Mathematik IV
(für IF, ET, Ph)

Oliver Ernst

Professur Numerische Mathematik

Sommersemester 2019

Studiengänge: B Informatik, B Angewandte Informatik,
B Elektrotechnik und Informationstechnik,
B Elektromobilität,
B Physik, B Computational Science

Mathematik!
TU Chemnitz

Inhalt I

Inhalt II

Inhalt

Stochastik

Der Begriff Stochastik umfasst die beiden mathematischen Disziplinen Wahrscheinlichkeitstheorie und Statistik. Grob gesagt:

- Wahrscheinlichkeitstheorie ist der rationale Umgang mit Unsicherheit und
- Statistik ist die Wissenschaft der Extraktion von Information und Erkenntnis aus Daten.

Beide benutzen ähnliche Grundbegriffe und Denkweisen.

- Stochastik wird zur angewandten Mathematik gezählt, d.h. ihre Techniken sind motiviert aus Fragestellungen der realen Welt: Risiken bewerten, Langzeitverhalten von zufälligen Phänomenen, Vorhersage.
- Als „Gründungsurkunde“ der Stochastik wird ein Brief Blaise Pascals aus dem Jahr 1654 an Pierre de Fermat angesehen, in dem er ihm das auf Luca Pacioli zurückgehendes Teilungsproblem beschreibt, also wie bei einem vorzeitig abgebrochenen Würfelspiel der Gewinneinsatz unter den beiden Spielern fair zu verteilen sei.
- Zuvor war ca. 1564 das Buch „Liber de Ludo Aleae“ (Buch der Glücksspiele) des Universalgelehrten Gerolamo Cardano erschienen, mit Anleitungen zum Rechnen mit Wahrscheinlichkeiten.

Stochastik

- Eine widerspruchsfreie mathematische Axiomatik für Wahrscheinlichkeit gelang erst Andrey Kolmogorov 1933.
- Stochastik ist heute ein sehr spannendes Gebiet, weil einerseits der rationale Umgang mit Unsicherheit lebensnotwendig ist, gleichzeitig die heute verfügbare mathematische Theorie, gepaart mit Rechenleistung und Datenverfügbarkeit in nie dagewesenem Umfang, ganz neue Aussagen und Vorhesragen ermöglichen.

Inhalt

- Die mathematischen Objekte, denen wir eine Wahrscheinlichkeit zuordnen werden, sind Mengen und heißen in der Stochastik Ereignisse. Genauer gesagt: sie sind Teilmengen einer Grundmenge Ω von Elementareregnissen.
- Einfachstes Beispiel sind die möglichen Ausgänge eines Münzwurfs (Kopf oder Zahl), welche wir mit $\Omega = \{0, 1\}$ modellieren.
- Interessanter ist die Augenzahl bei einem Wurf eines Würfels. Hier ist $\Omega = \{1, 2, 3, 4, 5, 6\}$.
  Elementarereignisse sind dier die Elemente von Ω, also $1, 2, 3, 4, 5, 6.$.
  Ein interessanteres Ereignis wäre etwa $A = \{2, 4, 6\}$, was wir umgangssprachlich mit „Augenzahl gerade“ beschreiben können.
- Um mit Ereignissen rechnen zu können benötigen wir Mengenoperationen wie Vereinigung, Durchschnitt oder Komplementbildung. Eine mathematische Struktur, die sich hierfür bestens eignet, ist die einer σ-Algebra.

Definition 15.1

Sei Ω eine Menge. Ein System $\mathfrak{A}$ von Teilmengen von Ω heißt σ-Algebra auf Ω, wenn folgende Bedingungen erfüllt sind:

1. $\Omega \in \mathfrak{A}$.
2. Mit A liegt auch das Komplement $A^c := \Omega \setminus A$ in $\mathfrak{A}$.
3. Gilt $A_k \in \mathfrak{A}$ für $k \in \mathbb{N}$, so ist auch $\cup_{k\in\mathbb{N}} A_k \in \mathfrak{A}$. ($\sigma$-Additivität)

Bemerkungen:

1. Mit $A, B \in \mathfrak{A}$ gilt dies also auch für $A \cup B$, $A \cap B$, A^c, $(A \cap B)^c$ etc.
2. Anwendung der de Morganschen Regeln zeigt: mit der abzählbar unendlichen Vereinigung liegt auch der abzählbar unendliche Durchschnitt einer Folge von Mengen aus $\mathfrak{A}$ wieder in $\mathfrak{A}$.
3. $\varnothing$ wird als unmögliches Ereignis, Ω als sicheres Ereignis bezeichnet.
4. Die Potenzmenge (das System aller Teilmengen) von Ω, geschrieben $\mathfrak{P}(\Omega)$ (manchmal auch als 2^Ω), ist eine σ-Algebra auf Ω.

Definition 15.2

a) Eine Menge Ω zusammen mit einer σ-Algebra auf Ω heißt Messraum $(\Omega, \mathfrak{A})$.

b) Eine Abbildung $\mu : \mathfrak{A} \to [0, \infty]$ heißt Maß, falls

   i) $\mu(\varnothing) = 0$

   ii) Sind $A_k \in \mathfrak{A}$ paarweise disjunkt, so gilt $\mu(\cup_{k \in \mathbb{N}} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$.

c) Ist $\mu(\Omega) < \infty$, so heißt μ endliches Maß, und ist $\mu(\Omega) = 1$, so heißt μ Wahrscheinlichkeitsmaß oder W-Maß.

Oft schreibt man $\mathbf{P}$ für Wahrscheinlichkeitsmaße und nennt $(\Omega, \mathfrak{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum.

Sind $A, B \in \mathfrak{A}$ und $\mathbf{P}$ ein Wahrscheinlichkeitsmaß auf $\mathfrak{A}$, so gelten

1. $\mathbf{P}(\varnothing) = 0$.
2. $A \cap B = \varnothing$, so gilt $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.
3. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
4. $A \subset B \;\Rightarrow\; 0 \leq \mathbf{P}(A) \leq \mathbf{P}(B) \leq 1$. (Isotonie)
5. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$. (Wahrscheinlichkeit des Gegenereignisses)

- Ist die Grundmenge endlich, d.h. $|\Omega| < \infty$, so kann man als σ-Algebra $\mathfrak{A}$ die Potenzmenge $\mathfrak{P}(\Omega)$ wählen.
- Jedes Wahrscheinlichkeitsmaß **P** auf $(\Omega, \mathfrak{A})$ ist dann charakterisiert durch

$$\mathbf{P}(\{\omega\}),\ \omega \in \Omega, \qquad \text{mit} \sum_{\omega\in\Omega} \mathbf{P}(\{\omega\}) = 1.$$

- Mit den Rechenregeln für Wahrscheinlichkeitmaße gilt dann für alle $A \subset \Omega$

$$\mathbf{P}(A) = \sum_{\omega\in A} \mathbf{P}(\{\omega\}).$$

Wir unterscheiden Anordnungsprobleme (mit Beachtung der Reihenfolge) und Auswahlprobleme (ohne Beachtung der Reihenfolge). Sei $|\Omega| = n < \infty$.

- Anordnung ohne Einschränkung: Die Menge Ω^k, $k \in \mathbb{N}$, der geordneten k-Tupel von Elementen aus Ω ist die Menge aller Anordnungen der Länge k, es gilt $|\Omega^k| = n^k$.
  Alternative Bezeichnungen: Variationen mit Wiederholung, geordnete Auswahlen von k Elementen mit Zurücklegen.
- Anordnung verschiedener Elemente: Die Menge aller geordneten k-Tupel von Elementen aus Ω bei denen alle Elemente verschieden sind (insbesondere $k \leq n$), besitzt die Mächtigkeit

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

  Alternative Bezeichnungen: Variationen ohne Wiederholungen, geordnete Auswahlen von k Elementen aus Ω ohne Zurücklegen.

- Permutationen einer Menge: Die Anzahl aller Permutationen (bijektive Abbildungen nach $\{1, \dots, n\}$) der Elemente von Ω beträgt $n!$.
- Auswahl einer Teilmenge: Die Anzahl aller möglichen Teilmengen bestehend aus $k \le n$ (verschiedenen) Elementen von Ω beträgt

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \qquad 0 \le k \le n.$$

- Gilt $|\Omega| = n < \infty$, so gilt $|\mathfrak{P}(\Omega)| = 2^n$ (oder auch $|2^\Omega| = 2^{|\Omega|}$).
- Die Anzahl aller k-elementigen Teilmengen ($0 \le k \le n$) von Ω ist gegeben durch
$$|\{A \subset \Omega : |A| = k\}| = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$
- Damit gilt
$$\binom{n}{k} = \binom{n}{n-k} \quad \text{und} \quad \sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Beispiel: (Urnenmodell mit Zurücklegen). In einer Urne sind insgesamt N Kugeln, W weiße und $S = N - W$ schwarze. In unserem Zufallsexperiment ziehen wir unabhängig, nacheinander n Kugeln. Was ist die Wahrscheinlichkeit dafür, dass wir:

1. n-mal die gleiche Kugel ziehen.
2. In den ersten k Ziehungen weiße Kugeln ziehen (Ereignis A_k).
3. Bei den n Zügen genau k weiße Kugeln ziehen, (Ereignis $A_{k,n}$).

Satz 15.3

Seien $(\Omega_1, \mathfrak{A}_1, \mathbf{P}_1), \dots, (\Omega_n, \mathfrak{A}_n, \mathbf{P}_n)$ *Wahrscheinlichkeitsräume. Dann heißt*

$$\mathfrak{A}_1 \otimes \cdots \otimes \mathfrak{A}_n := \{\textit{kleinste } \sigma\textit{-Algebra, die die Mengen } A_1 \times \cdots \times A_n,\ A_j \in \mathfrak{A}_j,\ \textit{enthält}\}$$

die Produkt σ-Algebra und es gibt genau ein Wahrscheinlichkeitsmaß $\mathbf{P}$ *auf* $\mathfrak{A}_1 \otimes \cdots \otimes \mathfrak{A}_n$ *mit*

$$\mathbf{P}(A_1 \times \cdots \times A_n) = \mathbf{P}_1(A_1) \cdot \ldots \cdot \mathbf{P}_n(A_n) \qquad \textit{für alle } A_j \in \mathfrak{A}_j, j = 1, \dots, n.$$

Wir schreiben $\mathbf{P} = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n$ *und nennen dieses Maß Produktmaß.*

- Wir haben diese Konstruktion schon gesehen im wichtigen Spezialfall, dass alle $(\Omega_j, \mathfrak{A}_j, \mathbf{P}_j)$ gleich sind.
- Im Allgemeinen ist dieser Satz ein tiefes, wichtiges Resultat und folgt aus dem Fortsetzungssatz von Carathéodory.

Für endliche $\{\Omega_j\}_{j=1}^n$ ist die Aussage einfach zu beweisen. Grundlage ist folgende Beobachtung:

Proposition 15.4

Sei $|\Omega| < \infty$. Dann gibt es zu jedem Maß μ auf dem Messraum $(\Omega, \mathfrak{P}(\Omega)$ genau eine Abbildung $p : \Omega \to [0,1]$ mit

$$\mu(\{\omega\}) = p(\omega)\ \forall \omega \in \Omega, \qquad \textit{und} \qquad \mu(A) = \sum_{\omega \in A} p(\omega)\ \forall A \in \mathfrak{P}(\Omega).$$

Anwendung auf Produktmaße: wir definieren $p_i(\omega_i) := \mathbf{P}_i(\{\omega_i\}), i = 1, \dots, n$ und setzen

$$p(\omega_1, \dots, \omega_n) := p_1(\omega_1) \cdot \ldots \cdot p_n(\omega_n).$$

Beispiel: (Urnenmodell ohne Zurücklegen) In unserer Urne sind N Kugeln: S schwarze und W weiße. Was ist die Wahrscheinlichkeit, dass bei n Ziehungen genau k Kugeln weiß sind? Dieses Ereignis bezeichnen wir mit $B_{n,k}$.

Inhalt

Beim Wurf eines (fairen) Würfels betrachten wir folgende beide Ereignisse:

$$A = \{\text{Augenzahl ungerade}\}, \qquad B = \{\text{ Augenzahl } \leq 3\}.$$

Welche Wahrscheinlichkeit besitzen diese?
Ändert sich $\mathbf{P}(A)$,durch das Wissen, dass B eingetreten ist?

Modellierung mit Wahrscheinlichkeitsraum

$$\Omega = \{1, 2, \dots, 6\}, \quad \mathfrak{A} = \mathfrak{P}(\Omega), \quad \mathbf{P}(\{\omega\}) = 1/6 \text{ für alle } \omega \in \Omega.$$

Dann ist

$$A = \{1, 3, 5\}, \qquad B = \{1, 2, 3\}.$$

Wir nehmen an, dass B eingetreten ist, und führen ein neues Wahrscheinlichkeitsmaß $\mathbf{P}_B$ auf $(B, \mathfrak{P}(B))$ ein: Für jede Menge $C \subset B$ definieren wir

$$\mathbf{P}_B(C) := \frac{|C|}{|B|}.$$

Elementarereignissen ω aus $\Omega \setminus B$ (die ja nicht eingetreten sein können) weisen wir die Wahrscheinlichkeit Null zu:

$$\mathbf{P}_B(\{\omega\}) = 0 \quad \forall \omega \in \Omega \setminus B.$$

Hiermit können wir $\mathbf{P}_B$ zu einem Wahrscheinlichkeitsmaß $\mathbf{P}(\cdot|B)$ auf ganz $\mathfrak{A}$ fortsetzen durch

$$\mathbf{P}(C|B) := \mathbf{P}_B(C \cap B) = \frac{|C \cap B|}{|B|} \quad \forall C \subset \Omega.$$

Damit ergibt sich nun für das Würfelbeispiel

$$\mathbf{P}(A|B) = \frac{|\{1,3\}|}{|\{1,2,3\}|} = \frac{2}{3}.$$

Definition 15.5 (Bedingte Wahrscheinlichkeit von Ereignissen)

Sei $(\Omega, \mathfrak{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und $B \in \mathfrak{A}$ mit $\mathbf{P}(B) > 0$. Dann ist für alle $A \in \mathfrak{A}$ die bedingte Wahrscheinlichkeit gegeben B definiert durch

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

Im Fall $\mathbf{P}(B) = 0$ definiert man oft (willkürlich) $\mathbf{P}(A|B) = 0$ für alle A.

Proposition 15.6

Ist $\mathbf{P}(B) > 0$, so ist (mit den obigen Bezeichnungen) $\mathbf{P}(\cdot|B)$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{A})$.

Definition 15.7 (Unabhängigkeit von Ereignissen)

Sei $(\Omega, \mathfrak{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Zwei Ereignisse $A, B \in \mathfrak{A}$ heißen unabhängig, falls

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B).$$

Bemerkung: Für $\mathbf{P}(B) > 0$ gilt offenbar: A, B unabhängig $\Leftrightarrow \mathbf{P}(A|B) = \mathbf{P}(A)$.

Proposition 15.8 (Formel der totalen Wahrscheinlichkeit)

Sei $(\Omega, \mathfrak{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und $B_1, \dots, B_n$ eine Zerlegung von Ω mit $\mathbf{P}(B_i) > 0$ für $i = 1, \dots, n$. Dann gilt für alle $A \in \mathfrak{A}$

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(A|B_i)\, \mathbf{P}(B_i).$$

Bemerkung: Die Aussage gilt auch für abzählbar viele $B_i, i \in \mathbb{N}$.

Proposition 15.9 (Bayessche Formel für Ereignisse)

Sei $(\Omega, \mathfrak{A}, \mathbf{P})$ *ein Wahrscheinlichkeitsraum,* $A, B \in \mathfrak{A}$ *und* $\mathbf{P}(A)\mathbf{P}(B) > 0$. *Dann gilt*

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A|B) \cdot \mathbf{P}(B)}{\mathbf{P}(A)}.$$

Mit Hilfe der Formel für die totale Wahrscheinlichkeit ergibt sich sofort folgende Verallgemeinerung: ist $(B_i)_{i\in\mathbb{N}}$ eine Zerlegung von Ω mit $\mathbf{P}(B_i) > 0$ für alle $i \in \mathbb{N}$, so gilt

$$\mathbf{P}(B_i|A) = \frac{\mathbf{P}(A|B_i)\,\mathbf{P}(B_i)}{\sum_{j\in\mathbb{N}} \mathbf{P}(A|B_j)\,\mathbf{P}(B_j)}, \quad i \in \mathbb{N}.$$

Diese Formel (Bayes-Formel für Zerlegungen) gibt an, wie sich die Wahrscheinlichkeit für alle Ereignisse B_i nach Bekanntwerden der Eintretens von Ereignis A verändert.

Bei einer Leichtathletikveranstaltung seien 50% der Athleten gedopt, 40% der Athleten positiv getestet und 95% der positiv getesteten seien tatsächlich gedopt.

- Mit welcher Wahrscheinlichkeit hat ein gedopter Athlet einen positiven Test?
- ... einen negativen Test?

Drei Maschinen stellen das gleiche Bauteil her, mit folgenden Ausschussraten:

Anteile	M_1	M_2	M_3
Produkt	20%	50%	30%
Ausschuss	1%	4%	5%

1. Wie groß ist die Wahrscheinlichkeit dafür, dass ein zufällig ausgewähltes Bauteil Ausschuss ist?
2. Ein Bauteil ist Ausschuss. Wie groß ist die Wahrscheinlichkeit dafür, dass es von Maschine 2 stammt?

Bemerkung: Seien $(\Omega_1, \mathfrak{A}_1, \mathbf{P}_1)$, $(\Omega_2, \mathfrak{A}_2, \mathbf{P}_2)$ Wahrscheinlichkeitsräume. Dann sind für $A_i \in \mathfrak{A}_i$ die Ereignisse $A_1 \times \Omega_2$ und $\Omega_1 \times A_2$ unabhängig, denn:

$$\begin{aligned}
(A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) &= A_1 \times A_2, \\
\mathbf{P}((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)) &= \mathbf{P}(A_1 \times A_2) = \mathbf{P}_1(A_1) \cdot \mathbf{P}_2(A_2) \\
&= \mathbf{P}(A_1 \times \Omega_2) \cdot \mathbf{P}(\Omega_1 \times A_2)
\end{aligned}$$

Entsprechendes gilt für Produkte von n Wahrscheinlichkeitsräumen. Dabei heißen $A_1, \dots, A_n$ unabhängig, wenn

$$\mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{i=1}^{k} \mathbf{P}(A_{i_j}) \qquad \text{für alle } \{i_1, \dots, i_k\} \subset \{1, \dots, n\}.$$

Inhalt

15 Stochastik

- Bernoulli-Verteilung: $\Omega = \{0,1\}$, $\mathbf{P}(\{1\}) = p$, $\mathbf{P}(\{0\}) = 1-p$, $p \in [0,1]$.
  Faire Münze: $p = 1/2$.

Wie kann man durch eine leichte Modifikation der Wurfregeln mit einer unfairen Münze eine faire simulieren? (Lösung wird John von Neumann zugeschrieben)

- Binomialverteilung: Ausführen von n unabhängigen Bernoulli-Experimenten.
  $p \in [0,1]$, $n \in \mathbb{N}$, $\Omega = \{0,1,2,\dots,n\}$ (# „positiver“ Ausgänge), mit
  $$\mathbf{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$
- Hypergeometrische Verteilung: Seien $0 \le M \le N$; $n, M, N \in \mathbb{N}_0$ und $\Omega = \{0,1,\dots,n\}$.
  $$\mathbf{P}(\{k\}) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$
  definiert die hypergeometrische Verteilung.
  (Kommt aus Urnenmodell ohne zurücklegen: M weiße, $N-M$ schwarze Kugeln, $\mathbf{P}(\{k\})$ = Wahrscheinlichkeit, dass genau k Kugeln weiß)

Inhalt

Definition 15.10

Eine reelle Zufallsvariable ist eine Funktion $X : \Omega \to \mathbb{R}$, wobei $(\Omega, \mathfrak{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum ist und X messbar, d.h. für jedes Intervall $I \subset \mathbb{R}$ ist

$$X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \in \mathfrak{A}.$$

Die Verteilungsfunktion einer Zufallsvariable X ist die reelle Funktion

$$F_X : \mathbb{R} \to \mathbb{R}, \quad F_X(t) := \mathbf{P}(X \leq t) = \mathbf{P}(\{\omega \in \Omega : X(\omega) \leq t\}).$$

Die Wahrscheinlichkeitsverteilung oder kurz Verteilung von X ist das Wahrscheinlichkeitsmaß $\mathbf{P}_X$ auf $(\mathbb{R}, \mathfrak{B}_1)$ gegeben durch

$$\mathbf{P}_X(I) := \mathbf{P}(X^{-1}(I));$$

dabei ist $\mathfrak{B}_1$ die Borel σ-algebra auf $\mathbb{R}$, die kleinste σ-algebra, die alle Intervalle enthält.

Beispiele für Zufallsvariablen:

1. Gesamtaugenzahl bei dreimal Würfeln.
2. Gewinn bei Glückspiel
3. Trefferzahl bei n Bernoulli-Experimenten.

Satz 15.11

Ist $(\Omega, \mathfrak{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, so definiert das Maß $\mathbf{P}$ ein Integral auf einer geeigneten Menge von Funktionen $L^1(\Omega, \mathfrak{A}, \mathbf{P})$, das durch

$$\int_\Omega \mathbb{1}_A \,\mathrm{d}\mathbf{P} = \mathbf{P}(A), \qquad A \in \mathfrak{A},$$

die Linearität, und eine geeignete Stetigkeitseigenschaft eindeutig bestimmt ist.

Definition 15.12

Ist $X \in L^1(\Omega, \mathfrak{A}, \mathbf{P})$ eine integrierbare Zufallsvariable, so heißt

$$\mathbf{E}[X] = \int X \, \mathrm{d}\mathbf{P}$$

der Erwartungswert von X.

Rechenregeln für Erwartungswerte: Sei $(\Omega, \mathfrak{A}, \mathbf{P})$, $X, Y \in L^1(\Omega, \mathfrak{A}, \mathbf{P})$, $c \in \mathbb{R}$.

1. $X + Y \in L^1(\Omega, \mathfrak{A}, \mathbf{P})$ und $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$.
2. $cX \in L^1(\Omega, \mathfrak{A}, \mathbf{P})$ und $\mathbf{E}[cX] = c\,\mathbf{E}[X]$.
3. $X + c \in L^1(\Omega, \mathfrak{A}, \mathbf{P})$ und $\mathbf{E}[X + c] = \mathbf{E}[X] + c$.

Dabei bedeuten (die ersten beiden (1) und (2) Regeln die Linearität des Integrals und die dritte folgt aus der Beobachtung, dass

$$\mathbf{E}[c] = c\,\mathbf{E}[1] = c\mathbf{E}[\mathbb{1}_\Omega] = c\mathbf{P}(\Omega) = c$$

zusammen mit (1) und (2).

Definition 15.13

Ist $X \in L^1(\Omega, \mathfrak{A}, \mathbf{P})$ und ist zusätzlich $X^2 \in L^1(\Omega, \mathfrak{A}, \mathbf{P})$, so heißt

$$\mathbf{Var}\, X := \mathbf{E}\left[(X - \mathbf{E}[X])^2\right]$$

die Varianz oder Streuung von X sowie $\sqrt{\mathbf{Var}\, X}$ die Standardabweichung.

Bemerkung: Besitzen zwei Zufallsvariablen $X : \Omega \to \mathbb{R}$ und $Y : \Omega' \to \mathbb{R}$ dieselbe Verteilung, so sind $\mathbf{E}[X] = \mathbf{E}[Y]$ und $\mathbf{Var}\, X = \mathbf{Var}\, Y$.

Beispiel: Eine Zufallsvariable auf $(\Omega, \mathfrak{A}, \mathbf{P})$ heißt Bernoulli-verteilt mit $p \in [0, 1]$, falls

$$\mathbf{P}(X = 0) = 1 - p, \quad \mathbf{P}(X = 1) = p,$$

d.h. X nimmt nur die Werte 0 oder 1 an. Es folgt

$$\mathbf{E}[X] = p, \qquad \mathbf{Var}\, X = p(1 - p).$$

Beispiel: Binomialverteilte Zufallsvariable: $\mathbf{E}[X] = np$, $\mathbf{Var}\, X = np(1 - p)$.

Bemerkung: (Zuffallsvariablen mit endlich vielen Werten)
Sei $(\Omega, \mathfrak{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable mit

$$X(\Omega) = \{x_1, \ldots, x_n\} \subset \mathbb{R}.$$

Dann sind $\Omega_i := X^{-1}(\{x_i\})$ paarweise disjunkt und mit

$$p_i := \mathbf{P}(\Omega_i)$$

gilt $X \in L^1(\Omega, \mathfrak{A}, \mathbf{P})$ und die Verteilungsfunktion ist gegeben durch eine Treppenfunktion mit Sprüngen p_i an den Stellen x_i.

Die Verteilung $\mathbf{P}_X$ ist das Wahrscheinlichkeitsmaß gegeben durch

$$\mathbf{P}_X(I) = \sum_{x_i \in I} p_i.$$

Zur einfachen Berechnung der Varianz hilft Unabhängigkeit:

Definition 15.14

Zwei Zufallsvariablen X, Y auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbf{P})$heißen unabhängig, wenn für alle Intervalle $I, J \subset \mathbb{R}$ die Ereignisse $\{X \in I\}$ und $\{Y \in J\}$ unabhängig sind.

Proposition 15.15

Sind X und Y integrierbar und unabhängig, so gilt $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$.

Bemerkung: Sind X, Y unabhängig und $\varphi, \psi : \mathbb{R} \to \mathbb{R}$, so sind auch $\varphi(X)$ und $\psi(Y)$ unabhängige Zufallsvariable. (einfach)

Proposition 15.16

Sind die Zufallsvariablen X_i, $i = 1, \ldots, n$ paarweise unabhängig, so ist $\mathbf{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbf{Var}\, X_i$.

Inhalt

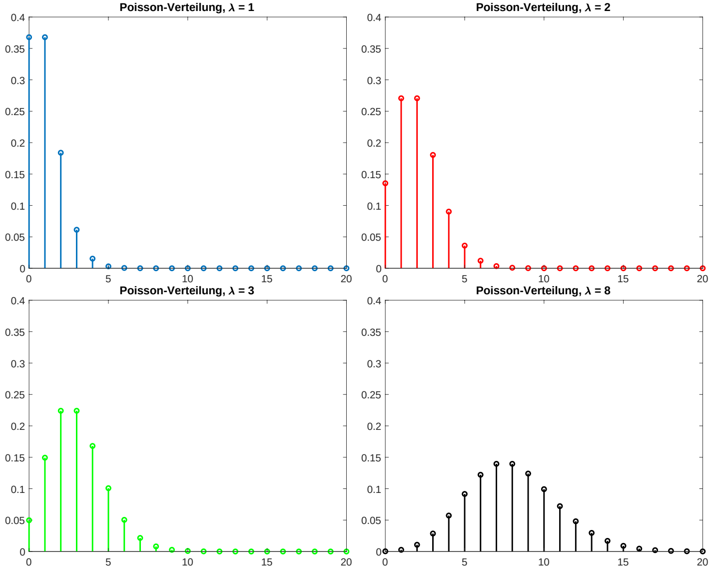
Sei $\lambda > 0$ und ein Wahrscheinlichkeitsmaß definiert auf $\mathbb{N}_0$ gegeben durch

$$\mathbf{P}(\{k\}) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \qquad k \in \mathbb{N}_0.$$

- Dieses Wahrscheinlichkeitsmaß heißt Poisson-Verteilung mit Intensität λ.
- Poisson-Veteilungen modellieren (seltene) unabhängige Ereignisse, welche in einem festen Zeitintervall mit einer mittleren Häufigkeit λ eintreten: # eingehender Mobilfunk-Anrufe pro Tag, # Webzugriffe pro Stunde, etc.
- Eine Binomialverteilung mit n groß und p klein ist näherungsweise Poisson-verteilt mit Intensität $\lambda = np$.
- Eine reelle Zufallsvariable heißt Poisson-verteilt, wenn $X(\Omega) = \mathbb{N}_0$ und

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

Es ist dann $\mathbf{E}[X] = \mathbf{Var}\, X = \lambda$.


Poisson-Verteilung, λ = 1
Poisson-Verteilung, λ = 2
Poisson-Verteilung, λ = 3
Poisson-Verteilung, λ = 8
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
5
10
15
20

Satz 15.17 (Poisson)

Die Wahrscheinlichkeiten der Binomialverteilung auf $\{0, 1, \dots, n\}$ mit $p_n = \lambda/n$ konvergieren für $n \to \infty$ gegen die Wahrscheinlichkeiten der Poisson-Verteilung.

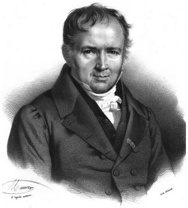
Beispiel: X hypergeometrisch verteilt, mit Parametern $N, N, n \in \mathbb{N}$; Urnenmodell ohne Zurücklegen, N Kugeln, $M \le N$ weiße, n Züge. Hier ist

$$\mathbf{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad 0 \le k \le n,$$

sowie $\mathbf{E}[X] = \frac{nM}{N}$ und $\mathbf{Var}\, X = \frac{nM}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$.

Satz 15.18

Für $N \to \infty$ und $\frac{M}{N} \to p$ konvergieren die Wahrscheinlichkeiten der hypergeometrischen Verteilung gegen die Bernoulli-Verteilung mit Parametern n und p.



- Die Poisson-Verteilung ist benannt nach dem französischen Mathematiker, Ingenieur und Physiker Siméon Denis Poisson (1781–1840).
- Die Poisson-Verteilung geriet lange in Vergessenheit und erhielt durch das Werk des russischen Ökonomen und Statistikers Ladislaus von Bortkiewicz (1868–1931) erneute Aufmerksamkeit.
- In seinem 1898 erschienenen Buch „Das Gesetz der kleinen Zahlen" findet sich z.B. folgendes Anwendungsbeispiel:
- 20 Jahre hindurch wurde festgehalten, wie oft Rekruten des kaiserlich-preußischen Heeres durch Hufschlag zu Tode kamen. (Zehn Regimenter über 20 Jahre, d.h. 200 Beobachtungen pro Regiment pro Jahr.) Gesamtzahl der Todesfälle 122, ergibt Rate von $\lambda = 122/200 \approx 0,61$.
- Klassische Poisson-Situation: seltenes Ereignis, beobachtet über wiederholte feste Zeitintervalle.

Wahrscheinlichkeiten nach Poisson-Verteilung für keine Hufschlagtote (HST) in einem Jahr:

$$\mathbf{P}(X = 0) = \exp(-\lambda)\frac{\lambda^0}{0!} \approx 0.543, \quad \lambda = 0.61.$$

Somit wäre zu erwarten, dass im Verlauf vom 200 Jahren $\mathbf{P}(X = 0) \cdot 200 \approx 109$ Jahre ohne Hufschlagtote verlaufen. Vollständige Tabelle:

# HST (k)	$\mathbf{P}(X = k)$	Erwartete # Jahre mit k HST	beobachtete # Jahre mit k HST
0	0.54335	108.67	109
1	0.33145	66.29	65
2	0.10110	20.22	22
3	0.02055	4.11	3
4	0.00315	0.63	1
5	0.00040	0.08	0
6	0.00005	0.01	0

Inhalt

Eine Zufallsvariable X heißt stetig verteilt, wenn es eine „schöne“ Funktion $f_X \geq 0$ auf $\mathbb{R}$ gibt mit der Eigenschaft

$$F_X(t) = \int_{-\infty}^{t} f_X(s)\,\mathrm{d}s.$$

Diese Funktion f_X heißt dann Dichte von X.

Beispiel: Zur Dichte

$$f_X(t) = \mathbb{1}_{[0,1]}(t) = \begin{cases} 1 & \text{falls } t \in [0,1], \\ 0 & \text{sonst,} \end{cases}$$

gehört die Verteilungsfunktion

$$F_X(t) = \begin{cases} 0 & \text{falls } t \leq 0, \\ t & \text{falls } 0 < t < 1, \\ 1 & \text{falls } t \geq 1. \end{cases}$$

Die zugehörige Zufallsvariable heißt gleichverteilt auf $[0,1]$.

Es gelten folgende Beziehungen/Rechenregeln für Zufallsvariablen X mit Dichten f_X:

- $$\int_{-\infty}^{\infty} f_X(t) \, \mathrm{d}t = \lim_{t \to \infty} F_X(t) = 1.$$
- F_X ist stetig.
- In allen Stetigkeitspunkten $t_0$von f_X ist F_X differenzierbar und es gilt
$$f_X(t_0) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(t)|_{t=t_0}.$$
- Für $-\infty < a < b < \infty$ gilt
$$\mathbf{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(s) \, \mathrm{d}s.$$
- $\mathbf{P}(X = x_0) = \mathbf{P}(x_0 \leq X \leq x_0) = F_X(x_0) - F_X(x_0) = 0.$
- $\mathbf{P}(X \in A) = \int_A f_X(t) \, \mathrm{d}t$ für messbares $A \subset \mathbb{R}$.

Für eine stetige Zufallsvariable X gilt $X \in L^1 \Leftrightarrow \int_{-\infty}^{\infty} |t| f_X(t)\ \mathrm{d}t < \infty$ und

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} t f_X(t)\ \mathrm{d}t.$$

Beispiel: Erwartungswert und Varianz einer gleichverteilten Zufallsvariable.

Beispiel: Exponentialverteilte Zufallsvariable

Eine Zufallsvariable heißt exponentialverteilt mit Parameter $\lambda > 0$, wenn sie die folgende Dichte besitzt

$$f_X(t) = \begin{cases} 0 & \text{für } t < 0, \\ \lambda\, \mathrm{e}^{-\lambda t} & \text{für } t \geq 0. \end{cases}$$

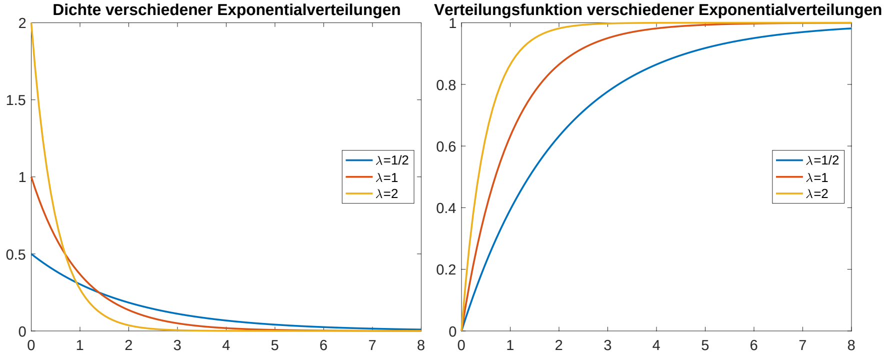
Es gilt dann

$$F_X(t) = \begin{cases} 0 & \text{für } t < 0, \\ 1 - \mathrm{e}^{-\lambda t} & \text{für } t \geq 0, \end{cases}$$

sowie

$$\mathbf{E}\,[X] = \frac{1}{\lambda}, \qquad \mathbf{Var}\, X = \frac{1}{\lambda^2}.$$

Diese Verteilung ergibt sich als einfaches Modell in der Zuverlässigkeitstheorie für den Fall einer konstanten Ausfallrate λ.



Dichte verschiedener Exponentialverteilungen
λ=1/2
λ=1
λ=2
Verteilungsfunktion verschiedener Exponentialverteilungen
λ=1/2
λ=1
λ=2

- Erinnerung: eine Poisson-verteilte Zufallsvariable X beschreibt die Häufigkeit eines seltenen Ereignisses innerhalb eines vorgegebenen Zeitraumes oder auch räumlichen Bereiches.
- Bei mittlerer Häufigkeit (Intensität) $\lambda > 0$ beträgt die Wahrscheinlichkeit des Auftretens von $k \in \mathbb{N}_0$ Ereignissen

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

- Interessiert man sich für den Zeitpunkt, bis zu dem das erste Ereignis eintritt, so lautet die Wahrscheinlichkeit, dass bis zum Zeitpunkt λt noch nichts passiert ist

$$\mathrm{e}^{-\lambda t} \frac{(\lambda t)^0}{0!} = \mathrm{e}^{-\lambda t}.$$

- Bezeichnet T die Zeit bis zum ersten Ereignis, so gilt

$$\mathbf{P}(T > t) = \mathrm{e}^{-\lambda t} \quad \text{oder} \quad \mathbf{P}(T \le t) = 1 - \mathrm{e}^{-\lambda t}.$$

- Charakteristisch für durch eine exponentialverteilte Zufallsvariable X modellierte Lebensdauer technischer Komponenten ist die Gedächtnislosigkeit der Exponentialverteilung: es gilt

$$\mathbf{P}(X > t) = \mathbf{P}(X > t + t_0 | X > t_0).$$

- Hat ein Bauteil bis zum Zeitpunkt t_0 überlebt, so ist es genauso wahrscheinlich, eine weitere Zeitdauer von t zu überleben wie von $t_0 = 0$ beginnend.
- Soll eine erhöhte Versagenswahrscheinlichkeit infolge von Abnutzung modelliert werden, so kann eine Gamma- oder Weibull-Verteilung herangezogen werden.

In einer Maschine sei eine Komponente mit mittlerer Betriebsdauer zwischen Ausfällen (engl. mean time between failures, MTBF) $1/\lambda = 5$ mehrfach verbaut. Wenn 6 dieser Komponenten verbaut sind, wir hoch ist die Wahrscheinlichkeit, dass nach 8 Jahren hiervon noch 2 im Betrieb sind?

Stochastik

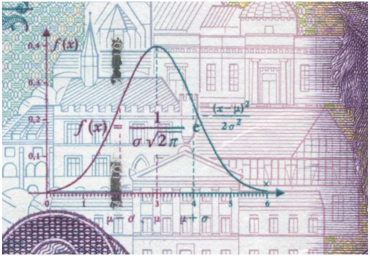
Stetige Verteilungen: Normalverteilung

Und nun die wichtigste Verteilung überhaupt:

Beispiel: Normalverteilung

Eine Zufallsvariable heißt normalverteilt mit Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$, geschrieben $X \sim \mathsf{N}(\mu, \sigma^2)$, wenn sie folgende Dichte besitzt

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right), \qquad t \in \mathbb{R}.$$



Es gilt

$$\mathbf{E}[X] = \mu, \qquad \mathbf{Var}\, X = \sigma^2.$$

Zeigen sie, dass im Fall $\mu = 0$, $\sigma = 1$ (Standardnormalverteilung)

$$f_X(t) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-t^2/2}$$

eine Dichte ist, dass also $\int_{-\infty}^{\infty} f_X(t)\, \mathrm{d}t = 1$.

Die Verteilungsfunktion der Standardnormalverteilung wird oft auch mit $\Phi(t)$ bezeichnet, also

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-\tau^2/2}\, \mathrm{d}\tau, \qquad t \in \mathbb{R}.$$

- Weibull-Verteilung[2]: Eine Zufallsvariable X mit Weibull-Verteilung ist charakterisiert durch die Dichte

$$f_X(x) = \begin{cases} \alpha\beta x^{\beta-1}\,\mathrm{e}^{-\alpha x^\beta}, & x > 0, \\ 0, & \text{sonst}, \end{cases} \qquad \alpha, \beta > 0.$$

(Speziallfall: Exponentialverteilung für $\alpha = \lambda, \beta = 1$)

- Prüfverteilungen in der Statistik: χ^2-Verteilung, t-Verteilung, F-Verteilung

$$\chi_n^2 \sim X_1^2 + \cdots + X_n^2, \quad X_k \sim \mathsf{N}(0,1),$$

$$F_{m,n} \sim \frac{\chi_m^2/m}{\chi_n^2/n},$$

$$t_n \sim \frac{X}{\sqrt{\chi_n^2/n}} \qquad X \sim \mathsf{N}(0,1) \qquad \text{Student } t\text{-Verteilung}^3$$

[2] Waloddi Weibull, 1939
[3] William Sealy Gosset a.k.a. Student, 1908

Pareto-Verteilung[4]: Eine Pareto-verteilte Zufallsvariable X mit Minimalwert $x_{\min} > 0$ besitzt die Verteilungsfunktion

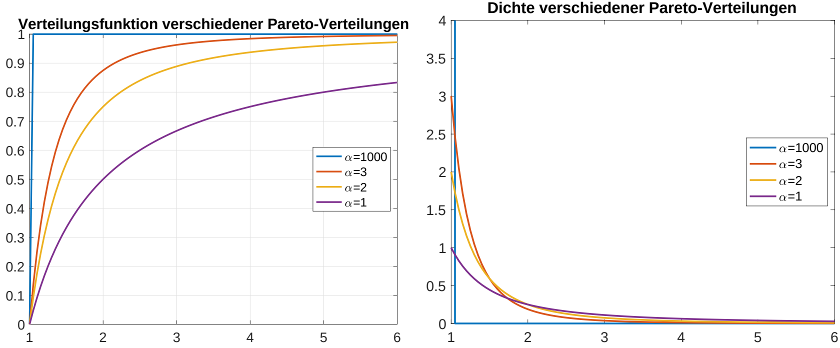
$$F_X(x) = \begin{cases} 0, & x < x_{\min}, \\ 1 - \left(\frac{x_{\min}}{x}\right)^\alpha, & x \geq x_{\min}, \end{cases} \qquad \alpha > 0.$$

Die zugehörige Dichte lautet

$$f_X(x) = \begin{cases} 0, & x < x_{\min}, \\ \frac{\alpha x_{\min}^\alpha}{x^{\alpha+1}}, & x \geq x_{\min}, \end{cases} \qquad \alpha > 0.$$

Einer Pareto-Verteilung folgen Größen wie die Dateigröße im Internetverkehr, Versicherungsschäden, die Größe von Städten, oder das Einkommen von Einzelpersonen einer Gesellschaft.

[4]Vilfredo Pareto, 1897



Pareto-Prinzip: 80% des Vermögens einer Gesellschaft befindet sich im Besitz von 20% der Bevölkerung. Dies wird von der Pareto-Verteilung mit $\alpha = \log_4 5 \approx 1.16$ beschrieben.[5]

[5] Laut Oxfam (2016) besitzen die 62 reichsten Personen der Welt soviel Vermögen wie die ärmste Hälfte der Weltbevölkerung.

- Zur Darstellung der Ungleichheit im Einkommen von Haushalten, welche einer Pareto-Verteilung folgt, wird oft die Lorenz-Kurve[6] verwendet. Hier werden auf der x-Achse alle Haushalte einer Gesellschaft nach aufsteigendem Einkommen angeordnet (0–100%) und auf der y-Achse deren kumulativen Anteil am Gesamteinkommen der Gesellschaft.
- Eine Lorenz-Kurve läßt sich parametrisieren als

$$L(F) = \frac{\int_{x_{\min}}^{x(F)} x f(x) \, \mathrm{d}x}{\int_{x_{\min}}^{\infty} x f(x) \, \mathrm{d}x} = \frac{\int_0^F x(F') \, \mathrm{d}F'}{\int_0^1 x(F') \, \mathrm{d}F'}, \qquad F \in [0,1].$$

  Hierbei bezeichnen f die Dichte sowie $x(F)$ die Umkehrfunktion der Verteilungsfunktion.
- Der Zähler summiert die Einkommen des Anteils F der Haushalte mit niedrigstem Einkommen, der Nenner stellt das Einkommen aller Haushalte dar.
- Bei perfekter Gleichverteilung erhält man als Lorenzkurve die Gerade $y = x$.

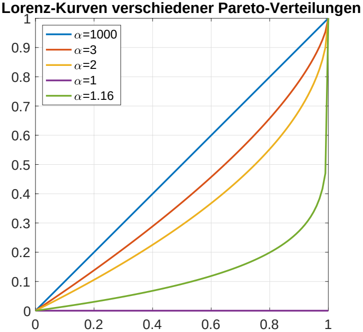
[6] Max O. Lorenz, 1905

Bei der Pareto-Verteilung ergibt sich als Lorenz-Kurve

$$L(F) = 1 - (1-F)^{1-1/\alpha}, \quad F \in [0,1].$$

Bei dem Wert von α, für den sich genau der 80-20 Pareto-Effekt einstellt, muss für die Lorenzkurve gelten

$$L(0.8) = 0.2,$$



d.h. die 80% aller Haushalte mit dem niedrigsten Einkommen verfügen zusammen über gerade einmal 20% des Gesamteinkommens.
Auflösen nach α ergibt

$$\alpha = \log 5 / \log 4 \approx 1.16.$$

Inhalt

- Das Gesetz der großen Zahlen (engl. law of large numbers, LLN) beschreibt das Verhalten bei beliebig ofter Wiederholung desselben Zufallsexperiments:
- Sei hierzu $(X_n)_{n\in\mathbb{N}}$ eine Folge reellwertiger Zufallsvariablen, welche paarweise unabhängig und identisch verteilt sind.
- In der Stochastik bezeichnet man solche Zufallsvariablen kurz i.i.d. (independent and identically distributed).

Satz 15.19 (Starkes Gesetz der großen Zahlen)

Seien $(X_k)_{k\in\mathbb{N}}$ i.i.d. auf $(\Omega, \mathfrak{A}, \mathbf{P})$ sowie $X \in L^1$. Dann gilt

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{fast sicher}} \mathbf{E}\left[X_1\right], \quad \textit{d.h.} \quad \mathbf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} X_k = \mathbf{E}\left[X_1\right]\right) = 1.$$

Bei Folgen $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbf{P})$ unterscheidet man verschiedene Konvergenzarten (gegen eine Zufallsvariable X):

- Fast sichere Konvergenz ist die stochastische Variante von punktweiser Konvergenz von Funktionen: (X_n) konvergiert fast sicher gegen X, in Zeichen $X_n \xrightarrow{\text{f.s.}} X$, falls $\mathbf{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1$.
- Die Folge (X_n) konvergiert in Wahrscheinlichkeit gegen X, in Zeichen $X_n \xrightarrow{\mathbf{P}} X$, falls $\lim_{n\to\infty} \mathbf{P}\{|X_n - X| > \epsilon\} = 0$ für alle $\epsilon > 0$.
- Die Folge (X_n) mit $X_n^p \in L^1$ konvergiert gegen X im p-ten Mittel ($p \geq 1$), in Zeichen $X_n \xrightarrow{L^p} X$, falls $\lim_{n\to\infty} \mathbf{E}\left[(X_n - X)^p\right] = 0$.
- Die schwache Konvergenz oder Konvergenz in Verteilung einer Folge (X_n) ist charakterisiert durch die Konvergenz der zugehörigen Verteilungsfunktionen F_{X_n}: (X_n) konvergiert in Verteilung gegen X, in Zeichen $X_n \xrightarrow{\text{d}} X$, falls

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t) \qquad \text{an allen Stetigkeitspunkten } t \text{ von } F_X.$$

Es gelten stets:

$$\boxed{\text{Konvergenz im } p\text{-ten Mittel}} \Downarrow \boxed{\text{Konvergenz in Wahrscheinlichkeit}} \Downarrow \boxed{\text{Konvergenz in Verteilung}}$$

sowie

$$\boxed{\text{Fast sichere Konvergenz}} \Downarrow \boxed{\text{Konvergenz in Wahrscheinlichkeit}} \Downarrow \boxed{\text{Konvergenz in Verteilung}}$$

Außerdem gilt für $1 \leq p < q$

$$\boxed{\text{Konvergenz im } q\text{-ten Mittel}} \Rightarrow \boxed{\text{Konvergenz im } p\text{-ten Mittel}}.$$

Bemerkung: Beim schwachen Gesetz der großen Zahlen liegt lediglich Konvergenz in Wahrscheinlichkeit vor, allerdings unter schwächeren Voraussetzungen als beim starken Gesetz.

- Eine weitere wichtige asymptotische Aussage, der zentrale Grenzwertsatz (engl. central limit theorem, CLT) zeigt die zentrale Bedeutung der Normalverteilung.
- Sei wieder $(X_k)_{k\in\mathbb{N}}$ eine Folge von i.i.d.-Zufallsvariablen mit $X_k \in L^1$ sowie zusätzlich $X_k^2 \in L^1$.
- Setze $\mu := \mathbf{E}[X_1]$ und $\sigma^2 := \mathbf{Var}\, X_1$.
- Für die Folge $(S_n)_{n\in\mathbb{N}}$ der Zufallsvariablen $S_n := \sum_{k=1}^n X_k$ gilt dann

$$\mathbf{E}[S_n] = n\mu, \qquad \mathbf{Var}\, S_n = n\sigma^2.$$

- Die S_n werden nun standardisiert zu

$$Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}, \qquad n \in \mathbb{N},$$

womit dann $\mathbf{E}[Z_n] = 0$, $\mathbf{Var}\, Z_n = 1$ für alle n.

Satz 15.20 (Zentraler Grenzwertsatz,Lindeberg-Lévy, 1925)

Bezeichnet $\Phi : \mathbb{R} \to [0,1]$ *die Verteilungsfunktion der Standardnormalverteilung, so gilt mit den obigen Bezeichnungen*

$$\lim_{n\to\infty} F_{Z_n}(t) = \Phi(t), \qquad t \in \mathbb{R}.$$

- Mit anderen Worten: Nach geeigneter Normierung verhält sich die Summe von i.i.d. Zufallsvariablen wie die Standard-Normalverteilung!
- Genauer: (Z_n) konvergiert in Verteilung gegen die Standard-Normalverteilung.
- Äquivalent: $(S_n - n\mu)/\sqrt{n} \xrightarrow{\mathsf{d}} \mathsf{N}(0, \sigma^2)$.
- Das Gesetz der große Zahlen beschreibt die Konvergenz des Mittelwerts, der zentrale Grenzwertsatz charakterisiert hingegen präzise die (asymptotische) Streuung um diesen Mittelwert.

- Sei (X_k) eine Folge paarweise unabhängiger Bernoulli-verteilter Zufallsvariable, d.h.

$$\mathbf{P}(X_k = 1) = p, \quad \mathbf{P}(X_k = 0) = 1 - p, \qquad p \in [0, 1].$$

- Die Zufallsvariablen $S_n := X_1 + \cdots + X_n$ folgen dann einer Binomialverteilung mit Parametern n und p, oder kurz $S_n \sim \mathsf{B}(n, p)$. Wie wir schon festgestellt haben, gelten hierfür

$$\mathbf{E}\,[S_n] = np, \qquad \text{sowie } \mathbf{Var}\, S_n = np(1-p).$$

- Anwendung von Satz 15.20 auf die Folge (X_k) bzw. (S_n) führt auf

Folgerung 15.21 (de Moivre (1733), Laplace (1812))

Seien $(X_k)_{k\in\mathbb{N}}$ unabhängig und $X_k \sim \mathsf{B}(n, p)$. Für die umskalierte Folge

$$Z_n := \frac{S_n - np}{\sqrt{np(1-p)}} \quad \textit{gilt} \quad \lim_{n\to\infty} F_{Z_n}(t) = \Phi(t), \quad t \in \mathbb{R}.$$

- Satz 15.21 ist somit ein Spezialfall des zentralen Grenzwertsatzes.
- Für große n kann also die (stetige) Normalverteilung als Approximation für die (diskrete) Binomialverteilung herangezogen werden.
- Erste Formulierung (für $p = 1/2$) durch Abraham de Moivre, später aufgenommen in 2. Band seiner Monografie *The Doctrine of Chances*, 1738.
- Verallgemeinerung auf $p \in [0, 1]$ durch Pierre-Simon Laplace in *Théorie Analytique des Probabilités*, 1812.

Inhalt

- Ziel der Statistik ist es, Daten übersichtlich aufzubereiten und Schlussfolgerungen aus ihnen zu ziehen.
- Man unterscheidet beschreibende und beurteilende Statistik
- Ursprünglich eher reine Buchführung hat sich die Statistik im 19./20. JH zunehmend die Methoden und Ergebnisse der Wahrscheinlichkeitstheorie zunutze gemacht.
- Angesichts der rasant wachsenden Datenfülle in heutiger Zeit kommt der Statistik eine enorme Bedeutung zu, sie bildet beispielsweise den mathematischen Kern bei den aktuellen Methoden der künstlichen Intelligenz.
- Auch für die Wahrnehmung und Interpretation unserer Welt kann man Statistik als elementare Kulturtechnik ansehen. Siehe hierzu etwa die Unstatistik des Monats am Max-Planck-Institut für Bildungsforschung, Berlin.

Sichtweise: bei der Statistik liegt typischerweise eine sehr große Grundgesamtheit vor, bei der gewisse Merkmale unterschiedlich oft auftreten. Die Verteilung solcher Merkmale wird als Zufallsvariable X modelliert.

Definition 15.22

Sei X eine Zufallsvariable. Unter einer Stichprobe vom Umfang n aus X versteht man den Vektor $(X_1, X_2, \dots, X_n)$ aus Zufallsvariablen, dessen Komponenten $X_1, X_2, \dots, X_n$ paarweise unabhängig sind und alle die gleiche Verteilung wie X besitzen.

Die Realisierung einer Stichprobe, also der Wert $X_1(\omega), \dots, X_n(\omega)$ für ein bestimmtes festes $\omega^\star \in \Omega$, auch konkrete Stichprobe genannt, wird bezeichnet mit

$$(x_1, \dots, x_n) = (X_1(\omega^\star), \dots, X_n(\omega^\star)).$$

Das Elementarereignis $\omega^\star$ wählt sozusagen die konkrete Belegung des Stichprobenvektors aus, etwa bei einer Umfrage die konkrete Auswahl von n Befragten.

Sei X diskrete Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbf{P})$ mit Werten $X(\Omega) = \{x_1^*, x_2^*, \dots, x_k^*\}$ (aufsteigend geordnet) sowie Wahrscheinlichkeiten

$$\mathbf{P}(X = x_i^*) = p_i > 0, \qquad i = 1, \dots, k.$$

Eine konkrete Stichproble $x_1, \dots, x_n$ (mit $n > k$) aus X enthalte

$$n_i \text{ Mal den Wert } x_i^*, \quad i = 1, \dots, k, \qquad \text{sodass } n_1 + \cdots + n_k = n.$$

Die relativen Häufigkeiten $\frac{n_i}{n}, i = 1, \dots, k$, können als Schätzung für die Wahrscheinlichkeiten p_i herangezogen werden, welche im Allgemeinen besser werden mit wachsendem n. Man spricht von der empirischen Häufigkeitsdichte

$$\left(x_1^*, \frac{n_1}{n}\right), \dots, \left(x_k^*, \frac{n_k}{n}\right)$$

als Approximation der Verteilung

$$(x_1^*, p_1), \dots, (x_k^*, p_k).$$

Mit den Summenhäufigkeiten

$$s_1 = \frac{n_1}{n}, \quad s_2 = s_1 + \frac{n_2}{n}, \quad s_k = s_{k-1} + \frac{n_k}{n} = 1$$

ergibt sich die empirische Verteilungsfunktion der Stichprobe

$$\hat{F}_{X,n}(t) := \begin{cases} 0, & t < x_1^*, \\ s_i, & x_i^* \le t < x_{i+1}^*,\ i = 1, \dots, k-1, \\ 1, & t \ge x_k^*, \end{cases} \qquad t \in \mathbb{R}.$$

Bestimmen Sie Häufigkeitsverteilung und empiriche Verteilungsfunktion folgender Stichprobe vom Umfang $n = 10$:

i	x_i^*	n_i
1	1	2
2	3	4
3	5	1
4	6	3

Sei X eine stetige Zufallsvariable mit (unbekannter) Dichte f_X.

- Eine konkrete Stichproble $x_1, \dots, x_n$ besteht im Allgemeinen aus n *verschiedenen* Werten x_i.
- Hier bildet man zur Beschreibung der Häufigkeitsverteilung unter der impliziten Annahme $X(\Omega) \in [\xi_1, \xi_{m+1}]$ Klassen[7] mit Intervallen

$$K_i = (\xi_i, \xi_{i+1}], \quad i = 1, \dots, m; \qquad \xi_1 < \xi_2 < \dots < \xi_{m+1}.$$

Somit gilt

$$\mathbf{P}(X \in K_i) = \mathbf{P}(\xi_i < X \leq \xi_{i+1}) = \int_{\xi_i}^{\xi_{i+1}} f_X(x) \, \mathrm{d}x.$$

- Die Klassen K_i spielen bei der stetigen Zufallsvariable X dieselbe Rolle wie die k möglichen Werte x_i^* im diskreten Fall.

[7] Man spricht auch von Behältern (engl. bins, binning) oder von Quantisierung.

- Analog zum diskreten Fall: bestimme Anzahl Stichprobenwerte in K_i.

$$
\begin{aligned}
&n_i && \text{: Anzahl Stichprobenwerte in } K_i \\
&\mathbf{P}(X \in K_i) \approx \frac{n_i}{n} && \text{: relative Häufigkeit der Stichprobenwerte in } K_i
\end{aligned}
$$

- Angenommen, Wahrscheinlichkeitsdichte f_X von X konstant in jedem K_i und besitzt dort den Wert am Mittelpunkt $\tilde{x}_i := \frac{1}{2}(\xi_i + \xi_{i+1})$, d.h.

$$f_X(x) = f_X(\tilde{x}_i), \quad x \in (\xi_i, \xi_{i+1}] \;\Rightarrow\; \frac{n_i}{n} \approx f_X(\tilde{x}_i)(\xi_{i+1} - \xi_i)$$

- Histogramm: empirische Dichte der Stichprobe, beziehe relative Häufigkeit auf Klassenbreite $\xi_{i+1} - \xi_i$:

$$f_X(\tilde{x}_i) \approx \hat{f}_{X,n}(\tilde{x}_i) := \frac{n_i}{n} \cdot \frac{1}{\xi_{i+1} - \xi_i}$$

- Auftragen über Klassenmittelpunkten $\tilde{x}_i$ oder über äquidistenten Indices.

- Fläche des Balkens der Höhe $f_X(\tilde{x}_i)$ über Klassenbreite $\xi_{i+1} - \xi_i$ beträgt $\frac{n_i}{n}$.
- Histogramm hängt von der gewählten Klasseneinteilung ab.
- Heuristische Empfehlung: $m \approx \sqrt{n}$, $m \geq 5$.

- Der Median (oder $\frac{1}{2}$-Quantil) $x_{1/2}$ einer Zufallsvariable X ist definiert durch die Beziehung
$$\mathbf{P}(X \leq x_{1/2}) = \mathbf{P}(X > x_{1/2}) = \frac{1}{2}.$$
- Der empirische Median $\hat{x}_{1/2}$ einer Stichprobe ist dadurch bestimmt, dass links und rechts von $\hat{x}_{1/2}$ gleich viele Stichprobenwerte liegen (bei ungeradem n der mittlere Stichprobenwert, bei geradem n das arithmetische Mittel der beiden mittleren Werte).
- Ein empirischer Modalwert einer Stichprobe ist eine Stelle (d.h. einer der Werte einer diskreten Zufallsvariable oder eine Klassenmitte bei stetigen), an dem die relative Häufigkeit ein relatives Maximum annimmt.
- Empirischer Median und empirischer Modalwert sind beides Funktionen der zufälligen Realisierung $(x_1, \ldots, x_n)$ der Stichprobe $(X_1, \ldots, X_n)$, also auch selbst Zufallsvariablen und können bei jeder Realisierung andere Werte annehmen.

Definition 15.23

Sei $(X_1, \dots, X_n)$ eine Stichprobe aus der reellwertigen Zufallsvariable X sowie $g : \mathbb{R}^n \to \mathbb{R}$ eine (messbare) Funktion. Dann heißt $Z(\omega) := (g \circ (X_1, \dots, X_n))(\omega) = g(X_1(\omega), \dots, X_n(\omega))$ Stichprobenfunktion.

- **Beispiele:**

$$\overline{X} := \frac{1}{n}(X_1 + \cdots + X_n), \qquad S^2 := \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$$

- Aus jeder Realisierung $(x_1, \dots, x_n)$ von $(X_1, \dots, X_n)$ ergibt sich eine Realisierung $\overline{x}$ von $\overline{X}(\omega)$ bzw. s^2 von $S^2(\omega)$.
- **Ziel:** Stichprobenfunktion so wählen, dass sie Parameter der Verteilung einer Zufallsvariable X approximiert.

- Sei q ein Parameter der Verteilung einer Zufallsvariable X, für den ein Schätzwert gesucht wird,

$$\text{z.B. } q = \mathbf{E}[X] \qquad \text{oder} \qquad q = \mathbf{Var}\, X.$$

- Eine Stichprobenfunktion $g(X_1, \dots, X_n)$, deren Realisierungen nach Vorliegen einer konkreten Stichprobe als Näherung für den tatsächlichen Wert q geeignet ist, heißt Punktschätzer oder Schätzfunktionfür q.
- Sucht man hingegen nach Intervallen, die den Parameter q in gewissem Sinn (etwa mit vorgegebener Wahrscheinlichkeit) enthalten, spricht man von einer Intervallschätzung (oder allgemeiner Bereichsschätzung).

Definition 15.24

Sei $(X_1, \ldots, X_n)$ eine Stichprobe aus der Zufallsvariable X und q ein Parameter der Verteilung von X. Eine Schätzfunktion $g(X_1, \ldots, X_n)$ für q heißt erwartungstreu, falls

$$\mathbf{E}\left[g(X_1, \ldots, X_n)\right] = q. \tag{15.1}$$

Sie heißt asymptotisch erwartungstreu, falls

$$\lim_{n\to\infty} \mathbf{E}\left[g(X_1, \ldots, X_n)\right] = q.$$

- Interpretation: bei Realisierung von $g(X_1, \ldots, X_n)$ durch eine große Zahl von Stichproben kann man erwarten, dass diese im Allgemeinen nahe an q liegen.
- Implizit: $g(X_1, \ldots, X_n)$ für jede Realisierung definiert, (15.1) gilt für jeden Wert von n. Somit ist eine erwartungstreue Schätzfunktion auch asymptotisch erwartungstreu.
- Ansonsten keine Voraussetzung über die Verteilung von X.
- Erwartungstreue bedeutet keine Aussage über die Güte der Schätzung einzelner Realisierungen.

Satz 15.25

Sei X eine Zufallsvariable mit Erwartungswert $\mathbf{E}[X]$ *sowie* $(X_1, \dots, X_n)$ *eine Stichprobe aus* X*. Dann ist die Schätzfunktion*

$$g(X_1, \dots, X_n) := \overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$$

ein erwartungstreuer Schätzer für $\mathbf{E}[X]$*.*

- $\overline{X}$ heißt auch (mathematisches) Stichprobenmittel (engl. *sample average*).
- Realisierungen $\overline{x}$ von $\overline{X}$ heißen Stichprobenmittelwerte.
- Auch eine einzelne Stichprobe X_1 $(n = 1)$ bildet einen erwartungstreuen Schätzer von $\mathbf{E}[X]$.
- Welchen Vorteil hat der Schätzer $\overline{X}$ mit $n > 1$?

Satz 15.26

Sei X eine Zufallsvariable mit $0 < \sigma_X^2 = \mathbf{Var}\, X < \infty$ und $(X_1, \dots, X_n)$ eine Stichprobe aus X. Für die Streuung $\sigma_{\overline{X}}^2$ von $\overline{X}$ gilt

$$\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}.$$

Definition 15.27

Ein Schätzer $G_1 = g_1(X_1, \dots, X_n)$ für q heißt wirksamer als ein Schätzer $G_2 = g_2(X_1, \dots, X_n)$ für q, falls

$$\mathbf{E}\left[(G_1 - q)^2\right] \leq \mathbf{E}\left[(G_2 - q)^2\right].$$

- Für erwartungstreue Schätzer bedeutet dies $\mathbf{Var}\, G_1 \leq \mathbf{Var}\, G_2$.
- Der Schätzer $\overline{X}$ für $\mathbf{E}[X]$ mit $n = n_1$ ist wirksamer als mit $n = n_2$ falls $n_1 > n_2$.

Sei $X \sim \mathsf{N}(\mu, \sigma^2)$. Wie groß ist, abhängig von n, die Wahrscheinlichkeit, dass eine Realisierung von $\overline{X}$ in einer σ-Umgebung von μ liegt, dass also

$$\mathbf{P}(|\overline{X} - \mu| < \sigma)?$$

- Mit $X_k \sim \mathsf{N}(\mu, \sigma^2), k = 1 \ldots, n$ gilt (vgl. Folie 153)

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) \sim \mathsf{N}\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathsf{N}(0, 1).$$

Satz 15.28

Ist $X \sim \mathsf{N}(\mu, \sigma^2)$, so besitzt der Schätzer $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ für $\mathbf{E}[X]$ die Verteilung $\mathsf{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

$\mathbf{P}(|\overline{X} - \mu| < \sigma)$ in Abhängigkeit von n:

- $n = 1$: Hier ist $\overline{X} = X_1 \sim \mathrm{N}(\mu, \sigma^2)$ und damit

$$\begin{aligned}\mathbf{P}(|X_1 - \mu| < \sigma) = \mathbf{P}(-\sigma < X_1 - \mu < \sigma) = \mathbf{P}\left(-1 < \frac{X_1 - \mu}{\sigma} < 1\right)\\ = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 0.682.\end{aligned}$$

- $n > 1$: Ähnlich schließt man

$$\begin{aligned}\mathbf{P}(|\overline{X} - \mu| < \sigma) = \mathbf{P}\left(-1 < \frac{\overline{X} - \mu}{\sigma} < 1\right) = \mathbf{P}\left(-1 < \frac{\overline{X} - \mu}{\sqrt{n}\sigma_{\overline{X}}} < 1\right)\\ = \mathbf{P}\left(-\sqrt{n} < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < \sqrt{n}\right).\end{aligned}$$

Wegen $\frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \sim \mathrm{N}(0,1)$ folgt nun

$$\mathbf{P}(|\overline{X} - \mu| = \Phi(\sqrt{n}) - \Phi(-\sqrt{n}) = 2\Phi(\sqrt{n}) - 1.$$

Aus einer Tabelle für die Werte der Standardnormalverteilung (oder per Software) erhält man für verschiedene Werte von n folgende Wahrscheinlichkeiten:

n	$\mathbf{P}(\lvert\overline{X} - \mu\rvert)$
1	0,682
2	0,842
3	0,916
4	0,954
5	0,974

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Beispiel für eine Tabelle zum Ablesen von Werten der Verteilungsfunktion $\Phi(z), z \in \mathbb{R}$, der Standardnormalverteilung $\mathsf{N}(0, 1)$.

Satz 15.29

Sei X eine Zufallsvariable mit **Var** $X = \sigma^2 < \infty$ *sowie $(X_1, \dots, X_n)$ eine Stichprobe aus X. Dann ist die Stichprobenfunktion*

$$S^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})^2$$

ein erwartungstreuer Schätzer für σ^2.

- Die Kenntnis von $\mathbf{E}[X]$ wird nicht vorausgesetzt, an dessen Stelle geht die Schätzfunktion $\overline{X}$ für $\mathbf{E}[X]$ in die Schätzfunktion S^2 für σ^2 mit ein.
- Bezeichnungen für die Schätzfunktion S^2: Stichprobenvarianz, Stichprobenstreuung, Stichprobendispersion.

- Der Faktor $\frac{1}{n-1}$ mag zunächst befremdlich erscheinen. Tatsächlich bildet aber die (vielleicht naheliegendere) Schätzfunktion

$$\tilde{S}^2 := \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2 = \frac{n-1}{n}S^2$$

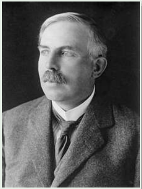
  für σ^2 lediglich einen asymptotisch erwartungstreuen Schätzer, aber keinen erwartungstreuen Schätzer. Es gilt nämlich

$$\mathbf{E}\left[\tilde{S}^2\right] = \frac{n-1}{n}\mathbf{E}\left[S^2\right] = \frac{n-1}{n}\sigma^2 \xrightarrow{n\to\infty} \sigma^2.$$

- Der Übergang von $\tilde{S}^2$ nach S^2 wird auch Bessel-Korrektur genannt.
- Bei *bekanntem* Erwartungswert $\mathbf{E}[X]$ hingegen ist die Schätzfunktion

$$S_0^2 := \frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbf{E}[X])^2$$

  für σ^2 erwartungstreu.



Die Physiker Rutherford und Geiger untersuchten die Emission von α-Teilchen aus einer radioaktiven Substanz. Die Anzahl der in einem bestimmten Zeitintervall emittierten α-Teilchen sei durch eine diskrete Zufallsvariable X modelliert. Rutherford und Geiger stellten fest, dass für Zeitintervalle der Länge 7,5 Sekunden die Zufallsvariable X die 11 Werte $0, 1, \dots, 10$ annehmen kann. Es wurde eine Stichprobe vom Umfang $n = 2608$ untersucht, d.h. es wurden die Werte von X in 2608 Intervallen der Dauer 7,5-Sekunden experimentell ermittelt. Die Anzahl der Zeitintervalle, in denen X den Wert i ($i = 0, 1, \dots, 10$) angenommen hat, sei n_i. Es ist $\sum_{i=1}^{10} n_i = 2608$.

i	0	1	2	3	4	5	6	7	8	9	10	
n_i	57	203	383	525	532	408	273	139	45	27	16	$n = 2068$

Fortsetzung →

1. Bestimmen Sie die empirische Häufigkeitsverteilung $(i, n_i/n)$.
2. Berechnen Sie die Summenhäufigkeiten s_i und die empirische Verteilungsfunktion $\hat{F}_n(x)$ (Schätzung für die Verteilungsfunktion $\mathbf{P}(X \le x)$).
3. Wie groß ist der empirische Median $i_{1/2}$? Geben Sie einen empirischen Modalwert i_m an.
4. Vergleich mit einer Poisson-verteilten Zufallsvariablen Y mit $\mathbf{P}(Y = i) = \frac{\lambda^i}{i!}\,\mathrm{e}^{-\lambda}$: Weil λ der Erwartungswert von Y ist ($\mathbf{E}[Y] = \lambda$), ist es naheliegend, zwecks möglichst guter Approximation von X durch die Poisson-verteilte Größe Y den Parameter λ mit dem Stichprobenmittelwert von X zu identifizieren, d.h.
$$\lambda = \frac{1}{n}\sum_{i=1}^{10} n_i.$$
zu setzen. Bestimmen Sie dieses λ, die daraus folgenden Wahrscheinlichkeiten $\mathbf{P}(Y = i) = \frac{\lambda^i}{i!}\,\mathrm{e}^{-\lambda}$ und vergleichen Sie $\mathbf{P}(Y = i)$ mit den empirischen Häufigkeiten $\frac{n_i}{n}$.

- Auf Folie 173: Wahrscheinlichkeit, dass Punktschätzer $\overline{X}$ für $\mathbf{E}[X]$ für $X \sim \mathsf{N}(\mu, \sigma^2)$ in $(\mu - \sigma, \mu + \sigma)$ enthalten.
- Dabei: μ, σ^2 bekannt. Ohne diese muss man anders verfahren.
- Intervallschätzungen: Unter- und Obergrenze eines Intervalls, welches interessierenden Parameter mit vorgegebener Wahrscheinlichkeit enthält, jeweils als Stichprobenfunktion.

Definition 15.30

Sei X eine Zufallsvariable, $(X_1, \ldots, X_n)$ ein Stichprobe aus X sowie g_u, g_o zwei reellwertige Stichprobenfunktionen mit $g_u(x_1, \ldots, x_n) \leq g_o(x_1, \ldots, x_n)$ für alle $(x_1, \ldots, x_n) \in X(\Omega)^n$. Das zufällige Intervall

$$[G_u(\omega), G_o(\omega)] \quad \text{mit} \quad G_u = g_u(X_1, \ldots, X_n), \quad G_o = g_o(X_1, \ldots, X_n) \tag{15.2}$$

heißt zufälliges, zweiseitiges Konfidenz- oder Vertrauensintervall für den Parameter q zum Konfidenzniveau $1 - \alpha \in [0, 1]$, wenn

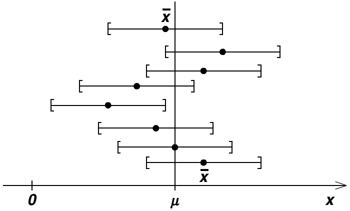
$$\mathbf{P}(q \in [G_u, G_o]) = 1 - \alpha.$$

G_u und G_o heißen untere bzw. obere Konfidenz- oder Vertrauensgrenzen für q.

- Manchmal ist es auch zweckmäßig, zufällige einseitige untere bzw. obere Konfidenzintervalle $[G_u, \infty)$ bzw. $(-\infty, G_o]$ zu betrachten, für die dann entsprechend

$$\mathbf{P}(q \geq G_u) = 1 - \alpha \quad \text{bzw.} \quad \mathbf{P}(q \leq G_o) = 1 - \alpha \text{ gilt .}$$

- Bei Vorliegen einer konkreten Stichprobe aus X wird aus dem zufälligen Intervall $[G_u, G_o]$ ein konkretes Intervall $[\gamma_u, \gamma_o]$, welches von Stichprobe zu Stichprobe variieren wird. Konfidenzniveau $1 - \alpha$ bedeutet, dass bei einer großen Zahl Stichproben $(1 - \alpha) \cdot 100\%$ der konkreten Konfidenzintervalle den wahren Wert q enthalten.



Quelle: Bärwolff (2017)

Ziel: Konfidenzintevalle der Form (15.2) für (unbekannte) Parameter μ, σ^2 einer Zufallsvariablen $X \sim \mathsf{N}(\mu, \sigma^2)$ basierend auf den Punktschätzern $\overline{X}$ und S^2.

Konfidenzintervall für Erwartungswert bei bekannter Varianz

Wie bereits gezeigt: Für Punktschätzer $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ für $\mu = \mathbf{E}[X]$ gelten

$$\mathbf{E}\left[\,\overline{X}\,\right] = \mu, \qquad \mathbf{Var}\,\overline{X} = \frac{\sigma^2}{n}, \qquad \overline{X} \sim \mathsf{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Entsprechend gilt für den standardisierten Stichprobenmittelwert $\overline{X}$

$$Z := \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathsf{N}(0, 1).$$

Sind nun zwei Zahlen $a \leq b$ so bestimmt, dass

$$\mathbf{P}\left(a \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq b\right) = 1 - \alpha,$$

so ist dies äquivalent mit

$$\mathbf{P}\left(\overline{X} - b\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} - a\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \tag{15.3}$$

Mit anderen Worten: aus dem (nichtzufälligen) Intervall $[a, b]$, welches mit Wahrscheinlichkeit $1 - \alpha$ die Zufallsvariable $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ enthält, wurde das zufällige Intervall $[\overline{X} - b\frac{\sigma}{\sqrt{n}}, \overline{X} - a\frac{\sigma}{\sqrt{n}}]$, welches mit derselben Wahrscheinlichkeit die (nichtzufällige) Größe μ enthält.

Die Forderung, dass für $Z \sim \mathsf{N}(0, 1)$ mit Wahrscheinlichkeit $1 - \alpha$ gilt $Z \in [a, b]$ ist äquivalent mit

$$\Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-\frac{\xi^2}{2}} \,\mathrm{d}\xi = 1 - \alpha.$$

Da die Dichte der Standardnormalverteilung symmetrisch zu $\xi = 0$ ist, legen wir den Mittelpunkt des Intervalls auf $\xi = 0$, was auf $a = -b$ führt.

Wegen $\Phi(-x) = 1 - \Phi(x)$ folgt schließlich

$$\Phi(b) - \Phi(-b) = 2\Phi(b) - 1 = 1 - \alpha$$

womit aufgrund der strengen Monotonie von Φ durch

$$\Phi(b) = 1 - \frac{\alpha}{2} \quad \text{und} \quad \Phi(a) = \Phi(-b) = \frac{\alpha}{2} \tag{15.4}$$

a und b eindeutig bestimmt sind.

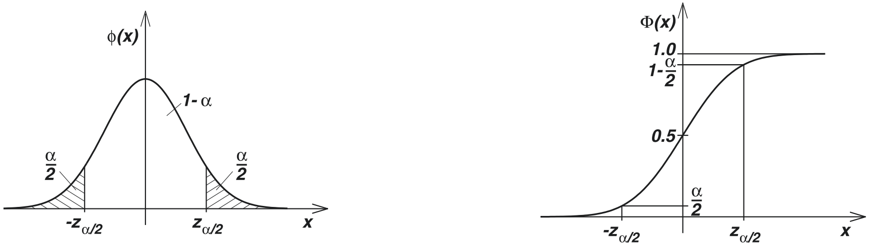
Definition 15.31

Ist F_X Verteilungsfunktion der Zufallsvariablen X und $p \in (0,1)$, so nennt man eine Zahl x_p mit $F_X(x_p) = p$ ein p-Quantil von X. Es gilt $\mathbf{P}(X \leq p) = p$.

Damit ist b das $(1 - \frac{\alpha}{2})$-Quantil einer $\mathsf{N}(0,1)$-verteilten Zufallsvariablen Z. Im Fall der Normalverteilung bezeichnet man allerdings das $(1-\frac{\alpha}{2})$-Quantil nicht mit $z_{1-\frac{\alpha}{2}}$, sondern mit $z_{\frac{\alpha}{2}}$. Für das $\frac{\alpha}{2}$-Quantil a gilt $a = -b$ und man schreibt $-z_{\frac{\alpha}{2}}$.

Es gilt also

$$\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}, \qquad \Phi(-z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}. \tag{15.5}$$



Quelle: Bärwolff (2017)

- Die Indizes der so bezeichneten Quantile liegen zwischen 0 und 1.
- Oft interessiert man sich speziell für die Konfidenzniveaus 90%, 95% und 99%, was den Werten $\alpha = 0,1$, $\alpha = 0,05$ bzw. $\alpha = 0,01$ (alos kleinen Werten von α) entspricht.

Quantile der Standardnormalverteilung zu verschiedenen Konfidenzniveaus (vgl. auch Folie 175):

$\frac{\alpha}{2}$	$1-\frac{\alpha}{2}$	Konfidenzniveau $1-\alpha$	Quantil $z_{\frac{\alpha}{2}}$
0,15	0,85	0,70	1,04
0,10	0,90	0,80	1,29
0,05	0,95	0,90	1,64
0,025	0,975	0,95	1,96
0,005	0,995	0,99	2,58

- Aus (15.3) und (15.4) folgt

$$\mathbf{P}\left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \tag{15.6}$$

- Die beiden Stichprobenfunktionen in (15.6) sind somit gegeben durch

$$G_u = \overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \qquad G_o = \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}.$$

- Für eine Realisierung $\overline{x}$ von $\overline{X}$ erhält man so mit
$$\left[\overline{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \tag{15.7}$$
ein konkretes, zweiseitiges $(1-\alpha)\cdot 100\%$-Konfidenzintervall für den Erwartungswert μ, welches den exakten Wert Mit Wahrscheinlichkeit $1-\alpha$ enthält.
- $1-\alpha$ nennt man daher Sicherheitswahrscheinlichkeit, α entsprechend Irrtumswahrscheinlichkeit.
- Die Lage des Konfidenzintervalls (15.7) ist zufällig.
- Seine Länge hängt im vorliegenden Fall (von bekanntem σ) nur von α und n ab, nicht aber von der Stichprobe.
- Einseitiges unteres $(1-\alpha)\cdot 100\%$-Konfidenzintervall: $\left[\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right)$.
Es gilt nämlich
$$\mathbf{P}\left(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu\right) = \mathbf{P}\left(\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \le z_\alpha\right) = 1-\alpha.$$

- Analog: einseitiges oberes $(1-\alpha)\cdot 100\%$-Konfidenzintervall: $\left(-\infty, \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right]$.
- Dabei ist z_α zwar auch für $\alpha \in (\frac{1}{2}, 1)$ definiert; in der Praxis sind aber, wie bei den zweiseitigen Intervallen, meistens kleine Werte von α von Interesse.

Ein Betrieb stellt Metallzylinder her mit Solldurchmesser $d_{\text{soll}} = 5,10$ cm. In Anbetracht von Fertigungsschwankungen kann der Durchmesser d der produzierten Zylinder als Zufallsvariable modelliert werden. Erfahrung über längere Zeit zeigt, dass für d die Annahme einer Normalverteilung mit Streuung $\sigma = 0,08$ cm gerechtfertigt ist. Aus der aktuellen Produktion wurde eine Stichprobe von 10 Zylindern entnommen. Der daraus bestimmte Stichprobenmittelwert betrug $\overline{d} = 5,14$ cm. In welchem Intervall liegt der mittlere Durchmesser $\mathbf{E}[d]$ aller aktuell produzierten Zylinder mit einer Sicherheit von $70\%, 80\%, 90\%, 95\%$ bzw. 99% ($\sigma = 0,15; 0,10; 0,05; 0,025; 0,005$)? Die Konfidenzintervalle für $\mathbf{E}[d]$ nach (15.7) sind

$$[G_u, G_o] = \left[5,14 - \frac{0,08}{3,162} z_{\frac{\alpha}{2}}, 5,14 + \frac{0,08}{3,162} z_{\frac{\alpha}{2}}\right]$$

Stochastik

Statistik: Konfidenzschätzung Parameter der Normalverteilung

Mit den aus der Tabelle auf Folie 186 ermittelten Werten von $z_{\frac{\alpha}{2}}$ ergibt sich

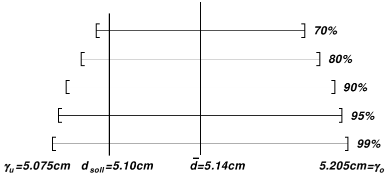
$1-\alpha$	G_u	G_o	$G_o - G_u$
0,70	5,114	5,166	0,052
0,80	5,107	5,173	5,173
0,90	5,099	5,181	0,082
0,95	5,090	5,190	0,100
0,99	5,075	5,205	0,130

Hieraus sieht man z.B., dass

$$5,090 \text{ cm} \leq \mathbf{E}[d] \leq 5,190 \text{ cm}$$

mit einer Wahrscheinlichkeit von $0,95$.
Mit derselben Wahrscheinlichkeit gilt

$$\mathbf{E}[d] - d_{\text{soll}} \in [-0,01 \text{ cm}, 0,09 \text{ cm}].$$



Quelle: Bärwolff (2017)

Variation der Länge $L = 2z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ eines Konfidenzintervalls mit Sicherheitswahrscheinlichkeit $(1-\alpha)$ in Abhängigkeit von α und n:

- Fordert man bei konstantem Stichprobenumfang n eine höhere Sicherheitswahrscheinlichkeit $1-\alpha$, so wächst auch L: dies liegt an der strengen Monotonie von Φ, denn mit wachsendem $1-\alpha$ wächst auch $1-\alpha/2$ und damit auch $z_{\frac{\alpha}{2}}$.
- Bei fester Sicherheitswahrscheinlichkeit $1-\alpha$ und wachsendem Stichprobenumfang n wird die Intervalllänge L kleiner.
- Für eine feste Intervalllänge L betrachte man die Folge α_n mit zugehöriger Längenfolge $L_n = 2z_{\alpha_n/2} \frac{\sigma}{\sqrt{n}}$. Die Forderung $L_m = L_n$ führt auf

$$2z_{\alpha_n/2} \frac{\sigma}{\sqrt{n}} = 2z_{\alpha_m/2} \frac{\sigma}{\sqrt{m}}, \qquad \text{also} \qquad z_{\alpha_m/2} = \sqrt{\frac{m}{n}}\, z_{\alpha_n/2}. \tag{15.8}$$

  Für $m > n$ ist dann $z_{\alpha_m/2} > z_{\alpha_n/2}$ und wegen (15.5) $\alpha_m < \alpha_n$.
  Bei fester Länge L führt also ein größerer Stichprobenumfang zu einer Erhöhung des Konfidenzniveaus.

Praktische Fragestellung: Wie groß muss der Stichprobenumfang n mindestens sein, damit bei vorgegebenem Konfidenzniveau $1-\alpha$ die Breite L des Konfidenzintervalls einen Maximalwert $L_{\max}$ nicht unterschreitet?

Antwort:

$$2z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le L_{\max} \quad \Leftrightarrow \quad n \ge 4\sigma^2 \left(\frac{z_{\alpha/2}}{L_{\max}}\right)^2. \tag{15.9}$$

Schrittweites Vorgehen: Bestimmung eines zweiseitigen Konfidenzintervalls für den Erwartungswert einer normalverteilten Zufallsvariable X bei bekannter Varianz σ^2:

1. Vorgabe des Konfidenzniveaus (Sicherheitswahrscheinlichkeit) $1-\alpha$.
2. Bestimmung von $z_{\alpha/2}$ aus $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$ aus Tabelle für Φ oder Software.
3. Festlegung des Stichprobenumfangs n
   - durch Vorgabe einer Mazimallänge $L_{\max}$ und $n \geq 4\sigma^2 \left(\frac{z_{\alpha/2}}{L_{\max}}\right)^2$
   - oder durch Vorgabe des Stichprobenumfangs n
4. Entnahme einer konkreten Stichprobe $(x_1, \ldots, x_n)$ aus X
5. Berechnung des konkereten Werts $\overline{x} = (x_1 + \cdots + x_n)/n$ des Schätzers $\overline{X}$
6. Das Konfidenzintervall lautet

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Konfidenzintervall für Erwartungswert bei unbekannter Varianz

- Konstruiere Konfidenzintervall für Erwartungswert $\mu = \mathbf{E}[X]$ einer normalverteilten Zufallsvariablen $X \sim \mathsf{N}(\mu, \sigma^2)$ ohne Kenntnis der Varianz σ^2
- Im Fall bekannter Varianz nutzten wir die Eigenschaft der Stichprobenfunktion

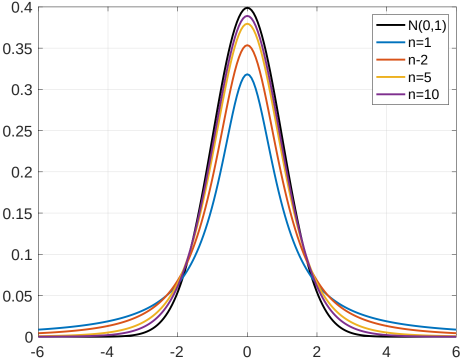
$$\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim \mathsf{N}(0, 1).$$

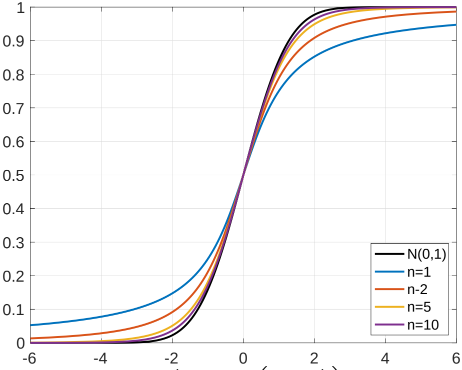
- Naheliegend: ersetze σ durch Punktschätzer S und erhalte so die Schätzfunktion

$$\frac{\overline{X} - \mu}{S}\sqrt{n}. \tag{15.10}$$

- Die Verteilung der Stichprobenfunktion (15.10) hängt nun auch vom Stichprogenumfang n ab, es handelt sich um die Student-t-Verteilung, genauer:
- Ist $(X_1, \dots, X_n)$ eine Stichprobe aus $X \sim \mathsf{N}(\mu, \sigma^2)$, so folgt die Stichprobenfunktion (15.10) einer t-Verteilung mit $n-1$ Freiheitsgraden.

Statistik: Dichte und Verteilungsfunktion der t-Verteilung





$$p_{t;n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

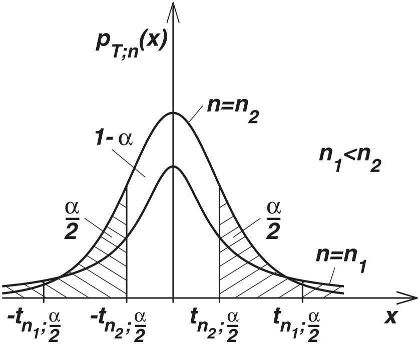
Γ bezeichnet die Gamma-Funktion.

$$F_{t;n}(x) = \frac{1}{2} + x\Gamma\left(\frac{n+1}{2}\right) \cdot \frac{{}_2F_1\left(\frac{1}{2}, \frac{n+1}{2}; \frac{3}{2}; \frac{-x^2}{n}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}$$

${}_2F_1$ bezeichnet die hypergeometrische Funktion.

Folgt die Zufallsvariable T_n einer t-Verteilung mit n Freiheitsgraden, so sind die analog zu (15.5) bezeichneten Quantile $t_{n;\alpha/2}$ sowie $-t_{n;\alpha/2}$ definiert durch

$$\mathbf{P}(-t_{n;\alpha/2} \le T_n \le t_{n;\alpha/2}) = \int_{-t_{n;\alpha/2}}^{t_{n;\alpha/2}} p_{t;n}(x)\,\mathrm{d}x = 1 - \alpha. \tag{15.11}$$



Quelle: Bärwolff (2017)

Quantile der t-Verteilung für zwei verschiedene Werte n_1, n_2 des Parameters n. Um bei einem geringeren Stichprobenunfang $n_1 < n_2$ denselben Flächeninhalt $1-\alpha$ abzudecken steigen die Quantile, d.h. $t_{n_1;\alpha/2} > t_{n_2;\alpha/2}$.

Analog zum Fall bekannter Varianz gehen wir zur Konstruktion eines Konfidenzintervalls für $\mu = \mathbf{E}[X]$ aus von der Forderung

$$\mathbf{P}\left(-t_{n-1;\alpha/2} \leq \frac{\overline{X}-\mu}{S}\sqrt{n} \leq t_{n-1;\alpha/2}\right) = 1-\alpha. \tag{15.12}$$

Umstellen führt hier wieder entsprechend auf

$$\mathbf{P}\left(\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha.$$

Somit ist durch

$$\left[\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right]$$

ein zufälliges zweiseitiges Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha$ gegeben. Nach Auswertung einer konkreten Stichprobe vom Umfang n erhält man mit den Realisierungen $\overline{x}, s$ der Punktschätzer $\overline{X}$ bzw. S das Intervall

$$\left[\overline{x} - t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} + t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\right]. \tag{15.13}$$

- Wie im Fall der Normalverteilung sind die Quantile $t_{n-1;\alpha/2}$ tabelliert bzw. werden von statistischer Software bereitgestellt.
- Für große n gilt $t_{n-1;\alpha/2} \approx z_{\alpha/2}$.
- Die Länge des Konfidenzintervalls ist $L = 2t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}$, in diesem Fall eine Zufallsvariable.
- Eine untere Schranke für den Stichprobenumfang analog zu (15.9) um eine Höchstbreite des Intervalls sicherzustellen ist hier nicht möglich aufgrund der Zufallsvariablen S, deren Wert s von der konkreten Realisierung $(x_1, \dots, x_n)$ der Stichprobe abhängt.

Stichprobe vom Umfang $n = 30$ einer normalverteilten Zufallsvariable ergibt Werte der Punktschätzer $\overline{X}$ und S^2 von

$$\overline{x} = 0,0433, \qquad s^2 = 1,1439.$$

Man bestimme konfidenzintervalle für μ mit Konfidenzniveau $80\%, 90\%, 95\%$ und 99%.

Lösung: Es ist $n - 1 = 29$. Das Konfidenzintervall gemäß (15.13) lautet somit

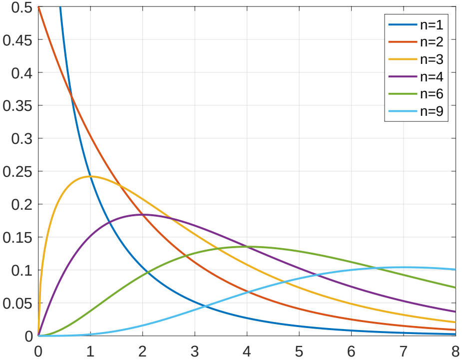
$$[G_u, G_o] = \left[0,0433 - t_{29;\alpha/2}\frac{1,0695}{5,4772}, 0,0433 + t_{29;\alpha/2}\frac{1,0695}{5,4772}\right]$$

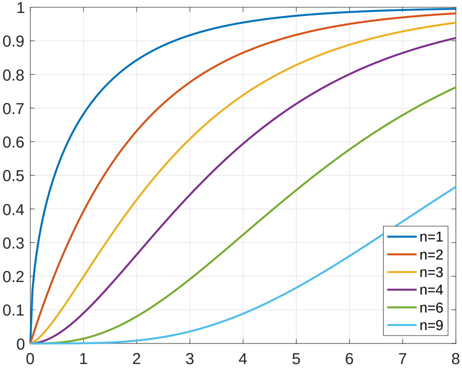
Ein kleines Matlab-Programm liefert folgende Tabelle

$1-\alpha$	$\alpha/2$	$t_{n-1;\alpha/2}$	G_u	G_o
0.80	0.100	1.313	-0.217	0.304
0.90	0.050	1.701	-0.295	0.381
0.95	0.025	2.048	-0.364	0.450
0.99	0.005	2.763	-0.506	0.592

Konfidenzintervall für Varianz

- Zur Bestimmung von σ^2 einer Stichprobe aus $X \sim \mathsf{N}(\mu, \sigma^2)$ wird als Stichprobenfunktion verwendet
$$\frac{(n-1)S^2}{\sigma^2}. \tag{15.14}$$
- Man kann zeigen: ist S^2 Stichprobenvarianz einer Stichprobe aus $X \sim \mathsf{N}(\mu, \sigma^2)$, so ist diese Größe Chi-Quadrat-verteilt (χ^2-verteilt) mit $n-1$ Freiheitsgraden.
- Allgemein ist die χ^2-Verteilung mit n Freiheitsgraden definiert als Verteilung der Summe χ_n^2 der Quadrate von n unabhängigen $\mathsf{N}(0,1)$-verteilten Zufallsvariablen $X_1, \dots, X_n$:
$$\chi_n^2 = X_1^2 + \cdots + X_n^2. \tag{15.15}$$
- Aus (15.15) folgt unmittelbar
$$\mathbf{E}\left[\chi_n^2\right] = n, \qquad \mathbf{Var}\, \chi_n^2 = 2n.$$

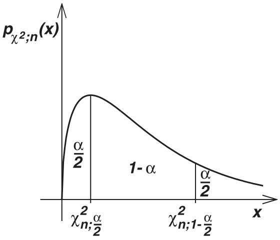




$$p_{\chi^2;n}(x) = \frac{x^{\frac{n}{2}-1}\,\mathrm{e}^{-\frac{x}{2}}}{2^{\frac{n}{2}}\,\Gamma(\frac{n}{2})}, \qquad x \geq 0, \quad n \in \mathbb{N}.$$

$$F_{\chi^2;n}(x) = P\left(\frac{n}{2}, \frac{x}{2}\right)$$

P bezeichnet hier die regularisierte unvollständige Gammafunktion.



Quelle: Bärwolff (2017)

- Die Dichtefunktion $p_{\chi^2;n}$ ist nicht symmetrisch bezüglich des Mittelwertes und verschwindet für $x < 0$.
- Daher werden die Quantile jetzt in Übereinstimmung mit der allgemeinen Definition 15.31 definiert und bezeichnet.
- Es kommen daher Quantile sowohl ≈ 0 als auch ≈ 1 in Betracht:
- $$\mathbf{P}\left(\chi^2_{n;\alpha/2} \leq \chi^2_n \leq \chi^2_{n;1-\alpha/2}\right) = \int_{\chi^2_{n;\alpha/2}}^{\chi^2_{n;1-\alpha/2}} p_{\chi^2_n}(x)\,\mathrm{d}x = 1 - \alpha.$$
- Da die Zufallsvariable (15.14) einer χ^2-Verteilung mit $n-1$ Freiheitsgraden genügt, ergibt sich durch Umstellen hier

$$\mathbf{P}\left(\chi^2_{n-1;\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1;1-\alpha/2}\right) = \mathbf{P}\left(\frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}\right)$$

- Somit ist
$$\left[\frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}\right]$$
ein zweiseitiges Konfidenzintervall für σ zum Konfidenzniveau $1-\alpha$.
- Nach Ziehen einer konkreten Stichprobe ist das konkrete Intervall gegeben durch
$$\left[\frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}\right].$$
- Wenn α fällt wird $\chi^2_{n-1;\alpha/2}$ kleiner und $\chi^2_{n-1;1-\alpha/2}$ größer.

Konfidenzintervall für die Wahrscheinlichkeit von Ereignissen

- **Ziel:** Quantifizierung der Genauigkeit, mit der die Eintrittswahrscheinlichkeit $p(A)$ eines Ereignisses A approximiert werden kann anhand beobachteter relativer Häufigkeiten $H_n(A)$ bei n Zufallsexperimenten.
- Betrachte $p = p(A)$ als Verteilungsparameter der Indikatorvariablen

$$X(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 1, & \text{falls } A \text{ eingetreten ist,} \\ 0, & \text{sonst.} \end{cases}$$

- Die Wahrscheinlichkeitsverteilung von $X = \mathbb{1}_A$ ist gegeben durch

$$\mathbf{P}(X = 1) = p, \qquad \mathbf{P}(X = 0) = 1 - p,$$

(Bernoulli-Schema) und somit

$$\mathbf{E}[X] = p, \qquad \mathbf{Var}\, X = p(1 - p).$$

- Es ist also ein Konfidenzintervall $[p_u, p_o]$ zu bestimmen, welches mit vorgegebener Wahrscheinlichkeit p enthält.

- Entnahme einer Stichprobe $(X_1, \ldots, X_n)$: entspricht n-maligem Ziehen (mit Zurücklegen) aus einer Urne mit $p \cdot 100\%$ der Kugeln gekennzeichnet mit einer „1“, die restlichen $(1-p) \cdot 100\%$ mit einer „0“.
- Realisierung dargestellt als binäre Folge, z.B.

$$(x_1, \ldots, x_n) = (0, 1, 1, 0, 0, 1, ..., 1, 0).$$

- # Eintritte von A bei n Versuchen:

$$Z_n = \sum_{k=1}^{n} X_k, \quad \text{Realisierung:} \quad z_n = \sum_{k=1}^{n} x_k.$$

Relative Eintrittshäufigkeit von A bei n Versuchen

$$\overline{X}_n = \frac{1}{n} Z_n = \frac{1}{n} \sum_{k=1}^{n} X_k, \qquad \text{Realisierung:} \quad H_n(A) := \frac{1}{n} z_n.$$

- Da $(X_k)_{k=1}^n$ i.i.d. erlaubt der zentrale Grenzwertsatz asymptotische Aussagen (große n).
- Wegen $\mathbf{E}[X_k] = p$ und $\mathbf{Var}\, X_k = p(1-p)$ für alle k gilt für große n

$$\frac{\overline{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n} \sim \mathsf{N}(0,1).$$

Wir erhalten somit ein (approximatives) Konfidenzintervall zum Konfidenzniveau $1-\alpha$ aus der Forderung

$$\mathbf{P}\left(-z_{\alpha/2} \leq \frac{\overline{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n} \leq z_{\alpha/2}\right) = 1-\alpha.$$

- Nach einigen mühseligen Umformungen erhalten wir $[G_u, G_o]$ mit

$$G_u = \frac{1}{1+\frac{1}{n}z_{\alpha/2}^2}\left[\frac{1}{2n}z_{\alpha/2}^2 + \overline{X}_n - \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{1}{4n}z_{\alpha/2}^2 + \overline{X}_n(1-\overline{X}_n)}\right],$$

$$G_o = \frac{1}{1+\frac{1}{n}z_{\alpha/2}^2}\left[\frac{1}{2n}z_{\alpha/2}^2 + \overline{X}_n + \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{1}{4n}z_{\alpha/2}^2 + \overline{X}_n(1-\overline{X}_n)}\right].$$

- Das entsprechende konkrete Konfidenzintervall $[p_u, p_o]$ ergibt sich durch Ersetzen von $\overline{X}_n$ durch $H_n(A)$.
- Als Näherung für große n erhalten wir

$$p_u = H_n(A) - \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{H_n(A)(1 - H_n(A))},$$

$$p_o = H_n(A) + \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{H_n(A)(1 - H_n(A))}.$$

Der französische Mathematiker und Naturforscher Georges-Louis Leclerc, Comte de Buffon (1707–1788) führte Münzwurfversuche durch und fand bei $n = 4040$ Versuchen eine relative Häufigkeit von

$$H_{4040}(A) = 0,5080.$$

Man bestimme ein 90%-Konfidenzintervall für $p(A)$. Hinweis: $z_{0,05} = 1,64$.

- Statistische Tests: dienen der Prüfung statistischer Hypothesen.
- Statistische Hypothese: Annahme oder Mutmaßung über die Verteilung einer Zufallsvariable.
- Parameterhypothese: Verteilung bekannt bis auf Parameterwerte, Annahme H_0 über Wert des Parameters q, etwa $H_0 : \mu = 2$ bei einer Verteilung $\mathsf{N}(\mu, \sigma^2)$. Zugehörige Tests heißen Parametertests.
- Vollständiges Testproblem: neben Hypothese H_0 auch Alternativhypothese H_1, teste „H_0 gegen H_1“.
- Zweiseitige Tests: z.B. $H_0 : q = q_0, H_1 : q \neq q_0$.
- Einseitige Tests: z.B. $H_0 : q = q_0$, $H_1 : q < q_0$.
- Ist H_1 die zu H_0 komplementäre Hypothese, so wird sie oft nicht explizit angegeben.
- Solche Parametertests heißen auch Signifikanztests.
- Ist auch die Form der Verteilung nicht bekannt, so spricht man von nichtparametrischen Tests oder von Anpassungstests.

Beispiel: Für eine Zufallsvariable X (Merkmal einer Grundgesamtheit) sei bekannt, dass sie $\mathsf{N}(\mu, 1)$-verteilt ist. Hypothese:

$$H_0 : \ \mu = 0.$$

Annahme oder Ablehnung von H_0 anhand einer Stichprobe aus X. Ablehnung bedeutet Annahme der Alternativhypothese $H_1 : \mu \neq 0$.

Beim vorliegenden Beispiel naheliegend: Verwendung von $\overline{X}$. Im Zusammenhang mit Parametertests heißen zufällige Stichprobenfunktionen Testfunktionen, deren Realisierungen Testgrößen.

Entnahme einer Stichprobe aus X vom Umfang $n = 30$ ergibt $\overline{x} = 0,04$.

Idee: Zur Entscheidung über $H_0 : \mu = 0$ angesichts von $\overline{x} = 0,04$ berechnen wir die Wahrscheinlichkeit $\mathbf{P}(|\overline{X} - 0| \geq 0,04)$.

Angenommen H_0 ist richtig. In diesem Fall gilt $\overline{X} \sim \mathsf{N}\left(0, \frac{1}{30}\right)$.
Äquivalent: $\sqrt{30}\,\overline{X} \sim \mathsf{N}(0,1)$ und damit

$$\begin{aligned}
\mathbf{P}(|\overline{X}| \geq 0,04) &= \mathbf{P}(|\overline{X}\sqrt{30}| \geq 0,04\sqrt{30}) \\
&= \mathbf{P}(\overline{X}\sqrt{30} \leq -0,04\sqrt{30}) + \mathbf{P}(\overline{X}\sqrt{30} \geq 0,04\sqrt{30}) \\
&= \Phi(-0,04\sqrt{30}) + \left[1 - \Phi(0,04\sqrt{30})\right] \\
&= 2\left[1 - \Phi(0,04\sqrt{30})\right] \approx 0,826.
\end{aligned}$$

Mit einer Wahrscheinlichkeit von etwa $82,6\%$ weicht also im Falle $\mu = 0$ bei einer Stichprobe vom Umfang $n = 30$ der Stichprobenmittelwert $\overline{X}$ um mindestens $0,04$ von $\mu = 0$ ab.

H_0 wird man in diesem Fall nur schwer verwerfen können.

Die gleiche Rechnung bei einer Realisierung $\overline{x} = 0,38$ ergibt

$$\begin{aligned}\mathbf{P}(|\overline{X}| \geq 0,38) &= \mathbf{P}(|\overline{X}\sqrt{30}| \geq 0,38\sqrt{30}) \\ &= 2\left[1 - \Phi(0,38\sqrt{30})\right] \approx 0,038.\end{aligned}$$

Wenn H_0 gilt, liegt die Wahrscheinlichkeit, dass sich $\overline{X}$ von $\mu = 0$ um mindestens $0,38$ unterscheidet also nur bei etwa $3,8\%$.

Intuitiv wird man im ersten Fall ($\overline{x} = 0,04$) H_0 akzeptieren und im zweiten Fall ($\overline{x} = 0,38$) ablehnen. In beiden Fällen kann man sich hierbei jedoch mit positiver Wahrscheinlichkeit irren.

Zur Entscheidung wird wir eine kritische Wahrscheinlichkeit, Irrtumswahrscheinlichkeit oder Signifikanzniveau α vorgegeben. Entsprechend heißt $1-\alpha$ Sicherheitswahrscheinlichkeit. Typische Werte: $\alpha = 0,001$, $\alpha = 0,01$, $\alpha = 0,05$, $\alpha = 0,1$.

Allgemeines Vorgehen bei Parametertests:

1. Zu prüfen: Hypothese $H_0 : q = q_0$ über Verteilungsparameter q einer Zufallsvariable (Grundgesamtheit) X.
2. Berechne Wert $t = g(x_1, \dots, x_n)$ einer Stichprobenfunktion $T = g(X_1, \dots, X_n)$ (Testgröße) angewandt auf konkrete Stichprobe.
3. Bestimme zu Signifikanzniveau α Zahl t_0 sodass $\mathbf{P}(|T| \geq t_0) \leq \alpha$. Das kleinste solche t kann bei stetigen Zufallsvariablen aus $\mathbf{P}(|T| \geq t_0) = \alpha$ bestimmt werden.
4. $W_1 := \{t : |t| \geq t_0 \}$ heißt kritischer Bereich oder Ablehnungsbereich. Im Fall $|t| \geq t_0$ Ablehung von H_0 mit Irrtumswahrscheinlichkeit α.
5. $W_0 := \{t : |t| < t_0\}$ heißt Annahmebereich (besser: Nichtablehnungsbereich) für H_0.
6. Interpretation: Ereignis $\{|T| \geq t_0\}$, welches bei Gültigkeit von H_0 nur mit Wahrscheinlichkeit α eintritt, wird als unmöglich behandelt.

Im obigen Beispiel:

- $X \sim \mathsf{N}(\mu, 1)$, als Testfunktion kann die (dimensionslose) Größe $T = \frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ verwendet werden.
- Gilt $H_0 : \mu = 0$ wird wegen $\sigma = 1$ hieraus $T = \overline{X}\sqrt{n}$.
- Gesucht: t_0 mit $\mathbf{P}(|\overline{X}\sqrt{n}| \geq t_0) = \alpha$.
- Wegen $\overline{X}\sqrt{n} \sim \mathsf{N}(0, 1)$ gilt

$$\mathbf{P}(|\overline{X}\sqrt{n}| \geq t_0) = 2[1 - \Phi(t_0)] \stackrel{!}{=} \alpha, \qquad \text{also } \Phi(t_0) = 1 - \frac{\alpha}{2}.$$

- Somit ist $t_0 = z_{\alpha/2}$ das $\left(1 - \frac{\alpha}{2}\right)$-Quantil der $\mathsf{N}(0, 1)$-Verteilung.
- Ablehnungsbereiche für $\alpha = 0,01$ und $\alpha = 0,05$:

α	t_0	W_1
$0,01$	$2,58$	$\{t : \|t\| \geq 2,58\}$
$0,05$	$1,96$	$\{t : \|t\| \geq 1,96\}$

- Fazit: im Beispiel $H_0 : \mu = 0$ mit $\overline{x} = 0,38$

$$\alpha = 0,05 \quad t_0 = 1,96 < 2,081 = |t| = |\overline{x}|\sqrt{30} \quad \text{(ablehnen)}$$
$$\alpha = 0,01 \quad t_0 = 2,58 > 2,081 = |t| = |\overline{x}|\sqrt{30} \quad \text{(nicht ablehnen)}$$

- Fazit: im Beispiel $H_0 : \mu = 0$ mit $\overline{x} = 0,04$

$$\alpha = 0,05 \quad t_0 = 1,96 > 0,219 = |t| = |\overline{x}|\sqrt{30} \quad \text{(nicht ablehnen)}$$
$$\alpha = 0,01 \quad t_0 = 2,58 > 0,219 = |t| = |\overline{x}|\sqrt{30} \quad \text{(nicht ablehnen)}$$

- In den letzten 3 Fällen widerspricht die Beobachtung also H_0 nicht.

- Idealer Test: Hypothese genau dann ablehnen, wenn sie falsch ist.
  (Ist mit Stichproben nicht erreichbar.)
- Unterscheidung zwischen

  Fehler 1. Art: Ablehnung von H_0, obwohl diese richtig.
  Testgröße t im kritischen Bereich, auch bei Gültigkeit von H_0 nicht ausgeschlossen. Wahrscheinlichkeit von Fehler 1. Art: α.

  Fehler 2. Art: Annahme von H_0, obwohl diese falsch.
  Testgröße t nicht im kritischen Bereich, obwohl H_0 falsch ist.
- Kombinationen:

	Annahme von H_0	Ablehnung von H_0
H_0 richtig	richtige Entscheidung	Fehler 1. Art (falscher Alarm)
H_0 falsch	Fehler 2. Art (versäumter Alarm)	richtige Entscheidung

Sei $X \sim \mathsf{N}(\mu, \sigma^2)$. Zu testen:

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0, \qquad \text{bei Signifikanzniveau } \alpha.$$

- Testfunktion (vgl. Konfidenzintervall bei unbekannter Varianz):

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

- Gilt H_0, so folgt T einer t-Verteilung mit $n-1$ Freiheitsgraden.
- Sei $t = \frac{\overline{x} - \mu_0}{s} \sqrt{n}$ die Testgröße aus einer konkreten Stichprobe.
- Annahme- und Ablehnungsbereiche:

$$W_0 = (-t_{n-1;\alpha/2}, t_{n-1;\alpha/2}), \qquad W_1 = (-\infty, -t_{n-1;\alpha/2}] \cup [t_{n-1;\alpha/2}, \infty)$$

- H_0 wird mit Irrtumswahrscheinlichkeit α abgelehnt, wenn $t \in W_1$, d.h. wenn $|t| \geq t_{n-1;\alpha/2}$.

Seien $X \sim \mathsf{N}(\mu_X, \sigma^2)$, $Y \sim \mathsf{N}(\mu_Y, \sigma^2)$ unabhängig, σ unbekannt. Zu testen:

$$H_0 : \mu_X = \mu_Y \qquad H_1 : \mu_X \neq \mu_Y .$$

- Zufällige Stichproben $(X_1, \dots, X_{n_X})$, $(Y_1, \dots, Y_{n_Y})$ aus beiden Grundgesamtheiten.
- Testfunktion

$$T_{XY} := \frac{\overline{X} - \overline{Y}}{\sqrt{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}} \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}}$$

- Gilt H_0, so folgt T einer t-Verteilung mit $n_X + n_Y - 2$ Freiheitsgraden.
- Aus zwei konkreten Stichproben: $\overline{x}, \overline{y}, s_X^2, s_Y^2$, Testgröße

$$t_{XY} := \frac{\overline{x} - \overline{y}}{\sqrt{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}} \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}}$$

- Bei zweiseitigem Test: Ablehnung von H_0 mit Irrtumswahrscheinlichkeit α, wenn
$$t_{XY} \in (-\infty, -t_{n_X+n_Y-2;\alpha/2}] \cup [t_{n_X+n_Y-2;\alpha/2}, \infty).$$
- Bei einseitigem Test
$$H_0 : \mu_X = \mu_Y, \qquad H_1 : \mu_X < \mu_Y$$
Ablehnungsbereich gegeben durch $W_1 = (-\infty, -t_{n_X+n_Y-2;\alpha/2}]$.
- Bei einseitigem Test
$$H_0 : \mu_X = \mu_Y, \qquad H_1 : \mu_X > \mu_Y$$
Ablehnungsbereich gegeben durch $W_1 = [t_{n_X+n_Y-2;\alpha/2}, \infty)$.
- In allen Fällen:
Ablehnung von H_0 mit Irrtumswahrscheinlichkeit α falls $t_{XY} \in W_1$.

- Keine parametrisierte Verteilungsfamilie von vorn hinein bekannt.
- Teste Vermutung, dass Stichprobe aus einer Verteilung mit Verteilungsfunktion F_0 stammt.
- χ^2-Anpassungstest: prüft, ob für eine Stichprobe aus X, mit unbekannter Verteilungsfunktion F, eventuell $F = F_0$ für gegebene Verteilungsfunktion F_0.
- Somit:

$$H_0 : F = F_0, \qquad H_1 : F \neq F_0, \qquad \text{bei Signifikanzniveau } \alpha$$

- Bei anderen Varianten des χ^2-Tests kann F_0 noch von weiteren Parametern abhängen.

Vorgehen:

1. Stichprobe $(X_1, \ldots, X_n)$ aus X, Realisierung $(x_1, \ldots, x_n)$.
2. Teile Bildbereich $W_X := X(\Omega)$ ein in m disjunkte Klassen:

   $$W_X = \Delta_1 \cup \cdots \cup \Delta_m, \qquad \Delta_i \cap \Delta_j = \varnothing \quad (i \neq j).$$

   Dann bilden die Ereignisse

   $$X \in \Delta_1, \ldots, X \in \Delta_m$$

   eine Zerlegung von Ω. Bei stetigem X, wähle Klassen als Intervalle

   $$\Delta_1 = (-\infty, a_1], \Delta_2 = (a_1, a_2], \ldots, \Delta_{m-1} = (a_{m-2}, a_{m-1}], \Delta_m = (a_{m-1}, \infty)$$

   Bei diskretem X mit endlich vielen Werten ist jede Zerlegung der Bildmenge möglich.

3. Sei $p_i := \mathbf{P}(X \in \Delta_i), i = 1, \dots, m$. Gilt H_0 und sind die Klassen halboffene Intervalle, so ist

$$p_i = F_0(a_i) - F_0(a_{i-1}), \qquad i = 1, \dots, m; \qquad a_0 = -\infty,\ a_m = \infty.$$

Bei diskretem X kann H_0 auch als Hypothese über die Wahrscheinlichkeitsfunktion formulieren:

$$H_0 : \mathbf{P}(X = \tilde{x}_j) = \mathbf{P}_0(X = \tilde{x}_j), \quad X(\Omega) = \{\tilde{x}_1, \tilde{x}_2, \dots\}.$$

Falls H_0 gilt ist dann

$$p_i = \mathbf{P}(X \in \Delta_i) = \sum_{j:\tilde{x}_j \in \Delta_i} \mathbf{P}_0(X = \tilde{x}_j), \qquad i = 1, \dots, m.$$

Ist F_0 stetig differenzierbar mit Dichte $f_0(x)$, so gilt

$$p_i = \int_{a_{i-1}}^{a_i} f_0(x)\, \mathrm{d}x. \qquad i = 1, \dots, m.$$

4. Sei n_i die zufällige Anzahl Stichprobenwerte in Δ_i, also $n = \sum_{i=1}^m n_i$. Gilt H_0, so nimmt X mit Wahrscheinlichkeit p_i einen Wert aus Δ_i an und mit Wahrscheinlichkeit $1 - p_i$ außerhalb. Bei n unabhängigen Stichproben ist n_i daher binomialverteilt mit Parameter p_i und n, also Erwartungswert np_i. Die Werte n_i heißen empirische Häufiggkeiten, die np_i theoretische Häufigkeiten.

5. Wenn H_0 gilt dürfen sich empirische und theoretische Häufigkeiten nicht signifikant unterscheiden. Testgröße:

$$t = \sum_{k=1}^{m} \frac{(n_k - np_k)^2}{np_k}.$$

Einfache Umformung:

$$t = n \sum_{k=1}^{m} \frac{1}{p_k} \left(\frac{n_k}{n} - p_k \right)^2 = \sum_{k=1}^{m} \frac{n_k^2}{np_k} - n.$$

6. Bei hinreichend großem n ist t (näherungsweise) Realisierung einer Zufallsvariablen T, welche χ^2-verteilt ist mit $m-1$ Freiheitsgraden.
7. **Entscheidung:** H_0 wird mit Irrtumswahrscheinlichkeit α abgelehnt, falls

$$t \geq \chi^2_{m-1;1-\alpha}$$

Dabei ist, wenn $p_{\chi^2;n}$ die Dichte der χ^2-Verteilung mit n Freiheitsgraden bezeichnet,

$$\int_{\chi^2_{m-1;1-\alpha}}^{\infty} p_{\chi^2;m-1}(x)\,\mathrm{d}x.$$

Die Quantile $\chi^2_{m-1;1-\alpha}$ erhält man aus Tabellen bzw. statistischer Software.