Statistik	Aufgaben	Techniken
Deskriptive Statistik	Beschreibung, graphische Darstellung und Validierung von Daten. ❗Keine Rückschlüsse auf Grundgesamtheit möglich.	Grafiken, Tabellen, Kennzahlen
Explorative Statistik	Suche nach Struktur in den Daten (ohne stochastische Methoden). Formulierung von Hypothesen für das den Daten zugrunde liegende stochastische Modell.	Iterative und interaktive Anwendung von Techniken aus der deskriptiven und induktiven Statistik.
Induktive Statistik	Ziehung von Schlüssen von den Daten (Stichprobe) auf Grundgesamtheit. Basierend auf stochastischen Modellen.	Statistische Modellierung, statistische Tests, Konfidenzintervalle, Schätzer

Datenerhebung & Messung

Ein Merkmal

Matrikelnummer	Name	Vorname	Geburtsdatum	Hauptfach	Nebenfach
xxxxx 234	Muster	Peter	01.01.2001	Statistik	Informatik
xxxxx 556	Schmid	Lena	31.10.2002	Informatik	Statistik
xxxxx 123	Müller	Jonas	27.08.1999	Mathematik	NA
xxxxx 767	Nguyen	Cho	24.12.2000	Medizin	Soziologie
xxxxx 111	Nagel	Cosima	26.10.1996	Jura	Ethik

Alle Merkmale

Merkmalsausprägung vom Merkmal "Nebenfach" bei der zweiten statistischen Einheit.

Eine Beobachtung

Grundgesamtheit: Studenten der LMU (über welche "Objekte" erhebe ich Daten?)

Stichprobe: z.B. Alle Statistik Studenten ❗Stichproben müssen nicht per Definition zufällig gewählt sein.

Statistische Einheit / Untersuchungseinheit (UE): Ein Student bzw. ein Element der Grundgesamtheit

Merkmal: Messbare Eigenschaft einer statistischen Einheit. In der Tabelle quasi der (sinnvolle) Spaltenname. z.B. Hauptfach ist ein Merkmal

Merkmalsausprägung: Der tatsächliche Wert des Merkmals bei einer statistischen Einheit. In der Tabelle ist das ein Wert in einer Zelle.

Beobachtung: Alle Merkmalsausprägungen einer statistischen Einheit zu einem Zeitpunkt. In der Tabelle sind das die Werte in einer Zeile.

Unterscheidung nach …		
… Quantifizierbarkeit der Ausprägungen	**Qualitative Merkmale**: • nur zuordenbar (einstufig) • Beispiele: Wohnort, Name	**Quantitative Merkmale**: • mess- oder zählbar • Beispiele: Alter, Körpergröße
… Anzahl der Ausprägungen*	**Diskrete Merkmale**: • höchstens abzählbar unendlich viele mögliche Ausprägungen • Beispiele: Gehaltsklassen, Kaufverhalten	**Stetige Merkmale**: • überabzählbar unendlich viele mögliche Ausprägungen • Beispiele: Geschwindigkeit, Gewicht
… Direktheit der Informationsgewinnung	**Beobachtbare Merkmale**: • können direkt erhoben werden • Beispiel: Abiturnote	**Latente Merkmale**: • Operationalisierung über Indikatoren/Items notwendig • Beispiele: Bildungsgrad, Kreativität, Nutzen

(*) Merkmale die eigentlich diskret sind, aber so viele Ausprägungen haben, dass sie wie stetige Merkmale behandelt werden können, nennt man auch quasi-stetig (z.B. Einkommen)

(*) Stetige Merkmale können durch Klassenbildung in diskrete Merkmal umgewandelt werden.

Skalenniveaus

Skalenniveau	Beispiele	Erlaubte Transformationen um Strukturen zu erhalten	natürliche Ordnung	sinnvolle Abstände	natürliche Null	natürliche Einheit	Berechenbare Kennzahlen
Nominalskala	Wohnort, Farbe	Bijektionen	✗	✗	✗	✗	Mode
Ordinal-Rangskala	Noten, Michelin-Sterne, Platzierung bei Sportevent	str. monoton steig. Abb.	✓	✗	✗	✗	Median
Intervallskala	Temperatur in C°, Jahreszahlen	affin lin. str. mon. steig. Abb.	✓	✓	✗	✗	Arithm. Mittel
Verhältnisskala	Preis, Länge, Gewicht, Temp. in K°	lineare str. mon. steig. Abb.	✓	✓	✓	✗	Geom. Mittel, Harm. Mittel
Absolutskala	Häufigkeit, Anzahl, Prozentpunkte	Identität	✓	✓	✓	✓	Alle

Datenerhebung

Methoden:

Beobachtung
Datengewinnung durch Erfassen von ungesteuerten Sachverhalten

Befragung
Fragebögen für mündliche / schriftliche / online Umfrage.

Experiment
Erzeugung der Daten durch Simulation von Situationen.

Umfang:

Vollerhebung
Alle stat. Einheiten einer GG werden untersucht.

Stichprobe (Teilerhebung)
Ein Teil der UE in einer GG wird untersucht.

Datenform:

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.

- Kohortenstudien in Medizin
- Mikrozensus

Wahrscheinlichkeitsrechnung

Elementarereignisse

Für eine Grundmenge Ω wird die ein-elementige Teilmenge $\{\omega\} \subseteq \Omega$ als Elementarereignis bezeichnet.

Ereignisse

Für eine Grundmenge Ω wird $A \subseteq \Omega$ als Ereignis bezeichnet.

Laplace - Wahrscheinlichkeit

Für eine abzählbare Grundmenge Ω und ein Ereignis $A \subseteq \Omega$ ist die Laplace-Wahrscheinlichkeit $\mathbb{P}(A) := \frac{|A|}{|\Omega|}$

Wahrscheinlichkeitsverteilung (Axiome von Kolmogorov) (vereinfacht)

Sei Ω eine Grundmenge und P ein Funktion auf $\mathcal{P}(\Omega)$. P heißt Wahrscheinlichkeitsverteilung oder Wahrscheinlichkeitsmaß auf Ω, wenn sie folgende Eigenschaft erfüllt: 1) $P(\Omega) = 1$ 2) $\forall A \subseteq \Omega : P(A) \geq 0$ 3) $\forall A, B \subseteq \Omega : A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Folgerungen

Korollar

- $P(\emptyset) = 0$
- $P(\bar{A}) = 1 - P(A)$
- $B \subseteq A \Rightarrow P(A \setminus B) = P(A) - P(B)$
- $B \subseteq A \Rightarrow P(B) \leq P(A)$
- $P\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{k=1}^{n} (-1)^{k+1} \cdot \sum_{1 \leq i_1 < \dots < i_k \leq n} P[A_{i_1} \cap \dots \cap A_{i_k}]$ ← Siebformel von Sylvester-Poincaré

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

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

Bedingte Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit von A gegeben B für Ereignisse $A, B \subseteq \Omega$ mit $P(B) > 0$ ist $P(A|B) := \frac{P(A \cap B)}{P(B)}$

Satz

- $P\left[\overset{\neq \emptyset}{\bigcap_{i=1}^{n} A_i}\right] = \prod_{i=1}^{n} P\left[A_i \,\middle|\, \bigcap_{j=1}^{i-1} A_j\right] = P[A_1] \cdot P[A_2 | A_1] \cdot P[A_3 | A_1 \cap A_2] \cdot \ldots \cdot P[A_n | A_1 \cap \dots \cap A_{n-1}]$ (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
  $P[B] = \sum_{i : P[A_i] > 0} P[B | A_i] \cdot P[A_i]$ (Satz von ~~totaler~~ Wahrscheinlichkeit)

! Spezialfall: $P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)$

Kombinatorik	mit Wiederholung / mit Zurücklegen	ohne Wiederholung / ohne Zurücklegen
Anzahl Kombinationen ohne Reihenfolge	$\binom{n}{m}$	$\binom{n+m-1}{m}$
Anzahl Kombinationen mit Reihenfolge	$\frac{n!}{(n-m)!}$	n^m
Anzahl Permutationen	$n!$	$\frac{n!}{n_1! \cdot \ldots \cdot n_s!}$

Stochastische Unabhängigkeit

Eine Kollektion von Ereignissen $(A_i)_{i \in I}$ heisst (stochastisch) unabhängig, wenn für jede endliche Kollektion $J \subseteq I$ gilt: $P\left[\bigcap_{i \in J} A_i\right] = \prod_{i \in J} P[A_i]$

- $A \perp B \Leftrightarrow A, B$ stochastisch unabhängig
- Paarweise Unabhängig $\not\Rightarrow$ Unabhängigkeit
- $A \perp B \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$

Satz

Die Ereignisse $(A_i)_{i \in I}$ seien unabhängig. Für jedes i sei $B_i = A_i \vee B_i = \bar{A}_i$. Dann sind die Ereignisse $(B_i)_{i \in I}$ unabhängig.

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 $P[A_i | B] = \frac{P[B | A_i] \cdot P[A_i]}{\sum_j P[B | A_j] \cdot P[A_j]}$

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

Wettverhältnis (Odds-Update)

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

Zufallsvariablen, Verteilungen & Häufigkeiten

Zufallsvariable

Ein Zufallsvariable X ist eine Abb. $X: \Omega \to \mathbb{R}$.

$T := X(\Omega)$ nennen wir Träger von X.

Wahrscheinlichkeitsfunktion einer diskreten ZV

$f(x_i) = P(X = x_i) = P(\{\omega \in \Omega \mid X(\omega) = x_i\}) \quad \forall x_i \in \mathbb{R}$.

f heisst auch **Wahrscheinlichkeitsdichte**.

Verteilung von X

$$P(X \in B) := P(X^{-1}(B)) = P(\{\omega \in \Omega \mid X(\omega) \in B\})$$

(diskret)

X ist eine diskrete ZV, wenn der Träger von X eine abzählbare Menge ist.

(stetig)

X ist eine stetige ZV, wenn es eine Funktion f gibt, für die gilt

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- $\forall b \in \mathbb{R}: F(b) = \int_{-\infty}^{b} f(x)\,dx$

Indikatorfunktion

Die Indikatorfunktion dient dazu zu checken ob ein Element in einer bestimmten Menge enthalten ist.

Sei $A \subseteq \mathbb{R}$.

$$I_A(x) = \begin{cases} 0, & \text{wenn } x \notin A \\ 1, & \text{wenn } x \in A \end{cases}$$

Andere Schreibweisen sind $\mathbb{1}_A$, $\mathbb{I}_A$, $I(x \in A)$

Verteilungsfunktion (diskret)

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

Verteilungsfunktion (stetig)

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

! Stringentere Definition später in Wahrscheinlichkeitstheorie (2. Semester)

Satz:

Jede Verteilungsfunktion F erfüllt folgende Eigenschaften:

i) Monotonie: $a \leq b \Rightarrow F(a) \leq F(b)$

ii) Rechtsstetig: $F(a) = \lim_{h \to 0^+} F(a+h) \quad \forall a \in \mathbb{R}$

iii) Normierung: $\lim_{a \to -\infty} F(a) = 0$ und $\lim_{a \to +\infty} F(a) = 1$

Umgekehrt ist jede Funktion F mit diesen drei Eigenschaften ein Verteilungsfunktion einer ZV. D.h. $\exists X: F = F_X$.

Notation & Terminologie

Bezeichnung in der Emperie	Notation/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)

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

- 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 $a_1, \dots, a_k$ für X und $b_1, \dots, b_m$ für Y. Eine $(k \times m)$-Kontingenztafel der absoluten Häufigkeiten besitzt die Form:

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

$h_{ij} = h(a_i, b_j)$ = 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 $a_1, \dots, a_k$ für X und $b_1, \dots, b_m$ für Y. Eine $(k \times m)$-Kontingenztafel der relativen Häufigkeiten besitzt die Form:

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

$f_{ij} = \frac{h_{ij}}{n}$ = 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

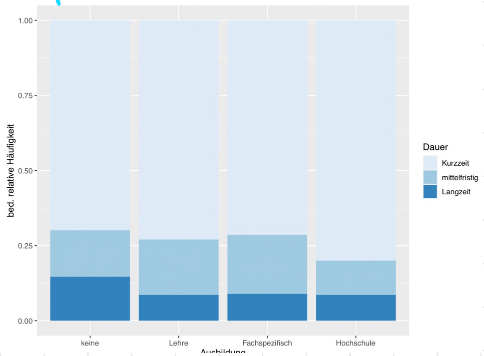
Die bedingte Häufigkeitsverteilung von Y unter der Bedingung $X = a_i$ $(Y | X = a_i)$ ist definiert als $f_Y(b_1 | a_i) = \frac{h_{i1}}{h_{i\cdot}}, \dots, f_Y(b_m | a_i) = \frac{h_{im}}{h_{i\cdot}}$

Satz

Wegen $\frac{h_{ij}}{h_{i\cdot}} = \frac{h_{ij}/n}{h_{i\cdot}/n} = \frac{f_{ij}}{f_{i\cdot}}$ gilt,

$f_Y(b_j | a_i) = \frac{f_{ij}}{f_{i\cdot}} \quad \forall i,j$

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 $\gamma(1,2 | X=1, X=2) := \frac{\gamma(1,2 | X=1)}{\gamma(1,2 | X=2)} = \frac{h_{11} \cdot h_{22}}{h_{21} \cdot h_{12}}$

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

$\gamma(Y=1, Y=2 | X=1, X=2) = \gamma(X=1, X=2 | Y=1, Y=2)$

! 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 $X=2$.

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 $\tilde{f}_{ij}$: $\tilde{f}_{ij} = f_{i\cdot} \cdot f_{\cdot j}$

χ^2-Koeffizient

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

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

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

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

χ^2 gross $\Leftrightarrow$ starker Zusammenhang

! χ^2 hängt von n, k und m ab

$\hookrightarrow$ schwer zu interpretieren.

Satz

Für eine Kontingenztafel der Form $\begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array}$ 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 $K^* := \frac{K}{\sqrt{\frac{\min\{k,m\} - 1}{\min\{k,m\}}}}$, $K^* \in [0,1]$

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 | ...

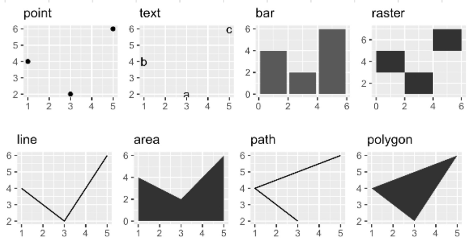
Skalen = Achsenabschnitte & Farbe & Legenden & Achsenbeschriftung & ...

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

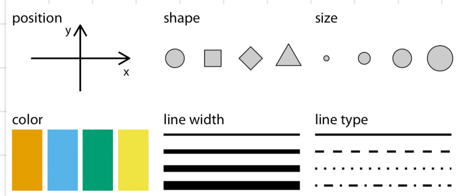
Facettierung = small multiples | lattice plot | plot | ...

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

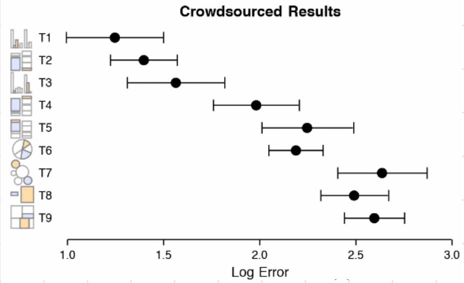
Beispiele Geometrien



Beispiele Ästhetiken



Wahrnehmung von Grafiken



Hierarchie der korrekten Interpret.:

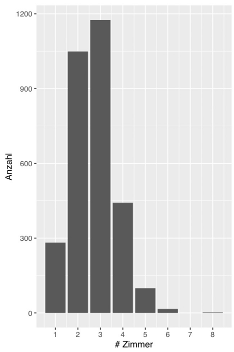
1. Position
2. Abstände / Längen
3. Steigung
4. Winkel
5. Flächen
6. Volumen
7. Farbe (Ton, Helligkeit, Sättigung)

Goldene Regel für Grafikgestaltung

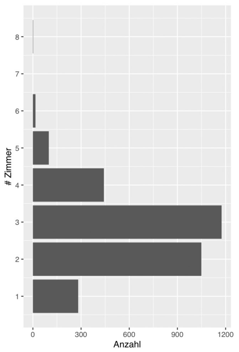
- Kommunikationsabsicht klarmachen
- Lesbarkeit maximieren

Visualisierung von Häufigkeiten & Verteilungen diskreter Merkmale

Säulendiagramm

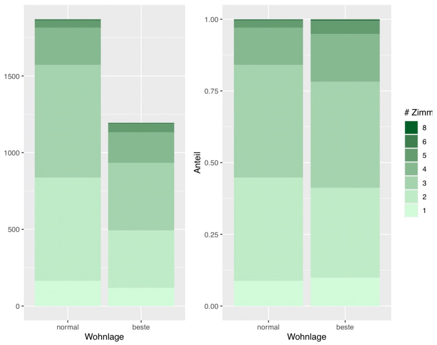


Balkendiagramm



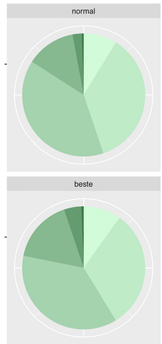
Für ordinale Merkmale, metrische Merkmale mit wenig Ausprägung und nominale Merkmale, wobei die Anordnung beliebig ist. ! Breite ist beliebig.

Stapeldiagramm (absolut & relativ)



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

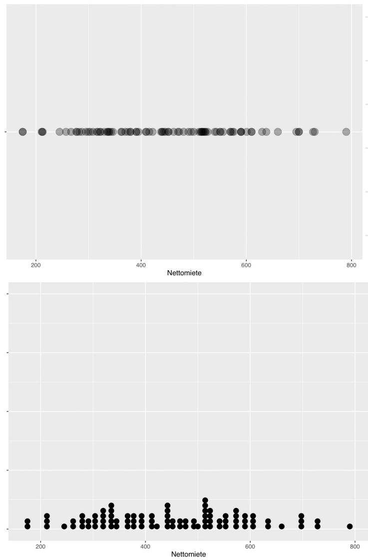
Kreisdiagramm



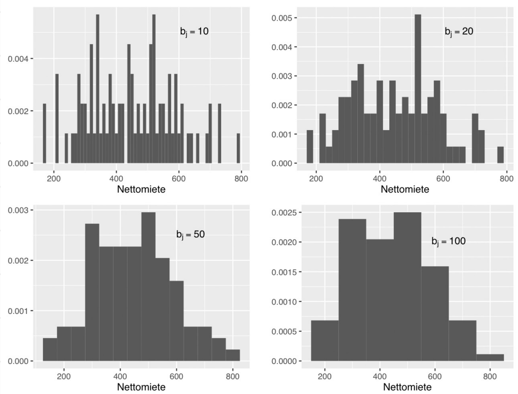
Grundsätzlich schwieriger zu interpretieren als Längendiagramme, aber enthält keine klare Ordnung.

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 $\underbrace{\text{Breite der Klasse } i}_{}$ ist.

Meist uniform

Daher nur bei metrischen Daten

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

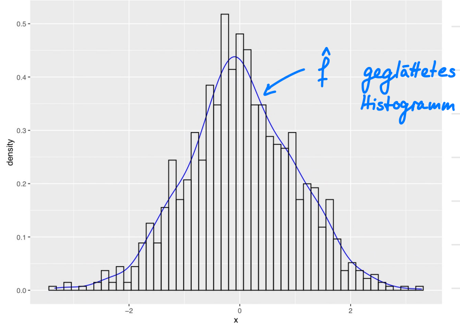
Kerndichteschätzung

Kerndichteschätzer

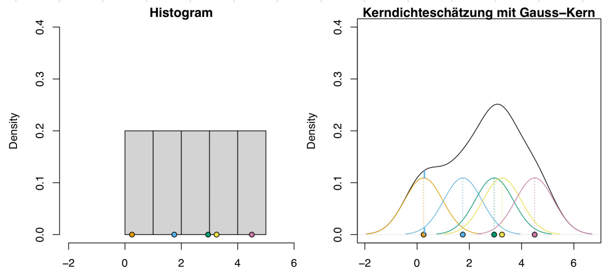
Sei K eine Kernfunktion, d.h. $\forall u: K(u) \geq 0$ und $\int_{-\infty}^{\infty} K(u)\,du = 1$.

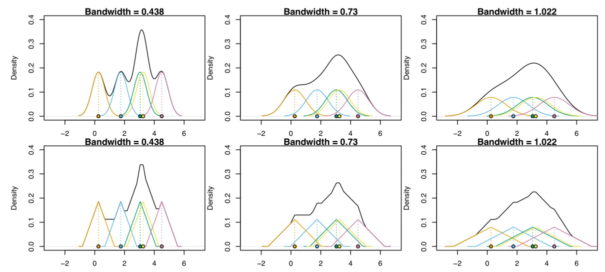
Dann ist der Kerndichteschätzer (KDE: kernel density estimator) definiert als

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



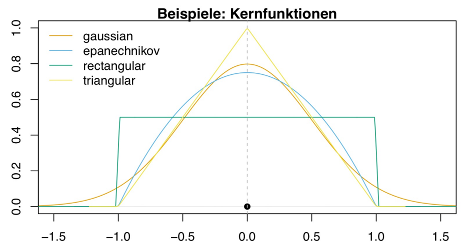
$\hat{f}$ geglättetes Histogramm





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: $K(u) := \max\{0, 1 - |u|\}$



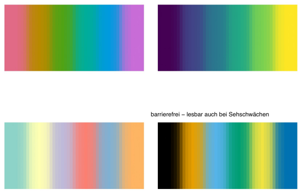
KDE > Histogramm bei größeren Datenmengen oder (quasi-)stetigen Merkmalen.

Vorteil:

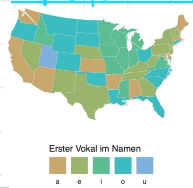
Kerndichteschätzungen berücksichtigen Entfernung der benachbarten Punkte mit abnehmender Gewichtung über Distanz.

Nachteil:

Abhängigkeit von Bandbreite h → Wird aus den Daten bestimmt.



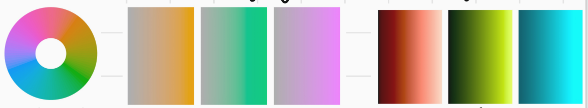
Beispiel qualitative Farbskala



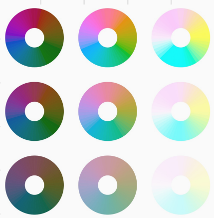
Farbskalen

Farbraum

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

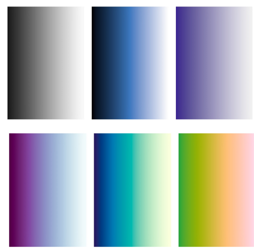


Farbraum



Farbskalentypen

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

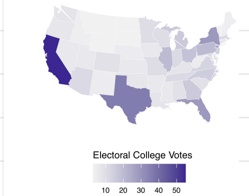


Farbe konstant, Sättigung & Helligkeit variiert.

Farbe, Helligkeit und Sättigung variiert.

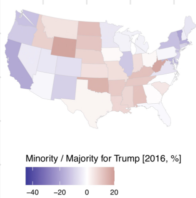
2 sequentielle Farbskalen mit je konstantem Farbton kombiniert.

Beispiel sequentielle Farbskala

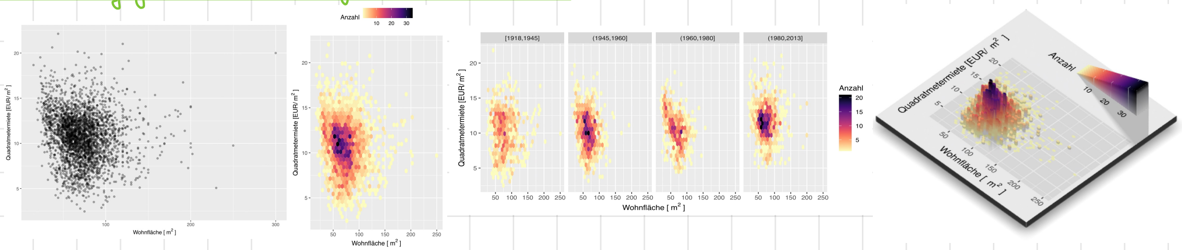


Sinnvoll, aber evtl. kleine Unterschiede verwischt

Beispiel divergente Farbskala



Visualisierung gemeinsamer Verteilungen metrischer Merkmale



Kennwerte & Verteilungeigenschaften

Lagemaßzahlen

> Wo liegt die Masse, Mitte und Mehrzahl der Daten?
> Welche Merkmalsausprägung ist typisch für die Verteilung?

Lagemaße

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.

Modus (Theorie)

Für ein ZV X ist der Modus definiert als $x_{mod} \in \mathbb{R}$ s.d. $f(x_{mod}) \geq f(x) \quad \forall x \in T_X$.

x_{mod} ist die Maximumstelle von f_X

(nicht zwingend eindeutig oder existent).

Jensen - Ungleichung

Für eine ZV X mit endlichem Erwartungswert und $g: \mathbb{R} \to \mathbb{R}$ konvex gilt $E[g(X)] \geq g(E[X])$

Markov - Ungleichung

Sei g eine nicht negative, monoton wachsende Funktion auf $\mathbb{R}$. Dann gilt $\forall c \in \mathbb{R}$ mit $g(c) > 0$

$$P[X \geq c] \leq \frac{E[g(X)]}{g(c)}$$

Der Mittelwert (Emperie)

Arithmetisches Mittel: $\bar{x}_{arith} := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{n} \cdot \sum_{i=1}^{k} h_i \cdot a_i = \sum_{i=1}^{k} f_i \cdot a_i$ (k: 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 $r = \max\{\hat{r} \in \mathbb{Z} \mid \hat{r} \leq n \cdot \alpha\}$.

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

Mittelwert / Erwartungswert (Theorie)

Diskrete ZV X: $E(X) := \sum_{\omega \in \Omega} P(\omega) \cdot X(\omega) = \sum x \cdot P(X = x) = \sum x \cdot f(x)$

Stetige ZV X: $E(X) := \int_{\mathbb{R}} x \cdot f(x) \, dx$

Satz

- $E(a + bX) = a + b \cdot E(X) \quad \forall a, b \in \mathbb{R}$
- $E(X + Y) = E(X) + E(Y)$ für alle ZV X, Y.
- $\exists c \in \mathbb{R} \; \forall x \in T_X : f(c - x) = f(c + x) \Rightarrow E(X) = c$.
- Für $g: \mathbb{R} \to \mathbb{R}$ und $Y = g(X)$, dann gilt $E(Y) = \begin{cases} \sum_{x \in T_X} g(x) f(x) & , X \text{ diskret} \\ \int_{\mathbb{R}} g(x) f(x) \, dx & , X \text{ stetig} \end{cases}$
- Ist $T_X \subseteq \mathbb{N}^+$, so gilt $E(X) = \sum_{k=1}^{\infty} P(X \geq k)$

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 $k = n \cdot p \in \mathbb{N}_0$)

Quantil (Theorie)

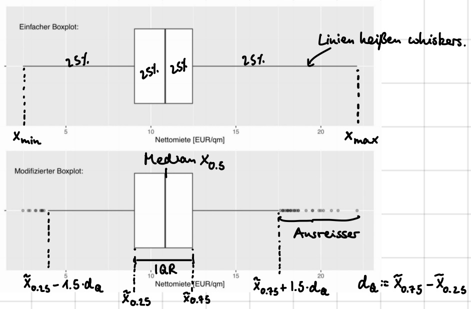
Das p-Quantil x_p einer ZV X mit $p \in [0,1]$ ist definiert als $x_p := \inf\{x \in \mathbb{R} \mid F_X(x) \geq p\}$

(Wenn $f(x) > 0 \; \forall x$ bzw. F_X str. mon. steigend, dann gilt $x_p = F_X^{-1}(p)$)

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

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

- 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:

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

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

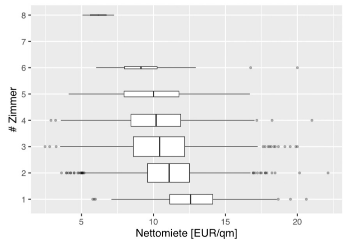
Schichtvariant: $\tilde{s}_{X_1}^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2$; $\tilde{s}_{X_2}^2 = \frac{1}{n_2} \cdot \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2)^2$

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

Gesamtstreuung = Streuung innerhalb der Schichten + Streuung zwischen den Schichten

Streuungszerlegung der Netto-Quadratmetermiete bezüglich Zimmerzahl:

Zimmer	n_j	$\bar{x}_j$	$\tilde{S}_{xj}^2$
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: $\tilde{S}_x^2 \approx 7.15$
Innerhalb: $\frac{1}{n} \sum_{j=1}^{r} n_j \tilde{S}_{xj}^2 = 6.7$
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 $MAD_X \leq \tilde{s}_X$

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

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

Varianz einer ZV

Die Varianz $Var(X)$ einer ZV X ist definiert als: ! $E(|X|) < \infty$

$$Var(X) := E[(X - E(X))^2] = \begin{cases} \sum_{x \in T_X} (x - E(X))^2 P(X = x) & , X \text{ diskret} \\ \int_{\mathbb{R}} (x - E(X))^2 f_X(x) \, dx & , X \text{ stetig} \end{cases}$$

$= E(X^2) - E(X)^2$

Beispiel (Pareto Verteilung)

$$f_X(x) = \begin{cases} \alpha \cdot \frac{x_0^\alpha}{x^{\alpha+1}} & , x > x_0 \\ 0 & , x < x_0 \end{cases}$$

finite $E(X)$, but infinite $Var(X)$ for $\alpha \in (1, 2]$

$$E(X) = \begin{cases} \infty & , \alpha \leq 1 \\ \frac{\alpha \cdot x_0}{\alpha - 1} & , \alpha > 1 \end{cases} \qquad Var(X) = \begin{cases} \infty & , \alpha \leq 2 \\ \frac{\alpha x_0^2}{(\alpha-1)^2 \cdot (\alpha - 2)} & , \alpha > 2 \end{cases}$$

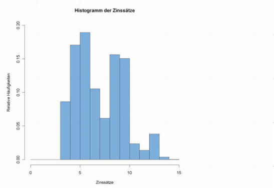
Chebyshev - Ungleichung

Für $Y = |X - E[X]|$ und $g(x) = (\max\{x, 0\})^2$ folgt durch die Markov Ungleichung

$$P[|X - E[X]| > c] \leq \frac{Var[X]}{c^2}$$

Verteilungseigenschaften

unimodal = eingipflig, multimodal = mehrgipflig



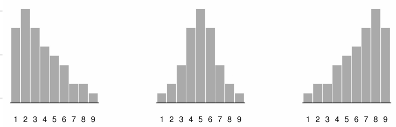
Das Histogramm der Zinssätze zeigt eine bimodale (trimodale...?) Verteilung.

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

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

$g_{0.25}$ nennt man auch Quartilskoeffizient.

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

(Zentrierter) Moment einer ZV

Für $p \in \mathbb{N}$ heisst

- p-tes Moment : $E[X^p]$
- p-tes absolutes Moment : $E[|X|^p]$
- p-tes zentriertes Moment : $E[(X - E[X])^p]$

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

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

wenn $E(X)<\infty$

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)

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

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.

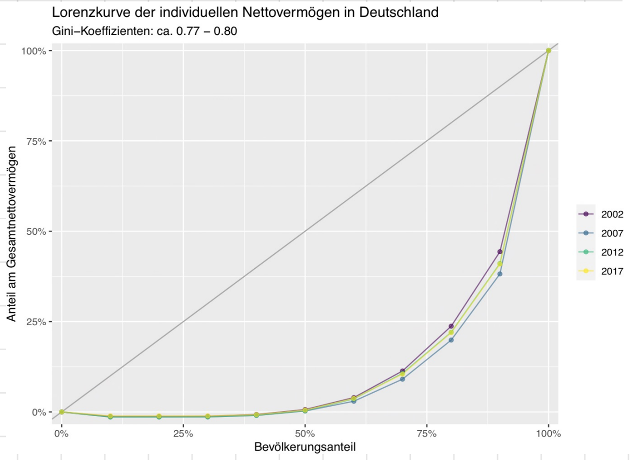
Berechnung

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

$u_{(j)} := \frac{j}{n}$ (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

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

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

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

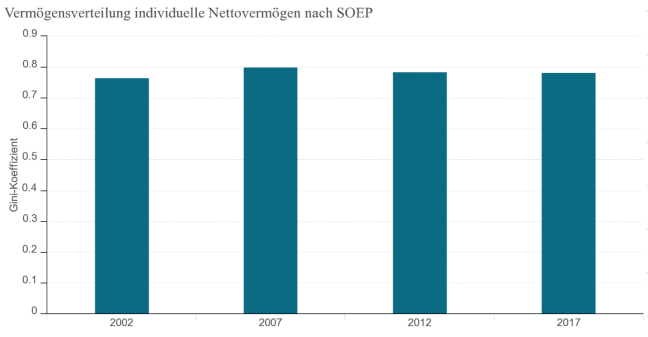
oder $G = 1 - \frac{1}{n}\sum_{i=1}^{n}(v_{(i-1)} + v_{(i)})$

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

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

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

Lorenzkurve muss

- stetig
- monoton steigend
- unter oder auf der Winkelhalbierenden
- konvex
- muss eine Funktion

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

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

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 $E(X) = \alpha/\beta > 0$ und $E(\ln(x)) = \frac{d}{dx} \ln(\Gamma(x))\big|_{x=\alpha} - \ln(\beta)$

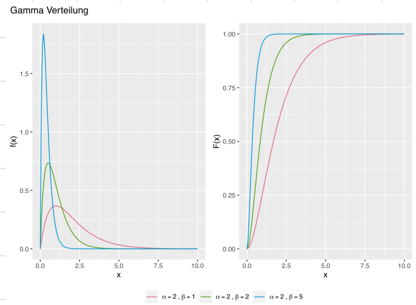
X ist Gamma-verteilt $X \sim \mathcal{G}(\alpha,\beta) \sim \Gamma(\alpha,\beta)$, wenn

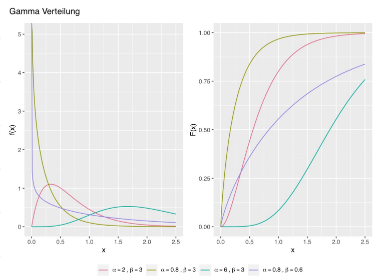
$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \cdot e^{-\beta x} \cdot \mathbb{1}_{(0,\infty)}(x)$. Γ ist die Gammafunktion:

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

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

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





Betaverteilung

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

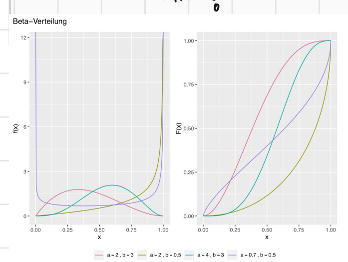
X ist Beta-verteilt $X \sim Be(\alpha,\beta)$, wenn

$f_{\alpha,\beta}(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1} \cdot (1-x)^{\beta-1}$, 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 → numerische Integration.

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

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 $\bar{X} = \frac{1}{n}\cdot\sum_{i=1}^n X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

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

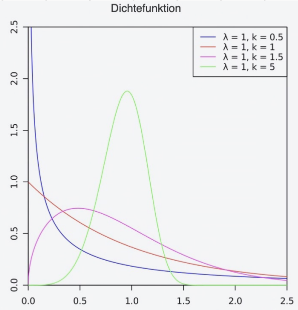
Beschreibung von Lebensdauer und Ausfallhäufigkeit von (mechanischen) Objekten.

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

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

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 $Y = X^2$.

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

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

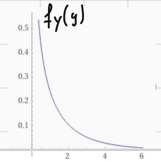
$g_1^{-1}(y) = -\sqrt{y}$ $\qquad g_2^{-1}(y) = \sqrt{y}$

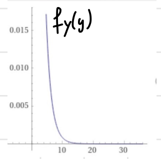
$$f_Y(y) = f_X(-\sqrt{y})\cdot\left|\frac{1}{-2\sqrt{y}}\right| + f_X(\sqrt{y})\cdot\left|\frac{1}{2\sqrt{y}}\right| = \frac{1}{\sqrt{2\pi\cdot y}}\cdot\exp\left(-\frac{1}{2}\cdot y\right), \text{ für } y>0$$

$= 0$, 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 $F_X(x)$.

Schätzung & Grenzwertesätze

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Wir betrachten die Summen $S_n = X_1 + \dots + X_n$ 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.

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

Schwaches Gesetz der grossen Zahlen

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von Zufallsvariablen, die Lebesgue integrierbar sind und mit Erwartungswert $E(X_i) = \mu \;\forall i \in \mathbb{N}$.
Wir sagen, dass das schwache Gesetz der grossen Zahlen gilt, wenn

$$\forall \varepsilon > 0: P[|\bar{X}_n - \mu| > \varepsilon] \to 0 \text{ für } n \to \infty \quad \text{bzw.} \quad \forall \varepsilon > 0 . \lim_{n\to\infty} P[|\bar{X}_n - \mu| < \varepsilon] = 1$$

Idee: Konvergenz in Wahrscheinlichkeit $(\bar{X}_n)_{n\in\mathbb{N}}$ gegen μ: Die Wahrscheinlichkeit für ein "ungewöhnliches" Ereignis, d.h. für ω mit $|\bar{X}_n(\omega) - \mu| > \varepsilon$, wird mit zunehmenden n kleiner. Aber es bleibt die Möglichkeit, dass unendlich oft $|\bar{X}_n(\omega) - \mu| > \varepsilon$ vorkommt (wenn auch unregelmäßig).

Starkes Gesetz der grossen Zahlen

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von Zufallsvariablen, die Lebesgue integrierbar sind und mit Erwartungswert $E(X_i) = \mu \;\forall i \in \mathbb{N}$.
Wir sagen, dass das starke Gesetz der grossen Zahlen gilt, wenn

$$P\left[\lim_{n\to\infty} \bar{X}_n = \mu\right] = 1 \quad \text{bzw.} \quad P\left[\{\omega \in \Omega \mid \lim_{n\to\infty} \bar{X}_n(\omega) = \mu\}\right] = 1$$

Idee: P-sichere Konvergenz für $\bar{X}_n \to \mu$: $\bar{X}_n$ konvergiert gegen μ bis auf eine Menge von Ereignissen ω deren Wahrscheinlichkeit Null ist ($P(\{\omega \in \Omega \mid \lim_{n\to\infty} \bar{X}_n(\omega) \neq \mu\}) = 0$).
Unterschied zu normaler Konvergenz ist, dass z.B. für Würfel $\omega = (6,6,6,\dots)$ vernachlässigbar ist als Folge für $\bar{X}_n$.

Korollar

Wenn die X_i i.i.d. sind, mit $E(X_i^2) < \infty$, so gilt das schwache Gesetz der grossen Zahlen.

Beweis: $E(X_i) = \mu \;\forall i \in \mathbb{N}$, da X_i i.i.d. Daraus folgt $E(\bar{X}_n) = \mu$. $Var(\bar{X}_n) = \frac{1}{n} Var(X_i)$. $Var(X_i)$ existiert wegen $E(X_i^2) < \infty$. Chebyshev-Ungleichung $\Rightarrow \forall \varepsilon > 0: P[|\bar{X}_n - \mu| > \varepsilon] \leq \frac{Var(X_i)}{n \cdot \varepsilon^2}$. Für $n \to \infty$ folgt der Satz.

Note: X_i i.i.d. ist nicht zwingend nötig. Es reicht auch gleich-verteilt und unkorreliert statt unabhängig mit $E(X_i) = \mu \;\forall i \in \mathbb{N}$. Es gilt dann immernoch $Var(\bar{X}_n) = o(n)$.

Satz

$E(|X_i|) < \infty \Rightarrow$ Starkes Gesetz der grossen Zahlen.

Satz

P-fastsichere Konvergenz impliziert Konvergenz in Wahrscheinlichkeit.
D.h. Starkes Gesetz $\Rightarrow$ Schwaches Gesetz. ! Keine Äquivalenz.

Zentraler Grenzwertsatz

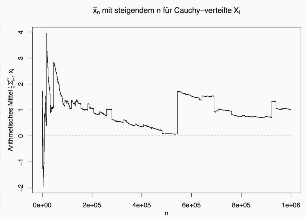
Seien $(X_i)_{i\in\mathbb{N}}$ i.i.d. Zufallsvariablen mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2 < \infty$. Dann gilt:

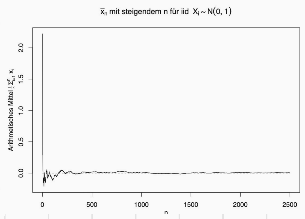
1) $\lim_{n\to\infty} P\left[\frac{S_n - n\mu}{\sqrt{n \cdot \sigma^2}} \leq x\right] = \Phi(x) \quad \forall x \in \mathbb{R}$. bzw. $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$

2) $\lim_{n\to\infty} P\left[\sqrt{n} \cdot \frac{(\bar{X} - \mu)}{\sigma} \leq x\right] = \Phi(x) \quad \forall x \in \mathbb{R}$. bzw. $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Beispiele

- Sei $X \sim Exp(1)$. Dann hat $Y \sim \sin(X) e^X \cdot \frac{1}{X}$ keinen Erwartungswert bzgl. Lebesgue-Integration. Verwendet man gewöhnliche Konvergenz und das Dirichlet Integral, so erhalten wir $E(Y) = \frac{\pi}{2}$. Daher gilt das schwache, aber nicht das starke Gesetz der grossen Zahlen
- 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 immer gleich stark $\forall n \in \mathbb{N}$. Somit ist das schwache Gesetz der grossen Zahlen nicht erfüllt.





Schlussfolgerung aus 'thinking fast and slow'

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

Zufallsvektoren & multivariate Verteilung

Gemeinsame Verteilungsfunktion

Seien $X_1, \dots, X_n$ Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.
Wir betrachten $\bar{X} = (X_1, \dots, X_n)^T$, $\mathcal{B}^n := \mathcal{B}(\mathbb{R}^n) := \sigma(\{A_1 \times \dots \times A_n \mid A_i \in \mathcal{B}\})$.
$\bar{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}^n)$. Das Bild $\mu_{\bar{X}}$ von P unter $\bar{X}$ heisst die gemeinsame Verteilung von $X_1, \dots, X_n$.
und $\mu_{\underline{X}} = P(\bar{X}^{-1}(A)) = P[\{\omega \in \Omega \mid X(\omega) \in A\}] = P[X \in A]$ für $A \in \mathcal{B}^n$.

Diskret. $\bar{X}(\Omega)$ abzählbar: $F_{\bar{X}}(A) = \sum_{x \in X(\Omega) \cap A} P[X = x] = \sum_{\substack{(x_1, \dots, x_n) \in A \\ x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)}} P[X_1 = x_1, \dots, X_n = x_n]$

Absolutstetig: $\exists f : \mathbb{R}^n \to \mathbb{R}$ messbar : $f \geq 0 \wedge \int_{\mathbb{R}^n} f(\vec{x})\, d\vec{x} = 1$ derart, dass $F_X(A) = \int_A f(\vec{x})\, d\vec{x}$

Transformationssatz

Sei $g : \mathbb{R}^n \to \mathbb{R}^n$ invertierbar und linear, d.h. $g(\vec{x}) = B\vec{x} + \vec{m}$ mit $\det(B) \neq 0$. Sei $\bar{Y} = g(\bar{X})$.
Wenn $\mu_{\bar{X}}$ absolutstetig ist, so ist auch $\mu_{\bar{Y}}$ absolutstetig und $f_{\bar{Y}}(\vec{y}) = \frac{1}{|\det(B)|} \cdot f_{\bar{X}}(B^{-1} \cdot (\vec{y} - \vec{m}))$

Beispiel Faltung

Sei $\bar{X} = (X_1, X_2)$ eine zweidimensionale absolutstetige ZV. Setze $\bar{Y} = g(\bar{X}) = (\bar{X}_1, \bar{X}_1 + \bar{X}_2)$.
Dann ist $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ bzw. $B^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. $\Rightarrow f_{\bar{Y}}(x_1, y) = \frac{1}{1} \cdot f_{\bar{X}}\left(B^{-1} \cdot \begin{pmatrix} x_1 \\ y \end{pmatrix}\right) = f_{\bar{X}}(x_1, y - x_1)$

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

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

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 $\bar{Y} = X_1 + X_2$. $T_{\bar{Y}} = \mathbb{N}_+ \setminus \{1\}$

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

Randverteilung

Sei $\bar{X} = (X_1, \dots, X_n)^T$ ein Zufallsvektor mit einer multivariaten Verteilung $P_{\bar{Z}}$. Dann heisst die Verteilung
$P_{X_i}(A) := P_{\bar{Z}}(\{\Omega_1 \times \dots \times \Omega_{i-1} \times A \times \Omega_{i+1} \times \dots \times \Omega_n\})$
die i-te Randverteilung von $\bar{Z}$.

Satz

Sei $g : \mathbb{R}^n \to \mathbb{R}$ messbar und $\bar{X}$ ein Zufallsvektor.

$$E(g(\bar{X})) = \begin{cases} \sum_{x_i \in X(\Omega)} g(x_1, \dots, x_n) \cdot P(X_1 = x_1, \dots, X_n = x_n) & \text{, diskreter Fall} \\ \int_{\mathbb{R}^n} g(\vec{x}) \cdot f_{\bar{X}}(\vec{x})\, d\vec{x} & \text{, stetiger Fall} \end{cases}$$

Multinomialverteilung

Ein Experiment mit d möglichen Ereignissen mit Wahrscheinlichkeit $\pi_1, \dots, \pi_d$ wird unabhängig voneinander n-mal wiederholt. Sei $\bar{X}$ ein d dimensionaler ZV dessen i-te Komponente angibt wie oft das i-te Ereignis eingetreten ist. Dann nennen wir
$\bar{X}$ multinomialverteilt $X \sim \mathcal{M}_d(n, \bar{\pi})$ mit $T_{\bar{X}} = \{\vec{x} \in \{0, .., n\}^d \mid \sum_{i=1}^{d} x_i = n\}$
$P_{n, \bar{\pi}}(X_1 = k_1, \dots, X_n = k_n) = \frac{n!}{k_1! \cdot \ldots \cdot k_n!} \cdot \pi_1^{k_1} \cdot \ldots \cdot \pi_n^{k_n}$

Negative Binomialverteilung

Wahrscheinlichkeit das man genau k Versuche für den n-ten Erfolg benötigt. X ist negativ binomialverteilt mit $X \sim \mathcal{NB}(n, p)$
$P_{n,\pi}(X = k) = \binom{k-1}{n-1} \cdot \pi^n \cdot (1-\pi)^{k-n} \cdot I(k \geq n)$

Beispiel

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

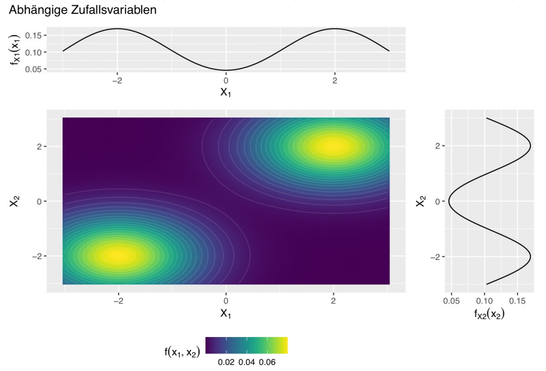
$$P(Y = n) = P(X_1 + X_2 = n) = \sum_{k=0}^{n} P(X_1 = k) \cdot P(X_2 = n - k) = \Big(\underbrace{\sum_{k=0}^{n} \binom{n}{k} \cdot \lambda_1^k \cdot \lambda_2^{n-k}}_{= (\lambda_1 + \lambda_2)^n}\Big) \cdot \frac{e^{-(\lambda_1 + \lambda_2)}}{n!}$$

Unabhängige Zufallsvariablen

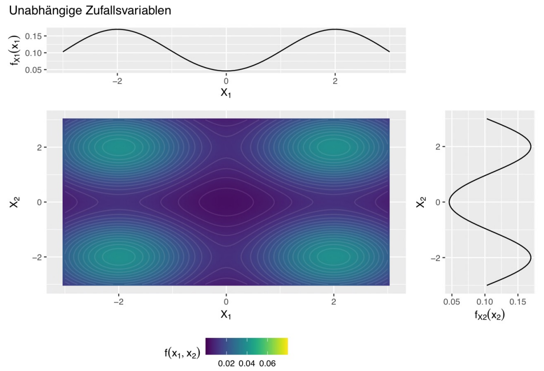
$X_1, \dots, X_n$ heissen stochastisch unabhängig, wenn
$\forall A_1, \dots, A_n \in \mathcal{B} : P[X_1 \in A_1, \dots, X_n \in A_n] = \prod_{i=1}^{n} P[X_i \in A_i]$
bzw. $F_{\bar{X}}(A_1 \times \dots \times A_n) = \prod_{i=1}^{n} F_{X_i}(A_i)$

Satz

Seien $X_1, \dots, X_n$ unabhängig. Dann ist $\mu_{\bar{X}}$ absolutstetig g.d.w. jedes μ_{X_i} absolutstetig ist. Ferner gilt dann $f_{\bar{X}}(\vec{x}) = \prod_{i=1}^{n} f_{X_i}(\vec{x})$



$f(x_1, x_2) \neq f_{X_1}(x_1) f_{X_2}(x_2)$



$f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$

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

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

Satz vom iterierten Erwartungswert

Für beliebige ZV X,Y und Funktion f gilt:

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

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

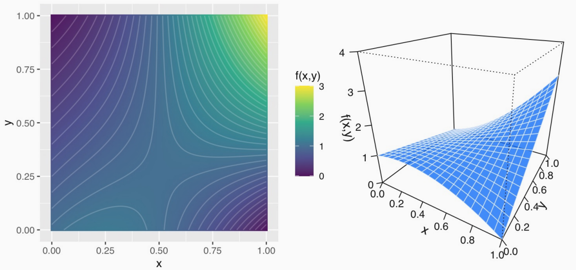
$$f_{X,Y}(x,y) = f_{X|Y}(x,y) \cdot f_Y(y) = f_{Y|X}(x,y) \cdot f_X(x)$$

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

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

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