze der großen Zahlen und Grenzwertsätze	34
2 Beschreibende Statistik	37
2.1 Merkmale und weitere wichtige Begriffe	37
2.2 Darstellung der Beobachtungsergebnisse	38
2.3 Statistische Maßzahlen	46
2.4 Streuungsmaße	47
2.5 Lineare Regression und Korrelation	50
3 Schließende Statistik	54
3.1 Grundbegriffe	54
3.2 Punktschätzungen	54
3.3 Intervallschätzungen	60
3.4 Statistische Testverfahren	71
Index	82
Beispiele	84

1 Wahrscheinlichkeitsrechnung

1.1 Einführung in die Kombinatorik

Bonus: Urnenmodell

Ein *Urnenmodell* ist ein Gedankenexperiment, das in der Wahrscheinlichkeitstheorie und in der Statistik verwendet wird, um verschiedene Zufallsexperimente auf einheitliche und anschauliche Weise zu modellieren.

Dazu wird ein fiktives Gefäß, Urne genannt, mit einer bestimmten Anzahl an Kugeln gefüllt, die anschließend zufällig gezogen werden. Damit ist gemeint, dass bei jedem Zug alle in der Urne befindlichen Kugeln die gleiche Wahrscheinlichkeit haben, ausgewählt zu werden. Dadurch kann die Bestimmung interessierender Wahrscheinlichkeiten auf die Lösung kombinatorischer Abzählprobleme zurückgeführt werden.

Definition: Permutation

Eine Anordnung von n verschiedenen Elementen in einer bestimmten Reihenfolge heißt eine *Permutation* der n Elemente.

Für eine n -elementige Menge gilt für die Anzahl P der Permutationen:

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

Befinden sich unter den n Elementen jeweils n_1, n_2, \dots, n_k gleiche, dann gilt:

$$P(n; n_1, n_2, \dots, n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!} \quad \text{mit } \sum_{i=1}^k n_i = n \quad k \leq n$$

Definition: Kombination

Eine *Kombination* oder ungeordnete Stichprobe ist in der Kombinatorik eine Auswahl von Objekten aus einer gegebenen Grundmenge, die (im Gegensatz zur Permutation) nicht alle Objekte der Grundmenge enthalten muss und bei der (im Gegensatz zur Permutation und Variation) die Reihenfolge unberücksichtigt bleibt.

Darf jedes Objekt nur einmal auftreten, spricht man von einer *Kombination ohne Wiederholung*.

Um alle *Kombinationen k-ter Ordnung ohne Wiederholung* zu erhalten, müssen alle Permutationen der n Kugeln betrachtet werden, wobei ein Vertauschen der Kugeln auf Plätzen mit demselben Merkmal (wird gezogen $\rightarrow k$ bzw. wird nicht gezogen $\rightarrow n - k$) keine neue Kombination ergibt:

$$C(n; k) = P(n; k; n - k) = \frac{n!}{k! \cdot (n - k)!} = \binom{n}{k}$$

Können Objekte mehrfach ausgewählt werden, so spricht man von einer *Kombination mit Wiederholung*.

Um alle *Kombinationen k-ter Ordnung mit Wiederholung* zu erhalten, gilt:^a

$$C_W(n; k) = P(n + k - 1; k; n - 1) = \frac{(n + k - 1)!}{k! \cdot (n - 1)!} = \binom{n + k - 1}{k}$$

^aZum Verständnis empfehle ich dieses Video: COMBINATIONS with REPETITION - DISCRETE MATHEMATICS

Definition: Variation

Eine *Variation* oder geordnete Stichprobe ist eine Auswahl von k Objekten aus einer Menge von n Objekten, wobei die Reihenfolge der Auswahl eine Rolle spielt.

Bei einer *Variation ohne Wiederholung* sollen k Objekte (mit $k \leq n$) auf k verfügbare Plätze platziert werden, wobei jedes Objekt nur höchstens einen Platz einnehmen darf. Es gibt für den ersten Platz n mögliche Objekte, für den zweiten Platz $n - 1$ Objekte usw. bis zum k -ten Platz, für den es noch $n - k + 1$ mögliche Objekte gibt. Insgesamt gilt also:

$$V(n; k) = k! \cdot C(n; k) = \frac{n!}{(n - k)!}$$

Bei einer *Variation mit Wiederholung* werden aus n Objekten k Objekte unter Beachtung der Reihenfolge ausgewählt, wobei Objekte auch mehrfach ausgewählt werden können. Nachdem jedes der n Objekte auf jedem der k Plätze der Auswahl erscheinen kann, gilt demzufolge:

$$V_W(n; k) = n^k$$

Bonus: Kombinatoriktabelle

	ohne Wiederholung	mit Wiederholung	
Kombinationen k -ter Ordnung	$C(n; k) = \binom{n}{k}$	$C_W(n; k) = \binom{n+k-1}{k}$	ungeordnete Stichproben
Variationen k -ter Ordnung	$V(n; k) = \frac{n!}{(n-k)!}$	$V_W(n; k) = n^k$	geordnete Stichproben
	Ziehung ohne Zurücklegen	Ziehung mit Zurücklegen	

1.2 Grundbegriffe

Definition: Zufallsexperiment

Damit ein Experiment ein *Zufallsexperiment* ist, muss es folgende Eigenschaften aufweisen:

- Es gibt einen genau festgelegten Plan zur Durchführung.
- Alle möglichen Ergebnisse des Experiments sind vorab bekannt.
- Das Ergebnis jedes einzelnen Experiments kann nicht vorhergesagt werden (Zufälligkeit).

Ein Zufallsexperiment kann einmalig und unwiederholbar sein oder auch Serien von Durchführungen mit gleichwertigen und von Durchführung zu Durchführung voneinander unabhängigen Versuchen ermöglichen.

Weiter kann ein Zufallsexperiment *einstufig* oder *mehrstufig* sein.

Definition: Ergebnismenge

Die Menge aller möglichen Ergebnisse eines Zufallsexperiments bezeichnen wir als *Ergebnismenge* Ω .

Definition: Ereignis

Interessieren wir uns nicht für alle möglichen Ergebnisse des Zufallsexperiments, sondern nur für bestimmte, z.B. „Würfeln einer geraden Zahl“, so sprechen wir von einem *Ereignis* ω .

Spezialfall: Einelementige Teilmengen von Ω sind *Elementarereignisse*

$$\omega \subseteq \Omega \text{ mit } |\omega| = 1$$

Ein Ereignis ist A ist entweder:

- Das *unmögliche* Ereignis (A enthält kein Element von Ω)
- Ein *Elementarereignis* (A enthält genau ein Element von Ω)
- Eine *Zusammenfassung* mehrerer Elementarereignisse (A enthält mehrere Elemente von Ω)
- Das *sichere* Ereignis (A enthält alle Elemente von Ω bzw. $A = \Omega$)

Definition: Ereignisraum

Die Menge aller Ereignisse, die sich aus der Ergebnismenge bilden lässt, heißt *Ereignisraum*.

Bonus: Zusammengesetzte Ereignisse

Da Ereignisse Teilmengen der Ergebnismenge sind, lassen sie sich auch wie Mengen verknüpfen. Wir erhalten dadurch zusammengesetzte Ereignisse:

1. Vereinigung:

$$A \cup B = \{\omega \mid \omega \in A \vee \omega \in B\}$$

2. Durchschnitt bzw. Schnittmenge:

$$A \cap B = \{\omega \mid \omega \in A \wedge \omega \in B\}$$

3. Gegenereignis bzw. Komplement:

$$\bar{A} = \{\omega \mid \omega \notin A\}$$

4. Differenz:

$$A \setminus B = \{\omega \mid \omega \in A \wedge \omega \notin B\}$$

5. Disjunkte Ereignisse:

$$A \cap B = \emptyset \iff A \text{ und } B \text{ sind disjunkt}$$

Analog zur Mengenalgebra gelten hier natürlich auch die Regeln von de Morgan^a.

^aSiehe: [De-morgansche Gesetze](#)

1.3 Wahrscheinlichkeit

Definition: Laplace-Experiment

Es gibt eine Reihe von Zufallsexperimenten, bei denen keines der Elementarereignisse gegenüber einem anderen bevorzugt ist, d.h. bei ausreichend häufiger Wiederholung des Experimentes tritt jedes Elementarereignis mit nahezu gleicher Häufigkeit auf. Ein derartiges Experiment bezeichnen wir als *Laplace-Experiment*.

Einem Elementarereignis ω_i aus einer Ergebnismenge Ω mit m möglichen Elementarereignissen wird definitionsgemäß die positive Zahl

$$P(\{\omega_i\}) = p(\omega_i) = \frac{1}{m} = \frac{1}{|\Omega|}$$

als *Wahrscheinlichkeit* zugeordnet. Damit gilt für ein Ereignis A direkt:

$$P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}) = p(\omega_i) = \frac{|A|}{m} = \frac{|A|}{|\Omega|}$$

Das gilt allerdings nur, wenn:

- Die Ergebnismenge Ω endlich ist.
- Alle Elementarereignisse gleichwahrscheinlich sind.

Definition: Absolute Häufigkeit

Die *absolute Häufigkeit* ist das Ergebnis einer einfachen Zählung von Objekten oder Ereignissen (besser Elementarereignissen). Sie gibt an, wie viele Elemente mit dem gleichen interessierenden Merkmal gezählt wurden.

Definition: Relative Häufigkeit

Die *relative Häufigkeit* gibt den Anteil der Elemente einer Menge wieder, bei denen eine bestimmte Merkmalsausprägung vorliegt. Sie wird berechnet, indem die absolute Häufigkeit eines Merkmals in einer zugrundeliegenden Menge durch die Anzahl der Objekte in dieser Menge geteilt wird. Die relative Häufigkeit ist also eine Bruchzahl und hat einen Wert zwischen 0 und 1.

Definition: Wahrscheinlichkeitsaxiome

Seien A, B, E Ereignisse. Es gilt allgemein:

- $P(\emptyset) = 0$
- $P(\bar{E}) = 1 - P(E)$
- $0 \leq P(E) \leq 1$
- $A \subseteq B \implies P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$

Definition: Ereignisalgebra

Es sei Ω eine nichtleere Menge. Eine Ereignisalgebra über Ω ist eine nichtleere Menge S von Teilmengen von Ω , für die gilt:

- Für jedes $E \in S$ ist $\bar{E} \in S$.
- Für jede Folge $E_1, E_2, \dots \in S$ ist $\bigcup_{E_i \in S} E_i \in S$.

Folgerungen:

- Für jede Folge $E_1, E_2, \dots \in S$ ist $\bigcap_{E_i \in S} E_i \in S$.
- $\emptyset \in S$ und $\Omega \in S$.
- Für jede nichtleere Menge Ω ist die Potenzmenge $\mathcal{P}(\Omega)$ eine Ereignisalgebra über Ω .

Ist Ω endlich oder abzählbar unendlich, wählt man als Ereignisalgebra stets die Potenzmenge.

Bonus: Kolmogoroff-Axiome

Es sei Ω eine Ergebnismenge und S eine Ereignisalgebra über Ω . Eine Zuordnungsvorschrift $P : S \rightarrow \mathbb{R}$ heißt *Wahrscheinlichkeitsmaß*, wenn gilt:

1. $\forall E \in S : 0 \leq P(E) \leq 1$.
2. $P(\Omega) = 1$
3. Falls E_1, E_2, \dots disjunkte Ereignisse sind, gilt

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i) \quad \text{„}\sigma\text{-Additivität“}$$

Das Tripel (Ω, S, P) mit Ω als Ergebnismenge, S als Ereignisalgebra und P als Wahrscheinlichkeitsmaß bezeichnet man als *Wahrscheinlichkeitsraum*.

Definition: Bedingte Wahrscheinlichkeit

In vielen Anwendungen ist das Eintreten eines Ereignisses A nicht unabhängig davon, ob vorher ein anderes Ereignis B eingetreten ist oder nicht. Man spricht dann von einer *bedingten Wahrscheinlichkeit* und schreibt $P(A | B)$. Es gilt:

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

Generell gilt $P(A | B) \neq P(B | A)$, aber:

$$\begin{aligned} P(A | B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(B \cap A)}{P(B)} = \frac{P(B \cap A)}{P(A)} \cdot \frac{P(A)}{P(B)} \\ \implies P(A | B) &= P(B | A) \cdot \frac{P(A)}{P(B)} \end{aligned}$$

Alle Axiome für eine Wahrscheinlichkeit werden von der bedingten Wahrscheinlichkeit erfüllt. Damit gilt auch:

- $E_1 \subseteq E_2 \implies P(E_1 | B) \leq P(E_2 | B)$
- $P(\bar{A} | B) = 1 - P(A | B)$

Definition: Multiplikationssatz

Löst man die Definition der bedingten Wahrscheinlichkeit auf nach der Wahrscheinlichkeit für das gleichzeitige Eintreten zweier Ereignisse A und B , so ergibt sich der *Multiplikationssatz* der Wahrscheinlichkeitsrechnung:

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

bzw.

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

Beispiel: Bedingte Wahrscheinlichkeit (Wahrscheinlichkeitsbaum)

Aufgrund des Praxisbezugs des Informatik Studiums, wurden Projektarbeiten eingeführt, welche in Teams zu je vier Studenten bearbeitet werden sollen. Die Teams sind für die Arbeitsaufteilung innerhalb der Gruppe selbst verantwortlich. Bei dem hier betrachteten Team 2 „jeder macht das was er am besten kann“ ergab sich folgende Tabelle:

Mitglied	Codeanteil (in %)	Fehler (in %)
1	20	1.25
2	20	2.5
3	20	5
4	40	1.875

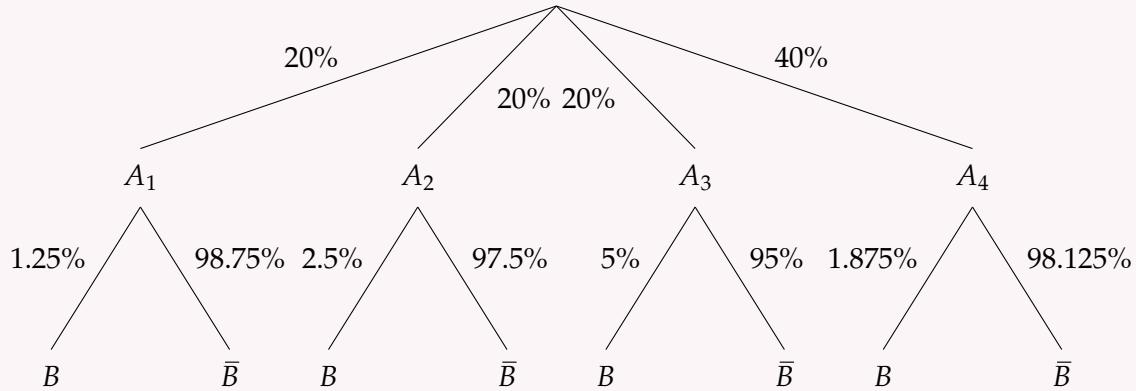
Dem fertigen Code ist nicht mehr anzusehen, von welchem Teammitglied er programmiert worden ist. Aus der Masse an Code wird rein zufällig eine Zeile herausgegriffen und auf Fehler überprüft. Folgende Ereignisse werden formuliert:

$$A_i = \{\text{Der Code wurde von Teammitglied } i \text{ programmiert}\} \quad i = 1, 2, 3, 4$$

$$B = \{\text{Der Code ist fehlerhaft}\}$$

- Formulieren Sie die Fehlerwahrscheinlichkeiten als bedingte Wahrscheinlichkeiten und zeichnen Sie den Wahrscheinlichkeitsbaum.
- Berechnen Sie die Wahrscheinlichkeit, dass die zufällig herausgegriffene Codezeile fehlerhaft ist.
- Wie groß ist die Wahrscheinlichkeit, dass ein fehlerhafter Code von Teammitglied i programmiert worden ist?

a)



Es gilt:

$$P(B | A_1) = 1.25\%$$

$$P(B | A_2) = 2.5\%$$

$$P(B | A_3) = 5\%$$

$$P(B | A_4) = 1.875\%$$

□

Beispiel: Bedingte Wahrscheinlichkeit (Berechnung)

b) Es gilt:

$$P(B) = \sum_{i=1}^4 P(A_i \cap B) = 0.2 \cdot 0.0125 + 0.2 \cdot 0.025 + 0.2 \cdot 0.05 + 0.4 \cdot 0.01875 = 0.025$$

□

c) Es gilt:

$$\begin{aligned} P(A_i | B) &= \frac{P(A_i \cap B)}{P(B)} \\ \implies P(A_1 | B) &= \frac{P(A_1 \cap B)}{P(B)} = \frac{0.2 \cdot 0.0125}{0.025} = 0.1 \\ \wedge \quad P(A_2 | B) &= \frac{P(A_2 \cap B)}{P(B)} = \frac{0.2 \cdot 0.025}{0.025} = 0.2 \\ \wedge \quad P(A_3 | B) &= \frac{P(A_3 \cap B)}{P(B)} = \frac{0.2 \cdot 0.05}{0.025} = 0.4 \\ \wedge \quad P(A_4 | B) &= \frac{P(A_4 \cap B)}{P(B)} = \frac{0.4 \cdot 0.01875}{0.025} = 0.3 \end{aligned}$$

□

Definition: Stochastische Unabhängigkeit

In einigen Anwendungen ist die Wahrscheinlichkeit für das Eintreten des Ereignisses A unabhängig davon, ob B eingetreten ist oder nicht:

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

Aus dem Multiplikationssatz ergibt sich damit als Definition für die *stochastische Unabhängigkeit* zweier Ereignisse A und B :

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

Bonus: Vierfeldertafel

	B	\bar{B}	Zeilensumme
A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A)$
\bar{A}	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A})$
Spaltensumme	$P(B)$	$P(\bar{B})$	1

Bei vollständiger Unabhängigkeit der Ereignisse A und B voneinander gilt:

	B	\bar{B}	Zeilensumme
A	$P(A) \cdot P(B)$	$P(A) \cdot P(\bar{B})$	$P(A)$
\bar{A}	$P(\bar{A}) \cdot P(B)$	$P(\bar{A}) \cdot P(\bar{B})$	$P(\bar{A})$
Spaltensumme	$P(B)$	$P(\bar{B})$	1

Bonus: Zusammenhangskoeffizient

Der *Zusammenhangskoeffizient* bzw. das *Assoziationsmaß* $Q \in [-1; 1]$ sei definiert als:

$$Q = \frac{P(A \cap B) \cdot P(\bar{A} \cap \bar{B}) - P(A \cap \bar{B}) \cdot P(\bar{A} \cap B)}{P(A \cap B) \cdot P(\bar{A} \cap \bar{B}) + P(A \cap \bar{B}) \cdot P(\bar{A} \cap B)}$$

Definition: Mehrstufiges Zufallsexperiment

Kompliziertere Zufallsprozesse bestehen häufig aus mehreren nacheinander ablaufenden Zufallsexperimenten bzw. einem *mehrstufigen Zufallsexperiment*.

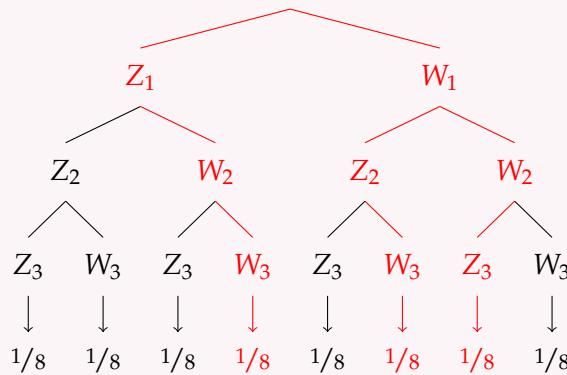
Ein wichtiges Hilfsmittel dabei sind *Ereignisbäume*.

Beispiel: Ereignisbaum

Ein Zufallsexperiment bestehe aus dem gleichzeitigen Werfen dreier *unterscheidbarer* Münzen.

Wie groß ist die Wahrscheinlichkeit dafür, dass einmal „Zahl“ und zweimal „Wappen“ erscheint?

Sei $A = \{\text{Es erscheint einmal „Zahl“ und zweimal „Wappen“}\}$.



Damit gilt:

$$P(A) = \underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}}_{P(Z_1 \cap W_2 \cap W_3)} + \underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}}_{P(W_1 \cap Z_2 \cap W_3)} + \underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}}_{P(W_1 \cap W_2 \cap Z_3)} = \frac{3}{8}$$

□

Definition: Totale Wahrscheinlichkeit

Sind nur bedingte Wahrscheinlichkeiten und die Wahrscheinlichkeiten des bedingenden Ereignisses bekannt, ergibt sich die totale Wahrscheinlichkeit von B aus:

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

Definition: Bayes'sche Formel

Für den Zusammenhang zwischen $P(A | B)$ und $P(B | A)$ ergibt sich direkt aus der Definition und dem Multiplikationssatz die *Bayes'sche Formel* bzw. der *Satz von Bayes*:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \cap A)}{P(B)} = \frac{P(B | A) \cdot P(A)}{P(B)} = \frac{P(B | A) \cdot P(A)}{\sum_{i=1}^n P(A_i) \cdot P(B | A_i)}$$

Die Wahrscheinlichkeit, dass B über einen bestimmten Pfad eintritt, ergibt sich als Verhältnis der Wahrscheinlichkeit für diesen Pfad zur totalen Wahrscheinlichkeit von B .

1.4 Wahrscheinlichkeitsverteilung einer Zufallsvariablen

Definition: Zufallsvariable

Eine *Zufallsvariable* X ordnet jedem Elementarereignis aus der Ergebnismenge eindeutig eine reelle Zahl zu, bildet also die Ergebnismenge auf die Menge der reellen Zahlen ab:

$$X : \Omega \rightarrow \mathbb{R}$$

X ist somit eine reellwertige Funktion mit

- Definitionsbereich: $D = \Omega$
- Wertebereich: $W_X = \{X | X = X(\omega) \wedge \omega \in \Omega\}$

X heißt *diskret*, wenn W_X endlich viele oder abzählbar unendlich viele reelle Werte enthält.

X heißt *stetig*, wenn W_X überabzählbar unendlich viele reelle Werte enthält.

Definition: Verteilungsfunktion

X sei eine Zufallsvariable. Die Funktion ($F_X : \mathbb{R} \rightarrow [0;1]$)

$$\forall x \in \mathbb{R} : F_X(x) := P(\{\omega \in \Omega | X(\omega) \leq x\}) := P(X \leq x)$$

heißt *Verteilungsfunktion* der Zufallsvariablen X .

Anschaulich: $F_X(x)$ (oft nur $F(x)$) ist die Wahrscheinlichkeit dafür, dass die Zufallsvariable X einen Wert $\leq x$ annimmt (höchstens x).

Verteilungsfunktionen besitzen folgende Eigenschaften:

1. $F(x)$ ist eine monoton wachsende Funktion mit $0 \leq F(x) \leq 1$.
2. $\lim_{x \rightarrow -\infty} F(x) = 0$ (unmögliches Ereignis)
3. $\lim_{x \rightarrow \infty} F(x) = 1$ (sicheres Ereignis)
4. $P(a < X \leq b) = F(b) - F(a)$

Beispiel: Verteilungsfunktion

Gegeben seien die folgenden, jeweils auf \mathbb{R} definierten Funktionen:

$$\text{a) } F_1(x) = \begin{cases} 0 & \text{für } x < 2 \\ x - 2 & \text{für } 2 \leq x < 4 \\ 1 & \text{für } x \geq 4 \end{cases}$$

$$\text{b) } F_2(x) = \begin{cases} 0 & \text{für } x < 0 \\ e^{-x} & \text{für } x \geq 0 \end{cases}$$

$$\text{c) } F_3(x) = e^{-e^{-x}} \text{ für } x \in \mathbb{R}$$

a) Für $x_1 = 7/2$ und $x_2 = 2$ gilt:

$$F_1(x_1) = \frac{3}{2} > 1 = F_2(x_2) \quad \wedge \quad x_1 < x_2 \quad \downarrow$$

Also ist $F_1(x)$ nicht monoton steigend und damit insgesamt keine Verteilungsfunktion. \square

b) Offensichtlich ist $F_2(x)$ (streng) monoton fallend für $x \geq 0$ und damit insgesamt keine Verteilungsfunktion. \square

c) Offensichtlich ist $F_3(x)$ monoton steigend und rechtsseitig stetig.

Es gilt:

$$\lim_{x \rightarrow \infty} F_3(x) = \lim_{x \rightarrow \infty} e^{-e^{-x}} = e^{-\lim_{x \rightarrow \infty} e^{-x}} = e^0 = 1$$

$$\lim_{x \rightarrow -\infty} F_3(x) = \lim_{x \rightarrow -\infty} e^{-e^{-x}} = e^{-\lim_{x \rightarrow -\infty} e^{-x}} = e^{-\infty} = 0$$

Also ist $F_3(x)$ damit insgesamt eine Verteilungsfunktion. \square

Definition: Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen

Bei einer diskreten Zufallsvariablen X gehört zu jedem Wert x_i , den sie annehmen kann, eine bestimmte Wahrscheinlichkeit:

$$P(X = x_i) = p_i$$

Damit ist die *Wahrscheinlichkeitsfunktion* einer diskreten Verteilung gegeben mit

$$f(x) = \begin{cases} p_i & \text{für } x = x_i \\ 0 & \text{sonst} \end{cases}$$

Die zugehörige Verteilungsfunktion ist dann

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

Anschaulich: $f(x)$ ist die Wahrscheinlichkeit dafür, dass die Zufallsvariable X den Wert x annimmt (genau x).

Die Wahrscheinlichkeitsfunktion $f(x)$ und die Verteilungsfunktion $F(x)$ besitzen folgende Eigenschaften:

1. $f(x_i) \geq 0$
2. $f(x)$ ist normiert, d.h. es gilt:

$$\sum_{i=1}^{\infty} f(x_i) = 1$$

3. $F(x)$ ist eine monoton wachsende Funktion mit $0 \leq F(x) \leq 1$.
4. $F(a < X \leq b) = F(b) - F(a)$

Definition: Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen

Die Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen X lässt sich durch die *Dichtefunktion* $f(x)$ oder durch die zugehörige Verteilungsfunktion

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

vollständig beschreiben.

Die Dichtefunktion $f(x)$ und die Verteilungsfunktion $F(x)$ besitzen folgende Eigenschaften:

1. $f(x_i) \geq 0$
2. $f(x)$ ist normiert, d.h. es gilt:

$$\int_{-\infty}^{\infty} f(x) \, dx = 1$$

3. $F(x)$ ist eine monoton wachsende Stammfunktion der Dichtefunktion $f(x)$, d.h. es gilt

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

4. $F(a \leq X \leq b) = \int_a^b f(x) \, dx = F(b) - F(a)$

Beispiel: Dichtefunktion

Handelt es sich bei den folgenden Funktionen um Dichtefunktionen? Begründen Sie Ihre Antwort.

a) $f_1(x) = \begin{cases} \sin(x) & \text{für } -\pi/2 \leq x \leq \pi/2 \\ 0 & \text{sonst} \end{cases}$

b) $f_2(x) = \begin{cases} e^{-x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$

a) Offensichtlich ist $f_1(x) = \sin(x) < 0$ für alle $-\frac{\pi}{2} < x < 0$.

Damit kann $f_1(x)$ keine Dichtefunktion sein. \square

b) Es muss gelten:

- f_2 ist nichtnegativ \checkmark (offensichtlich)
- f_2 ist integrierbar \checkmark (offensichtlich)
- f_2 ist normiert:

$$\int_{-\infty}^{\infty} f_2(t) dt = \int_0^{\infty} e^{-t} dt = [-e^{-t}]_0^{\infty} = 0 + 1 = 1 \quad \checkmark$$

Damit ist $f_2(x)$ eine Dichtefunktion. \square

Definition: Transformierte Zufallsvariable

Wenn eine reelle Zufallsvariable X auf dem Ergebnisraum Ω und eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ gegeben ist, dann ist auch $Y = g(X)$ eine Zufallsvariable auf demselben Ergebnisraum.

$g(X)$ wird auch als *Transformation der Zufallsvariablen X unter g* bezeichnet.

Die Verteilungsfunktion von Y lautet dann

$$F_Y(y) = P(g(X) \leq y)$$

Beispiele:

- Lineare Transformation:

$$g_1(X) = aX + b$$

- Betragsfunktion bzw. Gleichrichter:

$$g_2(X) = |X|$$

- Energiefunktion:

$$g_3(X) = aX^2$$

Algorithmus: Transformation der Verteilungsfunktion

Sei g streng monoton wachsend auf dem Wertebereich von X . Dann gilt für die Verteilungsfunktion:

$$F_Y(y) = P(\{\omega \in \Omega \mid Y(\omega) \leq y\}) = P(\{\omega \in \Omega \mid X(\omega) \leq g^{-1}(y)\}) = F_X(g^{-1}(y))$$

Für eine lineare Transformation gilt damit:

$$Y = g(X) = aX + b \implies F_Y(y) = F_X(g^{-1}(y)) = F_X\left(\frac{y-b}{a}\right)$$

Algorithmus: Transformation der Dichtefunktion

Falls $F_Y(y)$ geschlossen darstellbar ist, gilt für die Dichtefunktion:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = f_X(x) \cdot \frac{1}{g'(x)} = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}$$

Für eine lineare Transformation gilt damit:

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right) \implies f_Y(y) = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{|a|}$$

Beispiel: Transformation der Verteilungsfunktion und Dichtefunktion

Bestimmen Sie die Verteilungsfunktion $F_Y(y)$ und die Dichtefunktion $f_Y(y)$ für die transformierte Zufallsvariable Y , die sich als $Y = g(X)$ aus der ursprünglichen Zufallsvariablen X mit bekannter Verteilungsfunktion $F_X(x)$ und bekannter Dichtefunktion $f_X(x)$ ergibt:

$$g(X) = 3X - 1 \quad \text{mit } f_X(x) = \begin{cases} 4/27(3x^2 - x^3) & \text{für } 0 \leq x \leq 3 \\ 0 & \text{sonst} \end{cases}$$

Es gilt:

$$f_Y(y) = f_X\left(\frac{y+1}{3}\right) \cdot \frac{1}{|3|} = \begin{cases} \frac{1}{3} \cdot \frac{4}{27} \left(3\left(\frac{y+1}{3}\right)^2 - \left(\frac{y+1}{3}\right)^3\right) & \text{für } -1 \leq y \leq 8 \\ 0 & \text{sonst} \end{cases}$$

Und damit:

$$F_Y(y) = F_X\left(\frac{y+1}{3}\right) = \begin{cases} 0 & \text{für } y < -1 \\ \frac{4}{27} \left(\frac{y+1}{3}\right)^3 - \frac{1}{27} \left(\frac{y+1}{3}\right)^4 & \text{für } -1 \leq y \leq 8 \\ 1 & \text{für } y > 8 \end{cases}$$

□

^a $0 \leq x \leq 3 \wedge y = g(x) = 3x - 1 \implies -1 \leq y \leq 8$

Definition: Erwartungswert

Der wichtigste Lageparameter einer Verteilung ist der *Erwartungswert*. Der Erwartungswert einer Zufallsvariablen beschreibt die Zahl, die die Zufallsvariable im Mittel annimmt.

Für eine diskrete Zufallsvariable X gilt:

$$E(X) := \sum_i x_i \cdot P(X = x_i) \quad \text{bzw.} \quad E(X) := \sum_i X(\omega_i) \cdot P(\{\omega_i\}), \omega_i \in \Omega$$

Für eine stetige Zufallsvariable X gilt: (Schwerpunkt der Dichtefunktion)

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

Sehr oft wird statt $E(X)$ auch μ oder μ_X geschrieben.

Algorithmus: Transformation des Erwartungswertes auf eine neue Zufallsvariable

Bei einer Transformation von einer Zufallsvariable X auf eine neue Zufallsvariable $Y = g(X)$ gilt:

$$E(Y) = \int_{-\infty}^{\infty} y \cdot f_Y(y) \, dy = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \, dx$$

Speziell bei der linearen Transformation gilt:

$$E(aX + b) = aE(X) + b$$

Definition: Median

Der *Median* \tilde{x} ist der Wert, der die Verteilung „halbiert“.

Bei stetigen Zufallsvariablen existiert der Median immer und es gilt:

$$P(X \leq \tilde{x}) = F_X(\tilde{x}) = \int_{-\infty}^{\tilde{x}} f_X(x) \, dx = \frac{1}{2}$$

Bei diskreten Zufallsvariablen gibt es nur wenige Ausnahmen, bei denen ein x -Wert existiert mit genau $F(x) = \frac{1}{2}$. Daher gilt folgende Definition:

$$P(X < \tilde{x}) \leq \frac{1}{2} \quad \wedge \quad P(X \leq \tilde{x}) \geq \frac{1}{2}$$

Definition: Quantile

Ein *Quantil* ist ein Lagemaß. Ein bestimmter Anteil der Werte (einer Zufallsstichprobe) ist kleiner als das Quantil, der Rest ist größer. Das 25%-Quantil, bzw. das untere Quartil, beispielsweise ist der Wert, für den gilt, dass 25% aller Werte \leq sind als dieser Wert.

Bei stetigen Zufallsvariablen gilt:

$$P(X \leq x_\alpha) = F_X(x_\alpha) = \int_{-\infty}^{x_\alpha} f_X(x) \, dx = \alpha$$

Bei diskreten Zufallsvariablen gilt:

$$P(X < x_\alpha) \leq \alpha \quad \wedge \quad P(X \leq x_\alpha) \geq \alpha$$

Relevant sind insbesondere das *untere Quartil* ($\alpha = 1/4$), das *obere Quartil* ($\alpha = 3/4$) und der Median ($\alpha = 1/2$).

Der *Quartilabstand* $x_{3/4} - x_{1/4}$ ist ein Streuungsmaß, denn im Intervall zwischen unterem und oberem Quartil liegen die „inneren“ 50% der Verteilung.

Definition: Varianz

Die *Varianz* ist ein wichtiges Maß für die Streuung der Wahrscheinlichkeitsdichte um ihren Schwerpunkt. Mathematisch wird sie definiert als die mittlere quadratische Abweichung einer reellen Zufallsvariablen von ihrem Erwartungswert.

Bei diskreten Zufallsvariablen gilt:

$$\sigma^2 = \text{Var}(X) = \sum_i (x_i - \mu)^2 \cdot P(X = x_i)$$

Bei stetigen Zufallsvariablen gilt:

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) \, dx$$

Algorithmus: Transformation der Varianz auf eine neue Zufallsvariable

Bei einer Transformation von einer Zufallsvariable X auf eine neue Zufallsvariable $Y = g(X)$ gilt speziell bei der linearen Transformation:

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Beispiel: Transformation des Erwartungswertes und der Varianz

Gegeben ist ein Spannungssignal X mit Gaußscher Dichtefunktion $f_X(x)$, Erwartungswert $\mu_X = 1\text{V}$ und Varianz $\sigma_X^2 = 0.25\text{V}^2$. Das Signal wird durch die Funktion $Y = g(X) = 2X + 1.5\text{V}$ in ein Ausgangssignal Y transformiert.

Bestimmen Sie den Erwartungswert μ_Y sowie die Varianz σ_Y^2 des Ausgangssignals.

Es handelt sich hier im eine *lineare* Transformation.

Damit gilt:

$$E(Y) = E(2X + 1.5\text{V}) = 2E(X) + 1.5\text{V} = 2\mu_X + 1.5\text{V} = 2 \cdot 1\text{V} + 1.5\text{V} = 3.5\text{V}$$

$$\text{Var}(Y) = \text{Var}(2X + 1.5\text{V}) = 2^2 \text{Var}(X) = 4\sigma_X^2 = 4 \cdot 0.25\text{V}^2 = 1\text{V}^2$$

□

Definition: Standardabweichung

Die *Standardabweichung* ist ein weiteres Streuungsmaß und wird definiert als die Quadratwurzel der Varianz:

$$\sigma = \sqrt{\text{Var}(X)}$$

Algorithmus: Verschiebungssatz

Sei der Erwartungswert einer Zufallsvariable X gegeben mit $E(X) = \mu$. Dann gilt mit dem *Verschiebungssatz* für ein beliebiges $a \in \mathbb{R}$:

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

Für $a = 0$ erhalten wir eine alternative Formel zur Berechnung der Varianz:

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

Definition: Tschebyscheffsche Ungleichung

Mithilfe der *Tschebyscheffschen Ungleichung* lässt sich unter Verwendung der existierenden ersten beiden Momente (Erwartungswert und Varianz) die Wahrscheinlichkeit dafür abschätzen, dass die Zufallsvariable X Werte in bestimmten Intervallen der reellen Zahlengeraden annimmt, ohne jedoch die Verteilung von X zu kennen.

Sie lautet für eine Zufallsvariable X mit Erwartungswert μ und Varianz σ^2 :

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

Sie setzt keine besondere Verteilungsform voraus.

Ein Nachteil der Tschebyscheffschen Ungleichung ist, dass sie nur eine grobe Abschätzung liefert.

1.5 Spezielle Wahrscheinlichkeitsverteilungen

Definition: Gleichverteilung

Eine *Gleichverteilung* beschreibt eine Wahrscheinlichkeitsverteilung mit bestimmten Eigenschaften.

Im diskreten Fall tritt jedes mögliche Ergebnis mit der gleichen Wahrscheinlichkeit ein, im stetigen Fall ist die Dichte konstant.

Für eine im Intervall $[a, b]$ stetige gleichverteilte Zufallsvariable X gilt also:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in [a; b] \\ 0 & \text{sonst} \end{cases}$$

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

Der Grundgedanke einer Gleichverteilung ist, dass es keine Präferenz gibt.

Die Maßzahlen dieser Verteilung lauten:

- Erwartungswert: $\mu = \frac{a+b}{2}$
- Varianz: $\sigma^2 = \frac{(b-a)^2}{12}$
- Standardabweichung: $\sigma = \sqrt{\sigma^2} = \frac{b-a}{\sqrt{12}}$

Definition: Bernoulli-Experiment

Bei einem *Bernoulli-Experiment* gibt es stets nur zwei Ausgänge, Treffer oder Niete. Zudem muss die Wahrscheinlichkeit für einen Treffer, p , und somit auch die für eine Niete, $1 - p$, bei jedem der Experimente dieselbe sein.

Bei einem *Bernoulli-Experiment vom Umfang n* wird ein Bernoulli-Experiment n -fach ausgeführt mit der Voraussetzung, dass die Ergebnisse der einzelnen Studien voneinander unabhängig sind.

Definition: Binomialverteilung

Ein Bernoulli-Experiment mit den beiden sich gegenseitig ausschließenden Ergebnissen A und \bar{A} werde n -mal nacheinander ausgeführt (Bernoulli-Experiment mit Umfang n).

Dann genügt die diskrete Zufallsvariable

$X = \text{Anzahl der Versuche, in denen das Ereignis } A \text{ eintritt}$

der *Binomialverteilung* mit der Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} \quad x \in [0, n]$$

und der zugehörigen Verteilungsfunktion

$$F(x) = P(X \leq x) = \sum_{k \leq x} \binom{n}{k} \cdot p^k \cdot q^{n-k} \quad x \geq 0$$

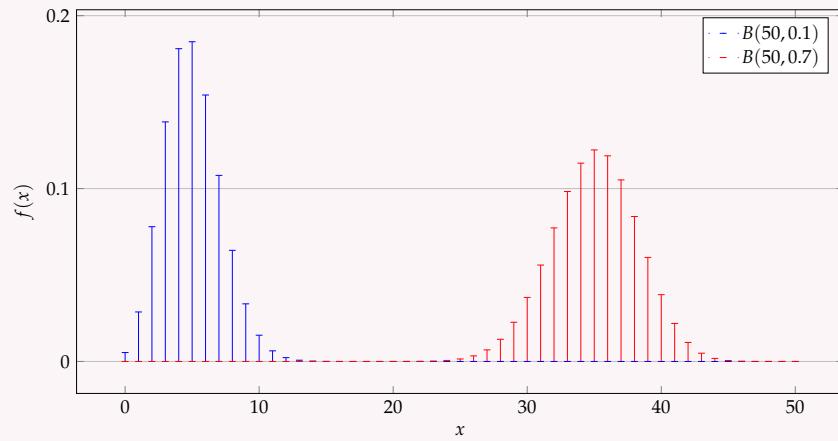
(für $x < 0$ ist $F(x) = 0$) n und p sind dabei die Parameter der Binomialverteilung.

Die Maßzahlen dieser Verteilung lauten:

- Erwartungswert: $\mu = np$
- Varianz: $\sigma^2 = npq = np(1-p)$
- Standardabweichung: $\sigma = \sqrt{npq} = \sqrt{np(1-p)}$

Man schreibt auch $X \sim B(n, p)$.

Beispiel: Binomialverteilung



Definition: Hypergeometrische Verteilung

Die *hypergeometrische Verteilung* beschreibt die Wahrscheinlichkeit dafür, dass bei N gegebenen Elementen (Grundgesamtheit des Umfangs N), von denen M die gewünschte Eigenschaft besitzen, beim Herausgreifen von n Probestücken (Stichprobe des Umfangs n) genau x Treffer erzielt werden, d. h. die Wahrscheinlichkeit für $X = x$ Erfolge in n Versuchen.

Dabei ist die Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad x \in [0, n]$$

und die zugehörige Verteilungsfunktion

$$F(x) = P(X \leq x) = \sum_{k \leq x} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad x \geq 0$$

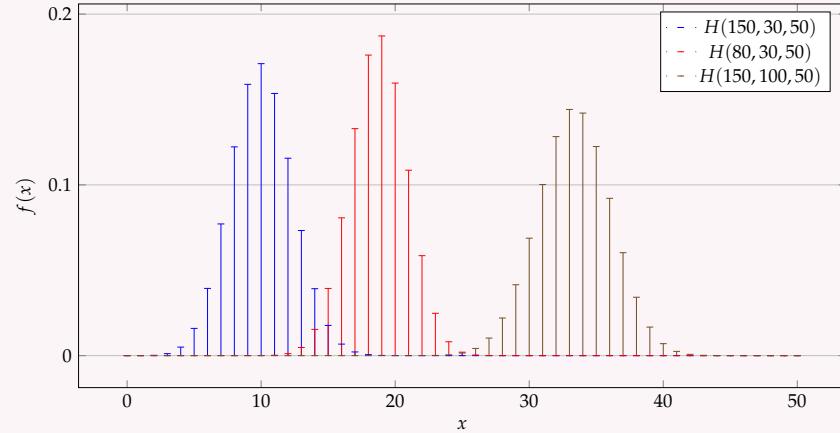
(für $x < 0$ ist $F(x) = 0$) N , M und n sind dabei die Parameter der hypergeometrischen Verteilung.

Die Maßzahlen dieser Verteilung lauten:

- Erwartungswert: $\mu = n \cdot \frac{M}{N}$
- Varianz: $\sigma^2 = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1} = \frac{nM(N-M)(N-n)}{N^2(N-1)}$

Man schreibt auch $X \sim H(N, M, n)$.

Beispiel: Hypergeometrische Verteilung



Definition: Poisson-Verteilung

Die *Poisson-Verteilung* ist eine Wahrscheinlichkeitsverteilung, mit der die Anzahl von Ereignissen modelliert werden kann, die bei konstanter mittlerer Rate unabhängig voneinander in einem festen Zeitintervall oder räumlichen Gebiet eintreten.

Die Wahrscheinlichkeitsfunktion einer diskreten poisson-verteilten Zufallsvariablen X ist gegeben mit

$$f(x) = P(X = x) = \frac{\mu^x}{x!} \cdot e^{-\mu} \quad x \in [0, \infty]$$

und die zugehörige Verteilungsfunktion mit

$$F(x) = P(X \leq x) = e^{-\mu} \cdot \sum_{k \leq x} \frac{\mu^k}{k!} \quad x \geq 0$$

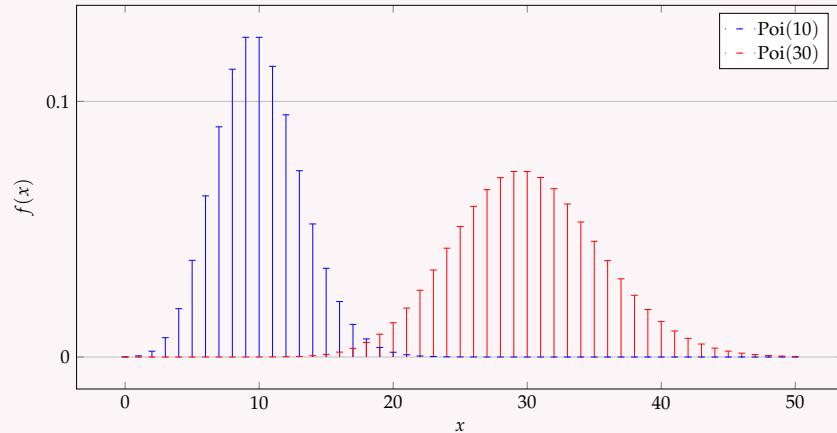
(für $x < 0$ ist $F(x) = 0$) μ ist dabei der Parameter der Poisson-Verteilung.

Die Maßzahlen dieser Verteilung lauten:

- Erwartungswert: μ
- Varianz: $\sigma^2 = \mu$
- Standardabweichung: $\sigma = \sqrt{\mu}$

Man schreibt auch $X \sim \text{Poi}(\mu)$.

Beispiel: Poisson-Verteilung



Definition: Lokaler Grenzwertsatz von de Moivre-Laplace

Nach dem *lokalen Grenzwertsatz von de Moivre-Laplace* konvergiert die Binomialverteilung für den Stichprobenumfang $n \rightarrow \infty$ und Wahrscheinlichkeiten $0 < p < 1$ gegen die Normalverteilung. Bei großem Stichprobenumfang kann daher die Normalverteilung $\mathcal{N}(0, \sigma^2)$ (mit $\sigma^2 = p(1-p)$) als Näherung der Binomialverteilung verwendet werden.

Also gilt:

$$B(k \mid p, n) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi n \sigma^2}} \cdot \exp \left(-\frac{n}{2\sigma^2} \left(\frac{k}{n} - p \right)^2 \right)$$

Definition: Normalverteilung

Die *Normalverteilung* ist eine der wichtigsten stetigen Wahrscheinlichkeitsverteilungen.

Eine stetige Zufallsvariable X hat eine Normalverteilung mit Erwartungswert $\mu \neq \pm\infty$ und Varianz $\sigma^2 > 0$, oft geschrieben als $X \sim \mathcal{N}(\mu, \sigma^2)$, wenn X die folgende Wahrscheinlichkeitsdichte hat:

$$f(x) = P(X = x) = \frac{1}{\sqrt{2\pi n\sigma^2}} \cdot \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad x \in \mathbb{R}$$

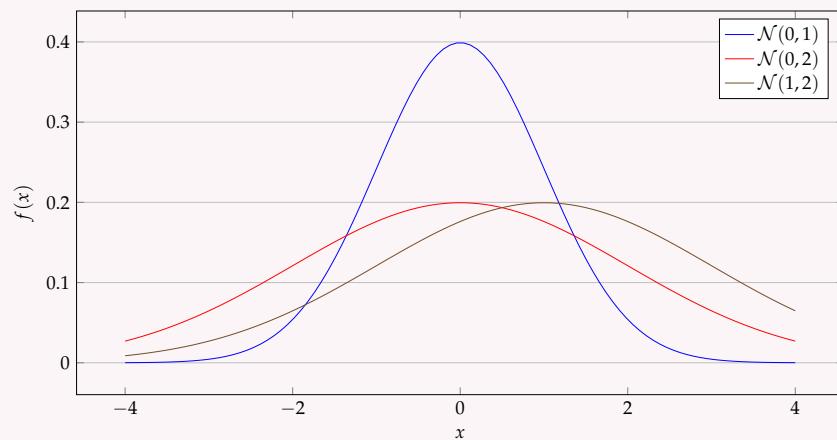
Die zugehörige Verteilungsfunktion ist dann gegeben mit

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Wichtige Eigenschaften der Normalverteilung sind:

- $f(x)$ ist symmetrisch um $x = \mu$
- $f(x)$ hat ein Maximum der Höhe $1/\sigma\sqrt{2\pi}$ bei $x = \mu$
- $f(x)$ hat Wendepunkte bei $x = \mu \pm \sigma$
- $F(x)$ ist nicht geschlossen lösbar!

Beispiel: Normalverteilung



Definition: Standardnormalverteilung

Die *Standardnormalverteilung* ist eine spezielle Variante der Normalverteilung. Dabei gilt $\mu = 0$, $\sigma^2 = 1$, bzw. $X \sim \mathcal{N}(0, 1)$.

Wichtige Eigenschaften der Standardnormalverteilung sind:

- Jede Normalverteilung lässt sich auf die Standardnormalverteilung transformieren.
- Die Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung ist tabelliert.

Algorithmus: Transformation auf standardisierte Variable

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt:

$$P(X \leq x) = F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \Phi(u)$$

wobei $u = \frac{x - \mu}{\sigma}$ eine standardisierte Variable ist.

Dabei gilt:

- $P(X \geq x) = 1 - \Phi(u)$
- $P(a \leq X \leq b) = \Phi\left(\frac{a - \mu}{\sigma}\right) - \Phi\left(\frac{b - \mu}{\sigma}\right)$
- $\Phi(-u) = 1 - \Phi(u)$

Definition: Geometrische Verteilung

Die *geometrische Verteilung* beschreibt das „Warten auf den ersten Erfolg“.

Dabei besitzt eine diskrete Zufallsvariable X eine geometrische Verteilung mit Parameter $0 < p < 1$, falls ihre Verteilung gegeben ist durch

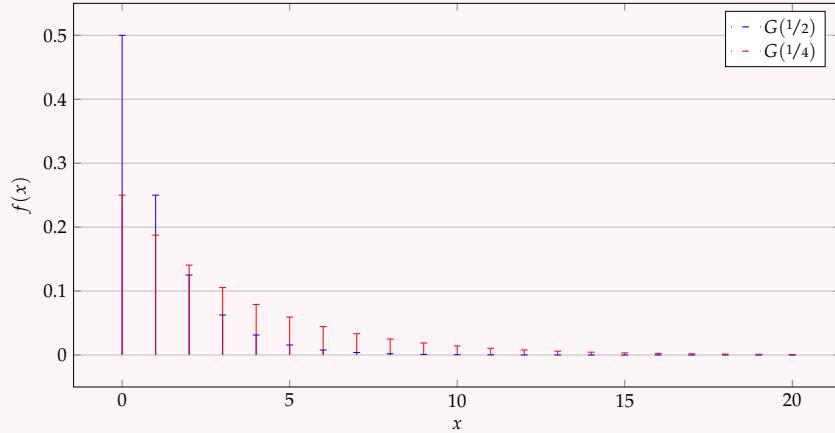
$$f(x) = P(X = x) = (1 - p)^x p \quad x \in \mathbb{N}_0$$

X beschreibt dann die Anzahl der „Nieten“ vor dem ersten „Treffer“, der mit der Wahrscheinlichkeit p eintritt.

Die Maßzahlen dieser Verteilung lauten:

- Erwartungswert: $\mu = \frac{1-p}{p}$
- Varianz: $\sigma^2 = \frac{1-p}{p^2}$

Beispiel: Geometrische Verteilung



Algorithmus: Approximation von Verteilungen

	Approximation durch eine ...		
	Binomialverteilung	Poisson-Verteilung	Normalverteilung
Binomial-verteilung		Bedingungen: $np \leq 10$ $n \geq 1500p$ Dann gilt: $\approx \text{Poi}(np)$	Bedingungen: $np(1-p) > 9$ Dann gilt: $\approx \mathcal{N}(np, np(1-p))$
Hypergeo-metrische Verteilung	Bedingungen: $0.1 < \frac{M}{N} < 0.9$ $n < 0.05N, n > 10$ Dann gilt: $\approx B(n, \frac{M}{N})$	Bedingungen: $0.1 < \frac{M}{N} < 0.9$ $n < 0.05N, n > 30$ Dann gilt: $\approx \text{Poi}(n \frac{M}{N})$	Bedingungen: $0.1 < \frac{M}{N} < 0.9$ $n < 0.05N, n > 30$ Dann gilt: $\approx \mathcal{N}(n \frac{M}{N}, n \frac{M}{N} (1 - \frac{M}{N}) \frac{N-n}{N-1})$
Poisson-Verteilung			Bedingungen: $\mu > 9$ Dann gilt: $\approx \mathcal{N}(\mu, \sqrt{\mu})$

Achtung:

Diskrete Verteilungen nehmen nur einzelne Werte an, stetige Verteilungen haben positive Wahrscheinlichkeit nur für Intervalle - Wenn wir also die diskrete Verteilung durch eine Stetige approximieren wollen, müssen wir aus dem einzelnen Wert a ein Intervall machen.

Weil diskrete Verteilungen oft Werte aus den ganzen Zahlen annehmen, teilt man die Intervalle zwischen den einzelnen Zahlen einfach gleich auf.

So wird aus a das Intervall $[a - 0.5, a + 0.5]$ und wir betrachten bei der Approximation

$$P(a - 0.5 \leq X \leq a + 0.5) \text{ statt } P(a)$$

Bonus: Zusammenfassung von Verteilungsfunktionen

Verteilung	Parameter	Wahrscheinlichkeits- bzw. Dichtefunktion	Art	Kennzahlen
Binomial-verteilung $B(n, p)$	$n \in \mathbb{N}$ $0 < p < 1$	$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$	diskret	$\mu = np$ $\sigma^2 = np(1-p)$
Hypergeometrische Verteilung $H(N, M, n)$	$N \in \mathbb{N}$ $M \in [0, N]$ $n \in [0, N]$	$f(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$	diskret	$\mu = n \frac{M}{N}$ $\sigma^2 = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$
Poisson-Verteilung $Poi(\mu)$	$\mu > 0$	$f(x) = \frac{\mu^x}{x!} \cdot e^{-\mu}$	diskret	$\mu = \mu$ $\sigma^2 = \mu$
(Gaußsche) Normal-verteilung $\mathcal{N}(\mu, \sigma^2)$	$\mu \in \mathbb{R}$ $\sigma^2 > 0$	$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}$	stetig	$\mu = \mu$ $\sigma^2 = \sigma^2$
Standard-normal-verteilung $\mathcal{N}(0, 1)$	$\mu = 0$ $\sigma^2 = 1$	$\phi(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} u^2}$	stetig	$\mu = 0$ $\sigma^2 = 1$

1.6 Mehrdimensionale Zufallsvariablen

Definition: Wahrscheinlichkeitsfunktion einer zweidimensionalen Verteilung

Die Wahrscheinlichkeitsfunktion einer zweidimensionalen Verteilung ist eine Funktion von zwei unabhängigen Variablen:

$$f(x, y) = P(X = x, Y = y)$$

Definition: Randverteilung

Die beiden eindimensionalen Wahrscheinlichkeitsfunktionen $f_1(x)$ und $f_2(y)$ der Zufallsvariablen X und Y werden als *Randverteilungen* der zweidimensionalen Verteilung (X, Y) bezeichnet.

Es gilt:

$$f_1(x) = \sum_y f(x, y)$$

$$f_2(y) = \sum_x f(x, y)$$

Beispiel: Zweidimensionale Zufallsvariable

Wir betrachten einen gleichzeitigen Wurf einer Münze und eines Würfels. Dabei seien die Zufallsvariablen X und Y wie folgt:

$$X = \text{„Anzahl Wappen bei der Münze“} \in \{0, 1\}$$

$$Y = \text{„Augenzahl beim Würfeln“} \in \{1, 2, 3, 4, 5, 6\}$$

Dann gibt es für die zweidimensionale Zufallsvariable (X, Y) offensichtlich 12 Elementareignisse:

\backslash	Y	1	2	3	4	5	6
X	0	(0, 1)	(0, 2)	(0, 3)	(0, 4)	(0, 5)	(0, 6)
	1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)

und es gilt die zweidimensionale Wahrscheinlichkeitsfunktion (Verteilungstabelle):

\backslash	Y	1	2	3	4	5	6	$f_1(x)$
X	0	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/2$
	1	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/2$
$f_2(y)$		$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

Definition: Verteilungsfunktion einer zweidimensionalen Verteilung

Die Wahrscheinlichkeitsfunktion einer zweidimensionalen Verteilung (X, Y) lässt sich vollständig durch die Verteilungsfunktion darstellen.

Für die *Verteilungsfunktion*

$$F(x, y) = P(X \leq x, Y \leq y)$$

gelten folgende Eigenschaften:

- $\lim_{x \rightarrow \infty, y \rightarrow \infty} F(x, y) = 1$
- $\lim_{x \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} F(x, y) = 0$
- $P(a < x \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$

Definition: Diskrete zweidimensionale Verteilung

(X, Y) heißt *diskret*, wenn beide Komponenten diskrete Zufallsvariablen sind.

Unter der Annahme, dass X und Y nur endlich viele Werte annehmen können, sei

$$P(X = x_i, Y = y_k) = p_{ik}$$

Für die Wahrscheinlichkeitsfunktion gilt:

$$f(x, y) = \begin{cases} p_{ik} & \text{für } x = x_i, y = y_k \\ 0 & \text{sonst} \end{cases}$$

mit der Normierung:

$$f(x, y) = \sum_{i=1}^m \sum_{j=1}^n f(x_i, y_j) = 1$$

Für die Randverteilungen gilt:

$$f_1(x) = \sum_y f(x, y)$$

$$f_2(y) = \sum_x f(x, y)$$

Für die Verteilungsfunktion gilt:

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_k \leq y} f(x_i, y_k)$$

Definition: Stetige zweidimensionale Verteilung

(X, Y) heißt *stetig*, wenn beide Komponenten stetige Zufallsvariablen sind.

Für die Verteilungsfunktion gilt:

$$F(x, y) = \int_{u=-\infty}^x \int_{v=-\infty}^y f(u, v) \, dv \, du$$

Für die Dichtefunktion gilt die Normierung:

$$f(x, y) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} f(x, y) \, dy \, dx = 1$$

Für die Randverteilungen gilt:

$$f_1(x) = \int_{y=-\infty}^{\infty} f(x, y) \, dy$$

$$f_2(y) = \int_{x=-\infty}^{\infty} f(x, y) \, dx$$

Definition: Stochastisch unabhängige Zufallsvariable

Die Zufallsvariablen X und Y mit den Verteilungsfunktionen $F_1(x)$ und $F_2(y)$ und der gemeinsamen Verteilungsfunktion $F(x, y)$ heißen *stochastisch unabhängig*, wenn die Bedingung

$$F(x, y) = F_1(x) \cdot F_2(y)$$

für alle (x, y) erfüllt ist.

Es gilt:

- Ist die Bedingung nicht erfüllt, sind die Zufallsvariablen stochastisch abhängig.
- Ist die Bedingung erfüllt, sind die beiden Ereignisse $X = x$ und $Y = y$ stochastisch unabhängig.
- Sind X und Y stochastisch unabhängige Zufallsvariablen, so gilt $f(x, y) = f_1(x) \cdot f_2(y)$
- Sinngemäße Erweiterung auf n Zufallsvariablen

Algorithmus: Summen von Zufallsgrößen

Die neue Zufallsvariable wird als Summe von anderen Zufallsvariablen definiert:

$$Z := X_1 + X_2 + \dots + X_n$$

Den Zahlenwerten $x_1 := X_1(\omega), \dots, x_n := X_n(\omega)$ eines Zufallsversuches mit Resultat ω wird die Zahl $z := x_1 + \dots + x_n$ als Ergebnis $Z(\omega)$ der neuen Zufallsgröße Z zugeordnet.

Sei $Z = X + Y$.

Dann gilt im diskreten Fall

$$P(Z = z) = \sum_i P(X = x_i, Y = z - x_i)$$

und im stetigen Fall

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z - x) \, dx$$

Sind X und Y stochastisch unabhängig, dann gilt insbesondere

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) \, dx$$

Das wird auch als „Faltung“ bezeichnet.

Algorithmus: Produkt und Quotient zweier stetiger Zufallsgrößen

Sei $Z = XY$. Dann gilt:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z/y, y) \cdot \frac{1}{|y|} \, dy$$

Sei $Z = X/Y$. Dann gilt:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(zy, y) \cdot |y| \, dy$$

Algorithmus: Erwartungswert und Varianz für die Summe von Zufallsgrößen

X und Y seien als Funktionen auf Ω explizit bekannt:

$$\implies (X + Y)(\omega) = X(\omega) + Y(\omega) \quad \wedge \quad (X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega)$$

Dann gilt für den Erwartungswert:

$$\forall \alpha \in \mathbb{R} : E(\alpha X + b) = \alpha E(X) + b$$

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$$

Und für die Varianz:

$$\forall \alpha \in \mathbb{R} : \text{Var}(\alpha X + b) = \alpha^2 \text{Var}(X)$$

$$\text{Var}(\alpha X + \beta Y) = \alpha^2 X + \beta^2 Y + 2\alpha\beta \text{Cov}(X, Y)$$

Falls X und Y stochastisch unabhängige Zufallsvariablen sind, gilt:

$$E(X \cdot Y) = E(X) \cdot E(Y) \implies \text{Cov}(X, Y) = 0 \implies \text{Var}(\alpha X + \beta Y) = \alpha^2 \text{Var}(X) + \beta^2 \text{Var}(Y)$$

Beispiel: Summen von Zufallsgrößen

Bei einem Produktionsvorgang werden Zylinder in den ausgefrästen Kreis eines Metallsockels eingepasst. Die beiden Teile werden rein zufällig aus den bisher produzierten Zylindern bzw. ausgefrästen Metallplatten ausgewählt. Der Durchmesser des Zylinders ist (in mm) nach $\mathcal{N}(24.9; (0.03)^2)$ -verteilt, der Durchmesser des in den Metallsockel eingefrästen Kreises ist nach $\mathcal{N}(25; (0.04)^2)$ -verteilt. Der Zylinder kann noch eingepasst werden, falls die lichte Weite der Durchmessers (= Durchmesser des gefrästen Kreises - Durchmesser des Zylinders) nicht mehr als 0.2mm beträgt.

a) Berechnen Sie

- i) den Erwartungswert
- ii) die Varianz
- iii) die Verteilung

der Zufallsvariablen „lichte Weite des Durchmessers“.

b) In wie viel Prozent aller Fälle lässt sich der Zylinder nicht in die Metallplatte einpassen?

a) i) Es gilt:

$$X = \{\text{Durchmesser des Zylinders [mm]}\} \sim \mathcal{N}(\mu_X, \sigma_X^2) = N(24.9, (0.03)^2)$$

$$Y = \{\text{Durchmesser des eingefrästen Kreises [mm]}\} \sim \mathcal{N}(\mu_Y, \sigma_Y^2) = N(25, (0.04)^2)$$

und wir definieren:

$$Z = Y - X$$

Damit gilt:

$$E(Z) = E(Y) - E(X) = \mu_Y - \mu_X = 0.1$$

□

ii) Da X und Y stochastisch unabhängig sind, gilt:

$$\text{Var}(Z) = \text{Var}(Y) + \text{Var}(X) - 2 \text{Cov}(X, Y) = \text{Var}(Y) + \text{Var}(X) = (0.05)^2$$

□

iii)

$$Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2) = \mathcal{N}(0.1, (0.05)^2)$$

□

b) Sei

$$u = \frac{x - \mu_Z}{\sigma_Z} = \frac{0.2 - 0.1}{0.05} = 2$$

Dann gilt damit:

$$P(X > 0.2) = 1 - P(X \leq 0.2) = 1 - \Phi(u) = 1 - \Phi(2) = 1 - 0.97725 = 0.02275$$

□

1.7 Kovarianz und Korrelation

Definition: Kovarianz

Die *Kovarianz* ein Zusammenhangsmaß für einen monotonen Zusammenhang zweier Zufallsvariablen mit gemeinsamer Wahrscheinlichkeitsverteilung. Der Wert dieser Kennzahl macht tendenzielle Aussagen darüber, ob hohe Werte der einen Zufallsvariablen eher mit hohen oder eher mit niedrigen Werten der anderen Zufallsvariablen einhergehen.

Die Kovarianz ist definiert als:

$$\text{Cov}(X, Y) = E((X - E(X)) \cdot (Y - E(Y)))$$

Es gilt:

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

Eigenschaften der Kovarianz:

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2. $\text{Cov}(X, X) = \text{Var}(X)$
3. $\text{Cov}(\alpha X + \beta, \gamma Y + \delta) = \alpha\gamma \cdot \text{Cov}(X, Y)$
4. X, Y stochastisch unabhängig $\implies \text{Cov}(X, Y) = 0$
- 5.

$$\text{Cov}\left(\sum_{i=1}^m \alpha_i X_i, \sum_{j=1}^n \beta_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j \text{Cov}(X_i, Y_j)$$

Zwei Zufallsvariablen X, Y heißen *unkorreliert*, falls $\text{Cov}(X, Y) = 0$ ist.

Im Allgemeinen darf aus der Unkorreliertheit zweier Zufallsvariablen nicht auf deren stochastische Unabhängigkeit geschlossen werden!

Definition: Korrelationskoeffizient

Der *Pearson'sche Korrelationskoeffizient* ist ein Maß für den Grad des linearen Zusammenhangs zwischen zwei Zufallsvariablen. Er kann Werte zwischen -1 und $+1$ annehmen. Bei einem Wert von $+1$ (bzw. -1) besteht ein vollständig positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen. Wenn der Korrelationskoeffizient den Wert 0 aufweist, hängen die beiden Merkmale überhaupt nicht linear voneinander ab. Allerdings können diese ungeachtet dessen in nichtlinearer Weise voneinander abhängen. Damit ist der Korrelationskoeffizient kein geeignetes Maß für die (reine) stochastische Abhängigkeit von Merkmalen.

Der Pearson'sche Korrelationskoeffizient ist definiert durch

$$\rho_{X,Y} := \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Von einer vorhandenen Korrelation zweier Zufallsvariablen darf aber nicht auf einen kausalen Zusammenhang geschlossen werden!

Beispiel: Kovarianz und Korrelationskoeffizient

In einem Beutel befinden sich 6 Münzen: eine 5-Cent-Münze, drei 2-Cent-Münzen und zwei 1-Cent-Münzen. Zufällig werden nacheinander - ohne Zurücklegen - 2 Münzen gezogen. X_1 gebe den Wert der ersten, X_2 den Wert der zweiten gezogenen Münzen an. Bestimmen Sie folgenden Werte:

- die zweidimensionale Wahrscheinlichkeitsfunktion $P(X_1 = i, X_2 = j)$ für $i, j \in \{1; 2; 5\}$.
- den Erwartungswert $E(X_i)$ und die Varianz $\text{Var}(X_i)$ ($i = 1, 2$).
- die Kovarianz $\text{Cov}(X_1, X_2)$ sowie den Korrelationskoeffizienten ρ_{X_1, X_2} .

a) Offensichtlich gilt:

$X_1 \backslash X_2$	1	2	5	$f_1(X_1)$
1	$1/15$	$1/5$	$1/15$	$1/3$
2	$1/5$	$1/5$	$1/10$	$1/2$
5	$1/15$	$1/10$	0	$1/6$
$f_2(X_2)$	$1/3$	$1/2$	$1/6$	1

□

b) Für X_1 gilt offensichtlich:

$$\mu_{X_1} = E(X_1) = \sum_i i \cdot f_1(i) = \dots = \frac{13}{6} \approx 2.1667$$

$$\sigma_{X_1}^2 = \text{Var}(X_1) = E((X_1 - \mu_{X_1})^2) = \sum_i f_1(i) \cdot (i - \mu_{X_1})^2 = \dots = \frac{65}{36} \approx 1.8056$$

In der Wahrscheinlichkeitsfunktion ist zu sehen, dass die Randverteilungen für beide Zufallsvariablen gleich sind. Damit gilt offensichtlich:

$$\mu_{X_2} = E(X_2) = \mu_{X_1} = \frac{13}{6} \approx 2.1667 \quad \wedge \quad \sigma_{X_2}^2 = \text{Var}(X_2) = \sigma_{X_1}^2 = \frac{65}{36} \approx 1.8056$$

□

c) Es gilt mit $i, j \in \{1, 2, 5\}$:

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E(X_1 X_2) - E(X_1) E(X_2) = \sum_i \sum_j i \cdot j \cdot f(i, j) - E(X_1) E(X_2) \\ &= \left(\frac{1}{15} + \frac{2}{5} + \frac{1}{3} + \frac{2}{5} + \frac{4}{5} + 1 + \frac{1}{3} + 1 + 0 \right) - \frac{13}{6} \cdot \frac{13}{6} \\ &= \frac{13}{3} - \frac{169}{36} = -\frac{13}{36} \approx -0.3611 \end{aligned}$$

und

$$\rho_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}} = \frac{-\frac{13}{36}}{\sqrt{\frac{65}{36}} \cdot \sqrt{\frac{65}{36}}} = -\frac{1}{5} = -0.2$$

□

1.8 Gesetze der großen Zahlen und Grenzwertsätze

Definition: Schwaches Gesetz großer Zahlen

Betrachte eine n -malige Wiederholung eines Zufallsexperimentes mit Messung der Zufallsvariablen (des Merkmals) X . Wir erhalten n stochastisch unabhängige Ergebnisse.

X_i sei die Realisierung der Zufallsvariablen im i -ten Versuch.

X_i ist dann eine Zufallsvariable mit der *gleichen Verteilung* wie X , insbesondere aber auch mit *gleichem Erwartungswert* und *gleicher Varianz*.

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

ist dann auch eine Zufallsvariable mit

$$\implies E(\bar{X}_{(n)}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n \cdot E(X) = E(X)$$

und

$$\implies \text{Var}(\bar{X}_{(n)}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n \cdot \text{Var}(X) = \frac{1}{n} \text{Var}(X)$$

Nach Tschebyscheff gilt dann:

$$P(|\bar{X}_{(n)} - \mu| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_{(n)})}{\epsilon^2} = \frac{\text{Var}(X)}{n} \cdot \frac{1}{\epsilon^2} = \frac{1}{n} \cdot \frac{\sigma^2}{\epsilon^2}$$

Mit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \cdot \frac{\sigma^2}{\epsilon^2} = 0 \quad \text{und} \quad \bar{X}_{(n)} = \frac{1}{n} \sum_{i=1}^n X_i$$

erhalten wir dann das *Schwache Gesetz großer Zahlen*

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \epsilon \right) = 0$$

Dieses besagt „stochastische Konvergenz“: Die Wahrscheinlichkeit, dass ein Mittelwert um mehr als ein beliebiges, festes $\epsilon > 0$ von μ abweicht, geht für wachsendes n gegen Null.

Es wird aber nicht ausgeschlossen, dass auch für unbeschränkt wachsende n gelegentlich noch größere Abweichungen als ϵ auftreten. Die Wahrscheinlichkeit für solche „Ausreißer“ geht mit wachsendem n aber gegen Null.

Definition: Starkes Gesetz großer Zahlen

Bei gleichen Voraussetzungen wie in der Herleitung für das schwache Gesetz großer Zahlen gilt auch das *Starke Gesetz großer Zahlen*:

$$P \left(\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| = 0 \right) = 1$$

Bei unbeschränkt oftmaliger Wiederholung der Messung einer Zufallsgröße konvergiert das Stichprobenmittel gegen den Erwartungswert von X .

Das Starke Gesetz großer Zahlen kann ebenfalls „Ausreißer“ nicht mit Sicherheit ausschließen.

Definition: Zentraler Grenzwertsatz

Die unabhängigen Zufallsgrößen X_1, X_2, \dots sollen die gleiche Verteilung haben mit

$$E(X_1) = E(X_2) = \dots = \mu \quad \text{und} \quad \text{Var}(X_1) = \text{Var}(X_2) = \dots = \sigma^2$$

Betrachte die Summe

$$S_n := \sum_{i=1}^n X_i$$

Dann gilt offensichtlich

$$E(S_n) = \sum_{i=1}^n E(X_i) = n \cdot \mu \quad \text{und} \quad \text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i) = n \cdot \sigma^2$$

Betrachte nun die „standardisierte Version“ von S_n als neue Zufallsgröße:

$$S_n^* = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sqrt{n} \cdot \sigma} = \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$$

Dann gilt der *Zentrale Grenzwertsatz* nach Lindeberg und Lévy, welcher besagt:

Für die Standardisierung

$$S_n^* = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}}$$

der Summe

$$S_n = \sum_{i=1}^n X_i$$

gilt die Grenzwertbeziehung

$$\lim_{n \rightarrow \infty} P(S_n^* \leq s) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^s e^{-\frac{x^2}{2}} dx \quad \forall s \in \mathbb{R}$$

Der zentrale Grenzwertsatz liefert also die Begründung für das Phänomen, dass sich bei der additiven Überlagerung vieler kleiner unabhängiger Zufallseffekte zu einem Gesamteffekt zumindest approximativ eine Normalverteilung ergibt.

Beispiel: Zentraler Grenzwertsatz

Bei der Verpackung von Kartoffeln in Beutel kann das Normalgewicht von 10kg i.A. nicht exakt eingehalten werden. Die Erfahrung zeigt, dass das Füllgewicht eines Beutels durch eine Zufallsvariable $Y = X + 10$ beschrieben werden kann, wobei X eine auf dem Intervall $[-0.25, 0.75]$ gleichverteilte Zufallsvariable ist.

- Berechnen Sie den Erwartungswert und die Varianz des Füllgewichtes eines Beutels.
- Die abgefüllten Beutel sollen mit einem Kleintransporter befördert werden. Berechnen Sie näherungsweise die Wahrscheinlichkeit dafür, dass die zulässige Nutzlast von 1020kg bei Zuladung von 100 Beuteln überschritten wird.

a) Offensichtlich gilt:

$$\begin{aligned} E(X) &= \frac{a+b}{2} = \frac{0.75 - 0.25}{2} = \frac{1}{4} \\ \text{Var}(X) &= \frac{a^2 + b^2 - (a+b)^2}{12} = \frac{1}{12} \end{aligned}$$

Und für $Y = X + 10$:

$$\mu_Y = E(Y) = E(X) + 10 = \frac{41}{4} \quad \wedge \quad \sigma_Y^2 = \text{Var}(Y) = \text{Var}(X) = \frac{1}{12}$$

□

b) Sei

$$S_n = Y_1 + \dots + Y_{100}$$

dann gilt nach dem Zentralen Grenzwertsatz:

$$E(S_n) = n\mu_Y = 100 \cdot \mu_Y = 1025 \quad \wedge \quad \text{Var}(S_n) = n\sigma_Y^2 = 100 \cdot \sigma_Y^2 = \frac{25}{3}$$

Sei dann die standardisierte Zufallsvariable Z_n gegeben mit

$$Z_n = \frac{S_n - E(S_n)}{\sigma_{S_n}} = \frac{S_n - n\mu_Y}{\sigma_Y \sqrt{n}} = \sqrt{3} \cdot \frac{S_n - 1025}{5}$$

dann besagt der Zentrale Grenzwertsatz, dass die Verteilungsfunktion von Z_n für $n \rightarrow \infty$ punktweise gegen die der Standardnormalverteilung $N(0, 1)$ konvergiert.

Wir wählen

$$u = \sqrt{3} \cdot \frac{1020 - 1025}{5} = -\sqrt{3} \approx -1.73$$

Damit gilt:

$$P(Z_n \geq u) = 1 - P(Z_n < u) \approx 1 - \Phi(-1.73) = 1 - (1 - \Phi(1.73)) = \Phi(1.73) = 95.82\%$$

□

^aSiehe: [Wikipedia](#)

2 Beschreibende Statistik

Bonus: Aufgabe der beschreibenden Statistik

Die *Aufgabe der beschreibenden Statistik* besteht darin, große und unübersichtliche Datenmengen so aufzuarbeiten, dass wenige aussagekräftige Kenngrößen und/oder Grafiken entstehen, in denen die gesamte Datenmenge „fokussiert“ ist.

2.1 Merkmale und weitere wichtige Begriffe

Definition: Beobachtungsmenge

Die *Beobachtungsmenge* bzw. *statistische Masse* bezeichnet diejenige Menge aller Objekte, über die eine Aussage getroffen werden soll, also die Menge aller statistischen Einheiten (auch Merkmalsträger, Untersuchungseinheit, Erhebungseinheit, Beobachtungseinheit) mit übereinstimmenden Identifikationskriterien (sachlich, räumlich und zeitlich).

Definition: Beobachtungseinheit

Die *Beobachtungseinheit* bzw. *statistische Einheit* ist Träger der Informationen für die statistische Untersuchung.

Statistische Einheiten können natürliche Einheiten (Personen, Tiere, Pflanzen, Werkstücke), aber auch künstliche Einheiten, zum Beispiel sozio-ökonomische Einheiten (Familien, Haushalte, Unternehmen) oder Ereignisse, sein.

Definition: Statistische Variable

Eine *statistische Variable* oder ein *statistisches Merkmal* ordnet einer Beobachtungseinheit (Untersuchungseinheit) eine Ausprägung oder einen Wert zu.

Eine statistische Variable liegt vor, wenn sich Ausprägungen bestimmter Merkmale durch eine Zahl oder durch Zahlenintervalle (Werte der Variablen) ausdrücken lassen und zu diesen Werten empirisch messbare Häufigkeiten gehören.

Offensichtlich müssen verschiedene Typen von Merkmalen unterschieden werden:

- *Qualitative Merkmale*: Die Werte brauchen keine physikalische Einheit.
 - *Qualitativ-nomiale Merkmale*: Merkmalsausprägungen sind nur dem Namen nach unterscheidbar, drücken aber keinerlei Wertung oder Intensität aus (z.B. Familienstand, Studienrichtung)
 - *Qualitativ-ordinale Merkmale*: Merkmalsausprägungen können zusätzlich noch in eine inhaltlich sinnvolle Rangordnung gebracht werden, aber keine definierte Skala (z.B. Interesse am Vorlesungsgegenstand)
- *Quantitative Merkmale*: auch „metrische“ oder „kardinale“ Merkmale
 - *Quantitativ-diskrete Merkmale*: Merkmale, die nur bestimmte, auf der Zahlengerade getrennt liegende Werte annehmen können (z.B. Anzahl Geschwister)
 - *Quantitativ-stetige Merkmale*: Werden durch Messung gewonnen und können jeden Wert innerhalb eines sinnvollen Intervall es annehmen (z.B. Körpergröße, -gewicht)

2.2 Darstellung der Beobachtungsergebnisse

Definition: Absolute Häufigkeit

Der Begriff *absolute Häufigkeit* ist gleichbedeutend mit dem umgangssprachlichen Begriff Anzahl.

Die absolute Häufigkeit ist das Ergebnis einer einfachen Zählung von Objekten oder Ereignissen (besser Elementarereignissen). Sie gibt an, wie viele Elemente mit dem gleichen interessierenden Merkmal gezählt wurden.

Für die absolute Häufigkeit n_i des Merkmalswertes a_i bei n beobachteten Merkmalswerten gilt also:

$$0 \leq n_i \leq n \quad \wedge \quad \sum_i n_i = n$$

Definition: Relative Häufigkeit

Die *relative Häufigkeit* gibt den Anteil der Elemente einer Menge wieder, bei denen eine bestimmte Merkmalsausprägung vorliegt. Sie wird berechnet, indem die absolute Häufigkeit eines Merkmals in einer zugrundeliegenden Menge durch die Anzahl der Objekte in dieser Menge geteilt wird.

Für die relative Häufigkeit h_i des Merkmalswertes a_i bei n beobachteten Merkmalswerten gilt also:

$$h_i := \frac{n_i}{n} \quad \wedge \quad 0 \leq h_i \leq 1 \quad \wedge \quad \sum_i h_i = 1$$

Beispiel: Stabdiagramm

In den 30 Museen der Stadt Artima gab es im letzten Monat jeweils X Neuerwerbungen pro Museum. Dabei sei folgende Urliste entstanden:

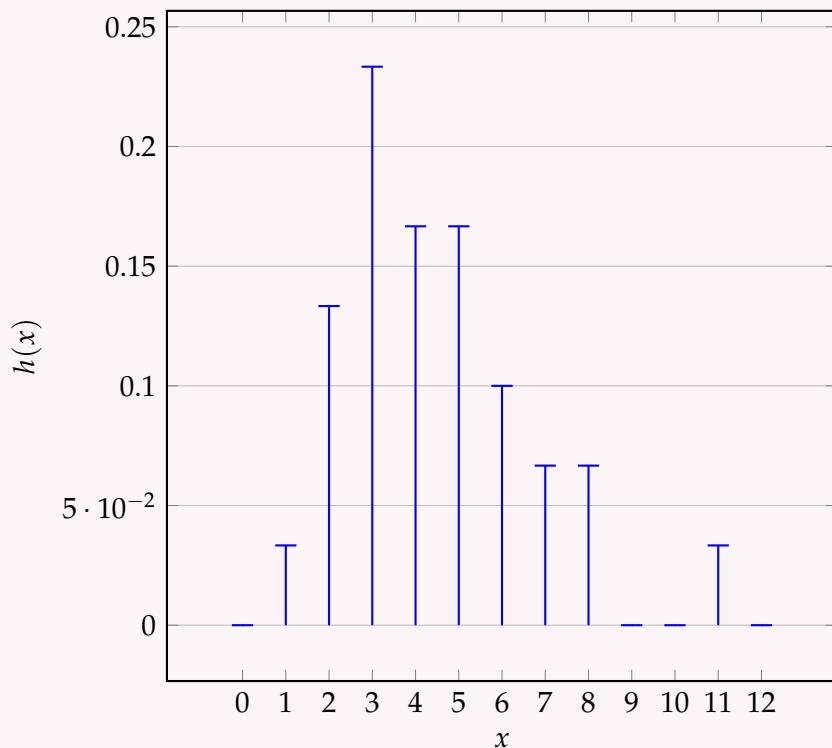
$$\begin{array}{cccccccccc} 2 & 4 & 3 & 5 & 5 & 2 & 3 & 1 & 5 & 6 \\ 4 & 7 & 8 & 3 & 2 & 8 & 3 & 6 & 4 & 6 \\ 5 & 7 & 3 & 3 & 2 & 5 & 4 & 4 & 3 & 11 \end{array}$$

Erstellen Sie eine Tabelle mit der absoluten und relativen Häufigkeit bzw. Summenhäufigkeit der Neuerwerbungen X pro Museum.

Es gilt:

x_i	n_i	h_i	H_i
1	1	$1/30$	$1/30$
2	4	$2/15$	$1/6$
3	7	$7/30$	$2/5$
4	5	$1/6$	$17/30$
5	5	$1/6$	$11/15$
6	3	$1/10$	$5/6$
7	2	$1/15$	$9/10$
8	2	$1/15$	$29/30$
11	1	$1/30$	1

Das zugehörige Stabdiagramm ist dann:



Definition: Summenhäufigkeit

Die *Summenhäufigkeit* oder *kumulierte Häufigkeit* gibt an, bei welcher Anzahl der Merkmalsträger in einer empirischen Untersuchung die Merkmalsausprägung kleiner ist als eine bestimmte Schranke. Die kumulierte Häufigkeit wird berechnet als Summe der Häufigkeiten der Merkmalsausprägungen von der kleinsten Ausprägung bis hin zu der jeweils betrachteten Schranke.

Für die absolute Summenhäufigkeit $F_{\text{abs}}(x)$ gilt also:

$$F_{\text{abs}}(x) := \sum_{i:a_i \leq x} n_i \quad x \in \mathbb{R}$$

Für die relative Summenhäufigkeit $F_{\text{rel}}(x)$ gilt also:

$$F_{\text{rel}}(x) := \sum_{i:a_i \leq x} h_i = \frac{F_{\text{abs}}}{n} \quad x \in \mathbb{R}$$

Definition: Empirische Verteilungsfunktion

Eine *empirische Verteilungsfunktion* oder *Summenhäufigkeitsfunktion* ist in der beschreibenden Statistik und der Stochastik eine Funktion, die jeder reellen Zahl x den Anteil der Stichprobenwerte, die kleiner oder gleich x sind, zuordnet.

Beispiel: Empirische Verteilungsfunktion

In den 30 Museen der Stadt Artima gab es im letzten Monat jeweils X Neuerwerbungen pro Museum. Dabei sei folgende Urliste entstanden:

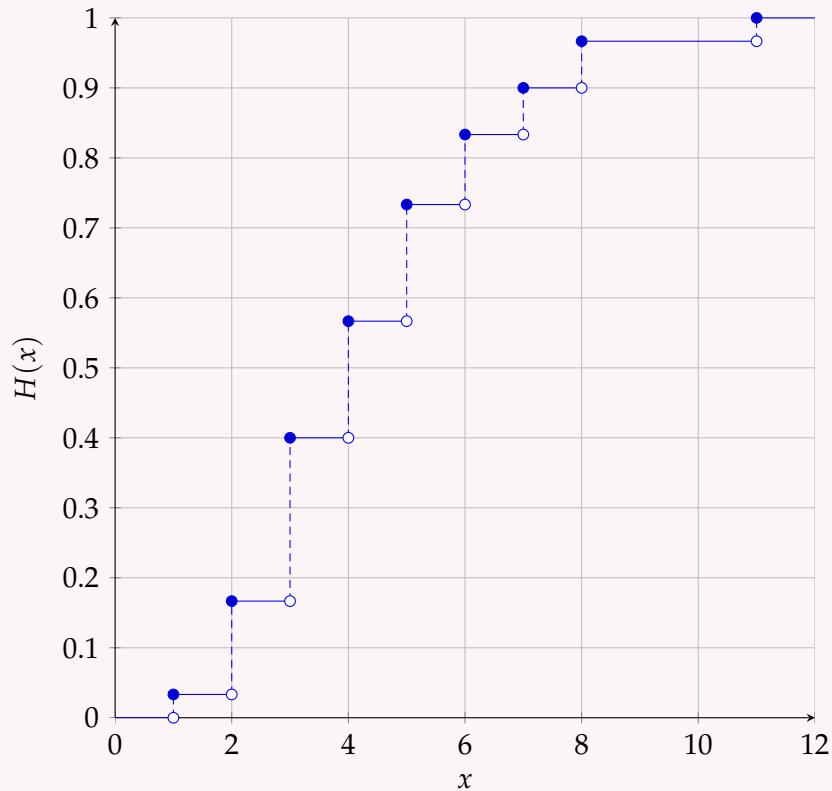
$$\begin{array}{cccccccccc} 2 & 4 & 3 & 5 & 5 & 2 & 3 & 1 & 5 & 6 \\ 4 & 7 & 8 & 3 & 2 & 8 & 3 & 6 & 4 & 6 \\ 5 & 7 & 3 & 3 & 2 & 5 & 4 & 4 & 3 & 11 \end{array}$$

Erstellen Sie eine Tabelle mit der absoluten und relativen Häufigkeit bzw. Summenhäufigkeit der Neuerwerbungen X pro Museum.

Es gilt:

x_i	n_i	h_i	H_i
1	1	$1/30$	$1/30$
2	4	$2/15$	$1/6$
3	7	$7/30$	$2/5$
4	5	$1/6$	$17/30$
5	5	$1/6$	$11/15$
6	3	$1/10$	$5/6$
7	2	$1/15$	$9/10$
8	2	$1/15$	$29/30$
11	1	$1/30$	1

Die zugehörige empirische Verteilungsfunktion:



Definition: Klasseneinteilung

Klasseneinteilung oder Klassierung bezeichnet in der Statistik die Einteilung von Merkmalswerten oder statistischen Reihen in getrennte Gruppen, Klassen oder Größenklassen.

Klassen sind disjunkte, d. h. nicht überlappende, aneinandergrenzende Intervalle von Merkmalswerten, die durch eine untere und eine obere Klassengrenze begrenzt und eindeutig festgelegt sind.

Da es bei statistischen Untersuchungen oft nicht möglich oder sinnvoll ist, alle einzelnen (verschiedenen) Merkmalsausprägungen oder Realisierungen der untersuchten Zufallsvariablen zu erheben oder zu verarbeiten, kann durch eine Klassierung eine bessere Übersicht über die Daten erreicht werden. Das trifft insbesondere auf stetige Merkmale zu.

Für den Gewinn an Übersichtlichkeit zahlt man mit einem Informationsverlust, denn über die Verteilung der Werte innerhalb einer Klasse ist dann nichts mehr bekannt.

Sei das Intervall $[a, b]$ gegeben. Dann definiert man eine Einteilung des Intervalls in disjunkte Klassen A_1, \dots, A_k , wobei gilt

$$A_i = (a_{i-1}, a_i] \quad a = a_0 < a_1 < \dots < a_k = b$$

Im Allgemeinen sind die Klassen äquidistant. Man definiert

$$\alpha_i = \frac{a_i + a_{i-1}}{2}$$

als Klassenmitten.

Sind Daten in Klassen eingeteilt, so bezeichnet man die absolute oder auch relative Häufigkeit der Werte einer Klasse als Klassenhäufigkeit.

Definition: Absolute Klassenhäufigkeit

Um die *absoluten Klassenhäufigkeiten* zu bestimmen, zählt man einfach nur, wie oft Werte aus der Stichprobe in der jeweiligen Klasse liegen.

Sei n_i die Anzahl des Vorkommens der Klasse A_i bei n beobachteten Merkmalswerten.

Es gilt:

$$0 \leq n_i \leq n \quad \wedge \quad \sum_i n_i = n$$

Definition: Relative Klassenhäufigkeit

Um die *relativen Klassenhäufigkeiten* zu bestimmen, teilt man jeweils die absolute Klassenhäufigkeit durch die Gesamtzahl n der Stichprobenwerte^a.

Sei h_i die relative Häufigkeit der Klasse A_i bei n beobachteten Merkmalswerten.

Es gilt:

$$h_i := \frac{n_i}{n} \quad \wedge \quad 0 \leq h_i \leq 1 \quad \wedge \quad \sum_i h_i = 1$$

^anicht die Anzahl der Klassen!

Definition: Häufigkeitsdichte

Die *Häufigkeitsdichte* spielt bei klassierten Merkmale eine Rolle. So gibt die Häufigkeitsdichte bei einem Histogramm die Höhe des Rechtecks an.

Ausgedrückt ist die Häufigkeitsdichte einer Klasse das Verhältnis der absoluten oder der relativen Häufigkeit einer Klasse zur entsprechenden Klassenbreite.

Es gilt:

$$h(x) := \frac{h_i}{|A_i|} = \frac{h_i}{a_i - a_{i-1}} \quad \wedge \quad x \in A_i = (a_{i-1}, a_i]$$

Definition: Histogramm

Der Graph von $h(x)$ ist ein *Histogramm*.

Es werden direkt nebeneinanderliegende Rechtecke von der Breite der jeweiligen Klasse gezeichnet, deren Flächeninhalte die (relativen oder absoluten) Klassenhäufigkeiten darstellen.

Die Höhe jedes Rechtecks stellt dann die (relative oder absolute) Häufigkeitsdichte dar, also die (relative oder absolute) Häufigkeit dividiert durch die Breite der entsprechenden Klasse.

Beispiel: Histogramm

In einer (kleinen) Bankfiliale werden die vergebenen Kredite untersucht:

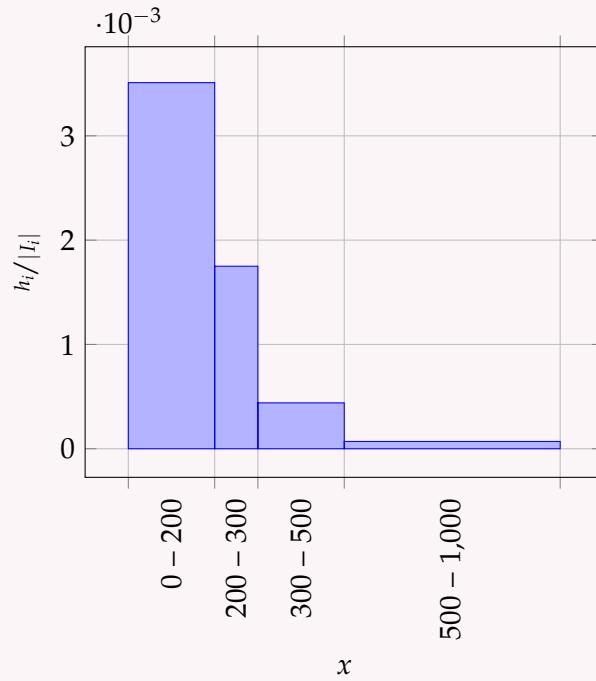
Kredithöhe (in Tausend €)	Anzahl der Kredite
[0; 200)	40
[200; 300)	10
[300; 500)	5
[500; 1000)	2

Erstellen Sie ein Histogramm.

Es gilt für das Histogramm:

I_i	n_i	h_i	$h_i/ I_i $
[0; 200)	40	40/57	1/285 ≈ 0.00351
[200; 300)	10	10/57	1/570 ≈ 0.00175
[300; 500)	5	5/57	1/2280 ≈ 0.00044
[500; 1000)	2	2/57	1/14250 ≈ 0.00007

Und damit:



□

Beispiel: Empirische Verteilungsfunktion (klassierte Daten)

In einer (kleinen) Bankfiliale werden die vergebenen Kredite untersucht:

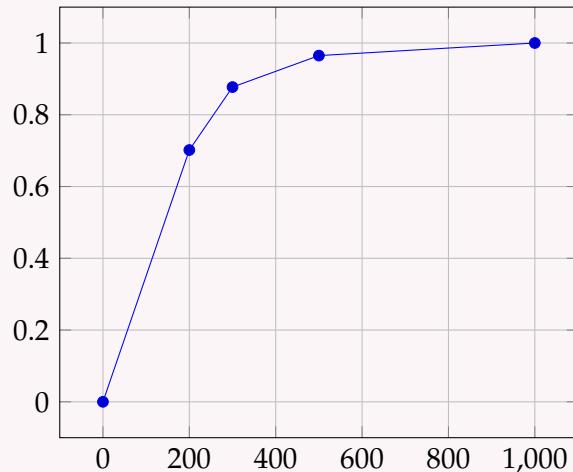
Kredithöhe (in Tausend €)	Anzahl der Kredite
[0; 200)	40
[200; 300)	10
[300; 500)	5
[500; 1000)	2

Skizzieren Sie die empirische Verteilungsfunktion.

Es gilt für die Verteilungsfunktion:

I_i	n_i	h_i	H_i
[0; 200)	40	$40/57$	$40/57$
[200; 300)	10	$10/57$	$50/57$
[300; 500)	5	$5/57$	$55/57$
[500; 1000)	2	$2/57$	1

Und damit:



□

2.3 Statistische Maßzahlen

Definition: Arithmetisches Mittel

Das *arithmetische Mittel* berechnet man, indem man die Summe der betrachteten Zahlen durch ihre Anzahl teilt und beschreibt damit das Zentrum einer Verteilung durch einen numerischen Wert und stellt einen Lageparameter dar.

Es gilt:

- Für Stichproben in einer Urliste, wobei x_i die Ausprägung des i -ten Elements darstellt:

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

- Für eine unklassierte Häufigkeitstabelle, wobei h_i die relative Häufigkeit des i -ten Elements a_i darstellt:

$$\bar{x} = \sum_{i=1}^n h_i \cdot a_i$$

- Für eine klassierte Häufigkeitstabelle, wobei h_i die relative Häufigkeit der i -ten Klasse A_i , und α_i die Klassenmitte darstellt:

$$\bar{x} \approx \sum_{i=1}^n h_i \cdot \alpha_i$$

Definition: Median

Der *Median* der Messwerte einer Urliste ist derjenige Messwert, der genau „in der Mitte“ steht, wenn man die Messwerte der Größe nach sortiert.

Im Allgemeinen teilt ein Median einen Datensatz, eine Stichprobe oder eine Verteilung so in zwei gleich große Teile, dass die Werte in der einen Hälfte nicht größer als der Medianwert sind und in der anderen nicht kleiner.

Es gilt:

- Für Stichproben in einer *sortierten* Urliste, wobei x_i die Ausprägung des i -ten Elements darstellt:

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

- Für eine unklassierte Häufigkeitstabelle, wobei H_i die kumulierte relative Häufigkeit bis zum i -ten Element a_i darstellt:
 - \tilde{x} ist das Merkmal a_i , bei dem $H_i = 0.5$ zum ersten Mal überschritten wird.
- Für eine klassierte Häufigkeitstabelle, wobei H_i die kumulierte relative Häufigkeit bis zur i -ten Klasse A_i darstellt:
 - \tilde{x} befindet sich in der *Einfallsklasse*, die zum ersten Mal $H_i = 0.5$ erreicht.
 - Sei $A_j = (a_j, b_j]$ die Einfallsklasse. Dann gilt:^a

$$\tilde{x} = x_{0.5} = a_j + \frac{0.5 - H_{j-1}}{H_j - H_{j-1}} \cdot (b_j - a_j)$$

^aSiehe *Quantil (empirisch)* für mehr Informationen.

Definition: Modalwert

Der *Modalwert* ist diejenige Merkmalsausprägung mit der größten (absoluten oder relativen) Häufigkeit. Er ist jedoch im Allgemeinen nicht eindeutig.

Es gilt:

- Für Stichproben ist \bar{x}_M das Element, das am häufigsten vorkommt.
- Bei klassierten Daten ist die Modalklasse diejenige Klasse mit der größten Häufigkeitsdichte.

Definition: Quantil (empirisch)

Für jede Zahl p zwischen 0 und 1 teilt ein *empirisches p-Quantil* die Stichprobe so, dass ein Anteil der Stichprobe von p kleiner als das empirische p -Quantil ist und ein Anteil von $1 - p$ der Stichprobe größer als das empirische p -Quantil ist.

Es gilt:

- Für Stichproben in einer *sortierten* Urliste, wobei x_i die Ausprägung des i -ten Elements darstellt:

$$x_p = \begin{cases} \frac{1}{2} (x_{n \cdot p} + x_{n \cdot p + 1}) & \text{falls } n \cdot p \text{ ganzzahlig} \\ x_{\lfloor n \cdot p + 1 \rfloor} & \text{falls } n \cdot p \text{ nicht ganzzahlig} \end{cases}$$

- Für eine unklassierte Häufigkeitstabelle, wobei h_i die relative Häufigkeit des i -ten Elements a_i darstellt:
 - x_p ist das Merkmal a_i , bei dem $H_i = p$ zum ersten Mal überschritten wird.
- Für eine klassierte Häufigkeitstabelle, wobei H_i die kumulierte relative Häufigkeit bis zur i -ten Klasse A_i darstellt:
 - x_p befindet sich in der *Einfallsklasse*, die zum ersten Mal $H_i = p$ erreicht.
 - Sei $A_j = (a_j, b_j]$ die Einfallsklasse. Dann gilt:

$$x_p = a_j + \frac{p - H_{j-1}}{H_j - H_{j-1}} \cdot (b_j - a_j)$$

2.4 Streuungsmaße

Definition: Spannweite

Die *Spannweite R* ist das einfachste Streuungsmaß in der Statistik und misst die Streuung in den Beobachtungen ordinalskalierter Merkmale.

Es gilt:

$$R = x_n - x_1 = x_{\max} - x_{\min}$$

Die Spannweite ist nicht robust gegenüber Ausreißern, sie hängt nur von den Extremwerten ab und verliert bei zunehmendem Stichprobenumfang an Informationsgehalt. Sie wird daher vor allem bei kleinen Stichprobenumfängen genutzt.

Definition: Quartilsabstand

Der *Quartilsabstand* bzw. *Interquartilsabstand* Q gibt für eine sortierte Stichprobe an, wie breit das Intervall ist, in dem die mittleren 50% der Stichprobeelemente liegen.

Es gilt:

$$Q = x_{0.75} - x_{0.25}$$

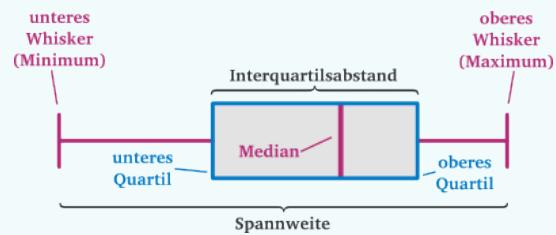
Definition: Box-Plot

Der *Box-Plot* ist ein Diagramm, das zur grafischen Darstellung der Verteilung eines mindestens ordinalskalierten Merkmals verwendet wird.

Es fasst dabei verschiedene robuste Streuungs- und Lagemaße in einer Darstellung zusammen.

Ein Box-Plot soll schnell einen Eindruck darüber vermitteln, in welchem Bereich die Daten liegen und wie sie sich über diesen Bereich verteilen. Deshalb werden alle Werte der sogenannten Fünf-Punkte-Zusammenfassung, also der Median, die zwei Quartile und die beiden Extremwerte, dargestellt.

Zusammengefasst lassen sich folgende Kennwerte ablesen:



Kennwert	Beschreibung	Lage im Box-Plot
Minimum	Kleinster Datenwert des Datensatzes	Ende eines Whiskers oder entferntester Ausreißer
Unteres Quartil	Die kleinsten 25% der Datenwerte sind kleiner als dieser oder gleich diesem Kennwert	Beginn der Box
Oberes Quartil	Die kleinsten 75% der Datenwerte sind kleiner als dieser oder gleich diesem Kennwert	Ende der Box
Maximum	Größter Datenwert des Datensatzes	Ende eines Whiskers oder entferntester Ausreißer
Spannweite	Gesamter Wertebereich des Datensatzes	Länge des gesamten Box-Plots (inklusive Ausreißer)
Quartilsabstand	Wertebereich, in dem sich die mittleren 50% der Daten befinden	Ausdehnung der Box

Definition: Empirische Varianz

Die *empirische Varianz* bzw. *Stichprobenvarianz* ist eine statistische Angabe für die Streubreite von Werten einer Stichprobe.

Sie gehört zu den Streuungsmaßen und beschreibt die mittlere quadratische Abweichung der einzelnen Messwerte vom empirischen Mittelwert.

Sie stellt damit eine Art durchschnittliches Abweichungsquadrat dar.

Die positive Wurzel der empirischen Varianz ist die *empirische Standardabweichung*. Die empirische Standardabweichung stellt das gebräuchlichste Streuungsmaß dar.

Es gilt:

- Für Stichproben in einer Urliste, wobei x_i die Ausprägung des i -ten Elements darstellt:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right)$$

- Für eine unklassierte Häufigkeitstabelle, wobei n_j die absolute Häufigkeit des j -ten Elements a_j darstellt:

$$s^2 = \frac{1}{n-1} \sum_{j=1}^m (a_j - \bar{x})^2 \cdot n_j = \frac{1}{n-1} \left(\sum_{j=1}^m a_j^2 \cdot n_j - \bar{x}^2 \right)$$

- Für eine klassierte Häufigkeitstabelle, wobei n_j die absolute Häufigkeit der j -ten Klasse A_j , und α_j die Klassenmitte darstellt:

$$s^2 \approx \frac{1}{n-1} \sum_{j=1}^m \alpha_j^2 \cdot n_j - \frac{1}{n(n-1)} \left(\sum_{j=1}^m \alpha_j \cdot n_j \right)^2$$

Insgesamt gilt natürlich jeweils für die empirische Standardabweichung:

$$s = \sqrt{s^2}$$

Bonus: Variationskoeffizient

Die Motivation für den *Variationskoeffizienten* ist, dass eine statistische Variable mit großem Mittelwert bzw. eine Zufallsvariable mit großem Erwartungswert im Allgemeinen eine größere Varianz aufweist als eine mit einem kleinen Mittel- bzw. Erwartungswert.

Da die Varianz und die daraus abgeleitete Standardabweichung nicht normiert sind, kann ohne Kenntnis des Mittelwerts nicht beurteilt werden, ob eine Varianz groß oder klein ist.

Der Variationskoeffizient ist eine Normierung der Varianz: Ist die Standardabweichung größer als der Mittelwert bzw. der Erwartungswert, so ist der Variationskoeffizient größer 1.

Es gilt:

$$\text{VarK} = \frac{s}{\bar{x}}$$

2.5 Lineare Regression und Korrelation

Definition: Empirische Kovarianz

Die *empirische Kovarianz* bzw. *Stichprobenkovarianz* ist eine nichtstandardisierte Maßzahl für den (linearen) Zusammenhang zweier statistischer Variablen.

Ist die Kovarianz positiv, dann gehen kleine Werte der einen Variable überwiegend einher mit kleinen Werten der anderen Variable und gleichfalls für große Werte. Für eine negative Kovarianz ist das genau umgekehrt.

Es werden benötigt:

- Empirische Varianzen der beiden Variablen X und Y
 - s_x^2 und s_y^2
- Arithmetische Mittelwerte der beiden Variablen X und Y
 - \bar{x} und \bar{y}

Es gilt:

$$\text{Cov}(X, Y) = s_{x,y} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y} \right)$$

Definition: Korrelationskoeffizient (empirisch)

Die Größe der Kovarianz lässt sich nicht sinnvoll interpretieren.

Der *empirische Korrelationskoeffizient* ist ein Maß für den Grad des linearen Zusammenhangs zwischen zwei Merkmalen, das nicht von den Maßeinheiten der Messung abhängt und somit dimensionslos ist.

Er kann Werte zwischen -1 und $+1$ annehmen. Bei einem Wert von $+1$ (bzw. -1) besteht ein vollständig positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen.

Wenn der Korrelationskoeffizient den Wert 0 aufweist, hängen die beiden Merkmale überhaupt nicht linear voneinander ab. Allerdings können diese ungeachtet dessen in nichtlinearer Weise voneinander abhängen.

Es gilt:

$$r_{x,y} = \frac{s_{x,y}}{s_x \cdot s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Klärung beliebter Missverständnisse:

- $r_{x,y}$ sagt nichts über die Größe der Geradensteigung aus.
- $r_{x,y} = 0$ bedeutet nur, dass zwischen X und Y kein *linearer* Zusammenhang besteht.
- $r_{x,y}$ nahe bei ± 1 bedeutet keinen kausalen Zusammenhang.
- $r_{x,y}$ gilt nur für quantitative Merkmale.

Bonus: Bestimmtheitsmaß

Das *Bestimmtheitsmaß* bewertet als Quadrat des Korrelationskoeffizienten die Anpassungsgüte der zu einem Datensatz ermittelten Regressionsgerade und hat einen Wert zwischen 0 und 1, wobei der Wert 1 die Situation beschreibt, dass alle Datenpaare auf einer Geraden liegen und damit perfekte Anpassung vorliegt.

Es gilt:

$$B_{x,y} = r_{x,y}^2$$

Algorithmus: Lineare Regression

Das Ziel einer *Regression* ist es, eine abhängige Variable durch eine oder mehrere unabhängige Variablen zu erklären.

Bei der *einfachen linearen Regression* wird eine abhängige Variable durch lediglich eine unabhängige Variable erklärt.

Das Modell der linearen Einfachregression geht daher von zwei metrischen Größen aus:

- einer Einflussgröße X (auch: erklärende Variable, Regressor oder unabhängige Variable)
- einer Zielgröße Y (auch: abhängige Variable, erklärte Variable oder Regressand)

Des Weiteren liegen n Paare $(x_1, y_1), \dots, (x_n, y_n)$ von Messwerten vor, die in einem funktionalen Zusammenhang stehen:^a

$$Y_i = \underbrace{f(x_i; \hat{a}, \hat{b}, \dots)}_{\text{systematische Komponente}} + \underbrace{\epsilon_i}_{\text{stochastische Komponente}}$$

Die stochastische Komponente beschreibt nur noch zufällige Einflüsse (z. B. zufällige Abweichungen wie Messfehler), alle systematischen Einflüsse sind in der systematischen Komponente enthalten.

Die lineare Einfachregression stellt den Zusammenhang zwischen der Einfluss- und der Zielgröße mithilfe von zwei festen, unbekannten, reellen Parametern \hat{a} und \hat{b} auf lineare Weise her.

Die Regressionsfunktion ist wie folgt definiert:

$$f(x_i; \hat{a}, \hat{b}) = \hat{a} + \hat{b}x_i$$

Wir schätzen mithilfe der Methode der kleinsten Quadrate und erhalten als Regressionsparameter \hat{a} und \hat{b} :

$$\begin{aligned}\hat{b} &= \frac{s_{x,y}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{a} &= \bar{y} - \hat{b}\bar{x}\end{aligned}$$

Damit erhalten wir die Regressionsgerade:^b

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

^aDiese werden meist als Streudiagramm dargestellt.

^bFür jeden Wert x_i können damit die Störgrößen bzw. Residuen berechnet werden mit $\hat{\epsilon}_i = y_i - \hat{y}_i$.

Beispiel: Lineare Regression, Empirische Kovarianz

Die Schülerinnen und Schüler wurden vor der Klausur anonym befragt, wie viele Stunden Schlaf sie vor der Klausur gehabt haben:

Note	2	4	3	5	6	6	1	2	5	4
Schlaf	9	5	6	5	1	2	9	8	4	7

Berechnen Sie aus den Daten:

- a) die empirische Kovarianz
- b) den empirischen Korrelationskoeffizient
- c) die lineare Regression

Erstellen Sie ein Streudiagramm und zeichnen Sie die Regressionsgerade ein.

a) Es gilt:

$$\text{Cov}(X, Y) = s_{XY} = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y} \right)$$

Wir berechnen zuerst die arithmetischen Mittel von X und Y wie folgt:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} \cdot (2 + 4 + 3 + 5 + 6 + 6 + 1 + 2 + 5 + 4) = \frac{19}{5} = 3.8$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{10} \cdot (9 + 5 + 6 + 5 + 1 + 2 + 9 + 8 + 4 + 7) = \frac{28}{5} = 5.6$$

Damit gilt:

$$\begin{aligned} \text{Cov}(X, Y) = s_{xy} &= \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y} \right) \\ &= \frac{1}{9} \left(\sum_{i=1}^n x_i y_i - 10 \cdot \frac{19}{5} \cdot \frac{28}{5} \right) \\ &= \frac{1}{9} \left(\sum_{i=1}^n x_i y_i - \frac{1064}{5} \right) \\ &= \frac{1}{9} \sum_{i=1}^n x_i y_i - \frac{1064}{45} \\ &= \frac{1}{9} \cdot 172 - \frac{1064}{45} \\ &= -\frac{68}{15} \approx -4.53 \end{aligned}$$

Beispiel: Lineare Regression, Empirischer Korrelationskoeffizient

b) Es gilt:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Wir berechnen zuerst die empirischen Standardabweichungen s_x und s_y wie folgt:

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{9} \sum_{i=1}^n \left(x_i - \frac{19}{5} \right)^2 = \frac{46}{15} \implies s_x = \sqrt{\frac{46}{15}} = \frac{\sqrt{690}}{15}$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{9} \sum_{i=1}^n \left(y_i - \frac{28}{5} \right)^2 = \frac{38}{5} \implies s_y = \sqrt{\frac{38}{5}} = \frac{\sqrt{190}}{5}$$

Damit gilt dann:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{-68/15}{\sqrt{690}/15 \cdot \sqrt{190}/5} = -\frac{34\sqrt{1311}}{1311} \approx -0.93903$$

Beispiel: Lineare Regression, Regressionsgleichung

c) Die Regressionsgerade ist gegeben mit:

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

Es gilt:

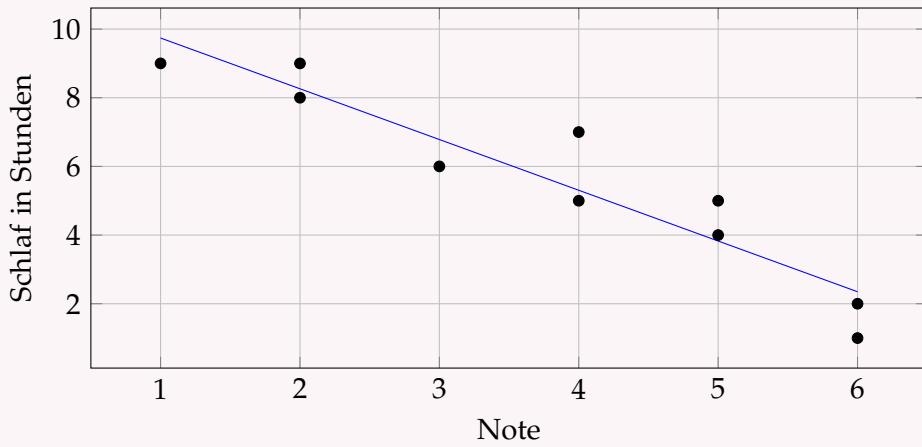
$$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{-68/15}{46/15} = -\frac{34}{23} \approx -1.478$$

$$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x} = \frac{28}{5} + \frac{34}{23} \cdot \frac{19}{5} = \frac{258}{23} \approx 11.217$$

Und damit erhalten wir:

$$\hat{y}_i = 11.217 - 1.478x_i$$

Beispiel: Lineare Regression, Regressionsgerade



3 Schließende Statistik

Bonus: Aufgabe der schließenden Statistik

Die *schließende Statistik* befasst sich mit dem Rückschluß von einer Stichprobe auf die Grundgesamtheit. Es muss eine repräsentative (d.h. nur zufallsbeeinflußte) Stichprobe aus der Grundgesamtheit gezogen werden.

Grundlage der schließenden Statistik ist die Wahrscheinlichkeitsrechnung.

3.1 Grundbegriffe

Definition: Grundgesamtheit

Unter einer **Grundgesamtheit** verstehen wir die Gesamtheit gleichartiger Objekte oder Elemente, die hinsichtlich eines bestimmten Merkmals untersucht werden sollen.

Das interessierende Merkmal beschreiben wir dabei durch eine Zufallsvariable X .

Definition: Stichprobe

Eine Stichprobe vom Umfang n der Zufallsvariablen X ist die Beobachtung von n unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n .

X_1, \dots, X_n nennt man *Stichprobenvariablen*, die beobachteten Werte x_1, \dots, x_n die *Stichprobenwerte*.

3.2 Punktschätzungen

Definition: Schätzfunktion

Eine Funktion $g(X_1, \dots, X_n)$ der Stichprobenvariablen heißt *Stichprobenfunktion* und ist wieder eine zufällige Variable.

Wird die Stichprobenfunktion zur Schätzung des Parameters θ verwendet, so heißt sie *Schätzfunktion* oder kurz *Schätzer* für θ und wird mit $\hat{\theta}$ bezeichnet.

Eine Schätzfunktion $\hat{\theta}$ für den Parameter θ heißt *erwartungstreu*, wenn gilt:

$$E(\hat{\theta}) = \theta$$

Sie heißt *konsistent*, wenn ihre Varianz mit wachsendem Stichprobenumfang n gegen 0 strebt:

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

Beispiel: Schätzung des Erwartungswertes

Das arithmetische Mittel ist eine *erwartungstreue* und *konsistente* Schätzfunktion des Erwartungswertes von X .

Es gilt (Erwartungstreue):

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

und (Konsistenz):^a

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

^aHier werden unter anderem der zentrale Grenzwertsatz und das starke Gesetz der großen Zahlen angewendet.

Beispiel: Schätzung der Varianz

Die Stichprobenvarianz ist eine *erwartungstreue* und *konsistente* Schätzfunktion der Varianz von X .

Es gilt (Erwartungstreue):

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right)^2 \\ &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2 - \frac{n}{n-1} (\bar{X} - \mu)^2\right) \\ &= \dots \\ &= \frac{1}{n-1} \cdot \sum_{i=1}^n E((X_i - \mu)^2) - \frac{n}{n-1} E((\bar{X} - \mu)^2) \\ &= \frac{n}{n-1} \cdot \sigma^2 - \frac{n}{n-1} \cdot \frac{\sigma^2}{n} \\ &= \frac{n-1}{n-1} \cdot \sigma^2 \\ &= \sigma^2 \end{aligned}$$

Auf einen Beweis der Konsistenz wird hier verzichtet.

Beispiel: Schätzfunktion (Erwartungstreue)

Die von einer Maschine für einen bestimmten Arbeitsvorgang benötigte Zeit sei eine Zufallsvariable X , für deren Dichtefunktion in Abhängigkeit von einem $\theta \in [0, 2]$ die Gestalt

$$f(x; \theta) = \begin{cases} \theta + 2(1 - \theta) \cdot x & \text{für } x \in [0, 1] \\ 0 & \text{sonst} \end{cases}$$

unterstellt wird. Zu X liege eine einfache Stichprobe X_1, \dots, X_n (die X_i sind unabhängig) vor.

a) Zeigen Sie, dass die Schätzfunktionen

- i) $\hat{\Theta}_1 = 4 - \frac{6}{n} \sum_{i=1}^n X_i$
- ii) $\hat{\Theta}_2 = 3 - \frac{6}{n} \sum_{i=1}^n X_i^2$

erwartungstreu für θ sind.

b) Überprüfen Sie zusätzlich, ob $\hat{\Theta}_1$ konsistent für θ ist.

a) i) Wir wissen, dass $\hat{\Theta}_1$ genau dann erwartungstreu ist, wenn $E(\hat{\Theta}_1) = \theta$ gilt.

Offensichtlich ist:

$$\begin{aligned} E(\hat{\Theta}_1) &= E\left(4 - \frac{6}{n} \sum_{i=1}^n X_i\right) \\ &= 4 - \frac{6}{n} \cdot \sum_{i=1}^n E(X_i) \\ &= 4 - 6E(X) \\ &= 4 - 6 \int_{-\infty}^{\infty} xf(x) dx \\ &= 4 - 6 \left(\int_0^1 x \cdot (\theta + 2(1 - \theta) \cdot x) dx \right) \\ &= \theta \end{aligned}$$

□

ii) Wir wissen, dass $\hat{\Theta}_2$ genau dann erwartungstreu ist, wenn $E(\hat{\Theta}_2) = \theta$ gilt.

Offensichtlich ist:

$$\begin{aligned} E(\hat{\Theta}_2) &= E\left(3 - \frac{6}{n} \sum_{i=1}^n X_i^2\right) \\ &= 3 - \frac{6}{n} E\left(\sum_{i=1}^n X_i^2\right) \\ &= 3 - 6E(X^2) \\ &= 3 - 6 \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= 3 - 6 \int_0^1 x^2 (\theta + 2(1 - \theta) \cdot x) dx \\ &= \theta \end{aligned}$$

□

Beispiel: Schätzfunktion (Konsistenz)

b) Wir wissen, dass $\hat{\Theta}_1$ genau dann konsistent ist, wenn $\lim_{n \rightarrow \infty} \text{Var}(\hat{\Theta}_1) = 0$ gilt.

Offensichtlich gilt:

$$\begin{aligned}\lim_{n \rightarrow \infty} \text{Var}(\hat{\Theta}_1) &= \lim_{n \rightarrow \infty} \text{Var}(\hat{\Theta}_1) \\&= \lim_{n \rightarrow \infty} \text{Var}\left(4 - \frac{6}{n} \sum_{i=1}^n X_i\right) \\&= \lim_{n \rightarrow \infty} \left(\frac{6}{n}\right)^2 \cdot \text{Var}\left(\sum_{i=1}^n X_i\right) \\&= 36 \cdot \lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot \sum_{i=1}^n \text{Var}(X_i) \\&= 36 \cdot \lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot n \cdot \text{Var}(X) \\&= 36 \cdot \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \text{Var}(X) \\&= 36 \cdot 0 \\&= 0\end{aligned}$$

□

Algorithmus: Maximum-Likelihood-Methode

Bei der *Maximum-Likelihood-Methode* wird von einer Zufallsvariablen X ausgegangen, deren Dichte- bzw. Wahrscheinlichkeitsfunktion f von einem unbekannten Parameter θ abhängig ist.

Liegt eine einfache Zufallsstichprobe mit n Realisierungen x_1, \dots, x_n von n unabhängig und identisch verteilten Zufallsvariablen X_1, \dots, X_n vor, so lässt sich die gemeinsame Dichtefunktion bzw. Wahrscheinlichkeitsfunktion wie folgt faktorisieren:

$$f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

Statt nun für einen festen Parameter θ die Dichte für beliebige Werte x_1, \dots, x_n auszuwerten, kann umgekehrt für beobachtete und somit feste Realisierungen x_1, \dots, x_n die gemeinsame Dichte als Funktion von θ interpretiert werden. Dies führt zur Likelihood-Funktion:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$$

Wird diese Funktion in Abhängigkeit von θ maximiert, also:^a

$$\frac{\partial L}{\partial \theta} = 0^b$$

so erhält man die Maximum-Likelihood-Schätzung für den unbekannten Parameter θ .

Es wird also der Wert von θ gesucht, bei dem die Stichprobenwerte die größte Dichte- bzw. Wahrscheinlichkeitsfunktion haben. Es ist naheliegend, einen Parameterwert θ als umso plausibler anzusehen, je höher die Likelihood.

Da das Ableiten bei Dichtefunktionen mit komplizierten Exponentenausdrücken sehr aufwändig werden kann, wird häufig die logarithmierte Likelihood-Funktion verwendet, da sie auf Grund der Monotonie des Logarithmus ihr Maximum an derselben Stelle wie die nichtlogarithmierte Dichtefunktion besitzt, jedoch einfacher zu berechnen ist:

$$L^*(\theta) = \log \prod_{i=1}^n f(x_i; \theta) = \sum_{i=1}^n \log f(x_i; \theta)$$

Die Maximum-Likelihood-Methode liefert nicht immer erwartungstreue Schätzer.

^aDies gilt meist. Manchmal ist L aber nicht differenzierbar.

^bBeachte, dass auch $\frac{\partial^2 L}{\partial \theta^2} < 0$ gelten muss!

Beispiel: Maximum-Likelihood-Methode

Aus Erfahrung sei bekannt, dass die Brenndauer einer Glühbirne einer bestimmten Sorte durch eine stetig verteilte Zufallsvariable X mit der Dichte

$$f(x) = \begin{cases} 2\theta \cdot x \cdot e^{-\theta x^2} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

beschrieben werden kann. Schätzen Sie das für diese Sorte passende θ aufgrund der folgenden 15 Brenndauern (in 1000 Stunden) mittels der Maximum-Likelihood-Methode:

1.5	1	2	1	1.5
1	2	0.5	2.5	1.5
1.5	1.5	1.5	2	1.5

Es gilt:

$$L(\theta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n 2\theta \cdot x_i \cdot e^{-\theta x_i^2} = (2\theta)^n \cdot e^{-\theta \sum_{i=1}^n x_i^2} \cdot \prod_{i=1}^n x_i$$

Damit erhalten wir $L^*(\theta)$ mit:

$$L^*(\theta) = \ln L(\theta) = n \cdot \ln 2\theta + \left(-\theta \cdot \sum_{i=1}^n x_i^2 \right) \cdot \ln e + \ln \prod_{i=1}^n x_i = n(\ln 2 + \ln \theta) - \theta \cdot \sum_{i=1}^n x_i^2 + \ln \prod_{i=1}^n x_i$$

$$\implies \frac{\partial L^*}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n x_i^2 \implies \frac{\partial L^*}{\partial \theta} = 0 \iff \theta = \frac{n}{\sum_{i=1}^n x_i^2}$$

Damit ist $\theta = \frac{n}{\sum_{i=1}^n x_i^2}$ ein potentieller Schätzer für θ .

Natürlich müssen wir überprüfen, ob es sich um ein Maximum handelt. Es gilt offensichtlich:

$$\frac{\partial^2 L^*(\theta)}{\partial \theta^2} = \frac{-n}{\theta^2} < 0 \quad \checkmark$$

Damit haben wir einen Schätzer und es gilt:

$$\theta = \frac{n}{\sum_{i=1}^n x_i^2} = \frac{15 \cdot 4}{149} = \frac{60}{149} \approx 0.403$$

□

3.3 Intervallschätzungen

Definition: Konfidenzintervall und Konfidenzniveau

Sei X_1, \dots, X_n eine Stichprobe zu einer Verteilung mit Parameter θ und $0 < \alpha < 1$.

Ist dann

$$I_n := [c_u, c_o]$$

ein Intervall, das von der Stichprobe abhängt und gilt

$$P(\theta \in I_n) \geq 1 - \alpha$$

dann nennt man I_n *Konfidenzintervall* für θ mit dem *Konfidenzniveau* $1 - \alpha$.^a

Typische Werte für α sind $\alpha = 0.05$, $\alpha = 0.01$ oder $\alpha = 0.001$.

Kann man ein Konfidenzintervall von oben und unten begrenzen, spricht man von einem *zweiseitigen Konfidenzintervall* mit

$$I_n = [c_u, c_o]$$

Kann man ein Konfidenzintervall nur von einer Seite begrenzen, spricht man von einem *einseitigen Konfidenzintervall* mit entweder

$$I_n = [c_u, \infty) \quad (\text{nach unten beschränkt})$$

oder

$$I_n = (-\infty, c_o] \quad (\text{nach oben beschränkt})$$

^aWenn ihr über Konfidenzintervalle und deren beschriebene Parameter redet, formuliert nicht den Satz: „Der Parameter liegt zu $n\%$ im Konfidenzintervall.“ Entweder liegt der Parameter im Intervall, oder nicht.

Bonus: Berechnung von Intervallen einer Standardnormalverteilung

Sei $0 < p < 1$ eine vorgegebene Wahrscheinlichkeit.

Dann sei u_p das zur Wahrscheinlichkeit p gehörige Quantil (obere Schranke) mit

$$\Phi(u_p) = p$$

Es gilt:

$$u_{1-p} = -u_p \iff u_p = -u_{1-p}$$

Der Wert von u_p kann dann aus folgender Tabelle abgelesen werden:

p	u_p	p	u_p
0.90	1.282	0.1	-1.282
0.95	1.645	0.05	-1.645
0.975	1.960	0.025	-1.960
0.99	2.326	0.01	-2.326
0.995	2.576	0.005	-2.576
0.999	3.090	0.001	-3.090

Für zu berechnende Intervalle gilt dann:

1. Einseitige Abgrenzung nach oben:

$$P(U \leq c) = \Phi(c) = p$$

$$\Phi(c) = p \implies c = u_p$$

2. Einseitige Abgrenzung nach unten:

$$P(U \geq c) = 1 - P(U \leq c) = 1 - \Phi(c) = 1 - p$$

$$\Phi(c) = 1 - p \implies c = u_{1-p}$$

3. Zweiseitige (symmetrische) Abgrenzung:

$$P(-c \leq U \leq c) = 2 \cdot \Phi(c) - 1 = p$$

$$\Phi(c) = \frac{1}{2}(1 + p) = u_{(1+p)/2}$$

Algorithmus: Konfidenzintervalle für den unbekannten Erwartungswert einer Normalverteilung bei bekannter Varianz

X sei eine normalverteilte Zufallsvariable mit *unbekanntem* Erwartungswert μ und *bekannter* Varianz σ^2 .

Wir wissen, dass das arithmetische Mittel ein Schätzer für den Erwartungswert μ ist:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Die normierte Zufallsvariable

$$U = \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma}$$

ist dann standardnormalverteilt.

Für U lässt sich dann schrittweise ein Konfidenzintervall konstruieren:

1. Wähle ein bestimmtes Konfidenzniveau $\gamma = 1 - \alpha$
2. Die Zufallsvariable U soll dann mit der gewählten Wahrscheinlichkeit γ einen Wert in dem *symmetrischen* Intervall $-c \leq U \leq c$ annehmen, also:

$$P(-c \leq U \leq c) = \gamma = 1 - \alpha$$

3. Mit dem vorgegebenen Konfidenzniveau γ können wir dann das Intervall wie folgt berechnen:

$$\begin{aligned}
 & -c \leq U \leq c \\
 \equiv & -u_\gamma \leq U \leq u_\gamma \\
 \equiv & -u_{(1+\gamma)/2} \leq U \leq u_{(1+\gamma)/2} \\
 \equiv & -u_{(1-\alpha)/2} \leq U \leq u_{(1-\alpha)/2} \\
 \equiv & -u_{1-\alpha/2} \leq U \leq u_{1-\alpha/2} \\
 \equiv & -u_{1-\alpha/2} \leq \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma} \leq u_{1-\alpha/2} \\
 \equiv & \bar{X} - u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\
 \implies & P\left(\bar{X} - u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = \gamma = 1 - \alpha
 \end{aligned}$$

4. Die Berechnung der Intervallgrenzen erfolgt dann anhand einer konkreten Stichprobe.
Das Konfidenzintervall

$$\bar{X} - u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

enthält den unbekannten Erwartungswert μ der normalverteilten Grundgesamtheit mit einem Konfidenzniveau von γ .

Beispiel: Konfidenzintervalle für den unbekannten Erwartungswert einer Normalverteilung bei bekannter Varianz

12 Versuchsflächen wurden mit einer neuen Weizensorte bestellt. Diese Flächen erbrachten folgende Hektarerträge (in Doppelzentner):

35.6 33.7 37.8 31.2 37.2 43.1 35.8 36.6 37.1 34.9 35.6 34.0

Aus Erfahrung weiß man, dass die Hektarerträge als eine Realisierung unabhängiger $\mathcal{N}(\mu, (\sqrt{3})^2)$ -verteilter Zufallsvariablen angesehen werden können.

Geben Sie für den Erwartungswert μ ein konkretes Konfidenzintervall zum Niveau 0.95 an.

Wir wissen, dass gilt:

$$X := \text{Hektarerträge (in Doppelzentner)} \sim \mathcal{N}(\mu, (\sqrt{3})^2)$$

Ein geeigneter Schätzer für μ ist bekanntermaßen

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

Für unsere Verteilung sind die relevanten Kennzahlen übrigens:

$$\bar{X} = \frac{1}{12} \cdot 432.6 = 36.05 \quad \wedge \quad \sigma^2 = 3 \implies \sigma = \sqrt{3}$$

Die normierte Zufallsvariable U mit

$$U = \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma}$$

ist dann standarnormalverteilt mit $\mathcal{N}(0, 1)$.

Es muss gelten:

$$P(-c \leq U \leq c) = \gamma = 1 - \alpha$$

Damit gilt:

$$\begin{aligned} P\left(\bar{X} - u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) &= \gamma \\ \equiv P\left(36.05 - 1.960 \cdot \frac{1}{2} \leq \mu \leq 36.05 + 1.960 \cdot \frac{1}{2}\right) &= 0.95 \\ \equiv P(35.07 \leq \mu \leq 37.03) &= 0.95 \end{aligned}$$

Damit erhalten wir unser Konfidenzintervall I_{12} mit

$$I_{12} = [35.07, 37.03]$$

□

Definition: Chi-Quadrat-Verteilung

X_1, \dots, X_n seien mit n mit $\mathcal{N}(0, 1)$ verteilte, stochastisch unabhängige Zufallsvariablen.

Betrachte die stetige Zufallsvariable Z mit Wertebereich $z \geq 0$:

$$Z = X_1^2 + \dots + X_n^2$$

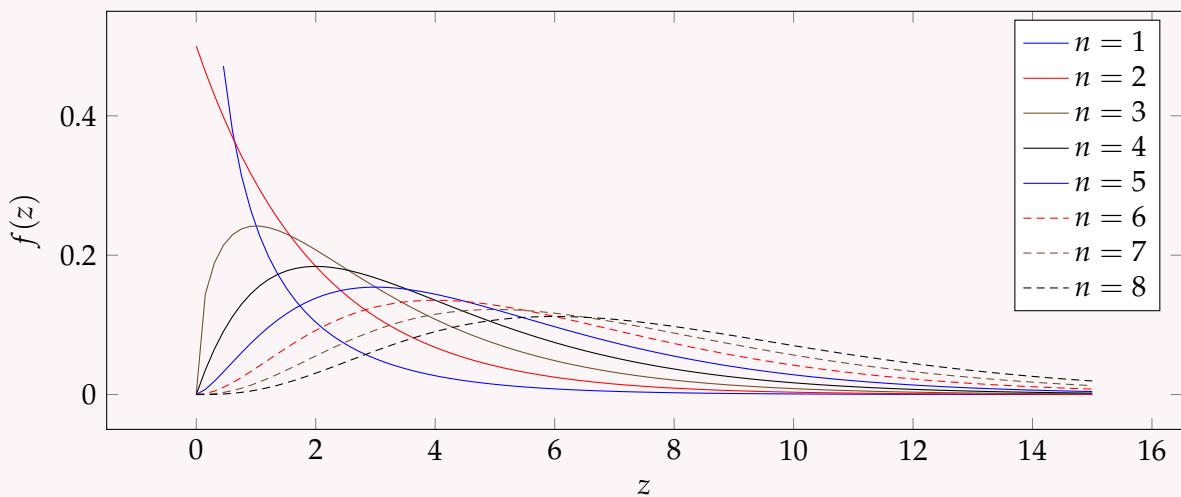
Z ist dann *Chi-Quadrat-verteilt* mit der Anzahl der Freiheitsgraden n , man schreibt $Z \sim \chi_n^2$.

Die Dichte- und Verteilungsfunktion der Chi-Quadrat-Verteilung sind an dieser Stelle nicht wirklich relevant. In den Teilen, in denen sie genutzt wird, muss lediglich der jeweilige Wert eines oder mehrerer $\chi_n^2(z)$ aus einer Tabelle abgelesen werden.

Eigenschaften der Chi-Quadrat-Verteilung:

- Die Dichtefunktion ist asymmetrisch.
- Die Dichtefunktion ist für $n \in \{1, 2\}$ streng monoton fallend.
- Die Dichtefunktion besitzt für $n > 2$ ein absolutes Maximum bei $z_{\max} = n - 2$.
- Für große Freiheitsgrade ($n > 1000$) lässt sich die Chi-Quadrat-Verteilung durch eine Normalverteilung $\mathcal{N}(n, 2n)$ annähern.

Beispiel: Chi-Quadrat-Verteilung



Bonus: Anzahl der Freiheitsgrade

Schätzungen statistischer Parameter können auf unterschiedlichen Mengen an Informationen oder Daten basieren. Die Anzahl unabhängiger Information, die in die Schätzung eines Parameters einfließen, wird als *Anzahl der Freiheitsgrade* bezeichnet.

Im Allgemeinen sind die Freiheitsgrade einer Schätzung eines Parameters gleich der Anzahl unabhängiger Einzelinformationen, die in die Schätzung einfließen, abzüglich der Anzahl der zu schätzenden Parameter, die als Zwischenschritte bei der Schätzung des Parameters selbst verwendet werden.

Beispielsweise fließen in die Berechnung der Stichprobenvarianz n Werte mit ein. Dennoch lautet die Anzahl der Freiheitsgrade $n - 1$, da als Zwischenschritt der Mittelwert geschätzt wird und somit ein Freiheitsgrad verloren geht.

Algorithmus: Konfidenzintervalle für die unbekannte Varianz einer Normalverteilung

X sei eine normalverteilte Zufallsvariable mit *unbekanntem* Erwartungswert μ und *unbekannter* Varianz σ^2 .

Wir wissen, dass das Stichprobenmittel ein Schätzer für die Varianz σ^2 ist:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Die Variable

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

ist dann Chi-Quadrat-verteilt mit $f = n - 1$ Freiheitsgraden.^a

Für Z lässt sich dann schrittweise ein Konfidenzintervall konstruieren:

1. Wähle ein bestimmtes Konfidenzniveau $\gamma = 1 - \alpha$
2. Die Zufallsvariable Z soll dann mit der gewählten Wahrscheinlichkeit γ einen Wert in dem Intervall $c_1 \leq Z \leq c_2$ annehmen, also:

$$P(c_1 \leq Z \leq c_2) = \gamma = 1 - \alpha$$

3. Mit dem vorgegebenen Konfidenzniveau γ können wir dann das Intervall wie folgt berechnen:

$$\begin{aligned} & c_1 \leq Z \leq c_2 \\ \equiv & \chi_{n-1}^2(\alpha/2) \leq Z \leq \chi_{n-1}^2(1-\alpha/2) \\ \equiv & \chi_{n-1}^2(\alpha/2) \leq (n-1) \cdot \frac{S^2}{\sigma^2} \leq \chi_{n-1}^2(1-\alpha/2) \\ \equiv & (n-1) \cdot \frac{S^2}{\chi_{n-1}^2(1-\alpha/2)} \leq \sigma^2 \leq (n-1) \cdot \frac{S^2}{\chi_{n-1}^2(\alpha/2)} \\ \implies & P \left((n-1) \cdot \frac{S^2}{\chi_{n-1}^2(1-\alpha/2)} \leq \sigma^2 \leq (n-1) \cdot \frac{S^2}{\chi_{n-1}^2(\alpha/2)} \right) = \gamma = 1 - \alpha \end{aligned}$$

4. Die Berechnung der Intervallgrenzen erfolgt dann anhand einer konkreten Stichprobe. Das Konfidenzintervall

$$(n-1) \cdot \frac{S^2}{\chi_{n-1}^2(1-\alpha/2)} \leq \sigma^2 \leq (n-1) \cdot \frac{S^2}{\chi_{n-1}^2(\alpha/2)}$$

enthält die unbekannte Varianz σ^2 der normalverteilten Grundgesamtheit mit einem Konfidenzniveau γ .

Durch Radizieren erhält man ein entsprechendes Konfidenzintervall für σ .

^a $n - 1$ Freiheitsgrade statt n , da hier μ über \bar{X} der Stichprobe geschätzt werden muss.

Beispiel: Konfidenzintervalle für die unbekannte Varianz einer Normalverteilung

Das Gewicht X , das ein Apfel einer bestimmten Sorte hat, sei normalverteilt. Die Untersuchung einer Stichprobe vom Umfang $n = 10$ ergab einen Mittelwert $\bar{x} = 98\text{g}$ und eine empirische Standardabweichung $s = 0.75\text{g}$. Geben Sie den Bereich an, in dem die Varianz mit 95%-iger Sicherheit liegt.

Wir wissen, dass gilt:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Ein geeigneter Schätzer für $\text{Var}(X) = \sigma^2$ ist bekanntermaßen

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

Für unsere Verteilung gilt dann:

$$s^2 = 0.75^2 = 0.5625$$

Die Zufallsvariable Z mit

$$Z = (n-1) \cdot \frac{s^2}{\sigma^2} = \frac{1}{\sigma^2} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2$$

ist dann χ^2 -verteilt mit $n-1$ Freiheitsgraden.

Es muss gelten:

$$P(c_1 \leq Z \leq c_2) = \gamma = 1 - \alpha$$

Damit gilt:

$$\begin{aligned} P\left((n-1) \cdot \frac{s^2}{\chi_{n-1}^2(1-\frac{\alpha}{2})} \leq \sigma^2 \leq (n-1) \cdot \frac{s^2}{\chi_{n-1}^2(\frac{\alpha}{2})}\right) &= \gamma \\ \equiv P\left(9 \cdot \frac{0.5625}{\chi_9^2(0.975)} \leq \sigma^2 \leq 9 \cdot \frac{0.5625}{\chi_9^2(0.025)}\right) &= 0.95 \\ \equiv P\left(9 \cdot \frac{0.5625}{19.02} \leq \sigma^2 \leq 9 \cdot \frac{0.5625}{2.70}\right) &= 0.95 \\ \equiv P(0.2662 \leq \sigma^2 \leq 1.875) &= 0.95 \end{aligned}$$

Damit erhalten wir unser Konfidenzintervall I_{10} mit

$$I_{10} = [0.2662, 1.875]$$

□

Definition: t-Verteilung

X und Y seien zwei stochastisch unabhängige Zufallsvariablen mit

$$X \sim \mathcal{N}(0, 1) \quad \wedge \quad Y \sim \chi_n^2$$

Dann ist die stetige Zufallsvariable T mit

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

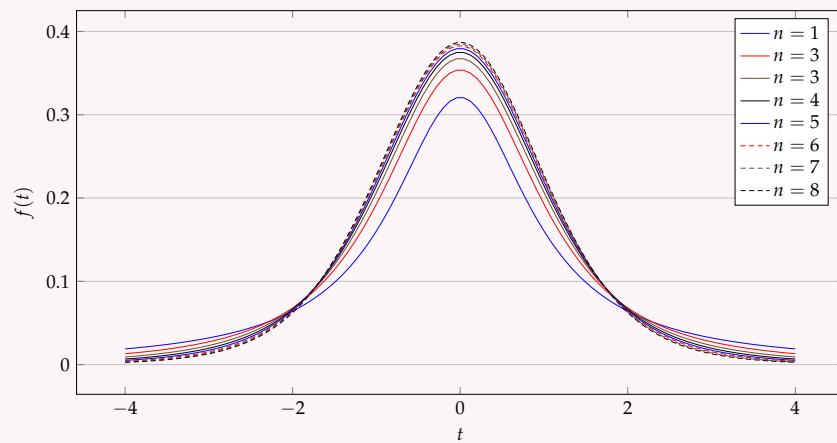
t-verteilt mit n Freiheitsgraden, man schreibt $T \sim t_n$.

Die Dichte- und Verteilungsfunktion der t-Verteilung sind an dieser Stelle nicht wirklich relevant. In den Teilen, in denen sie genutzt wird, muss lediglich der jeweilige Wert eines oder mehrerer $t_n(t)$ aus einer Tabelle abgelesen werden.

Eigenschaften der t-Verteilung:

- Die Dichtefunktion ist eine gerade Funktion (achsensymmetrisch).
- Die Dichtefunktion besitzt an der Stelle $t_{\max} = 0$ ein absolutes Maximum.
- Die Dichtefunktion nähert sich für $t \rightarrow \pm\infty$ asymptotisch der t -Achse.
- Für große Freiheitsgrade ($n > 30$) lässt sich die t-Verteilung durch die Standardnormalverteilung $\mathcal{N}(0, 1)$ annähern.

Beispiel: t-Verteilung



Algorithmus: Konfidenzintervalle für den unbekannten Erwartungswert einer Normalverteilung bei unbekannter Varianz

X sei eine normalverteilte Zufallsvariable mit *unbekanntem* Erwartungswert μ und *unbekannter* Varianz σ^2 .

Wir wissen, dass das arithmetische Mittel ein Schätzer für den Erwartungswert μ ist:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Wir wissen auch, dass die Wurzel des Stichprobenmittels ein Schätzer für die Standardabweichung σ ist:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Die normierte Variable

$$T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$$

ist dann t-verteilt mit $f = n - 1$ Freiheitsgraden.

Für T lässt sich dann schrittweise ein Konfidenzintervall konstruieren:

1. Wähle ein bestimmtes Konfidenzniveau $\gamma = 1 - \alpha$
2. Die Zufallsvariable T soll dann mit der gewählten Wahrscheinlichkeit γ einen Wert in dem *symmetrischen* $-c \leq U \leq c$ annehmen, also:

$$P(-c \leq T \leq c) = \gamma = 1 - \alpha$$

3. Mit dem vorgegebenen Konfidenzniveau γ können wir dann das Intervall wie folgt berechnen:

$$\begin{aligned} & -c \leq T \leq c \\ \equiv & -t_{n-1}(1 - \alpha/2) \leq T \leq t_{n-1}(1 - \alpha/2) \\ \equiv & -t_{n-1}(1 - \alpha/2) \leq \sqrt{n} \cdot \frac{\bar{X} - \mu}{S} \leq t_{n-1}(1 - \alpha/2) \\ \equiv & \bar{X} - t_{n-1}(1 - \alpha/2) \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1}(1 - \alpha/2) \cdot \frac{S}{\sqrt{n}} \\ \implies & P\left(\bar{X} - t_{n-1}(1 - \alpha/2) \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1}(1 - \alpha/2) \cdot \frac{S}{\sqrt{n}}\right) = \gamma = 1 - \alpha \end{aligned}$$

4. Die Berechnung der Intervallgrenzen erfolgt dann anhand einer konkreten Stichprobe. Das Konfidenzintervall

$$\bar{X} - t_{n-1}(1 - \alpha/2) \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1}(1 - \alpha/2) \cdot \frac{S}{\sqrt{n}}$$

enthält den unbekannten Erwartungswert γ der normalverteilten Grundgesamtheit mit einem Konfidenzniveau von γ .

Beispiel: Konfidenzintervalle für den unbekannten Erwartungswert einer Normalverteilung bei unbekannter Varianz

Das Umweltreferat einer Großstadt will Aufschluss darüber gewinnen, wie viele Asbestfasern pro Kubikmeter Luft im Freien in ca. einem Meter Abstand von asbestzementhaltigen Gebäudeteilen zu erwarten sind. Bei $n = 14$ diesbezüglichen Messungen traten die Werte

$$\begin{array}{ccccccc} 980 & 1340 & 610 & 750 & 880 & 1250 & 2410 \\ 1100 & 470 & 1040 & 910 & 1860 & 730 & 820 \end{array}$$

auf, die als Ergebnisse unabhängiger normalverteilter Stichprobenvariablen angesehen werden.

Führen Sie für den Erwartungswert μ der Anzahl X der unter den obigen Bedingungen vorhandenen Asbestfasern eine Intervallschätzung zum Konfidenzniveau 0.95 durch.

Wir wissen, dass gilt:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Geeignete Schätzer für μ bzw. σ sind bekanntermaßen

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad \text{bzw.} \quad S = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

Für unsere Verteilung gilt dann:

$$\bar{X} = \frac{1}{14} \cdot 15150 = \frac{7575}{7} \quad \text{bzw.} \quad S = \sqrt{\frac{1}{13} \cdot \frac{24126450}{7}} = \frac{5\sqrt{87820278}}{91} \approx 514.9$$

Die Zufallsvariable T mit

$$T = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$$

ist dann t -verteilt mit $n - 1$ Freiheitsgraden.

Es muss gelten:

$$P(-c \leq T \leq c) = \gamma = 1 - \alpha$$

Damit gilt:

$$\begin{aligned} & P\left(\bar{X} - t_{n-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1}\left(-\frac{\alpha}{2}\right) \cdot \frac{S}{\sqrt{n}}\right) = \gamma \\ & \equiv P\left(\frac{7575}{7} - t_{13}(0.975) \cdot \frac{514.9}{\sqrt{14}} \leq \mu \leq \frac{7575}{7} + t_{13}(0.975) \cdot \frac{514.9}{\sqrt{14}}\right) = 0.95 \\ & \equiv P\left(\frac{7575}{7} - 2.160 \cdot \frac{514.9}{\sqrt{14}} \leq \mu \leq \frac{7575}{7} + 2.160 \cdot \frac{514.9}{\sqrt{14}}\right) = 0.95 \\ & \equiv P(784.897 \leq \mu \leq 1379.389) = 0.95 \end{aligned}$$

Damit erhalten wir unser Konfidenzintervall I_{14} mit

$$I_{14} = [784.897, 1379.389]$$

□

Algorithmus: Konfidenzintervalle für den unbekannten Anteilswert p (Parameter p einer Binomialverteilung)

Sei $p = P(A)$ die Wahrscheinlichkeit für ein Ereignis A in einem Bernoulli-Experiment.

Dann ist

$$X := \text{Anzahl des Eintretens von } A$$

bei einer n -fachen Ausführung des Bernoulli-Experiments eine *binomialverteilte Zufallsvariable*, die bei umfangreichen Stichproben näherungsweise normalverteilt ist^a mit

$$\mathbb{E}(X) = \mu = np \quad \wedge \quad \text{Var}(X) = \sigma^2 = np(1-p)$$

Eine erwartungstreue Schätzfunktion für p ist:

$$\hat{p} = \frac{X}{n} \implies X = n \cdot \hat{p}$$

Die Zufallsvariable

$$U = \frac{X - \mu}{\sqrt{\text{Var}(\hat{p})}} = \frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} = \frac{n\hat{p} - np}{\sqrt{n\hat{p}(1-\hat{p})}}$$

ist dann näherungsweise standardnormalverteilt.

Für p lässt sich dann schrittweise ein Konfidenzintervall konstruieren:

1. Wähle ein bestimmtes Konfidenzniveau $\gamma = 1 - \alpha$
2. Die Zufallsvariable U soll dann mit der gewählten Wahrscheinlichkeit γ einen Wert in dem *symmetrischen* $-c \leq U \leq c$ annehmen, also:

$$P(-c \leq U \leq c) = \gamma = 1 - \alpha$$

3. Mit dem vorgegebenen Konfidenzniveau γ können wir dann das Intervall wie folgt berechnen:

$$\begin{aligned} & -c \leq U \leq c \\ & \equiv \dots \\ & \equiv \hat{p} - \frac{u_{1-\alpha/2}}{n} \sqrt{n\hat{p}(1-\hat{p})} \leq p \leq \hat{p} + \frac{u_{1-\alpha/2}}{n} \sqrt{n\hat{p}(1-\hat{p})} \\ \implies & P\left(\hat{p} - \frac{u_{1-\alpha/2}}{n} \sqrt{n\hat{p}(1-\hat{p})} \leq p \leq \hat{p} + \frac{u_{1-\alpha/2}}{n} \sqrt{n\hat{p}(1-\hat{p})}\right) = \gamma = 1 - \alpha \end{aligned}$$

4. Die Berechnung der Intervallgrenzen erfolgt dann anhand einer konkreten Stichprobe.^b
Das Konfidenzintervall

$$\hat{p} - \frac{u_{1-\alpha/2}}{n} \sqrt{n\hat{p}(1-\hat{p})} \leq p \leq \hat{p} + \frac{u_{1-\alpha/2}}{n} \sqrt{n\hat{p}(1-\hat{p})}$$

enthält den unbekannten Anteilswert p der binomialverteilten Grundgesamtheit mit einem Konfidenzniveau von γ .

^aSiehe Grenzwertsatz von Moivre und Laplace

^bIn dieser gilt dann $\hat{p} = k/n$, wobei k die Anzahl der eingetretenen Elemente mit Eigenschaft A darstellt.

3.4 Statistische Testverfahren

Definition: Hypothesentest

Ein *Hypothesentest* dient dazu, anhand vorliegender Beobachtungen eine begründete Entscheidung über die Gültigkeit oder Ungültigkeit einer Hypothese zu treffen.

Da die vorhandenen Daten Realisierungen von Zufallsvariablen sind, lässt sich meist nicht mit Sicherheit sagen, ob eine Hypothese stimmt oder nicht. Man versucht daher, die Wahrscheinlichkeiten für Fehlentscheidungen zu kontrollieren (Signifikanzniveau).

Algorithmus: Planung und Durchführung eines Parametertests

Die Wahrscheinlichkeitsverteilung einer Zufallsvariablen X sei von der Art her bekannt, enthalte jedoch noch einen oder mehrere unbekannte Parameter.

1. Formulieren der *Nullhypothese* H_0 und *Alternativhypothese* H_1 :

- Einseitiger Test:
 - Nullhypothese: $H_0 : \theta \leq \theta_0$ bzw. $H_0 : \theta \geq \theta_0$
 - einseitige Alternativhypothese: $H_1 : \theta > \theta_0$ bzw. $H_1 : \theta < \theta_0$
- Zweiseitiger Test:
 - Nullhypothese: $H_0 : \theta = \theta_0$
 - zweiseitige Alternativhypothese: $H_1 : \theta \neq \theta_0$

2. Festlegung des Signifikanzniveaus α .^a (typisch: $\alpha = 0.05$, $\alpha = 0.01$, $\alpha = 0.001$)

3. Bestimmung einer Test- bzw. Prüfvariablen $T = g(X_1, X_2, \dots, X_n)$ ^b

4. Auf Basis des Signifikanzniveaus α werden kritische Grenzen berechnet, so dass Testvariable T mit Wahrscheinlichkeit $\gamma = 1 - \alpha$ Werte aus jeweiligem Intervall annimmt:

- Einseitiger Test:

$$P(c_u \leq T \leq c_o) = \gamma = 1 - \alpha$$

- Zweiseitiger Test:

$$P(T \leq c) = \gamma = 1 - \alpha \quad \text{bzw.} \quad P(T \geq c) = \gamma = 1 - \alpha$$

5. Berechnung eines Test- bzw. Prüfwertes von T aus einer konkreten Stichprobe:

$$x_1, \dots, x_n \implies \hat{T} = g(x_1, \dots, x_n)$$

6. Testentscheidung anhand des Wertes von \hat{T} :

- \hat{T} fällt in den nicht-kritischen Bereich der Testvariablen T (Bereich von H_0):
 - H_0 wird nicht verworfen
 - Unterschied zwischen Schätzwert \hat{T} und angenommenen Wert θ_0 ist zufallsbedingt.
- \hat{T} fällt in den kritischen Bereich der Testvariablen T (Bereich von H_1):
 - H_0 wird zugunsten der Alternativhypothese H_1 verworfen
 - Schätzwert \hat{T} weicht signifikant von dem angenommenen Wert θ_0 ab.

Faustregel:

- H_0 enthält immer das Gleichheitszeichen
- H_0 nicht verwerbar, wenn durch Stichprobe gestützt

^a α gibt die Wahrscheinlichkeit an, die Nullhypothese abzulehnen, obwohl sie zutrifft. Siehe Fehler 1. Art.

^b Stichprobenfunktion, deren Wahrscheinlichkeitsverteilung als bekannt vorausgesetzt werden muss.

Definition: Fehler 1. Art

Eine an sich *richtige Nullhypothese* H_0 wird *verworfen*.

Die Wahrscheinlichkeit für einen Fehler 1. Art wird mit α bezeichnet.

Über die Entscheidung für ein Signifikanzniveau α wird also auch die Wahrscheinlichkeit für das Auftreten des Fehlers 1. Art festgelegt.

Definition: Fehler 2. Art

Eine an sich *falsche Nullhypothese* H_0 wird *nicht verworfen*.

Die Wahrscheinlichkeit für einen Fehler 2. Art wird mit β bezeichnet.

Im Gegensatz zum Fehler 1. Art lässt sich die Wahrscheinlichkeit eines Fehler 2. Art meist nicht vorab bestimmen.^a

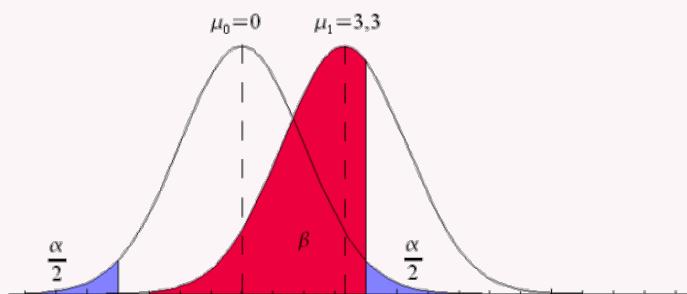
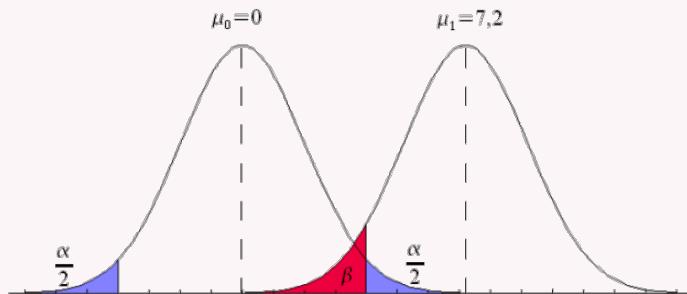
Zur Berechnung von β ist die tatsächliche Verteilung bzw. der tatsächliche Parameter θ_1 nötig.

^aGrund dafür ist, dass die Alternativhypothese im Gegensatz zur Nullhypothese i.d.R. sehr unbestimmter bzw. globaler Natur ist.

Beispiel: Fehler 1. und 2. Art

Es seien zwei Hypothesentests gegeben:

- μ_0 sei zu bestimmen
- H_0 behauptet in beiden Fällen, dass μ_0 gleich 0 ist
- Der tatsächliche Wert μ_1 sei im zweiten Test „näher“ am behaupteten Wert μ_0
- *Man erkennt:* Je kleiner die Abweichung des tatsächlichen vom behaupteten Wert μ_0 , desto größer paradoxerweise die Wahrscheinlichkeit, einen Fehler zu machen, wenn man aufgrund des Testergebnisses weiterhin dem behaupteten Wert μ_0 Glauben schenkt.



Definition: Operationscharakteristik und Gütfunktion

Die *Operationscharakteristik* ist ein Konzept, mit dem ein funktionaler Zusammenhang zwischen der Wahrscheinlichkeit eines Fehlers 2. Art und der tatsächlichen Lage des unbekannten Parameters θ einer Verteilungsfunktion $F(x | \theta)$ hergestellt wird.

Sei θ_1 der *tatsächliche* Parameter. Dann berechnet sich der Fehler 2. Art als Wahrscheinlichkeit, dass eine Testvariable \hat{t} in den nicht-kritischen Bereich der Nullhypothese fällt, obwohl in Wahrheit θ_1 statt θ_0 die Verteilung bestimmt:

$$\beta = P(\hat{t} \in \text{nicht-kritischer Bereich von } H_0 \mid \theta_1)$$

β hängt also von θ_1 ab und kann daher als Funktion von θ_1 definiert werden:

$$\beta = f(\theta_1) \quad \text{bzw.} \quad \beta = \beta(\theta_1)$$

Diese Funktion wird als *Operationscharakteristik*, bezeichnet.

Die Gegenwahrscheinlichkeit zu β ist, dass H_0 abgelehnt und dafür H_1 akzeptiert wird, wenn θ_1 der wahre Parameter ist.

Hier ist die Ablehnung von H_0 zu Gunsten von H_1 also erwünscht, weshalb die entsprechende Funktion

$$\gamma(\theta_1) = 1 - \beta(\theta_1)$$

auch *Gütfunktion* genannt wird.

Algorithmus: Bestimmung der Grenze bei einem einfachen Alternativtest

- Nullhypothese $H_0 : p = p_0$, Alternativhypothese $H_1 : p = p_1^a$
- Stichprobe des Umfangs n
- Ergebnis: N Elemente mit relevanter Eigenschaft (z.B. defekte Schrauben in der Stichprobe)

Gesucht ist eine Signifikanzschranke bzw. Grenze k , mit der wir uns bei einem Ergebnis N für H_0 , bzw. H_1 bei einer Stichprobe von n Elementen entscheiden können:

$$N > k \implies \text{Entscheidung für } H_1 \quad \text{bzw.} \quad N \leq k \implies \text{Entscheidung für } H_0$$

Für die Wahrscheinlichkeit für einen Fehler 1. Art gilt (z.B. bei einer Binomialverteilung):

$$\alpha = \alpha(k) = P(N > k \mid p = p_0) = \sum_{l=k+1}^n \binom{n}{l} p_0^l (1-p_0)^{n-l}$$

Für die Wahrscheinlichkeit für einen Fehler 2. Art gilt (z.B. bei einer Binomialverteilung):

$$\beta = \beta(k) = P(N \leq k \mid p = p_1) = \sum_{l=0}^k \binom{n}{l} p_1^l (1-p_1)^{n-l}$$

^aWir nehmen an, $p_1 > p_0$

Algorithmus: Bestimmung der Grenze bei einem einseitigen Hypothesentest

- Nullhypothese $H_0 : p \leq p_0$, Alternativhypothese $H_1 : p > p_0$
- Stichprobe des Umfangs n
- Ergebnis: N Elemente mit relevanter Eigenschaft (z.B. defekte Schrauben in der Stichprobe)

Gesucht ist eine Signifikanzschranke bzw. Grenze k , mit der wir uns bei einem Ergebnis N für H_0 , bzw. H_1 bei einer Stichprobe von n Elementen entscheiden können:

$$N > k \implies \text{Entscheidung für } H_1 \quad \text{bzw.} \quad N \leq k \implies \text{Entscheidung für } H_0$$

Um möglichst selten eine falsche Entscheidung zu treffen, legt man eine obere Schranke α_0 für die Fehlerwahrscheinlichkeit 1. Art fest:

$$\alpha \leq \alpha_0$$

Für die Wahrscheinlichkeit für einen Fehler 1. Art gilt (z.B. bei einer Binomialverteilung):

$$\alpha = \alpha(p) = P(N > k \mid p) = \sum_{l=k+1}^n \binom{n}{l} p^l (1-p)^{n-l} \leq \alpha_0 \quad \forall p \leq p_0$$

$\alpha(p)$ ist streng monoton wachsend in p . Damit gilt:

$$\alpha(p) \leq \alpha(p_0) \quad \forall p \leq p_0$$

Das gesuchte k ist dann das kleinste k , das die folgende Ungleichung erfüllt:

$$\sum_{l=k+1}^n \binom{n}{l} p_0^l (1-p_0)^{n-l} \leq \alpha_0$$

Algorithmus: Bestimmung der Grenzen bei einem zweiseitigen Hypothesentest

- Nullhypothese $H_0 : p = p_0$, Alternativhypothese $H_1 : p \neq p_0$
- Stichprobe des Umfangs n
- Ergebnis: N Elemente mit relevanter Eigenschaft (z.B. defekte Schrauben in der Stichprobe)

Gesucht sind eine Signifikanzschranken bzw. Grenzen k_1 und k_2 , mit denen wir uns bei einem Ergebnis N für H_0 , bzw. H_1 bei einer Stichprobe von n Elementen entscheiden können:

$$N < k_1 \vee N > k_2 \implies \text{Entscheidung f\"ur } H_1 \quad \text{bzw.} \quad k_1 \leq N \leq k_2 \implies \text{Entscheidung f\"ur } H_0$$

Um m\"oglichst selten eine falsche Entscheidung zu treffen, legt man eine obere Schranke α_0 f\"ur die Fehlerwahrscheinlichkeit 1. Art fest:

$$\alpha \leq \alpha_0$$

F\"ur die Wahrscheinlichkeit f\"ur einen Fehler 1. Art gilt (z.B. bei einer Binomialverteilung):

$$\begin{aligned} \alpha = \alpha(p) &= P(N < k_1 \vee N > k_2 \mid p = p_0) \\ &= \sum_{l=0}^{k_1-1} \binom{n}{l} p_0^l (1-p_0)^{n-l} + \sum_{l=k_2+1}^n \binom{n}{l} p_0^l (1-p_0)^{n-l} \leq \alpha_0 \end{aligned}$$

Wir m\"ussen die Wahrscheinlichkeiten auf beide Seiten symmetrisch verteilen und erhalten:

$$\sum_{l=0}^{k_1-1} \binom{n}{l} p_0^l (1-p_0)^{n-l} \leq \frac{\alpha}{2} \quad \text{mit gr\"o\3sten } k_1$$

$$\sum_{l=k_2+1}^n \binom{n}{l} p_0^l (1-p_0)^{n-l} \leq \frac{\alpha}{2} \quad \text{mit kleinsten } k_2$$

Algorithmus: Zweiseitiger Einstichproben-Gauß-Test

- Normalverteilung mit unbekanntem Erwartungswert μ und bekannter Varianz σ^2
- Nullhypothese $H_0 : \mu = \mu_0$, Alternativhypothese $H_1 : \mu \neq \mu_0$
- Test:

$$|\bar{X} - \mu_0| \begin{cases} > c & \Rightarrow \text{Entscheidung für } H_1 \\ \leq c & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

1. Vorgabe des Signifikanzniveaus α (Wahrscheinlichkeit für Fehler 1. Art)
2. Berechnung der kritischen Grenze c unter der Annahme, dass H_0 gültig ist:

$$\begin{aligned} \alpha &= P(\bar{X} - \mu_0 > c) + P(\bar{X} - \mu_0 < -c) \\ &= P\left(\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma} > \sqrt{n} \cdot \frac{c}{\sigma}\right) + P\left(\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma} < -\sqrt{n} \cdot \frac{c}{\sigma}\right) \end{aligned}$$

Die Variable

$$u = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma}$$

ist standardnormalverteilt. Es gilt weiterhin:

$$\begin{aligned} \alpha &= 1 - \Phi\left(\sqrt{n} \cdot \frac{c}{\sigma}\right) + \Phi\left(\sqrt{n} \cdot \frac{-c}{\sigma}\right) = 2 - 2 \cdot \Phi\left(\sqrt{n} \cdot \frac{c}{\sigma}\right) \\ \Rightarrow \Phi\left(\sqrt{n} \cdot \frac{c}{\sigma}\right) &= 1 - \frac{\alpha}{2} \Rightarrow \sqrt{n} \cdot \frac{c}{\sigma} = u_{1-\alpha/2} \Rightarrow c = u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \end{aligned}$$

3. Testentscheidung:

$$|\bar{X} - \mu_0| \begin{cases} > u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} & \Rightarrow \text{Entscheidung für } H_1 \\ \leq u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

4. Gütfunktion:

$$\begin{aligned} \gamma(\mu) &= P(\bar{X} - \mu_0 > c) + P(\bar{X} - \mu_0 < -c) = P(\bar{X} > c + \mu_0) + P(\bar{X} < -c + \mu_0) \\ &= P\left(\sqrt{n} \cdot \frac{|\bar{X}| - \mu}{\sigma} > \sqrt{n} \cdot \frac{c + \mu_0 - \mu}{\sigma}\right) + P\left(\sqrt{n} \cdot \frac{|\bar{X}| - \mu}{\sigma} < \sqrt{n} \cdot \frac{-c + \mu_0 - \mu}{\sigma}\right) \\ &= 1 - \Phi\left(\sqrt{n} \cdot \frac{c + \mu_0 - \mu}{\sigma}\right) + \Phi\left(\sqrt{n} \cdot \frac{-c + \mu_0 - \mu}{\sigma}\right) \\ &= 1 - \Phi\left(\sqrt{n} \cdot \frac{c}{\sigma} + \frac{\sqrt{n}}{\sigma} \cdot (\mu_0 - \mu)\right) + \Phi\left(\sqrt{n} \cdot \frac{-c}{\sigma} + \frac{\sqrt{n}}{\sigma} \cdot (\mu_0 - \mu)\right) \\ &= 1 - \Phi\left(u_{1-\alpha/2} + \frac{\sqrt{n}}{\sigma} \cdot (\mu_0 - \mu)\right) + \Phi\left(-u_{1-\alpha/2} + \frac{\sqrt{n}}{\sigma} \cdot (\mu_0 - \mu)\right) \end{aligned}$$

Definition: Konsistenz eines Tests

Bei ausreichend großen Stichprobenumfängen wird H_0 immer abgelehnt, falls dies berechtigt ist. Mit wachsendem Stichprobenumfang wächst also die Wahrscheinlichkeit, einen Fehler 2. Art zu vermeiden. Der Parametertest heißt *konsistent*.

Algorithmus: Berechnung des Mindeststichprobenumfangs

- Signifikanzniveau α
- Wert $\mu = \mu_1$

Gesucht sei der Mindeststichprobenumfang n_0 so, dass damit die Wahrscheinlichkeit für den Fehler 2. Art höchstens

$$\beta_0 = \beta(\mu_1)$$

ist. Typisch sind z.B. $\alpha = 0.01$ und $\beta_0 = 0.1$.

Es gilt für einen *zweiseitigen Test*:

$$n_0 = (u_{1-\alpha/2} + u_{1-\beta_0})^2 \cdot \frac{\sigma^2}{(\mu_1 - \mu_0)^2}$$

Bei den typischen Werten $\alpha = 0.01$ und $\beta_0 = 0.1$ gilt vereinfacht:

$$n = 14.88 \cdot \frac{\sigma^2}{(\mu_1 - \mu_0)^2} \approx 15 \cdot \frac{\sigma^2}{(\mu_1 - \mu_0)^2}$$

Es gilt für einen *einseitigen Test*:

$$n_0 = (u_{1-\alpha} + u_{1-\beta_0})^2 \cdot \frac{\sigma^2}{(\mu_1 - \mu_0)^2}$$

Bei den typischen Werten $\alpha = 0.01$ und $\beta_0 = 0.1$ gilt vereinfacht:

$$n = 13.02 \cdot \frac{\sigma^2}{(\mu_1 - \mu_0)^2} \approx 15 \cdot \frac{\sigma^2}{(\mu_1 - \mu_0)^2}$$

Algorithmus: Einseitiger Einstichproben-Gauß-Test (Abgrenzung nach oben)

- Normalverteilung mit unbekanntem Erwartungswert μ und bekannter Varianz σ^2
- Nullhypothese $H_0 : \mu \leq \mu_0$, Alternativhypothese $H_1 : \mu > \mu_0$

$$|\bar{X} - \mu_0| \begin{cases} > c & \Rightarrow \text{Entscheidung für } H_1 \\ \leq c & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

1. Vorgabe des Signifikanzniveaus α (Wahrscheinlichkeit für Fehler 1. Art)
2. Berechnung der kritischen Grenze c unter der Annahme, dass H_0 gültig ist:

$$1 - \alpha = P(\bar{X} - \mu_0 \leq c) = P\left(\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma} \leq \sqrt{n} \cdot \frac{c}{\sigma}\right)$$

Die Variable

$$u = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma}$$

ist standardnormalverteilt. Es gilt weiterhin:

$$1 - \alpha = \Phi\left(\sqrt{n} \cdot \frac{c}{\sigma}\right) \quad \Rightarrow \quad \sqrt{n} \cdot \frac{c}{\sigma} = u_{1-\alpha} \quad \Rightarrow \quad c = u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$$

3. Testentscheidung:

$$|\bar{X} - \mu_0| \begin{cases} > u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} & \Rightarrow \text{Entscheidung für } H_1 \\ \leq u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

Algorithmus: Einseitiger Einstichproben-Gauß-Test (Abgrenzung nach unten)

- Normalverteilung mit unbekanntem Erwartungswert μ und bekannter Varianz σ^2
- Nullhypothese $H_0: \mu \geq \mu_0$, Alternativhypothese $H_1: \mu < \mu_0$

$$|\bar{X} - \mu_0| \begin{cases} < c & \Rightarrow \text{Entscheidung für } H_1 \\ \geq c & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

1. Vorgabe des Signifikanzniveaus α (Wahrscheinlichkeit für Fehler 1. Art)
2. Berechnung der kritischen Grenze c unter der Annahme, dass H_0 gültig ist:

$$\alpha = P(\bar{X} - \mu_0 \leq c) = P\left(\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma} \leq \sqrt{n} \cdot \frac{c}{\sigma}\right)$$

Die Variable

$$u = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma}$$

ist standardnormalverteilt. Es gilt weiterhin:

$$\alpha = \Phi\left(\sqrt{n} \cdot \frac{c}{\sigma}\right) \quad \Rightarrow \quad \sqrt{n} \cdot \frac{c}{\sigma} = u_\alpha \quad \Rightarrow \quad c = u_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

3. Testentscheidung:^a

$$|\bar{X} - \mu_0| \begin{cases} < u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} & \Rightarrow \text{Entscheidung für } H_1 \\ \geq u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

^aEs gilt $|u_{1-\alpha}| = |u_\alpha|$

Bonus: Zusammenfassung Einstichproben-Gauß-Test

H_0	H_1	Kritischer Bereich für Signifikanzniveau α
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n} \cdot \frac{ \bar{X} - \mu_0 }{\sigma} > u_{1-\alpha/2}$
$\mu \geq \mu_0$ $\mu \leq \mu_0$	$\mu < \mu_0$ $\mu > \mu_0$	$\sqrt{n} \cdot \frac{ \bar{X} - \mu_0 }{\sigma} > u_{1-\alpha}$

Algorithmus: Zweiseitiger Einstichproben-t-Test

- Normalverteilung mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2
- Nullhypothese $H_0 : \mu = \mu_0$, Alternativhypothese $H_1 : \mu \neq \mu_0$

1. Vorgabe des Signifikanzniveaus α (Wahrscheinlichkeit für Fehler 1. Art)
2. Berechnung der Test- oder Prüfvariablen:

$$T = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{S}$$

ist t-verteilt mit $f = n - 1$ Freiheitsgraden.

Es gilt weiterhin:

$$1 - \alpha = P(-c \leq T \leq c) \implies c = t_{n-1}(1 - \alpha/2)$$

3. Testentscheidung anhand eines konkreten Stichprobenergebnisses

$$\hat{t} = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{S}$$

$$|\hat{t}| \begin{cases} > t_{n-1}(1 - \alpha/2) & \implies \text{Entscheidung für } H_1 \\ \leq t_{n-1}(1 - \alpha/2) & \implies \text{Entscheidung für } H_0 \end{cases}$$

Algorithmus: Einseitiger Einstichproben-t-Test (Abgrenzung nach oben)

- Normalverteilung mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2
- Nullhypothese $H_0 : \mu \leq \mu_0$, Alternativhypothese $H_1 : \mu > \mu_0$

$$|\bar{X} - \mu_0| \begin{cases} > c & \implies \text{Entscheidung für } H_1 \\ \leq c & \implies \text{Entscheidung für } H_0 \end{cases}$$

1. Vorgabe des Signifikanzniveaus α (Wahrscheinlichkeit für Fehler 1. Art)
2. Berechnung der Test- oder Prüfvariablen:

$$T = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{S}$$

ist t-verteilt mit $f = n - 1$ Freiheitsgraden.

Es gilt weiterhin:

$$1 - \alpha = P(T \leq c) \implies c = t_{n-1}(1 - \alpha)$$

3. Testentscheidung anhand eines konkreten Stichprobenergebnisses

$$\hat{t} = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{S}$$

$$|\hat{t}| \begin{cases} > t_{n-1}(1 - \alpha) & \implies \text{Entscheidung für } H_1 \\ \leq t_{n-1}(1 - \alpha) & \implies \text{Entscheidung für } H_0 \end{cases}$$

Algorithmus: Einseitiger Einstichproben-t-Test (Abgrenzung nach unten)

- Normalverteilung mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2
- Nullhypothese $H_0 : \mu \geq \mu_0$, Alternativhypothese $H_1 : \mu < \mu_0$

$$|\bar{X} - \mu_0| \begin{cases} < c & \Rightarrow \text{Entscheidung für } H_1 \\ \geq c & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

- Vorgabe des Signifikanzniveaus α (Wahrscheinlichkeit für Fehler 1. Art)
- Berechnung der Test- oder Prüfvariablen:

$$T = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{S}$$

ist t-verteilt mit $f = n - 1$ Freiheitsgraden.

Es gilt weiterhin:

$$1 - \alpha = P(T \geq c) \quad \Rightarrow \quad c = t_{n-1}(\alpha)$$

- Testentscheidung anhand eines konkreten Stichprobenergebnisses^a

$$\hat{t} = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{S}$$

$$|\hat{t}| \begin{cases} < t_{n-1}(\alpha) & \Rightarrow \text{Entscheidung für } H_1 \\ \geq t_{n-1}(\alpha) & \Rightarrow \text{Entscheidung für } H_0 \end{cases}$$

^aEs gilt: $|t_{n-1}(\alpha)| = |t_{n-1}(1 - \alpha)|$

Bonus: Zusammenfassung Einstichproben-t-Test

H_0	H_1	Kritischer Bereich für Signifikanzniveau α
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n} \cdot \frac{ \bar{X} - \mu_0 }{S} > t_{n-1}(1 - \alpha/2)$
$\mu \geq \mu_0$ $\mu \leq \mu_0$	$\mu < \mu_0$ $\mu > \mu_0$	$\sqrt{n} \cdot \frac{ \bar{X} - \mu_0 }{S} > t_{n-1}(1 - \alpha)$

Index

- Absolute Häufigkeit, 5, 38
Absolute Klassenhäufigkeit, 42
Anzahl der Freiheitsgrade, 64
Approximation von Verteilungen, 24
Arithmetisches Mittel, 46
Aufgabe der beschreibenden Statistik, 37
Aufgabe der schließenden Statistik, 54
- Bayes'sche Formel, 10
Bedingte Wahrscheinlichkeit, 6
Beobachtungseinheit, 37
Beobachtungsmenge, 37
Berechnung des Mindeststichprobenumfangs, 76
Berechnung von Intervallen einer Standardnormalverteilung, 60
Bernoulli-Experiment, 19
Bestimmtheitsmaß, 50
Bestimmung der Grenze bei einem einfachen Alternativtest, 73
Bestimmung der Grenze bei einem einseitigen Hypothesentest, 73
Bestimmung der Grenzen bei einem zweiseitigen Hypothesentest, 74
Binomialverteilung, 19
Box-Plot, 48
- Chi-Quadrat-Verteilung, 63
- Diskrete zweidimensionale Verteilung, 27
- Einseitiger Einstichproben-Gauß-Test
(Abgrenzung nach oben), 77
Einseitiger Einstichproben-Gauß-Test
(Abgrenzung nach unten), 78
Einseitiger Einstichproben-t-Test
(Abgrenzung nach oben), 80
Einseitiger Einstichproben-t-Test
(Abgrenzung nach unten), 80
Empirische Kovarianz, 50
Empirische Varianz, 48
Empirische Verteilungsfunktion, 40
Ereignis, 4
Ereignisalgebra, 6
Ereignisraum, 4
Ergebnismenge, 4
Erwartungswert, 15
- Erwartungswert und Varianz für die Summe von Zufallsgrößen, 29
- Fehler 1. Art, 71
Fehler 2. Art, 72
- Geometrische Verteilung, 24
Gleichverteilung, 19
Grundgesamtheit, 54
- Histogramm, 43
Hypergeometrische Verteilung, 20
Hypothesentest, 71
Häufigkeitsdichte, 42
- Klasseneinteilung, 41
Kolmogoroff-Axiome, 6
Kombination, 2
Kombinatoriktabelle, 3
Konfidenzintervall und Konfidenzniveau, 60
Konfidenzintervalle für den unbekannten Anteilswert p (Parameter p einer Binomialverteilung), 69
Konfidenzintervalle für den unbekannten Erwartungswert einer Normalverteilung bei bekannter Varianz, 61
Konfidenzintervalle für den unbekannten Erwartungswert einer Normalverteilung bei unbekannter Varianz, 67
Konfidenzintervalle für die unbekannte Varianz einer Normalverteilung, 64
Konsistenz eines Tests, 76
Korrelationskoeffizient, 32
Korrelationskoeffizient (empirisch), 50
Kovarianz, 32
- Laplace-Experiment, 5
Lineare Regression, 51
Lokaler Grenzwertsatz von de Moivre-Laplace, 22
- Maximum-Likelihood-Methode, 57
Median, 16, 46
Mehrstufiges Zufallsexperiment, 10
Modalwert, 46

- Multiplikationssatz, 7
 Normalverteilung, 22
 Operationscharakteristik und Gütfunktion, 72
 Permutation, 2
 Planung und Durchführung eines Parametertests, 71
 Poisson-Verteilung, 21
 Produkt und Quotient zweier stetiger Zufallsgrößen, 29
 Quantil (empirisch), 47
 Quantile, 16
 Quartilsabstand, 47
 Randverteilung, 26
 Relative Häufigkeit, 6, 38
 Relative Klassenhäufigkeit, 42
 Schwaches Gesetz großer Zahlen, 34
 Schätzfunktion, 54
 Spannweite, 47
 Standardabweichung, 18
 Standardnormalverteilung, 23
 Starkes Gesetz großer Zahlen, 34
 Statistische Variable, 37
 Stetige zweidimensionale Verteilung, 28
 Stichprobe, 54
 Stochastisch unabhängige Zufallsvariable, 28
 Stochastische Unabhängigkeit, 9
 Summen von Zufallsgrößen, 29
 Summenhäufigkeit, 39
 t-Verteilung, 66
 Totale Wahrscheinlichkeit, 10
 Transformation auf standardisierte Variable, 23
 Transformation der Dichtefunktion, 15
 Transformation der Varianz auf eine neue Zufallsvariable, 17
 Transformation der Verteilungsfunktion, 14
 Transformation des Erwartungswertes auf eine neue Zufallsvariable, 16
 Transformierte Zufallsvariable, 14
 Tschebyscheffsche Ungleichung, 18
 Urnenmodell, 2
 Varianz, 17
 Variation, 3
 Variationskoeffizient, 49
 Verschiebungssatz, 18
 Verteilungsfunktion, 11
 Verteilungsfunktion einer zweidimensionalen Verteilung, 27
 Vierfeldertafel, 9
 Wahrscheinlichkeitsaxiome, 6
 Wahrscheinlichkeitsfunktion einer zweidimensionalen Verteilung, 26
 Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen, 12
 Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen, 13
 Zentraler Grenzwertsatz, 35
 Zufallsexperiment, 4
 Zufallsvariable, 11
 Zusammenfassung Einstichproben-Gauß-Test, 79
 Zusammenfassung Einstichproben-t-Test, 81
 Zusammenfassung von Verteilungsfunktionen, 25
 Zusammengesetzte Ereignisse, 4
 Zusammenhangskoeffizient, 9
 Zweiseitiger Einstichproben-Gauß-Test, 75
 Zweiseitiger Einstichproben-t-Test, 79

Beispiele

Bedingte Wahrscheinlichkeit (Berechnung), 8
Bedingte Wahrscheinlichkeit
 (Wahrscheinlichkeitsbaum), 7
Binomialverteilung, 20

Chi-Quadrat-Verteilung, 64

Dichtefunktion, 13

Empirische Kovarianz, 52
Empirische Verteilungsfunktion, 40
Empirische Verteilungsfunktion (klassierte
 Daten), 44
Empirischer Korrelationskoeffizient, 53
Ereignisbaum, 10

Fehler 1. und 2. Art, 72

Geometrische Verteilung, 24

Histogramm, 43
Hypergeometrische Verteilung, 21

Konfidenzintervalle für den unbekannten
 Erwartungswert einer
 Normalverteilung bei bekannter
 Varianz, 62
Konfidenzintervalle für den unbekannten
 Erwartungswert einer
 Normalverteilung bei unbekannter
 Varianz, 68
Konfidenzintervalle für die unbekannte
 Varianz einer Normalverteilung, 65

Kovarianz und Korrelationskoeffizient, 32

Lineare Regression, Empirische Kovarianz, 51
Lineare Regression, Empirischer
 Korrelationskoeffizient, 52
Lineare Regression, Regressionsgerade, 53
Lineare Regression, Regressionsgleichung, 53

Maximum-Likelihood-Methode, 58

Normalverteilung, 23

Poisson-Verteilung, 22

Regressionsgerade, 53
Regressionsgleichung, 53

Schätzfunktion (Erwartungstreue), 55
Schätzfunktion (Konsistenz), 56
Schätzung der Varianz, 55
Schätzung des Erwartungswertes, 54
Stabdiagramm, 38
Summen von Zufallsgrößen, 30

t-Verteilung, 67
Transformation der Verteilungsfunktion und
 Dichtefunktion, 15
Transformation des Erwartungswertes und
 der Varianz, 17

Verteilungsfunktion, 11

Zentraler Grenzwertsatz, 35
Zweidimensionale Zufallsvariable, 27