

Stocha Notizen

Inhalt

Grundbegriffe	2
Wahrscheinlichkeitsmaß	3
Rechenregeln für Wahrscheinlichkeitsmaß	4
Laplace-Raum	5
Kombinatorisches Abzählprinzip	5
Urnenmodelle	6
1. Ziehen in Reihenfolge mit Zurücklegen	6
2. Ziehen in Reihenfolge ohne Zurücklegen	6
3. Ziehen ohne Reihenfolge ohne Zurücklegen	6
4. Ziehen ohne Reihenfolge mit Zurücklegen	6
Ereignisalgebra / σ -Algebra	7
Bedingte Wahrscheinlichkeit	8
Rechenregel	8
Satz von der totalen Wahrscheinlichkeit	8
Satz von Bayes	8
Mehrstufige Wahrscheinlichkeitsmodelle	9
Pfadregel	9
Produktsatz	9
Totale und paarweise Unabhängigkeit	9
Zufallsvariablen	10
Diskrete Zufallsvariable	10
Verteilung einer Zufallsvariable	11
Wahrscheinlichkeitsfunktion/ Zähl-dichte	12
Verteilungsfunktion	13
Quantilfunktion	14
Stetige Zufallsvariablen/Dichtefunktion	15
Dichtetransformationssatz	16
Unabhängige Zufallsvariablen	17
Kriterium für diskrete Zufallsvariablen	17
Kriterium für stetige Zufallsvariablen	17
Zufallsstichprobe	18
Erwartungswert	19
Erwartungswert für diskrete ZV	19
Erwartungswert für stetige ZV	19
Rechenregeln für den Erwartungswert	19
Varianz	20
Verschiebungssatz	20
Rechenregeln für Varianz	20
Momente, Kurtosis, Exzess	21
Bernoulli-Verteilung	22
Binomialverteilung	23
Faltung	23
Negative Binomialverteilung	23
Geometrische Verteilung	24
Poisson-Verteilung	25
Poisson-Grenzwertsatz	25
Verteilung	25
Rechenregeln	25
Stetige Gleichverteilung (uniforme Verteilung)	26
Exponentialverteilung	27
Normalverteilung	28
Eigenschaften der Normalverteilung	28
Rechenregeln	28
Zufallsvektoren	29
Verteilung eines Zufallsvektors (X, Y)	29
Produktverteilung / Produktmodell	30
Bedingte Verteilung/ Unabhängigkeit	31
Bedingte Dichte	32
Kriterien für Unabhängigkeit	33
Bedingter Erwartungswert	34
Erwartungsvektor	35
Kovarianz	36
Rechenregeln	36
Multivariate Normalverteilung	37
Das Gesetz der großen Zahlen	38
Tschebyschow-Ungleichung	38
Schwaches Gesetz der großen Zahlen	38
Starkes Gesetz der großen Zahlen	38
Hauptsatz der Statistik	39
Zentraler Grenzwertsatz (ZGWS)	40
Stochastische Konvergenz	41
Rechenregeln	41
Fast sichere Konvergenz	41
Grundbegriffe Statistik	42
Visualisierung von Zahlenmaterial	43
Sortierte Beobachtungen	44
Gruppierung (Klassierung) von Daten	45
Histogramm	46
Median	47
Berechnung	47
Eigenschaften	47
Arithmetisches Mittel	48
Eigenschaften	48
Entropie	49
Eigenschaften	49
Stichprobenvarianz/ empirische Varianz	50
Rechenregeln	50
Verschiebungssatz	51
Quantile	52
Berechnung	52
Quartile	53
Fünfpunkte-Zusammenfassung und Boxplot	54
Schließende Statistik	55
Grundbegriffe	55
Stichprobe	55
Verteilungsmodell	55
Statistik, Schätzfunktion, Schätzer	56
Statistik	56
Empirische Verteilungsfunktion	57
Likelihood-Funktion	58
Likelihood-Prinzip	58
Maximum-Likelihood-Schätzer	58
Likelihood für Dichten	58
Likelihood für Stichproben	58
Schätzer	59
Erwartungstreue	60
Anschauung	60
Verzerrung (Bias)	61
Gütekriterien	62
(Asymptotische) Erwartungstreue, Unverfälschtheit	62
Konsistenz	62
Effizienz	62
Mittlerer quadratischer Fehler/ MSE	62
Der t -Test (für eine Stichprobe)	63
Einseitiger t -Test (1)	63
Einseitiger t -Test (2)	63
Zweiseitiger t -Test	63
p-Wert	64
Einseitige Tests	64
Zweiseitiger Test	64
Einseitige Tests gegeben p_{zweis}	64
Gütefunktion	65
2-Stichproben-Tests	66
Verbundenes Design	66
Unverbundenes Design	66
Test auf Varianzhomogenität	66
Test auf Lageunterschied	66
Welch-Test auf Lageunterschied	67
Fallzahlplanung	68
Zweiseitiger Test	68
Einseitiger Test	68
t -Verteilung	69
χ^2 -Verteilung	70
Verteilung der Varianzschätzer	71
F-Verteilung	72
Konfidenzintervall	73
Konfidenzintervall für μ	73
Konfidenzintervall für σ^2	73
Konfidenzintervall für p	73
Statistische Testtheorie	74
Testproblem, Nullhypothese, Alternative	74
Statistischer Test	74
Fehler 1. und 2. Art	74
Signifikanzniveau, Test zum Niveau α	74
Schärfe (Power)	74
Hypothesen (über den Erwartungswert μ)	75
Gauß-Test	76
Einseitiger Gauß-Test	76
Zweiseitiger Gauß-Test	76

Grundbegriffe

Ergebnismenge/Grundmenge Ω

Ereignis A : Teilmenge der Ergebnismenge

Wahrscheinlichkeit eines **Ereignisses** $P(A)$

Versuchsausgang $\omega \in \Omega$

Elementar**ereignis** $\{\omega\}, \omega \in \Omega$

Sicheres **Ereignis**: $P(A) = 1$

Unmögliches **Ereignis**: $P(A) = 0$

Ereignisse disjunkt: $A_i \cap A_j = \emptyset = \emptyset$

Zufallsexperiment: (Ω, P)

Wahrscheinlichkeitsraum: (Ω, \mathcal{A}, P)

Wahrscheinlichkeitsmaß

Ein/e **Wahrscheinlichkeitsmaß/Wahrscheinlichkeitsverteilung** ist eine Abbildung, die jedem **Ereignis** $A \subseteq \Omega$ eine Zahl $P(A) \in \mathbb{R}$ zuordnet ($P : \text{Pot}(\Omega) \rightarrow \mathbb{R}$), sodass gilt:

1. $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. Sind A_1, A_2, \dots paarweise disjunkt, dann gilt

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots = \sum_{k=1}^{\infty} P(A_k)$$

Rechenregeln für **Wahrscheinlichkeitsmaß**

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

2. Für $A \subseteq B$ gilt: $P(A \setminus B) = P(B) - P(A)$

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

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

5. $P(A \cup B) \leq P(A) + P(B)$

Laplace-Raum

(Ω, P) heißt Laplace-Raum, wenn $\Omega = \{\omega_1, \dots, \omega_K\}$ endlich ist und das **Wahrscheinlichkeitsmaß** durch

$$p(\omega) = P(\{\omega\}) = \frac{1}{K} \quad \omega \in \Omega$$

gegeben ist. P heißt auch diskrete **Gleichverteilung** auf Ω . Dann berechnen sich **Wahrscheinlichkeiten** durch

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl möglicher Fälle}}{\text{Anzahl günstiger Fälle}}$$

Kombinatorisches Abzählprinzip

Ist $\Omega = \Omega_1 \times \dots \times \Omega_k$ für ein $k \in \mathbb{N}$ und $A = A_1 \times \dots \times A_k \subseteq \Omega$, dann ist

$$|A| = |A_1| \cdot |A_2| \cdot \dots \cdot |A_k|$$

Urnenmodelle

Sei $A = \{1, \dots, N\}$.

1. Ziehen in Reihenfolge mit Zurücklegen

$$\Omega = \{(\omega_1, \dots, \omega_n) \mid \omega_1, \dots, \omega_n \in A\}, \quad |\Omega| = N^n$$

2. Ziehen in Reihenfolge ohne Zurücklegen

$$\Omega = \{(\omega_1, \dots, \omega_n) \mid \omega_1, \dots, \omega_n \in A, \omega_i \neq \omega_j \text{ für } i \neq j\}$$

Es gilt:

$$|\Omega| = N \cdot (N - 1) \cdot \dots \cdot (N - n + 1) = \frac{N!}{(N - n)!}$$

3. Ziehen ohne Reihenfolge ohne Zurücklegen

$$\Omega = \{\{\omega_1, \dots, \omega_k\} : \omega_1, \dots, \omega_k \in \{1, \dots, n\}, \omega_i \neq \omega_k, (i \neq j)\}$$

Es gilt:

$$|\Omega| = \frac{1}{k!} \frac{n!}{(n - k)!} = \binom{n}{k}$$

4. Ziehen ohne Reihenfolge mit Zurücklegen

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i \in A, i = 1, \dots, n, \omega_1 \leq \dots \leq \omega_n\}$$

Es gilt:

$$|\Omega| = \binom{N - 1 + n}{n}$$

Ereignisalgebra / σ -Algebra

Ein Mengensystem $\mathcal{A} \subseteq \text{Pot}(\Omega)$ von Teilmengen von Ω heißt **Ereignisalgebra** (σ -Algebra), wenn die folgenden Eigenschaften gelten:

1. Die Ergebnismenge Ω und die leere Menge \emptyset gehören zu \mathcal{A} .
2. Mit A ist auch \bar{A} Element von \mathcal{A} .
3. Sind A_1, A_2, \dots Mengen aus \mathcal{A} , dann ist auch $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots$ ein Element von \mathcal{A} .

Die Elemente von \mathcal{A} heißen **Ereignisse**.

Bedingte Wahrscheinlichkeit

Es seien A, B Ereignisse $P(B) > 0$. Dann heißt

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

bedingte Wahrscheinlichkeit von A gegeben B .

Liegt ein Laplace-Raum vor, dann ist $P(A | B)$ der Anteil der für das Ereignis $A \cap B$ günstigen Fälle, bezogen auf die möglichen Fälle, welche die Menge B bilden:

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

Rechenregel

A, B seien Ereignisse mit $P(B) > 0$. Dann gilt:

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

Satz von der totalen Wahrscheinlichkeit

Sei A_1, \dots, A_k eine disjunkte Zerlegung von Ω :

$$\Omega = A_1 \cup \dots \cup A_k, \quad A_i \cap A_j = \emptyset, i \neq j$$

Dann gilt:

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

Satz von Bayes

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

Mehrstufige Wahrscheinlichkeitsmodelle

$$\Omega = \Omega_1 \times \dots \times \Omega_n$$

Startverteilung $p(\omega_1)$, $\omega_1 \in \Omega_1$

Bedingte Wahrscheinlichkeiten $p(\omega_j \mid \omega_1, \dots, \omega_{j-1})$

Pfadregel

für $\omega = (\omega_1, \dots, \omega_n)$:

$$P(\{\omega\}) = p(\omega_1)p(\omega_2 \mid \omega_1) \cdots p(\omega_n \mid \omega_1, \dots, \omega_{n-1})$$

Produktsatz

Zwei Ereignisse heißen stochastisch unabhängig, wenn

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

gilt.

k Ereignisse erfüllen den Produktsatz, wenn gilt:

$$P(\cap_{i=1}^k A_i) = \prod_{i=1}^k P(A_i)$$

Totale und paarweise Unabhängigkeit

- $A_1, \dots, A_n \subseteq \Omega$ heißen (total) stochastisch unabhängig, wenn für jede Teilauswahl A_{i_1}, \dots, A_{i_k} von $k \in \mathbb{N}$ Ereignissen der Produktsatz gilt.
- A_1, \dots, A_n heißen paarweise stochastisch unabhängig, wenn alle Paare A_i, A_j mit $(i \neq j)$ stochastisch unabhängig sind.

Zufallsvariablen

Eine Abbildung $X : \Omega \rightarrow \mathcal{X} \subseteq \mathbb{R}$, $\omega \mapsto X(\omega)$

Ω abzählbar, in die reellen Zahlen heißt **Zufallsvariable** (mit Werten in \mathcal{X}).

$x = X(\omega)$: Realisation

Zusatz: Allgemeines $\Omega : X$ muss messbar sein:

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \text{ für alle Ereignisse } B \text{ von } \mathcal{X}.$$

Diskrete Zufallsvariable

X heißt **diskrete Zufallsvariable**, wenn

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

eine diskrete Menge (endlich oder abzählbar) ist.

Notiz: Ω diskret \Rightarrow Alle ZV sind diskret.

Verteilung einer Zufallsvariable

Die Zuordnung, die jedem Ereignis A die Wahrscheinlichkeit $P(X \in A)$ zuordnet, heißt **Verteilung von X** . Formal:

$$P_X : A \mapsto P_{X(A)} = P(X \in A)$$

für Ereignisse $A \subseteq \mathcal{X}$.

Hinweis: Unterscheide P , das **Wahrscheinlichkeitsmaß** auf Ω und P_X , das **Wahrscheinlichkeitsmaß** auf \mathcal{X} .

- Punktförmige Ereignisse $\{x\}, x \in \mathcal{X}$:

$$P_X(\{x\}) = P(X = x)$$

- Intervallförmige Ereignisse $(a, b], a \leq b$

$$P_X((a, b]) = P(X \in (a, b]) = P(a < X \leq b)$$

Wahrscheinlichkeitsfunktion/ Zähl-dichte

X sei diskrete Zufallsvariable mit Werten $\mathcal{X} = \{x_1, x_1, \dots\} \subseteq \mathbb{R}$. Dann heißt die Funktion

$$p_X(x) = P(X = x), \quad x \in \mathbb{R},$$

Wahrscheinlichkeitsfunktion oder **Zähl-dichte** von X , Es gilt:

$$\sum_{x \in \mathcal{X}} p_X(x) = \sum_{i=0}^{\infty} p_x(x_i) = 1$$

Sie bestimmt eindeutig die **Verteilung** von X .

Die **Zähl-dichte** kann durch die **Punkt-wahrscheinlichkeiten**

$$p_i = P(X = x_i), \quad i = 1, 2, \dots$$

festgelegt werden: Es gilt $p_X(x_i) = p_i$ und $p_X(x) = 0$, wenn $x \notin \mathcal{X}$. Kann X nur endlich viele Werte x_1, \dots, x_k annehmen, dann heißt (p_1, \dots, p_k) auch **Wahrscheinlichkeitsvektor**.

Verteilungsfunktion

Die Funktion $F_x : \mathbb{R} \rightarrow [0, 1]$,

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}$$

heißt **Verteilungsfunktion** von X . $F_X(x)$ ist monoton wachsend, rechtsstetig und es gilt:

$$F(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = 0, \quad F(\infty) := \lim_{x \rightarrow \infty} F_X(x) = 1$$

Ferner gilt: $P(X < x) = F(x -) = \lim_{z \uparrow x} F(z)$ und

$$P(X = x) = F(x) - F(x -).$$

Allgemein heißt jede monoton wachsende und rechtsstetige Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit $F(-\infty) = 0$ und $F(\infty) = 1$ **Verteilungsfunktion (auf \mathbb{R})** und besitzt obige Eigenschaften.

Quantilfunktion

$F(x)$ sei eine **Verteilungsfunktion**.

Die Funktion $F^{-1} : [0, 1] \rightarrow \mathbb{R}$,

$$F^{-1}(p) = \min\{x \in \mathbb{R} : F(x) \geq p\}, \quad p \in (0, 1),$$

heißt **Quantilfunktion von F** .

Ist $F(x)$ stetig und streng monoton steigend, dann ist $F^{-1}(p)$ die Umkehrfunktion von $F(x)$.

Für ein festes p heißt $F^{-1}(p)$ (theoretisches) p -Quantil.

Stetige Zufallsvariablen/Dichtefunktion

Eine ZV X heißt **stetig (verteilt)**, wenn es eine integrierbare nicht-negative Funktion $f(x)$ gibt, so dass für alle Intervalle $(a, b] \subseteq \mathbb{R}$ gilt:

$$P_X((a, b]) = P(a < X \leq b) = \int_a^b f(x) dx$$

$f_X(x) = f(x)$ heißt dann **Dichtefunktion von X** (kurz: **Dichte**).

Allgemein heißt jede Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f(x) \geq 0, x \in \mathbb{R}, \text{ und } \int_{-\infty}^{\infty} f(x) dx = 1$$

Dichtefunktion.

Notation: X hat **Dichte** $f_X(x)$:

$$X \sim f_X$$

Verteilungsfunktion aus **Dichte**:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad x \in \mathbb{R}$$

Dichte aus **Verteilungsfunktion**:

$$f_X(x) = F'_X(x), \quad x \in \mathbb{R}$$

Dichtetransformationssatz

X sei eine stetige **Zufallsvariable** mit Werten in $\mathcal{X} = (a, b)$, $a < b$, und mit **Dichte**funktion $f_X(x)$.

Weiter sei $y = g(x)$ eine stetig differenzierbare Funktion mit Umkehrfunktion $x = g^{-1}(y)$, so dass $(g^{-1})' \neq 0$ gilt.

Dann hat die **Zufallsvariable** $Y = g(X)$ die **Dichte**funktion

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

Unabhängige Zufallsvariablen

Zwei **Zufallsvariablen** X und Y heißen stochastisch unabhängig, wenn die **Ereignisse** $\{X \in A\}$ und $\{Y \in B\}$ stochastisch unabhängig sind, für alle **Ereignisse** $A \subseteq \mathbb{R}$ und $B \subseteq \mathbb{R}$, d.h.

$$P(X \in A, Y \in B) = P(\{X \in A\} \cap \{Y \in B\}) = P(X \in A) \cdot P(Y \in B)$$

n **Zufallsvariablen** X_1, \dots, X_n mit Werten in Mengen $\mathcal{X}_1, \dots, \mathcal{X}_n$ heißen (total) **stochastisch unabhängig**, wenn für alle **Ereignisse** $A_1 \subseteq \mathcal{X}_1, \dots, A_n \subseteq \mathcal{X}_n$ die **Ereignisse** $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$ stochastisch unabhängig sind. D.h.: Für alle $i_1, \dots, i_k \in \{1, \dots, n\}$ gilt:

$$P(X_{i_1} \in A_{i_1}, \dots, X_{i_k} \in A_{i_k}) = P(X_{i_1} \in A_{i_1}) \cdots P(X_{i_k} \in A_{i_k})$$

Kurz: Stets gilt der Produktsatz für gemeinsame **Wahrscheinlichkeiten** (d.h. von Schnitten)

Kriterium für diskrete Zufallsvariablen

Zwei diskrete **Zufallsvariablen** X und Y sind stochastisch unabhängig, wenn für alle Realisationen x_i von X und y_j von Y die **Ereignisse** $\{X = x_i\}$ und $\{Y = y_j\}$ stochastisch unabhängig sind, d.h.

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

Dann gilt ferner

$$P(X = x_i | Y = y_j) = P(X = x_i), \quad \text{und} \quad P(Y = y_j | X = x_i) = P(Y = y_j)$$

Kriterium für stetige Zufallsvariablen

Zwei stetige **Zufallsvariablen** X und Y sind stochastisch unabhängig, wenn für alle Intervalle $(a, b]$ und $(c, d]$ die **Ereignisse**

$$\{a < X \leq b\} \quad \text{und} \quad \{c < Y \leq d\}$$

unabhängig sind, d.h.

$$\begin{aligned} P(a < X \leq b, c < Y \leq d) &= \int_a^b f_X(x) dx \cdot \int_c^d f_Y(y) dy \\ &= \int_a^b \int_c^d f_X(x) f_Y(y) dy dx \end{aligned}$$

Zufallsstichprobe

Das Gesamtexperiment sei wie folgt beschrieben:

- n -fache Wiederholung eines Zufallsexperiments beschrieben durch $X : \Omega \rightarrow \mathcal{X}$.
- Die Wiederholungen erfolgen unter identischen Bedingungen.
- Die Ergebnisse hängen nicht voneinander ab.

Stochastisches Modell:

- n **Zufallsvariablen** $X_1, \dots, X_n : \Omega \rightarrow \mathcal{X}$.
- X_i repräsentiert das Ergebnis der i -ten Wiederholung.

X_1, \dots, X_n bilden eine (**einfache**) **Zufallsstichprobe**, wenn gilt:

1. X_1, \dots, X_n sind stochastisch unabhängig und
2. X_1, \dots, X_n sind identisch verteilt, d.h. alle X_i besitzen dieselbe **Verteilung**:

$$P(X_i \in A) = P(X_1 \in A), i = 1, \dots, n \text{ für alle } \text{Ereignisse } A.$$

Sei $F(x) = F_X(x)$ die **Verteilungsfunktion** der X_i , so schreibt man kurz:

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F(x)$$

i.i.d. (engl.: independent an identically distributed) steht hierbei für unabhängig und identisch verteilt.

Erwartungswert

Erwartungswert für diskrete ZV

$X \sim p_X$ diskrete ZV mit Werten in \mathcal{X} , verteilt nach der Zähldichte p_X . Dann heißt die reelle Zahl

$$E(X) = \sum_{x \in \mathcal{X}} x \cdot p_X(x)$$

Erwartungswert von X , sofern $\sum_{x \in \mathcal{X}} |x| \cdot p_X(x) < \infty$.

Wichtiger Spezialfall: $\mathcal{X} = \{x_1, \dots, x_k\}$ endlich. Dann ist

$$E(X) = x_1 \cdot p_X(x_1) + x_2 \cdot p_X(x_2) + \dots + x_k \cdot p_X(x_k).$$

Erwartungswert für stetige ZV

$X \sim f_X$ stetige ZV, verteilt nach der Dichtefunktion $f_X(x)$.

Dann heißt die reelle Zahl

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

Erwartungswert von X (sofern $\int_{-\infty}^{\infty} |x| \cdot f_X(x) dx < \infty$).

Rechenregeln für den Erwartungswert

Seien X, Y ZVen (mit $E(|X|), E(|Y|) < \infty$) und $a, b \in \mathbb{R}$.

1. $E(X + Y) = E(X) + E(Y)$,
2. $E(aX + b) = aE(X) + b$,
3. $E(|X + Y|) \leq E(|X|) + E(|Y|)$
4. **Jensen-Ungleichung:** Ist $g(x)$ konvex, dann gilt:

$E(g(X)) \geq g(E(X))$ und $E(g(X)) > g(E(X))$, falls $g(x)$ strikt konvex ist. Ist $g(x)$ konkav bzw. strikt konkav, dann kehren sich die Ungleichheitszeichen um.

Produkteigenschaft Seien X, Y stochastisch unabhängige ZVen.

Für alle Funktionen $f(x)$ und $g(y)$ (mit $E(|f(x)|) < \infty$) gilt:

$$E(f(X) \cdot g(Y)) = E(f(X)) \cdot E(g(Y))$$

Insbesondere $E(XY) = E(X) \cdot E(Y)$

Notiz

X, Y unabhängig $\Rightarrow E(XY) - E(X) \cdot E(Y) = 0$.

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

ist ein gängiges Maß für Abhängigkeit.

Varianz

Sei X eine **Zufallsvariable**. Dann heißt

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

Varianz von X , sofern $E(X^2) < \infty$. Die Wurzel aus der **Varianz**,

$$\sigma_X = \sqrt{\text{Var}(X)},$$

heißt **Standardabweichung von X** .

Verschiebungssatz

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

Rechenregeln für Varianz

X, Y **Zufallsvariablen**

1. $\text{Var}(aX) = a^2 \text{Var}(X)$
2. Falls $E(X) = 0$, dann gilt $\text{Var}(X) = E(X^2)$
3. Sind X und Y stochastisch unabhängig, dann gilt:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Momente, Kurtosis, Exzess

Sei X eine **Zufallsvariable** und $k \in \mathbb{N}$ und es gelte $E(|X^k|) < \infty$.

- $m_k = E(X^k)$ ist das **k -te Moment** von X
- $m_k^* = E(|X|^k)$ ist das **k -te absolute Moment** von X
- $m_{k(a)} = E(X - a)^k$ ist das **k -te Moment um a**
- $m_k^*(a) = E(|X - a|^k)$ ist das **k -te absolute Moment um a**
- $\beta_2 = (E(X^*)^4)$ heißt Kurtosis von X (misst die Wölbung)
- $\gamma_2 = \beta_2 - 3$ heißt Exzess. $\gamma_2 > 0$: **Verteilung** 'spitzer' als **Gaussverteilung**, $\gamma_2 < 0$: 'flacher'

Bernoulli-Verteilung

Sei A ein Ereignis. Beobachte, ob A eintritt oder nicht:

$$X = \mathbb{1}_A = \begin{cases} 1, & A \text{ tritt ein} \\ 0, & A \text{ tritt nicht ein.} \end{cases}$$

Träger: $= \{0, 1\}$ (binär). Verteilung gegeben durch $p = P(X = 1) = P(A)$, $q = 1 - p = P(X = 0)$

p : Erfolgswahrscheinlichkeit

$$X \sim \text{Ber}(p), \quad X \sim \text{Bin}(1, p)$$

- Erwartungswert: $E(X) = p$,
- Varianz: $\text{Var}(X) = p(1 - p)$

Binomialverteilung

- Modell: X_1, \dots, X_n i.i.d. $\sim \text{Ber}(p)$
- Anzahl der Erfolge gegeben durch:

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

$\Rightarrow Y$ ist binomialverteilt.

Y heißt binomialverteilt, $Y \sim \text{Bin}(n, p)$, wenn

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n$$

Erwartungswert: $E(Y) = np$

Varianz: $\text{Var}(Y) = np(1-p)$

Zähldichte: $p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, \dots, n\}$

Faltung

$X \sim \text{Bin}(n_1, p)$ und $Y \sim \text{Bin}(n_2, p)$ unabhängig, dann folgt:

$$X + Y \sim \text{Bin}(n_1 + n_2, p)$$

Negative Binomialverteilung

Die **Verteilung** der Summe

$$S = T_1 + \dots + T_k = \sum_{i=1}^k T_i$$

von k i.i.d. $\text{Geo}(p)$ -verteilten **Zufallsvariablen** T_1, \dots, T_k heißt **negativ binomialverteilt**. S_k ist die Anzahl der erforderlichen Versuche, um k Erfolge zu beobachten.

Geometrische Verteilung

T heißt **geometrisch verteilt** mit Parameter $p \in (p, 1]$. Notation: $T \sim \text{Geo}(p)$

$$P(W = n) = p(1 - p)^n, \quad n = 0, 1, \dots$$

$$P(T = n) = p(1 - p)^{n-1}, \quad n = 1, 2, \dots$$

- Erwartungswerte:

$$E(T) = \frac{1}{p}, \quad E(W) = \frac{1}{p} - 1$$

- Varianzen:

$$\text{Var}(T) = \frac{1 - p}{p^2}, \quad \text{Var}(W) = (1 - p)$$

Poisson-Verteilung

Es werden punktförmige **Ereignisse** in einem Zeitintervall $[0, T]$ gezählt:

$$X_t = \begin{cases} 1, & \text{Ereignis zur Zeit } t \\ 0, & \text{kein Ereignis zur Zeit } t \end{cases}$$

Die X_t sind unabhängig und identisch verteilt. Zerlege nun $[0, T]$ in n gleichbreite Teilintervalle:

$$X_{ni} = \begin{cases} 1, & \text{Ereignis im } i\text{-ten Teilintervall} \\ 0, & \text{kein Ereignis im } i\text{-ten Teilintervall} \end{cases}$$

Dann gilt: X_{n1}, \dots, X_{nn} i.i.d. **Bin**(1, p_n) mit

$$p_n = \lambda \cdot \frac{T}{n}$$

Dabei ist λ die Proportionalitätskonstante.

Anzahl:

$$Y = X_{n1} + \dots + X_{nn} \sim \text{Bin}(n, p_n)$$

Poisson-Grenzwertsatz

Sind $Y_n \sim \text{Bin}(n, p_n)$, $n = 1, 2, \dots$, binomialverteilte **Zufallsvariablen** mit $np_n \rightarrow \lambda$, $n \rightarrow \infty$, dann gilt für ein festes k :

$$\lim_{n \rightarrow \infty} P(Y_n = k) = p_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Verteilung

Y heißt **poissonverteilt** mit Parameter λ . Notation: $Y \sim \text{Poi}(\lambda)$, wenn

$$P(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

- **Erwartungswert:** $E(Y) = \lambda$
- **Varianz:** $\text{Var}(Y) = \lambda$
- **Zähldichte:** $p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}_0$

Rechenregeln

- $X \sim \text{Poi}(\lambda)$ und $Y \sim \text{Poi}(\mu)$ unabhängig, dann $X + Y \sim \text{Poi}(\lambda + \mu)$
- $X \sim \text{Poi}(\lambda)$ die Anzahl in $[0, T]$ und Y die Anzahl im Teilintervall $[0, r \cdot T]$, so ist $Y \sim \text{Poi}(r \cdot \lambda)$

Für (sehr) kleine p : $Y \sim \text{Bin}(n, p)$ wird mit $\lambda = np$ die **Binomialverteilung** approximiert:

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

Stetige Gleichverteilung (uniforme Verteilung)

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

X heißt dann **stetig gleichverteilt auf dem Intervall** $[a, b]$. Notation: $X \sim U[a, b]$. Für die **Verteilungsfunktion** ergibt sich:

$$F(x) = \frac{x-a}{b-a}, \quad x \in [a, b]$$

sowie $F(x) = 0$, wenn $x < a$ und $F(x) = 1$ für $x > b$

- **Erwartungswert:** $E(X) = \frac{a+b}{2}$
- **Varianz:** $\text{Var}(X) = \frac{(b-a)^2}{12}$

Exponentialverteilung

X heißt **exponentialverteilt**, $X \sim \text{Exp}(\lambda)$ mit Parameter $\lambda > 0$, wenn X die **Dichtefunktion**

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

und $f(x) = 0$ für $x \leq 0$ besitzt.

- **Erwartungswert:** $E(X) = \frac{1}{\lambda}$
- **Varianz:** $E((X - \frac{1}{\lambda})^2) = E(X^2) - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$

Normalverteilung

X heißt normalverteilt mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 \in (0, \infty)$, falls X die Dichte

$$\varphi_{(\mu, \sigma^2)}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

hat (**Gauß'sche Glockenkurve**)

Verteilungsfuntion:

$$\Phi_{(\mu, \sigma^2)}(x) = \int_{-\infty}^x \varphi_{(\mu, \sigma^2)}(t) dt = p$$

Quantilfunktion (Umkehrfunktion):

$$x = \Phi_{(\mu, \sigma^2)}^{-1}(p)$$

Eigenschaften der Normalverteilung

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

Rechenregeln

- $X \sim N(\mu_1, \sigma_1^2)$ und $Y \sim N(\mu_2, \sigma_2^2)$ unabhängig, dann gilt: $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- Ist $X \sim N(\mu, \sigma^2)$ und sind $a, b \in \mathbb{R}$, dann gilt: $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- Ist $X \sim N(\mu, \sigma^2)$, dann gilt:

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

mit X^* : Standardisierte Version

- Ist $X^* \sim N(0, 1)$, dann gilt

$$\mu + \sigma \cdot X^* \sim N(\mu, \sigma^2)$$

Sind $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ unabhängig, dann ist das arithmetische Mittel normalverteilt mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

und

$$\bar{X}^* = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1)$$

Zufallsvektoren

Wenn Ω abzählbar ist, dann heißt jede Abbildung

$$\mathbf{X} : \Omega \rightarrow \mathbb{R}^n, \quad \omega \mapsto \mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

in den n -dimensionalen Raum \mathbb{R}^n Zufallsvektor. Realisationen von $\mathbf{X} = (X_1, \dots, X_n)$ sind Vektoren x im \mathbb{R}^n : $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Verteilung eines Zufallsvektors (X, Y)

Diskrete Zufallsvektoren: Verteilung durch Zähldichten gegeben

$$p(x, y) = P(X = x, Y = y), \quad (x, y) \in \mathbb{R}^2$$

- Stäbe über der (x, y) -Ebene an denjenigen Stellen (x, y) mit $P(X = x, Y = y) > 0$, sonst 0.
- $P(X \in A, Y \in B) = \sum_{(x,y) \in A \times B} p(x, y)$ (Summe der Stäbe, die in $A \times B$ stehen.)

Stetige Zufallsvektoren: Verteilung gegeben durch Dichte $f(x, y)$

$$f(x, y) \geq 0, (x, y) \in \mathbb{R}^2, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$$

- ‘Gebirge’ über der (x, y) -Ebene
- $P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f(x, y) \, dy \, dx$

Produktverteilung / Produktmodell

Produkt-Verteilungsfunktion

$$F(x, y) = F_X(x) \cdot F_Y(y), \quad (x, y) \in \mathbb{R}^2$$

Produkt-Zähldichte

$$p(x, y) = p_X(x) \cdot p_Y(y), \quad (x, y) \in \mathbb{R}^2$$

Produkt-Dichtefunktion

$$f(x, y) = f_X(x) \cdot f_Y(y), \quad (x, y) \in \mathbb{R}^2$$

Dies entspricht der stochastischen Unabhängigkeit von X und Y .

Bedingte Verteilung/ Unabhängigkeit

X, Y diskret mit Werten in $\mathcal{X} = \{x_1, x_2, \dots\}$ bzw. $\mathcal{Y} = \{y_1, y_2, \dots\}$.

1. Bedingte **Wahrscheinlichkeit** von $X = x_i$ gegeben $Y = y_j$:

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

$p(x_i, y_j)$: gemischte Zähl**dichte** $p_Y(y_j)$: Zähl**dichte** von Y .

2. Definiert die **bedingte Zähl**dichte****

$$p(x | y) = p_{X|Y}(X = x | Y = y) = \begin{cases} \frac{p(x, y)}{p_Y(y)}, & y \in \{y_1, y_2, \dots\} \\ p_X(x), & y \notin \{y_1, y_2, \dots\} \end{cases}$$

3. Endlicher Fall: $p(x_i, y_j)$: Tabelle (Kontingenztafel), $p_Y(y_j)$: Rand
4. Entsprechend definiert man die bedingte **Wahrscheinlichkeit** von $Y = y_j$ gegeben $X = x_j$.

Bedingte Dichte

X und Y stetig verteilt: $(X, Y) \sim f(x, y)$.

Bedingte Dichte von X gegeben $Y = y$ (y fest):

$$f(x | y) = f_{X|Y}(x | y) = \begin{cases} \frac{f(x, y)}{f_Y(y)}, & f_Y(y) > 0 \\ f_X(x), & f_Y(y) = 0 \end{cases}$$

Dies ist als Funktion von x eine **Dichte** (für jedes $y \in \mathbb{R}$).

Notation: $X | Y = y \sim f_{X|Y}(x | y)$

Kriterien für Unabhängigkeit

1. Diskreter Fall: X und Y sind genau dann stochastisch unabhängig, wenn für alle x und y gilt:

$$p_{X|Y}(x|y) = p_X(x) \quad \text{bzw.} \quad p_{Y|X}(y|x) = p_Y(y)$$

bzw. $p_{x,y} = p_X(x)p_Y(y)$ (gemischte Zähl**dichte** = **Produkt-Zähl**dichte)

2. Stetiger Fall: X und Y genau dann stochastisch unabhängig, wenn für alle x und y gilt:

$$f_{X|Y}(x) = f_X(x) \quad \text{bzw.} \quad f_{Y|X}(y) = f_Y(y)$$

bzw. $f(x,y) = f_X(x)f_Y(y)$ (gemischte **Dichte** = **Produkt**dichte)

3. X, Y sind genau dann stochastisch unabhängig, wenn für die gemeinsame **Verteilungsfunktion** $F_{X,Y}(x,y)$ für alle $x, y \in \mathbb{R}$ gilt:

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$$

Bedingter Erwartungswert

Berechne den **Erwartungswert** mit bedingter **Verteilung**.

1. Sei (X, Y) nach der **Zähldichte** $p(x, y)$ verteilt.

Der **bedingte Erwartungswert** ist gegeben durch

$$E(X | Y = y) = \sum_{x \in \mathcal{X}} x \cdot p_{X|Y}(x | y)$$

2. Sei $(X, Y) \sim f_{X|Y}(x, y)$ stetig. Bedingter **Erwartungswert**:

$$E(X | Y = y) = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x | y) dx$$

3. $g(Y) = E(X | Y = y)$ ist eine Funktion von y .
4. Einsetzen der **Zufallsvariable** Y liefert **bedingte Erwartung von X gegeben Y** .
Notation: $E(X | Y) := g(Y)$

Erwartungsvektor

$\mathbf{X} = (X_1, \dots, X_n)'$ Zufallsvektor.

Gelte: $\mu_i = E(X_i)$, $i = 1, \dots, n$ existieren.

Der (Spalten-)Vektor $\boldsymbol{\mu} = (E(X_1), \dots, E(X_n))'$ heißt **Erwartungsvektor von \mathbf{X}** .

Die Rechenregeln übertragen sich, z.B. gilt für $a, b \in \mathbb{R}$ und Zufallsvektoren \mathbf{X} und \mathbf{Y} :

$$E(a \cdot \mathbf{X} + b \cdot \mathbf{Y}) = a \cdot E(\mathbf{X}) + b \cdot E(\mathbf{Y})$$

Kovarianz

Wenn X und Y **Zufallsvariablen** mit existierenden **Varianzen** sind, dann heißt

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

Kovarianz von X und Y .

- X, Y heißen **unkorreliert**, falls $\text{Cov}(X, Y) = 0$
- Ist $\mathbf{X} = (X_1, \dots, X_n)$ ein Zufallsvektor, dann heißt die symmetrische $(n \times n)$ -Matrix $\text{Var}(\mathbf{X}) = (\text{Cov}(X_i, X_j))_{i,j}$ der n^2 **Kovarianzen** **Kovarianzmatrix von \mathbf{X}**
- Wenn alle X_i paarweise unkorreliert sind, ist $\text{Var}(\mathbf{X})$ eine Diagonalmatrix
- Korrelation:

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

Rechenregeln

Seien X, Y und Z **Zufallsvariablen** mit endlichen **Varianzen**. Dann gelten für alle $a, b \in \mathbb{R}$ die folgenden Rechenregeln:

- $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, Y) = 0$, wenn X und Y unabhängig sind
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
- Cauchy-Schwarz-Ungleichung: $\sigma_X^2 = \text{Var}(X), \sigma_Y^2 = \text{Var}(Y)$

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)} = \sigma_X\sigma_Y$$

Multivariate Normalverteilung

Wenn X_1, \dots, X_n unabhängig und identisch $N(0, 1)$ -verteilte **Zufallsvariablen** sind, dann ist die gemeinsame **Dichtefunktion** des Zufallsvektors $\mathbf{X} = (X_1, \dots, X_n)'$ gegeben durch

$$\varphi(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2 \right), \quad x_1, \dots, x_n \in \mathbb{R}$$

\mathbf{X} heißt **multivariat** oder n -dimensional **standardnormalverteilt**. Notation: $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{1})$

- **Erwartungswertvektor**: $\boldsymbol{\mu} = E(\mathbf{X}) = \mathbf{0} = (0, \dots, 0)' \in \mathbb{R}^n$
- **Kovarianzmatrix**: $\boldsymbol{\Sigma}$ hat die Form einer Einheitsmatrix

Satz: Wenn $\mathbf{X} = (X_1, \dots, X_n)' \sim N_n(\boldsymbol{\mu}, \mathbf{I})$ und $\mathbf{a} = (a_1, \dots, a_n)' \in \mathbb{R}^n$, dann folgt

$$\mathbf{a}'\mathbf{X} \sim N_n(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\mathbf{a})$$

Das Gesetz der großen Zahlen

Tschebyschow-Ungleichung

Seien X_1, \dots, X_n i.i.d. mit **Varianz** $\sigma^2 \in (0, \infty)$ und **Erwartungswert** μ . Dann gilt:

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Schwaches Gesetz der großen Zahlen

Das arithmetische Mittel $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ konvergiert gegen den **Erwartungswert** μ , d.h. für jede Toleranzabweichung $\varepsilon > 0$ gilt: $P(|\bar{X}_n - \mu| > \varepsilon) \rightarrow 0$ für $n \rightarrow \infty$

Starkes Gesetz der großen Zahlen

Seien X_1, \dots, X_n i.i.d. mit $E(|X_1|) < \infty$ und **Erwartungswert** μ . Dann konvergiert das arithmetische Mittel mit **Wahrscheinlichkeit** 1 gegen μ , d.h.

$$P(\bar{X}_n \rightarrow \mu) = P(\{\omega \mid \bar{X}_{n(\omega)} \text{ konvergiert gegen } \mu\}) = 1$$

Hauptsatz der Statistik

Die **Zufallsvariablen** X_1, \dots, X_n, \dots seien unabhängig und identisch nach der **Verteilungsfunktion** F verteilt.

Dann konvergiert der (maximale) Abstand zwischen der **empirischen Verteilungsfunktion** $F_n(x)$ und der **wahren Verteilungsfunktion** $F(x)$ mit **Wahrscheinlichkeit** 1 gegen 0:

$$P\left(\lim_{n \rightarrow \infty} \max_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0\right) = 1$$

Zentraler Grenzwertsatz (ZGWS)

Seien X_1, \dots, X_n i.i.d. mit

$$\mu = E(X_1), \quad \sigma^2 = \text{Var}(X_1) \in (0, \infty)$$

Dann gilt: \bar{X}_n ist asymptotisch $N\left(\mu, \frac{\sigma^2}{n}\right)$ -verteilt,

$$\bar{X}_n \sim_{\text{approx}} N\left(\mu, \frac{\sigma^2}{n}\right)$$

in dem Sinne, dass die **Verteilungsfunktion** der standardisierten Version gegen die **Verteilungsfunktion** der $N(0, 1)$ -**Verteilung** konvergiert:

$$P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} < x\right) \rightarrow \Phi(x), \quad n \rightarrow \infty$$

Die Aussage des zentralen Grenzwertsatzes bleibt gültig, wenn die in der Praxis meist unbekannte Streuung σ durch die empirische Standardabweichung

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

ersetzt wird.

Stochastische Konvergenz

Eine Folge X_1, X_2, \dots von Zufallsvariablen **konvergiert stochastisch** oder in **Wahrscheinlichkeit**

- gegen die **Zufallsvariable** X , wenn für alle $\varepsilon > 0$ gilt:

$$P(|X_n - X| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty$$

Notation: $X_n \xrightarrow{P} X, n \rightarrow \infty$

- gegen die Konstante a , wenn für all $\varepsilon > 0$ gilt:

$$P(|X_n - a| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty$$

Rechenregeln

X_1, X_2, \dots und Y_1, Y_2, \dots seien Folgen von **ZVen**.

- Aus $X_n \xrightarrow{P} a, n \rightarrow \infty$ und $Y_n \xrightarrow{P} b, n \rightarrow \infty$ folgt für $\lambda, \mu \in \mathbb{R}$:

$$\lambda \cdot X_n \pm \mu \cdot Y_n \xrightarrow{P} \lambda \cdot a \pm \mu \cdot b, \quad n \rightarrow \infty$$

- Aus $X_n \xrightarrow{P} X, n \rightarrow \infty$ und $Y_n \xrightarrow{P} b, n \rightarrow \infty$ folgt:

$$X_n \cdot Y_n \xrightarrow{P} X \cdot b, \quad n \rightarrow \infty$$

und, falls $b \neq 0$ und ein $n_0 \in \mathbb{N}$ ex. mit $P(Y_n \neq 0) = 1$ für $n > n_0$:

$$\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{b}, \quad n \rightarrow \infty$$

Aus $X_n \xrightarrow{P} X, n \rightarrow \infty$ und $Y_n \xrightarrow{P} b, n \rightarrow \infty$

- folgt für jede stetige Funktion $f : \mathbb{R} \rightarrow \mathbb{R} : f(X_n) \xrightarrow{P} f(X), n \rightarrow \infty$
- folgt für jede stetige Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R} : f(X, Y)$ und $f(X_n, Y_n)$ definiert sind für alle $n \in \mathbb{N}$, dann gilt: $f(X_n, Y_n) \xrightarrow{P} f(X, Y), n \rightarrow \infty$

Fast sichere Konvergenz

Eine Folge X_1, X_2, \dots von **Zufallsvariablen** **konvergiert fast sicher** oder mit

Wahrscheinlichkeit 1

- gegen die **Zufallsvariable** X , wenn gilt:

$$P\left(\lim_{n \rightarrow \infty} |X_n - X| = 0\right) = 1$$

Notation: $X_n \xrightarrow{\text{f.s.}} X, n \rightarrow \infty$

- gegen die Konstante a , wenn gilt:

$$P\left(\lim_{n \rightarrow \infty} |X_n - a| = 0\right) = 1$$

Grundbegriffe Statistik

Statistische Analyse von Daten:

1. Definition der relevanten **statistischen Einheiten** (**Untersuchungseinheiten, Merkmalsträger**)
2. Die **Grundgesamtheit** G ist die Menge aller statistischen Einheiten.
3. Erhebe Daten (**Merkmale**, Variablen) an allen (Totalerhebung) oder ausgewählten Einheiten.
4. Werden die Daten durch Experimente gewonnen, dann heißen die $g \in G$ auch **Versuchseinheiten** (experimental units). Werden die Daten durch Beobachtung gewonnen, so spricht man von **Beobachtungseinheiten** (observational units).
5. **Merkmale** X nehme gewissen **Merkmalsausprägungen** M an. Formal:

$$X : G \rightarrow M, \quad g \mapsto X(g)$$

(o.E. (durch Kodieren) $M \subseteq \mathbb{R}$)

6. Zufallsauswahl: Ziehe n mal aus der 'Urne' G mit Zurücklegen:

$$\Omega = G \times \dots \times G$$

7. **Zufallsstichprobe**: $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$, unabhängig und identisch verteilte **Zufallsvariablen** (Zufallsvektoren, wenn mehrere Variablen erhoben werden).
8. Bei Experimenten werden den gezogenen $g \in G$ gewisse **Ausprägungen** zugeordnet (z.B. Kontrollgruppe/ Behandlungsgruppe). Diese haben i.d.R. nur wenige mögliche **Ausprägungen** (z.B. binär 0/1).
9. Deskriptive Statistik betrachtet eine Realisation $(x_1, \dots, x_n)'$ als Input, die **Datenmatrix**.

Skalenniveaus: $X : G \rightarrow M$

Diskrete Merkmale: M endlich oder abzählbar unendlich.

Stetige Merkmale: $M \subseteq \mathbb{R}$ Intervall (oder ganz \mathbb{R}).

In der Praxis werden stetige **Merkmale** oft vergrößert (komprimiert) durch **Gruppierung**.

Klassifikation von **Merkmalen** aufgrund des Skalenniveaus:

- **Nominalskala**: **Ausprägungen** nur unterscheidbar (Labels)
- **Ordinalskala**: **Ausprägungen** können verglichen werden (Schulnoten, Grad der Zustimmung 1-5, ...)
- **Metrische Skala (Kardinalskala, Intervallskala, Ratioskala)**:
 - Kardinalskala: Messe Vielfache einer Grundeinheit.
 - Intervallskala: Nullpunkt willkürlich. Dann können Quotienten nicht interpretiert werden (Temperatur).
 - Verhältnis-, Quotienten- o. Ratioskala: Nullpunkt physikalisch zwingend (Längen, Gewichte, Geld, Anzahlen)

Visualisierung von Zahlenmaterial

Ausgangspunkt: **Rohdaten (Primärdaten, Urliste)** nach der Erhebung.

Allgemeine Situation: Erhebe p **Merkmale** an n statistischen Einheiten.

Darstellung der Daten in der **Datenmatrix** (Tabelle):

stat. Einheit Nr.	Geschlecht	Alter	Größe	Messwert
1	M	18	72.6	10.2
2	W	21	18.7	9.5
⋮				⋮
n	W	19	15.6	5.6

i -te Zeile: Werte der p Variablen für die i -te statistische Einheit

j -te Spalte: **Stichprobe** der n beobachteten Werte des j -ten **Merkmals**.

Zeilen = Beobachtungen, Spalten = Variablen

Selektierte Spalte:

→ **Stichprobe** x_1, \dots, x_n

→ Datenvektor $\mathbf{x} = (x_1, \dots, x_n)'$

Nominale/ordinale Daten:

Zähle aus, wie oft die **Ausprägungen** (a_1, \dots, a_k) im Datensatz vorkommen.

Die absoluten **Häufigkeiten** (engl.: frequencies, counts) h_1, \dots, h_k sind durch

$$\begin{aligned} h_j &= \text{Anzahl der } x_i \text{ mit } x_i = a_j \\ &= \sum_{i=1}^n \mathbf{1}(x_i = a_j) \end{aligned}$$

$j = 1, \dots, k$ gegeben. Die (tabellarische) Zusammenstellung der absoluten **Häufigkeiten** h_1, \dots, h_k heißt **absolute Häufigkeitsverteilung**.

Es gilt:

$$n = h_1 + \dots + h_k$$

Dividiert man die absoluten **Häufigkeiten** durch den **Stichprobenumfang** n , so erhält man die **relativen Häufigkeiten** f_1, \dots, f_k . Für $j = 1, \dots, k$ berechnet sich f_j durch

$$f_j = \frac{h_j}{n}$$

f_j ist der Anteil der Beobachtungen, die den Wert a_j haben.

Die (tabellarische) Zusammenstellung der f_1, \dots, f_k heißt **relative Häufigkeitsverteilung**.

Die relativen **Häufigkeiten** summieren sich zu 1 auf:

$$f_1 + \dots + f_k = 1$$

Darstellung durch Stab-, Balken- oder Kreisdiagramme.

Sortierte Beobachtungen

Die sortierten Beobachtungen werden mit $x_{(1)}, \dots, x_{(n)}$ bezeichnet. Die Klammer um den Index deutet somit den Sortiervorgang an.

Es gilt:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$x_{(i)}$ heißt ***i*-te Ordnungsstatistik**,

$(x_{(1)}, \dots, x_{(n)})$ heißt **Ordnungsstatistik** der **Stichprobe** x_1, \dots, x_n . Das **Minimum** $x_{(1)}$ wird auch mit x_{\min} bezeichnet, das **Maximum** entsprechend mit x_{\max} .

Messbereich (range): $[x_{\min}, x_{\max}]$ (kleinstes Intervall, das alle Daten enthält)

Gruppierung (Klassierung) von Daten

Lege k Intervalle

$$I_1 = [g_1, g_2], \quad I_2 = (g_2, g_3], \quad \dots, \quad I_k = (g_k, g_{k+1}]$$

fest, welche den Messbereich überdecken.

I_j heißt j -te **Gruppe** oder **Klasse** und ist für $j = 2, \dots, k$ gegeben durch $I_j = (g_j, g_{j+1}]$. Die Zahlen g_1, \dots, g_{k+1} heißen Gruppengrenzen. Des Weiteren führen wir noch die k **Gruppenbreiten**

$$b_j = g_{j+1} - g_j, \quad j = 1, \dots, k$$

und die k **Gruppenmitten**

$$m_j = \frac{g_{j+1} + g_j}{2}, \quad j = 1, \dots, k$$

ein.

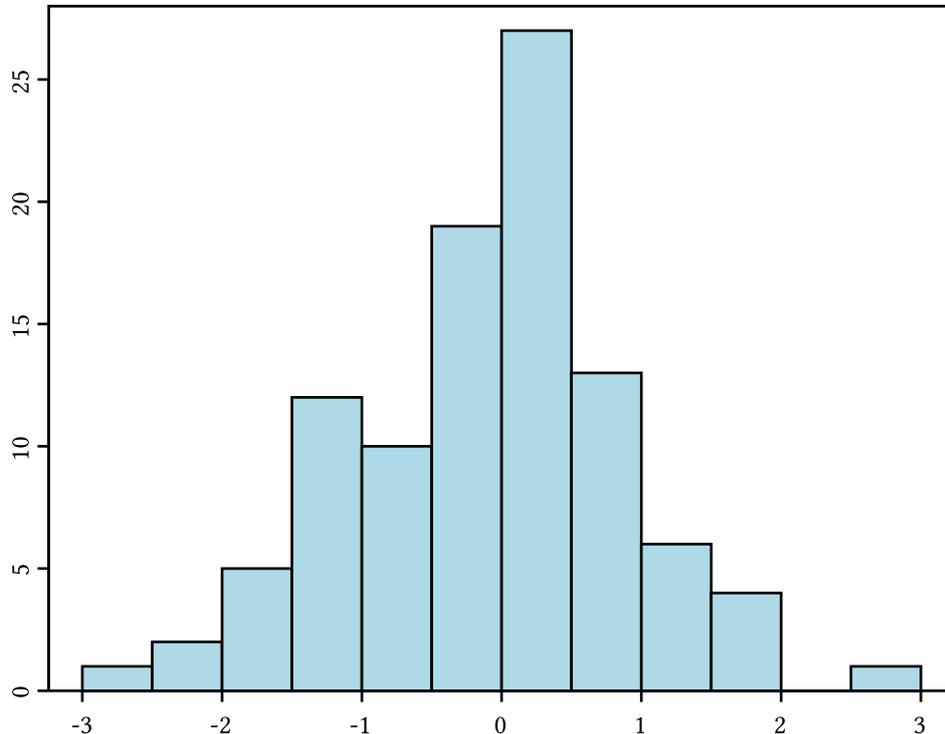
Histogramm

Das Histogramm ist eine grafische Darstellung der relativen **Häufigkeitsverteilung**, die dem Prinzip der Flächentreue folgt.

1. Gruppiere in k Klassen mit Gruppengrenzen $g_1 < \dots < g_{k+1}$.
2. Berechne zugehörige relative **Häufigkeiten** f_1, \dots, f_k .
3. Zeichne über Gruppe j ein Rechteck der Fläche f_j .

Hierzu bestimmen wir die Höhe I_j des j -ten Rechtecks so, dass die Fläche $F_j = b_j I_j$ des Rechtecks der relativen **Häufigkeit** f_j entspricht:

$$F_j = b_j I_j \stackrel{!}{=} f_j \Rightarrow I_j = \frac{f_j}{b_j}, \quad j = 1, \dots, k$$



Der obere Rand des Histogramms definiert eine Treppenfunktion $\hat{f}(x)$, die über dem j -ten Intervall der Gruppeneinteilung den konstanten Funktionswert I_j annimmt. Außerhalb der Gruppeneinteilung setzt man $\hat{f}(x)$ auf 0.

\hat{f} heißt **Häufigkeitsdichte** oder auch **Dichteschätzer**.

→ Die aus dem Histogramm abgeleitete **Häufigkeitsdichte** ist ein Schätzer für die **Wahrscheinlichkeitsdichte** $f(x)$ des **Merkmals**.

Die **Häufigkeitsdichte** ist selbst eine **Wahrscheinlichkeitsdichte**:

1. $\hat{f} \geq 0$ für alle $x \in \mathbb{R}$
2. Für $x \in (g_j, g_{j+1}]$ ist sie konstant mit Wert

$$\hat{f}(x) = I_j = \frac{f_j}{g_{j+1} - g_j}$$

so dass

$$\int_{g_j}^{g_{j+1}} \hat{f}(x) dx = (g_{j+1} - g_j) \hat{f}(x) = f_j$$

Summation über j liefert daher den Wert 1 und somit

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{f}(x) dx &= \int_{g_1}^{g_{k+1}} \hat{f}(x) dx \\ &= \int_{g_1}^{g_2} \hat{f}(x) dx + \dots + \int_{g_k}^{g_{k+1}} \hat{f}(x) dx \\ &= \sum_{j=1}^k f_j = 1 \end{aligned}$$

Median

x_{med} heißt **Median** von x_1, \dots, x_n , wenn

- mind. 50% der Daten kleiner oder gleich x_{med} sind und
- mind. 50% der Daten größer oder gleich x_{med} sind

Berechnung

- n ungerade: $x_{\text{med}} = x_{(k)}$, $k = \frac{n+1}{2}$
- n gerade: Jede Zahl des Intervalls $[x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}]$

Konvention:

$$x_{\text{med}} = \begin{cases} x_{(\frac{n+1}{2})}, & n \text{ ungerade} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & n \text{ gerade} \end{cases}$$

Eigenschaften

- Vollzieht affin-lineare Transformation nach

$$y_i = a + b \cdot x_i, \quad i = 1, \dots, n$$

Dann gilt: $y_{\text{med}} = a + b \cdot x_{\text{med}}$

- Vollzieht monotone Transformationen $f(x)$ nach

$$y_i = f(x_i), \quad i = 1, \dots, n$$

Dann gilt: $y_{\text{med}} = f(x_{\text{med}})$

- x_{med} minimiert $Q(m) = \sum_{i=1}^n |x_i - m|$

Arithmetisches Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + \dots + x_n}{n}$$

heißt **arithmetisches Mittel** oder **arithmetischer Mittelwert**.

Bei gruppierten Daten mit

- f_1, \dots, f_k : relative **Häufigkeit**
- m_1, \dots, m_k : Gruppenmitten

verwendet man:

$$\bar{x}_g = f_1 m_1 + \dots + f_k m_k$$

Eigenschaften

- Schwerpunkteigenschaft
- Hochrechnung
- Verhalten unter affin-linearen Transformationen
- \bar{x} minimiert $Q(m) = \sum_{i=1}^n (x_i - m)^2$, $m \in \mathbb{R}$

Entropie

Die Kennzahl

$$H = - \sum_{j=1}^k f_j \log(f_j)$$

heißt **Shannon-Wieder-Index** oder **Shannon-Entropie**.

$$J = \frac{H}{\log(k)}$$

heißt **relative Entropie**

Eigenschaften

- $p \leq H \leq \log(k)$
- $0 \leq J \leq 1$
- Minimalwert: 1-Punkt-**Verteilung**
- Maximalwert: **Gleichverteilung** auf k Kategorien

Stichprobenvarianz/ empirische Varianz

Stichprobenvarianz:

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

Bei gruppierten Daten:

$$s_g^2 = \sum_{j=1}^k f_j (m_j - \bar{x}_g)^2$$

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

Eigenschaften

Maßstabänderung von Datenvektoren $\mathbf{x} = (x_1, \dots, x_n)$

$$b \cdot \mathbf{x} = (b \cdot x_1, \dots, b \cdot x_n)$$

Lageänderung

$$\mathbf{x} + a = (x_1 + a, \dots, x_n + a)$$

Rechenregeln

- Invarianz unter Lageänderung

$$\text{var}(a + \mathbf{x}) = \text{var}(\mathbf{x})$$

- Quadratische Reaktion auf Maßstabänderung

$$\text{var}(b \cdot \mathbf{x}) = b^2 \cdot \text{var}(\mathbf{x})$$

Verschiebungssatz

Es gilt:

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

sowie

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

In der Praxis wird die Formel

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

verwendet.

Quantile

Ein **empirisches p-Quantil**, $p \in (0, 1)$, von x_1, \dots, x_n ist jede Zahl \tilde{x}_p , sodass

- mindestens $100 \cdot p\%$ der Datenpunkte $\leq \tilde{x}_p$ und
- mindestens $100 \cdot (1 - p)\%$ der Datenpunkte $\geq \tilde{x}_p$

sind.

1

Berechnung

- np ganzzahlig: Jede Zahl aus $[x_{(np)}, x_{(np+1)}]$. Nicht immer ist das Merkmal metrisch skaliert. Dann sind mitunter nur bestimmte x -Werte interpretierbar, nicht jedoch 'Zwischenwerte'. Dann sind (nur) $x_{(np)}$ und $x_{(np+1)}$ Quantile
- sonst $\tilde{x}_p = x_{(\lfloor np \rfloor + 1)}$.

Konvention: Intervallmitte: $\frac{1}{2}(x_{(np)} + x_{(np+1)})$

¹Formelsammlung S. 3

Quartile

Quartile:

$Q_1 = \tilde{x}_{0.25}$: unteres Quartil (grenzt das untere Viertel ab)

$Q_2 = \tilde{x}_{0.5}$: Median (grenzt die untere Hälfte ab, teilt die **Verteilung**)

$Q_3 = \tilde{x}_{0.75}$ oberes Quartil (grenzt das obere Viertel ab)

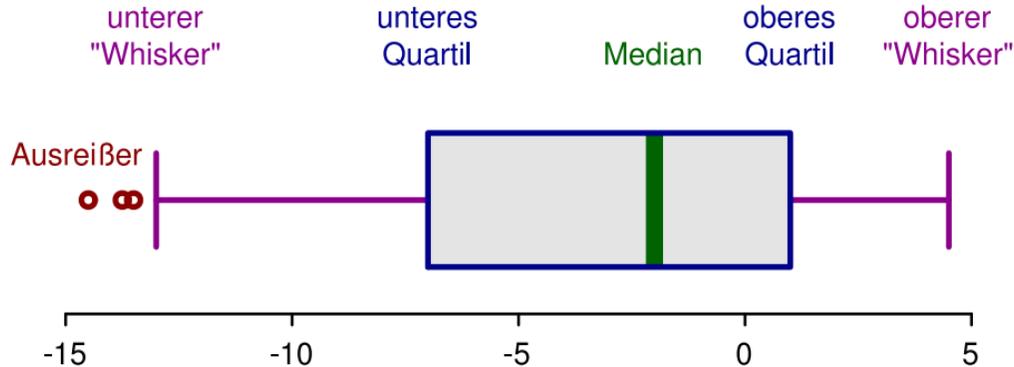
Zwischen Q_1 und Q_3 liegen die zentralen 50% der Datenpunkte (die Mitte)!

$Q_3 - Q_1$ heißt **Interquartilabstand IQR** und ist ein robustes Streuungsmaß.

Fünfpunkte-Zusammenfassung und Boxplot

Die 5 Statistiken (Kennzahlen) x_{\min} , Q_1 , $\tilde{x} = x$, Q_3 , x_{\max} heißen **Fünfpunkte-Zusammenfassung**.

Boxplot: Grafische Darstellung der 5-Punkte-Zusammenfassung:



Schließende Statistik

- Gegeben: Verrauschte (zufallsbehaftete) Daten $X_1, \dots, X_n \sim F_\vartheta$.
- Gesucht: Das Modell F_ϑ also ϑ .
- Ziel: Schließe aus den Daten auf das zugrunde liegende Modell.
- Relevante Schritte:
 1. Gute Modellklasse für die Daten finden.
 2. Schätzen des Modells aus den Daten.
 3. Testen: Gilt $\vartheta \in \Theta_0$ oder $\vartheta \in \Theta_1$?
 4. Untersuche, ob das Modell die Daten gut erklärt.

Modellierung
Schätzen
Testen
Modellvalidierung

Grundbegriffe

Stichprobe

X_1, \dots, X_n heißt **Stichprobe** vom **Stichprobenumfang** n , wenn

$$X_1, \dots, X_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$$

Zufallsvariablen auf einem **Wahrscheinlichkeitsraum** (Ω, \mathcal{A}, P) sind. Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n)$ nimmt Werte im **Stichprobenraum**

$$\mathcal{X} = \{\mathbf{X}(\omega) : \omega \in \Omega\} \subset \mathbb{R}^n$$

an. Realisierungen: Vektoren $(x_1, \dots, x_n) \in \mathcal{X}$.

Hinweis:

In der Statistik interessiert i.d.R. der zugrunde liegende **Wahrscheinlichkeitsraum** (Ω, \mathcal{A}, P) nicht, sondern lediglich der **Stichprobenraum** \mathcal{X} und die **Verteilung** $P_{\mathbf{X}}$ von $\mathbf{X} = (X_1, \dots, X_n)'$ hierauf!

Verteilungsmodell

Eine Menge \mathcal{P} von (möglichen) **Verteilungen** auf \mathbb{R}^n (für die **Stichprobe** (X_1, \dots, X_n)) heißt **Verteilungsmodell**.

\mathcal{P} heißt **parametrisches Verteilungsmodell**, falls

$$\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$$

für eine Menge $\Theta \in \mathbb{R}^k$ von Parametervektoren.

Θ : **Parameterraum** d.h. Es gibt eine Bijektion $\mathcal{P} \leftrightarrow \Theta$

Ein **Verteilungsmodell**, das nicht durch einen endlichdimensionalen Parameter parametrisiert werden kann, heißt **nichtparametrisches Verteilungsmodell**.

Statistik, Schätzfunktion, Schätzer

Statistik

Sei X_1, \dots, X_n eine **Stichprobe** (und o.E. $\mathcal{X} = \mathbb{R}^n$).

Eine Abbildung

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^d$$

mit $d \in \mathbb{N}$ (oft: $d = 1$) heißt **Statistik**. Bildet T in den Paramterraum ab, d.h.

$$T : \mathbb{R}^n \rightarrow \Theta$$

dann heißt T **Schätzfunktion** oder kürzer **Schätzer** (für ϑ).

Allgemein: Schätzung von Funktionen $g(\vartheta)$ von θ durch Statistiken

$T : \mathbb{R}^n \rightarrow \Gamma$ mit $\Gamma = g(\Theta) = \{g(\vartheta) \mid \vartheta \in \Theta\}$.

Empirische Verteilungsfunktion

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(X_i), \quad x \in \mathbb{R}$$

Hierbei: $\mathbf{1}_{(-\infty, x]}(X_i) = \mathbf{1}(X_i \leq x)$

$\hat{F}_n(x)$: Anteil der Beobachtungen, die kleiner oder gleich x sind.

- Die Anzahl $n\hat{F}_n(x)$ der Beobachtungen $\leq x$ ist binomialverteilt mit Parametern n und $p(x) = E(\mathbf{1}(X_i \leq x)) = F(x)$
- Daher folgt:

$$E(\hat{F}_n(x)) = P(X_i \leq x) = F(x), \quad \text{Var}(\hat{F}_n(x)) = \frac{F(x)(1 - F(x))}{n}$$

- Nach dem **Hauptsatz der Statistik** konvergiert $\hat{F}_n(x)$ mit **Wahrscheinlichkeit** 1 gegen $F(x)$ (gleichmäßig in x).

Likelihood-Funktion

Sei $p_\vartheta(x)$ eine Zähl**dichte** (in $x \in \mathcal{X}$) und $\vartheta \in \Theta$ ein Parameter. Für eine gegebene (feste) Beobachtung $x \in \mathcal{X}$ heißt die Funktion

$$L(\vartheta | x) = p_\vartheta(x), \quad \vartheta \in \Theta$$

Likelihood-Funktion.

Likelihood-Prinzip

Ein **Verteilungsmodell** ist bei gegebenen Daten plausibel, wenn es die Daten mit hoher **Wahrscheinlichkeit** erzeugt. Entscheide dich für das plausibelste **Verteilungsmodell!**

Maximum-Likelihood-Schätzer

$p_\vartheta(x)$ sei Zähl**dichte** (in $x \in \mathcal{X}$).

$\vartheta \in \Theta \subset \mathbb{R}^k, k \in \mathbb{N}$.

Dann heißt $\hat{\vartheta} = \hat{\vartheta}(x) \in \Theta$ **Maximum-Likelihood-Schätzer (ML-Schätzer)**, wenn für festes x gilt:

$$p_{\hat{\vartheta}} \geq p_\vartheta(x) \quad \text{für alle } \vartheta \in \Theta$$

(Falls Maximum nicht eindeutig, so wähle eines aus)

Hierdurch ist eine Funktion $\hat{\vartheta} : \mathcal{X} \rightarrow \Theta$ definiert.

Likelihood für Dichten

Sei $f_\vartheta(x)$ eine **Dichte**funktion (in x), $\vartheta \in \Theta \subset \mathbb{R}^k, k \in \mathbb{N}$

Für ein festes x heißt die Funktion

$$L(\vartheta | x) = f_\vartheta(x), \quad \vartheta \in \Theta$$

Likelihood-Funktion. $\hat{\vartheta} \in \Theta$ heißt **Maximum-Likelihood-Schätzer**, wenn bei festem x gilt:

$$f_{\hat{\vartheta}}(x) \geq f_\vartheta, \quad \text{für alle } \vartheta \in \Theta$$

Likelihood für Stichproben

Kompakt: $X \sim f_\vartheta(x), f_\vartheta$ eine Zähl**dichte** oder **Dichte**funktion. Dann ist

$$L_\vartheta(x) = f_\vartheta(x)$$

Sei nun speziell $\mathbf{X} = (X_1, \dots, X_n)'$ mit

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F$$

Dann ist die gemeinsame (Zähl)-**Dichte** die Produkt-(Zähl)-**Dichte**.

Also:

$$L_\vartheta(x) = f_\vartheta \cdots f_\vartheta(x_n)$$

(Gilt für Zähl**dichten** und **Dichte**funktionen).

Schätzer

Unterscheide (konzeptionell):

- Die Abbildung $\hat{\vartheta}_n = \hat{\vartheta}_n(x)$,

$$x = (x_1, \dots, x_n) \mapsto \hat{\vartheta}_n(x)$$

die jeder Realisation x des **Stichproben**raums \mathcal{X} einen Schätzwert zuordnet; gedanklich nach Durchführung des Zufallsexperiments.

- Die Abbildung $\hat{\Theta}_n = \hat{\vartheta}_n(X)$,

$$X = (X_1, \dots, X_n) \mapsto \hat{\vartheta}_n(X)$$

die jedem (zufälligen) Vektor X die Zufallsgröße $\hat{\vartheta}_n(X)$ zuordnet; gedanklich vor Durchführung des Zufallsexperiments.

Es ist üblich, in beiden Fällen $\hat{\vartheta}_n$ zu schreiben und von einem ‘Schätzer’ zu sprechen. Ob die Statistik (als **Zufallsvariable** bzw. Zufallsvektor) oder eine Realisation derselben gemeint ist, muss aus dem Kontext erschlossen werden.

Erwartungstreue

Ein Schätzer $\hat{\vartheta}_n$ heißt **erwartungstreu für** ϑ , wenn für alle $\vartheta \in \Theta$ gilt:

$$E(\hat{\vartheta}_n) = \vartheta$$

$g(\hat{\vartheta}_n)$ heißt **erwartungstreu für** $g(\vartheta)$, wenn für alle $\vartheta \in \Theta$ gilt:

$$E(g(\hat{\vartheta}_n)) = g(\vartheta)$$

Sinngemäß gelten diese Definitionen auch für nichtparametrische Modelle: T_n heißt erwartungstreu für eine Kenngröße $g(F)$, wenn $E(T_n) = E_F(T_n) = g(F)$ für alle **Verteilungsfunktionen** F der betrachteten **Verteilungsklasse**. Hierbei deutet $E_F(\cdot)$ an, dass der **Erwartungswert** unter der Annahme $X_i \sim F$ berechnet wird.

Anschauung

- Wende erwartungstreuen Schätzer N Mal auf **Stichproben** vom Umfang n an.
- N Schätzungen: $\hat{\vartheta}_n(1), \dots, \hat{\vartheta}_n(N)$
- Wende Gesetz der großen Zahlen an:

$$\frac{1}{N} \sum_{i=1}^N \hat{\vartheta}_n(i) \rightarrow E(\hat{\vartheta}_n(i)) = E(\hat{\vartheta}_n)$$

- $\hat{\vartheta}_n$ erwartungstreu: rechte Seite ist ϑ unabhängig von $\vartheta \in \Theta$
- sonst: rechte Seite $\neq \vartheta$

Werden Schätzungen aus einer täglichen **Stichprobe** vom Umfang n über einen langen Zeitraum gemittelt, so schwankt dieses Mittel um $E(\hat{\vartheta})$. Bei einer erwartungstreuen Schätzfunktion also um den wahren Wert ϑ .

Verzerrung (Bias)

Die Verzerrung (engl.: Bias) wird gemessen durch

$$\text{Bias}(\hat{\vartheta}_n; \vartheta) = E_{\vartheta}(\hat{\vartheta}) - \vartheta$$

Gütekriterien

(Asymptotische) Erwartungstreue, Unverfälschtheit

Ein Schätzer $\hat{\vartheta}$ für einen Parameter ϑ heißt **asymptotisch erwartungstreu für ϑ** , wenn für alle ϑ

$$E_{\vartheta}(\hat{\vartheta}) \rightarrow \vartheta$$

gilt.

Konsistenz

Ein Schätzer $\hat{\vartheta}_n = T(X_1, \dots, X_n)$ basierend auf einer **Stichprobe** vom Umfang n heißt **(schwach) konsistent für ϑ** , falls

$$\hat{\vartheta}_n \xrightarrow{P}, \quad n \rightarrow \infty$$

Gilt sogar fast sichere Konvergenz, dann heißt $\hat{\vartheta}_n$ **stark konsistent für ϑ** .

- Ist $\hat{\vartheta}_n$ konsistent für ϑ und ist g stetig, dann ist $g(\hat{\vartheta}_n)$ konsistent für den abgeleiteten Parameter $g(\vartheta)$.
- Die obige Aussage gilt auch für vektorwertige Parameter und ihre Schätzer. Insbesondere folgt aus der Konsistenz von $\hat{\vartheta}_n$ für ϑ und $\hat{\xi}_n$ für ξ die Konsistenz von $\hat{\vartheta} \pm \hat{\xi}_n$ für $\vartheta \pm \xi$.

Effizienz

1. Sind T_1 und T_2 zwei erwartungstreue Schätzer für ϑ und gilt $\text{Var}(T_1) < \text{Var}(T_2)$, so heißt T_1 **effizienter** als T_2 .
2. T_1 ist **effizient**, wenn T_1 effizienter als jede andere erwartungstreue Schätzfunktion ist.

Mittlerer quadratischer Fehler/ MSE

Der MSE ist das wichtigste Gütemaß für Bewertung und Vergleiche von Schätzern. Er integriert die **Varianz** (als Streuungsmaß) und den Bias in einer Kennzahl.

$$\text{MSE}(\hat{\vartheta}_n; \vartheta) = E_{\vartheta} \left((\hat{\vartheta}_n - \vartheta)^2 \right)$$

Additive Zerlegung:

Ist $\hat{\vartheta}_n$ eine Schätzfunktion mit $\text{Var}_{\vartheta}(\hat{\vartheta}_n) < \infty$, dann gilt die additive Zerlegung

$$\text{MSE}(\hat{\vartheta}_n; \vartheta) = \text{Var}_{\vartheta}(\hat{\vartheta}_n) + [\text{Bias}(\hat{\vartheta}_n; \vartheta)]^2$$

MS-Effizienz:

1. Sind T_1 und T_2 zwei Schätzer für ϑ und gilt $\text{MSE}(T_1; \vartheta) < \text{MSE}(T_2; \vartheta)$, so heißt T_1 **effizienter** als T_2 .
2. T_1 ist **effizient**, wenn T_1 effizienter als jede andere erwartungstreue Schätzfunktion ist.

Der t -Test (für eine Stichprobe)

Einseitiger t -Test (1)

Der einseitige t -Test verwirft die Nullhypothese $H_0 : \mu \leq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu > \mu_0$, wenn $T > t(n-1)_{1-\alpha}$.

Einseitiger t -Test (2)

Der einseitige t -Test verwirft die Nullhypothese $H_0 : \mu \geq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu < \mu_0$, wenn $T < -t(n-1)_{1-\alpha} = t(n-1)_{\alpha}$.

Zweiseitiger t -Test

Der zweiseitige t -Test verwirft die Nullhypothese $H_0 : \mu = \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu \neq \mu_0$, wenn $|T| > t(n-1)_{1-\alpha/2}$.

p-Wert

Durchführung eines statistischen Tests:

1. Formuliere H_0 und H_1
2. Wähle Signifikanzniveau α .
3. Bestimme kritischen Wert c_{krit}
4. Berechne t_{obs}
5. Vergleiche t_{obs} mit c_{krit}

Einseitige Tests

Testproblem: $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$

$$p = P_{\mu_0}(T > t_{\text{obs}})$$

Testproblem: $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$

$$p = P_{\mu_0}(T < t_{\text{obs}})$$

Lehne H_0 genau dann ab, wenn $p < \alpha$.

Zweiseitiger Test

Testproblem: $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$

$$p_{\text{zweis}} = P_{\mu_0}(|T| > |t_{\text{obs}}|)$$

Lehne H_0 genau dann ab, wenn $p_{\text{zweis}} < \alpha$

Einseitige Tests gegeben p_{zweis}

Gegeben: t_{obs} und p_{zweis} .

1. Lehne $H_0 : \mu \leq \mu_0$ zugunsten von $H_1 : \mu > \mu_0$ ab, falls

$$t_{\text{obs}} \geq 0 \text{ und } \frac{p_{\text{zweis}}}{2} \stackrel{t_{\text{obs}} > 0}{=} P_{\mu_0}(T > t_{\text{obs}}) < \alpha$$

2. Lehne $H_0 : \mu \geq \mu_0$ zugunsten von $H_1 : \mu < \mu_0$ ab, falls

$$t_{\text{obs}} \leq 0 \text{ und } \frac{p_{\text{zweis}}}{2} \stackrel{t_{\text{obs}} < 0}{=} P_{\mu_0}(T > t_{\text{obs}}) < \alpha$$

Gütefunktion

Die Funktion

$$G(\mu) = P_{\mu}(„H_1“) = P(„H_1“ \mid \mu, \sigma^2), \quad \mu \in \mathbb{R}$$

heißt Gütefunktion (an der Stelle μ).

Formel für die Güte des einseitigen Gaußtests:

$$G(\mu) = \Phi\left(-z_{z-\alpha} + \frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$$

Analog für den zweiseitigen Test:

$$G_{\text{zweis.}}(\mu) = 2\Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$$

Hinweis: In der Praxis wird mit σ aus Trainingsdaten (historischen Daten) durch S geschätzt.

2-Stichproben-Tests

Verbundenes Design

Für $i = 1, \dots, n$ erhebe

X_i : Messung an der i -ten Versuchseinheit vorher

Y_i : Messung an der i -ten Versuchseinheit nachher

Modell: Bivariate einfache **Stichprobe**

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

von normalverteilten Zufallsvektoren mit

$$\mu_X = E(X_i) \quad \text{und} \quad \mu_Y = E(Y_i)$$

Betrachte die Differenzen (nachher - vorher):

$$D_i = Y_i - X_i, \quad i = 1, \dots, n$$

Erwartungswert der Differenzen:

$$E(D_i) = E(Y_i) - E(X_i) = \mu_Y - \mu_X = \delta$$

Es sei $\sigma_D^2 = \text{Var}(D_1) = \dots = \text{Var}(D_n)$ unbekannt.

Verwerfe dann

$$H_0 : \delta = 0 \Leftrightarrow \mu_X = \mu_Y \quad (\text{kein Effekt})$$

zugunsten von

$$H_1 : \delta \neq 0 \Leftrightarrow \mu_X \neq \mu_Y \quad (\text{Effekt vorhanden})$$

falls

$$|T| > t(n-1)_{1-\frac{\alpha}{2}}$$

wobei

$$T = \sqrt{n} \frac{\bar{D}}{S_D} \quad \text{mit} \quad \bar{D} = \frac{1}{n} \sum_{i=1}^n D_i, \quad S_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2}$$

Unverbundenes Design

Modell: Zwei unabhängige **Stichproben**

$$X_{11}, \dots, X_{1n_1} \stackrel{\text{i.i.d.}}{\sim} N(\mu_1, \sigma_1^2)$$

$$X_{21}, \dots, X_{2n_2} \stackrel{\text{i.i.d.}}{\sim} N(\mu_2, \sigma_2^2)$$

Schritte:

1. Test auf **Varianzhomogenität**: Gilt $\sigma_1^2 = \sigma_2^2$?

2. Test auf **Lageunterschied**: Gilt $\mu_1 = \mu_2$?

Test auf Varianzhomogenität

Testproblem $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$ **Varianzschätzungen:**

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2, \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2$$

Teststatistik: $F = \frac{S_1^2}{S_2^2}$. Unter $H_0 : \sigma_1^2 = \sigma_2^2$ ist F F -verteilt.

Test

H_0 ablehnen, falls

$$F < F(n_1 - 1, n_2 - 1)_{\frac{\alpha}{2}} \quad \text{oder} \quad F > F(n_1 - 1, n_2 - 1)_{1-\frac{\alpha}{2}}$$

Äquivalent: Nummeriere so, dass $S_1^2 \leq S_2^2$ und lehne H_0 ab, falls $F < F(n_1 - 1, n_2 - 1)_{\frac{\alpha}{2}}$.

Test auf Lageunterschied

Annahme: $\sigma_1 = \sigma_2 =: \sigma$ (**Varianzhomogenität**).

Testproblem (zweiseitig):

$$H_0 : \mu_1 = \mu_2 \quad (\text{kein Lageunterschied})$$

versus

$$H_1 : \mu_1 \neq \mu_2 \quad (\text{Lageunterschied})$$

Testprobleme (einseitig):

$$H_0 : \mu_1 \geq \mu_2 \quad \text{versus} \quad H_1 : \mu_1 < \mu_2$$

bzw.

$$H_0 : \mu_1 \leq \mu_2 \quad \text{versus} \quad H_1 : \mu_1 > \mu_2$$

Teststatistik:

$$T = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} S} \quad \text{mit}$$

$$S^2 = \frac{n_1 - 1}{n_1 + n_2 - 2} S_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{1j} - \bar{X}_2)^2 \right)$$

1. Lehne $H_0 : \mu_1 = \mu_2$ zugunsten von $H_1 : \mu_1 \neq \mu_2$ ab, wenn $|T| > t(n_1 + n_2 - 2)_{1-\frac{\alpha}{2}}$.

2. Lehne $H_0 : \mu_1 \geq \mu_2$ zugunsten von $H_1 : \mu_1 < \mu_2$ ab, wenn $T > t(n_1 + n_2 - 2)_{\alpha}$.

3. Lehne $H_0 : \mu_1 \leq \mu_2$ zugunsten von $H_1 : \mu_1 > \mu_2$ ab, falls $T < t(n_1 + n_2 - 2)_{1-\alpha}$.

Welch-Test auf Lageunterschied

Bei **Varianz**inhomogenität $\sigma_1^2 \neq \sigma_2^2$ verwendet man den Welch-Test.

Teststatistik:

$$T = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Lehne $H_0 : \mu_1 = \mu_2$ auf dem Niveau α ab, wenn $|T| > t(df)_{1-\frac{\alpha}{2}}$, wobei

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2 \frac{1}{n_1-1} + \left(\frac{S_2^2}{n_2}\right)^2 \frac{1}{n_2-1}}$$

Falls $df \notin \mathbb{N}$, dann vorher auf nächste ganze Zahl abrunden.

Fallzahlplanung

Für $n = n_1 = n_2$ kann man folgende Näherungen verwenden:

Zweiseitiger Test

Wähle

$$n \geq \frac{\sigma^2}{\Delta^2} (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2$$

um eine Schärfe von $1 - \beta$ bei einer Abweichung von $\Delta = |\mu_A - \mu_B|$ näherungsweise zu erzielen.

Einseitiger Test

Wähle

$$n \geq \frac{\sigma^2}{\Delta^2} (z_{1-\alpha} + z_{1-\beta})^2$$

um eine Schärfe von $1 - \beta$ bei einer Abweichung von $\Delta = |\mu_A - \mu_B|$ näherungsweise zu erzielen.

t-Verteilung

Die **Verteilung** von

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

heißt **t-Verteilung** mit $n - 1$ **Freiheitsgraden**.

Notation: $t(n - 1)$

p -Quantil: $t(n - 1)_p$

χ^2 -Verteilung

Sind U_1, \dots, U_k i.i.d. $\sim N(0, 1)$, dann heißt die Verteilung von

$$Q = \sum_{i=1}^k U_i^2$$

χ^2 -Verteilung mit k Freiheitsgraden.

Momente: Es gilt: $E(Q) = k$ und $\text{Var}(Q) = 2k$.

Gilt mit einer Konstanten $c > 0$:

$$\frac{T}{c} \sim \chi^2(k)$$

dann heißt T **gestreckt χ^2 -verteilt mit k Freiheitsgraden**. Man schreibt auch: $T \sim c \cdot \chi^2(k)$

Verteilung der Varianzschätzer

Annahme: Normalverteilungsmodell, d.h. $X_1, \dots, X_n \stackrel{d}{\sim} N(\mu, \sigma^2)$

Fall 1: μ bekannt. Verwende $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$. Dann gilt

$$\frac{n}{\sigma^2} \hat{\sigma}_n^2 \sim \chi^2(n)$$

Fall 2: μ unbekannt. Verwende $\hat{\sigma}_n^2 := S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Dann

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

F-Verteilung

Seien $Q_1 \sim \chi^2(n_1)$ und $Q_2 \sim \chi^2(n_2)$ unabhängig. Dann heißt die **Verteilung** des Quotienten

$$F = \frac{Q_1 / n_1}{Q_2 / n_2}$$

F-Verteilung mit n_1 und n_2 Freiheitsgraden.

Notation: $F(n_1, n_2)$.

p -Quantil: $F(n_1, n_2)_p$.

Momente: $E(F) = \frac{n_2}{n_2-2}$, $\text{Var}(F) = \frac{2n_2^2(n_2+2+n_1-2)}{n_1(n_2-2)^2(n_2-4)}$

Konfidenzintervall

Ein Intervall $[L, U]$ mit datenabhängigen Intervallgrenzen

$$L = L(X_1, \dots, X_n)$$

$$U = U(X_1, \dots, X_n)$$

heißt **Konfidenzintervall (Vertrauensbereich) zum Konfidenzniveau $1 - \alpha$** , wenn für alle $\vartheta \in \Theta$ gilt:

$$P([L, U] \ni \vartheta) \geq 1 - \alpha$$

Konfidenzintervall für μ

Zweiseitiges Konfidenzintervall, σ unbekannt:

$$\left[\bar{X} - t_{(n-1)_{1-\frac{\alpha}{2}}} \frac{S}{\sqrt{n}}, \bar{X} + t_{(n-1)_{1-\frac{\alpha}{2}}} \frac{S}{\sqrt{n}} \right]$$

Einseitiges unteres Konfidenzintervall:

$$\left[-\infty, \bar{X} + t_{(n-1)_{1-\alpha}} \frac{S}{\sqrt{n}} \right]$$

Einseitiges oberes Konfidenzintervall:

$$\left[\bar{X} - t_{(n-1)_{1-\alpha}} \frac{S}{\sqrt{n}}, \infty \right]$$

Falls σ bekannt ist: Ersetze in den Formeln:

1. S durch σ .
2. $t_{(n-1)_{1-\frac{\alpha}{2}}}$ durch $z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$
3. $t_{(n-1)_{1-\alpha}}$ durch $z_{1-\alpha}$

$z_{1-\alpha}$: $(1 - \alpha)$ -Quantil der $N(0, 1)$ -**Verteilung**.

Konfidenzintervall für σ^2

Das zweiseitige Konfidenzintervall für σ^2 liefert:

$$\left[\frac{n-1}{\chi^2(n-1)_{1-\alpha/2}} \hat{\sigma}^2, \frac{n-1}{\chi^2(n-1)_{\alpha/2}} \hat{\sigma}^2 \right]$$

Konfidenzintervall für p

Modell: $Y \sim \text{Bin}(n, p)$

Approximatives Konfidenzintervall (aus ZGWS):

$$L = \hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$U = \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Statistische Testtheorie

Testproblem, Nullhypothese, Alternative

Sind f_0 und f_1 zwei mögliche **Verteilungen** für eine **Zufallsvariable** X , dann wird das **Testproblem**, zwischen $X \sim f_0$ und $X \sim f_1$ zu entscheiden, in der Form

$$H_0 : f = f_0 \quad \text{gegen} \quad H_1 : f = f_1$$

notiert, wobei f die wahre **Verteilung** von X bezeichnet. H_0 heißt **Nullhypothese** und H_1 **Alternative (Alternativhypothese)**.

Statistischer Test

Ein (**statistischer**) **Test** ist eine Entscheidungsregel, die basierend auf T entweder zugunsten von H_0 (Notation: „ H_0 “) oder zugunsten von H_1 („ H_1 “) entscheidet.

Fehler 1. und 2. Art

Entscheidung für H_1 , obwohl H_0 richtig ist, heißt **Fehler 1. Art**. H_0 wird dann fälschlicherweise verworfen. Eine Entscheidung für H_0 , obwohl H_1 richtig ist, heißt **Fehler 2. Art**. H_0 wird fälschlicherweise akzeptiert.

Insgesamt sind vier Konstellationen möglich, die in der folgenden Tabelle zusammengefasst sind:

	H_0	H_1
„ H_0 “	✓	Fehler 2. Art
„ H_1 “	Fehler 1. Art	✓

Signifikanzniveau, Test zum Niveau α

Bezeichnet „ H_1 “ eine Annahme der Alternative und „ H_0 “ eine Annahme der Nullhypothese durch eine Entscheidungsregel, dann ist durch diese Regel ein **statistischer Test zum Signifikanzniveau (Niveau) α** gegeben, wenn

$$P_{H_0}(\text{„}H_1\text{“}) \leq \alpha$$

Genauer ist die linke Seite das tatsächliche Signifikanzniveau des Tests und die rechte Seite das vorgegebene **nominale** Signifikanzniveau.

Hinweis: Die **Wahrscheinlichkeit** eines Fehlers 2. Art wird nicht unbedingt kontrolliert. Dies erfordert eine Planung der **Stichprobengröße**.

Schärfe (Power)

Die **Wahrscheinlichkeit** eines Fehlers 2. Art wird üblicherweise mit β bezeichnet. Die **Gegenwahrscheinlichkeit**,

$$1 - \beta = P_{H_1}(\text{„}H_1\text{“}) (= E_{H_1}(1 - \phi)),$$

dass der Test die Alternative H_1 tatsächlich aufdeckt, heißt **Schärfe (Power)** des Testverfahrens.

Hypothesen (über den Erwartungswert μ)

Einseitiges Testproblem:

$$H_0 : \mu \leq \mu_0 \quad \text{gegen} \quad H_1 : \mu > \mu_0$$

bzw.

$$H_0 : \mu \geq \mu_0 \quad \text{gegen} \quad H_1 : \mu < \mu_0$$

Zweiseitiges Testproblem:

$$H_0 : \mu = \mu_0 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0$$

Wichtig: Der Grenzfall “=” wird immer H_0 zugeschlagen.

Gauß-Test

Gegeben: $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ mit bekannter Varianz $\sigma^2 \in (0, \infty)$

Teststatistik:

$$T = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \quad (\mu_0 \in \mathbb{R} \text{ vorgegebener Sollwert})$$

Verteilung der Teststatistik:

$$T \sim N(0, 1) \quad \text{für} \quad \mu = \mu_0$$

Einseitiger Gauß-Test

1. Der einseitige Gaußtest verwirft die Nullhypothese $H_0 : \mu \leq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu > \mu_0$ wenn $T > z_{1-\alpha}$.
2. Der einseitige Gaußtest verwirft die Nullhypothese $H_0 : \mu \geq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu < \mu_0$, wenn $T < -z_{1-\alpha} = z_\alpha$.

Zweiseitiger Gauß-Test

Der zweiseitige Gauß-Test verwirft die Nullhypothese $H_0 : \mu = \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu \neq \mu_0$, wenn $|T| > z_{1-\frac{\alpha}{2}}$.