Sadistik

Einführung

Teilgebiet	Beschreibung	Techniken
Deskriptive Statistik	Zusammenfassung & Visualisierung von Daten	Grafiken, Tabellen, Kennzahlen
Explorative Statistik	Suche nach Struktur in Daten Schnittstelle zwischen der deskriptiven & induktiven Stat.	Iterative und interaktive Anwendung von Techniken aus der deskriptiven und induktiven Statistik.
Induktive Statistik	Schliesst von beobachteten Daten auf allgemeine Phänomene.	Statistische Modellierung, Statistische Tests, Inferenz

Datenerhebung & Messung

Statistische Einheit, Untersuchungseinheit (UE)
Objekte, an denen Grössen erfasst werden.

$\in$

Grundgesamtheit, Population (GG)
Menge aller, für eine Fragestellung relevanten, statistischen Einheiten.

$\supseteq$

Teilgesamtheit, Stichprobe
Teilmenge der GG

Merkmal, Variable
Messbare Eigenschaft einer UE.

Messung
Abbildung von der Menge der UE zu einer Menge von Zahlen/Symbolen.

Eine gültige Messung ist ein Homomorphismus bzgl. dieser Relation

Merkmalsausprägung
Konkreter Wert eines Merkmals für eine bestimmte UE.

→ Durch die Messbarkeit lässt sich eine Relation zwischen den UE definieren

Beobachtung
Die Gesamtheit der beobachteten Merkmalsausprägungen der gemessenen Merkmale einer UE zu einem bestimmten Zeitpunkt.

Gütekriterien für Messungen

Validität
(Gültigkeit) Wird das gemessen, was gemessen werden soll?

↳ Bias = 0

Reliabilität
(Zuverlässigkeit) Ist die Messung zuverlässig?

↳ $r = \frac{\sigma_X^2}{\sigma_{X^*}^2}$, $X^* = X + U$

X^*: beobachtete Messwerte; U: Messfehler

Skalenniveaus

Skalenniveau	Beispiele	Erlaubte Transformationen um Strukturen zu erhalten	Werte	Ordnung	Differenz	Quotient	Lageparameter
Nominalskala	Wohnort, Farbe	Bijektionen	✓	✗	✗	✗	Mode
Ordinal - Rangskala	Soziale Schicht, Schulbildung	str. monoton steig. Abb.	✓	✓	✗	✗	Median
Intervallskala *	Temperatur, IQ, Jahreszahlen	affin lin. str. mon. steig. Abb.	✓	✓	✓	✗	Arith. Mittel
Verhältnisskala * (Intervallskala mit natürlicher Null) D.h. minimum der Skala ist Null.	Preis, Länge, Gewicht	lineare str. mon. steig. Abb.	✓	✓	✓	✓	Geom. Mittel Harm. Mittel
Absolutskala	Häufigkeit, Anzahl	Identität	✓	✓	✓	✓	〃

* Äquivalenzrelation $S := \text{Skala},\ D := S \times S,\ R = \{(m,n) \in D \times D \mid (m_1,m_2) \in D \wedge (n_1,n_2) \in D \wedge m_1 - m_2 = n_1 - n_2\} \subseteq D \times D$ & Ordnungsrelation

* Äquivalenzrelation $R = \{(m,n) \in D \times D \mid (m_1,m_2) \in D \wedge (n_1,n_2) \in D \wedge \frac{m_1}{m_2} = \frac{n_1}{n_2}\} \subseteq D \times D$ & Ordnungsrelation

Indexbildung

Zusammenfassung von Einzelindikatoren zu einer neuen Variable

Datenerhebung

Vollerhebung

Alle stat. Einheiten einer GG werden untersucht.

Stichprobe

Ein Teil der UE in einer GG wird untersucht.

Querschnittdaten

Eine Beobachtung pro UE.

- Noten, Aktivitäten, Geschlecht, können zu bestimmtem Zeitpunkt von UE erhoben werden und z.B. mittels Regression auf Zusammenhänge untersucht werden.

Zeitreihe

Mehrere Beobachtungen einer UE.

- Temperatur, Wind & Luftfeuchtigkeit werden in regelmässigen Abständen gemessen um Prognosen über die zeitliche Entwicklung der UE 'Wetter' zu machen

Längsschnittdaten

Mehrere Beobachtungen mehrerer UE.

- Kohorentstudien in Medizin
- Mikrozensus

Grundräume

Ergebnisraum

$\Omega \neq \emptyset$ heißt Grundgesamtheit oder Ergebnisraum

$A \subseteq \Omega$ heißt Ereignis.

$\omega_i \in \Omega$ heißt Elementarereignis oder Ergebnis.

σ-Algebra

Ein Mengensystem $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ heißt σ-Algebra über Ω oder abgeschlossenes Mengensystem, falls gilt:

1) $\Omega \in \mathcal{F}$ 2) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$ 3) $A_i \in \mathcal{F}, i \in I \Rightarrow \bigcup_{i \in I} A_i \in \mathcal{F}$

Erzeugende σ-Algebra

Ist $A \subseteq \Omega$, so heißt die Menge $\sigma(A) = \{\emptyset, A, \bar{A}, \Omega\}$ die von A erzeugte σ-Algebra und ist die kleinste σ-Algebra die A enthält.

Satz

Sei $I \neq \emptyset$ eine beliebige Indexmenge und $\mathcal{F}_i$ eine σ-Algebra über Ω $\forall i \in I$.

Dann ist auch $\mathcal{F} := \bigcap_{i \in I} \mathcal{F}_i$ eine σ-Algebra über Ω.

Meßraum / meßbar

Sei $\mathcal{F}$ eine σ-Algebra über Ω und $A \in \mathcal{F}$, dann heißt A meßbar bezüglich $\mathcal{F}$. Das Tupel $(\Omega, \mathcal{F})$ heißt Meßraum (measurable space)

Erzeuger

Sei $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ und Σ die Menge aller σ-Algebren über Ω, die $\mathcal{E}$ enthalten. Dann wird die σ-Algebra $\sigma(\mathcal{E}) := \bigcap_{\mathcal{F} \in \Sigma} \mathcal{F}$ als die von $\mathcal{E}$ erzeugte σ-Algebra $\sigma(\mathcal{E})$ bezeichnet. Gilt umgekehrt für eine σ-Algebra $\mathcal{A}$, $\sigma(\mathcal{E}) = \mathcal{A}$, so heißt $\mathcal{E}$ Erzeuger von $\mathcal{A}$.

Borelsche σ-Algebra

Sei $\Omega = \mathbb{R}$ und $\mathcal{O} = \{(a,b) \mid a,b \in \mathbb{R}, a<b\}$. Dann heißt $\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{O})$ die borelsche σ-Algebra über $\mathbb{R}$. $A \in \mathcal{B}(\mathbb{R})$ heißt borelsche Menge. Für $\Omega = \mathbb{R}^n$ setzen wir $\mathcal{O}^n = \{U \subset \mathbb{R}^n \mid U \text{ offen}\}$ und $\mathcal{B}^n := \mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{O}^n)$

Eigenschaften von $\mathcal{B}$

i) $\emptyset \in \mathcal{B}$, $\mathbb{R} \in \mathcal{B}$

ii) $\forall a,b \in \mathbb{R}, a<b$: $[a,b] \in \mathcal{B} \wedge [a,b) \in \mathcal{B} \wedge (a,b] \in \mathcal{B}$

iii) $\forall a \in \mathbb{R}$: $\{a\} \in \mathcal{B}$

iv) $\mathbb{N} \in \mathcal{B}$, $\mathbb{Q} \in \mathcal{B}$, $\bar{\mathbb{Q}} \in \mathcal{B}$

Wahrscheinlichkeitsmaß

Mathematisches Maß

Maß und Maßraum

Sei $(\Omega, \mathcal{F})$ ein Meßraum. Eine Funktion $\mu: \mathcal{F} \to \bar{\mathbb{R}}$ heißt Maß auf $\mathcal{F}$, falls gilt:

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

ii) $\forall A \in \mathcal{F}: \mu(A) \geq 0$

iii) σ-Additivität: $\mu\left(\dot{\bigcup}_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$

Wir nennen $(\Omega, \mathcal{F}, \mu)$ Maßraum (measure space).

Reduziertes Maß

Sei μ ein Maß auf $\mathcal{F}$ und $A \in \mathcal{F}$. Dann heißt $\mu_{|A}$ reduziertes Maß und ist definiert durch $\mu_{|A}(B) := \mu(A \cap B) \, \forall B \in \mathcal{F}$.

Maßeindeutigkeitssatz

Seien μ, ν zwei Maße auf $(\Omega, \mathcal{F})$ und $\mathcal{E}$ ein durchschnittsstabiler Erzeuger von $\mathcal{F}$ mit den Eigenschaften:

i) $\mu(E) = \nu(E)$ $\forall E \in \mathcal{E}$.

ii) Es gibt eine Folge $(E_n)_{n \in \mathbb{N}}$ disjunkter Mengen aus $\mathcal{E}$ mit $\mu(E_n) = \nu(E_n) < \infty$ und $\dot{\bigcup}_{n=1}^{\infty} E_n = \Omega$.

Dann sind die beiden Maße identisch: $\mu = \nu$.

Indikatorfunktion

Sei $A \subseteq \Omega$. $I_A: \Omega \to \mathbb{R}$, $I_A(\omega) := \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$

Satz

Seien $A, B \subseteq \Omega$. Dann gilt

i) $I_{\bar{A}} = 1 - I_A$ ii) $I_{A \cap B} = I_A \cdot I_B$

iii) $A \cap B = \emptyset \Rightarrow I_{A \cup B} = I_A + I_B$ iv) $A \subseteq B \Rightarrow I_{B \setminus A} = I_B - I_A$

Terminologie

Gibt es eine Folge $(A_i)_{i \in \mathbb{N}}$ in $\mathcal{F}$, s.d. $\bigcup_{i=1}^{\infty} A_i = \Omega$ und $\mu(A_i) < \infty$ $\forall i \in \mathbb{N}$, so heißt μ σ-endlich.

Ist außerdem $\mu(\Omega) < \infty$, so heißt μ endlich.

Ist außerdem $\mu(\Omega) = \infty$, so heißt μ normiert.

Zählmaß

$$\mu_Z(A) := \begin{cases} |A|, & A \text{ endlich} \\ +\infty, & \text{sonst} \end{cases}$$

! μ_Z ist σ-endlich, wenn Ω abzählbar ist.

Durchschnittsstabil / π-System

Ein Mengensystem $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ heißt durchschnittsstabil, oder π-System, wenn gilt:

$A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.

! $\mathcal{O}$ ist durchschnittsstabil

Lebesgue-Maß

Für $B \in \mathcal{O}^n$ mit $B = \prod_{i=1}^{n} (a_i, b_i)$ definieren wir $\mathrm{vol}(B) := (b_1 - a_1) \cdot \ldots \cdot (b_n - a_n)$. Für $A \subseteq \mathbb{R}^n$ definieren wir $\lambda^*(A) = \inf\left\{ \sum_{B \in \mathcal{C}} \mathrm{vol}(B) \mid A \subseteq \bigcup_{i=1}^{\infty} B_i = \mathcal{C} \right\}$.

Das Lebesgue-Maß ist definiert als $\lambda(A) = \lambda^*(A)$ für alle Lebesgue-messbaren Mengen A, d.h. für alle $A \subseteq \mathbb{R}^n$ für die gilt $\forall S \subseteq \mathbb{R}^n$: $\lambda^*(S) = \lambda^*(S \cap A) + \lambda^*(S \setminus A)$

Dirac Maß / Dirac Verteilung

Sei $\delta_\omega: \mathcal{F} \to \mathbb{R}$, $\delta_\omega(A) := I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$.

δ_ω heißt Dirac Maß.

Für $\omega \in \Omega$ heißt $\mathbb{P}(\omega) = \delta_\omega$ Dirac Verteilung oder Einpunktverteilung.

Es gilt also $\mathbb{P}(\omega) = 1 \wedge \mathbb{P}(\omega') = 0$ $\forall \omega' \in \Omega, \omega' \neq \omega$.

Wahrscheinlichkeitsmaß

Wahrscheinlichkeitsmaß / Verteilung

Sei $(\Omega, \mathcal{F})$ ein Meßraum und $\mathbb{P}: \mathcal{F} \to [0,1]$ ein Maß auf $(\Omega, \mathcal{F})$ mit $\mathbb{P}(\Omega) = 1$.
Dann heißt $\mathbb{P}$ Wahrscheinlichkeitsmaß oder Verteilung und ordnet jedem Ereignis $A \in \mathcal{F}$ die Wahrscheinlichkeit $\mathbb{P}(A)$ zu.
$(\Omega, \mathcal{F}, \mathbb{P})$ heißt Wahrscheinlichkeitsraum.

μ-fast-überall / P-fast-sicher

Ist $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $A \in \mathcal{F}$ mit $\mu(A) = 0$, so heißt A μ-Nullmenge.
Eine Eigenschaft E gilt μ-fast-überall in Ω, wenn:
$\exists N \in \mathcal{F}: \mu(N) = 0 \wedge \forall \omega \in \Omega \setminus N$: E gilt für ω.
Wir sagen E gilt $\mathbb{P}$-fast-sicher, wenn es für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ E $\mathbb{P}$-fast-überall gilt.

Bedingte Wahrscheinlichkeit

Ist $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathbb{P}(B) > 0$ für $B \in \mathcal{F}$, so heißt $\mathbb{P}(\cdot \mid B): \mathcal{F} \to [0,1]$, $A \mapsto \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
die bedingte Wahrscheinlichkeitsverteilung bzw. bedingte Wahrscheinlichkeit von A gegeben B.

Satz

$\mathbb{P}(\cdot \mid B)$ ist ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$.

Laplace-Wahrscheinlichkeit

Sei $(\Omega, \mathcal{F})$ ein Meßraum, $|\Omega| < \infty$ und $\mathbb{P}: \mathcal{F} \longrightarrow [0,1]$ ein Maß auf $(\Omega, \mathcal{F})$ mit $\mathbb{P}[A] := \frac{|A|}{|\Omega|}$. Wir nennen $\mathbb{P}$ die Laplace-Wahrscheinlichkeit.

Eigenschaften

Eigenschaften eines Maßes

Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $A, B, A_n \in \mathcal{F}$ $\forall n \in \mathbb{N}$. Dann gilt:

i) Endliche Additivität: $A \cap B = \emptyset \Rightarrow \mu(A \cup B) = \mu(A) + \mu(B)$.
ii) Subadditivität: $A \subseteq B \wedge \mu(A) < \infty \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$
iii) Monotonie: $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$
iv) Sub-σ-Additivität: $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

Eigenschaften eines Wahrscheinlichkeitsmaßes

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B, A_n \in \mathcal{F}$, $\forall n \in \mathbb{N}$.
Dann gilt:
i) $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$
ii) $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ und $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$
iii) $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
iv) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Stetigkeit des Maßes

Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $A, A_n \in \mathcal{F}$, $n \in \mathbb{N}$.
Dann gilt:
i) Stetigkeit von unten: Sei $A_1 \subseteq A_2 \subseteq \ldots \Rightarrow \mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k \to \infty} \mu(A_k)$
ii) Stetigkeit von oben: Sei $A_1 \supseteq A_2 \supseteq \ldots$ und $\mu(A_1) < \infty \Rightarrow \mu\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k \to \infty} \mu(A_k)$

Stetigkeit des Wahrscheinlichkeitsmaßes

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein W-maßraum und $A, A_1, A_2, \ldots \in \mathcal{F}$
Dann gilt:
i) Sei $A_1 \subseteq A_2 \subseteq \ldots \Rightarrow \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k \to \infty} \mathbb{P}(A_k)$
ii) Sei $A_1 \supseteq A_2 \supseteq \ldots \Rightarrow \mathbb{P}\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k \to \infty} \mathbb{P}(A_k)$

Satz

- $P[\bigcap_{i=1}^{n} A_i]^{\neq\emptyset} = \prod_{i=1}^{n} P[A_i \mid \bigcap_{j=1}^{i-1} A_j] = P[A_1] \cdot P[A_2 \mid A_1] \cdot P[A_3 \mid A_1 \cap A_2] \cdot \ldots \cdot P[A_n \mid A_1 \cap \ldots \cap A_{n-1}]$ (Multiplikationssatz)
- Sei $(A_i)_{i \in I}$ eine disjunkte Zerlegung von Ω. D.h. $\Omega = \dot{\bigcup}_{i \in I} A_i$. Dann gilt für beliebiges B
  $P[B] = \sum_{i: P[A_i] > 0} P[B \mid A_i] \cdot P[A_i]$ (Satz von totaler Wahrscheinlichkeit)
  Spezialfall: $P(B) = P(B \mid A) \cdot P(A) + P(B \mid A^c) \cdot P(A^c)$

Siebformel von Sylvester-Poincaré

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein W-maßraum und $A_1, A_2, \ldots \in \mathcal{F}$,
dann gilt die Verallgemeinerung von (iii) von oben:

$$P\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{k=1}^{n} (-1)^{k+1} \cdot \sum_{1 \leq i_1 < \ldots < i_k \leq n} P[A_{i_1} \cap \ldots \cap A_{i_k}]$$

$$= \sum_{i=1}^{n} P[A_i] - \sum_{1 \leq i < j \leq n} P[A_i \cap A_j]$$

$$+ \sum_{1 \leq i < j < k \leq n} P[A_i \cap A_j \cap A_k] - \ldots + (-1)^{n+1} \cdot P\left[\bigcap_{i=1}^{n} A_i\right]$$

Spezialfall ($n=2$): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Satz von Bayes

$(A_i)_{i \in I}$ sei so, dass $\Omega = \dot{\bigcup}_{i \in I} A_i$. B sei so, dass $P[B] \neq 0$.

Dann ist $P[A_i \mid B] = \frac{P[B \mid A_i] \cdot P[A_i]}{\sum_j P[B \mid A_j] \cdot P[A_j]}$

Spezialfall: $P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} = \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid A^c) \cdot P(A^c)}$

Wettverhältnis (Odds-Update)

$$\underbrace{\frac{P[B \mid A]}{P[B^c \mid A]}}_{\text{a-posteriori Verhältnis}} = \overbrace{\frac{P[A \mid B]}{P[A \mid B^c]}}^{\text{Faktor der neuen Information}} \cdot \underbrace{\frac{P[B]}{P[B^c]}}_{\text{a-priori-Verhältnis}}$$

Sensitivität: $P(T^+ \mid K^+)$
Spezifität: $P(T^- \mid K^-)$
Prävalenz: $P(K)$
Relevant: $P(K \mid T)$, $P(K^- \mid T^-)$

Zufallsvariablen (Theorie)

Urbild

Seien X, Y nicht-leere Mengen. Für eine Abbildung $f: X \to Y$ und $A \in \mathcal{P}(Y)$ definieren wir $f^{-1}(A) = \{x \in X \mid f(x) \in A\} = \{f \in A\}$. f^{-1} heißt Urbild von A.

Eigenschaften:

i) $f^{-1}(A^c) = (f^{-1}(A))^c \quad \forall A \in \mathcal{P}(Y)$.

ii) $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B) \quad \forall A, B \in \mathcal{P}(Y)$

iii) $f^{-1}\left(\bigcup_{k=1}^{\infty} A_k\right) = \bigcup_{k=1}^{\infty} f^{-1}(A_k) \quad \forall \{A_k\}_{k \in \mathbb{N}} \subseteq \mathcal{P}(Y)$

Satz

Sei $(Y, \mathcal{F}_1)$ ein Meßraum, dann ist $(X, f^{-1}(\mathcal{F}_1))$ ein Meßraum. D.h. $f^{-1}(\mathcal{F}_1) := \{f^{-1}(A) \mid A \in \mathcal{F}_1\}$ ist eine σ-Algebra.

Messbare Abbildung

Seien $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ Meßräume. Eine Abbildung $f: \Omega_1 \to \Omega_2$ heißt $\mathcal{F}_1$-$\mathcal{F}_2$-messbar, wenn $f^{-1}(\mathcal{F}_2) \subseteq \mathcal{F}_1$.
Man spricht abgekürzt von einer messbaren Abbildung f, falls die involvierten σ-Algebren eindeutig bekannt sind.

Satz

Seien $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ Meßräume, wobei $\mathcal{F}_2 = \sigma(\mathcal{E})$ von einem Mengensystem $\mathcal{E}$ erzeugt ist. Die Abb. $f: \Omega_1 \to \Omega_2$ ist genau dann $\mathcal{F}_1$-$\mathcal{F}_2$-messbar, wenn $f^{-1}(\mathcal{E}) \subseteq \mathcal{F}_1$.

Bildmaß

Ist $(\Omega_1, \mathcal{F}_1, \mu)$ ein Maßraum und $(\Omega_2, \mathcal{F}_2)$ ein Meßraum und $f: \Omega_1 \to \Omega_2$ messbar, so heißt $\mu_f: \mathcal{F}_2 \to [0, +\infty)$, $\mu_f(A) := \mu(f^{-1}(A))$ das Bildmaß μ_f von μ unter f. $(\Omega_2, \mathcal{F}_2, \mu_f)$ ist ein Maßraum.

Beispiel

Für Meßräume $(\Omega_1, \mathcal{F})$ und $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$ heisst die Abb. $f: \Omega_1 \to \overline{\mathbb{R}}$ $\mathcal{F}$-$\overline{\mathcal{B}}$-messbar, wenn $f^{-1}(\{-\infty\}), f^{-1}(\{+\infty\}) \in \mathcal{F}$ und $f^{-1}((-\infty, a)) \in \mathcal{F} \; \forall a \in \mathbb{R}$.

Zufallsvariablen

Zufallsvariable

Ist $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\Omega_2, \mathcal{F}_2)$ ein Meßraum, so heißt eine $\mathcal{F}_1$-$\mathcal{F}_2$-messbare Abb. $X: \Omega_1 \to \Omega_2$ Zufallsvariable (ZV).

Verteilung einer Zufallsvariable

Ist $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\Omega_2, \mathcal{F}_2)$ ein Meßraum und $X: \Omega_1 \to \Omega_2$ eine ZV, so heißt das Bildmaß $\mathbb{P}_X$ von $\mathbb{P}$ unter X die Verteilung von X.
Das Bildmaß $\mathbb{P}_X$ von $\mathbb{P}$ unter X, $\mathbb{P}_X(A) := \mathbb{P}(\overbrace{X^{-1}(A)}^{\in \mathcal{F}_1}) \in [0,1] \; \forall A \in \mathcal{F}_2$, ist wieder ein Wahrscheinlichkeitsmaß.

! Unterschied zu Dichte f in Statistik 1:
$f: \Omega \to \mathbb{R}$, $f(x) := \mathbb{P}(X^{-1}(\{x\}))$ → funktioniert nicht für stetige ZV.
$\mathbb{P}_X: \mathcal{F} \to \mathbb{R}$, $\mathbb{P}_X(\{x\}) := \mathbb{P}(X^{-1}(\{x\}))$

Beispiel

Sei $\Omega = \{0,1\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$. $\mathbb{P}(\{\vec{\omega}\}) = p^k (1-p)^{n-k}$, wobei k die Anzahl 1er und p die Wahrscheinlichkeit für eine 1 beschreiben.
Sei $X: \Omega \to \mathbb{N}_0$ die messbare Abb. die jedem $\vec{\omega} \in \Omega$ die Anzahl 1er zuordnet.
Dann ist $\mathbb{P}_X(\{k\}) = \mathbb{P}(X^{-1}(\{k\})) = \sum_{\vec{\omega} \in X^{-1}(\{k\})} \mathbb{P}(\{\vec{\omega}\}) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$

Satz

Sind $(\Omega_i, \mathcal{F}_i)$, $i = 1,2,3$ Meßräume und $f: \Omega_1 \to \Omega_2$ und $g: \Omega_2 \to \Omega_3$ messbar. Dann ist auch $f \circ g: \Omega_1 \to \Omega_3$ messbar.

Satz

Sei $X: \Omega \to \mathbb{R}^n$ eine ZV und $g: \mathbb{R}^n \to \mathbb{R}^m$ meßbar. Dann ist $g \circ X: \Omega \to \mathbb{R}^m$ eine ZV.

Satz

Es sei $(\Omega, \mathcal{F})$ ein Meßraum. $f, g: \Omega \to \mathbb{R}$ meßbare Funktionen und $\alpha, \beta \in \mathbb{R}$. Dann sind die Funktionen a) $\alpha \cdot f + \beta \cdot g$ b) $f \cdot g$ c) f/g, falls $g(\omega) \neq 0 \; \forall \omega \in \Omega$ ebenfalls meßbar.

Unabhängigkeit

Unabhängigkeit von Ereignissen

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Ereignisse $A_i \in \mathcal{F}$, $i \in I \neq \emptyset$ heißen paarweise stochastisch unabhängig, wenn $\forall J \subseteq I$ nicht-leer gilt:

$$\mathbb{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbb{P}(A_j)$$

Satz

i) Seien $A, B \in \mathcal{F}$ stoch. unabh. und $\mathbb{P}(B) > 0 \Rightarrow \mathbb{P}(A \mid B) = \mathbb{P}(A)$

ii) Seien $A, B \in \mathcal{F}$ stoch. unabh. $\Rightarrow A, \bar{B}$ sind stoch. unabhängig.

iii) Seien $A, B, C \in \mathcal{F}$. A, B sto. unab., B, C sto. unab. und A, C sto. unab. $\not\Rightarrow A, B, C$ sto. unabhängig.

Stochastisch unabhängige Mengensysteme

Eine Familie von Mengensysteme $\mathcal{F}_i \subseteq \mathcal{F}$, $i \in I \neq \emptyset$, heißt stochastisch unabhängig, wenn dies für jede Familie von Ereignissen $A_j \in \mathcal{F}_j$, $j \in J$, für alle $i \in I$ gilt.

Stochastisch unabhängige Zufallsvariablen

Eine Familie von ZV's $X_i: \Omega \to \Omega_i$ auf einem W-raum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt stochastisch unabhängig, wenn dies für die Familie der von den ZV's X_i erzeugten σ-Algebren $\sigma(X_i) = \sigma(X_i^{-1}(\mathcal{F}_i)) \quad \forall i \in I$ gilt.

Satz

Seien $X_1, \dots, X_n$ reelle ZV's. $X_1, \dots, X_n$ sind stochastisch unabhängig, genau dann wenn $\forall c \in \mathbb{R}^n$: $P(X_1 \leq c_1, \dots, X_n \leq c_n) = \prod_{i=1}^n P(X_i \leq c_i) = \prod_{i=1}^n F_{X_i}(c_i)$

Zufallsvariablen (Emperie)

Notation & Terminologie

Bezeichnung in der Emperie	Bezeichnung in der Theorie
Merkmal	Zufallsvariable $X, Y, \ldots$
Anzahl Untersuchungseinheiten	n
Merkmalsausprägung von Merkmal X der i-ten UE.	x_i, $i \in \{1, \ldots, n\}$
Rohdaten / Urliste	$x_1, \ldots, x_n$
(evtl. geordnete) verschiedene Werte aus der Urliste	$a_1, \ldots, a_k$, $k \leq n$, $a_i \in \{x_1, \ldots, x_n\}$ (Eindeutige Elemente der Urliste)
relative Häufigkeit f_j	Wahrsch.fkt. bzw. Dichte $f_X(a_j)$
kum. rel. Häufigkeit $F(x)$	Verteilungsfunktion $F_X(x)$

Absolute Häufigkeit

Die absolute Häufigkeit von a_j ist die Anzahl der x_i aus der Urliste mit $x_i = a_j$

$h(a_j) = h_j$

Relative Häufigkeit

Die relative Häufigkeit von a_j ist der Anteil der x_i an der Urliste für die gilt $x_i = a_j$. $f(a_j) = f_j = \frac{h_j}{n}$

Absolute / relative Häufigkeitsverteilung

$h_1, \ldots, h_k$ heisst absolute Häufigkeitsverteilung

$f_1, \ldots, f_k$ heisst relative Häufigkeitsverteilung

Kumulative relative Häufigkeit / empirische Verteilungsfunktion

(Sinnvoll bei ordinal oder metrisch)

$F(x) = (\text{Anteil UE mit } x_i \leq x) = \sum_{a_i \leq x} f(a_i)$ (ECDF)

$\longrightarrow$

- monoton wachsende Treppenfkt. mit Sprüngen bei $a_1, \ldots, a_k$.
- Sprunghöhe $f_1, \ldots, f_k$
- rechtsseitig stetig
- $F(x) = 0$ für $x < a_1$, $F(x) = 1$ für $x \geq a_k$

Zusammenhangsmaße für diskrete ZV

Kontingenztafel der absoluten Häufigkeiten

Seien X, Y diskrete Merkmale mit Ausprägungen $a_1, \ldots, a_k$ für X und $b_1, \ldots, b_m$ für Y. Eine $(k \times m)$-Kontingenztafel der absoluten Häufigkeiten besitzt die Form:

$$\begin{array}{c|ccc|c} & b_1 & \cdots & b_m & \\ \hline a_1 & h_{11} & \cdots & h_{1m} & h_{1\cdot} \\ a_2 & h_{21} & \cdots & h_{2m} & h_{2\cdot} \\ \cdot & \cdot & & \cdot & \cdot \\ a_k & h_{k1} & \cdots & h_{km} & h_{k\cdot} \\ \hline & h_{\cdot 1} & \cdots & h_{\cdot m} & n \end{array}$$

$h_{ij} = h(a_i, b_j)$ = Absolute Häufigkeit der Kombination (a_i, b_j)

$h_{i\cdot}$ = Randhäufigkeit von a_i in X ; $h_{\cdot j}$ = Randhäufigkeit von b_j in Y

Kontingenztafel der relativen Häufigkeiten

Seien X, Y diskrete Merkmale mit Ausprägungen $a_1, \ldots, a_k$ für X und $b_1, \ldots, b_m$ für Y. Eine $(k \times m)$-Kontingenztafel der relativen Häufigkeiten besitzt die Form:

$$\begin{array}{c|ccc|c} & b_1 & \cdots & b_m & \\ \hline a_1 & f_{11} & \cdots & f_{1m} & f_{1\cdot} \\ \cdot & \cdot & & \cdot & \cdot \\ a_k & f_{k1} & \cdots & f_{km} & f_{k\cdot} \\ \hline & f_{\cdot 1} & \cdots & f_{\cdot m} & 1 \end{array}$$

$f_{ij} = \frac{h_{ij}}{n}$ = Relative Häufigkeit der Kombination (a_i, b_j)

$f_{i\cdot} = \frac{h_{i\cdot}}{n}$ = relative Randhäufigkeit von a_i in X;

$f_{\cdot j} = \frac{h_{\cdot j}}{n}$ = relative Randhäufigkeit von b_j in Y

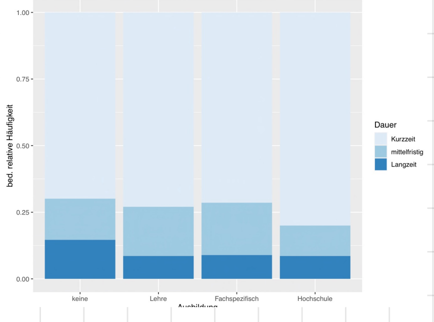
Bedingte Häufigkeitsverteilung

Die bedingte Häufigkeitsverteilung von Y unter der Bedingung $X = a_i$ $(Y | X = a_i)$ ist definiert als $f_Y(b_1 | a_i) = \frac{h_{i1}}{h_{i\cdot}}, \ldots, f_Y(b_m | a_i) = \frac{h_{im}}{h_{i\cdot}}$

Satz

Wegen $\frac{h_{ij}}{h_{i\cdot}} = \frac{h_{ij}/n}{h_{i\cdot}/n} = \frac{f_{ij}}{f_{i\cdot}}$ gilt, $f_Y(b_j | a_i) = \frac{f_{ij}}{f_{i\cdot}}$ $\forall i,j$

Beispiel



Bedingte Odds & Odds ratio

Für festes $X = a_i$ bezeichnen wir $\gamma(1,2 | X = a_i) = \frac{h_{i1}}{h_{i2}}$ als bedingte Odds.

Als relative Chancen (Odds ratio) bezeichnen wir $\gamma(1,2 | X = 1, X = 2) := \frac{\gamma(1,2 | X = 1)}{\gamma(1,2 | X = 2)} = \frac{h_{11} \cdot h_{22}}{h_{21} \cdot h_{12}}$

Erwartete absolute/relative Häufigkeit

Unter empirischer Unabhängigkeit wird erwartet, dass $f_Y(b_j|a_i) = f_Y(b_j) \quad \forall i,j$ gilt.

Daher gilt für die erwartete absolute Häufigkeit $\tilde{h}_{ij}$: $\tilde{h}_{ij} = \frac{h_{i\cdot} \cdot h_{\cdot j}}{n}$ (nicht immer $\in \mathbb{N}$)

Und für die erwartete relative Häufigkeit $\tilde{f}_{ij}$: $\tilde{f}_{ij} = f_{i\cdot} \cdot f_{\cdot j}$

Odds ratio (Kreuzproduktverhältnis) ist symmetrisch bezüglich der Wahl von X, Y.

$\gamma(Y=1, Y=2 \mid X=1, X=2) = \gamma(X=1, X=2 \mid Y=1, Y=2)$! Symmetrisches Maß und kann als Risikofaktor interpretiert werden:

- $\gamma = 1$: Odds in beiden Populationen gleich.
- $\gamma > 1$: Odds in $X=1$ höher als in $X=2$
- $\gamma < 1$: Odds in $X=1$ niedriger als in $X=2$.

χ^2-Koeffizient

Zusammenhangsmaß zum quantifizieren von "Abstand" zwischen beobachteten und erwarteten gemeinsamen Häufigkeiten.

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}} = n \cdot \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(f_{ij} - \tilde{f}_{ij})^2}{\tilde{f}_{ij}}$$

$\chi^2 \in [0, n \cdot \min(k,m) - 1]$

$\chi^2 = 0 \iff X, Y$ empirisch unabhängig

χ^2 gross $\iff$ starker Zusammenhang

! χ^2 hängt von n, k und m ab

↳ schwer zu interpretieren.

Satz

Für eine Kontingenztafel der Form

a	b
c	d

gilt

$$\chi^2 = \frac{n \cdot (a \cdot d - c \cdot b)^2}{(a+b)(a+c)(c+d)(b+d)}$$

(Korregierter) Kontingenzkoeffizient

! Misst nur Stärke des Zusammenhangs, Nicht die Richtung, wie bei γ

Normierung von χ^2: Kontingenzkoeffizient $K := \sqrt{\frac{\chi^2}{n + \chi^2}}$, $K \in \left[0, \sqrt{\frac{\min\{k,m\}-1}{\min\{k,m\}}}\right]$

Korrigierter Kontingenzkoeffizient $K^* := \frac{K}{\sqrt{\frac{\min\{k,m\}-1}{\min\{k,m\}}}}$, $K^* \in [0,1]$

Verteilungsfunktion

Verteilungsfunktion

Ist $\mathbb{P}: \mathcal{B} \to [0,1]$ ein W-maß auf $\mathbb{R}$, so heißt

$F_{\mathbb{P}}: \mathbb{R} \to [0,1]$, $x \mapsto F_{\mathbb{P}}(x) := \mathbb{P}[(-\infty, x)]$

die Verteilungsfunktion von $\mathbb{P}$.

Alternative Definition

Jede monoton wachsende, rechtsstetige Funktion $F: \mathbb{R} \to [0,1]$ mit $\lim_{x \to -\infty} F(x) = 0$ und $\lim_{x \to +\infty} F(x) = 1$ heißt Verteilungsfunktion.

Beispiel Binomialverteilung

$\mathbb{P}(\{k\}) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \cdot I_{\{0,\dots,n\}}(k)$

Dann: $F_{\mathbb{P}}(x) = \mathbb{P}[(-\infty, x)] = \mathbb{P}(\dot{\bigcup}_{k \leq x} \{k\}) = \sum_{k=0}^{\lfloor x \rfloor} \mathbb{P}(\{k\})$

Korrespondenssatz

Für jede Verteilungsfkt. F ist $\lambda_F((a,b)) = F(b) - F(a)$ ein W-maß mit $F_{\lambda_F} = F$. Umgekehrt gibt es für jedes reelles W-maß $\mathbb{P}$ eine Verteilungsfkt. F.

Lebesgue-Stieltjes-Maß von F

Eigenschaften von Verteilungsfunktionen

Eine Verteilungsfunktion $F_{\mathbb{P}}$ ist:

i) monoton wachsend

ii) rechtsstetig

iii) und es gilt $\lim_{x \to -\infty} F(x) = 0$ und $\lim_{x \to +\infty} F(x) = 1$

Transformation der Verteilungsfunktion

Für umkehrbare, streng monoton steigende Transformationen g gilt: $F_{g(X)}(y) = F_X(g^{-1}(y))$

Median

Gegeben sei eine eindimensionale ZV mit X mit Verteilung $\mathbb{P}_X$. $m \in \mathbb{R}$ heißt Median von $\mathbb{P}_X$ bzw. X, wenn

$\mathbb{P}_X[(-\infty, m]] \geq 0.5$ und $\mathbb{P}_X[[m, +\infty)] \geq 0.5$

Quantile

Gegeben sei eine eindimensionale ZV mit X mit Verteilung $\mathbb{P}_X$. $q \in \mathbb{R}$ heißt p-Quantil von $\mathbb{P}_X$ bzw. X, wenn

$\mathbb{P}_X[(-\infty, q]] \geq p$ und $\mathbb{P}_X[[q, +\infty)] \geq 1-p$

Lebesgue Integral

Einfache Funktionen

Sei $f: \Omega \to \mathbb{R}$ eine Funktion mit $|f(\Omega)| < \infty$ (also mit nur endlich vielen Funktionswerten). Dann heißt f einfache Funktion und lässt sich schreiben als

$f = \sum_{i=1}^{n} y_i \cdot \mathbb{1}_{A_i}$, wobei y_i ein Funktionswert ist und A_i die Menge für die $f(A_i) = y_i$ gilt.

f^+, f^-

$f^+ := \max\{0, f\}$, $f^- := -\min\{0, f\}$, $f = f^+ + f^-$, $|f| = f^+ + f^-$

Lebesgue-Integral für einfache Funktionen

Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und f eine <u>nicht-negative</u>, einfache und $\mathcal{F}$-$\mathcal{B}$-meßbare Funktion.

Dann definieren wir das Lebesgue-Integral von f als $\int_\Omega f \, d\mu = \int_\Omega \sum_{i=1}^{n} y_i \cdot \mathbb{1}_{A_i} \, d\mu := \sum_{i=1}^{n} y_i \cdot \mu(A_i)$ (geht auch für $n = \infty$).

Sei g eine einfache, $\mathcal{F}$-$\mathcal{B}$-meßbare Funktion und $\int_\Omega g^+ \, d\mu < \infty$ oder $\int_\Omega g^- \, d\mu < \infty$, dann heißt g μ-quasi-integrierbar und wir definieren $\int_\Omega g \, d\mu := \int_\Omega g^+ \, d\mu - \int_\Omega g^- \, d\mu$

Wenn $\int_\Omega g^+ \, d\mu < \infty$ <u>und</u> $\int_\Omega g^- \, d\mu < \infty$, so heißt g μ-integrierbar. ! Auf $\overline{\mathbb{R}}^n$ wählt man i.d.R. $\mu = \lambda$

Lebesgue-Integral

Sei $f: \Omega \to \overline{\mathbb{R}}$ eine $\mathcal{F}$-$\mathcal{B}$-meßbare Funktion. Wir definieren das Ober-Integral als

$\overline{\int_\Omega} f \, d\mu := \inf\{ \int_\Omega g \, d\mu \mid g \text{ ist eine } \mu\text{-integrierbare, einfache Funktion mit } g \geq f \ \mu\text{-fast-überall} \}$

Wir nennen f μ-integrierbar, wenn $\overline{\int_\Omega} f \, d\mu$ existiert und definieren $\int_\Omega f \, d\mu := \overline{\int_\Omega} f \, d\mu$

Idee: Annäherung an f mittels einfachen, messbaren, μ-int.baren Funktionen.

Satz

Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f: \Omega \to [0, +\infty]$ meßbar, dann ist f μ-integrierbar. Ausserdem gilt dann

- $\int_\Omega f \, d\mu = 0 \Leftrightarrow f = 0$ μ-fast-überall

Satz

Sei $X: \Omega \to \mathbb{R}^n$ eine n-dim. reelle ZV und $f: \mathbb{R}^n \to \mathbb{R}$ eine meßbare Fkt., s.d. $f \circ X: \Omega \to \mathbb{R}$ int.bar ist. Dann gilt

$\int f \, d\mathbb{P}_X = \int f \circ X \, d\mathbb{P}$

Monotonie und Linearität

Seien $f_1, f_2: \Omega \to \overline{\mathbb{R}}$ μ-int.bare Fkt. und $\alpha, \beta \in \mathbb{R}$, dann gilt

i) $f_1 \geq f_2$ μ-fast-überall $\Rightarrow \int_\Omega f_1 \, d\mu \geq \int_\Omega f_2 \, d\mu$

ii) $\alpha f_1 + \beta f_2$ ist μ-int.bar und $\int_\Omega \alpha f_1 + \beta f_2 \, d\mu = \alpha \cdot \int_\Omega f_1 \, d\mu + \beta \cdot \int_\Omega f_2 \, d\mu$

Weitere Eigenschaften

Sei $f: \Omega \to \overline{\mathbb{R}}$ eine μ-int.bare Fkt. mit $\Omega = A_1 \dot{\cup} A_2$ und $A_1, A_2 \in \mathcal{F}$. Dann gilt

i) $\int_\Omega f \, d\mu = \int_{A_1} f \, d\mu + \int_{A_2} f \, d\mu$

ii) $f|_{A_1}$ und $f \cdot I_{A_1}$ sind μ int.bar mit $\int_{A_1} f|_{A_1} \, d\mu = \int_\Omega f \cdot I_{A_1} \, d\mu$

Lebesgue-Integral bzgl. Dirac-Maß

Sei $\omega \in \Omega$ und δ_ω das Dirac-Maß. Sei f eine meßbare Fkt. Dann ist f δ_ω-int.bar, g.d.w. $f(\omega) < \infty$ und es gilt $\int_\Omega f \, d\delta_\omega = f(\omega)$

Lebesgue-Integral bzgl. Zählmaß

Für den Maßraum $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_Z)$ ist eine Fkt. $f: \mathbb{N} \to \mathbb{R}$ μ-int.bar, g.d.w. die Reihe $\sum_{n=1}^{\infty} f(n)$ absolut konvergiert. Es gilt $\int_{\mathbb{N}} f \, d\mu_Z = \sum_{n=1}^{\infty} f(n)$

→ Integration mit Zählmaß ist gleich Summe der y-Werte

Riemann- und Lebesgue-Integral

Sei $f: [a,b] \to \mathbb{R}$ eine beschränkte, meßbare und Riemann-integrierbare Fkt., dann ist f λ-integrierbar und $\int_a^b f(x) \, dx = \int_{[a,b]} f \, d\lambda$

Korollar

Sei $f: \Omega \to \overline{\mathbb{R}}$ eine μ-int.bare Fkt. und $A_1 \in \mathcal{F}$ s.d. $\mu(A_1) = 0$. Dann gilt

i) $|\int_\Omega f \, d\mu| \leq \int_\Omega |f| \, d\mu$ (Dreiecksungleichung)

ii) $\int_{A_1} f \, d\mu = 0$ (Lebesgue-integral von Nullmengen ist Null)

Konvergenzsätze für Lebesgue-Integrale

Monotone Konvergenz | Satz von Beppo Levi

Sei $(f_n)_{n\in\mathbb{N}}$ eine Folge von meßbaren, numerischen Funktionen, s.d. $\forall n\in\mathbb{N}: f_n \le f_{n+1}$. Dann gilt $\lim\limits_{n\to\infty}\int_\Omega f_n \, d\mu = \int_\Omega \lim\limits_{n\to\infty} f_n \, d\mu$.

Dominierte Konvergenz / Satz von Lebesgue

Sei $g:\Omega\to[0,+\infty]$ μ-integrierbar und $f:\Omega\to\overline{\mathbb{R}}$, $\{f_k\}_{k\in\mathbb{N}}:\Omega\to\mathbb{R}$ μ-meßbar.
Angenommen $|f_k|\le g$ und $f_k \xrightarrow{k\to\infty} f$ μ-fast-überall. Dann folgt daraus:

i) $\lim\limits_{k\to+\infty}\int_\Omega |f_k - f| \, d\mu = 0$ ii) $\lim\limits_{k\to+\infty}\int_\Omega f_k \, d\mu = \int_\Omega f \, d\mu$

Gegenbeispiel

Warum ist es nötig, dass die Folge f_n von g majorisiert/dominiert wird ($|f_n|\le g$) um den limes ins Integral zu ziehen?
Wir betrachten die Folge $f_n(x) = \frac{1}{2n}\cdot \mathbb{1}_{[-n,n]}(x)$ f_n konvergiert punktweise gegen $f\equiv 0$.
$\int_\Omega f_n \, d\lambda = 1$ $\forall n\in\mathbb{N}$. Also existiert eine Verteilungsfkt. $F_n(x)$ zu f_n. Würden wir nun den Limes einfach ins Integral ziehen dürfen (ohne Einschränkungen), bekämen wir $\int f \, d\lambda = \int \lim\limits_{n\to\infty} f_n \, d\lambda = \lim\limits_{n\to\infty} \underbrace{\int f_n \, d\lambda}_{=1} = \lim\limits_{n\to\infty} 1 = 1$. Also hätte $f\equiv 0$ eine Verteilungsfkt. $F(x)$. Offensichtlich ist das aber Schwachsinn, denn $\int f \, d\lambda = \int 0 \, d\lambda = 0$.
Es ist hier auch ersichtlich, dass es kein λ-int bares $g:\Omega\to[0,\infty]$ gibt, s.d. $\forall n\in\mathbb{N}: 0\le|f_n|\le g$ gilt.
Beweis wäre, dass $\sum_{n=1}^{\infty} 2\cdot\frac{1}{2n} = \infty$ ergibt.

Dichte

Lemma

$\nu:\mathcal{F}\to\overline{\mathbb{R}}$, $\nu(A) := \int_A f \, d\mu$ ist ein Maß und $\int g \, d\nu = \int g\cdot f \, d\mu$

Absolute Stetigkeit von Maßes

Sind ν und μ zwei Maße auf $(\Omega,\mathcal{F})$, so heißt ν absolut stetig bezüglich μ, wenn $\forall A\in\mathcal{F}$ gilt:
$\mu(A)=0 \Rightarrow \nu(A)=0$. Notation: $\nu \ll \mu$ (μ dominiert ν).

Satz von Radon-Nikodým | Dichte

Ist $(\Omega,\mathcal{F},\mu)$ ein Maßraum mit σ-endlichen Maßen μ und ν. Wenn $\nu\ll\mu$, dann existiert eine meßbare Fkt. $f:\Omega\to\overline{\mathbb{R}}$, s.d. $\forall A\in\mathcal{F}$: $\nu(A) = \int_A f \, d\mu$. Notation: $\nu = f\odot\mu$
Die Funktion f ist eindeutig bis auf eine μ-Nullmenge. D.h. sei g eine Fkt. mit obigen Eigenschaften, dann gilt $g=f$ μ-fast-sicher.
f wird auch geschrieben als $\frac{d\mu}{d\nu}$ und wird Radon-Nikodým-Ableitung genannt.
Wir nennen f auch Wahrscheinlichkeitsdichte von einer ZV bzgl. μ.

Satz

Sei $(\Omega,\mathcal{F},\mathbb{P})$ ein W-raum. Sei μ ein Maß mit $\mathbb{P}\ll\mu$. Dann existiert die Dichte f mit $\mathbb{P} = f\odot\mu$ (Dichte der Verteilung $\mathbb{P}$). Sei $\mathcal{F}=\mathcal{B}$, dann gilt $F(x) = \mathbb{P}[(-\infty,x]] = \int_{(-\infty,x]} f \, d\mu$

> Diskrete Verteilungen werden von $\mu=\mu_Z$ dominiert.
> (Absolut-) stetige Verteilungen werden von $\mu=\lambda$ dominiert.

Beispiel für gemischte Verteilung

Von der verkauften Ware verursacht der Anteil $\alpha>0$ keine Kosten und der Rest verursacht exponentialverteilte Kosten mit Erwartungswert β. Wie sind die Kosten verteilt?
$\mathbb{P}(\{0\}) = \alpha$, $\mathbb{P}$ ist auf $\mathbb{R}^+$ absolut stetig. $\mathbb{P}$ wird nicht von λ dominiert, da $\mathbb{P}(\{0\}) = \alpha > 0 = \lambda(\{0\})$.
$\mathbb{P}$ wird nicht von μ_Z dominiert, da μ_Z nicht σ-endlich auf $(\mathbb{R},\mathcal{B})$ ist.
Wir wählen $\mu = \lambda+\delta_0$. μ ist σ-endlich und dominiert $\mathbb{P}$: Sei $B\in\mathcal{B}$ eine μ-Nullmenge, dann $\lambda(B)+\delta_0(B)=0$ und somit $\lambda(B)=\delta_0(B)=0$. Offensichtlich ist $\{0\}\not\subseteq B$. Somit folgt $\mathbb{P}(B) = \int_B f \, d\lambda \underset{\uparrow}{=} 0$. Also gilt $\mu(B)=0 \Rightarrow \mathbb{P}(B)=0$.
weil B λ-Nullmenge ist
somit gilt $\mathbb{P}\ll\mu$ und daher existiert $f^*\in M^+$, s.d. $\mathbb{P}(A) = \int_A f^* \, d\mu$ $\forall A\in\mathcal{B}$. Dieses f^* konstruieren wir durch

$$f^*(x) = \begin{cases}(1-\alpha)f(x) & , x\neq 0\\ \alpha & , x=0\end{cases}$$

Diskrete Verteilungen

Diskreter Wahrscheinlichkeitsraum

Sei $(\Omega,\mathcal{F},\mathbb{P})$ ein Wahrscheinlichkeitsraum. Gibt es eine abzählbare Menge $T\in\mathcal{F}$ mit $\mathbb{P}(T)=1$, so heißt
- $(\Omega,\mathcal{F},\mathbb{P})$ diskreter Wahrscheinlichkeitsraum
- $\mathbb{P}$ diskretes W-maß
- T abzählbarer Träger von $\mathbb{P}$ ❗ Ω muss nicht abzählbar sein.

Zähldichte

$f:\Omega\to[0,1]$, $f(\omega) = \mathbb{P}(\{\omega\})\cdot I_T(\omega)$ heißt Zähldichte.
$\mathbb{P}(A) = \sum_{\omega\in A\cap T} f(\omega)$ $\forall A\in\mathcal{F}$

Stetige Verteilungen

Stetig verteilte ZV

Eine reelle ZV X heißt stetig verteilt, wenn es eine nicht-negative Borel-messbare Fkt. $f:\mathbb{R}\to\mathbb{R}$ mit der Eigenschaft $\int_\mathbb{R} f(t)\,dt = 1$ gibt, so dass $\mathbb{P}_X(B) = \mathbb{P}(X\in B) = \int_B f(t)\,dt$, $B\in\mathcal{B}$.
f heißt stetige Dichte (bzgl. des Lebesgue-Maßes) von X bzw. $\mathbb{P}_X$.
❗ f muss nicht stetig sein, aber es muss eine stetige Fkt. g geben, mit $f = g\cdot I_{(a,b)}$

Träger

Sei $\mathbb{P}$ eine (beliebige) Verteilungen mit Dichte $f(x)$, dann ist $T=\{x\in\Omega \mid f(x)>0\}$ der Träger der Verteilung.

Dichtetransformationssatz

Sei $X:\Omega\to\mathbb{R}$ eine ZV mit stetiger Verteilungsfkt. F_X und Dichte $f_X(x) = \frac{\partial F_X(x)}{\partial x}$ bzgl. des λ-Maßes. Sei $g:\mathbb{R}\to\mathbb{R}$ bijektiv und stetig diff.bar mit $g'(x)\neq 0$. Dann hat $g\circ X$ die Dichte $f_{g\circ X}(y) = f_X(g^{-1}(y))\cdot|(g^{-1})'(y)|$

Wichtige parametrische Verteilungen

Diskrete parametrische Verteilungen

Bernoulli-Verteilung

Wahrscheinlichkeit für Erfolg/Miserfolg.

X ist Bernoulli verteilt $X \sim \mathcal{B}(\pi)$, wenn

$P_\pi(X=x) = \pi^x \cdot (1-\pi)^{1-x}$ für $x \in \{0,1\}$

Dann ist $F_X(x) = 1-\pi$, für $x \in [0,1)$

Diskrete Gleichverteilung

Laplace Wahrscheinlichkeiten

X ist diskret gleichverteilt $X \sim \mathcal{U}_D(a,b)$, wenn

$T_X = \{x_1, \dots, x_n\} \subseteq [a,b]$ und

$P_{a,b}(X=x) = \frac{1}{b-a} \quad \forall x \in T_X$.

Dann ist $F_X(x) = \sum_{i=1}^{k} x_{(i)}$ für $x \in [x_{(k)}, x_{(k+1)})$

Geometrische-Verteilung (Variante A)

Wahrscheinlichkeit das man genau x Versuche für den ersten Erfolg benötigt. X ist geometrisch verteilt mit $X \sim \mathcal{G}(p)$.

$P_\pi(X=x) = (1-\pi)^{x-1} \cdot \pi$, $x \in \mathbb{N}^+$, $\pi \in (0,1)$

Dann ist $F_X(x) = 1 - (1-\pi)^{\lfloor x \rfloor}$, für $x \in [1, +\infty)$

Geometrische-Verteilung (Variante B)

Wahrscheinlichkeit k Fehlversuche vor dem ersten Erfolg zu haben. X ist geometrisch verteilt mit $X \sim \mathcal{G}(p)$.

$P_\pi(X=k) = (1-\pi)^k \cdot \pi$, $k \in \mathbb{N}_0$, $\pi \in (0,1)$

Dann ist $F_X(x) = 1 - (1-p)^{\lfloor x \rfloor + 1}$, für $x \in [0, +\infty)$

Binomial-Verteilung

Wahrscheinlichkeit für x Erfolge bei n Bernoulli-Versuchen.

X ist Binomial-verteilt $X \sim \mathcal{B}(n,\pi)$, wenn

$P_{n,\pi}(X=x) = \binom{n}{x} \cdot \pi^x \cdot (1-\pi)^{n-x}$, für $x \in \mathbb{N}_0$

Dann ist $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$, für $x \geq 0$

Hypergeometrische Verteilung

Wahrscheinlichkeit für x Erfolge beim "ziehen" von n Elementen aus einer Stichprobe der Größe N, wobei es insgesamt M günstige Elemente gibt.

X ist hypergeometrisch verteilt $X \sim H(N,M,n)$, wenn

$P_{N,M,n}(X=x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$, für $x \in \{\max\{0, n-(N-M)\}, \dots, \min\{n,M\}\}$

Dann ist $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} P_{N,M,n}(X=k) \cdot \mathbb{1}_{[0,\infty)}$

Poisson-Verteilung

Wahrscheinlichkeit für x Erfolge innerhalb eines Zeitintervalls, wobei $\lambda > 0$ die Rate ist, mit der Erfolge in diesem Zeitintervall auftreten.

X ist Poisson-verteilt $X \sim \mathcal{P}(\lambda)$, wenn

$P(X=x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$, für $x \in \mathbb{N}_0$

Dann gilt $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} P(X=k) = \frac{\Gamma(\lfloor x+1 \rfloor, \lambda)}{\lfloor x \rfloor !}$

Satz

Verteilung	$E(X)$	$Var(X)$
$\mathcal{B}(n,\pi)$	$n \cdot \pi$	$n \cdot \pi(1-\pi)$
$\mathcal{G}(\pi)$ (A)	$\frac{1}{\pi}$	$\frac{1-\pi}{\pi^2}$
$\mathcal{G}(\pi)$ (B)	$\frac{1-\pi}{\pi}$	$\frac{1-\pi}{\pi^2}$
$\mathcal{U}_D(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
$\mathcal{H}(N,M,n)$	$n \cdot \frac{M}{N}$	$n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{n-1}$
$\mathcal{P}(\lambda)$	λ	λ

Stetige parametrische Verteilungen

Stetige Gleichverteilung

X ist stetig gleichverteilt $X \sim \mathcal{U}(a,b)$, wenn

$f_{a,b}(x) = \frac{1}{b-a} \cdot \mathbb{1}_{[a,b]}(x)$.

Dann gilt $F_X(x) = \frac{x-a}{b-a} \cdot \mathbb{1}_{[a,b)}(x) + \mathbb{1}_{[b,+\infty)}(x)$

Exponentialverteilung

Wartezeit zwischen Erfolgen einer Poisson-verteilten ZV.

X ist exponential verteilt $X \sim Exp(\lambda)$, wenn

$f_\lambda(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{(0,\infty)}$

Dann gilt $F_X(x) = (1 - e^{-\lambda x}) \cdot \mathbb{1}_{(0,\infty)}$

Gammaverteilung

Verallgemeinerung der Exp-Verteilung. Die Gamma-Verteilung ist die Maximum-Entropie-Wahrscheinlichkeitsdichte für eine ZV X mit $E(X) = \frac{\alpha}{\beta} > 0$ und $E(\ln(x)) = \frac{d}{dx} \ln(\Gamma(x))\Big|_{x=\alpha} - \ln(\beta)$

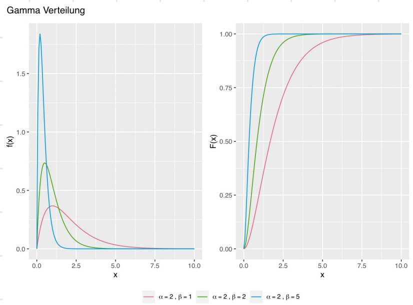
X ist Gamma-verteilt $X \sim \mathcal{G}(\alpha,\beta) \sim \Gamma(\alpha,\beta)$, wenn

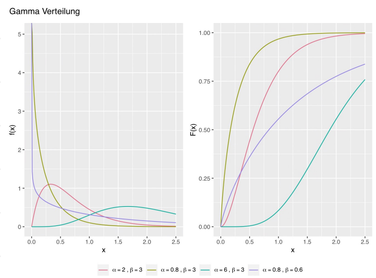
$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \cdot e^{-\beta x} \cdot \mathbb{1}_{(0,\infty)}(x)$. Γ ist die Gammafunktion:

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx$. Insbesondere $\Gamma(x+1) = x! \quad \forall x \in \mathbb{N}_0$.

Dann gilt $F_X(x) = \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\beta x} t^{\alpha-1} \cdot e^{-t} \cdot \mathbb{1}_{[0,\infty)}(t) \, dt$

<u>Note</u>: $Exp(\lambda) = \Gamma(\alpha=1, \beta=\lambda)$. $\chi^2(d) = \Gamma(\alpha = \frac{d}{2}, \beta = \frac{1}{2})$





Betaverteilung

Familie von Verteilungen auf $[0,1]$, parametrisiert durch zwei Parameter α, β.

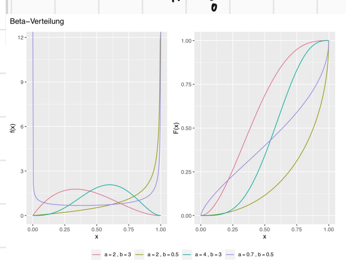
X ist Beta-verteilt $X \sim Be(\alpha,\beta)$, wenn

$f_{\alpha,\beta}(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1} \cdot (1-x)^{\beta-1}$, für $x \in (0,1)$.

B ist die Betafunktion und dient zum normieren.

$B(\alpha,\beta) := \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \, dx$

Dann gilt $F_X(x) = \frac{1}{B(\alpha,\beta)} \cdot \int_0^x t^{\alpha-1} \cdot (1-t)^{b-1} \, dt$



Normalverteilung

X ist normalverteilt $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $E(X) = \mu$ und $Var(X) = \sigma^2$, wenn

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right), \text{ für } x \in \mathbb{R}.$$

Eine geschlossene Form für $F_X(x)$ existiert nicht $\rightarrow$ numerische Integration.

Für $\mu = 0$ und $\sigma^2 = 1$ nennen wir $X \sim \mathcal{N}(0,1)$ standartnormalverteilt.

Für $F_X(x)$ wird dann meist als $\Phi(x)$ notiert, wobei sich diese Werte in Tabellen finden lassen.

! Linearität: $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y = aX + b \Rightarrow Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

↳ Es gilt ausserdem $X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$.

Additivität: Seien $X_1, \dots, X_n$ sto. unabhängig mit $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ $\forall i \in \{1, \dots, n\}$

i) $\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$ (umkehrung gilt auch)

ii) Sind alle μ_i gleich μ und alle $\sigma_i^2 = \sigma$, dann gilt $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

! t-Verteilung und Invers-Gamma-Verteilung hat nicht mehr reingepasst. Siehe dafür Folien/Skript.

Cauchy-Verteilung

Verteilung von Verhältnis (ratio) zweier unabhängiger normalverteilter ZV mit Erwartungswert 0. X ist Cauchy-verteilt $X \sim \mathcal{C}$, wenn

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \text{ für } x \in \mathbb{R}.$$

Dann gilt $F_X(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}$

Note: Sehr viel Wahrscheinlichkeitsmasse in den "Tails". Momente sind undefiniert.

Satz

Verteilung	$E(X)$	$Var(X)$
$\mathcal{U}(a,b)$	$\frac{1}{2} \cdot (a+b)$	$\frac{1}{12} \cdot (b-a)^2$
$Exp(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\Gamma(\alpha, \beta)$	α/β	α/β^2
$Be(\alpha, \beta)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha \cdot \beta}{(\alpha+\beta)^2 \cdot (\alpha+\beta+1)}$
$\mathcal{N}(\mu, \sigma^2)$	μ	σ^2
$\mathcal{C}$	nicht existent	nicht existent
$Wei(\lambda, k)$	$\lambda^{-1} \cdot \Gamma(1 + 1/k)$	$\lambda^{-2} \cdot (\Gamma(1 + 2/k) - \Gamma(1 + 1/k))$

Survivorfunktion und Hazardrate

Sei T eine stetige, nicht-negative ZV (Wartezeit auf ein Ereignis). Dann ist

$S(t) := \mathbb{P}(T \geq t) = 1 - F_T(t)$ die Survivorfunktion (Wahrscheinlichkeit bis Zeitpunkt t zu überleben) und

$h(t) := \lim_{\Delta t \to 0} \frac{\mathbb{P}(T \geq t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$ die Hazardrate (Likelihood / "Wahrscheinlichkeit", dass das Ereignis im nächsten Moment eintritt).

Weibull-Verteilung

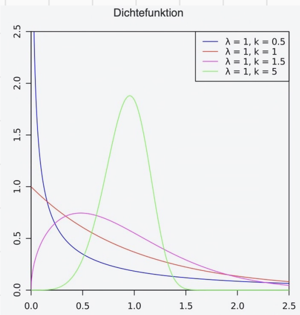
Beschreibung von Lebensdauer und Ausfallhäufigkeit von (mechanischen) Objekten.

X ist Weibull-verteilt $X \sim Wei(\lambda, k)$ mit Parametern λ und k, wenn

$$f_{\lambda,k}(x) = \lambda k \cdot (\lambda x)^{k-1} \cdot e^{-(\lambda x)^k} \cdot \mathbb{1}_{\mathbb{R}^+}(x).$$

Dann gilt:

$$F_{\lambda,k}(x) = \left(1 - e^{-(\lambda x)^k}\right) \cdot \mathbb{1}_{\mathbb{R}^+}(x)$$



Für $T \sim Wei(\lambda, k)$ ist für $t > 0$ $h(t) = \lambda k \cdot (\lambda t)^{k-1}$

! Der Formparameter $k > 0$ beschreibt das zeitliche Verhalten der Ausfälle.

$k < 1$: Frühausfall (z.B. Kinderkrankheiten)

$k = 1$: Exp-Verteilung (zufällige Ausfälle)

$k > 1$: Ermüdungs- und Verschleißausfälle nach langer Nutzung

Dichtetransformationssatz

Transformationssatz (für monotone Transformationen)

Sei X eine ZV mit Dichte $f_X(x)$ und der stetigen Verteilung $F_X(x)$.

Sei $Y = g(X)$ eine streng monotone und differenzierbare Fkt. auf X mit $g'(x) > 0$. Dann hat Y eine Dichtefunktion $f_Y(y)$:

$$f_Y(y) = f(g^{-1}(y)) \cdot \left| \frac{1}{g'(g^{-1}(y))} \right| = f(g^{-1}(y)) \cdot \left| (g^{-1})'(y) \right|$$

und es gilt $F_Y(y) = F_X(g^{-1}(y))$

Transformationssatz (für nicht monotone Transformationen)

Sei X eine ZV mit Dichte $f_X(x)$ und der stetigen Verteilung $F_X(x)$.

Sei $Y = g(X)$. Angenommen es existiert eine Partition $A_0, A_1, \dots, A_K$ von X, s.d. $P(X \in A_0) = 0$ und $f_X(x)$ stetig auf allen A_i ist. Angenommen es existieren $g_1(x), \dots, g_K(x)$ Fkt. entsprechend definiert auf $A_1, \dots, A_K$, für die gilt:

i) $g(x) = g_i(x)$ $\forall x \in A_i$. ii) $g_i(x)$ ist streng monoton auf A_i iii) $g_i(A_i) = T_Y$ $\forall i$

iv) g_i^{-1} ist stetig differenzierbar auf Y ($g'(x) \neq 0$ $\forall x \in A_i$)

Dann hat Y eine Dichtefunktion $f_Y(y)$ mit:

$$f_Y(y) = \sum_{i=1}^{k} f(g_i^{-1}(y)) \cdot \left| \frac{1}{g_i'(g_i^{-1}(y))} \right| = \sum_{i=1}^{k} f(g_i^{-1}(y)) \cdot \left| (g_i^{-1})'(y) \right|, \quad \forall y \in T_Y$$

$$= 0, \quad y \notin T_Y$$

und es gilt $F_Y(y) = \sum_{i=1}^{k} F_X(g_i^{-1}(y))$

Beispiel

$X \sim \mathcal{N}(0,1)$ und $Y = X^2$.

$f_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} \cdot x^2\right)$. $A_0 = \{0\}$, $A_1 = (-\infty, 0)$, $A_2 = (0, +\infty)$

$g_1: A_1 \to \mathbb{R}^+, x \mapsto x^2$ $\quad$ $g_2: A_2 \to \mathbb{R}^+, x \mapsto x^2$.

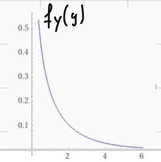
$g_1^{-1}(y) = -\sqrt{y}$ $\quad$ $g_2^{-1}(y) = \sqrt{y}$

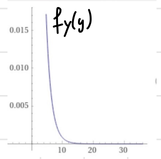
$$f_Y(y) = f_X(-\sqrt{y}) \cdot \left| \frac{1}{-2\sqrt{y}} \right| + f_X(\sqrt{y}) \cdot \left| \frac{1}{2\sqrt{y}} \right| = \frac{1}{\sqrt{2\pi \cdot y}} \cdot \exp\left(-\frac{1}{2} \cdot y\right), \text{ für } y > 0$$

$$= 0, \text{ sonst}$$

$\int_0^\infty f_Y(y)\, dy = 1$ ✓

$Y \sim \mathcal{G}(1/2, 1/2)$





Inversions Methode

Die Inversionsmethode ist ein Simulationsverfahren, um aus gleichverteilten Zufallszahlen andere Wahrscheinlichkeitsverteilungen zu erzeugen.

1. Erzeuge $u_1, \dots, u_n$ mit $u_i \overset{i.i.d.}{\sim} \mathcal{U}(0,1)$.
2. Berechne $x_i = F_X^{-1}(u_i)$, $i = 1, \dots, n$

Die x_i sind dann Zufallszahlen aus der gewünschten Verteilung $F_X(x)$.

Momente (Theorie)

Momente

Sei X eine int.bare ZV, $n \in \mathbb{N}$, s.d. X^n quasi-int.bar ist. Dann heißt

$m_n(X) = E[X^n]$ n-tes Moment

$m_{(n)}(X) = E[|X|^n]$ n-tes absolutes Moment

$m_n^\circ(X) = E[(X - E(X))^n]$ n-tes zentriertes Moment

Erwartungswert

Ist $X: \Omega \to \mathbb{R}$ eine quasi-int.bare ZV, so heißt

$m_1(X) = \mathbb{E}[X] := \int X \, d\mathbb{P} = \int x \, d\mathbb{P}_X(x)$

Erwartungswert von X (1-tes Moment).

Satz

Sei $\mathbb{P}_X(A) = \int_A f_X \, d\mu$, dann ist $\mathbb{E}[X] = \int X \, d\mathbb{P} = \int x \, d\mathbb{P}_X = \int x \cdot f_X \, d\mu = \begin{cases} \sum_{x \in T_X} x \cdot f_X(x) & \text{, wenn } \mu = \mu_Z \\ \int_{-\infty}^{\infty} x f_X(x) \, dx & \text{, wenn } \mu = \lambda \end{cases}$

$\mathbb{E}[g(X)] = \int g \circ X \, d\mathbb{P} = \int g \, d\mathbb{P}_X = \begin{cases} \sum_{x=-\infty}^{\infty} g(x) f_X(x) & \text{, } \mathbb{P} \text{ diskret} \\ \int_{\mathbb{R}} g(x) f_X(x) \, dx & \text{, } \mathbb{P} \text{ stetig} \end{cases}$

Eigenschaften von $\mathbb{E}[X]$

Analog zum Lebegue-Integral:

(i) $E\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i E[X_i]$ (ii) $X \leq Y \Rightarrow E[X] \leq E[Y]$

(iii) Sind $X_1, \dots, X_n$ unabhängige int.bare ZV, so gilt:

$\mathbb{E}\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n \mathbb{E}[X_i]$

Varianz

Ist X eine ZV mit $\mathbb{E}[|X|] < \infty$, so heißt

$m_2^{(\circ)}(X) = Var[X] := \mathbb{E}[(X - \mathbb{E}(X))^2]$

Varianz von X (2-tes zentriertes Moment)

Verschiebungssatz

$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

! Insbesondere gilt also $E[X]^2 \leq E[X^2]$

Eigenschaften von Var[X]

$Var[aX + b] = a^2 \cdot Var[X]$

Spezialfall Beispiel

$Var[X] = 0 \Rightarrow X \equiv E[X]$ bzw. $X \sim \delta_{E[X]}$

Kovarianz / Korrelation

Seien $X, Y: \Omega \to \mathbb{R}$ zwei ZV mit $Var[X] > 0$ und $Var[Y] > 0$, dann ist die Kovarianz

$Cov(X,Y) := E[(X - E(X)) \cdot (Y - E(Y))]$ und

die Korrelation $\rho(X,Y) := \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

! Unabhängig $\underset{\not\Leftarrow}{\Rightarrow}$ Unkorreliert

$|\rho(X,Y)| = 1 \Leftrightarrow Y = a + bX$, $a, b \in \mathbb{R}$

Eigenschaften von Kovarianz

(i) $Cov(X,X) = Var(X)$

(ii) $|Cov(X,Y)| \leq Var(X) \cdot Var(Y)$

(iii) $Cov(aX + b, cY + d) = a \cdot c \cdot Cov(X,Y)$

iv) $Cov(X,Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$

v) $Var(X+Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y)$

vi) $\hookrightarrow X, Y$ sto. unabhängig $\Rightarrow Var(X+Y) = Var(X) + Var(Y)$

Schiefe

Ist X^3 quasi-int.bar, dann heißt

$\gamma(X) := \frac{m_3^\circ(X)}{m_2^\circ(X)^{3/2}} = \frac{E((X - E(X))^3)}{\sqrt{Var(x)^3}}$

Schiefe von X.

Kurtosis

Ist X^4 quasi-int.bar, dann heißt

$K(X) := \frac{m_4^\circ(X)}{m_2^\circ(X)^2} = \frac{E((X - E(X))^4)}{Var(x)^2} = \frac{E((X - E(X))^4)}{E((X - E(X))^2)^2}$

Kurtosis von X.

Symmetrische Verteilung

Sei $\mathbb{P}$ eine Verteilung auf $(\mathbb{R}, \mathcal{B})$.

$\mathbb{P}$ heißt symmetrisch um $a \in \mathbb{R}$, wenn

$\mathbb{P}((-\infty, a - x]) = \mathbb{P}([a + x, +\infty))$ $\forall x \in \mathbb{R}$.

Eigenschaften symmetrischer Verteilungen

Sei $\mathbb{P}$ eine symmetrische Verteilung auf $(\mathbb{R}, \mathcal{B})$. $X \sim \mathbb{P}_X = \mathbb{P}$ und g eine int.bare Fkt. Dann gilt:

(i) g int.bar $\Rightarrow \int g \, d\mathbb{P}_X = \int g(2a - x) \, d\mathbb{P}_X(x)$

(ii) $m_1(X) = a$ (iii) $m_{2n+1}^{(\circ)}(X) = 0$ $\forall n \in \mathbb{N}$ (iv) a ist Median von X

Modus

Gegeben sei eine eindim. ZV X mit Verteilung $\mathbb{P}_X$ und Dichte $f(x)$.

m heißt Modus von X bzw. $\mathbb{P}$, wenn $f(m)$ ein lok. Max. von $f(x)$ ist.

Median

Gegeben sei eine eindimensionale ZV mit X mit Verteilung $\mathbb{P}_X$. $m \in \mathbb{R}$ heißt Median von $\mathbb{P}_X$ bzw. X, wenn

$\mathbb{P}_X[(-\infty, m]] \geq 0.5$ und $\mathbb{P}_X[[m, +\infty)] \geq 0.5$

Quantile

Gegeben sei eine eindimensionale ZV mit X mit Verteilung $\mathbb{P}_X$. $q \in \mathbb{R}$ heißt p-Quantil von $\mathbb{P}_X$ bzw. X, wenn $\mathbb{P}_X[(-\infty, q]] \geq p$ und $\mathbb{P}_X[[q, +\infty)] \geq 1 - p$

Markov- und Chebyshev-Ungleichungen

Sei $X: \Omega \to \mathbb{R}$ eine reelle ZV. Dann gilt

$\forall \varepsilon > 0$: $\mathbb{P}(|X| > \varepsilon) \leq \frac{1}{\varepsilon^n} \cdot E[|X|^n]$

Markov-Ungleichung ($n = 1$): $\mathbb{P}(|X| > \varepsilon) \leq \frac{1}{\varepsilon} \cdot E[|X|]$

Chebyshev-Ungleichung ($n = 2$): $\mathbb{P}(|X - E[X]| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \cdot Var(X)$

p-Norm / L^p-Raum

Sei $f: \Omega \to \mathbb{R}$, $\mathcal{F}$-$\mathcal{B}$-meßbar auf $(\Omega, \mathcal{F}, \mu)$. Dann heißt

$\|f\|_p = \left(\int |f|^p \, d\mu\right)^{1/p} \in [0, +\infty)$ die p-Norm von f.

Für $p \geq 1$ ist der L^p-Raum

$L^p := L^p(\Omega, \mathcal{F}, \mu) := \{f: \Omega \to \mathbb{R} \mid f \ \mathcal{F}\text{-}\mathcal{B}\text{-meßbar mit } \|f\|_p < \infty\}$

Wahrscheinlichkeitserzeugende Funktion

Für eine ZV $X: \Omega \to \mathbb{N}_0$ heißt $m_X: [0,1] \to [0,1]$ mit

$m_X(t) := \sum_{k=0}^{\infty} t^k \cdot P(X = k)$ die wahrscheinlichkeitserzeugende Funktion.

! Es gilt $m_X(t) = E[t^X]$

Multiplikationsformel

Sind X, Y unabhängige $\mathbb{N}_0$-wertige ZV, so gilt

$m_{X+Y}(t) = m_X(t) \cdot m_Y(t)$, $|t| \leq 1$

Jensen-Ungleichung

Sei X eine int.bare ZV und $g: \mathbb{R} \to \mathbb{R}$ konvex.

Dann gilt $E[g(X)] \geq g(E[X])$

Ist $f: \mathbb{R} \to \mathbb{R}$ konkav, so gilt $f(E[X]) \geq E[f(X)]$

Hölder-Ungleichungen

1) Für $1 < p, q < \infty$ mit $\frac{1}{p} + \frac{1}{q} = 1$ und $f, g: \Omega \to \mathbb{R}$ meßbar gilt $\|f \cdot g\|_1 \leq \|f\|_p \cdot \|g\|_q$

2) Ist $\mu(\Omega) < \infty$ und $p > q \geq 1$, so ist $L^q \subseteq L^p$ und es gibt $c \geq 0$, s.d. $\|f\|_p \leq c \cdot \|f\|_q$ $\forall f \in L^q$

3) Sind $f, g: \Omega \to \mathbb{R}$ meßbar und $p \geq 1$, so gilt $\|f + g\|_p \leq \|f\|_p + \|g\|_p$

Beispiel

Sei $X_i \sim B(1,p)$ und $Y = \sum_{i=1}^n X_i \sim B(n,p)$. Dann ist $m_{X_i}(t) = \sum_{k=0}^{\infty} t^k \cdot P[X = k] = 1 - p + t \cdot p$.

$m_Y(t) = \prod_{i=1}^n m_{X_i}(t) = (1 - p + tp)^n$. $E[Y] = \lim_{t \to 1} n \cdot (1 - p + tp)^{n-1} \cdot p = n \cdot p$

Momenterzeugung

Ist $X: \Omega \to \mathbb{N}_0$, dann gilt:

$E[X] = \lim_{t \to 1} \frac{d}{dt} m_X(t)$

Mehrdimensionale Momente (benötigt Thema Zufallsvektoren)

Erwartungswert und Kovarianzmatrix

Sei $\underline{X}: \Omega \to \mathbb{R}^n$, $\underline{X} = (X_1, \dots, X_n)$ eine n-dim. ZV, dann heißt

$\mathbb{E}[\underline{X}] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$ der (n-dimensionale) Erwartungswert von $\underline{X}$.

$\mathbb{V}[\underline{X}] := \mathbb{E}[(\underline{X} - \mathbb{E}[\underline{X}]) \cdot (\underline{X} - \mathbb{E}[\underline{X}])^T)$ die Kovarianzmatrix von $\underline{X}$.

Anmerkung: $\mathbb{V}[\underline{X}]_{i,i} = Var[X_i]$. $\mathbb{V}[\underline{X}]_{i,j} = Cov(X_i, X_j) = \mathbb{V}[\underline{X}]_{j,i}$

! $\mathbb{V}[\underline{X}]$ ist positiv semidefiniert

Korrelationsmatrix

Sei Σ die Kovarianzmatrix einer mehrdimensionalen ZV $\underline{X} = (X_1, \dots, X_n)$ und $diag(\Sigma)$ die Matrix der Diagonalelemente von Σ. Dann heißt

$R := Corr(\underline{X}) := diag(\Sigma)^{-\frac{1}{2}} \cdot \Sigma \cdot diag(\Sigma)^{-\frac{1}{2}}$

die Korrelationsmatrix von $\underline{X}$.

Standardisierung

Sei $\underline{X} = (X_1, \dots, X_n)$ ein n-dim. ZV mit $\mathbb{E}[\underline{X}] = \underline{\mu}$ und positiv definiter Kov.matrix Σ.

Dann existiert eine invertierbare Matrix $B \in \mathbb{R}^{n \times n}$, so dass gilt:

Für $\underline{Y} := B^{-1} \cdot (\underline{X} - \underline{\mu})$ ist $\mathbb{E}[\underline{Y}] = \underline{0}$ und $\mathbb{V}[\underline{Y}] = I_n$

Satz

Sei $\underline{X}: \Omega \to \mathbb{R}^n$ eine n-dim. ZV. $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$.

(i) $\mathbb{E}[A\underline{X} + b] = A\mathbb{E}[\underline{X}] + b$

(ii) $\mathbb{V}[A\underline{X} + b] = A \cdot \mathbb{V}[\underline{X}] A^T$

Erzeugende Funktion

Momenterzeugende Funktion

Ist X eine reelle ZV und $D := \{s \in \mathbb{R} \mid \mathbb{E}[\exp(sX)] < \infty\}$, so heißt die Fkt.

$M: D \to \mathbb{R}$, $s \mapsto \mathbb{E}[\exp(sX)] = \int \exp(sx) \, d\mathbb{P}_X = \int \exp(sx) \cdot f_X(x) \, d\mu$

momenterzeugende Funktion.

Satz

Sei X eine ZV mit momenterzeugender Funktion $M: D \to \mathbb{R}$.

Ist $(-a, a) \subseteq D$ für ein beliebiges $a > 0$, so gilt $\forall n \in \mathbb{N}$:

i) $\mathbb{E}[X^n] < \infty$ ii) $M(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} \mathbb{E}[X^n]$ iii) $\frac{\partial^n M}{\partial^n s}(0) = \mathbb{E}[X^n]$

Charakteristische Funktion

Sei X eine ZV. Die Funktion $\varphi_X: \mathbb{R} \to \mathbb{C}$ mit

$\varphi_X(t) := \mathbb{E}[\exp(itX)] = \int \exp(itX) \, d\mathbb{P} = \int f_X(x) \cdot \exp(itx) \, d\mu$

heißt charakteristische Fkt. von X.

Beispiel

i) $X \sim B(n, p)$: $\varphi_X(t) = (1 + p + p \cdot e^{it})^n$

ii) $X \sim Po(\lambda)$: $\varphi_X(t) = \exp(-\lambda \cdot (1 - e^{it}))$

iii) $X \sim \mathcal{N}(0,1)$: $\varphi_X(t) = \exp(-t^2/2)$

Eigenschaften von charakteristischen Funktionen

Seien φ_X und φ_Y charakteristische Funktionen und $a, b \in \mathbb{R}$. Dann gilt:

i) Eindeutigkeit: $\varphi_X = \varphi_Y \Leftrightarrow \mathbb{P}_X = \mathbb{P}_Y$

ii) Linearität: $\varphi_{aX+b}(t) = \exp(itb) \cdot \varphi_X(a \cdot t)$

iii) φ_X ist eine gleichmäßig stetige Fkt.

iv) $\mathbb{E}[|X|^k] < \infty \Rightarrow \mathbb{E}[X^k] = \frac{\varphi_X^{(k)}(0)}{i^k}$ — k-te Ableitung.

Dichte von X mittels der charakteristischen Fkt.

Sei φ_X eine charakteristische Funktion. Dann gilt: Die Verteilung von X hat die Dichte $f_X(x) = \frac{1}{2\pi} \cdot \int \exp(-itx) \cdot \varphi_X(t) \, dt$

Momente (Emperie) / Kennwerte & Verteilungeigenschaften

Lagemaße

Lagemaßzahlen

> Wo liegt die Masse, Mitte und Mehrzahl der Daten?

> Welche Merkmalsausprägung ist typisch für die Verteilung?

Modus (Emperie)

Häufigster Wert in einer Stichprobe.

- oft nicht eindeutig
- nur bei gruppierten Daten oder Merkmalen mit wenig Ausprägungen sinnvoll.

\+ stabil bei allen injektiven Transformationen

\+ geeignet für jedes Skalenniveau.

Der Mittelwert (Emperie)

Arithmetisches Mittel: $\bar{x}_{arith} := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{n} \cdot \sum_{i=1}^{k} h_i \cdot a_i = \sum_{i=1}^{k} f_i \cdot a_i$

Anzahl Klassen

- instabil gegenüber Ausreißern
- mindestens intervallskalierte Daten

\+ bekanntestes Lagemaß und wichtiger theoretischer Nutzen.

Gewichtetes Mittel: $\bar{x}_W := \frac{1}{\sum_{i=1}^{n} w_i} \cdot \sum_{i=1}^{n} w_i \cdot x_i$, mit $w_i \geq 0 \ \forall i$

Geometrisches Mittel: $\bar{x}_G := \sqrt[n]{\prod_{i=1}^{n} x_i} = \exp\left(\frac{1}{n} \sum_{i=1}^{n} \log(x_i)\right)$.

- Mindestens Verhältnisskala ($x_i > 0$)

\+ Anwendbar auf multiplikative Faktoren

Harmonisches Mittel: $\bar{x}_H := \frac{1}{\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{x_i}} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{-1}\right)^{-1}$

- Mindestens Verhältnisskala ($x_i > 0$)

\+ Anwendbar auf Quotienten/Verhältnisse.

Getrimmtes Mittel: Sei $\alpha \in (0,1)$, $x_{(1)}, \dots, x_{(n)}$ die geordnete Stichprobe und $r = \max\{\tilde{r} \in \mathbb{Z} \mid \tilde{r} \leq n \cdot \alpha\}$.

Wir definieren das α-getrimmte Mittel als $\bar{x}_\alpha := \frac{1}{n-2r} \cdot \sum_{i=r+1}^{n-r} x_{(i)}$.

+/- α-Anteil der extremsten Werte wird abgeschnitten

Alternativ: $\bar{x}_\alpha := \frac{1}{n} \cdot \left(\sum_{i=r+1}^{n-r} x_{(i)} + r \cdot \tilde{x}_\alpha + r \cdot \tilde{x}_{1-\alpha}\right)$

α-Quantil

Quantil/Perzentil (Emperie)

$x_{(1)}, \dots, x_{(k)}$ sind die geordneten Werte der Stichprobe.

Das p-Quantil ($p \in [0,1]$) ist der Wert $\tilde{x}_p$ für den gilt:

Anteil p der Daten sind $\leq \tilde{x}_p$ & Anteil $1-p$ der Daten sind $\geq \tilde{x}_p$.

$$\tilde{x}_p := \begin{cases} x_{(k)} & \text{, falls } n \cdot p \notin \mathbb{N}_0 \text{ und } k > n \cdot p \\ \frac{1}{2} \cdot (x_{(k)} + x_{(k+1)}) & \text{, falls } k = n \cdot p \in \mathbb{N}_0 \end{cases}$$

(Alternativ beliebigen Wert $\tilde{x}_p \in [x_{(k)}, x_{(k+1)}]$ für $k = n \cdot p \in \mathbb{N}_0$)

Median (Emperie)

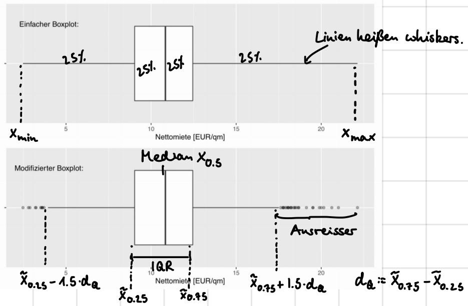
$\tilde{x}_{med} := \tilde{x}_{0.5} :=$ 50%-Perzentil

\+ anschaulich

\+ stabil gegenüber monotonen Transformationen

\+ geeignet für mindestens ordinale Daten

\+ stabil gegenüber Ausreißern



Streuungsmaße

Streuungsmaßzahlen

> Über welchen Bereich erstrecken sich die Ausprägungen?

> Wie groß ist die Schwankung der beobachteten Werte?

> Wie eng beieinander liegen die beobachteten Werte?

Stichprobenvarianz & Stichprobenstandartabweichung

$s_x^2 := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$. $\quad s_x := \sqrt{s_x^2}$.

- Empfindlich gegen Ausreißer

Standartabweichung einer ZV

Die Standartabweichung $\sigma(X)$ einer ZV X ist definiert als

$\sigma(X) = \sqrt{Var(X)}$ bzw. $s_x = \sqrt{s_x^2}$ für Stichproben

Satz

Sei $Y = a + bX$, für $a, b \in \mathbb{R}$. Dann gilt:

$\sigma_Y^2 = b^2 \cdot \sigma_X^2$; $\sigma_Y = |b| \cdot \sigma_X$; $s_Y^2 = b^2 \cdot s_X^2$; $s_Y = |b| \cdot s_X$

Verschiebungssatz

$\forall c \in \mathbb{R}: \sum_{i=1}^{n} (x_i - c)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n \cdot (\bar{x} - c)^2$

$\overset{c=0}{\Rightarrow} Var(X) = E(X^2) - E(X)^2$

Variationskoeffizient

$v = \frac{\sigma_X}{\bar{x}}$ mit $\bar{x} > 0$.

Skalierungsunabhängige Maßzahl für relative Schwankungen um $\bar{x}$.

Streuungszerlegung I

Seien die Daten in r Schichten aufgeteilt:

$$x_1, \dots, x_{n_1}, x_{n_1+1}, \dots, x_{n_1+n_2}, \dots, x_n \quad \text{mit } n = \sum_{j=1}^{r} n_j$$

Schichtmittelwerte: $\bar{x}_1 = \frac{1}{n_1} \cdot \sum_{i=1}^{n_1} x_i \; ; \; \bar{x}_2 = \sum_{i=n_1+1}^{n_1+n_2} x_i \; ; \dots$

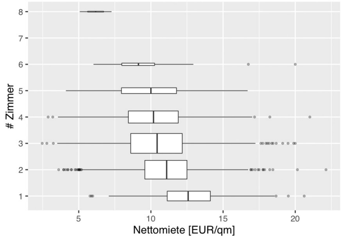
Schichtvarianz: $\tilde{s}^2_{X_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 \; ; \; \tilde{s}^2_{X_2} = \frac{1}{n_2} \cdot \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2)^2$

Dann gilt: $$\tilde{s}^2_X = \frac{1}{n} \cdot \sum_{j=1}^{r} n_j \cdot \tilde{s}^2_{X_j} + \frac{1}{n} \cdot \sum_{j=1}^{r} n_j \cdot (\bar{x}_j - \bar{x})^2$$

Gesamtstreuung = Streuung innerhalb der Schichten + Streuung zwischen den Schichten

Streuungszerlegung der Netto-Quadratmetermiete bezüglich Zimmerzahl:

Zimmer	n_j	$\bar{x}_j$	$\tilde{S}^2_{xj}$
8	2	6.2	1.2
6	16	10.0	13.2
5	99	9.9	7.3
4	442	10.2	7.0
3	1175	10.4	7.1
2	1049	10.9	6.1
1	282	12.6	6.2



Höhe der Boxplots proportional zur Gruppengröße

Gesamtvarianz: $\tilde{S}^2_x \approx 7.15$
Innerhalb: $\frac{1}{n} \sum_{j=1}^{r} n_j \tilde{S}^2_{xj} = 6.7$
Zwischen: $\frac{1}{n} \sum_{j=1}^{r} n_j (\bar{x}_j - \bar{x})^2 = 0.45$

$\Longrightarrow$ nur $\frac{0.45}{7.15} = 6.25\%$ der Gesamtvarianz der Quadratmetermiete entfallen auf Unterschiede zwischen Wohnungen mit unterschiedlicher Zimmerzahl.

Mittlere absolute Abweichung (L_1-Loss)

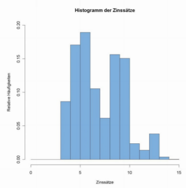
$MAD_X := \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - \bar{x}|$. Es gilt $MAD_X \leq \tilde{s}_X$

$MedAD_X := \text{Median}(|x_i - \tilde{x}_{med}|)$

+ einfacher zu interpretieren als s_X.
+ Weniger Ausreißer-empfindlich.
- weniger schöne theoretische Eigenschaften als s_X.

Verteilungseigenschaften

unimodal = eingipflig, **multimodal** = mehrgipflig



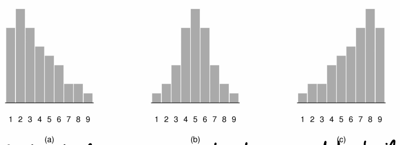
Das Histogramm der Zinssätze zeigt eine bimodale (trimodale...?) Verteilung.

Symmetrie & Schiefe

symmetrisch $\Leftrightarrow$ Annähernd spiegelsymm.

linkssteil (rechtsschief) $\Leftrightarrow$ Verteilung fällt nach links deutlich steiler und nach rechts langsamer ab.

rechtssteil (linksschief) $\Leftrightarrow$ Verteilung fällt nach rechts deutlich steiler und nach links langsamer ab.



(a) linkssteil: $\bar{x} > x_{med} > x_{mod}$

(b) symmetrisch: $\bar{x} \approx x_{mod} \approx x_{med}$

(c) rechtssteil: $\bar{x} < x_{med} < x_{mod}$

Quantilskoeffizient

$$g_p := \frac{(\tilde{x}_{1-p} - \tilde{x}_{med}) - (\tilde{x}_{med} - \tilde{x}_p)}{\tilde{x}_{1-p} - \tilde{x}_p}, \quad p \in (0, \tfrac{1}{2})$$

$g_{0.25}$ nennt man auch Quartilskoeffizient.

symmetrisch: $g_p = 0$
linkssteil: $g_p > 0$
rechtssteil: $g_p < 0$

Fisher's Momentkoeffizient für Schiefe (3. normiertes Moment)

$$g_{m_3} = E\left[\left(\frac{X - E(X)}{\sigma_X}\right)^3\right] = \frac{E[(X - E(X))^3]}{E[(X - E(X))^2]^{\frac{3}{2}}} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^3}{\sigma_X^3} \overset{\text{wenn } E(X) < \infty}{=} \frac{E(X^3) - 3E(X) \cdot \sigma_X^2 - E(X)^3}{\sigma_X^3}$$

symmetrisch: $g_{m_3} = 0$
linkssteil: $g_{m_3} > 0$
rechtssteil: $g_{m_3} < 0$

Corrected: skew $\tilde{g}_m = \frac{n \cdot \sum_{i=1}^{n} (x - \bar{x})^3}{(n-1)(n-2) \cdot \tilde{s}_X^3}$

Kurtosis (4. normiertes Moment)

$$k_X := \text{Kurt}(X) := E\left[\left(\frac{X - E(X)}{\sigma_X}\right)^4\right] = \frac{E[(X - E(X))^4]}{E[(X - E(X))^2]^2} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^4}{\sigma_X^4}$$

Exzess-Kurtosis: $k^*_X = k_X - 3$

Sample kurtosis: $\tilde{k} = \frac{n \cdot (n+1)}{(n-1)(n-2)(n-3)} \cdot \frac{\sum_{i=1}^{n} (x - \bar{x})^4}{\tilde{s}_X^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$

mesokurtisch: $k^*_X \approx 0$.
leptokurtisch: $k^*_X > 0$. Viele extreme Werte
platykurtisch: $k^*_X < 0$. Wenig extreme Werte

Konzentrationsmaße

Lorenzkurve

Das Merkmal darf nur positive Werte annehmen.

$x_{(1)}, \dots, x_{(n)}$ sei die geordnete Stichprobe.

Die Lorenzkurve verbindet Punktepaare bestehend aus den Teilsummen von $x_{(i)}$ (d.h. $\sum_{i=0}^{k} x_{(i)}$) und dem relativen Anteil an Individuen, die diese Teilsumme besitzen.

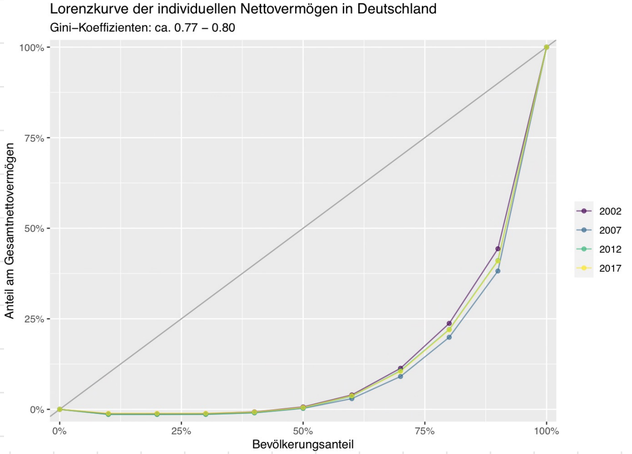
Berechnung

$u_{(0)} := 0$, $v_{(0)} := 0$, $j = 1, \dots, n$.

$u_{(j)} := \frac{j}{n}$ (Aufteilung der x-Achse)

$v_{(j)} := \sum_{i=1}^{j} x_{(i)} \cdot \frac{1}{\sum_{i=1}^{n} x_{(i)}}$ (y-Werte)

$v_{(j)}$ ist monoton steigend



Gini - Koeffizient / Lorenz'sches Konzentrationsmaß

Der Gini-Koeffizient ist eine Maßzahl, die das Ausmaß der Konzentration beschreibt. Er ist definiert als

$$G = 2 \cdot F \quad ; \quad G \in \left[0, \frac{n-1}{n}\right]$$

wobei F die Fläche zwischen $y=x$ und der Lorenzkurve ist.

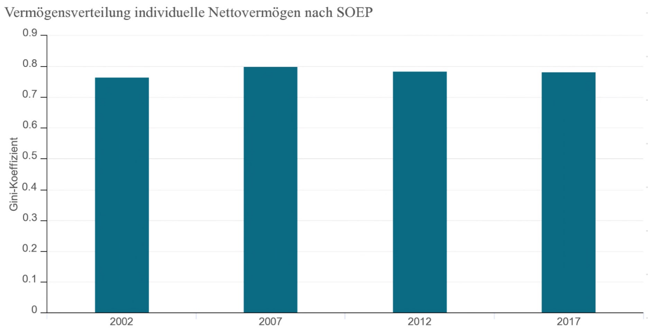
$$G = \frac{2 \cdot \sum_{i=1}^{n} i \cdot x_{(i)} - (n+1) \cdot \sum_{i=1}^{n} x_{(i)}}{n \cdot \sum_{i=1}^{n} x_{(i)}} \quad \text{oder } G = 1 - \frac{1}{n} \sum_{i=1}^{n} \left(v_{(i-1)} + v_{(i)}\right)$$

$$= \frac{n-1}{n} - \frac{2}{n} \cdot \sum_{i=1}^{n-1} v_{(i)}$$

Normierter Gini-Koeffizient : $G^{+} = \frac{n}{n-1} \cdot G \quad \in [0,1]$

$G^{+} = 0$ bedeutet keine Konzentration (Gleichverteilung)

$G^{+} = 1$ bedeutet volle Konzentration (Monopol)



Herfindahl - Index

Seien $x_1, \dots, x_n$ Daten mit $x_i \geq 0$. $p_i := \frac{x_i}{\sum_j x_j}$

Der Herfindahl-Index ist

$$H := \sum_{i=1}^{n} p_i^2 \quad \in \left[\frac{1}{n}, 1\right]$$

Statistische Grafiken

Grammatik von Grafiken

Grafik = Daten + geometrische Elemente + Ästhetische Zuordnung
+ Datentransformationen + Skalen + Koordinatensysteme
+ Facettierung + Theme + {Grafik}

Geometrische Elemente = Punkte | Linien | Rechtecke | Boxplots | Dichtefkt. | ...

Ästhetische Zuordnung = Position & Farbe & Größe & Form

Datentransformationen = id | Mittelwerte | Anteile | ...

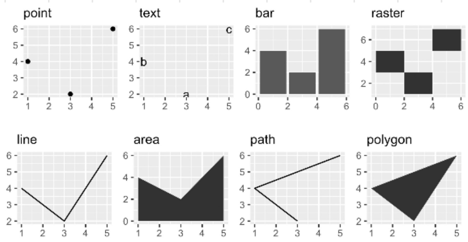
Skalen = Achsenabschnitte & Farbe & Legenden & Achsenbeschriftung & ...

Koordinatensysteme = kartesisch | logarithmisch | Polarkoord. | Kartenproj. | ...

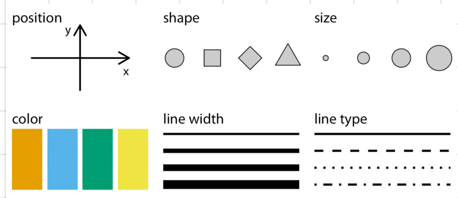
Facettierung = small multiples | lattice plot | plot | ...

Theme = Font & Gitterlinienien & Hintergrundfarben & Layout von Text & ...

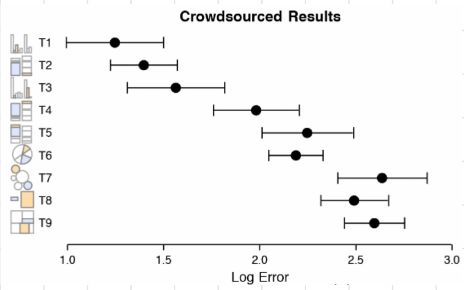
Beispiele Geometrien



Beispiele Ästhetiken



Wahrnehmung von Grafiken



Hierarchie der korrekten Interpret.:

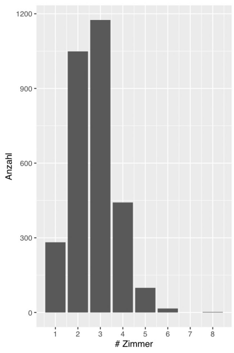
1. Position
2. Abstände / Längen
3. Steigung
4. Winkel
5. Flächen
6. Volumen
7. Farbe (Ton, Helligkeit, Sättigung)

Goldene Regel für Grafikgestaltung

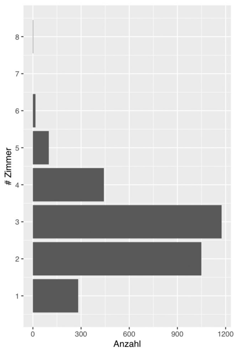
- Kommunikationsabsicht klarmachen
- Lesbarkeit maximieren

Visualisierung von Häufigkeiten & Verteilungen diskreter Merkmale

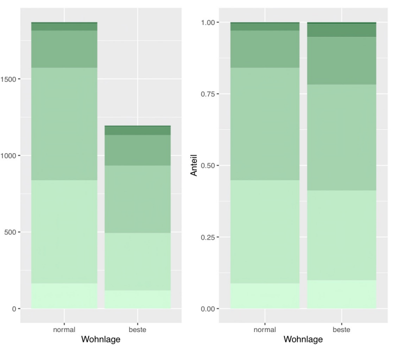
Säulendiagramm



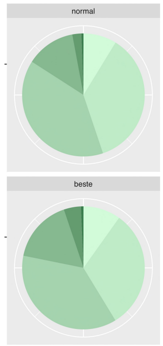
Balkendiagramm



Stapeldiagramm (absolut & relativ)



Kreisdiagramm



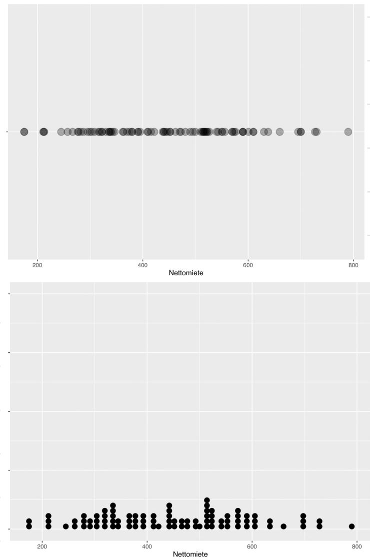
Für ordinale Merkmale, metrische Merkmale mit wenig Ausprägung und nominale Merkmale, wobei die Anordnung beliebig ist. ! Breite ist beliebig.

Anwendbar für die Gleichen Merkmale wie zuvor. Besonder geeignet für den Vergleich verschiedener Gruppen (bedingte Häufigkeiten).

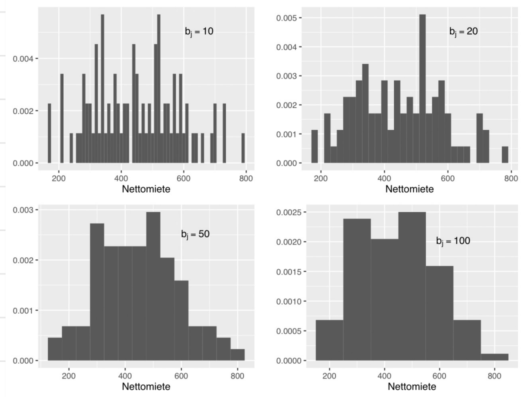
Grundsätzlich schwieriger zu interpretieren als Längendiagramme, aber enthält keine klare Ordnung.

Visualisierung von Häufigkeiten & Verteilungen metrischer Merkmale

Dotplots



Histogramme



Im Gegensatz zu Säulendiag. sind Histogramme für mindestens Intervallskalierte Daten gedacht. Relative Häufigkeiten werden durch Flächen dargestellt (Flächentreue). Vorgehen: Aufteilung in Klassen und Bestimmung der relativen Häufigkeiten $f_i = \frac{n_i}{n}$. Höhe y_i des Balkens i bestimmen mittels $b_i \cdot y_i = f_i$, wobei b_i die $\underbrace{\text{Breite der Klasse } i}_{\text{Daher nur bei metrischen Daten}}$ ist. ← Meist uniform

! Nachteil:
- Interpretation der Höhe bei unterschiedlichen Breiten nicht sinnvoll.
- Visueller Eindruck hängt von Klassenbreiten ab.
- Vorsicht bei Rändern.

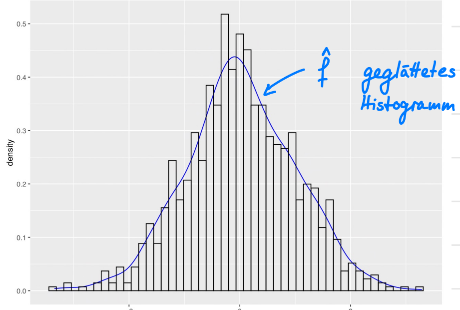
Kerndichteschätzung

Kerndichteschätzer

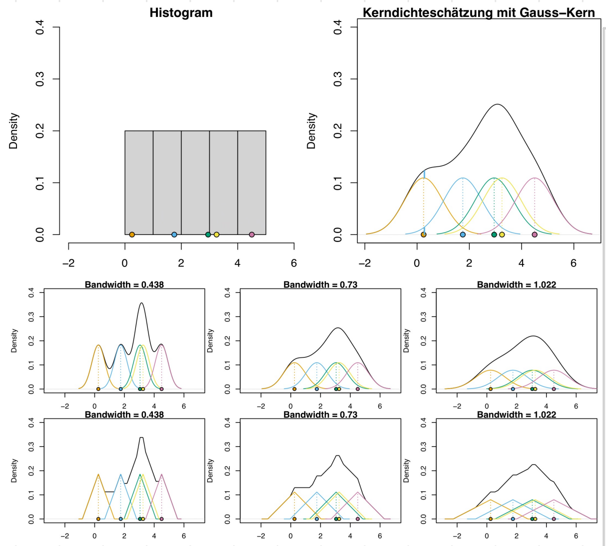
Sei K eine Kernfunktion, d.h. $\forall u: K(u) \geq 0$ und $\int_{-\infty}^{\infty} K(u)\, du = 1$.

Dann ist der Kerndichteschätzer (KDE: kernel density estimator) definiert als

$$\hat{f}(x) := \frac{1}{n \cdot h} \cdot \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$



$\hat{f}$ geglättetes Histogramm



Beziehung zu Fourier-Transformation

For the sample $(x_1, \dots, x_n)$ the characteristic function $\phi(t) = E[e^{itX}]$ can be estimated by $\hat{\varphi}(t) = \frac{1}{n} \cdot \sum_{j=1}^{n} e^{itx_j}$. Knowing ϕ we can get the corresponding density func. f trough the Fourier transform formula. ($\to f_X$ is Fourier transform of ϕ_X) But because $\hat{\varphi}$ is only an estimate and diverges for $|t| \to \infty$ we need to "dampen" $\hat{\varphi}$ by multiplying with $\Psi_h(t) = \Psi(ht)$, whereas $\Psi(0) = 1$ and $\Psi(ht) \to 0$ for $|t| \to \infty$. The bandwidth h controls the speed of convergence to 0.

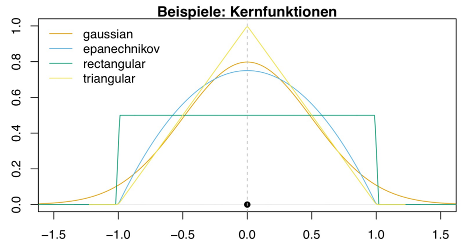
Most common is $\Psi(t) = \chi_{[-1,1]}$ or $\Psi(t) = e^{-\pi t^2}$.

$$\hat{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\varphi}(t) \cdot \Psi_h(t)\, e^{-itx}\, dt = \dots = \frac{1}{n \cdot h} \cdot \sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right)$$

where K is the Fourier transform of Ψ.

Beispiele für Kernfunktionen

- Gauß-Kern: $K(u) := \frac{1}{\sqrt{2\pi}} \cdot \exp(-\frac{1}{2} u^2)$
- Epanechnikov-Kern: $K(u) := \max\{0, \frac{3}{4} \cdot (1 - u^2)\}$
- Dreieck-Kern: $K(u) := \max\{0, 1 - |u|\}$



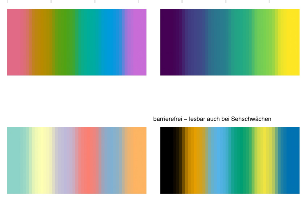
KDE > Histogramm bei größeren Datenmengen oder (quasi-)stetigen Merkmalen.

Vorteil:

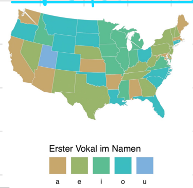
Kerndichteschätzungen berücksichtigen Entfernung der benachbarten Punkte mit abnehmender Gewichtung über Distanz.

Nachteil:

Abhängigkeit von Bandbreite h → Wird aus den Daten bestimmt.



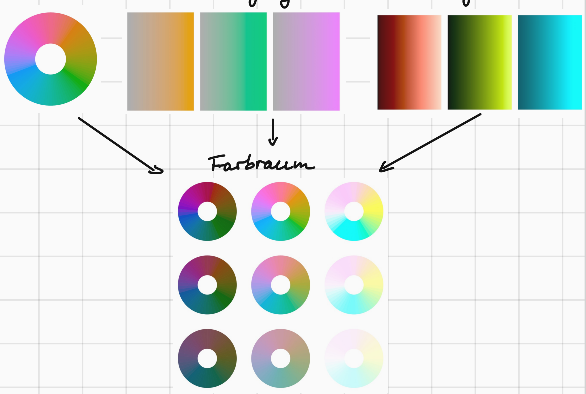
Beispiel qualitative Farbskala



Farbskalen

Farbraum

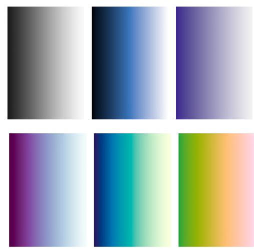
Der Farbraum ist definiert durch die Wahlmöglichkeiten für Farbton, Farbsättigung und Helligkeit.



Farbskalentypen

- Qualitativ: (eher) nur für nominales Skalenniveau.
- Sequentiell: mindestens ordinales Skalenniveau. ⟶
- Divergent: mindestens ordinales Skalenniveau mit "neutralem" mittleren Wert.

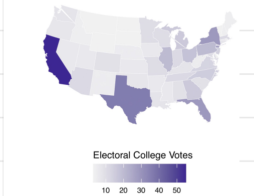
2 sequentielle Farbskalen mit je konstantem Farbton kombiniert.



Farbe konstant, Sättigung & Helligkeit variiert.

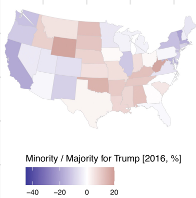
Farbe, Helligkeit und Sättigung variiert.

Beispiel sequentielle Farbskala

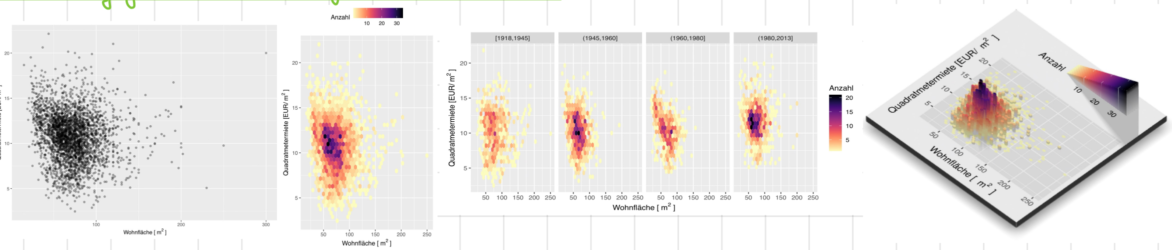


Sinnvoll, aber evtl. kleine Unterschiede verwischt

Beispiel divergente Farbskala



Visualisierung gemeinsamer Verteilungen metrischer Merkmale



Konvergenz von Zufallsvariablen und Verteilungen

Punktweise und gleichmäßige Konvergenz

Punktweise Konvergenz: $\forall x \in \mathbb{R}\ \forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n > N: \ |f_n(x) - f(x)| < \varepsilon \quad (\forall x \in \mathbb{R}: \lim_{n\to\infty} f_n(x) = f(x))$

Gleichmäßige Konvergenz: $\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall x \in \mathbb{R}\ \forall n > N: \ |f_n(x) - f(x)| < \varepsilon \quad (\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |f_n(x) - f(x)| = 0)$

Fast sichere Konvergenz

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X_n: \Omega \to \mathbb{R}$ eine ZV $\forall n \in \mathbb{N}$.
Dann konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ fast sicher gegen $X: \Omega \to \mathbb{R}$, falls $\mathbb{P}(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$. Wir schreiben $X_n \xrightarrow{f.s.} X$.

Idee: Wir wollen punktweise Konvergenz von $X_n(\omega) \to X(\omega)$, aber wollen "unmögliche" Ereignisse ignorieren. Wir sagen X_n konvergiert fast sicher gegen X, wenn die punktweise Konvergenz fast-sicher gilt.

Beispiel

Sei $X_1 \sim \mathcal{U}[0,1]$ und $X_n \sim X_1^n$. Dann gilt $X_n \xrightarrow{f.s.} 0$ ← Funktion die konstant 0 ist.

Beweis: Für alle $\omega \in \Omega$ mit $X(\omega) \in [0,1)$ gilt $\lim_{n\to\infty} X(\omega)^n = 0$. Für $\omega' \in \Omega$ mit $X(\omega') = 1$ gilt $\lim_{n\to\infty} X(\omega')^n = 1$. Also gilt $\mathbb{P}[\{\omega \in \Omega \mid \lim_{n\to\infty} X(\omega)^n = 0] = \mathbb{P}[X(\omega) \in [0,1)] = \mathbb{P}[X<1] = 1$.

Stochastische Konvergenz (Konvergenz in Wahrscheinlichkeit)

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X_n: \Omega \to \mathbb{R}$ eine ZV $\forall n \in \mathbb{N}$.
Dann konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ in Wahrscheinlichkeit gegen $X: \Omega \to \mathbb{R}$, falls $\forall \varepsilon > 0: \lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0$. Wir schreiben $X_n \xrightarrow{\mathbb{P}} X$ und sprechen von stochastischer Konvergenz.

Idee: Wir wollen gleichmäßige Konvergenz von $X_n(\omega) \to X(\omega)$, aber wollen "unmögliche" Ereignisse ignorieren. Wir sagen X_n konvergiert stochastisch gegen X, wenn die Wahrscheinlichkeit, dass $\sup_{\omega\in\Omega} |X_n(\omega) - X(\omega)| > \varepsilon > 0$ ist, gegen 0 konvergiert.

Satz

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von ZV $X_i: \Omega \to \mathbb{R}$ und $X: \Omega \to \mathbb{R}$. Dann gilt
$X_n \xrightarrow{f.s.} X \implies X_n \xrightarrow{\mathbb{P}} X$

Beispiel für $X_n \xrightarrow{\mathbb{P}} X$, aber $X_n \not\xrightarrow{f.s.} X$

(anderes Beispiel in Folie 412)

Sei $\{X_n\}_{n\in\mathbb{N}}$ ein Folge von $\overset{\text{unabhängigen}}{\text{ZV}}$ mit $\mathbb{P}[X_n = 1] = \frac{1}{n}$ und $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n}$. Behauptung: $X_n \xrightarrow{\mathbb{P}} 0$

Für $0 < \varepsilon \leq 1$ gilt $\lim_{n\to\infty} \mathbb{P}[|X_n - 0| \geq \varepsilon] = \lim_{n\to\infty} \mathbb{P}[|X_n| \geq \varepsilon] = \lim_{n\to\infty} \frac{1}{n} = 0$. Für $\varepsilon > 1$ ist es trivial. Also gilt $X_n \xrightarrow{\mathbb{P}} 0$.

Behauptung: $X_n \not\xrightarrow{f.s.} 0$. $\sum_{n=1}^{\infty} \mathbb{P}[X_n = 1] = \infty$ (harmonische Reihe). D.h. das Ereignis $X_n = 1$. Laut (Borel-Cantelli) Lemma gilt für Folgen von unabhängigen Ereignissen, deren Wahrscheinlichkeit sich zu ∞ summiert, dass das Ereignis unendlich oft eintritt.

Also gibt es kein $N \in \mathbb{N}$, ab dem $\forall n > N$ gilt $X_n = 0$. Somit ist $\mathbb{P}[\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = 0\}] = 0 \neq 1$. Also $X_n \not\xrightarrow{f.s.} 0$.

Konvergenz im Moment

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X_n: \Omega \to \mathbb{R}$ eine ZV $\forall n \in \mathbb{N}$.
Dann konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ im r-ten Moment $(r \in \mathbb{N})$ gegen $X: \Omega \to \mathbb{R}$, wenn $\mathbb{E}[|X_n^r|] < \infty\ \forall n \in \mathbb{N}$ und $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0$. Wir schreiben $X_n \xrightarrow{r} X$.

Satz

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X_n: \Omega \to \mathbb{R}$ eine ZV $\forall n \in \mathbb{N}$ und $X: \Omega \to \mathbb{R}$. Dann gilt
$\forall r > s \geq 1: \ X_n \xrightarrow{r} X \implies X_n \xrightarrow{s} X$.

Beispiel

Sei $r > s \geq 1$ und $\mathbb{P}[X_n = n] = n^{-\frac{1}{2}(r+s)}$ und $\mathbb{P}[X_n = 0] = 1 - n^{-\frac{1}{2}(r+s)}$
$\mathbb{E}[|X_n|^s] = n^s \cdot n^{-\frac{1}{2}(r+s)} = n^{-\frac{1}{2}\overbrace{(r-s)}^{>0}} \xrightarrow{n\to\infty} 0$. Also $X_n \xrightarrow{s} 0$.

Satz

Gilt $X_n \xrightarrow{r} X$ und $\sum_{n=1}^{\infty} \mathbb{E}[|X_n - X|^r] < \infty$, so folgt $X_n \xrightarrow{f.s.} X$.

Konvergenz in Verteilung

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von ZVen mit Verteilungsfkt. F_i.
Dann konvergiert $(X_i)_{i\in\mathbb{N}}$ in Verteilung gegen $X: \Omega \to \mathbb{R}$ mit Verteilungsfkt. F, wenn
$\forall x \in \{x \in \Omega \mid F \text{ ist stetig in } x\}: \lim_{i\to\infty} \underbrace{F_i(x)}_{\mathbb{P}(X_n \leq x)} = \underbrace{F(x)}_{\mathbb{P}(X \leq x)}$. Wir schreiben $X_i \xrightarrow{\mathcal{D}} X$.

! Gleiche Definition für Zufallsvektoren $(\underline{X}_i)_{i\in\mathbb{N}}$.

Satz von Slutsky

Seien $X_n \xrightarrow{\mathcal{D}} X$, $A_n \xrightarrow{\mathbb{P}} a$, $B_n \xrightarrow{\mathbb{P}} b$,
dann gilt $A_n + B_n X_n \xrightarrow{\mathcal{D}} a + bX$

Satz

Sei $(X_n)_{n\in\mathbb{N}} \sim \mathcal{B}(n, p_n)$ und $(p_n)_{n\in\mathbb{N}}$ eine Folge mit $\lim_{n\to\infty} n \cdot p_n = \lambda > 0$, dann gilt $X_n \xrightarrow{\mathcal{D}} \mathrm{Po}(\lambda)$

Zusammenhang

$$\begin{tikzcd}
X_n \xrightarrow{f.s.} X \arrow[dr, Rightarrow] & & X_n \xrightarrow{r} X \arrow[ll, Rightarrow, "\sum_{k=1}^{\infty} \mathbb{E}[|X_k - X|^r] < \infty"'] \arrow[dl, Rightarrow] \\
& X_n \xrightarrow{\mathbb{P}} X \arrow[d, Rightarrow] & \\
& X_n \xrightarrow{\mathcal{D}} X &
\end{tikzcd}$$

Gesetz der großen Zahlen und zentraler Grenzwertsatz

i.i.d.-Modell

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von ZVen. Wir nennen die X_i der Folge i.i.d. (independent identically distributed), wenn sie (i) alle unabhängig voneinander sind und (ii) die gleiche Verteilung besitzen.

Idee

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.
Wir betrachten die Summen $S_n = X_1 + \ldots + X_n$ und interessieren uns für das Verhalten von S_n für $n \to \infty$.
Die Gesetze der grossen Zahlen beschreiben die Konvergenz von $\frac{1}{n} \cdot S_n$ für $n \to \infty$, während der zentrale GWS eine Angabe über die asymptotische Form der Verteilung von S_n macht.

Schwaches Gesetz der großen Zahlen

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von ZVen mit $\mathbb{E}[X_i]=\mu \ \forall i \in \mathbb{N}$.
Wir sagen, dass das schwache Gesetz der grossen Zahlen gilt, wenn $\bar{X}_n := \frac{1}{n}\cdot\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mu$. Also $\forall \varepsilon > 0: \lim_{n\to\infty} \mathbb{P}[|\bar{X}_n - \mu| > \varepsilon] = 0$

Satz

heißt: Erwartungswert und Varianz existieren.
Wenn die X_i i.i.d. sind, mit $\mathbb{E}[X_i^2] < \infty \ \forall n \in \mathbb{N}$, so gilt das schwache Gesetz der großen Zahlen.

Beweis: Wegen $(X_n)_{n\in\mathbb{N}}$ i.i.d. und $\mathbb{E}[X_n]=\mu$ und $Var[X_n]=\sigma^2 \ \forall n\in\mathbb{N}$ folgt: $\mathbb{E}[\bar{X}_n]=\mu$ und $Var[\bar{X}_n] = \frac{1}{n} Var[X_n]$. $\mathbb{P}[|\bar{X}_n - \mu| > \varepsilon] \overset{\text{Tschebychev}}{\leq} \frac{Var(X_n)}{n \cdot \varepsilon^2}$. Für $n\to\infty$ folgt der Satz.

Anmerkung: X_n i.i.d. ist nicht zwingend nötig. Es reicht gleich-verteilt und unkorreliert (statt unabhängig) und das der Erwartungswert $\mathbb{E}[X_i]$ existiert.

Starkes Gesetz der großen Zahlen

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von ZVen mit $\mathbb{E}[X_i]=\mu \ \forall n \in \mathbb{N}$.
Wir sagen, dass das starke Gesetz der grossen Zahlen gilt, wenn $\bar{X}_n := \frac{1}{n}\cdot\sum_{i=1}^{n} X_i \xrightarrow{f.s.} \mu$. Also $\mathbb{P}[\lim_{n\to\infty} \bar{X}_n = \mu] = 1$

Satz

heißt: Erwartungswert existiert und ist endlich.
Wenn die X_i i.i.d. sind, mit $\mathbb{E}[|X_i|] < \infty \ \forall n \in \mathbb{N}$, so gilt das starke Gesetz der großen Zahlen.

Intuition: Schwaches Gesetz besagt, dass endliche Stichproben für grosses n, bis auf ein paar unglückliche Stichproben, "ausgeglichen" sind. Starkes Gesetz besagt, dass unendliche Stichproben fast-immer "ausgeglichen" sind und das die Menge der "unausgeglichenen" Stichproben Wahrscheinlichkeit 0 hat.

zentraler Grenzwertsatz

Seien $(X_n)_{n\in\mathbb{N}}$ i.i.d. ZVen mit $\mathbb{E}[X_n]=\mu$ und $Var[X_n] = \sigma^2 < \infty$. Sei $S_n = \sum_{i=1}^{n} X_i$. Dann gilt:

1) $\lim_{n\to\infty} P\left[\frac{S_n - n\mu}{\sqrt{n\cdot\sigma^2}} \leq x\right] = \mathcal{N}(0,1)$ bzw. $\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{D} N(0,1)$ bzw. $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$

2) $\lim_{n\to\infty} P\left[\sqrt{n}\cdot\frac{(\bar{X}_n - \mu)}{\sigma} \leq x\right] = \mathcal{N}(0,1)$ bzw. $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)$ bzw. $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

zentraler Grenzwertsatz mehrdimensional

Seien $(\underline{X}_n)_{n\in\mathbb{N}}$ i.i.d. k-dim. ZV mit $\mathbb{E}[\underline{X}_n \underline{X}_n^T] < \infty$. Sei $\mathbb{E}[\underline{X}_n] = \underline{\mu}$ und $\mathbb{V}[\underline{X}_n] = \Sigma$, dann gilt:

$\lim_{n\to\infty} P\left[\frac{\sum_{j=1}^{n} \underline{X}_j - n\cdot\underline{\mu}}{\sqrt{n}} \leq \underline{x}\right] = \mathcal{N}_k(\underline{0}, \Sigma)$ bzw. $\frac{1}{\sqrt{n}}\left(\sum_{j=1}^{n} \underline{X}_j - n\cdot\underline{\mu}\right) \xrightarrow{D} \mathcal{N}_k(\underline{0}, \Sigma)$

Schlussfolgerung aus 'thinking fast and slow'

- Große Stichproben liefern präzisere Ergebnisse als kleine Stichproben.
- Kleine Stichproben führen häufiger zu extremen Ergebnissen als grosse Stichproben.

Satz von Gliwenko-Cantelli / Hauptsatz der Statistik

Seien $X_1, \ldots, X_n$ i.i.d. ZVen mit $X_i \sim F$. Wir definieren die empirische Verteilungsfkt. als $\hat{F}(x) := \frac{1}{n}\cdot\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$. Es gilt:

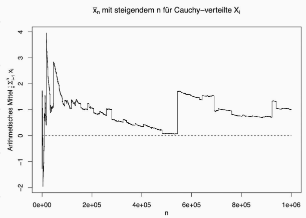
(i) $\hat{F} \xrightarrow{f.s.} F \quad \forall x \in \mathbb{R}$

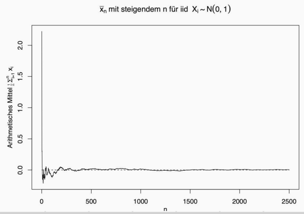
(ii) $\sqrt{n}\left(\hat{F}_n(x) - F(x)\right) \xrightarrow{D} \mathcal{N}(0, F(x)\cdot(1-F(x))) \quad \forall x\in\mathbb{R}$

(iii) $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{f.s.} 0$

Beispiel für den Fall, dass das Gesetz d.g. Zahlen nicht gilt

Sei $(X_i)_{i\in\mathbb{N}}$ i.i.d. mit $X_i \sim \mathcal{C}$. Dann kann man zeigen, dass auch $\bar{X}_n \sim \mathcal{C}$. Somit ist $E(\bar{X}_n)$ nicht existent. $\bar{X}_n$ streut also stark $\forall n \in \mathbb{N}$.
Somit ist das schwache Gesetz der grossen Zahlen nicht erfüllt.





Zufallsvektoren

Zufallsvektor

Ist $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ Meßraum, so heißt eine $\mathcal{F}_1$-$\mathcal{B}^n$-messbare Abb. $\underline{X}: \Omega_1 \longrightarrow \mathbb{R}^n$ n-dimensionaler Zufallsvektor.

n-dim. σ-Algebra

Ist $\Omega = \bigtimes_{i=1}^{n} \Omega_i$ und $\mathcal{F}_i$ σ-Algebra von Ω_i für alle i, dann lässt sich eine σ-Algebra für Ω definieren als: $\bigotimes_{i=1}^{n} \mathcal{F}_i := \sigma\left(\left\{\bigtimes_{i=1}^{n} A_i \,\middle|\, A_i \in \mathcal{F}_i,\ i=1,\dots,n\right\}\right)$

Borelsche σ-Algebra im n-dim.

Siehe Def. Borelsche σ-Algebra. $\mathcal{O}^n = \bigtimes_{i=1}^{n} \mathcal{O}$

Produktmaß

Das Produktmaß $\bigotimes_{i=1}^{n} \mu_i$ ist definiert durch $\left(\bigotimes_{i=1}^{n} \mu_i\right)\left(\bigtimes_{i=1}^{n} A_i\right) := \prod_{i=1}^{n} \mu_i(A_i) \quad \forall A_i \in \mathcal{F}_i,\ i \in \{1,\dots,n\}$

Produktmaßraum

Seien $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i=1,\dots,n$ Maßräume. Ihr Produktmaßraum ist der Maßraum $\left(\bigtimes_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^{n} \mathcal{F}_i, \bigotimes_{i=1}^{n} \mu_i\right)$

Mehrdimensionale Verteilungsfunktion

n-dim. Verteilungsfunktion

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sei $\underline{X}: \Omega \longrightarrow \mathcal{B}^n$ eine n-dim. reelle ZV, $\underline{X} = (X_1, \dots, X_n)$.

Dann heißt $F_{\underline{X}}: \mathcal{B}^n \longrightarrow [0,1]$, $\underline{x} = (x_1, \dots, x_n) \mapsto \mathbb{P}[\underline{X} \le \underline{x}] = \mathbb{P}[X_1 \le x_1, \dots, X_n \le x_n] = \mathbb{P}[\{\omega \in \Omega \mid X_i(\omega) \le x_i \ \forall i\}]$

Satz

Sei $\underline{X}: \Omega \longrightarrow \mathbb{R}^n$ eine n-dim. ZV und $g: \mathbb{R}^n \longrightarrow \mathbb{R}^k$ $\mathcal{B}^n$-$\mathcal{B}^k$-meßbar.

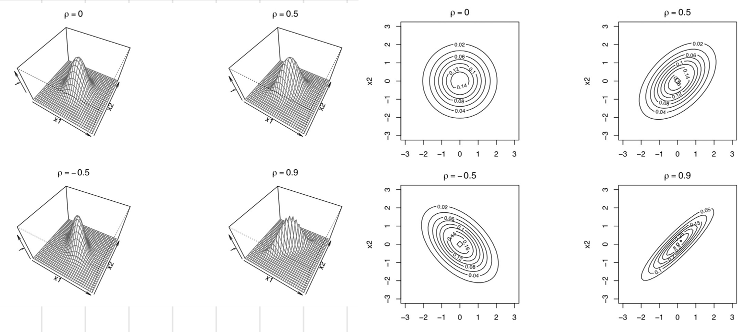
Dann ist $g \circ \underline{X}: \Omega \longrightarrow \mathbb{R}^k$, $\omega \longmapsto g(\underline{X}(\omega))$ eine k-dimensionale reelle ZV mit Verteilung $\mathbb{P}_{g(\underline{X})}[A] = \mathbb{P}[g(\underline{X}) \in A] = \mathbb{P}[\{\omega \in \Omega \mid g(\underline{X}(\omega)) \in A\}]$

2-dimensionale Normalverteilung

Sei $\underline{X}: \Omega \to \mathbb{R}^2$ eine ZV mit Dichte

$$f_{\underline{X}}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right)\right)$$

Dann ist $\underline{X} \sim \mathcal{N}_2\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$



Satz von Fubini

Seien $(\Omega_1, \mathcal{F}_1, \mu)$ und $(\Omega_2, \mathcal{F}_2, \nu)$ zwei Maßräume mit σ-endlichen Maßen μ und ν.

Ist $f: \Omega_1 \times \Omega_2 \to \mathbb{R}$ eine nicht-negative $(\mathcal{F}_1 \otimes \mathcal{F}_2)$-$\mathcal{B}$-meßbare Funktion bzw. eine $(\mu \otimes \nu)$-int.bare Fkt, so gilt

$$\int f \, d(\mu \otimes \nu) = \int_{\Omega_1}\int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu\, d\nu = \int_{\Omega_2}\int_{\Omega_1} f(\omega_1, \omega_2)\, d\nu\, d\mu$$

Randverteilung

Ist $g: \mathbb{R}^n \longrightarrow \mathbb{R}$ mit $g(\underline{x}) = x_j$ für ein festes $j \in \{1,\dots,n\}$, so gilt $g(\underline{X}) = X_j$, $\mathbb{P}_{X_j}[A] = \mathbb{P}_{\underline{X}}[\{\underline{x} \in \mathbb{R}^n \mid x_j \in A\}] = \mathbb{P}[X_j \in A]$ für $A \in \mathcal{B}$.

$\mathbb{P}_{X_j}$ heißt Randverteilung oder marginale Verteilung von X_j.

Satz

Besitzt $\underline{X}$ die Dichte f, so hat X_j die marginale Dichte bzw. Randdichte $f_j: \mathbb{R} \to \mathbb{R}$

$$f_j(y) = \int \dots \int f(x_1, \dots, x_{j-1}, y, x_{j+1}, \dots, x_n)\, dv_1 \dots dv_{j-1}\, dv_{j+1} \dots dv_n$$

Beispiel

Für die 2-dim. Norm.verteilte ZV $\underline{X}$ gilt:

$f_{X_1}(x_1) = \int_{\mathbb{R}} f(x_1, x_2)\, dx_2 = \dots = \frac{1}{\sqrt{2\pi\sigma_1^2}} \cdot \exp\left(-\frac{1}{2}(x_1-\mu_1)^2\right)$. Also ist $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$. Genauso gilt $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

Multinomialverteilung

Urnenmodell mit k Farben und n gezogenen Kugeln ohne zurücklegen. Die ZV $X = (X_1, \dots, X_k)$ beschreibt die gezogenen Farben, wobei X_i die Anzahl Kugeln mit Farbe i ist.

Sei $k \le n \in \mathbb{N}$ und $p_1, \dots, p_k \in [0,1]$ mit $\sum_{i=1}^{n} p_i = 1$. Die Verteilung mit der Zähldichte $f(n_1, \dots, n_k) = \frac{n_1 + \dots + n_k}{n_1! \cdot \ldots \cdot n_k!} \cdot p_1^{n_1} \cdot \ldots \cdot p_k^{n_k}$ heißt Multinomialverteilung.

Beispiel

$f(n_1, \dots, n_k) = \frac{n_1 + \dots + n_k}{n_1! \cdot \ldots \cdot n_k!} \cdot p_1^{n_1} \cdot \ldots \cdot p_k^{n_k}$

$f_{X_1}(n_1) = \sum_{n_2 + \dots n_k = n - n_1} f(n_1, \dots, n_k) = \dots = \frac{n!}{n_1!(n-n_1)!} p_1^{n_1} \cdot \left(\sum_{i=2}^{k} p_i\right)^k$

Unabhängigkeit

Satz

Sei $\underline{X} = (X_1, \dots, X_n)$ eine reelle n-dimensionale ZV. Dann gilt

$X_1, \dots, X_n$ stochastisch unabhängig $\iff \forall \underline{x} \in \mathbb{R}^n: \mathbb{P}[\underline{X} \le \underline{x}] = \prod_{i=1}^{n} \mathbb{P}[X_i \le x_i]$

Satz

Sind $X_1, \dots, X_n$ reelle ZV mit Dichten $f_{X_1}, \dots, f_{X_n}$.

$X_1, \dots, X_n$ sind unabhängig $\iff$ Für $\underline{X} = (X_1, \dots, X_n)$ gilt $\forall \underline{x}: f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$

Satz

Seien $\underline{X}_1 : \Omega \to \mathbb{R}^{n_1}$, $\underline{X}_2 : \Omega \to \mathbb{R}^{n_2}$ zwei stochastisch unabhängige ZV und seien $h_1 : \mathbb{R}^{n_1} \to \mathbb{R}^{m_1}$ und $h_2 : \mathbb{R}^{n_2} \to \mathbb{R}^{m_2}$ meßbare Funktionen.

Dann sind $\underline{Y}_1 := h_1 \circ \underline{X}_1$ und $\underline{Y}_2 := h_2 \circ \underline{X}_2$ stoch. unabhängig.

Transformationssatz

Transformationssatz für lineare Transformationen

Sei $g : \mathbb{R}^n \to \mathbb{R}^n$ invertierbar und linear, d.h. $g(\underline{x}) = B\underline{x} + \underline{m}$ mit $\det(B) \neq 0$. Sei $\underline{Y} = g(\underline{X})$.

Wenn $f_{\underline{X}}$ absolutstetig ist, so ist auch $f_{\underline{Y}}$ absolutstetig und $f_{\vec{Y}}(\vec{y}) = \frac{1}{|\det(B)|} \cdot f_{\vec{X}}(B^{-1} \cdot (\vec{y} - \vec{m}))$

Allgemeiner Transformationssatz

Sei $\underline{X} : \Omega \to \mathbb{R}^n$ ein ZV mit stetiger Dichte $f_{\underline{X}}$. Sei $g : \mathbb{R}^n \to \mathbb{R}^n$ meßbar. Sei $G_m \in \mathcal{B}^n$ mit $\dot{\bigcup}_{m \in M} G_m = \mathbb{R}^n$ (Partition von $\mathbb{R}^n$.), so dass

$g_m := g|_{G_m}$ bijektiv und stetig diff.bar. Sei $J_{g_m^{-1}}(\underline{y}) = \left(\frac{\partial g_{m,i}^{-1}(\underline{y})}{\partial y_j} \right)_{\substack{i=1,\dots,n \\ j=1,\dots,n}}$ die Jacobimatrix von g_m^{-1}. Dann gilt:

$$f_{g(\underline{X})}(\underline{x}) = \sum_{m \in M} v_m(\underline{x}) \quad , \quad v_m(\underline{x}) = \begin{cases} f_{\underline{X}}(g_m^{-1}(\underline{x})) \cdot \det(J_{g_m^{-1}}(\underline{x})) & \text{, falls } g_m^{-1}(\underline{x}) \in G_m \\ 0 & \text{, sonst} \end{cases}$$

Beispiel Faltung

Sei $\vec{X} = (X_1, X_2)$ eine zweidimensionale absolutstetige ZV. Setze $\vec{Y} = g(\vec{X}) = (\bar{X}_1, \bar{X}_1 + \bar{X}_2)$.

Dann ist $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ bzw. $B^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. $\Rightarrow f_{\vec{Y}}(x_1, y) = \frac{1}{1} \cdot f_{\vec{X}}\left(B^{-1} \cdot \binom{x_1}{y}\right) = f_{\vec{X}}(x_1, y - x_1)$

Für $\vec{Z} = \bar{X}_1 + \bar{X}_2$ gilt also: $f_{\vec{Z}}(\vec{z}) = \int_{\mathbb{R}} f_{\vec{Y}}(x_1, y)\, dx_1 = \int_{\mathbb{R}} f_{\vec{X}}(x_1, y - x_1)\, dx_1$.

Für unabhängige X_1, X_2 gilt $f_{\vec{X}}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$. $\Rightarrow f_{\vec{Z}}(\vec{z}) = \int_{\mathbb{R}} f_{X_1}(x_1) \cdot f_{X_2}(y - x_1)\, dx_1 = f_{X_1} * f_{X_2}$

Faltung beschreibt wie die Form von einer Funktion durch eine andere Funktion modifiziert wird. Anschaulich bedeutet $f * g$, dass jeder Wert von f durch das mit g gewichtete Mittel der ihn umgebenden Werte ersetzt wird.

Beispiel

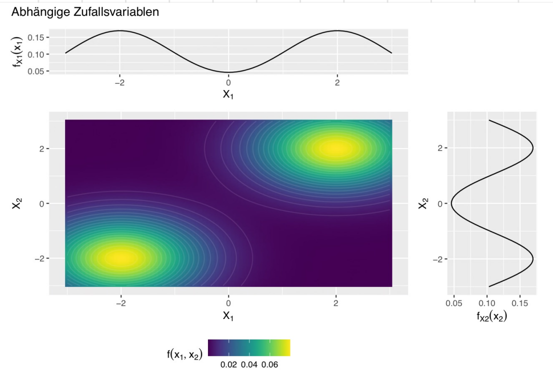
Seien $X_1, X_2 \sim \mathcal{G}(\pi)$ und unabhängig. Wir betrachten $\bar{Y} = X_1 + X_2$. $T_{\bar{Y}} = \mathbb{N}_+ \setminus \{1\}$

$$P(Y = n) = P(X_1 + X_2 = n) = \sum_{k=1}^{n-1} P(X_1 = k) \cdot P(X_2 = n-k) = \sum_{k=1}^{n-1} (1-\pi)^{k-1} \cdot \pi \cdot (1-\pi)^{n-k-1} \cdot \pi = \sum_{k=1}^{n-1} (1-\pi)^{n-2} \cdot \pi^2 = (n-1) \cdot (1-\pi)^{n-2} \cdot \pi^2$$

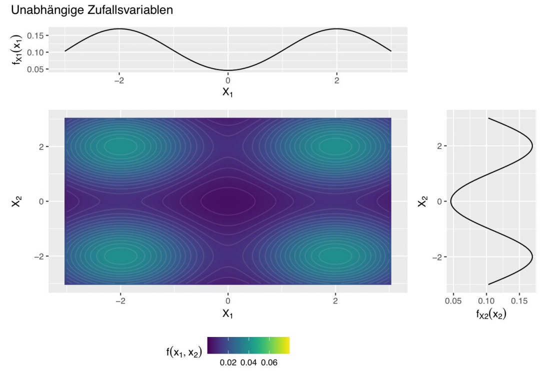
Beispiel

Seien $X_1 \sim P(\lambda_1)$, $X_2 \sim P(\lambda_2)$ und unabhängig. Wir betrachten $\bar{Y} = X_1 + X_2$.

$$P(Y = n) = P(X_1 + X_2 = n) = \sum_{k=0}^{n} P(X_1 = k) \cdot P(X_2 = n-k) = \underbrace{\left(\sum_{k=0}^{n} \binom{n}{k} \cdot \lambda_1^k \cdot \lambda_2^{n-k} \right)}_{= (\lambda_1 + \lambda_2)^n} \cdot \frac{e^{-(\lambda_1 + \lambda_2)}}{n!}$$



$f(x_1, x_2) \neq f_{X1}(x_1) f_{X2}(x_2)$



$f(x_1, x_2) = f_{X1}(x_1) f_{X2}(x_2)$

Bedingte Verteilungs- & Dichtefunktion

Die bedingte Verteilungsfunktion $F_{X|Y}(x|y)$ und die bedingte Wahrscheinlichkeits- bzw. Dichtefunktion $f_{X|Y}(x,y)$, gegeben $Y=y$ mit $y \in T_Y$, sind definiert als:

$$F_{X|Y} = P(X \leq x \mid Y=y) = \begin{cases} \dfrac{P(X \leq x, Y=y)}{P(Y=y)} &, X,Y \text{ diskret} \\ \displaystyle\int_{-\infty}^{x} \frac{f_{X|Y}(u,y)}{f_Y(y)}\, du &, X,Y \text{ stetig} \end{cases}$$

$$f_{X|Y}(x,y) = \begin{cases} \dfrac{P(X=x, Y=y)}{P(Y=y)} &, X,Y \text{ diskret} \\ \dfrac{f_{X,Y}(x,y)}{f_Y(y)} &, X,Y \text{ stetig} \end{cases}$$

Beispiel

Sei $X \sim Be(\alpha,\beta)$ und $Y|X = B(n, \pi = X)$.

Dann gilt $f(x,y) = \frac{1}{B(\alpha,\beta)} \cdot \binom{n}{y} \cdot x^{y-\alpha-1} \cdot (1-x)^{n-y+\beta-1}$

$\Rightarrow X|Y \sim Be(\alpha+y, \beta+n-y)$

$f_Y(y) = \binom{n}{y} \cdot \frac{B(y+\alpha, n-y+\beta)}{B(\alpha,\beta)}$ für $y = 0, \dots, n$

$Y \sim BB(n,\alpha,\beta)$ Beta-Binomialverteilung

Beta-Binomialverteilung

Die Anzahl Erfolge in n unabhängigen Bernoulli-Experimenten mit Beta-verteilten Erfolgswahrscheinlichkeiten ist Beta-Binomialverteilt mit $X \sim BeB(n,\alpha,\beta)$ und $n \in \mathbb{N}$, $\alpha, \beta \in \mathbb{R}^+$. Es gilt

$$P_{n,\alpha,\beta}(X=x) = \binom{n}{x} \cdot \frac{B(\alpha+x, b+n-x)}{B(\alpha,\beta)}$$

wobei B die Betafunktion ist.

Bedingte Momente

$$E(X|Z=z) := \begin{cases} \displaystyle\sum_{x \in T_X} x \cdot P(X=x \mid Z=z) &, X \text{ diskret} \\ \displaystyle\int_{\mathbb{R}} x \cdot f_{X|Z}(x|Z=z)\, dx &, X \text{ stetig} \end{cases}$$

$E(X|Z)$ ist ein ZV
$E(X|Z=z)$ ist ein Zahlenwert

$$Var(X|Z=z) := E\big(X - E(X|Z=z) \mid Z=z\big) = \begin{cases} \displaystyle\sum_{x \in T_X} \big(x - E(X|Z=z)\big)^2 \cdot P(X=x \mid Z=z) &, X \text{ diskret} \\ \displaystyle\int_{\mathbb{R}} \big(x - E(X|Z=z)\big)^2 f_{X|Z}(x|Z=z)\, dx &, X \text{ stetig} \end{cases}$$

Satz vom iterierten Erwartungswert

Für beliebige ZV X,Y und Funktion f gilt:

$$E\big[E(f(X)|Z)\big] = E(f(X))$$

Satz von der totalen Varianz

Für beliebige ZV X,Y gilt:

$$Var(X) = \underbrace{E(Var(X|Z))}_{\text{erwartete bedingte Varianz}} + \underbrace{Var(E(X|Z))}_{\text{Varianz des bedingten Erwartungswertes}}$$

Satz

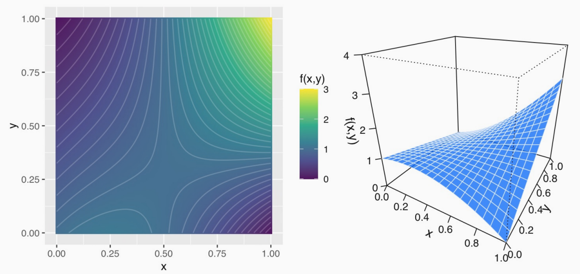
$$f_{X,Y}(x,y) = f_{X|Y}(x,y) \cdot f_Y(y) = f_{Y|X}(x,y) \cdot f_X(x)$$

X,Y sind unabhängig $\Leftrightarrow f_{X|Y}(x,y) = f_X(x)\ \forall x,y \in X,Y \Leftrightarrow f_{Y|X}(x,y) = f_Y(y)\ \forall x,y \in X,Y$

Beispiel

Sei $(X,Y) \in [0,1] \times [0,1]$ ein stetiger Zufallsvektor mit gemeinsamer Dichte

$$f_{X,Y}(x,y) = (1 + x - y - 2x^2 + 4x^2 y) \cdot I(x \in [0,1]) \cdot I(y \in [0,1]):$$



$$f_X(x) = \int_0^1 f_{X,Y}(x,y)\,dy = \int_0^1 1+x-y-2x^2+4x^2 y\, dy = 1 + x - 2x^2 + (4x^2-1) \cdot \int_0^1 y\, dy = \tfrac{1}{2} + x$$

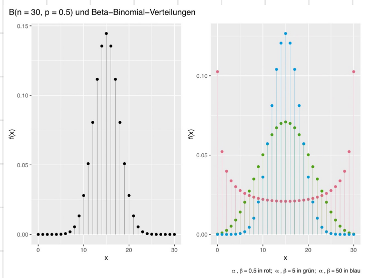
$$f_{Y|X}(x,y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1+x-y-2x^2+4x^2 y}{\frac{1}{2}+x} = 2 \cdot (2xy - x - y + 1)$$

Beispiel

Seien X,Y zwei unabhängige ZV. $X \sim \mathcal{P}(\lambda_1)$, $Y \sim \mathcal{P}(\lambda_2)$, $Z = X + Y \sim \mathcal{P}(\lambda_1+\lambda_2)$

$X|Z = z \sim B\left(z, \pi = \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

$$P(X=x|Z=z) = \frac{P(X=x, Y=z-x)}{P(Z=z)} = \frac{P(X=x) \cdot P(Y=z-x)}{P(Z=z)} = \binom{z}{x} \cdot \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x \cdot \left(1 - \frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{z-x}$$



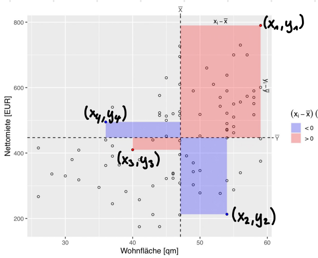
Zusammenhangsmaße für metrische Merkmale

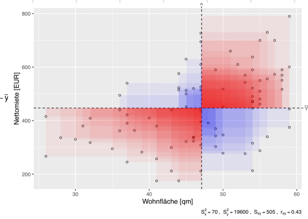
Kovarianz (Emperie)

Die Kovarianz ist eine Kennzahl für Stärke & Richtung des linearen Zusammenhangs zweier metrischer Merkmale.

Für die Daten $(x_i, y_i)_{i=1,\dots,n}$ ist die Kovarianz definiert als:

$$S_{XY} := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$





Bravais-Pearson-Korrelationskoeffizient

Für die Daten $(x_i, y_i)_{i=1,\dots,n}$ ist der BPK-Koeffizient definiert als:

$$r_{XY} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{S_X} \frac{(y_i - \bar{y})}{S_Y} = \frac{S_{XY}}{S_X S_Y} \in [-1, 1]$$

$r_{XY} > 0$: gleichsinniger linearer Zusammenhang

$r_{XY} < 0$: gegenläufiger linearer Zusammenhang

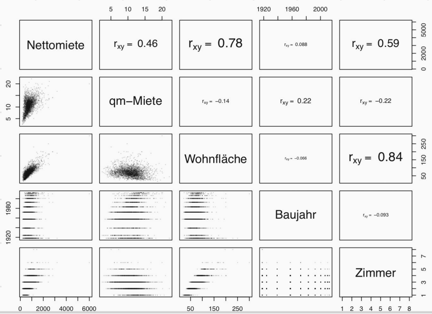
$r_{XY} = 0$: kein linearer Zusammenhang (unkorreliert)

!• Misst nur linearen Zusammenhang.

• $r_{XY} = 0$ bedeutet nicht automatisch Unabhängigkeit.

• Korrelation & Kovarianz sind Ausreisser empfindlich.

Beispiel multivariater Zusammenhänge

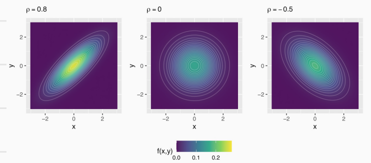


Beispiel - Bivariate Standartnormalverteilung

Die **bivariate Standardnormalverteilung** mit Parameter ρ ($|\rho| < 1$) hat Träger $T = \mathbb{R} \times \mathbb{R}$ und Dichtefunktion

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

Die Randverteilungen von X, Y sind für jedes ρ Standartnormalverteilt. $\rho(X, Y) = \rho$. Wenn $\rho = 0$, so folgt in diesem Spezialfall Unabhängigkeit von X, Y.



Spearman-/Rang-Korrelationskoeffizient

X, Y (mindestens) ordinal.

Idee: Berechne den Bravais-Pearson-Korrelationskoeffizienten für die Ränge statt für die Merkmalswerte.

$$r_{XY}^{SP} = \frac{\sum_{i=1}^{n} (R(x_i) - \bar{R}_X)(R(y_i) - \bar{R}_Y)}{\sqrt{\sum_{i=1}^{n} (R(x_i) - \bar{R}_X)^2 \cdot \sum_{i=1}^{n} (R(y_i) - \bar{R}_Y)^2}} = r_{R(X)R(Y)}$$

Seien $(x_{(i)}, y_{(i)})$ die nach Grösse sortierten Daten (aufsteigend), dann ist $R(x_{(i)}) = i$, wenn $x_{(i)}$ unique ist. Für mehrere $\underbrace{\text{gleiche Werte}}_{\text{Ties}}$ $x_{(i)} = x_{(i+1)} = \dots$ nimmt man den Durchschnitt der Indizes als Rang.

$r_{XY}^{SP} > 0$: gleichsinniger monotoner Zusammenhang

$r_{XY}^{SP} < 0$: gegenläufiger monotoner Zusammenhang

$r_{XY}^{SP} = 0$: kein monotoner Zusammenhang (unkorreliert)

$$\bar{R}_X = \bar{R}_Y = \frac{1}{n} \sum_{i=1}^{n} i = \frac{n+1}{2}$$

Wenn alle $R(x_i), R(y_i)$ voneinander verschieden sind, so kann man r_{XY}^{SP} verkürzt berechnen durch

$$r_{XY}^{SP} = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{(n^2-1) \cdot n}, \quad d_i := R(x_i) - R(y_i)$$

Paarvergleichsmaße: Kendall's Tau

Betrachte Paare von Beobachtungen (x_i, y_i), (x_j, y_j). Ein Paar heißt

konkordant falls $\binom{x_i}{y_i} < \binom{x_j}{y_j}$ oder $\binom{x_i}{y_i} > \binom{x_j}{y_j}$.

diskordant falls $x_i < x_j$ und $y_i > y_j$ oder $x_i > x_j$ und $y_i < y_j$

N_c ist die Anzahl konkordanter Paare. N_d die Anzahl diskordanter Paare.

Kendalls Tau ist definiert als $\tau_{XY} := \frac{N_c - N_D}{\binom{n}{2}} = 2 \cdot \frac{N_c - N_D}{n \cdot (n-1)}$

Varianten:
- Goodman & Kruskal γ-Koeffizient ignoriert Bindungen (Ties).

$$\gamma_{XY} := \frac{N_c - N_D}{N_c + N_D}$$

- Somers D wird typischerweise Verwendet, wenn Y binär ist.
  $\Rightarrow$ Viele Paare mit Bindungen in Y.

$$D_{XY} := \frac{N_c - N_D}{\text{Anzahl Paare mit ungleichem } y}$$

Distanzkovarianz & Distanzkorrelation

Modernes Zusammenhangsmaß für (fast) beliebige Zusammenhänge (nicht nur monoton/lin.) basierend auf Produkten der normierten Distanzen der Beobachtungen untereinander, nicht nur Distanzen zu Mittelwerten.

$$dS_{XY} := \frac{1}{n^2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij}^X \cdot D_{ij}^Y$$

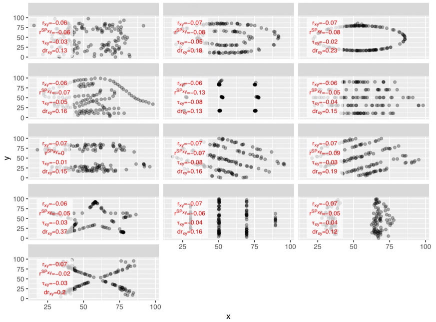
$$D_{ij}^X := d_{ij}^X + \bar{d}^X - \bar{d}_{\cdot j}^X - \bar{d}_{i \cdot}^X, \quad d_{ij}^X := |x_i - x_j|,$$

$$\bar{d}_{\cdot j}^X := \frac{1}{n} \sum_{i}^{n} d_{ij}, \quad \bar{d}_{i \cdot}^X := \frac{1}{n} \sum_{j}^{n} d_{ij}^X, \quad \bar{d}^X = \frac{1}{n^2} \sum_i \sum_j d_{ij}^X$$

$$dr_{XY} := \sqrt{\frac{dS_{XY}^2}{\sqrt{dS_{XX}^2 dS_{YY}^2}}} \in [0, 1]$$

! Misst nur Stärke und nicht Richtung.
Lässt sich mit geeigneter Metrik auf fast alles anwenden.

! $dr_{XY} = 0 \Leftrightarrow X, Y$ unabhängig



Zusammenhangsmaße für dichotome und ordinale/metrische Merkmale

Beispiel

> Medizin: X = Biomarker oder diagnostischer Score
> Y = krank vs. nicht krank

> Marketing: X = Kundeneigenschaften
> Y = Kaufentscheidung (ja/nein)

Sensitiviät und Spezifität

$Y \in \{0,1\}$ (Zielgröße), X mindestens ordinalskaliert

$Y=1$: "positiver" Fall , $Y=0$: "negativer" Fall

Sei $\hat{y}_i$ die Prognose für y_i auf Basis von x_i. Es soll gelten

$\hat{y}_i = 1 \Leftrightarrow x_i \geq c$.

	$y_i = 0$	$y_i = 1$	
Vorhersage $\hat{y}_i = 0$	wahr negativ	falsch negativ	# negative Vorhersagen
Vorhersage $\hat{y}_i = 1$	falsch positiv	wahr positiv	# positive Vorhersagen
	# negative	# positive	

Sensitivität: $TPR(c) = f(\hat{Y}=1 \mid Y=1) = f(X \geq c \mid Y=1) = \frac{\#\text{wahr pos.}}{\#\text{positiv}}$

true pos. rate

true neg. rate

Spezifität: $TNR(c) = f(\hat{Y}=0 \mid Y=0) = 1 - f(X \geq c \mid Y=0) = \frac{\#\text{wahr neg.}}{\#\text{negativ}} = 1 - FPR(c)$

False pos. rate: $FPR(c) = f(\hat{Y}=1 \mid Y=0) = f(X \geq c \mid Y=0) = \frac{\#\text{falsch pos.}}{\#\text{negativ}}$

ROC-Kurve

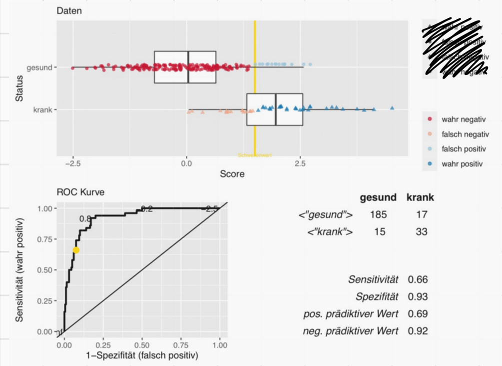
Die ROC-Kurve zeigt die Zuverlässigkeit der Vorhersagen für alle möglichen Schwellenwerte c an.

Verbindet die Punkte $(FPR(c), TPR(c))$ $\forall c \in [x_{(1)}, x_{(n)}]$

Für $c < x_{(1)} \Rightarrow \hat{y}_i = 1 \ \forall i \Rightarrow (FPR(c), TPR(c)) = (1,1)$

Für $c > x_{(n)} \Rightarrow \hat{y}_i = 0 \ \forall i \Rightarrow (FPR(c), TPR(c)) = (0,0)$

Beispiel



AUC-Maß

Maß zur Bewertung der ROC Kurve.

$$AUC := \frac{N_C + \frac{N_E}{2}}{N}$$

, N_C := Anzahl konkordanter Paare

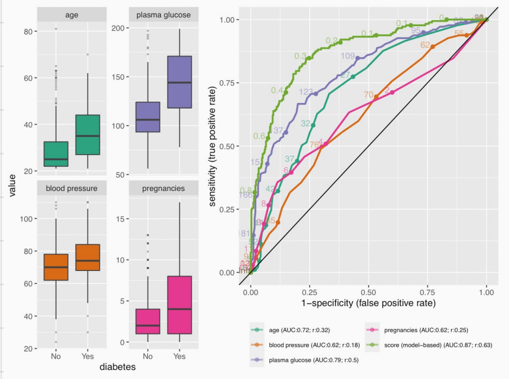
, N_E := Anzahl Paare mit Bindung in X.
$= |\{(i,j) : x_i = x_j , y_i \neq y_j\}|$

> perfekte Trennbarkeit der Y-Gruppen durch c bedeutet $AUC = 1$

> X, Y unabhängig $\Rightarrow AUC \approx 0$

! Behandelt Spezifität & Sensitivität gleichwertig und beachtet positiv/negativ prädikaten Wert nicht.

Beispiel



Positiv/negativ prädikater Wert (ppV, npV)

$ppV := f(Y=1 \mid \hat{Y}=1) = \frac{\#\text{wahr pos.}}{\#\text{pos. Vorhersagen}}$

$npV := f(Y=0 \mid \hat{Y}=0) = \frac{\#\text{wahr neg.}}{\#\text{neg. Vorhersagen}}$

Beispiel

NIPT-test für Trisomie 21. Prävalenz 8 in 10000.

NIPT-Sensitivität. 99.2% , NIPT-Spezifität: 99.9%

Erwartete Häufigkeiten:

	Kind krank	Kind gesund	
NIPT: "Kind krank"	794	999	1793
NIPT: "Kind gesund"	6	998 201	998207
	800	999 200	1 000 000

$npV = 0.99994$, $ppV = 0.443$

$\Rightarrow$ negative NIPT-Diagnosen sind nahezu sicher korrekt, aber von den positiven sind nur 44% korrekt.

		True condition			
	Total population	Condition positive	Condition negative	Prevalence = Σ Condition positive / Σ Total population	Accuracy (ACC) = (Σ True positive + Σ True negative) / Σ Total population
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = Σ True positive / Σ Predicted condition positive	False discovery rate (FDR) = Σ False positive / Σ Predicted condition positive
Predicted condition	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = Σ False negative / Σ Predicted condition negative	Negative predictive value (NPV) = Σ True negative / Σ Predicted condition negative
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = Σ True positive / Σ Condition positive	False positive rate (FPR), Fall-out, probability of false alarm = Σ False positive / Σ Condition negative	Positive likelihood ratio (LR+) = TPR / FPR	Diagnostic odds ratio (DOR) = LR+ / LR− ; F₁ score = 2 · (Precision · Recall) / (Precision + Recall)
		False negative rate (FNR), Miss rate = Σ False negative / Σ Condition positive	Specificity (SPC), Selectivity, True negative rate (TNR) = Σ True negative / Σ Condition negative	Negative likelihood ratio (LR−) = FNR / TNR	

Korrelation & Kausalität

$A \longrightarrow B$: "A ist Ursache von B"

$A \rightsquigarrow B$: "A und B korrelieren (nicht stoch. unabhängig)"

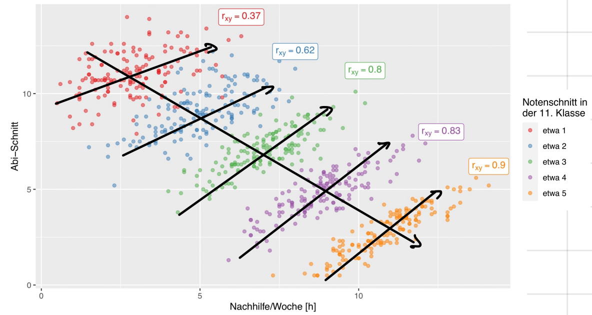
Fork: $Y \leftarrow X \rightarrow Z$, Collider: $Y \rightarrow X \leftarrow Z$, Pipe: $Y \rightarrow X \rightarrow Z$ führen alle zu $\Rightarrow$ $X \rightsquigarrow Y$, $X \rightsquigarrow Z$, $Y \rightsquigarrow Z$

Still a Fork: $Y \leftarrow X \rightarrow Z$, $Y \dashrightarrow Z$, still a Collider: $Y \rightarrow X \leftarrow Z$, $Y \dashrightarrow Z$, still a Pipe: $Y \rightarrow X \rightarrow Z$, $Y \dashrightarrow Z$ führen alle zu $\Rightarrow$ $X \rightsquigarrow Y$, $X \rightsquigarrow Z$, $Y \rightsquigarrow Z$

Confounding - Scheinkorrelation über Drittvariable

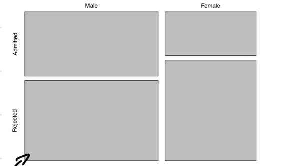
Sonnentage/Monat → Ertrunkene/Monat, Sonnentage/Monat → Eiscremeabsatz/Monat; Ertrunkene/Monat ⌇ Eiscremeabsatz/Monat

Alter → Schuhgröße, Alter → Wortschatz; Schuhgröße ⌇ Wortschatz

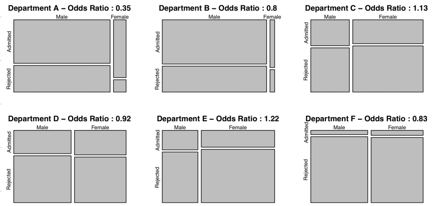


Scheinkorrelation über Aggregation

Aufnahmequote Uni Berkley



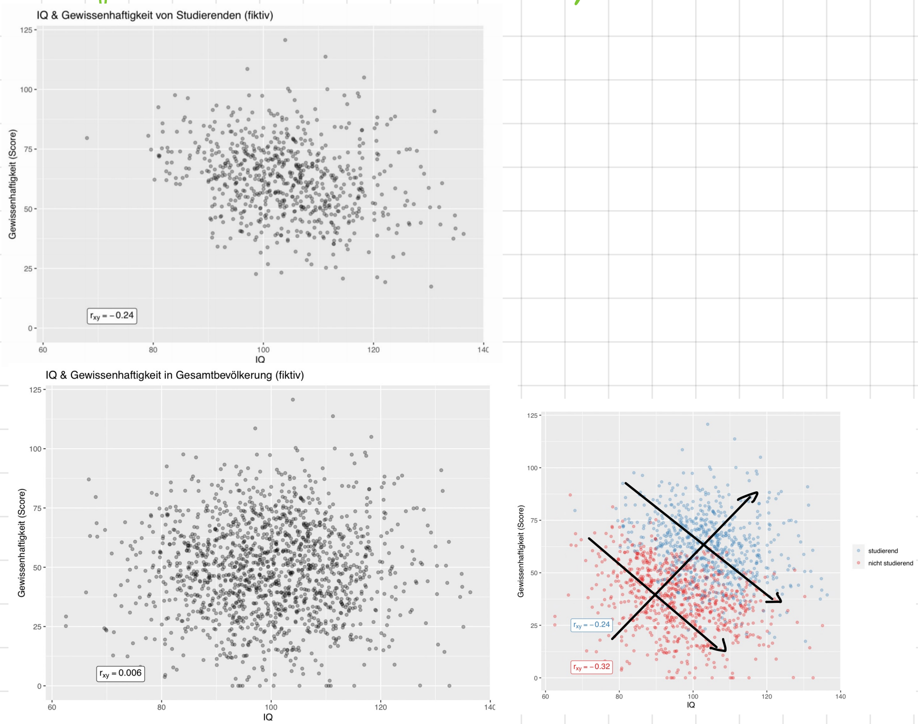
γ(Admitted, Rejected | Male, Female) = 1.83



Fach → Geschlecht, Fach → Aufnahme; Geschlecht ↓? Aufnahme

Durch Mediator 'Fach' wird Assoziation zwischen Geschlecht und Aufnahme komplexer

Conditioning on Collider - Scheinkorrelation durch Stichprobenwahl



Schlau → Studiert ← Fleißig

Scheinkorrelation zwischen 'Fleiß' und 'Schläue' durch conditioning auf 'Studieren'.
Nur diejenigen, die eher schlau oder fleißig sind können studieren:
Dadurch entsteht negative Korrelation zwischen Schläue und Fleiß in der Subpopulation der Studierenden.

Ob eine UE in der beobachteten Stichprobe ist, hängt von einer Variable ab, die von den untersuchten Variablen kausal abhängt: Selection-distortion-effect