Stochastik

GianAndrea Müller

15. Januar 2018

Inhaltsverzeichnis

1 Definitionen

- Elementarereignis ω: Möglicher Ausgang eines Zufallsexperiments.
- Grundraum Ω: Menge aller Elementarereignisse.
- Ereignis A: Kollektion gewisser Elementarereignisse.
- Disjunkte Mengen A und B $\rightarrow A \cap B = \{\}$.

1.0.1 Wahrscheinlichkeitsbaum

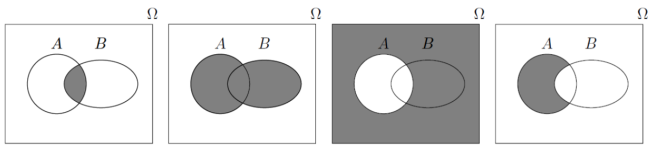
- In jeder Verzweigung ist die Summe der Wahrscheinlichkeiten 1.
- Wahrscheinlichkeiten für spezifische Kombinationen erhält man durch Multiplizieren der Wahrscheinlichkeiten eines bestimmten Pfades. Pfade des gleichen Ereignisses werden dann aufaddiert.

1.1 Mengentheorie

Name	**Symbol**	**Bedeutung**
Durchschnitt	$A \cap B$	„A **und** B“
Vereinigung	$A \cup B$	„A **oder** B“
Komplement	A^c	„**nicht** A“
Differenz	$A \backslash B = A \cap B^c$	„A **ohne** B“

1.1.1 De Morgan'sche Regeln

- $(A \cap B)^c = A^c \cup B^c$
- $(A \cup B)^c = A^c \cap B^c$



1.2 Axiome der Wahrscheinlichkeitsrechnung (Kolmogorov)

(A1) $0 \leq \mathbb{P}(A) \leq 1$

(A2) $\mathbb{P}(\Omega) = 1$

(A3) $\mathbb{P}(A \cup B) = P(A) + P(B) \qquad \forall A, B : A \cap B = \emptyset$

$\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \sum_{i \geq 1} \mathbb{P}(A_i) \qquad \forall A_k \cap A_l = \emptyset,\ k \neq l$

1.2.1 Weitere Rechenregeln

$$
\begin{aligned}
\mathbb{P}(A^c) &= 1 - \mathbb{P}(A) && \forall A \\
\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) && \forall A, B \\
\mathbb{P}(A_1 \cup \ldots \cup A_n) &\leq \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n) && \\
\mathbb{P}(B) &\leq \mathbb{P}(A) && \forall A, B :\ B \subseteq A \\
\mathbb{P}(A \backslash B) &= \mathbb{P}(A) - \mathbb{P}(B) && \forall A, B :\ B \subseteq A
\end{aligned}
$$

1.3 Modell von Laplace

$\mathbb{P}(\{\omega_k\}) = \frac{1}{|\Omega|},\ k \geq 1$

$$\boxed{\mathbb{P}(A) = \sum_{k:\omega_k \in A} \mathbb{P}(\{\omega_k\}) = \sum_{k:\omega_k \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}}$$

1.4 Unabhängigkeit von Ereignissen

Wenn man die Wahrscheinlichkeiten $\mathbb{P}(A)$ und $\mathbb{P}(B)$ kennt, so kann man $\mathbb{P}(A \cap B)$ im Allgemeinen nicht berechnen.

$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)}$ für unabhängige Ereignisse

1.5 Bedingte Wahrscheinlichkeiten

$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$ bedingte Wahrscheinlichkeit von A gegeben B

1.5.1 Rechenregeln

$$
\begin{aligned}
&0 \leq \mathbb{P}(A|B) \leq 1 && \forall A \\
&\mathbb{P}(B|B) = 1 && \\
&\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) && \forall A_1, A_2 :\ A_1 \cap A_2 = \nvDash \\
&\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B) && \forall A
\end{aligned}
$$

1.5.2 Redefinition der Unabhängigkeit

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

$$\boxed{\text{A,B unabhängig} \Leftrightarrow\ \mathbb{P}(A|B) = \mathbb{P}(A)\ \Leftrightarrow\ \mathbb{P}(B|A) = \mathbb{P}(B)}$$

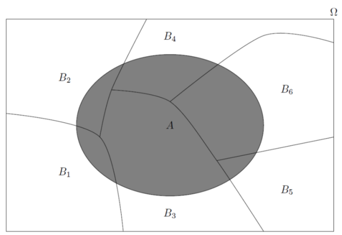
Achtung: im Allgemeinen:

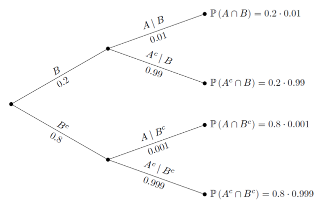
$$\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$$
$$\mathbb{P}(A|B^c) \neq 1 - \mathbb{P}(A|B)$$

1.6 Satz der totalen Wahrscheinlichkeit

$$\boxed{\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{k} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

$B_1 \cup \ldots \cup B_k = \Omega$





1.7 Satz von Bayes

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

$$\boxed{\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{l=1}^{k} \mathbb{P}(A|B_l)\mathbb{P}(B_l)}}$$

2 Wahrscheinlichkeitsverteilungen

2.1 Definitionen

Zufallsvariable	X
Wertebereich	W
realisierter Wert	x

$\boxed{\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)}$ Unabhängigkeit

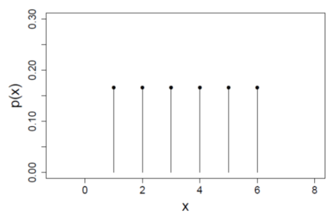
2.2 Wahrscheinlichkeitsfunktion

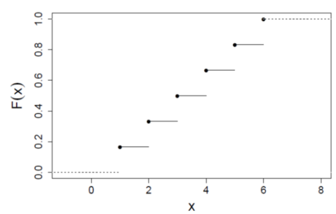
$$\boxed{p(x_K) = \mathbb{P}(X = x_k), k \geq 1}$$

Eigenschaften:

- $\sum_{k \geq 1} p(x_k) = 1$
- $\mathbb{P}(X \in A) = \sum_{k: x_k \in A} p(x_k) \qquad A \subset W$
- $$\begin{aligned} \mathbb{P}(()a < X \leq b) &= \mathbb{P}(()X \in (a, b]) \\ &= \mathbb{P}(()X \in (-\infty, b]) - \mathbb{P}(()X \in (-\infty, a]) \\ &= F(b) - F(a) \end{aligned}$$
- $\mathbb{P}(()X > x) = 1 - \mathbb{P}(()X \leq x) = 1 - F(x)$
- F ist monoton steigend
- $\lim_{x \to -\infty} F(x) = 0$ und $\lim_{x \to \infty} F(x) = 1$
- F ist rechts-stetig $\rightarrow \lim_{x \searrow a} F(x) = F(a)$

$\boxed{F(x) = \mathbb{P}(()X \leq x)}$ kumulative Wahrscheinlichkeitsfunktion





2.2.1 Kennzahlen

$$\boxed{\mu_x = \mathbb{E}[X] = \sum_{k \geq 1} x_k p(x_k)} \text{ Erwartungswert}$$

für eine beliebige Transformation $Y = g(X),\ g : \mathbb{R} \to \mathbb{R}$

$$\boxed{\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_{k \geq 1} g(x_k) p(x_k)}$$

$$\boxed{Var(X) = \sigma_x^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{k \geq 1} (x_k - \mu_X)^2 p(x_k)} \text{ Varianz}$$

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

Rechenregeln:

$$\begin{aligned}
\mathbb{E}[a + bX] &= a + b \cdot \mathbb{E}[X],\ a, b \in \mathbb{R} \\
Var(X) &= \mathbb{E}[X^2] - E[X]^2 \\
Var(a + bX) &= b^2 Var(X),\ a, b \in \mathbb{R} \\
Var(a) &= 0,\ a \in \mathbb{R} \\
\mathbb{E}[a + bX + cY] &= a + b \cdot \mathbb{E}[X] + c \cdot \mathbb{E}[Y],\ a, b, c \in \mathbb{R}
\end{aligned}$$

2.2.2 Bernoulliverteilung [Bernoulli(p)]

$$\boxed{X = \begin{cases} 1 & \text{Wahrscheinlichkeit p} \\ 0 & \text{Wahrscheinlichkeit 1-p} \end{cases}}$$

Es gilt: $\mathbb{E}[X] = p \qquad Var(X) = p \cdot (1 - p)$

$T_1, T_2 \sim \text{Bern(p)}$, unabhängig $\to \min(T_1, T_2) \sim \text{Bern}(p^2)$

2.2.3 Binomialverteilung [Bin(n,p)]

Verteilung der Anzahl Erfolge bei n (unabhängigen) Wiederholungen eines Experiments. $W = \{0, 1, \ldots, n\}$

$$\boxed{p(x) = \binom{n}{x} p^x (1 - p)^{n - x},\ x \in W}$$

$$\mathbb{E}[X] = np \qquad Var(X) = n \cdot p \cdot (1 - p)$$

$$\boxed{\binom{n}{x} = \frac{n!}{x!(n - x)!}}$$

$$\boxed{\mathbb{P}(X \leq x) \approx \Phi\left(\frac{x - np}{\sqrt{np(1 - p)}}\right)} \text{ Normalapproximation}$$

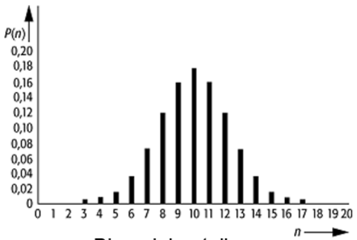
Beispiel

Wie gross ist die Wahrscheinlichkeit, dass bei 1000 Münzwürfen maximal 530 Mal Kopf erscheint?

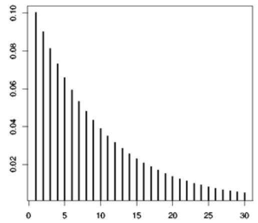
$$X \sim \text{Bin}(1000, 0.5) \to X \sim \mathcal{N}(\mu, \sigma^2)$$

$$\mu = 1000 \cdot 0.5 \qquad \sigma^2 = 1000 \cdot 0.5 \cdot (1 - 0.5) = 250$$

$$\boxed{\mathbb{P}(X \leq 530) = \Phi\left(\frac{530 - 500}{\sqrt{250}}\right) = \Phi(1.897) \approx 0.97}$$



Binomialverteilung



geometrische Verteilung

2.2.4 Geometrische Verteilung [Geom(p)]

Anzahl Wiederholungen von unabhängigen Bernoulli(p) Experimenten bis zum ersten Erfolg.

$$\boxed{p(x) = p \cdot (1 - p)^{x - 1}}$$

$$\mathbb{E}[X] = \frac{1}{p} \qquad Var(X) = \frac{1 - p}{p^2} \qquad F(n) = 1 - (1 - p)^n$$

2.2.5 Poissonverteilung [Pois(λ)]

$$\boxed{p(x) = e^{-\lambda} \frac{\lambda^x}{x!},\ x \in W}$$

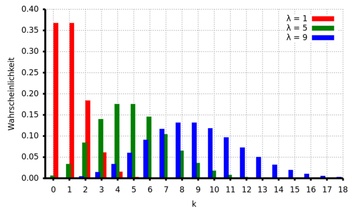
$$\mathbb{E}[X] = \lambda \qquad Var(X) = \lambda \qquad \mathbb{P}(X \geq 1 | X \leq 1) = \frac{\lambda}{\lambda + 1}$$

Für grosse n und ein kleines p mit $np = \lambda$ nähert sich die Poissonverteilung der Binominalverteilung an.

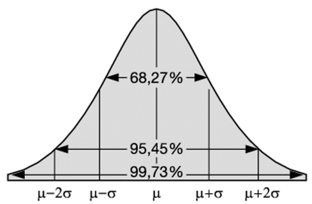
$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1 - p)^{n - x} \approx e^{-\lambda} \frac{\lambda^x}{x!}$$

$X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2)$ mit X und Y unabhängig, dann gilt:

$$\boxed{X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)}$$



Poissonverteilung



Normalverteilung

3 Stetige Verteilungen

3.1 Wahrscheinlichkeitsdichte

$$f(x) = \lim_{h \to 0} \frac{\mathbb{P}(x < X \leq x + h)}{h} = \lim_{h \to 0} \frac{F(x + h) - F(x)}{h} = F'(x) = f(x)$$

$\mathbb{P}(x < X \leq x + h) \approx hf(x) \qquad \mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$

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

Bestimmen der Verteilungsfunktion

$F(t)$ Verteilungsfunktion für $Y = g(X)$

$$\boxed{F(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(x \leq g^{-1}(t)) = \int_0^{g^{-1}(t)} X dx}$$

$F(u)$ Verteilungsfunktion für $U = X + Y$

$$\boxed{F(u) = \mathbb{P}(U \leq u) = \mathbb{P}(X + Y \leq u) = \int_0^u \int_0^{u - y} f_X(x) dx f_Y(y) dy}$$

3.2 Kennzahlen

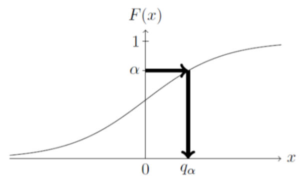
$$\boxed{\mathbb{E}[X] = \mu_X = \int_{-\infty}^{\infty} x f(x) dx}$$

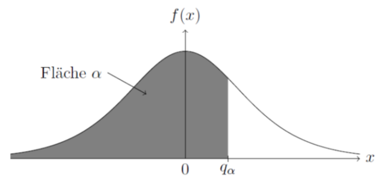
$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

$$\boxed{\text{Var}(X) = \sigma_X^2 = \mathbb{E}[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx}$$

$$\boxed{\alpha = \mathbb{P}(X \leq q_\alpha) = F(q_\alpha) \qquad q_\alpha = F^{-1}(\alpha)} \ (\alpha \times 100)\text{%-Quantil}$$

Der Median ist das 50%-Quantil.





3.3 Uniforme Verteilung [Uni(a,b)]

$$\boxed{\begin{aligned}
f(x) &= \begin{cases} \frac{1}{b - a} & a \leq x \leq b \\ 0 & else \end{cases} \\
F(x) &= \begin{cases} 0 & x < a \\ \frac{x - a}{b - a} & a \leq x \leq b \\ 1 & x > b \end{cases}
\end{aligned}}$$

$W = [a, b]$

$$\begin{aligned}
\mathbb{E}[X] &= \frac{a + b}{2} \\
\text{Var}(X) &= \frac{(b - a)^2}{12}
\end{aligned}$$

3.4 Normalverteilung $[\mathcal{N}(\mu,\sigma^2)]$

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

$W = \mathbb{R}$

$$\begin{aligned}\mathbb{E}[X] &= \mu\\ \mathrm{Var}(X) &= \sigma^2\end{aligned}$$

- Dichte der Normalverteilung ist symmetrisch um den Erwartungswert μ.
- Je grösser σ, desto flacher oder breiter wird die Dichte.
- Fläche über dem Intervall $[\mu-\sigma,\mu+\sigma]$ ist ungefähr 2/3.
- Fläche über dem Intervall $[\mu-2\sigma,\mu+2\sigma]$ ist ungefähr 0.95.
- Wahrscheinlichkeit weniger als eine Standardabweichung vom Erwartungswert zu liegen: ca. 66%.

3.4.1 Standardnormalverteilung $[\mathcal{N}(0,1)]$

$$\boxed{\begin{aligned}\varphi(x) &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}\\ \Phi(x) &= \int_{-\infty}^{x}\varphi(u)du\end{aligned}}$$

$$\begin{aligned}z_\alpha &= \Phi^{-1}(\alpha),\ \alpha \in (0,1)\\ F(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right)\end{aligned}$$

Die Verteilungsfunktion $\mathcal{N}(\mu,\sigma^2)$ kann als Transformation von Φ berechnet werden.

3.5 Lognormalverteilung

$$X \sim \mathcal{N}(\mu,\sigma^2) \qquad Y = e^X$$

Eine Zufallsvariable Y heisst lognormalverteilt wenn ihr Logarithmus normalverteilt ist.

$$\boxed{f_Y(y) = \begin{cases}0 & \leq 0\\ \frac{1}{\sqrt{2\pi}\sigma y}\exp\left\{-\frac{1}{2}\left(\frac{\log(y)-\mu}{\sigma}\right)^2\right\} & y>0\end{cases}}$$

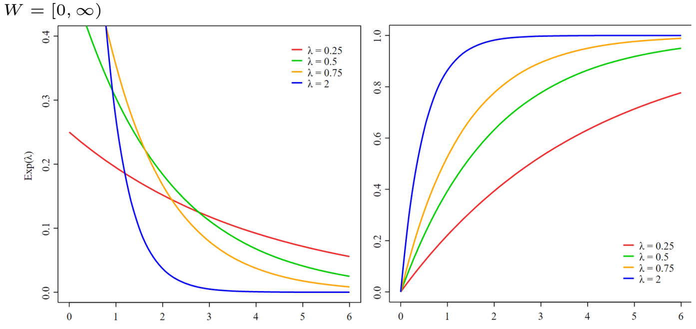
3.6 Exponentialverteilung $[\exp(\lambda)]$

Die Exponentialverteilung ist das einfachste Modell für Wartezeiten auf Ausfälle und eine stetige Version der geometrischen Verteilung.

$$\boxed{\begin{aligned}f(x) &= \begin{cases}0 & x<0\\ \lambda e^{-\lambda x} & x \geq 0\end{cases}\\ F(x) &= \begin{cases}0 & xy0\\ 1-e^{-\lambda x} & x \geq 0\end{cases}\end{aligned}}$$

$$\boxed{\begin{aligned}\mathbb{E}[X] &= 1/\lambda\\ \mathrm{Var}(X) &= 1/\lambda^2\end{aligned}}$$

$W = [0,\infty)$



4 Transformationen

4.1 Linear

$g(x) = a + bx,\ b > 0$

$$\begin{aligned}F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(a+bX \leq y)\\ &= \mathbb{P}(X \leq \tfrac{y-a}{b})\\ &= F_X(\tfrac{y-a}{b})\end{aligned}$$

$b < 0$

$$F_Y(y) = 1 - F_X(\tfrac{y-a}{b})$$

$$\boxed{f_Y(y) = \frac{1}{|b|}f_X\left(\frac{y-a}{b}\right)}$$

4.1.1 Standardisierung

Eine Zufallsvariable kann immer so transformiert werden, dass sie Erwartungswert 0 und Varianz 1 hat.

$$\boxed{g(x) = \frac{x-\mu_X}{\sigma_X}}$$

$$\begin{aligned}\mathbb{E}[Z] &= 0\\ \mathrm{Var}(Z) &= 1\end{aligned}$$

Beispiel:

$$\begin{aligned}\mathbb{P}(X \leq 3) &= \mathbb{P}(\tfrac{X-\mu}{\sigma} \leq \tfrac{3-\mu}{\sigma})\\ &= \mathbb{P}(Z \leq \tfrac{3-\mu}{\sigma})\\ &= \Phi\left(\frac{3-\mu}{\sigma}\right)\end{aligned}$$

4.1.2 Allgemeiner monotoner Fall

g eine beliebige differenzierbare, streng monotone Funktion so hat $Y = g(X)$ die Dichte:

$$\boxed{f_Y(y) = \left|\frac{1}{g'(g^{-1}(y))}\right| f_X(g^{-1}(y)),\ y \in W_Y}$$

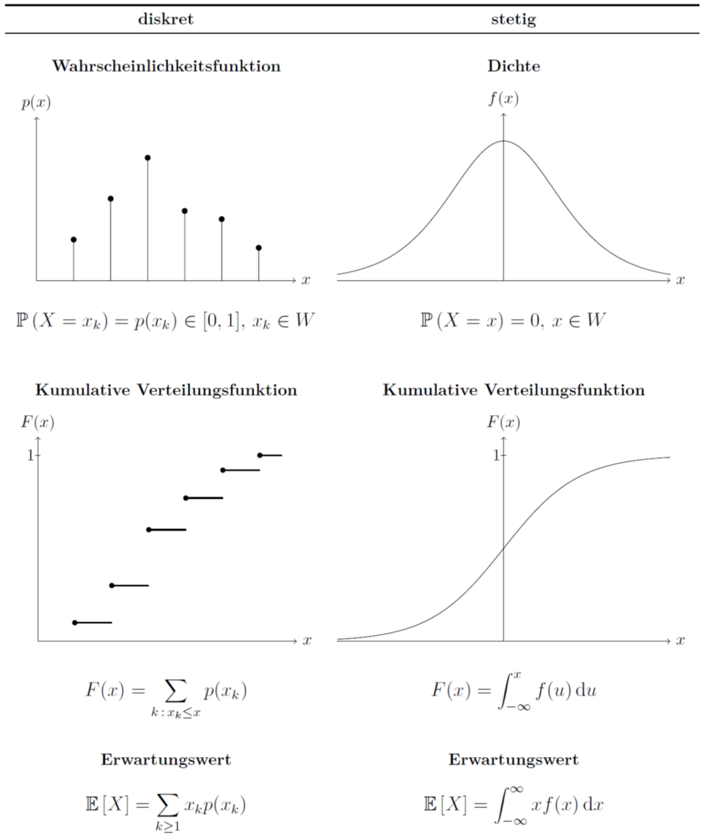
$$W_Y = g(W_X) = \{g(x), x \in W_X\}$$

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

Man braucht also für den Erwartungswert von Y die transformierte Dichte f_Y nicht. Falls g konvex ist gilt: $\mathbb{E}[g(x)] \geq g(\mathbb{E}[X])$.

Die Quantile transformieren bei monoton wachsenden Funktionen mit:

$$\alpha = \mathbb{P}(X \leq q_\alpha) = \mathbb{P}(g(X) \leq g(q_\alpha)) = \mathbb{P}(Y \leq g(q_\alpha))$$



5 Deskriptive Statistik

Verfügbare Daten sind eine **Stichprobe** der **Grundgesamtheit**. Damit die Stichprobe repräsentativ ist, muss sie zufällig aus der Grundgesamtheit entnommen werden $\rightarrow$ **Zufallsstichprobe**.

5.1 Kennzahlen

$\boxed{\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)}$ arithmetisches Mittel

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

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

empirisches ($\alpha \times 100\%$)-Quantil

$$\boxed{q_\alpha = \begin{cases} \frac{1}{2}(x_{(\alpha \cdot n)} + x_{(\alpha \cdot n + 1)}) & \text{falls } \alpha \cdot n \ \in \mathbb{Z} \\ x_{(\lceil \alpha \cdot n \rceil)} & \text{sonst} \end{cases}}$$

$$\boxed{q_{0.5} = \begin{cases} \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)}) & \text{falls } n \text{ gerade} \\ x_{(\frac{n+1}{2})} & \text{falls } n \text{ ungerade} \end{cases}}$$ empirischer Median

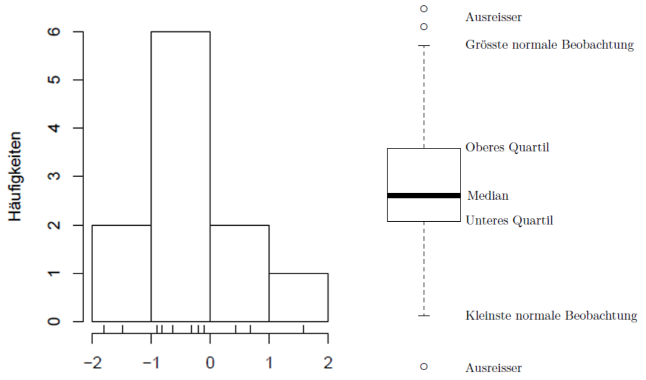
$q_{.25}$ / $q_{.75}$ unteres / oberes Quartil

$\boxed{q_{.75} - q_{.25}}$ Quartilsdifferenz (IQR)

5.2 Grafische Darstellungen

5.2.1 Histogramm

- Aufteilung des Wertebereichs in Klassen $(c_{k-1}, c_k]$.
- Auftragen von Balken mit Höhe proportional zu $\frac{h_k}{c_k - c_{k-1}}$.
- Fläche der Balken proportional zu der Anzahl Beobachtungen im entsprechenden Interval.
- Grössere Klassenbreite führt zu Erosion. Gipfel werden abgetragen, Täler werden aufgefüllt.
- Sturges Rule: Aufteilung in $\lceil 1 + \log_2(n) \rceil$ gleich breite Intervalle.
- Beim Vergleich beachten: Diskret oder nicht?



5.2.2 Boxplot

- Rechteck aus unterem und oberem Quartil.
- Median im Rechteck als Strich.
- Linien vom Rechteck zum grössten und kleinsten normalen Wert.
- Normaler Wert höchstens 1.5 mal die Quartilsdifferenz von einem der beiden Quartile entfern.

5.2.3 Empirische kumulative Verteilungsfunktion

$$\boxed{F_n(x) = \frac{1}{n}\text{Anzahl}\{i | x_i \leq x\} \in [0,1]}$$

5.3 Mehrere Messgrössen

$\boxed{r = \frac{s_{xy}}{s_x s_y} \in [-1,1]}$ empirische Korrelation

$\boxed{\text{empirische Kovarianz}}$

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

- $r = +1$ genau dann, wenn $y_i = a + bx_i$ für ein $a \in \mathbb{R}$ und $b > 0$
- $r = -1$ genau dann, wenn $y_i = a + bx_i$ für ein $a \in \mathbb{R}$ und $b < 0$

$\boxed{\text{Korrelation nie blind berechnen! } \rightarrow \text{ Streudiagramm betrachten!}}$

6 Mehrdimensionale Verteilungen

6.1 Gemeinsame, Rand- und bedingte Verteilungen

Gmsm. Wahr.funkt.	$\mathbb{P}(X = x, Y = y),\ x \in W_X,\ y \in W_Y$
Randverteilung	$\mathbb{P}(X = x) = \sum_{y \in W_Y} \mathbb{P}(X = x, Y = y),\ x \in W_X$
Bed. Verteilung	$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}$
Erwartungswert	$\mathbb{E}[g(x,y)] = \sum_{\substack{x \in W_X \\ y \in W_Y}} g(x,y)p(x,y)$
Bed. Erwartungswert	$\mathbb{E}[Y \mid X = x] = \sum_{y \in W_Y} y\mathbb{P}(Y = y \mid X = x)$
Gmsm. Wahr.dichte	$\mathbb{P}((X,Y) \in A) \int\int_A f_{X,Y}(x,y)dxdy$
Randdichte	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
Bed. Dichte	$f_{Y \mid X=x}(y) = f_Y(y \mid X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
Integralgrenzen beachten!	
Erwartungswert	$\mathbb{E}[g(x,y)] = \int\int_{\mathbb{R}^2} g(x,y) f_{X,Y}(x,y)dxdy$
Bed. Erwartungswert	$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y f_{Y \mid X=x}(y)dy$

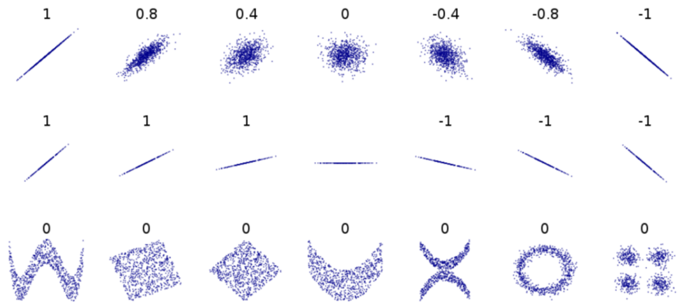
Wenn X und Y unabhängig sind, kann man aus den Randverteilungen auf die gemeinsame Verteilung schliessen.

$$\boxed{\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y),\ x \in W_X,\ y \in W_Y}$$
Unabhängigkeit Diskret

$\boxed{f_{X,Y}(x,y) = f_X(x) f_Y(y), x, y \in \mathbb{R}}$ Unabhängigkeit Kontinuierlich

$\boxed{\mathbb{E}[a + bX + cY] = a + b \cdot \mathbb{E}[X] + c \cdot \mathbb{E}[Y],\ a, b, c \in \mathbb{R}}$ für eine Linearkombination

6.2 Kovarianz und Korrelation



$$\boxed{\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]}\ \text{Kovarianz}$$

$$\boxed{\mathrm{Corr}(X,Y) = \rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}}\ \text{Korrelation}$$

$$\boxed{-1 \leq \mathrm{Corr}(X,Y) \leq 1}$$

$$\mathrm{Corr}(X,Y) = +1 \Leftrightarrow Y = a + bX,\ a \in \mathbb{R},\ b > 0$$
$$\mathrm{Corr}(X,Y) = -1 \Leftrightarrow Y = a + bX,\ a \in \mathbb{R},\ b < 0$$

$|\mathrm{Corr}(X,Y)| = 1 \rightarrow$ perfekt linearer Zusammenhang.
$\mathrm{Corr}(X,Y) = 0 \rightarrow$ X und Y sind unkorreliert.

$$\boxed{X \text{ und } Y \text{ sind unabhängig} \Rightarrow \mathrm{Corr}(X,Y) = 0}$$

Umkehrung gilt im allgemeinen **nicht**!

$$\boxed{\mathrm{Var}(X) = \mathrm{Cov}(X,X)}$$

$$\boxed{\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}$$

$$\boxed{\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]}\ \text{bei Unabhängigkeit}$$

$$\boxed{\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathrm{Cov}(X_i, Y_j),\ a_i, b_j \in \mathbb{R}}$$

$$\mathrm{Cov}(a + bX, c + dY) = bd\mathrm{Cov}(X,Y)$$

$$\mathrm{Corr}(a + bX, c + dY) = \mathrm{sign}(b)\mathrm{sign}(d)\mathrm{Corr}(X,Y)$$

$$\boxed{\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j}^{n} \mathrm{Cov}(X_i, X_j)}$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

Falls X_i unabhängig ist die Varianz der Summe die Summe der Varianzen:

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$$

6.3 Zweidimensionale Normalverteilung

$$\boxed{f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{det(\Sigma)}} \exp\left\{-\frac{1}{2}(x-\mu_X, y-\mu_Y)\Sigma^{-1}\begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix}\right\}}$$

where

$$\Sigma = \begin{pmatrix} \mathrm{Cov}(X,X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Cov}(Y,Y) \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \sigma_Y^2 \end{pmatrix}$$

Wenn $\mathrm{Cov}(X,Y) = 0$ gilt, so ist Σ eine Diagonalmatrix. Dann lässt sich die Wahrscheinlichkeitsdichte als Produkt der beiden Randdichten berechnen. Im Falle der 2D Normalverteilung gilt als dass aus Unkorreliertheit Unabhängigkeit folgt. Im Allgemeinen gilt das jedoch nicht.

7 Grenzwertsätze

7.1 i.i.d.

n Zufallsvariabeln $X_1, \ldots, X_n$, X_i: i-te Wiederholung.
Alle Variablen unabhängig und gleich verteilt.

$$\boxed{\text{i.i.d. : independent and identically distributed}}$$

7.2 Summen und arithmetische Mittel von Zufallsvariabeln

$$S_n = X_1 + \ldots + X_n$$

$$\bar{X}_n = \frac{1}{n} S_n$$

$$\boxed{\begin{array}{lll} \mathbb{E}[S_n] = n\mathbb{E}[X_i] & \mathrm{Var}(S_n) = n\mathrm{Var}(X_i) & \sigma_{S_n} = \sqrt{n}\sigma_{x_i} \\ \mathbb{E}[\bar{X}_n] = \mathbb{E}[X_i] & \mathrm{Var}(\bar{X}_n) = \frac{1}{n}\mathrm{Var}(X_i) & \sigma_{\bar{X}_n} = \frac{1}{\sqrt{n}}\sigma_{X_i} \end{array}}$$

7.3 Gesetz der grossen Zahlen / Zentraler Grenzwertsatz

$$\boxed{\bar{X}_n \overset{n\to\infty}{\to} \mu}$$

$\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_i]$ $\quad$ $\mathrm{Var}((\bar{X}_n)) \rightarrow 0 (n \rightarrow \infty)$

$$f_n(A) \overset{n\to\infty}{\to} \mathbb{P}(A)$$

$$\boxed{S_n \approx \mathcal{N}(n\mu, n\sigma^2)}$$

$$\boxed{\bar{X}_n \approx \mathcal{N}(\mu, \frac{\sigma^2}{n})}$$

Falls $X_i, \ldots, X_n$ i.i.d. mit Erwartungswert μ und Varianz σ^2. Meistens ist $n \geq 30$ schon genug.

Ausnahmen:

- $X_i \sim \mathrm{Bern(p)} \rightarrow X_i \in \{0,1\} \rightarrow S_n \sim \mathrm{Bin}(n,p)$ mit $p = \mathbb{P}(X_i) = 1$.
- $X_i \sim \mathrm{Pois}(\lambda) \rightarrow S_n \sim \mathrm{Pois}(n\lambda)$.
- $X_i \sim \mathcal{N}(\mu, \sigma^2) \rightarrow S_n \sim \mathcal{N}(n\mu, n\sigma^2)$ und $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Weiterhin:
Bin(n,p) $\sim \mathcal{N}$ für grosse n, nicht zu kleine p
Pois(λ) $\sim \mathcal{N}$ für grosse λ

Beispiel

$$X_i \sim \mathrm{Exp}(0.5) \quad I_i \sim \mathrm{Ber}(0.2)$$

X_i und I_i unabhängig

$$\mathbb{P}(\sum_{i=1}^{25} X_i I_i \geq 20) = ?$$

$\mathbb{E}[X_i I_i] = 0.4$ $\quad$ $\mathbb{E}[(X_i I_i)^2] = 1.6$

$$\mathbb{P}(\sum_{i=1}^{25} X_i I_i \geq 20) = \mathbb{P}(\frac{\sum_{i=1}^{25} X_i I_i - 25\mathbb{E}[X_1 I_1]}{\sqrt{25\mathrm{Var}(X_1 I_1)}} \geq \frac{20 - 25\mathbb{E}[X_1 I_1]}{\sqrt{25\mathrm{Var}(X_1 I_1)}})$$

$$\mathbb{P}(\sum_{i=1}^{25} X_i I_i \geq 20) \approx 1 - \Phi\left(\frac{20 - 25\mathbb{E}[X_1 I_1]}{5\sqrt{\mathrm{Var}(X_1 I_1)}}\right) = 1 - \Phi(1.667) = 0.05$$

Für Produkte

X_i i.i.d.

$$Y = \prod_i X_i$$

$$\log(Y) = \sum_i \log(Y_i)$$

$\rightarrow \log(Y)$ normalverteilt $\rightarrow Y$ **lognormalverteilt**

8 Parameterschätzungen

$F(\Theta)$ Verteilungsfamilie

Θ Parameter der Familie

$\hat{}$ Schätzwert

8.1 QQ-Plot

Grundidee: Die empirischen Quantile müssen den theoretischen entsprechen, wenn die Daten tatsächlich aus der entsprechenden Verteilung kommen.

$$\alpha_k = \frac{k-0.5}{n}, k = 1, \ \forall \ \ldots, n$$

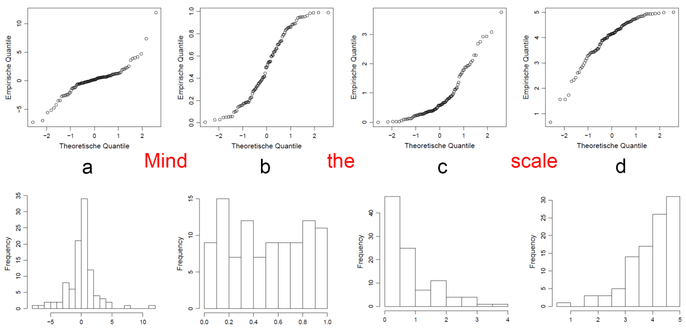
$\alpha n = k - 0.5$ $\quad \lceil \alpha_k n \rceil = k$ also $q_{\alpha_k} = X_{(k)}$

$F^{-1}(\alpha_k)$ theoretisches Quantil

wobei F: kumulative Verteilungsfunktion

$$\boxed{\{F^{-1}(\alpha_k), x_{(x)}\},\ k = 1, \ldots, n}$$ QQ-Plot

- Wenn die geschätzten Parameter stimmen liegen die Punkte des QQ-Plots auf der Linie $x = y$.
- Zur Berechnung der Quantile muss die Verteilung der Daten quasi schon bekannt sein → Man geht von einer Normalverteilung aus und prüft die Annahme im Plot.
- Wenn man die theoretischen Quantile aus $\mathcal{N}(0,1)$ → **Normalplot**
- $X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow q_\alpha = \mu + \sigma z_\alpha$



a) Empirische Quantile weiter weg als theoretische → Langschwänzige Verteilung.

b) Empirische Quantile näher als theoretische → Kurzschwänzige Verteilung.

c) Tiefe emp. Q. näher → links kurz, hohe emp. Q. weiter weg → rechts lang → **Schiefe Verteilung**

d) Tiefe emp. Q. weiter weg → links lang, hohe emp. Q. näher → recht kurz → **Schiefe Verteilung**

8.2 Momenten Methode

$$\boxed{\mu_k(\Theta) = \mathbb{E}[X^k],\ \forall\ k = 1, 2, \ldots}$$ Momente von X

$$\boxed{m_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k}$$ empirische Momente von X

Jetzt sollen die Parameter so gewählt werden das die empirischen mit den theoretischen Momenten übereinstimme.

$$\boxed{\begin{aligned} \mu_1(\Theta) &= m_1 \\ \mu_2(\Theta) &= m_2 \\ &\vdots \\ \mu_r(\Theta) &= m_r \end{aligned}}$$

- Der Momentenschätzer ist dann die Lösung dieses Gleichungssystems.
- Nicht unbedingt eindeutig und kann auch unsinnige Resultate liefern.
- Beispiel: $\begin{aligned} m_1 &\overset{!}{=} \mu_1 = \mathbb{E}[X] = \mu \\ m_2 &\overset{!}{=} \mu_2 = \mathbb{E}[X^2] = \sigma^2 + \mu^2 \end{aligned}$

8.3 Maximum Likelihood Methode

$$\boxed{L(\Theta) = p_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\Theta) = \prod_{i=1}^{n} p_X(x_i|\Theta)}$$

Die Likelihoodfunktion beschreibt wie wahrscheinlich es unter einem Parametersatz Θ ist die beobachteten Grössen zu erhalten.

$$\boxed{l(\Theta) = \log(L(\Theta)) = \sum_{i=1}^{n} \log(p_X(x_i|\Theta))}$$

Man kann auch alternativ die log-Likelihoodfunktion maximieren, da der Logarithmus monoton wachsend ist.

$$\boxed{\frac{\partial}{\partial\Theta} l(\Theta, x_1, \ldots, x_n) = 0 \qquad \hat{\Theta} = \hat{\Theta}(x_1, \ldots, x_n)}$$

Wenn nötig 2. Ableitung überprüfen!

8.4 Schätzer für Parameter der Binomialverteilung

$$\boxed{\hat{p} = \frac{n_{erfolge}}{n_{messungen}} = \frac{x}{n}}$$

$$\boxed{\hat{p} \pm \Phi^{-1}(1 - \tfrac{\alpha}{2}) \cdot \sqrt{\hat{p}(1-\hat{p})\tfrac{1}{n}}}$$ Vertrauensintervall

8.5 Allgemeine Schätzer für Erwartungswert und Varianz

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} F(\Theta)$$

F **und** Θ unbekannt.

$$\hat{\mu}_X = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \mathbb{E}[\hat{\mu}_X] = \mathbb{E}[X] = \mu$$

$$\mathrm{Var}(\hat{\mu}_X) = \frac{1}{n}\sigma_X^2$$

$$\hat{\sigma}_X^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \qquad \mathbb{E}[\hat{\sigma}_X^2] = \sigma_X^2$$

8.5.1 Genauigkeit von Schätzern

$$X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\hat{\mu}_X \sim \mathcal{N}(\mu, \tfrac{\sigma^2}{n})$$

Mit Wahrscheinlichkeit 95% liegt der Schätzer dann im Intervall:

$p\left(\mu - z_{0.975}\frac{\sigma}{\sqrt{n}} \leq \hat{\mu} \leq \mu + z_{0.975}\frac{\sigma}{\sqrt{n}}\right) = 95\%$

$p\left(-z_{0.975}\frac{\sigma}{\sqrt{n}} \leq \hat{\mu} - \mu \leq z_{0.975}\frac{\sigma}{\sqrt{n}}\right) 95\%$

$p\left(-z_{0.975} \leq \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} \leq z_{0.975}\right) = 95\%$

$p\left(\hat{\mu} \in [\mu - z_{0.975}\frac{\sigma}{\sqrt{n}}, \mu + z_{0.975}\frac{\sigma}{\sqrt{n}}]\right) = 95\%$

oder

$p\left(|\hat{\mu} - \mu| \leq z_{0.975}\frac{\sigma}{\sqrt{n}}\right) = 95\%$

oder

$$\boxed{p(\mu \in \hat{\mu} \pm 2\tfrac{\sigma}{\sqrt{n}}) \approx p(\mu \in \underbrace{\hat{\mu} \pm z_{0.975}\frac{\sigma}{\sqrt{n}}}_{\text{Vertrauensinterval}}) = 95\%}$$

$1.96 = z_{0.975} = 0.975 \cdot 100\%$-Quantil von $\mathcal{N}_0^1$
Links und rechts ein 2.5% Anteil.

Merkregel: „Schätzung ± 2 × Standardfehler“

9 Statistische Tests

9.1 Binominaltest

$$H_0 \;:\; p = p_0 \quad \text{Nullhypothese}$$

$$\begin{array}{ll} p \neq p_0 & \text{(„zweiseitig“)} \\ p > p_0 & \text{(„einseitig nach oben“)} \quad \text{Alternativhypothese} \\ p < p_0 & \text{(„einseitig nach untern“)} \end{array}$$

		Entscheidung	
		H_0	H_A
Wahrheit	H_0	Kein Fehler	Fehler 1. Art
	H_A	Fehler 2. Art	Kein Fehler

$$\boxed{\mathbb{P}_{p_0}(X \geq c) = \sum_{k=c}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k}}$$

$$\mathbb{P}_{p_0}(X \leq c) \leq \alpha$$

$$\boxed{\alpha = 0.05 \text{ oder } 0.01} \quad \text{Signifikanzniveau}$$

Falls $x \geq c(\alpha)$ ist die Alternativhypothese statistisch nachgewiesen. Die Abweichung von der Nullhypothese ist **signifikant**. $\rightarrow$ Die Wahrscheinlichkeit einen Fehler 1. Art zu begehen ist kleiner α.
Wobei c der Wert der Realisierung x von X ist ab dem wir davon ausgehen können das die Realisierung mit einem Signifikanzniveau von α H_0 widerspricht, d.h. kein Zufall ist.

$$K = \{c, c+1, \dots, n\} \quad \text{Verwerfungsbereich}$$

Komplementär dazu der **Akzeptanzbereich**.
Anleitung

1. Geeignetes Modell für die Daten wählen.
2. H_0
3. H_A
4. Signifikanzniveau α wählen.
   $\Rightarrow \mathbb{P}(\text{Fehler 1. Art}) \leq \alpha$
5. Verwerfungsbereich finden

Absence of evidence is not evidence of absence.

Falls H_0 nicht verworfen werden kann ist das nicht der Nachweis der Wahrheit von H_0.

9.2 Z-Test (σ bekannt)

Statistischer Test für normalverteilte Daten mit bekannter Streuung.

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2, /n)$$

$$H_0 : \mu = \mu_0$$

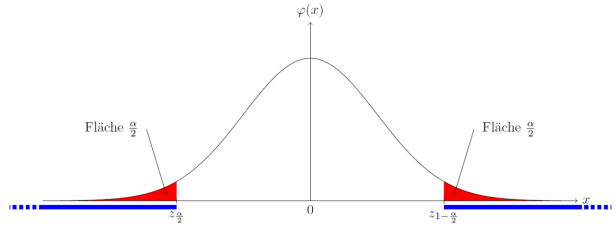
$$H_A : \quad \begin{array}{l} A: \ \mu \neq \mu_0 \\ B: \ \mu > \mu_0 \\ C: \ \mu < \mu_0 \end{array}$$

$$\boxed{Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{\text{beobachtet - erwartet}}{\text{Standardfehler}}} \quad \text{Teststatistik}$$

Eine **Teststatistik** ist eine (spezielle) Zufallsvariable die verwendet wird um die Testentscheidung zu treffen.

$\lvert z\rvert \geq z_{1-\alpha/2}$	$\Leftrightarrow z \in K = (-\infty, -z_{1-\alpha/2}] \cup [z_{1+\alpha/2}, \infty)$	für $H_A : \mu \neq \mu_0$
$z \geq z_{1-\alpha}$	$\Leftrightarrow z \in K = [z_{1-\alpha}, \infty)$	für $H_A : \mu > \mu_0$
$z \leq z_\alpha = -z_{1-\alpha}$	$\Leftrightarrow z \in K = (-\infty, z_\alpha] = (-\infty, -z_{1-\alpha}]$	für $H_A : \mu < \mu_0$

Begründung für Fall 1: Die Wahrscheinlichkeit für einen Fehler 1. Art ist $\mathbb{P}(|Z| \geq z_{1-\frac{\alpha}{2}})_{\mu_0} = \alpha$



9.3 t-Test (σ unbekannt)

Statistischer Test für normalverteilte Daten mit unbekannter Streuung. Wir ersetzen daher σ mit einem Schätzer $S_n/\sqrt{n}$

$$\boxed{T = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} = \frac{\text{beobachtet-erwartet}}{\text{geschätzter Standardfehler}}} \quad \text{Teststatistik}$$

Durch den Schätzer haben wir eine zusätzliche Variationsquelle $\rightarrow$ T ist keine Standardnormalverteilung sondern folgt einer **t-Verteilung** mit n-1 Freiheitsgraden.

$$T \sim t_{n-1}$$

Eine t-Verteilung ist symmetrisch um 0, nimmt aber eher betragsmässig grosse Werte an / ist langschwänziger.
Pro Beobachtung ein Freiheitsgrad, pro Parameter der uns interessiert müssen wir einen bezahlen $\rightarrow$ n-1 Freiheitsgrade (n Beobachtungen - μ gesucht).

$$\boxed{t = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}}$$

$\lvert t\rvert \geq t_{n-1,1-\frac{\alpha}{2}}$	$\Leftrightarrow t \in K = (-\infty, -t_{n-1,1-\frac{\alpha}{2}}] \cup [t_{n-1,1-\frac{\alpha}{2}})$	$H_A : \mu \neq \mu_0$
$t \geq t_{n-1,1-\alpha}$	$\Leftrightarrow t \in K = [t_{n-1,1-\alpha}, \infty)$	$H_A : \mu > \mu_0$
$t \leq t_{n-1,\alpha}$	$\Leftrightarrow t \in K = (-\infty, t_{n-1,\alpha}] = (-\infty, -t_{n-1,1-\alpha}]$	$H_A : \mu < \mu_0$

10 Allg. Eigenschaften von stat. Tests

10.1 Macht

$$\boxed{\mathbb{P}(\text{Fehler 1. Art}) \leq \alpha}$$

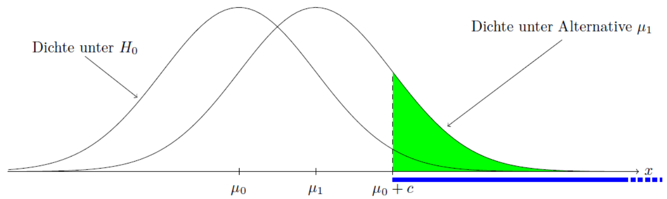
$$\boxed{\beta(\theta) = \mathbb{P}(\text{Fehler 2. Art})}$$

$$\boxed{1 - \beta(\theta) = \mathbb{P}(\text{Test verwirft richtigerweise } H_0 \text{ für } \theta \in H_A)} \quad \text{Macht}$$

Berechnung von β

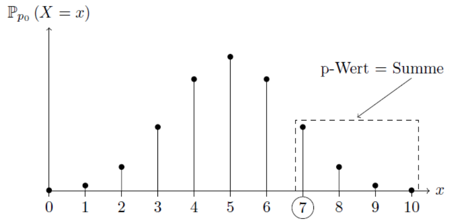
1. Verwerfungsbereich K auf Niveau α festlegen.
2. Annahme treffen für $\theta \in H_A$
3. Verteilung von X bezüglich θ berechnen.
4. β ist dann die Wahrscheinlichkeit, dass $X(\theta)$ in $K(\alpha)$ fällt.

Je weiter μ von μ_0 weg liegt, desto grösser die Macht. Je grösser die Stichprobe desto grösser die Macht.

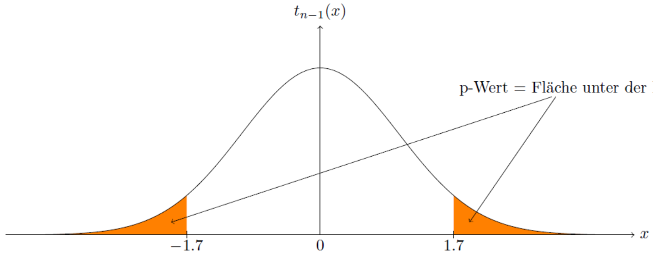


10.2 P-Wert

Je kleiner das Signifikanzniveau α desto schwieriger wird es H_0 zu verwerfen.
Der p-Wert is die Wahrscheinlichkeit, unter der Nullhypothese ein mindestens so extremes Ereignis zu beobachten wie das aktuell beobachtete.



Bin(10,0.5)-Verteilung, $H_0 : p = 0.5$, $H_A : p > 0.5$, betrachteter Wer: x=7



Zweiseitiger t-Test, t=1.7

- Der p-Wert ist **nicht** die Wahrscheinlichkeit, dass die Nullhypothese stimmt.
- Der p-Wert ist **nicht** grösser als die Wahrscheinlickeit eines Fehlers 1. Art.
- Der p-Wert sagt nichts über die **Effektgrösse**. Betrachte das Vertrauensintervall.
- Falls man genügend Tests macht findet man immer irgendwann einen signifikanten p-Wert.

10.3 Multiple testing correction

K tests, $\mathbb{P}(\text{Fehler 1.Art in mindestens einem Test}) \leq \alpha$
$\rightarrow$ jeden einzelnen Test zum Niveau α/K machen!

10.4 Vertrauensintervalle

$$I = \{\theta_0 : \text{Nullhypothese } H_0 : \theta = \theta_0 \text{ wird nicht verworfen }\}$$

- Das Vertrauensintervall ist der Wertbereich für θ, den wir aufgrund der vorliegenden Daten als plausibel betrachten.
- Das Vertrauensintervall ist **zufällig**, denn es hängt indirekt von den Realisierungen der Zufallsvariablen ab.
- Im Vertrauensintervall sieht man automatisch welche Nullhypothesen verworfen werden.
- Je enger das Vertrauensintervall desto genauer die Parameterschätzung.
- Das Vertrauensintervall ist **nicht** der Annahmebereich des Tests. Der Annahmebereich geht von einer konkreten Nullhypothese aus, das Vertrauensintervall definiert annehmbare Nullhypothesen.
- Informationsgehalt: Testentscheid$\preceq$P-Wert$\preceq$Vertrauensintervall

$$\mathbb{P}(I \ni \theta) = 1 - \alpha \quad \text{Alternativinterpretation}$$

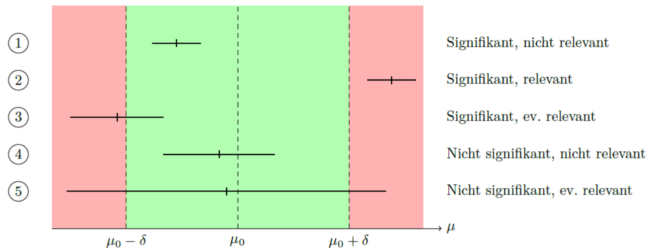
$$\boxed{I = \bar{X}_n \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}} \text{ z-Test}$$

$$\boxed{I = \bar{X}_n \pm \frac{S_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}} \text{ t-Test}$$

Länge eines Vertrauensintervalls: $1 - \alpha$ Interval $\rightarrow L_\alpha = 2\frac{\sigma}{\sqrt{n}} z_{1-\alpha_2}$

10.5 Statistische Signifikanz und Fachliche Relevanz

Beispiel: Abfüllmaschine
Abweichungen vom Sollgewicht, 1000 g bis zu 5 g spielen keine Rolle.
Irrelevanzbereich: $(995, 1005)$ Alle Abweichungen ausserhalb sind relevant.
Vertrauensintervall: $[1001.75, 1003.51]$
Es ist also möglich das eine Abweichung als statistisch signifikant behandelt wird, obwohl sie nicht relevant ist.



10.6 Vorzeichen-Test

Für nicht normalverteilte Daten

$$\mu \quad \text{Median}$$

$$H_0 : \mu = \mu_0$$

$$H_A : \mu \neq \mu_0$$

Idee: Wenn $\mu = \mu_0$ tatsächlich stimmt, dann sollten wir gleich viele Werte grösser μ_0 wie Werte kleiner μ_0 beobachten.

$$\text{Q} = \text{Anzahl Werte grösser } \mu_0$$

Unter H_0 folgt dann Q einer Bin(n,0.5)-Verteilung, so dass man einen Binominaltest durchführen kann.

10.7 Wilcoxon-Test

$$X_1, \ldots, X_n \sim \mathcal{F}(\mu)$$

wobei $\mathcal{F}(\mu)$ eine **symmetrische** und **stetige** Verteilung mit Mittelwert bzw. Median μ ist.

$$\text{Rang}(|X_i - \mu_0|) = k$$

$$V_i = \begin{cases} 1 & X_i > \mu_0 \\ 0 & \text{sonst} \end{cases}$$

$$\boxed{\text{Teststatistik}} \; W = \sum_{i=1}^{n} \text{Rang}(|X_i - \mu_0|) V_i$$

Der Verwerfungsbereich ergibt sich aus Tabelle 7.4 im Skript.

11 Vergleich zweier Stichproben

11.1 Gepaarte und ungepaarte Stichproben

Gepaarte Stichprobe: Beide Versuchsbedingungen an derselben Versuchseinheit eingesetzt.
Ungepaarte Stichprobe: Versuchseinheiten beider Gruppen sind unabhängig.

11.2 Versuchsplanung

Es muss sichergestellt werden, dass Unterschiede zwischen den Testgruppen tatsächlich durch verschidene Versuchsbedingungen und nicht durch andere Störgrössen verursacht werden.
Randomisierung: Zuordnung von Versuchseinheit und Versuchsbedingung wird zufällig gewählt und in zufälliger Reihenfolge ausgemessen.
Für gepaarte Stichproben kann nur die Reihenfolge / Platzierung der Versuchsbedingungen und der Ausmessung randomisiert werden.
Wenn die eine Versuchsbedingung als Kontrollgruppe dient → **randomisierte kontrollierte Sudie**.
Randomisierung ist so mächtig, da sie eine systematische Aufteilung der Testgruppen ausschliesst und somit jeder Unterschied der Testgruppen auf die verschiedenen Versuchsbedingungen zurückzuführen ist.
Bekannte Eigenschaften, von denen man im Voraus weiss, dass sie Einfluss auf das Ergebnis haben sollen ausgenutzt werden um homogene Gruppen zu bilden (Blocks).

Block what you can, randomize what you cannot.

Doppelblindheit: Weder die Versuchseinheit, noch der Messer weiss unter welchen Versuchsbedingungen gemessen wird.
Beobachtungsstudie: Wenn Versuchsbedingungen nicht randomisiert werden können (Rauchen - Lungenkrebs).
Cofounder: Umstand der einen Einfluss auf die Messergebnisse hat, der nicht von den Versuchsbedingungen herrührt. Wird automatisch durch Randomisierung ausgeschlossen.

11.3 Gepaarte Vergleiche

Gepaarte Stichprobe → Differenz innerhalb der Paare: $u_i = x_i - y_i$
Kein Unterschied → $\mathbb{E}[U_i] = 0$

$$H_0 : \mathbb{E}[U_i] = 0$$

$$H_A : \mathbb{E}[U_i] \neq 0$$

Geeignete Tests:

1. t-Test
2. Vorzeichen-Test, falls die Normalverteilung nicht gerechtfertigt scheint.
3. Wilcoxon-Test

11.4 Zwei-Stichproben Tests

Bei unabhängigen Stichproben lassen sich keine Paare bilden, dann hat man 2 i.i.d. Zufallsvariablen:

$$X_1, \ldots, X_n \qquad Y_1, \ldots, Y_m$$

im einfachsten Fall:

$$X_1, \ldots, X_n \; i.i.d. \sim \mathcal{N}(\mu_X, \sigma^2) \qquad Y_1, \ldots, Y_m \; i.i.d. \sim \mathcal{N}(\mu_Y, \sigma^2)$$

Dann haben wir folgende Hypothesen:

$$H_0 : \mu_X = \mu_Y$$

$$H_A : \mu_X \neq \mu_Y$$

$$Z = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} 1 \frac{1}{m}}}$$

Erste Definition der Teststatistik Z.
Da $\text{Var}(\bar{X}_n - \bar{Y}_m) = \sigma^2(\frac{1}{n} + \frac{1}{m})$ ist $Z \sim \mathcal{N}(0,1)$!
Aber σ ist in der Praxis unbekannt. Also $\sigma \to S_{pool}$

$$\boxed{\begin{aligned} S_{pool}^2 &= \frac{1}{n+m-2} \left(\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 + \sum_{i=1}^{m} (Y_i - \bar{Y}_m)^2 \right) \\ &= \frac{1}{n+m-2} \left((n-1) S_X^2 + (m-1) S_Y^2 \right) \end{aligned}}$$

S_{pool} ist ein gewichtetes Mittel der Schätzungen der Varianzen in den beiden Gruppen.

$$\boxed{T = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}}} \text{ Teststatistik}$$

$\lvert t\rvert \geq t_{*,1-\frac{\alpha}{2}}$	$\Leftrightarrow t \in K = (-\infty, -t_{*,1-\frac{\alpha}{2}}] \cup [t_{*,1-\frac{\alpha}{2}}, \infty)$	$H_A : \mu_X \neq \mu_Y$
$t \geq t_{*,1-\alpha}$	$\Leftrightarrow t \in K = [t_{*,1-\alpha}, \infty)$	$H_A : \mu_X > \mu_Y$
$t \leq -t_{*,1-\alpha}$	$\Leftrightarrow t \in K = (-\infty, t_{*,\alpha}] = (-\infty, -t_{*,1-\alpha}]$	$H_A : \mu_X < \mu_Y$

$t_* = t_{n+m-2}$

12 Übersicht

12.1 Die wichtigsten 1D Verteilungen

Verteilung	$p(x)$	W_X	$\mathbb{E}[X]$	$\text{Var}(X)$
Bernoulli(p)	$p^x(1-p)^{1-x}$	$\{0, 1\}$	p	$p(1-p)$
Bin(n,p)	$\binom{n}{k} p^x (1-p)^{n-x}$	$\{0, \ldots, n\}$	np	$np(1-p)$
Geom(p)	$p(1-p)^{x-1}$	$\{1, 2, \ldots\}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Pois(λ)	$e^{-\lambda} \frac{\lambda^x}{x!}$	$\{0, 1, \ldots\}$	λ	λ
Uni(a,b)	$\frac{1}{b-a}$	$[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exp(λ)	$\lambda e^{-\lambda x}$	$\mathbb{R}_+$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma(α, λ)	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$	$\mathbb{R}_+$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	$\mathbb{R}$	μ	σ^2

12.2 Rechenregeln

1. $\mathbb{E}[a + bX] = a + b \cdot \mathbb{E}[X], \; a, b \in \mathbb{R}$
2. $\mathbb{E}[a + bX + cY] = a + b \cdot \mathbb{E}[X] + c \cdot \mathbb{E}[Y], \; a, b, c \in \mathbb{R}$
egal ob X,Y unabhängig sind oder nicht
3. $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
4. $\text{Var}(a + bX) = b^2 \text{Var}(X), \; a, b \in \mathbb{R}$
5. $\sigma_{a+bX} = |b| \sigma_X, \; b \in \mathbb{R}$
6. $\text{Var}(a) = 0, \; a \in \mathbb{R}$
7. $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
8. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
9. $\text{Cov}(X, X) = \text{Var}(X)$
10. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$
11. $\text{Cov}(X, a) = 0, \; a \in \mathbb{R}$
12. $\text{Cov}(a + bX, c + dY) = bd\text{Cov}(X, Y), \; a, b, c, d \in \mathbb{R}$
13. $\text{Corr}(a + bX, c + dY) = \text{sign}(b)\text{sign}(d)\text{Corr}(X, Y), \; a, b, c, d \in \mathbb{R}$
14. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
15. Sind X und Y **unabhängig** so gilt:
    - $\text{Cov}(X, Y) = 0$
    - $\text{Corr}(X, Y) = 0$

    Achtung: Die Umkehrung gilt im Allgemeinen nicht! D.h. aus Unkorreliertheit folgt nicht Unabhängigkeit.
16. X und Y **unabhängig** (oder allgemeiner: unkorreliert), so gilt:
    - $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
    - $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$
    - $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Oder allgemeiner für mehrere Zufallsvariablen:

17. $\mathbb{E}\left[a_0 + \sum_{i=1}^{n} a_i X_i\right] = a_0 + \sum_{i=1}^{n} a_i \mathbb{E}[X_i], \; a_i \in \mathbb{R}$
18. $\text{Cov}\left(a_0 + \sum_{i=1}^{n} a_i X_i, b_0 + \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \text{Cov}(X_i, Y_j), \; a_i, b_j \mathbb{R}$
19. $\text{Var}\left(a_0 + \sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{Cov}(X_i, X_j), \; a_i \in \mathbb{R}$
20. $X_1, \ldots, X_n$ **unabhängig** (oder allgemeiner unkorreliert), so gilt:

$$\text{Var}\left(a_0 + \sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i)$$

für $a_i \in \mathbb{R}$ (konstanter Term fällt weg, es verbleibt die Summe der Varianzen.