Stochastik

Michael van Huffel

Version: March 3, 2022

This summary has been written based on the Lecture of Stochastik (401-0603-00L) by Prof. Dr. Patrick Cheridito (Autumn 21) and the summary of Till Richter. There is no guarantee for completeness and/or correctness regarding the content of this summary. Use it at your own discretion

0 Contents

1 Grundlage Wahrscheinlichkeitrechnung

- Ω: Ereignisraum, Grundraum
- ω: Elementarereignis (sich gegenseitig ausschliessende Ergebnisse)
- A, B, C: Ereignis, Teilmenge von Ω
- Schnittmenge: $A \cap B$ heisst " A und B "
  Falls $A \cap B = \emptyset$: A,B *disjunkt*
- Vereinigung: $A \cup B$ heisst "' A oder B "'.
- Komplement: $A^c = \overline{A}$ heisst "'nicht A"'
- Differenz: $A \backslash B = A \cap B^c$ heisst "' A ohne B "'

1.1 Laplace Modell:

Uniforme Verteilung:

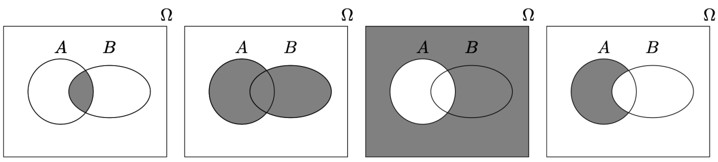
$$\mathbb{P}[\omega] = 1/\Omega \quad \text{fur alle } \omega$$

Alle Elementarereignisse haben gleiche Wahrscheinlichkeit!

$$\mathbb{P}[A] = \frac{|A|}{|\Omega|} = \frac{\#\omega \in A}{\#\omega \in \Omega}$$

Theorem: De Morgan'sche Regeln

1. $(A \cap B)^c = A^c \cup B^c$
2. $(A \cup B)^c = A^c \cap B^c = 1 - A \cup B$



1.1.1 Rechenregeln:

- $A \cup B = B \cup A$
- $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$
- $\mathbb{P}[A^c] + \mathbb{P}[A] = 1$
- $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B] = \mathbb{P}[A \cap B^c] + \mathbb{P}[B]$
- $\mathbb{P}[A \cap B] = \mathbb{P}[A|B] \cdot \mathbb{P}[B] = \mathbb{P}[B|A] \cdot \mathbb{P}[A]$
- $\mathbb{P}[A \cap B]^c = \mathbb{P}[A^c] + \mathbb{P}[B^c]$

Theorem: Kolmgorov'sche Axiome

1. $0 \leqslant \mathbb{P}[A] \leqslant 1$
2. $\mathbb{P}[\Omega] = 1 \qquad \mathbb{P}[\emptyset] = 0$
3. $\mathbb{P}\left(\cup_{i=1}^{m} A_i\right) = \mathbb{P}[A_1 \cup ... \cup A_m] \leqslant \mathbb{P}[A_1] + ... + \mathbb{P}[A_m]$ (wenn ausschliessend)

1.2 Unabhaengigkeit:

Eintreten von A beeinflusst WK von B nicht. (Kein kausaler Zusammenhang.)

$$\boxed{\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]}$$

disjunkt $\Rightarrow$ abhängig (falls $\mathbb{P}[A], \mathbb{P}[B] > 0$. Es gilt $\mathbb{P}[A \cap B] = 0 \neq \mathbb{P}[A]\mathbb{P}[B]$)

1.3 Kombinatorik:

n: # in Grundmenge / k: # in Elemente

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

	Variation mit Reihenfolge	Kombination ohne Reihenfolge
mit Zurücklegen	n^k	$\binom{n+k-1}{k}$
ohne Zurücklegen	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Hypergeometrische Verteilung: Total N, betrachte C, davon K richtig in $B \subset N$, $C - K$ richtig in $N - B$

$$\frac{\binom{B}{K} \cdot \binom{N-B}{C-K}}{\binom{N}{C}}$$

1.4 Zufallsvariable:

Funktion von Ω nach $\mathbb{R} : \omega \rightarrow X(\omega)$

- X **diskret**, falls ω "'abzählbar"'
- X **stetig**, falls ω "'in einem Intervall"'

W := Wertebereich der Zufallsvariable

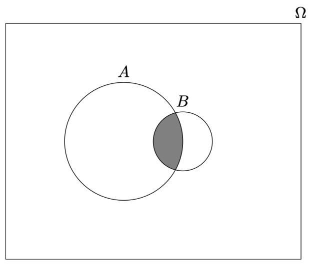
1.5 Bedingte Wahrscheinlichkeit:

$$\boxed{\mathbb{P}[A|B] = \mathbb{P}[A \text{ gegeben } B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}}$$

$$\mathbb{P}[A|B] \cdot \mathbb{P}[B] = \mathbb{P}[A \cap B] = \mathbb{P}[B \cap A] = \mathbb{P}[B|A] \cdot \mathbb{P}[A]$$

Falls A und B unabhaengig:

$$\mathbb{P}[A|B] = \mathbb{P}[A|B^c] = \mathbb{P}[A]$$



Note: Das Ereignis A kann geschrieben werden als $A = (A \cap B) \cup (A \cap B^c)$

1.6 Rechenregeln:

- $\mathbb{P}(B|B) = 1$
- $\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) - \mathbb{P}(A_1 \cap A_2)$
- $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$
- $\mathbb{P}(G \cap (G \cup B)) = \mathbb{P}(G)$ and $\mathbb{P}(B \cap (G \cup B)) = \mathbb{P}(B)$
- $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$, siehe Satz von Bayes
- $\mathbb{P}(A|B^c) = 1 - \mathbb{P}(A^c|B^c)$
- $\max\{0, 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c)\} \leq \mathbb{P}(A \cap B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\}$

Theorem: Satz der totalen Wahrscheinlichkeit

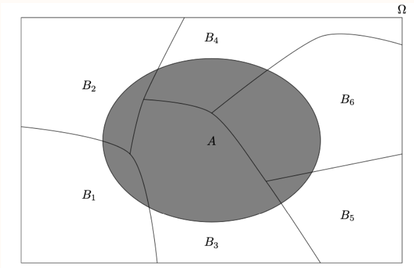
k disjunkte Ereign. $B_1, \ldots, B_k$, wobei $B_1 \cup \ldots \cup B_k = \Omega$:

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

Theorem: Satz von Bayes

Fuer B_j disjunkte Ereignisse es gilt:

$$\mathbb{P}[B_i|A] = \frac{\mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]}{\mathbb{P}[A|B_1] \cdot \mathbb{P}[B_1] + ... + \mathbb{P}[A|B_k] \cdot \mathbb{P}[B_k]}$$



2 Diskrete Verteilung

2.1 Wahrscheinlichkeitsfunktion p:

$$p(x_i) = \mathbb{P}[X = x_i] = \mathbb{P}(\{\omega : X(\omega) = x_i\})$$

Eigenschaften: $\quad p(x_i) \geqslant 0, \qquad \sum p(x_i) = 1$

2.2 Kumultative Verteilungsfunktion:

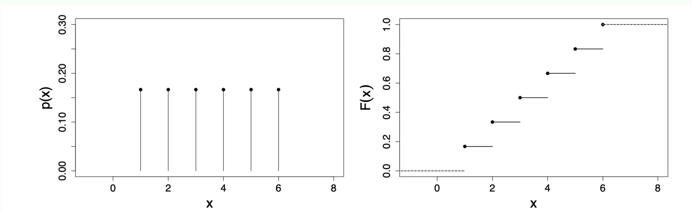
Es gilt:

- $F(b) = \mathbb{P}[X \leqslant b] = \mathbb{P}(X \in (-\infty, b]) = \sum_{x_i < x} p(x_i)$
- $\mathbb{P}[X \in (a, b]] = \mathbb{P}(a < X \leq b) = F(b) - F(a)$
- $\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = 1 - F(x)$

Die kumulative Verteilungsfunktion F erfuelt zudem immer:

1. F ist **monoton steigend**: $F(x_1) \leqslant F(x_2) \ \forall \ x_2 > x_1$
2. F ist **rechtsstetig**: $\lim_{h \to 0} F(x+h) = F(x)$
3. $\lim_{x \to \infty} F(x) = 1$ und $\lim_{x \to 0} F(x) = 0$

Example: Verteilung Wahscheinlichkeit Wurfel



2.3 Erwartungswert und Varianz (diskret):

Erwartungswert (mittlere Lage):

$$\mu_x = \mathbb{E}(X) = \sum_{x_i \in W} x_i \cdot p(x_i), \quad \mathbb{E}(g(X)) = \sum_{x_k \in W} g(x_k) \cdot p(x_k)$$

Varianz (Streuung):

$$Var(X) = \sigma_x^2 = \mathbb{E}[(X - \mathbb{E}(X))^2] = \sum (x_i - \mu_i)^2 p(x_i)$$

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

Standardabweichung:

$$\sigma_x = \sqrt{Var(X)}$$

2.3.1 Rechenregeln:

- $\mathbb{E}[a + bX] = a + b\mathbb{E}[X]$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- $Var(a + bX) = b^2 Var(X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

2.4 Bernoulli $X \sim BER(p)$:

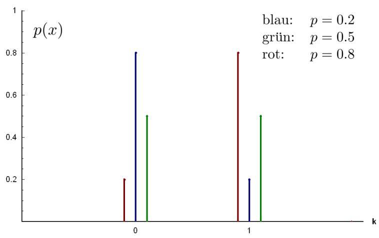
Experimente mit ja/nein Ergebnis, $p(x = \{0, 1\})$

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

Note: $Var[X(1 - X)] = 0$

Example: Bauteilen

Wenn $X \sim BER(p) \Rightarrow X^n \sim BER(p)$



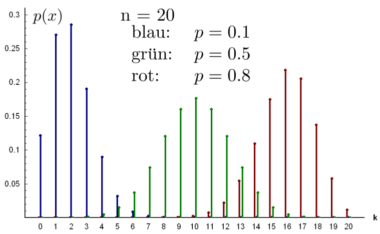
2.5 Binomialverteilung $X \sim BIN(n, p)$:

Wiederholung eines Bernoulliexperimentes
#Wiederholungen: n, WK Erfolg: p, #Erfolge: x
Note: if n big, wegen ZGWS: $BIN(n, p) \approx \mathcal{N}(np, np(1 - p))$

Example: Bauteilen

Aus 10 Bauteilen WK, dass 1 defekt hat:

$$\mathbb{P}(X \geq 1) = \mathbb{P}(X = 1) + \ldots + \mathbb{P}(X = 10) = 1 - \mathbb{P}(X = 0)$$



2.6 Geometrische Verteilung $X \sim GEO(p)$:

Anzahl Wiederholungen x bis Erfolg mit $BER(p)$ (WK p).

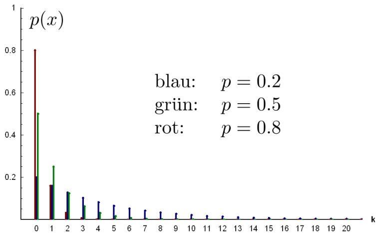
$$F(x) = \mathbb{P}(X \leq x) = \sum_{i=1}^{x} p(i) = 1 - (1 - p)^x$$

Example: Mindeste Versuche

Mindeste # Versuche bis 50% Chance Erfolg?

$F(n) \geq 0.5 \quad$, solve for n

Note: $1 - F(n)$ ist die Chance kein Erfolg bis n.



2.7 Poissonverteilung $X \sim POI(\lambda)$:

Verteilung # Ereignisse fuer seltene Ereignisse (p klein) bei viele (n gross) unabhangige Versuche.
Note:

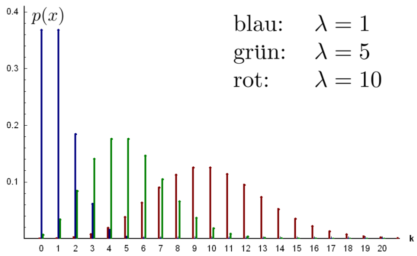
- n gross, p klein: $\text{BIN}(n,p) \sim \text{POI}(\lambda = np)$
- if λ gross, wegen ZGWS: $X \approx \mathcal{N}(\lambda, \lambda)$

Anzahl pro Zeiteinheit: x, Schnitt pro Zeiteinheit: λ
Bsp: Radioaktiver Zerfall, Callcenter

$$X \sim POI(\lambda_1) \quad Y \sim POI(\lambda_2) \quad (X+Y) \sim POI(\lambda_1 + \lambda_2)$$

Note: $\mathbb{E}(X \cdot Y) = \lambda_1 \lambda_2$ wenn X, Y unabhangig sind.

Note: Wenn wir aber $\frac{1}{2}(X+Y)$ betrachten, so liegt keine Poissonverteilung vor mit Parameter $\frac{1}{2}(\lambda_1 + \lambda_2)$. Der Grund ist ganz einfach: Nur schon der Wertebereich stimmt nicht fur eine Poissonverteilung! Der Erwartungswert ist aber $\frac{1}{2}(\lambda_1 + \lambda_2)$.



2.8 Übersicht diskreter Verteilungsfunktionen:

Verteilung	$p(x)$	W_X	$\mathbb{E}[X]$	σ^2 $Var(X)$
BER(p)	$p^x(1-p)^{1-x}$	{0,1}	p	$p(1-p)$
BIN(n,p)	$\binom{n}{x}p^x(1-p)^{n-x}$	$\{0,\dots,n\}$	np	$np(1-p)$
GEO (p)	$p(1-p)^{x-1}$	$\{1,2,\dots\}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
POI(λ)	$\frac{\lambda^x}{x!}e^{-\lambda}$	$\{0,1,\dots\}$	λ	λ

3 Stetige Verteilung

Da $\mathbb{P}[X = x] \approx 0$ betrachtet man $\mathbb{P}[x \le X \le x+h]$
Note: If X_1, X_2 identisch verteilt, so ist $\mathbb{P}(X_1 \le X_2) = 0.5$.

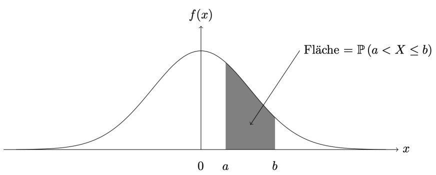
Wahrscheinlichkeitsdichte f**:**

$$f(b) = F'(b) = lim_{h\to 0}\frac{\mathbb{P}[x \leqslant b \leqslant x+h]}{h}$$

Mit der **kumulativen Verteilungsfunktion** $F(x)$**:**

$$F(x) = \mathbb{P}[X \leqslant x] = \int_{-\infty}^{x} f(u)du$$

$$\mathbb{P}(a < X \le b) = F(b) - F(a) = \int_a^b f(x)\,dx$$



Note: $F_X^2(x) = \mathbb{P}[X^2 \le x] = \mathbb{P}[X \le \sqrt{x}] = \displaystyle\int_{-\infty}^{\sqrt{x}} f_X(x)\,dx$

Eigenschaften:

1. $f(x) \geqslant 0 \quad F$ ist steigend.
2. $\int_{-\infty}^{\infty} f(x)dx = 1, \quad F(-\infty) = 0 \quad F(\infty) = 1$
3. $f(x) > 1$ ist möglich!

3.1 Erwartungswert und Varianz (stetig):

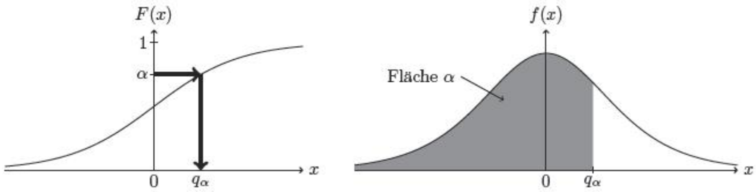
- $\mathbb{E}[X] = \mu_x = \int_{-\infty}^{\infty} x \cdot f(x)dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)dx$
- $Var(X) = \sigma_x^2 = \int_{-\infty}^{\infty}(x-\mu_x)^2 \cdot f(x)dx$
- $Var(X) = \mathbb{E}\left((x - \mathbb{E}[X])^2\right) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

3.2 Quantile $q(\alpha)$:

$$\mathbb{P}[X \leqslant q(\alpha)] = \alpha \qquad q(\alpha) = F^{-1}(\alpha)$$

Note:

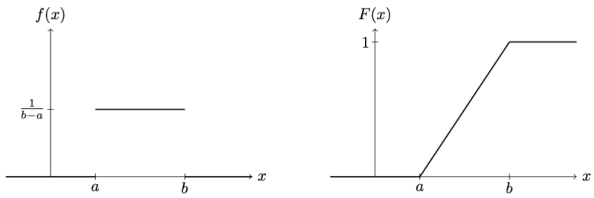
- Median wenn $\alpha = 1/2$
- Verteilung symmetrisch $\Rightarrow q_{\frac{1}{2}} = E(X)$
- Median ist robuster gegen Ausreisser als Erwartungswert
- **Fuer empirische Quantile siehe 4.1.**



3.3 Uniforme Verteilung $X \sim UNI(a,b)$:

Stetige Version des Laplace-Modells, benutzt für Messfehler, $W = [a,b]$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{else} \end{cases} \qquad F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}$$



Eigenschaften:

- $\mathbb{E}[X] = \dfrac{(a+b)}{2}$
- $Var(X) = \sigma_X^2 = \dfrac{(b-a)^2}{12}$
- $Median(X) = \frac{a+b}{2}$
- Auf einen Kres: $f(x,y|r) = \begin{cases} \frac{1}{r^2\pi}, & x^2+y^2 \le r^2 \\ 0, & \text{else} \end{cases}$
- k-tes Moment

$$m_k = \frac{1}{k+1}\sum_{i=0}^{k} a^i b^{k-i} = \frac{1}{k+1}\sum_{i=0}^{k}\left(\mu - \sqrt{3\sigma^2}\right)^i\left(\mu + \sqrt{3\sigma^2}\right)^{k-i}$$

- k-tes zentrales Moment

$$\mu_k = \begin{cases} \frac{(b-a)^k}{2^k(k+1)} & \text{k gerade} \\ 0 & \text{k ungerade} \end{cases} = \begin{cases} \frac{\sqrt{3}^k\sigma^k}{(k+1)} & \text{k gerade} \\ 0 & \text{k ungerade} \end{cases}$$

3.3.1 Summe gleichverteilter UNI Zufallsvariablen:

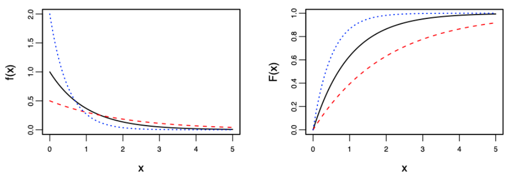
Wenn Breite gleich ⇒ Dreiecksverteilt, sondern trapezförmige Verteilung. Geg: Intervall [a,b], die andere auf dem Intervall [c,d]. Sei $\alpha = \min\{d-c, b-a\}$und $\beta = \max\{d-c, b-a\}$:

$$f\colon \mathbb{R} \to \mathbb{R}, x \longmapsto \begin{cases} 0 & x \notin [a+c, b+d] \\ \frac{x}{\alpha\beta} - \frac{a+c}{\alpha\beta} & x \in [a+c, a+c+\alpha] \\ \frac{1}{\beta} & x \in [a+c+\alpha, a+c+\beta] \\ \frac{b+d}{\alpha\beta} - \frac{x}{\alpha\beta} & x \in [a+c+\beta, b+d] \end{cases}$$

3.4 Exponentialverteilung $X \sim EXP(\lambda)$:

Einfachstes Modell für Wartezeiten auf Ausfälle, $W = [0, \infty)$.

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0 \qquad F(x)] = 1 - e^{-\lambda x}, \quad x \geq 0$$



Eigenschaften:

- $\mathbb{E}[x] = \sigma_X = \frac{1}{\lambda}$
- $Var(X) = \sigma^2 = \frac{1}{\lambda^2}$
- $q_\alpha = \frac{\ln \frac{1}{1-\alpha}}{\lambda}$
- k-tes Moment: $\mathbb{E}(X^k) = \frac{k!}{\lambda^k}$
- Sind $X_1 \sim \text{Exp}(\lambda_1), \ldots, X_n \sim \text{Exp}(\lambda_n)$ stochastisch unabhängig, so ist $\min(X_1, \ldots, X_n) \sim \text{Exp}(\lambda_1 + \cdots + \lambda_n)$
- Sind $X_1 \sim \text{Exp}(\lambda_1), \ldots, X_n \sim \text{Exp}(\lambda_n)$ stochastisch unabhängig, so ist $X_1 + \ldots + X_n$ eine Linearkombination von Exponentialverteilungen, λ_i alle gleich, so ist die Summe $\sim \text{Erl}(n, \lambda)$ **Erlang-verteilt:** $(\mathbb{E}(x) = \frac{n}{\lambda}, Var(x) = \frac{n}{\lambda^2})$

$$f_{\sum_{i=1}^n X_i} = \frac{\lambda^n \cdot x^{n-1}}{(n-1)!} e^{-\lambda x} \quad F(x) = \begin{cases} 1 - \mathrm{e}^{-\lambda x} \sum_{i=0}^{n-1} \frac{(\lambda x)^i}{i!} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Example: Poissonprozesse und Ausfall von n Sys.

1. Sei $N(t)$ die Anzahl Ereignisse im Zeitintervall $[0, t]$, $t \in \mathcal{R}$. Fur einen sogenannten homogenen Poissonprozess gilt $N(t) \sim POI(\lambda t)$. T_1 = Zeitpunkt des ersten Ereignisses. Es gilt:

   $$\{T_1 > t\} = \{\text{Kein Ereignis in [0, t]}\} = \{N(t) = 0\}$$

   $$\mathbb{P}(T_1 \leq t) = 1 - e^{-\lambda t}$$

   Zeit bis ersten Ereignis: $T_1 \sim Exp(\lambda)$ ⇒ wegen Unabhanignkeit gilt im algemein: Zeiten zwischen zwei aufeinanderfolgenden Ereignissen exponential-verteilt sind.
2. Ausfall von n Komponenten $\mathbb{P}(x) = 1 - e^{-n\lambda x}$

3.4.1 Überlebenswahrscheinlichkeit:

Es ergibt sich unmittelbar die auf einen Zeitpunkt x_0 bezogene bedingte Überlebenswahrscheinlichkeit:

$$\mathbb{P}(X > x_0 + x | X > x_0) = \frac{\mathrm{e}^{-\lambda(x_0+x)}}{\mathrm{e}^{-\lambda x_0}} = \mathrm{e}^{-\lambda x}$$

3.5 Normal- / Gaussverteilung $X \sim \mathcal{N}(\mu, \sigma^2)$:

Modell für Verteilung von Messwerten, $W = \mathbb{R}$.

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

$$\mathbb{E}[X] = \mu \qquad \sigma_X = \sigma$$

Verteilfunktion $F(x)$ nicht geschlossen darstellbar. Wenn X_1, X_2 Unabhangig:

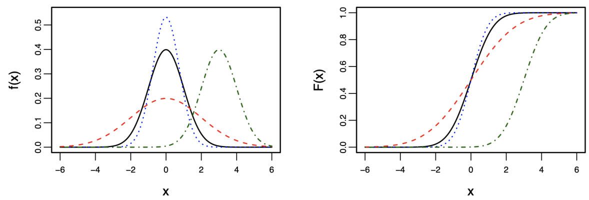
$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2), \quad X_1 \pm X_2 \sim \mathcal{N}(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$$

Bei der $\mathcal{N}(0,1)$-Verteilung gilt:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\} \quad \text{and} \quad \Phi(x) = \int_{-\infty}^{x} \varphi \tilde{x} d\tilde{x}$$

$$F(x) = \mathbb{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(u)$$

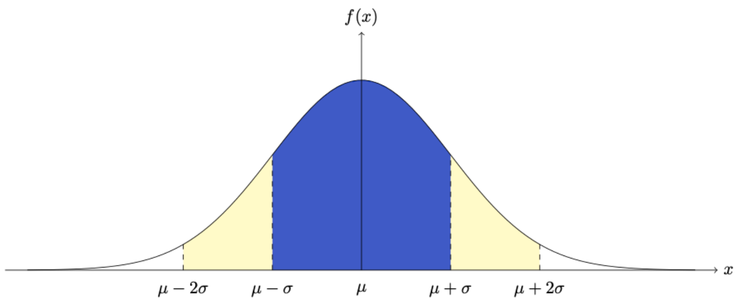
$$f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$



Eigenschaften:

- $\mathbb{P}(X \geq x) = 1 - \Phi(u)$
- $\Phi(u) = 1 - \Phi(-u)$
- $\mathbb{P}(a \leq X \leq b) = F(b) - F(a)$
- $\mathbb{P}(|X - \mu| \leq k \cdot \sigma) = 2\Phi(k) - 1$
- Normalapproximation:

$$\mathbb{P}[\underbrace{s}_{\sim\mathcal{N}(\mu,\sigma^2)} \leq k] = \mathbb{P}\left[\underbrace{\frac{s-\mu}{\sigma}}_{=\tilde{s}\sim\mathcal{N}(0,1)} \leq \frac{k-\mu}{\sigma}\right] = \Phi\left(\frac{k-\mu}{\sigma}\right)$$



Note: Ca. 66% der Flache befindet sich im Intervall $[\mu - \sigma, \mu + \sigma]$, ca. 95% der Flache im Intervall $[\mu - 2\sigma, \mu + 2\sigma]$.

3.6 Transformationen $Y = g(X)$:

Linear: $g(x) = a + b \cdot x, \quad \rightarrow \quad Y = a + b \cdot X$

$$\mathbb{E}[Y] = a + b \cdot \mathbb{E}[X] \qquad Var(Y) = (|b|)^2 \cdot Var(X)$$

$$F_Y(y) = F_X(\frac{x-a}{b}), \quad \text{falls } b > 0$$

$$F_Y(y) = 1 - F_X(\frac{x-a}{b}), \quad \text{falls } b < 0$$

$$f_Y(x) = \frac{1}{|b|} f_x(\frac{x-a}{b})$$

Standardisierung:

$$g(x) = \frac{x-\mu}{\sigma}, \qquad \mathbb{E}[Z] = 0, \qquad Var(Z) = 1$$

Allgemein (monoton steigend):

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

$$Var(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$$

$$f_Y(y) = F'_Y(y) = \left|\frac{1}{g'(g^{-1}(b))}\right| \cdot f_X(g^{-1}(y))$$

Quantile transformieren einfach mit: $q_{\alpha,X} \Rightarrow q_{\alpha,Y} = g(q_{\alpha,X})$

Example: Verteilungsfunktion von eine Transformation

Sei: $f_Z(z)$ fur $z > 0$ definiert. Die Verteilungsfunktion von $\ln(1+Z)$ ist gleich (fur $t > 0$, sonst $F(t) = 0$):

$$F(t) = \mathbb{P}(\ln(1+Z) \leq t) = \mathbb{P}(Z \leq e^t - 1) = \int_0^{e^t-1} f_Z(z) dz$$

Lognormalverteilt:

$$Y \sim LOG(\mu, \sigma^2) \leftrightarrow log(Y) \sim N(\mu, \sigma^2)$$

$$Y = e^x \text{ aus } X \sim N(\mu, \sigma^2)$$

Es gilt:

- $\mathbb{E}[Y] = \exp(\frac{\mu+\sigma^2}{2})$
- $\mathbb{E}[\ln(X)] = \mu$
- $P[a < X < b] = P[\ln a < \ln X < \ln b] = \Psi\left(\frac{\ln b - \mu}{\sigma}\right) - \Psi\left(\frac{\ln a - \mu}{\sigma}\right)$

Paretoverteilt:

$$Y = e^x \text{ aus } X \sim EXP(\alpha)$$

Es gilt:

- $P[X \geqslant x] = x^{-\alpha}$
- $f_y(x) = \alpha \cdot x^{-(\alpha+1)}$
- $F_y(x) = 1 - x^{-\alpha}$

3.7 Simulation von Verteilungen:

Sei $X \sim U(0,1)$ und $Y = F^{-1}(X)$ mit beliebigem F!
Y hat gerade die Verteilungsfunktion F.

1. $X \sim U(0,1) \Rightarrow A = \{x_1, x_2, \ldots, x_n\}$
2. $F(Y) = F^{-1}(X) = F^{-1}(A)$

3.8 Übersicht stetiger Verteilungsfunktionen:

Verteilung	$p(x)$	W_X	$\mathbb{E}[X]$	$\overset{\sigma^2}{Var(X)}$
UNI(a,b)	$\frac{1}{b-a}$	$[a,b]$	$(a+b)/2$	$(b-a)^2/12$
EXP(λ)	$\lambda e^{-\lambda x}$	$\mathbb{R}_+$	$1/\lambda$	$1/\lambda^2$
Gamma (α,λ)	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$	$\mathbb{R}_+$	α/λ	α/λ^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

4 Deskriptive Statistik

4.1 Kennzahlen:

Arithmetisches Mittel, abhängig von Ausreissern:

$$\overline{x} = \frac{1}{n}(x_1 + \ldots x_n)$$

Empirische Varianz:

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

Empirische Kovarianz:

$$s_{xy} = \frac{1}{n-1}\sum_{n-1}^{n}(x_i - \overline{X})(y_i - \overline{Y})$$

Empirischer Korrelationskoeffizient:

$$r = \frac{s_{xy}}{s_x \cdot s_y} \in [-1,1], \qquad \begin{matrix} r = +1 \Leftrightarrow y_i = a + bx_i, b > 0 \\ r = -1 \Leftrightarrow y_i = a + bx_i, b < 0 \end{matrix}$$

$sign(r)$ gibt Richtung,$|r|$ die Stärke der linearen Abhängigkeit.

Empirische kumulative Verteilungsfunktion:

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

Empirisches α-Quantil q_α, $(0 < \alpha < 1)$:
Bei **geordneten Werten** $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$ ist $q_\alpha = x_{(k)}$, wobei mit k die kleinste ganze Zahl, so dass $x_{(k)} > \alpha n$.
Die Werte werden also etwa im Verhältnis $\alpha : (1-\alpha)$ aufgeteilt.
Das Quantil ist *robust gegenüber Ausreissern*.

$$q_\alpha = \begin{cases} \frac{1}{2}\left(x_{(\alpha \cdot n)} + x_{(\alpha \cdot n + 1)}\right), & \text{if } n \text{ gerade} \\ x_{(\lceil \alpha \cdot n \rceil)}, & \text{else} \end{cases}$$

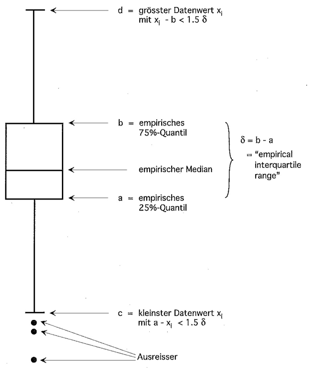
Median	$q_{0.5}$
unteres Quartil	$q_{0.25}$
oberes Quartil	$q_{0.75}$
Quartilsdifferenz	QD $= q_{0.75} - q_{0.25}$ robuste Kennzahl für Streuung

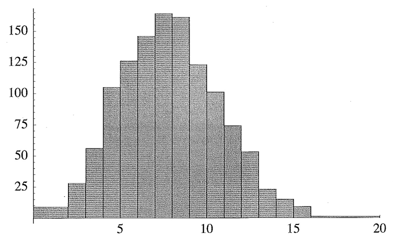
4.2 Histogramm und Boxplot:

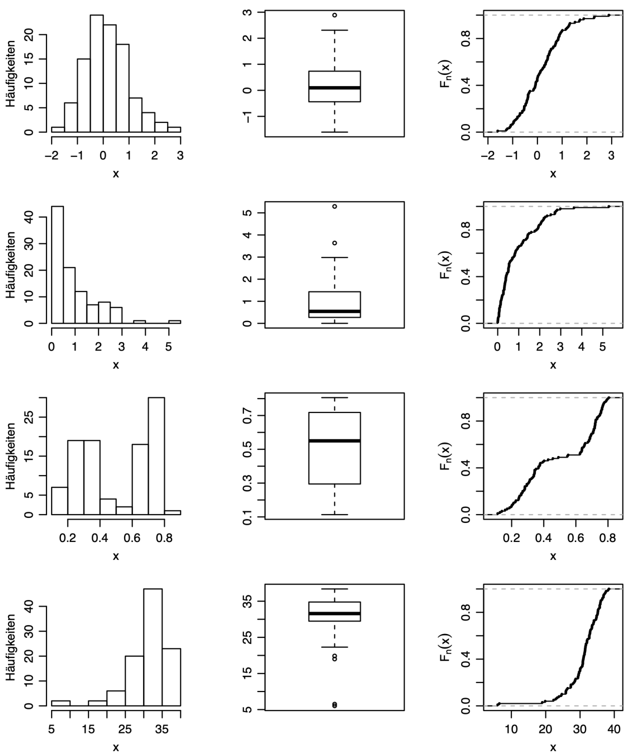
Histogramm: Einteilung der Werte in Klassen (Gesamtflaeche Histogramm = 1)

- Intervalle $(c_{k-1}, c_k]$
- Fläche der Balken $\sim h_k$ mit h_k: Häufigkeit, Anzahl Werte im Intervall
- Höhe der Balken $\sim \frac{h_k}{c_k - c_{k-1}}$

Boxplot: Rechteck begrenzt durch das 25%- und 75%-Quantil und dickem Strich für den Median, Linien von grösstem bis kleinstem "'normalen"' Wert (max. 1.5 mal die Quartilsdifferenz), Ausreisser: Sterne







5 Mehrdimensionale Verteilungen

5.1 Diskret:

Wahrscheinlichkeitsfunktion

$$\mathbb{P}(X = x, Y = y) \rightarrow \text{Tabelle}$$

$$\mathbb{P}(X \leq x, Y \leq y) = \sum_{\substack{x_i \in W_x | x_i \leq x \\ y_i \in W_y | y_i \leq y}} \mathbb{P}(X = x_i, Y = y_i)$$

Randverteilungen: WK'funktionen

$$\mathbb{P}(X = x) = \sum_{y \in W_Y} \mathbb{P}(X = x, Y = y)$$

$$\mathbb{P}(Y = y) = \sum_{x \in W_X} \mathbb{P}(X = x, Y = y)$$

Zweidimensionale Verteilungstabelle

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_k \leq y} p(x_i, y_k)$$

Bedingte Verteilungen: WK'funktionen

$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

$$\mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)}$$

Note: Die Randverteilung von X (gleich fuer Y) ist also:

$$\mathbb{P}(X = x) = \sum_{y \in W_Y} \mathbb{P}(X = x | Y = y) \mathbb{P}(Y = y)$$

Erwartungswert von $g(X, Y)$, $(g : \mathbb{R}^2 \rightarrow \mathbb{R})$

$$\mathbb{E}[g(X, Y)] = \sum_{x \in W_x, y \in W_y} g(x, y) \mathbb{P}(X = x, Y = y)$$

Bedingter Erwartungswert

$$\mathbb{E}[X | Y = y] = \sum_{x \in W_x} x \cdot \mathbb{P}(X = x | Y = y)$$

$$\mathbb{E}[Y | X = x] = \sum_{y \in W_y} y \cdot \mathbb{P}(Y = y | X = x)$$

Note: Wenn X, Y unabhängig $\Leftrightarrow$ (alle Aussage equivalent)

- $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \forall x \in W_x, y \in W_y$
- $\mathbb{P}(X + Y \in [a, b]) = \sum_{n \in [a,b]} \sum_{k=0}^{n} \mathbb{P}(X = k) \cdot \mathbb{P}(Y = n - k)$
- $\mathbb{P}(X = x | Y = y) = \mathbb{P}(X = x), \forall x \in W_x, y \in W_y$
- $\mathbb{P}(Y = y | X = x) = \mathbb{P}(Y = y), \forall x \in W_x, y \in W_y$

5.2 Stetig:

$$\mathbb{P}[(X, Y) \in A] = \int\int_A f_{X,Y}(x, y) dx dy$$

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

Dichte (oder gemeinsame Verteilung):

$$f(x, y), \qquad f : \mathbb{R}^2 \rightarrow \mathbb{R}$$

Randverteilungen: Dichten

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Bedingte Verteilungen: Dichten

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

Erwartungswert von $g(X, Y)$, $(g : \mathbb{R}^2 \rightarrow \mathbb{R})$

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y) dx dy$$

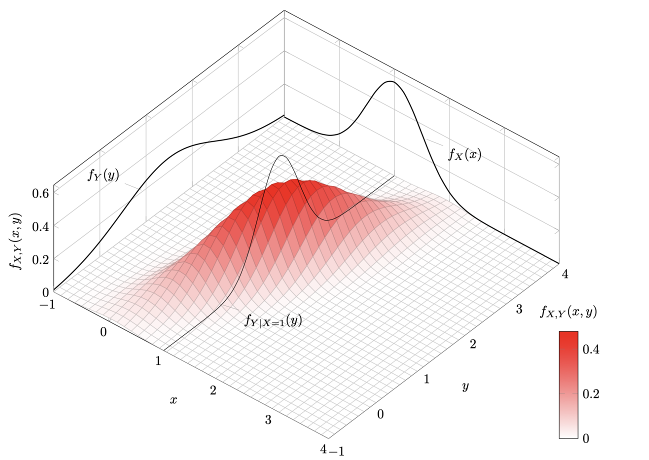
Bedingter Erwartungswert

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

$$\mathbb{E}[Y | X = x] = \int_{-\infty}^{\infty} y f_Y(y | X = x) dy$$

Note: Wenn X, Y unabhängig $\Leftrightarrow$ (alle Aussage equivalent)

- $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \qquad F(x_1, x_2) = F(x_1) \cdot F(x_2)$
- $f_X(x | Y = y) = f_X(x), \forall x \in W_x, y \in W_y$
- $f_Y(y | X = x) = f_Y(y), \forall x \in W_x, y \in W_y$

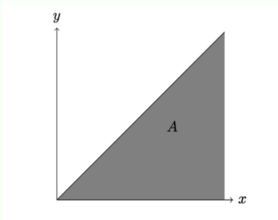


Example: Machinen Lebensdauer

Seien X und Y unabhangig und $Exp(\lambda_i)$ verteilt. So ist:

$$f_{X,Y}(x,y) = \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 x}$$

$$\mathbb{P}(Y < X) = \int_0^\infty \int_0^x f_{X,Y}(x,y)dydx$$



Example: Wurfel und Mutze Experiment

Experiment: Zuerst Wurfel einmal geworfen (Resultat = Y). Dann wird Y-mal Mutze geworfen. X bezeichne die Anzahl "Kopf" aus den Y Mutzenwurfen.

$$\mathbb{P}(X = k|Y = j) = \begin{cases} BIN(j, 0.5)(k), & k \in \{0, \dots, j\} \\ 0, & k > j \end{cases}$$

Note:

- $\mathbb{P}(Y = j) = \frac{1}{6}$
- $\mathbb{P}(X = k) = \sum_{i=k}^{j} \mathbb{P}(X = k|Y = i) \cdot \mathbb{P}(Y = i)$
- $\mathbb{E}(X) = \sum_{k=1}^{6} k\mathbb{P}(X = k) = \sum_{k=1}^{6} \sum_{j=1}^{6} k\mathbb{P}(X = k|Y = j)\mathbb{P}(Y = j)$

5.3 Transformationen:

Seien X und Y **unabhängige** ZV mit Dichten f_X, f_Y.

Summe: $f_{X+Y}(t) = f_X(t) * f_Y(t) = \int_{\mathbb{R}} f_X(u) \cdot f_Y(t-u)\, du$

$$\mathbb{P}(X + Y \le z) = \int_0^z \int_0^t f_X(u) \cdot f_Y(t-u)dudt$$

Produkt: $f_{X_1 \cdot X_2}(z) = \int\limits_{X_1 \cdot X_2 = z} \frac{1}{|t|} f_{X_1}(t) \cdot f_{X_2}\left(\frac{z}{t}\right)\, dt$

Quotient: $f_{X_1/X_2}(z) = \int\limits_{X_1/X_2 = z} |t| f_{X_1}(z \cdot t) f_{X_2}(t)\, dt$

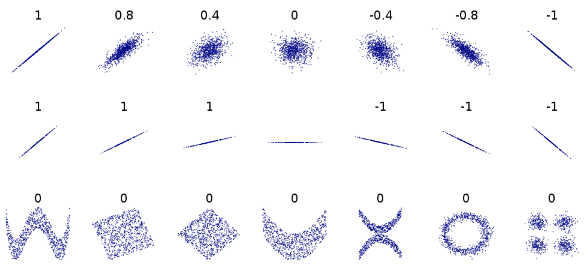
5.4 Kovarianz und Korrelation:

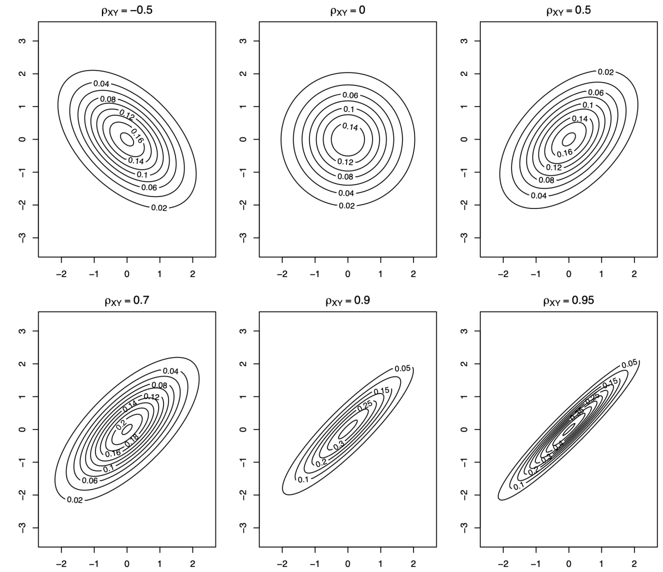
Kovarianz und Korrelation sind Kennzahlen, welche die Abhängigkeit von Zufallsvariabeln beschreiben.

Kovarianz: $Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$

Die Korrelation misst Starke und Richtung der **linearen Abhangigkeit** zwischen X und Y.

$$Corr(X,Y) = \rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$





Note: Im Bild stehen Beispiele fur empirische Korrelationen

Example: Kovarianz aus Wahrscheilichkeitstabelle

Gegeben sei die Tabelle der Waahscheinlichkeit $\mathbb{P}(X = k, Y = j)$ fuer $k, j = 0, 1$. Dann ist:

- $\mathbb{E}[X] = \mathbb{P}(X = 1) = \mathbb{P}(X = 1|Y = 0) + \mathbb{P}(X = 1|Y = 1)$
- $\mathbb{E}[Y] = \mathbb{P}(Y = 1) = \mathbb{P}(X = 0|Y = 1) + \mathbb{P}(X = 1|Y = 1)$
- $Cov(Y_1, Y_2) = \underbrace{\mathbb{E}[XY]}_{=\mathbb{P}(X=1|Y=1)} - \mathbb{E}[X]\mathbb{E}[Y]$

5.4.1 Rechenregeln::

- $\mathbb{E}[a_0 + \sum_{i=1}^n a_i X_i] = a_0 + \sum_{i=1}^n a_i \mathbb{E}[X_i]$
- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$
- $Var(\sum_{i=1}^n a_i Xi) = \sum_{i=1}^n a_i^2 Var(X_i) +$

  $+2\sum_{i \le j}^n sign(a_i) sign(a_j) Cov(a_i X_i, a_j X_j)$
- $Cov(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j)$
- $Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[Y] \cdot \mathbb{E}[X]$
- $Cov(a + bX, c + dY) = b \cdot d \cdot Cov(X,Y)$
- $Cov(Z+X, Z+Y) = Cov(Z,Z) + Cov(Z,Y) + Cov(X,Z) + Cov(X,Y)$
- $Corr(a + bX, c + dY) = sign(b) \cdot sign(d) \cdot Corr(X,Y)$
- $Corr(X,Y) = +1$ iff $Y = a + bX$ fur $b > 0$
- $Corr(X,Y) = -1$ iff $Y = a + bX$ fur $b < 0$

Wenn X,Y unabhangig:

- $Corr(X,Y) = Cov(X,Y) = 0$

5.4.2 Korrelation - Unabhängigkeit:

korreliert	$\Rightarrow$	abhängig
abhängig	$\not\Rightarrow$	korreliert
unabhängig	$\Rightarrow$	unkorreliert
unkorreliert	$\not\Rightarrow$	unabhängig
korreliert	$\Leftrightarrow$	Kovarianz $\neq 0$
unkorreliert	$\Leftrightarrow$	Kovarianz $= 0$
Korrelation $= 0$	$\Leftrightarrow$	$Cov(X,Y) = 0$

Ausnahme: Für 2D-Norm, Ver.: abhängig $\Leftrightarrow$ korreliert

5.5 Zwei-dimensionale Normalverteilung:

Kovarianz-Matrix:

$$\Sigma = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}$$

Normalverteilung:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \Sigma\right)$$

Dichte:

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

5.6 Lineare Prognose:

Lineare Prognose von Y gestutzt mit Ansatz: $\widehat{Y} = a + bX$:

$$\widehat{Y} = \mu_Y + \frac{Cov(X,Y)}{Var(X)}(X - \mu_X)$$

$$\mathbb{E}[(Y-\widehat{Y})^2] = (1-\rho_{XY}^2) \cdot Var(Y)$$

6 Grenzwertsätze

Note: Die n-fache Wiederholung eines Zufallsexperimentes ist selber wieder ein Zufallexperiment.

6.1 Independent and identically distributed:

- Ergebnisse im ursprünglichen Experiment $A_1, ..., A_n$ sind unabhangig
- $\mathbb{P}[A_1] = \ldots = \mathbb{P}[A_n] = \mathbb{P}[A]$ gleiche WK
- Zufallsvariabeln $X_1, ..., X_n$ sind **unabhangig**
- alle X_i haben **dieselbe Verteilung**

Mit der i.i.d. Annahme gilt:

1. $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot P[B]$
2. $F_{x_1} = F_{x_n} \to \mathbb{P}(x_1 \leqslant t) = \mathbb{P}(x_n \leqslant t)$
3. $\mathbb{E}[X_1 \cdot X_2] = \mathbb{E}[X_1] \cdot \mathbb{E}[X_2]$

6.2 Funktionen von Zufallsvariablen:

Anstelle von $X_1, ...X_n$ werden neue Zufallsvariabeln als Funktion der alten gebildet.

- Summe: $S_n = X_1 + ... + X_n$
- arithmetisches Mittel: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i = \frac{S_n}{n}$ (keine stetige Variable $\forall n$)

Sonderflälle von S_n: bei denen die Bestimmung einfach ist:

1. Wenn $X_i \in \{0,1\}$ (Bernoulliexperiment), dann ist $S_n \sim BIN(n,p)$ mit $p = P[X_i = 1]$
2. Wenn $X_i \sim POI(\lambda)$, dann ist $S_n \sim POI(n \cdot \lambda)$
3. Wenn $X_i \sim N(\mu, \sigma^2)$, dann ist $S_n \sim \mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$
4. Wenn $X_1, ..., X_n$ i.i.d und Normalverteilt mit $\mu_1, ..., \mu_n$ und $\sigma_1, ..., \sigma_n$, dann ist $S_n \sim \mathcal{N}(\sum \mu, \sum \sigma^2)$

6.2.1 Rechenregeln:

Streuung der Summe nimmt zu:

- $\mathbb{E}[S_n] = n \cdot \mathbb{E}[X_i]$
- $Var(S_n) = n \cdot Var(X_i)$ $\Rightarrow \sigma_{S_n} = \sqrt{n} \cdot \sigma_{X_i}$

Streuung des arithm. Mittels nimmt ab:

- $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i]$
- $Var(\overline{X}_n) = Var(X_i)/n$ $\Rightarrow \sigma_{\overline{X}_n} = \sigma_{X_i}/\sqrt{n}$

Note: $\uparrow n \Rightarrow$ "'dünnerer"' Dichte: $\lim_{n\to\infty} Var(\bar{X}_n) = 0$

6.3 GGZ und ZGWS:

Theorem: Gesetz der Groessen Zahlen (GGZ)

Seien $X_1, ..., X_n$ i.i.d mit μ
Arithmetische Mittel strebt gegen den Erwartungswert:

$$\overline{X}_n \to \mu \ (n \to \infty)$$

Dichte strebt gegen die Wahrscheinlichkeit von A:

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

Theorem: Zentraler Grenzwertsatz (ZGWS)

Seien $X_1, ..., X_n$ i.i.d. mit μ und σ^2, dann für grosse n:

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

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

Note: $Z = \frac{1}{\sqrt{n}}\frac{S_n - n\mu}{\sigma} \sim \mathbb{N}(0,1) \Rightarrow (1-\alpha) \times \%$ Konfidenzintervall:

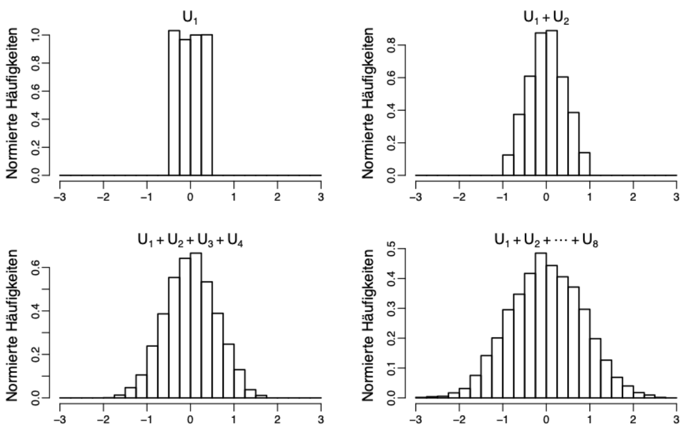
$$1 - \alpha = \mathbb{P}(\underbrace{-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}}_{Intervall})$$

Theorem: Chebychev Ungleichung

Sei X eine Zufallsvariable mit Erwartungswert μ und und endlicher Varianz σ, Dann gilt $\forall k > 0$:

$$\mathbb{P}(|\overline{X}_n - \mu| \geq k) \leqslant \frac{\sigma^2}{k^2} \text{ oder } \mathbb{P}(|X-\mu| < k) \geq 1 - \frac{\sigma^2}{k^2}.$$

Mit dieser ist man stets auf der sicheren Seite, dafur aber meistens ziemlich grob.



Example: Casinò Roulette

In ein Casinò mit 25 Roulette (37 Zahlen), $10^4 \frac{\#\text{"rot"}}{\text{month}}$.

1. Erwartete gewinn des Casinò?
2. Wahrscheinlichkeit das Casinò am ende Monat mehr als 4000.- Gewinn hat?
3. Wievel mal muss 1.- auf "rot" gesetzt damit durchschlittliche Gewinn Casinò per Wette mit 95% Wahrscheilichkeit 2 cent oder mehr ist?

Solution:

1. Sei X_i das Gewinn des Casinò bei i-ten Spiel:

$$\begin{cases}\mathbb{P}(X_i = 1) = \frac{19}{37} \\ \mathbb{P}(X_i = -1) = \frac{18}{37}\end{cases} \quad \begin{cases}\mathbb{E}(X_i) = \frac{19}{37} - \frac{18}{37} = \frac{1}{37} \\ Var(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = 1 - \frac{1}{37^2}\end{cases}$$

Gewinn Casinò $= S_n \Rightarrow \mathbb{E}(S_1) = \mathbb{E}(X_1) \cdot n$

2. $S_n \sim \mathcal{N}(\mathbb{E}(X_1) \cdot n, Var(X_1) \cdot n) \Rightarrow \mathbb{P}(S_n > 4000) = \ldots$ normalisieren
3. $\mathbb{P}(\bar{X}_n > 0.02) \geq 0.95$, mit $\bar{X}_n \sim \mathcal{N}(\mathbb{E}(X_1), \frac{Var(X_1)}{n})$

7 Parameterschatztungen

7.1 Wahl der Verteilungsfamilie (QQ-Plot):

Q-Q-Plot: "'Quantil-Quantil-Plot"', es wird das theoretische Quantil gegenüber dem empirischen Quantil aufgetragen. Aussage möglich, wie gut die Daten dem Modell entsprechen. Wir definieren:

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

die entsprechenden empirischen $(\alpha \times 100)\%$ - Quantile.
Note: Die Beobachtung $x_{(k)}$ gerade das entsprechende empirische Quantil. Das entsprechende (theoretische) $(\alpha \times 100)\%$ - Quantile ist gegeben durch $F^{-1}(\alpha_k)$.

Falls die Daten wirklich von unserer Modellverteilung generiert wurden, sollten die empirischen Quantile ungefahr den theoretischen Quantilen entsprechen (**Gerade auf ungefahr Winkelhalblierende!**).

- **Vorteil** von QQ-Plot: die Abweichungen am Rand der Verteilung sind einfacher zu erkennen als z.B. bei einem Histogramm.
- **Nachteil** von QQ-Plot: die Parameter der Modellverteilung schon kennen mussen, sonst konnen wir die (theoretischen) Quantile nicht berechnen!

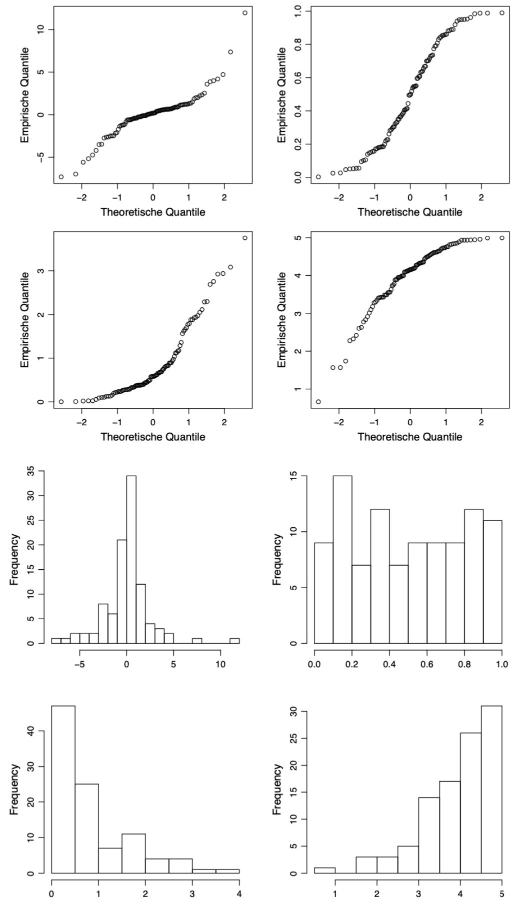
7.1.1 Exponential QQ-Plot:

Angenommen wird ein F gemass 3.4: im Plot ist zu sehen eine Gerate mit Steigung $\frac{1}{\lambda}$ durch Nullpunkt.

7.1.2 Normalverteilt QQ-Plot:

QQ-Plot, bei dem die Modellverteilung F die Standardnormalverteilung $\mathcal{N}(0, 1)$ ist. Das Plot (wenn Normalverteilt) ist eine Gerade mit Steigung σ und y-Achsenabschnitt μ.
Falls der Normalplot keine schone Gerade zeigt, kann man trotzdem etwas uber die zugrunde liegende Verteilung sagen:

- **Langschwanzigen Verteilung** (verglichen mit einer Normalverteilung): die empirischen Quantile in einem Bereich zwar auf einer Geraden liegen, aber im oberen Bereich nach oben und im unteren Bereich nach unten "wandern". Der QQ-Plot → "invertiertes S". Das bedeutet, dass das empirische 99% Quantil viel grosser ist als man von der Normalverteilung erwarten wurde. (1% grossten Werte der (empirischen) Verteilung sind grosser als von der Normalverteilung erwartet).
- **Kurzschwanzigen Verteilung** (verglichen mit einer Normalverteilung): gleich umgekehrt wie beim Langschwanzigen Verteilung: der QQ-Plot zeigt dann eine "S-Form".
- **Schiefe Verteilung** (verglichen mit einer Normalverteilung): zeigen sich durch "durchgebogene" Kurven. Bei recht schiefer Verteilung (Plot unten rechts) ist der empirische Median grosser als das arithmetische Mittel (umgekehrt beim links-schiefer)



Note: Im Bild stehen Beispiele fur QQ-Plots von langschwanzigen (oben links), kurzschwanzigen (oben rechts) und schiefen Verteilungen (unten).

7.2 Methode der Parameterschatzung:

Die Verteilung von X_i sei bekannt bis auf einen unbekannten Parameter θ, dabei kann θ auch mehrere Komponenten haben.

7.2.1 Momentenmethode:

Das k-te Moment von X ist definiert als:

$$\mu_k = \mathbb{E}[X^k]$$

Das k-te empirische Moment von $x_1, \ldots, x_n$ gegeben ist durch:

$$m_k = \hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

Achtung: Momentenmethode ist nicht eindeutig und nicht optimal und kann manchmal auch unsinnige Resultate liefern.

Vorgehen: Momenthenmethode

1. Unbekannte Parameter θ als Funktion der Momente $\mu_k = \mathbb{E}\left(X^k\right)$ schreiben
2. Vergleiche wahre μ_k durch empirische Momente m_k und lose fur $\hat{\theta}_i$

Note: Binomial: $\theta = p$, Poisson: $\theta = \mu$, Standardnormalverteilung: $\theta = (\mu, \sigma^2)^T$

Example: Anteil der kranken Population

Sei X_i eine Zufallsvariable, die 1 ist, falls die i-te Testperson positiv getestet ist und sonst 0. Sei A_i das Ereignis, dass die i-te Testperson die Krankheit bereits hatte. Gegeben sei:

- $m_1 = \frac{\#\text{Leute Positiv bei Test}}{\#\text{Leute}}$
- $p_1 = \mathbb{P}(X_i = 1|A_i)$ und $p_2 = \mathbb{P}(X_i = 1|A_i^c)$
- $\mathbb{P}(A_I) = q$

Gesucht ist das Schatzer q der Anteil der Kranken Leuten:

$$\mathbb{E}(X_i) = 1 \cdot \mathbb{P}(X_i = 1) = p_1 q + p_2(1-q)$$

Vergleichen von $\mathbb{E}(X_i)$ und m_1, das erste empirische Moment liefert die Losung.

7.2.2 Maximum-likelihood Methode (MLE):

Sei die Maximum-Likelihood definiert als:

$$\hat{\theta} = \text{argmax}_\theta \;\; L(\theta) = \text{argmax}_\theta \;\; l(\theta)$$

mit

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

und (log- Likelihoodfunktion)

$$l(\theta) = \ln(L(\theta)) = \sum_{i=1}^n \ln(p_X(x_i|\theta)) = \ln(p_X(x_1|\theta)) + \cdots + \ln(p_X(x_n|\theta))$$

Note: Fur stetige funktionen: $p_X(x_i|\theta) \Rightarrow f_X(x_i|\theta)$

Note: Bei Bernoulli: $\hat{p} = \frac{\sum x_i}{n}$, bei POI: $\hat{\lambda} = \frac{\sum x_i}{n}$, bei $UNI[0,\theta] : \hat{\theta} = \max\{x_1, \ldots, x_n\}$

7.2.3 Allgemeine Schatzer fur Erwartungswert und Varianz:

Gegeben seien $x_1, \ldots, x_n$ i.i.d Beobachtungen von einer ZV X. Die Schatzer lauten:

$$\hat{\mu}_X = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$$

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

Ein Schätzer $\hat{\theta}$ ist auch eine Zufallsvariable und heisst **erwartungstreu**, wenn $\mathbb{E}\left(\hat{\theta}\right)$ = wahrer Parameterwert.

Wir haben also, dass

- $\mathbb{E}\left(\hat{\mu}_X\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(X_i\right) = \frac{1}{n} \cdot n \cdot \mu = \mu \rightarrow$ erwartungstreu!
- $\mathbb{E}(\hat{\sigma}_X^2) = \sigma_X^2 \rightarrow$ erwartungstreu!
- $Var(\hat{\mu}_X) = \frac{1}{n}\sigma_X^2 \rightarrow$ nicht erwartungstreu!

Example: $X \sim BIN(n,p)$ Schatzer

Schätzer $\hat{p} = \frac{x}{n}$ (x: gemessene Erfolge)

$$\mathbb{E}\left(\hat{p}\right) = \mathbb{E}\left(\frac{x}{n}\right) = \frac{1}{n}\mathbb{E}\left(x\right) = \frac{1}{n} \cdot np = p$$

$\rightarrow$ erwartungstreu!

8 Statistische Tests und Vertrauensintervalle fur eine Stichprobe

8.1 Das Testproblem:

Nullhypothese: $H_0 : p = p_0$

Alternativen:

- $H_A : p \neq p_0$ (zweiseitig)
- $H_A : p > p_0$ (einseitig nach oben, rechtsseitig)
- $H_A : p < p_0$ (einseitig nach unten, linksseitig)

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

Fehler 1. Art: Verwerfen Nullhypothese, obwohl sie richtig ist.

$$\mathbb{P}(\text{Fehler 1. Art}) = \mathbb{P}_{p_0}(X \geq c) \leq \alpha$$

Fehler 2. Art: kein Verwerfen Nullhypothese, obwohl sie falsch ist.

$$\mathbb{P}(\text{Fehler 2. Art}) = \beta(\theta) \qquad \text{(siehe 8.2.1)}$$

Vorgehen: Statistischen Tests

1. Wahle ein geeignetes Modell fur die Daten und ein Signifikanzniveau α.
2. Lege die Nullhypothese $H_0 : \theta = \theta_0$ fest. θ bezeichnet hier allgemein einen Parameter in einem Modell.
3. Spezifiziere die Alternativehypothese:

$$\begin{aligned} H_A : &\theta \neq \theta_0 \text{ ("zweiseitig")} \\ &\theta > \theta_0 \text{ ("einseitig nach oben")} \\ &\theta < \theta_0 \text{ ("einseitig nach unten")} \end{aligned}$$

4. Konstruiere den Verwerfungsbereich K fur H_0, so dass gilt:

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

5. Betrachte, ob die Beobachtung x in den Verwerfungsbereich fallt.

8.2 Eigenschaften von statistischen Tests:

8.2.1 Macht eines Tests:

Ein statistischer Test kontrolliert direkt die Wahrscheinlichkeit eines Fehlers 1. Art durch das Signifikanzniveau α:

$$\mathbb{P}(\text{Fehler 1. Art}) = \mathbb{P}(\text{Test verwirft } H_0\text{, obwohl } H_0 \text{ stimt}) \leq \alpha$$

Note: Bei stetigen Verteilungen ist obige Ungleichung eine Gleichung, da wir das Niveau exakt kontrollieren konnen.

Wahrscheinlichkeit eines **Fehler 2. Art**:

$$\beta(\theta) = \mathbb{P}(\text{Test akzeptiert } H_0 \text{ obschon ein } \theta \in H_A \text{ stimmt})$$

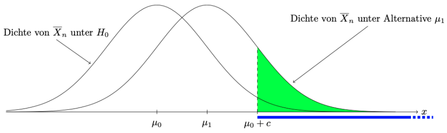
Die **Macht** eines Testes ist die Wahrscheinlichkeit, dass dieser für ein $p \neq p_0$ richtig verwirft:

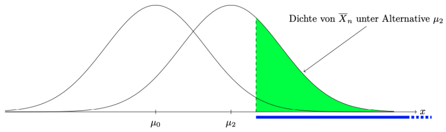
$$\text{Macht} = 1 - \beta(\theta) = 1 - \mathbb{P}(\text{Fehler 2. Art})$$

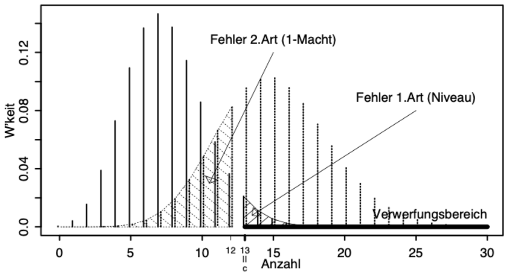
Note:

- Wenn Macht $\uparrow$, $\alpha \uparrow$, $\#n \uparrow$, $\mathbb{P}$(Fehler 1. Art) $\uparrow$, grosse $K \uparrow$, $\mathbb{P}(H_0$ verwerfen$) \uparrow$.

- Je kleiner der Unterschied zwischen p_0 und p, desto kleiner die Macht und desto schwieriger ist es für den Test richtig zu entscheiden.







8.2.2 P-Wert:

P-Wert = Wahrscheinlichkeit, unter der Nullhypothese einen mindestens so extremen Wert der Teststatistik zu beobachten (in Richtung der Alternative), wie der aktuell beobachtete.

P-Wert $< \alpha \rightarrow H_0$ verwerfen

Note:

- P-Wert ist grosser bei beidseitig Hypothese als bei einseitig.
- Verglichen mit dem reinen Testentscheid enthalt der p-Wert aber mehr Information, da man direkt sieht, "wie stark" die Nullhypothese verworfen wird.
- P-Wert von den Daten abhangt $\rightarrow$ zufallig $\rightarrow$ P-Wert $\sim UNI(0,1)$
- Der P-Wert ist insbesondere nicht die Wahrscheinlichkeit, dass die Nullhypothese stimmt!
- Ein (sehr) kleiner P-Wert bedeutet nicht zwangslaufig, dass ein fachlich relevantes Resultat gefunden wurde, da der P-Wert nichts uber eine Effektgrosse aussagt.
- Man muss wissen, ob der Test einseitig oder zweiseitig ist.

8.2.3 Multiples Testen:

Wir schreiben $H_{0,j}$ für die j-te Nullhypothese, $j = 1, \dots, m$. Mit m bezeichnen wir also die Anzahl Tests. Wenn wir annehmen, dass alle Nullhypothesen stimmen und alle Tests **unabhangig** voneinander sind, dann haben wir:

$$\begin{aligned}\mathbb{P}(\geq \text{ ein } H_{0,j} \text{ verworfen}) &= 1 - \mathbb{P}(\text{Kein } H_{0,j} \text{ verworfen}) \\ &= 1 - \mathbb{P}\left(\bigcap_{j=1}^{m} \{H_{0,j} \text{ nicht verworfen}\}\right) \\ &= 1 - \prod_{j=1}^{m} \mathbb{P}(H_{0,j} \text{ nicht verworfen}) \\ &= 1 - (1-\alpha)^m\end{aligned}$$

Wenn wir **keine Unabhangigkeit** annehmen, dann:

$$\begin{aligned}\mathbb{P}(\text{Mind. ein } H_{0,j} \text{ verworfen}) &= \mathbb{P}\left(\bigcup_{j=1}^{m} \{H_{0,j} \text{ verworfen}\}\right) \\ &= \sum_{j=1}^{m} \mathbb{P}(H_{0,j} \text{ verworfen}) \\ &= \alpha \cdot m\end{aligned}$$

Theorem: Bonferroni-Korrektur

Gegeben seien m Tests wo wir keine Unabhangigkeit annehmen: fur jeden einzelnen Test definieren wir das striktere Niveau $\alpha^* = \frac{\alpha}{m}$, dann:

$$\mathbb{P}(\text{Mind. ein } H_{0,j} \text{ verworfen}) \leq \alpha$$

Note: Universell gultig. Der Nachteil ist, dass man ein sehr striktes Niveau verwenden muss und daher Macht verliert im Gegensatz zu anderen Korrektur-Methoden. In der Praxis sollte man nur einen definierten Test durchfuhren, oder falls mehreren Tests , eine entsprechende Korrektur-Methode anwenden.

8.3 Vertrauens- /Konfidenzintervalle:

Ein Vertrauensintervall (oder Konfidenz-) I fur den Parameter θ zum Niveau $1-\alpha$ besteht aus allen Parameterwerten zum Signifikanzniveau α die vertraglich sind.
Note: Das Vertrauensintervall ist zufallig (hangt indirekt von unseren Beobachtungen ab, die wir als Realisierungen von Zufallsvariablen betrachten). Fur andere Realisierungen werden wir also ein (leicht) anderes Vertrauensintervall erhalten!

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

Hier ist I zufallig und θ fix. Das heisst, wenn wir ein Experiment viele Male wiederholen, dann fangt das Vertrauensintervall den wahren (unbekannten) Parameter im Schnitt in $(1-\alpha) \times 100\%$ der Falle ein.

Die Vertrauensintervalle enthalten auch eine Angabe uber die Genauigkeit der Parameterschatzung!

Testentscheid $\prec$ P-Wert $\prec$ Vertrauensintervall

wobei sich die Relation "$\prec$" auf den Informationsgehalt bezieht.

Zweiseitiges $(1-\alpha)$ Vertrauensintervall allgemein:

$$I = \hat{\theta} \pm \left(1 - \frac{\alpha}{2}\right)_{\text{Quantil}} \cdot \sqrt{Var(\hat{\theta})}$$

Rechtsseitig ($\theta_A > \theta_0$):

$$I = \in \left[\hat{\theta} - (1-\alpha)_{\text{Quantil}} \cdot \sqrt{Var(\hat{\theta})}\, , \, \infty\right)$$

Linksseitig ($\theta_A < \theta_0$):

$$I = \left(-\infty \, , \, \hat{\theta} + (1-\alpha)_{\text{Quantil}} \cdot \sqrt{Var(\hat{\theta})}\right]$$

8.3.1 Andere Verteilungen Vertrauensintervalle:

- $X \sim Poi(\lambda) \approx \mathcal{N}(\lambda, \lambda)$, $\hat{\theta} = \lambda$:

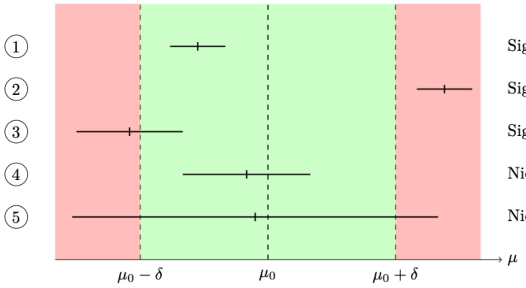
$$I = \lambda \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\lambda}$$

- Hypergeometrisch (Urnenmodell ohne Zurücklegen), n gross, N sehr gross:

$$I = p \pm z\left(1 - \tfrac{\alpha}{2}\right)\sqrt{\tfrac{p(1-p)}{n}}\sqrt{\tfrac{N-n}{N-1}}$$

8.3.2 Statistische Signifikanz und fachliche Relevanz:

Nehmen wir an, dass Abweichungen bis δ vom μ_0 keine Rolle spielen, also nicht relevant sind. Wir haben also einen "**irrelevanten Bereich**", der von $\mu_0 - \delta$ bis $\mu_0 + \delta$ geht. Ausserhalb sprechen wir vom "**Relevanzbereich**":



8.4 Binominalverteilte Test:

Wir haben, dass:

- $H_0 : p = p_0$
- $H_A : p > p_0$

Einseitig Test: Unter H_0 ist die Wahrscheinlichkeit, die Nullhypothese falschlicherweise abzulehnen (d.h. die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen) gegeben durch:

$$\boxed{\mathbb{P}(\text{Fehler 1 Art}) = \mathbb{P}_{p_0}[X \geqslant c] = \sum_{k=c}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k}}$$

Note: $\mathbb{P}(\text{Fehler 2. Art}) = 1 - \sum_{k=c}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k}$

Bei **Zweiseitige Alterhative** $H_A : p \neq p_0$ verwerfen die Nullhypothese für $c_2 \leq x \leq c_1$ wenn:

- $\mathbb{P}_{p_0}[X \leq c_1] = \sum_{k=0}^{c_1} \binom{n}{k} p_0^k (1-p_0)^{n-k} \leqslant \frac{\alpha}{2}$ and
- $\mathbb{P}_{p_0}[X \geq c_2] = \sum_{k=c_2}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} \leqslant \frac{\alpha}{2}$

die Verwerfungsbereich ist also gegeben in diesen Fall bei:

$$K = \{0, 1, 2, \ldots, c_1\} \cup \{c_2, c_2 + 1, \ldots, n\}$$

Note: Bin-Verteilung nur bei $p = 0.5$ symmetrisch!

Example: Muntze Wurf

Ich werfe eine Muntze 10 Mal und bekomme 8 mal Kopf, $\alpha = 0.05$.
Wir wissen, dass $H_0 : X \sim BIN(10, 0.5)$

- $\mathbb{P}_{H_0}(X = 9) + \mathbb{P}_{H_0}(X = 10) < \alpha$
- P-Wert $= \mathbb{P}_{H_0}(X = 8) + \mathbb{P}_{H_0}(X = 9) + \mathbb{P}_{H_0}(X = 10) > \alpha \Rightarrow K = \{9, 10\}$

Macht wenn in der Tat $p = 0.75 \rightarrow H_{wahr} : X \sim BIN(10, 0.75)$

$\mathbb{P}(X \in K) = \mathbb{P}(X \geq 9) = \mathbb{P}_{H_{wahr}}(X = 9) + \mathbb{P}_{H_{wahr}}(X = 10) = 0.24$

Die Macht der Test ist also $0.24 = \mathbb{P}(\text{significant Testresultat})$

Wenn n gross, $X \sim Bin(n, p) \approx \mathcal{N}(np, np(1-p))$.
Das $(1-\alpha) \times 100\%$ **Vertrauensintervall** ist gegeben als:

$$\boxed{I = \hat{p} \pm \Phi^{-1}(1 - \frac{\alpha}{2}) \sqrt{\hat{p}(1-\hat{p}) \frac{1}{n}}}, \hat{p} = \frac{x}{n}, x = \text{gemessene Erfolge}$$

Note: $Var(\hat{p}) = Var\left(\frac{x}{n}\right) = \frac{1}{n^2} Var(X) = \frac{\hat{p} \cdot (1-\hat{p})}{n}$

8.5 Tests Stichprobe bei normalverteilten Daten:

Gegeben seien n voneinander unabhangige Beobachtungen $x_1, \ldots, x_n$ einer ZV $X \sim \mathcal{N}(\mu, \sigma^2)$. Als Schatzer fur die unbekannten Parameter der Normalverteilung betrachten wir die erwartungstreu- en Schatzer:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

Wir fixieren je nach Problemstellung ein $\mu_0 \in \mathbb{R}$ unter Nullhypothese $H_0 : \mu = \mu_0$ gegen eine der moglichen Alternativen:

$$\begin{aligned} H_A : &\mu \neq \mu_0 \text{ ("zweiseitig")} \\ &\mu > \mu \text{ ("einseitig nach oben")} \\ &\mu < \mu_0 \text{ ("einseitig nach unten")} \end{aligned}$$

8.5.1 Z-Test:

Das Z-Test ist das Machtiger Test die wir machen konnen. Wir nehmen, dass $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, μ unbekannt, aber σ bekannt.

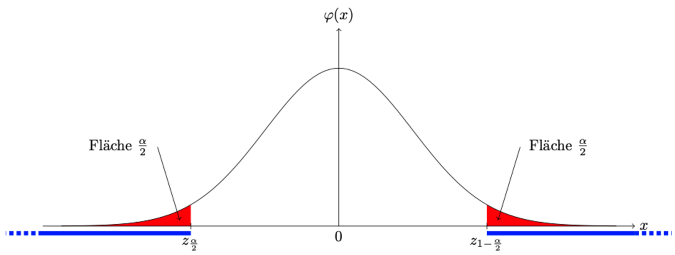
Vorgehen: Z-Test

Wir definieren also die **Teststatistik** als:

$$\boxed{Z = \frac{\bar{X}_n - \mu_0}{\sigma / \sqrt{n}}} = \frac{\text{beobachtet} - \text{erwartet}}{\text{geschatzter Standardfehler}} \sim \mathcal{N}(0, 1)$$

Wir verwerfen H_0 falls:

1. $H_A : \mu \neq \mu_0$:
   $|Z| \geq z_{1-\frac{\alpha}{2}} \iff Z \in K = (-\infty, -z_{1-\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}, \infty)$
2. $H_A : \mu > \mu_0$:
   $$Z \geq z_{1-\alpha} \iff Z \in K = [z_{1-\alpha}, \infty)$$
3. $H_A : \mu < \mu_0$:
   $Z \leq z_\alpha = -z_{1-\alpha} \iff Z \in K = (-\infty, z_\alpha] = (-\infty, -z_{1-\alpha}]$



Note:

- $z_{1-\alpha} = \Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$
- Im Bild: Dichtefunktion der Teststatistik Z mit Verwerfungsbereich (blau) des zweiseitigen Z-Tests zum Niveau α.
- $\mathbb{P}(\text{Fehler 1. Art}) = \mathbb{P}(|Z| \geq z_{1-\frac{\alpha}{2}}) = \alpha$
- Wenn wahre Wert $\mu = \mu_{wahr} \rightarrow Z \sim \mathcal{N}(\frac{\mu_{wahr} - \mu_0}{\sigma/\sqrt{n}}, 1) = \mathcal{N}(\tilde{\mu}, 1)$

$(1-\alpha) \times 100\%$ **Vertrauensintervall** fur μ_0 ist gegeben durch:

$$\boxed{I = \bar{X}_n \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2})}$$

Diese hat also die Form:

$$\text{Schatzung} \pm \text{Quantil} \cdot (\text{geschatzter Standardfehler})$$

Macht: Sei K das Verwerfungsbereich von Z. Wenn in Wahreit die Daten mit μ_{wahr} zentriert sind definieren wir die neue Teststatistik als:

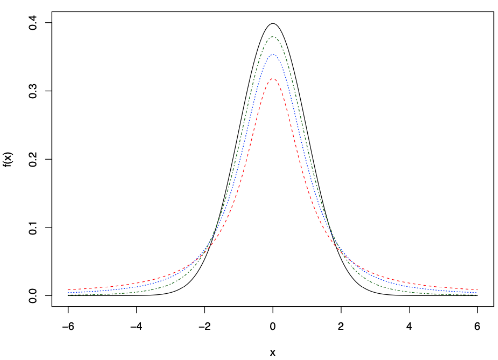
$$Z_A = \frac{\bar{X}_n - \mu_{wahr}}{\sigma/\sqrt{n}} = Z - \frac{\mu_{wahr} - \mu_0}{\sigma/\sqrt{n}}$$

So ist die Macht der Z-Test gegeben durch (am bsp fur einseitig unten Test):

$$\begin{aligned}
&\text{Macht} = \mathbb{P}(Z \in K|H_A) = \\
&= \mathbb{P}(Z \leq z_\alpha|H_A) = \mathbb{P}(Z_A \leq z_\alpha - \frac{\mu_{wahr} - \mu_0}{\sigma/\sqrt{n}}|H_A) = \\
&= \Phi(z_\alpha - \frac{\mu_{wahr} - \mu_0}{\sigma/\sqrt{n}}) = \Phi(z_\alpha - \tilde{\mu})
\end{aligned}$$

8.5.2 t-Test:

Das t-Test ist wenig machter als das Z-Test. Wir nehmen an, dass $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, μ, σ unbekannt.
Note; Die t-Verteilung ist wie die Standardnormalverteilung symmetrisch um 0, hat aber eher die Tendenz, grosse Werte anzunehmen (langschwanziger).



Vorgehen: t-Test

Die realisierte **Teststatistik** ist:

$$\tilde{T} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} \sim t_{n-1}, \qquad \text{unter } H_0,\ n-1 \text{ DOF}$$

Wir verwerfen H_0 falls:

1. $H_A : \mu \neq \mu_0$:

$$|\tilde{T}| \geq t_{n-1,1-\frac{\alpha}{2}} \iff \tilde{T} \in K = (-\infty, -t_{n-1,1-\frac{\alpha}{2}}] \cup [t_{n-1,1-\frac{\alpha}{2}}, \infty)$$

2. $H_A : \mu > \mu_0$:

$$\tilde{T} \geq t_{n-1,1-\alpha} \iff \tilde{T} \in K = [t_{n-1,1-\alpha}, \infty)$$

3. $H_A : \mu < \mu_0$:

$$\tilde{T} \leq t_{n-1,\alpha} = -t_{n-1,1-\alpha} \Leftrightarrow \tilde{T} \in K = (-\infty, \underbrace{t_{n-1,\alpha}}_{=-t_{n-1,1-\alpha}}]$$

$(1-\alpha)\cdot 100\%$ **Vertrauensintervall** fur μ_0 ist gegeben durch:

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

Zweiseitig P-Wert:

- P-Wert $= 2 \cdot \mathbb{P}(T \geq \tilde{T})$
  Note: Wenn P-Wert $> 2 \cdot \mathbb{P}(T \geq t_{n-1,1-\frac{\alpha}{2}}) = \alpha$ kann die Nullhypothese nicht verworfen werden.
- $\tilde{T} = t_{n-1,1-\frac{\alpha}{2}}$, lose fur α durch Interpolation $\Rightarrow \alpha =$ P-Wert

Macht: Sei K das Verwerfungsbereich von $\tilde{T}$. Wenn in Wahreit die Daten mit μ_{wahr} zentriert sind definieren wir die neue Teststatistik als:

$$\tilde{T}_A = \frac{\bar{X}_n - \mu_{wahr}}{S_n/\sqrt{n}} = \tilde{T} - \frac{\mu_{wahr} - \mu_0}{S_n/\sqrt{n}}$$

So ist die Macht der t-Test gegeben durch (am bsp fur zweisetig Test):

$$\begin{aligned}
&\text{Macht} = \mathbb{P}(\tilde{T} \in K|H_A) = \\
&= \mathbb{P}(\tilde{T} > t_{n-1,1-\frac{\alpha}{2}}|H_A) + \mathbb{P}(\tilde{T} < -t_{n-1,1-\frac{\alpha}{2}}|H_A) = \\
&= \mathbb{P}(\tilde{T}_A > t_{n-1,1-\frac{\alpha}{2}} - \frac{\mu_{wahr} - \mu_0}{S_n/\sqrt{n}}|H_A) + \ldots
\end{aligned}$$

Note: Die Wahrschenlichkeit $\mathbb{P}(\tilde{T}_A > \ldots|H_A) = 1 - \mathbb{P}(\tilde{T}_A \leq \ldots|H_A)$ lese ich von der Tabelle der t-Verteilung.

8.6 Tests Stichprobe bei nicht normalverteilten Daten:

In der Praxis sind unsere Messdaten nicht immer normalverteilt. Auch fur diese Situationen gibt es entsprechende Tests mit weniger starken Annahmen.

8.6.1 Vorzeichen-Test:

Wir betrachten allgemeiner Beobachtungen $x_1, \ldots, x_n$ von i.i.d ZV $X_1, \ldots, X_n \sim \mathcal{F}(\mu)$, wobei $\mathcal{F}$ eine beliebige stetige Verteilung mit Median μ ist.
Wir stellen die Null- und Alternativhypothese bzgl. dem Median:

- $H_0 : \mu = \mu_0$
- $H_A : \mu \neq \mu_0$

Note: Bei symmetrischen Verteilungen entspricht der Median dem Erwartungswert. Das Vorzeichen Test hat auch eine sehr tiefe Macht!
Wenn $\mu = \mu_0$ tatsachlich stimmt, dann beobachten wir mit 50% Wahrscheinlichkeit einen Wert grosser als μ_0 (Definition des Medians). Wenn wir also allzu viele (oder zu wenige) Werte haben, die grosser als μ_0 sind, dann spricht dies gegen die Nullhypothese und fur die Alternative.

Vorgehen: Vorzeichen-Test

Wir betrachten:

$$Q = \#\text{Positive } X_i - \mu_0$$

Die Anzahl positiver Vorzeichen Q, folgt unter H_0 einer $BIN(n, 0.5)$-Verteilung $\rightarrow H_0 : p_0 = 0.5$

Wir verwerfen H_0 falls:

1. $H_A : p_A \neq 0.5$:

$$\text{P-Wert} = \mathbb{P}(X \leq n-Q) \cup \mathbb{P}(X \geq Q) \overset{p_0=0.5}{=} 2\cdot\mathbb{P}(X \geq Q) < \alpha$$

2. $H_A : p_A > 0.5$:

$$\text{P-Wert} = \mathbb{P}(X \geq Q) = \sum_{i=Q}^{n} BIN(n, 0.5)(i) < \alpha$$

3. $H_A : p_A < 0.5$:

$$\text{P-Wert} = \mathbb{P}(X \leq Q) = \sum_{i=0}^{Q} BIN(n, 0.5)(i) < \alpha$$

Note: Bin-Verteilung nur bei $p = 0.5$ symmetrisch!

Macht: siehe Beispiel 8.4

8.6.2 Wilcoxon-Test:

Kompromiss: setzt weniger voraus als t-Test, nutzt Daten aber besser aus als Vorzeichen-Test. Wir betrachten allgemeiner Beobachtungen $x_1, \dots, x_n$ von i.i.d ZV $X_1, \dots, X_n \sim \mathcal{F}(\mu)$, wobei $\mathcal{F}$ eine beliebige stetige **symmetrische** Verteilung mit Median μ ist.
Wir stellen die Null- und Alternativhypothese bzgl. dem Median:

- $H_0 : \mu = \mu_0$
- $H_A : \mu \neq \mu_0$

Vorgehen: Wilcoxon-Test

1. Wir definieren: $U_i = X_i - \mu_0$
2. Range bilden: $\text{Rang}(|U_i|) = k$
   Note: $k = 1$ bedeutet, dass $|U_i|$ den kleinsten Wert unter $|U_1|, \dots, |U_n|$ hat.
   Achtung: Wenn zwei $|U_i|$ gleich sind, teilt man die Range auf (siehe unten)
3. V_i ist Indikator, ob U_i positiv ist: $\begin{cases} V_i = 1, & \text{falls } U_i > 0 \\ V_i = 0, & \text{falls } U_i < 0 \end{cases}$
4. Verwerfung der Nullhypothese falls W zu gross, zu klein oder beides ist (Tabellen).

$$W = \sum_{i=1}^{n} Rang(|U_i|) \cdot V_i$$

Note:

- Der Wilcoxon-Test halt das Niveau α exakt, falls $\mathcal{F}$ um 0 symmetrisch und x_i i.i.d ist. Beim t- Test wird zwar das Niveau auch ungefahr eingehalten, falls die Daten nicht ganz normalverteilt sind, aber die Wahrscheinlichkeit ein Fehler 2.Art von t-Test ist oft viel grosser als Fehler 2.Art von Wilcoxon-Test
- Der Wilcoxon-Test ist in der Praxis dem t- oder Vorzeichen-Test vorzuziehen, da Daten meist nicht Normalverteilt sind.

Example: Gleiche Range bei Wilcoxon-Test

Wenn wir diese Folge von Beobachtungen hatten:

$$|X_3| < |X_2| = |X_5| < |X_6| < |X_1| = |X_4| = |X_8| < |X_7|$$

Dass wurde:

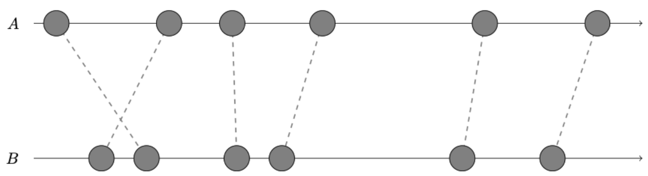
- $\text{Rank}(|X_3|) = 1$
- $\text{Rank}(|X_2|) = \text{Rank}(|X_5|) = \frac{2+3}{2} = 2.5$
- $\text{Rank}(|X_6|) = 4$
- $\text{Rank}(|X_1|) = \text{Rank}(|X_4|) = \text{Rank}(|X_8|) = \frac{5+6+7}{3} = 6$
- $\text{Rank}(|X_7|) = 8$

9 Vergleich zweier Stichproben

9.1 Gepaarte und ungepaarte Stichproben:

Ungepaarte Stichproben: Zufallig gewahlte Reihenfolge der Versuche, verschiedene Versuchseinheiten unter zwei verschiedenen Versuchbedingungen (**Randomisierung**). Einzelne Tests mussen nicht gleiche Stichprobengrosse haben.
Bsp: 2 Medikamente an verschiedenen Stichproben getestet.

Gepaarte Stichproben: beide Versuchsbedingungen an derselben Versuchseinheit getestet. Notwendigerweise mussen die beiden Stichprobengrossen gleich sein.
Bsp: 15 Prufkorper von 2 Labors ausgesmesen, Alte und neue Schuhe von den selben Atlhleten probiert.



Theorem: George Box Merkregel

Die Merkregel lautet:

"Block what you can, randomize what you cannot."

In dem Sinne sind also gepaarte Stichproben (falls realisierbar) unabhangigen Stichproben vorzuziehen.

Note: Sobald Menschen involviert sind $\Rightarrow$ Experiment wenn moglich doppelblind durchgefuhrt wird (weder die Person noch die Versuchsperson die Gruppenzugehorigkeit kennen). Dies ist wichtig, um den Effekt von Voreingenommenheit bei der Beurteilung auszuschalten.

9.2 Gepaarte Vergleiche:

Differenz innerhalb der Paare:

$$u_i = x_i - y_i$$

mit Zufallsvariablen i.i.d $U_1, \dots, U_n$.
Note: P-Wert kleiner als bei zwei Stichprobe t-Test, kein Unterschied zwischen Versuchsreihen: $\mathbb{E}[U_i] = 0$. Wir haben also:

- $H_0 : \mathbb{E}(U_i) = 0 \Rightarrow \mu_0 = 0$
- $H_A : \mathbb{E}(U_i) \neq 0 \Rightarrow \mu_0 \neq <> 0$

Verschiedene mogliche Tests:

- t-Test, wenn Normalverteilt und unabhangig, QQ-Plot
- Vorzeichen-Test, wenn beliebig verteilt
- Wilcoxon-Test, wenn symmetrisch verteilt, Histogramm

9.3 Zwei-Stichproben Test (ungepaarte Vergleiche):

Ungepaarte Stichproben, unabhangige i.i.d Zufallsvariablen.

9.3.1 Mit σ bekannt:

Annahmen: Normalverteilt mit gleicher Varianz, $\sigma_X = \sigma_Y = \sigma$

$$\left.\begin{array}{ll} X_i \overset{iid}{\sim} \mathcal{N}(\mu_X, \sigma^2), & i = 1, \dots, n \\ Y_i \overset{iid}{\sim} \mathcal{N}(\mu_Y, \sigma^2), & j = 1, \dots, m \end{array}\right\} \mu_X, \mu_Y, \sigma \text{ bekannt}$$

$$Var(\bar{X}_n - \bar{Y}_m) = \sigma^2 (\frac{1}{n} + \frac{1}{m})$$

Wir testen also zweiseitig:

- $H_0 : \mu_X = \mu_Y$
- $H_A : \mu_X \neq \mu_Y$

Wir definieren die Teststatistik als:

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

und machen einen **Z-Test** um die Nullhypothese zu verwerfen.
Note: $Z \sim \mathcal{N}\left(\frac{\mu_X - \mu_Y}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}, 1\right)$, falls $\mu_X = \mu_Y \Rightarrow Z \sim \mathcal{N}(0, 1)$

9.3.2 Mit σ unbekannt:

In der Praxis kennen wir σ nicht und so wir definieren:

$$S_{pool}^2 = \frac{1}{n+m-2}\left(\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 + \sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2\right)$$
$$= \frac{1}{n+m-2}\left((n_X-1)\cdot S_X^2 + (n_Y-1)\cdot S_Y^2\right)$$

Note: S_{pool} ist nichts anderes als ein gewichtetes Mittler der Schatzungen der Variablen.
Die Teststatistik lautet:

$$T = \frac{(\bar{X}_n-\bar{Y}_m)-(\mu_X-\mu_Y)}{S_{pool}\sqrt{\frac{1}{n}+\frac{1}{m}}} \quad \text{mit} \quad T \sim t_{n+m-2}$$

Note:

- Erinnerung: $-t_{n+m-2,1-\alpha} = t_{n+m-2,\alpha}$
- Wir mussen hier 2 DOF abziehen, weil die Parameter μ_X und μ_Y geschatzt haben.

Wie bei 9.3.1 wir testen fur $H_0 : \mu_X = \mu_Y$.

Vorgehen: Zwei Stichprobe t-Test

Wir definieren die Teststatistik als:

$$\tilde{T} = \frac{\bar{X}_n-\bar{Y}_m}{S_{pool}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

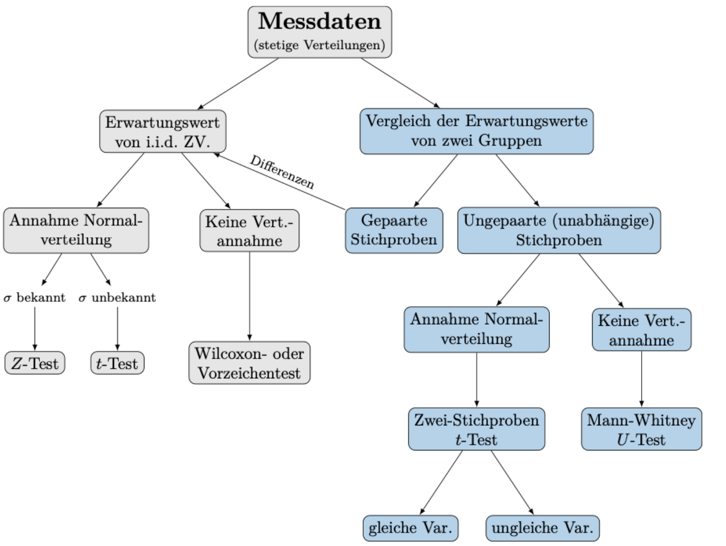
Wir verwerfen H_0 falls:

1. $H_A : \mu_X \neq \mu_Y$:
$|\tilde{T}| \geq t_{n+m-2,1-\frac{\alpha}{2}} \Leftrightarrow \tilde{T} \in K = (-\infty, -t_{n+m-2,1-\frac{\alpha}{2}}] \cup [t_{n+m-2,1-\frac{\alpha}{2}}, \infty)$
2. $H_A : \mu_X > \mu_Y$:
$\tilde{T} \geq t_{n+m-2,1-\alpha} \Longleftrightarrow \tilde{T} \in K = [t_{n+m-2,1-\alpha}, \infty)$
3. $H_A : \mu_X < \mu_Y$:
$\tilde{T} \leq t_{n+m-2,\alpha} = -t_{n+m-2,1-\alpha} \Leftrightarrow \tilde{T} \in K = (-\infty, \underbrace{t_{n+m-2,\alpha}}_{=-t_{n+m-2,1-\alpha}}]$

Das $(1-\alpha)\times 100\%$ **Vertrauensinterval** fur $d = \mu_X - \mu_Y$ ist:

$$I = \bar{X}_n - \bar{Y}_m \pm S_{pool}\sqrt{\frac{1}{n}+\frac{1}{m}} \cdot t_{n+m-2,1-\frac{\alpha}{2}}$$

Note: Enthaltet das Interval das Nullpunkt nicht, so verwerfen wir H_0 (fuer das zweiseitige Test),



10 Lineare Regression

10.1 Einfache lineare Regression:

Wir betrachten die vorliegenden Daten (x_i, y_i), $\quad i = \ldots, n$

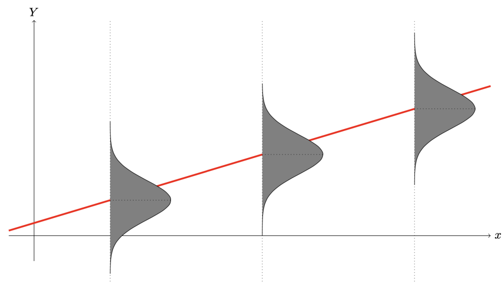
10.1.1 Modell:

Das einfache lineare Regressionsmodell lautet $(i = 1, \ldots n)$

$$Y_i = \beta_0 + \beta_1 x_i + E_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$$

Mit:

- Y_i ist die Zielvariable (unabhangig)
- x_i ist die erklarende Variable, typischerweise nicht-zufallig
- E_i i.i.d $\sim \mathcal{N}(0, \sigma^2)$, σ^2 = Storparameter



Note: Das Modell heisst "einfach", weil nur eine erklarende Variable vorhanden ist. Zudem heisst das Modell "linear", da die Parameter β_i linear in der Modellgleichung vorkommen.

10.1.2 Parameterschatzungen:

Maximum-Likelihood Schatzer

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2}(\frac{y_i-\beta_0-\beta_1 x_i}{\sigma})^2\}$$

Kleinste Quadrat Schatzer

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

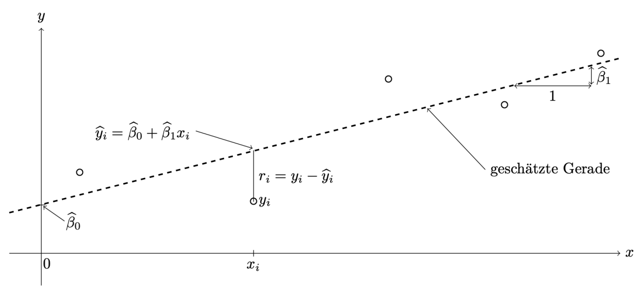
Wir definieren also:

$$\hat{\beta}_0, \hat{\beta}_1 = \text{argmin}_{\beta_0,\beta_1} L(\beta_0, \beta_1) = \text{argmin}_{\beta_0,\beta_1} S(\beta_0, \beta_1)$$

Es gilt so:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Note: Wir koennen auch $\hat{\beta}_1$ schreiben als $\hat{\beta}_1 = r\frac{s_y}{s_x}$ mit r, s_x, s_y aus 4.1.



Wir definieren also die **Residuen** als $r_i = y_i - \hat{y}_i$.
Daher schatzen wir die Varianz des Fehlerterms mit $\sigma^2 = \frac{1}{n-2}\sum_{i=1}^{n} r_i^2$

10.1.3 Tests und Vertrauensintervalle:

Es gilt:

$$\hat{\beta}_1 = \mathcal{N}\left(\beta_1, \frac{\sigma^2}{SS_X}\right)$$

$$\hat{\beta}_0 = \mathcal{N}\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SS_X}\right)\right)$$

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

Note: Die Parameterschatzer sind also Zufallsvariablen und erwartungstreu. Die Standardabweichung eines Parameterschatzers entspricht den **Standardfehler**.
Setzt man die Schatzung $\hat{\sigma}$ ein, so erhalt man den geschatzten Standardfehler.

Vorgehen: *t*-Test bei Lineare Regression

Wird definieren unsere Teststatistik als:

$$\boxed{\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{SS_X}} \sim t_{n-2}}$$

Wir haben also erneut eine eine Funktion der Form:

$$\frac{\text{beobachtet} - \text{erwartet}}{\text{geschatzter Standardfehler}}$$

Beim Test verwenden wir:

- $H_0 : \beta_1 = 0$: "Es gibt **keinen** linearen Zusammenhang zwischen x und y."
- $H_A : \beta_1 \neq 0$: "Es gibt **einen** linearen Zusammenhang zwischen x und y."

Wir berechnen:

$$1 - \alpha = \mathbb{P}\left(\frac{|\hat{\beta}_1 - \beta_1|}{\hat{\sigma}/\sqrt{SS_X}} \leq t_{n-2,1-\frac{\alpha}{2}}\right) =$$

$$= \mathbb{P}\left(|\hat{\beta}_1| \leq t_{n-2,1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{SS_X}}\right)$$

So, wir verwerfen H_0 falls:

$$\left|\frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{SS_X}}\right| \geq t_{n-2,1-\frac{\alpha}{2}}$$

$(1-\alpha) \times 100\%$ **Vertrauensintervall** fur β_1 ist gegeben durch:

$$\boxed{I = \hat{\beta}_1 \pm t_{n-2,1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{SS_X}}}$$

Diese hat also auch die Form:

$$\text{Schatzung} \pm \text{Quantil} \cdot (\text{geschatzter Standardfehler})$$

10.1.4 Computer Output Regression:

Einen computer liefert das folgendes Resultat:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.3681     0.7191   8.856 0.000305
tiefe       -13.6373     1.0111 -13.487 4.01e-05
```

Wir lesen ab:

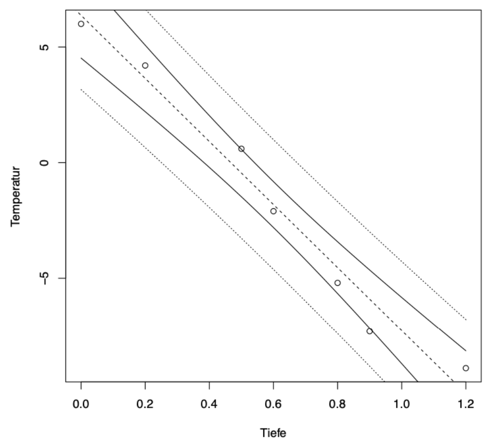
$$\begin{bmatrix} \hat{\beta}_0 & \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SS_X}\right) & t_{wert,\hat{\beta}_0} & \text{P-Wert}_{\hat{\beta}_0} \\ \hat{\beta}_1 & \frac{\sigma^2}{SS_X} & t_{wert,\hat{\beta}_1} & \text{P-Wert}_{\hat{\beta}_1} \end{bmatrix}$$

Note: Das "Error" column entspricht den Standardfehlern von $\hat{\beta}_0$ und $\hat{\beta}_1$. Das "$Pr(>|t|)$" column entspricht den P-Wert fur den zweiseitiger t Test.

10.1.5 Vertrauensintervalle fur den Erwartungswert und Prognoseintervalle:

Vertrauensintervall fur den wahren Wert der Geraden an der Stelle x: der Wert der Modellgerade adS x ist nichts anderes als der Erwartungswert der Zielgrosse, wenn wir ein x fixieren und ist gleich $\beta_0 + \beta_1 x$.

Prognoseintervall: Intervall, das mit hoher Wahrscheinlichkeit eine neue (zufallige) Beobachtung Y adS x einfangt. Im Verhaltnis zu Vertrauensintervall ist das Prognoseintervall ist immer breiter. Dies ist auch einleuchtend, denn es muss noch die Variabilitat einer Beobachtung "abdecken" (wegen Fehlerterm E). Beide Bander sind ubrigens gekrummt, beim Prognoseband ist dies nur viel schlechter sichtbar.



Note: Daten, geschatzte Gerade (gestrichelt), Vertrauensintervalle (durchgezogen) und Prognoseintervall (gepunktet) im Beispiel mit der Bohrung in Permafrostboden.

Annahmen:

1. $\mathbb{E}[E_i] = 0 \Rightarrow$ keinen systematischen Fehler im Modell (die Modellgleichung ist korrekt)
2. Die $E_1, \ldots, E_n$ sind i.i.d. Die Fehler sind also unabhangig voneinander und folgen der gleichen Verteilung (Varianz gleich gross).
3. Die $E_1, \ldots, E_n$ sind normalverteilt.

Note: Je deutlicher die Modellannahmen verletzt sind, desto weniger vertrauenswurdig sind die Resultate (p- Werte der Tests,Vertrauensintervalle,...). Bei einer nur "leichten" Verletzung der Modellannahmen sind die Resultate sicher noch brauchbar.

10.1.6 Typ von Plots:

Tukey-Anscombe Plot (TA-Plot)
Man plottet die Residuen r_i gegen $\hat{y}_i$ auf (bei linearen Regressio r_i auch gegen x_i aufzeichnen). Idealfall $\Rightarrow$ Streuung der Punkte um die x-Achse sehen. Mogliche Szenarien sind:

- Falls systematische Abweichungen von der x-Achse erkennbar sind, so spricht dies gegen die Annahme 1 von 10.1.5. Denn in diesem Fall gibt es Bereiche, wo der Fehler im Schnitt nicht 0 ist (z.B. falls man einen quadratischen Effekt im Modell vergessen hat).
- Ist die Streuung stark unterschiedlich (z.B. trichterformiges Bild), so spricht dies gegen Annahme 2 von 10.1.5.
- Ev. sieht man auch "Ausreisserpunkte".

Normalplot
Man erstellt einen Normalplot der Residuen. Es sollten keine groben Abweichungen von einer Geraden vorliegen, sonst ware dies eine Verletzung der Modellannahme 3 von 10.1.5.

Serial Correlation Plot
Um die Unabhangigkeit der $E_1, \ldots, E_n$ zu uberprufen, kann man die Residuen r_i gegen die entsprechende Beobachtungsnummer i aufzeichnen. Sinnvoll, wenn die Beobachtungen in dieser zeitlichen Reihenfolge aufgenommen wurden.
Im Idealfall sollte es keine Regionen geben, wo sich die Residuen ahnlich verhalten (d.h. wo z.B. alle positiv sind).

10.2 Multiple lineare Regression:

In der Praxis hat man oft nicht nur eine, sondern mehrere erklarende Variablen $x^{(1)}, \ldots, x^{(m)}, \quad m > 1$.

10.2.1 Modell:

Das multiple lineare Regressionsmodell ist gegeben durch:

$$Y_i = \beta_0 + \beta_1 x_i^{(1)} + \ldots + \beta_p x_i^{(m)} + E_i, \quad i = 1, \ldots, n$$

wobei wieder wie fruher E_i i.i.d $\sim \mathcal{N}(0, \sigma^2)$ angenommen wird.

Note:

- Bei $x^{(j)} \Rightarrow$ j-te erklarende Variable der i-ten Beobachtung.
- Total haben wir also $p = m + 1$ verschiedene β-Parameter.
- Der Parameter β_j ist der Effekt der erklarenden Variable $x^{(j)}$ auf die Zielgrosse Y , wenn alle anderen erklarenden Variablen fest gehalten werden und nur $x^{(j)}$ variiert wird.
- Modell linear, weil die Parameter linear in der Modellgleichung vorkommen (Bsp: $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + E_i$)

Matrix-Schreibweise: (n Variablen):

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \ldots & x_1^{(m)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \ldots & x_2^{(m)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & \ldots & x_n^{(m)} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, E = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix}$$

Das Model ist also gleich: $Y = X\beta + E$.

Note:

- Die Matrix X heisst Designmatrix.
- i-ten Zeile = erklarenden Variablen der i-ten Beobachtung.
- Erste Spalte = Achsenabschnitt.

Multiple lineare Regressionsmodelle mit:

- **Lineare Variablen:** $Y_i = \beta_0 + \beta_1 x_i + E_i$

$$p = 2, \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- **Quadratischen Variablen:** $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + E_i$

$$p = 3, \quad X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

- **Transform Variablen:** $Y_i = \beta_0 + \beta_1 \log(x_i^{(1)}) + \beta_2 \sin(\pi x_i^{(2)}) + E_i$

$$p = 3, \quad X = \begin{bmatrix} 1 & \log(x_1^{(1)}) & \sin(\pi x_1^{(2)}) \\ 1 & \log(x_2^{(1)}) & \sin(\pi x_2^{(2)}) \\ \vdots & \vdots & \vdots \\ 1 & \log(x_n^{(1)}) & \sin(\pi x_n^{(2)}) \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Note: Alle diese 3 Modelle sind lineares, obwohl nicht immer die $x_i^{(i)}$ linear sind.

10.2.2 Parameterschatzung:

Methode der kleinste Quadrate: falls die Matrix X vollen Rang hat, so ist die Losung geschlossen darstellbar:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

wobei wir hier mit y den Vektor der beobachteten Werte der Zielgrosse bezeichnen. Fur $\hat{\sigma}^2$ verwendet man:

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i^{(1)} - \ldots - \hat{\beta}_m x_i^{(m)} \right)^2$$

Der Nenner bei der Schatzung von $\hat{\sigma}^2$ hat die Form #Beobachtungen − #Parameter und ist erwartungstreu.

10.2.3 Test und Vertrauensintervalle:

Individuelle Tests
Wir konnen fur jede Parameter β_j ein Test durchfuhren:

- $H_{0,j} : \beta_j = 0$
- $H_{A,j} : \beta_j \neq 0$

fur $j = 0, \ldots, m$. Wir erhalten eine t-Verteilung, mit $n - m$ DOF. Die # DOF hat also auch hier die Form "Anzahl Beobachtungen − Anzahl Parameter".
Note: Ein individueller Test beantwortet die Frage, ob man eine einzelne Variable weglassen kann. Wenn sich zwei erklarende Variablen sehr ahnlich sind (d.h. wenn sie stark korreliert sind), so man kann beide (einzeln) weglassen (sie ergeben gleich Informationen).

F-Test
Testet ob es plausibel ist, dass alle Variablen weggelassen werden konnen (typischerweise ausser dem Achsenabschnitt).

- $H_0 : \beta_1 = \beta_2 = \ldots = \beta_m = 0$ ("keine Variable hat einen Einfluss")
- H_A : Mindestens ein $\beta_j \neq 0, \quad j \in \{1, \ldots, m\}$

Note: Dieser Test basiert sich auf die F-Verteilung. Computer ergibt p-Wert des entsprechenden Tests, den wir auch als Globaltest bezeichnen.