

Zusammenfassung Wahrscheinlichkeit & Statistik

Grundlagen

Binomischer Lehrsatz

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

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

Ausserdem:

$$\begin{aligned} (a+b)^2 &= a^2 + 2ab + b^2 & (a-b)^2 &= a^2 - 2ab + b^2 \\ (a+b)(a-b) &= a^2 - b^2 & a^3 - b^3 &= (a-b)(a^2 + ab + b^2) \end{aligned}$$

Mitternachtsformel

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Potenzgesetze / Logarithmus / Wurzelgesetze

$$\begin{aligned} a^0 &= 1 & a^1 &= a \\ a^m \cdot a^n &= a^{m+n} & (a^n)^m &= a^{nm} \\ \frac{a^n}{a^m} &= a^{n-m} & a^{-n} &= \frac{1}{a^n} \end{aligned}$$

$$a^{\frac{b}{n}} = \sqrt[n]{a^b}$$

$$\log(0) = \text{undef.}$$

$$\log(1) = 0$$

$$x = \log_a(y) \Leftrightarrow a^x = y$$

$$-\log(x) = \log\left(\frac{1}{x}\right)$$

$$\log(x) - \log(y) = \log\left(\frac{x}{y}\right)$$

$$\frac{\log(x)}{\log(a)} = \log_a(x) \quad n \log x = \log x^n$$

$$\log_a(u) + \log_a(v) = \log_a(u \cdot v)$$

$$\sqrt[n]{a} \cdot \sqrt[n]{b} = \sqrt[n]{a \cdot b} \quad \frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$$

$$\sqrt[m]{\sqrt[n]{a}} = \sqrt[m \cdot n]{a} \quad (\sqrt[n]{a})^m = \sqrt[n]{a^m}$$

Wahrscheinlichkeitstheorie

Der Ereignisraum oder Grundraum $\Omega \neq \emptyset$ ist die Menge aller möglichen Ereignisse des betrachteten Zufallsexperimentes. Die Elemente $\omega \in \Omega$ heissen Elementarerignisse oder Ausgänge des Experimentes. Die Potenzmenge von Ω , bezeichnet mit $\mathcal{P}(\Omega)$ oder 2^Ω ist die Menge aller Teilmengen von Ω . Die Klasse aller (beobachtbaren) Ereignisse ist \mathcal{F} , wobei \mathcal{F} eine Teilmenge von $\mathcal{P}(\Omega)$ ist. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ ist eine σ -Algebra und muss somit folgende Axiome erfüllen:

1. $\Omega \in \mathcal{F}$
2. Für jedes $A \in \mathcal{F}$ ist auch das Komplement $A^c \in \mathcal{F}$
3. Für jede Folge $(A_n)_{n \in \mathbb{N}}$ mit $A_n \in \mathcal{F}$ für alle $n \in \mathbb{N}$, gilt $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Wir gehen allgemein davon aus, dass wir bei den Zufallsexperimenten jeweils genau ein Elementarereignis ω erhalten. Man sagt dann, dass das Ereignis A eintritt, wenn $\omega \in A$. Mithilfe von Mengenoperationen können aus $A, B \in \mathcal{F}$ neue Ereignisse gebildet werden, nämlich $A \cap B$ (Schnittmenge), $A \cup B$ (Vereinigung) und A^c (Komplement). Ein Wahrscheinlichkeitsmass ist eine Abbildung $P: \mathcal{F} \rightarrow [0, 1]$, welche die folgenden Bedingungen erfüllt:

1. $P[A] \geq 0$ für alle $A \in \mathcal{F}$
2. $P[\Omega] = 1$
3. $P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$, falls die $A_i \in \mathcal{F}$ paarweise disjunkt sind, d.h. $A_i \cap A_k = \emptyset$ für $i \neq k$. Dies wird auch als $P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$ notiert.

Aus den Axiomen (und der Definition der bedingten Wahrscheinlichkeit) lassen sich die folgenden Rechenregeln herleiten:

1. $P[A^c] = 1 - P[A]$
2. $P[\emptyset] = 0$
3. Für $A \subseteq B$ gilt $P[A] \leq P[B]$
4. Für beliebige A, B gilt:
 $P[A \cup B] = P[A] + P[B] - P[A \cap B]$
5. $P[A \cap B] = P[B|A]P[A]$

Diskrete Wahrscheinlichkeitsräume

Ein Wahrscheinlichkeitsraum heisst ein diskreter Wahrscheinlichkeitsraum, wenn die Ergebnismenge Ω endlich oder abzählbar unendlich ist und die σ -Algebra die Potenzmenge ist. In diesem Fall reicht es, die Wahrscheinlichkeiten $p_i = P[\{\omega_i\}]$ aller Elementarerignisse anzugeben, denn es gilt:

$$P[A] = P\left[\bigcup_{\substack{i \text{ mit} \\ \omega_i \in A}} \{\omega_i\}\right] = \sum_{\substack{i \text{ mit} \\ \omega_i \in A}} P[\{\omega_i\}] = \sum_{\substack{i \text{ mit} \\ \omega_i \in A}} p_i$$

Wenn Ω endlich (nicht im abzählbaren Fall!) ist und sind $\omega_1, \dots, \omega_N$ alle gleich wahrscheinlich (also $p_1 = p_2 = \dots = p_N = \frac{1}{N}$), so heisst Ω ein Laplace-Raum und P die diskrete Gleichverteilung auf Ω . Für beliebige $A \subseteq \Omega$ gilt dann $P[A] = \frac{|A|}{|\Omega|}$.

Bedingte Wahrscheinlichkeiten

Seien A, B Ereignisse und $P[A] > 0$. Die bedingte Wahrscheinlichkeit von B unter der Bedingung, dass A eintritt ("gegeben A ") wird definiert durch:

$$P[B|A] := \frac{P[B \cap A]}{P[A]}$$

Im diskreten Fall haben wir also insbesondere:

$$p_i^* = P^*[\{\omega_i\}] = P[\{\omega_i\} | A] = \begin{cases} \frac{p_i}{P[A]} & \text{für } \omega_i \in A \\ 0 & \text{für } \omega_i \in A^c \end{cases}$$

Der Satz der totalen Wahrscheinlichkeit besagt: Sei A_1, \dots, A_n eine Zerlegung von Ω (in paarweise disjunkte Ereignisse), also $\bigcup_{i=1}^n A_i = \Omega$. Dann gilt für beliebige Ereignisse B :

$$P[B] = \sum_{i=1}^n P[B|A_i]P[A_i]$$

Das Resultat gilt auch für eine abzählbare Zerlegung $(A_i)_{i \in \mathbb{N}}$.

Die Formel von Bayes besagt: Ist A_1, \dots, A_n eine Zerlegung von Ω mit $P[A_i] > 0$ für $i = 1, \dots, n$ und B ein Ereignis mit $P[B] > 0$, so gilt für jedes k :

$$P[A_k|B] = \frac{P[B|A_k]P[A_k]}{\sum_{i=1}^n P[B|A_i]P[A_i]}$$

Insbesondere also auch:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[B|A]P[A]}{P[B|A]P[A] + P[B|A^c]P[A^c]}$$

Unabhängigkeit

Zwei Ereignisse A, B heissen (stochastisch) unabhängig, wenn

$$P[A \cap B] = P[A]P[B]$$

Falls $P[A] = 0$ oder $P[B] = 0$, ist A und B somit immer unabhängig. Für $P[A] \neq 0$ gilt, dass A, B genau dann unabhängig sind, wenn $P[B|A] = P[B]$ (und symmetrisch für $P[B] \neq 0$). Für n Ereignisse gilt: Die Ereignisse A_1, \dots, A_n heissen (stochastisch) unabhängig, wenn für jede endliche Teilfamilie die Produktformel gilt, d.h. für $m \in \mathbb{N}$ und $\{k_1, \dots, k_m\} \subseteq \{1, \dots, n\}$ gilt immer:

$$P\left[\bigcap_{i=1}^m A_{k_i}\right] = \prod_{i=1}^m P[A_{k_i}]$$

Anstatt endlich viele Ereignisse kann man in obiger Definition auch eine beliebige Familie (abzählbar oder überabzählbar) betrachten; die Produktformel wird jedoch immer nur für alle endlichen Teilmengen gefordert.

Zufallsvariablen, die unabhängig sind und alle dieselbe Verteilung haben, nennt man i.i.d (independent identically distributed) bzw. auf deutsch u.i.v. (unabhängig und identisch verteilt).

Diskrete Zufallsvariablen und Verteilungen

Eine (genauer: reellwertige) diskrete Zufallsvariable (ZV) auf Ω ist eine Funktion $X : \Omega \rightarrow \mathbb{R}$. Mit Ω ist auch der Wertebereich $\mathcal{W}(X) = \{x_1, x_2, \dots\}$ von X endlich oder abzählbar. Die Verteilungsfunktion von X ist die Abbildung $F_X : \mathbb{R} \rightarrow [0, 1]$, definiert durch:

$$t \mapsto F_X(t) := P[X \leq t] := P[\{\omega | X(\omega) \leq t\}] = \sum_{\substack{k \text{ mit} \\ x_k \leq t}} p_X(x_k)$$

Die Gewichtsfunktion oder diskrete Dichte von X ist die Funktion $p_X : \mathcal{W}(X) \rightarrow [0, 1]$, definiert durch

$$p_X(x_k) := P[X = x_k] = P[\{\omega | X(\omega) = x_k\}]$$

Für jede Menge $B \subseteq \mathcal{W}(X)$ gilt

$$\mu_X(B) = P[X \in B] = \sum_{x_k \in B} p_X(x_k)$$

$\mu_X(B)$ ist dabei die (diskrete) Verteilung. Für jede Teilmenge $A \subseteq \Omega$ ist die Indikatorfunktion I_A von A definiert durch:

$$\sum_{x_k \in \mathcal{W}(X)} p_X(x_k) = P[X \in \mathcal{W}(X)] = 1$$

Der Erwartungswert von X ist definiert als:

$$E[X] := \sum_{x_k \in \mathcal{W}(X)} x_k p_X(x_k) = \sum_{\omega_i \in \Omega} p_i X(\omega_i)$$

falls die Reihe absolut konvergiert, also

$\sum_{x_k \in \mathcal{W}(X)} |x_k| p_X(x_k) < \infty$ gilt. Für $Y = g(X)$, wobei $g : \mathbb{R} \rightarrow \mathbb{R}$, haben wir:

$$E[Y] = E[g(X)] = \sum_{x_k \in \mathcal{W}(X)} g(x_k) p_X(x_k)$$

falls die Reihe absolut konvergiert. Der Erwartungswert hat folgende Eigenschaften:

1. **Monotonie:** Ist $X \leq Y$ (also $X(\omega) \leq Y(\omega)$ für alle ω), so gilt auch $E[X] \leq E[Y]$.
2. **Linearität:** Für beliebige $a, b \in \mathbb{R}$ gilt $E[aX + b] = aE[X] + b$.
3. Wenn X nur Werte in $N_0 = \{0, 1, 2, \dots\}$ annimmt, so gilt

$$E[X] = \sum_{j=1}^{\infty} P[X \geq j] = \sum_{\ell=0}^{\infty} P[X > \ell]$$

Wenn $E[X^2] < \infty$, so ist die Varianz von X folgendermassen definiert:

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

$\sqrt{\text{Var}[X]} = \text{sd}(X) = \sigma(X)$ heisst die Standardabweichung von X . Es gilt:

$$\text{Var}[X] = \sum_{x_k \in \mathcal{W}(X)} (x_k - E[X])^2 p_X(x_k)$$

Für die Varianz gilt weiter:

1. $\text{Var}[X] = E[X^2] - (E[X])^2$
2. $\text{Var}[Y] = \text{Var}[aX + b] = a^2 \text{Var}[X]$

Gemeinsame Verteilungen, unabhängige ZV

Seien X_1, \dots, X_n beliebige Zufallsvariablen. Die gemeinsame Verteilungsfunktion von X_1, \dots, X_n ist die Abbildung $F : \mathbb{R}^n \rightarrow [0, 1]$, definiert durch

$$(x_1, \dots, x_n) \mapsto F(x_1, \dots, x_n) := P[X_1 \leq x_1, \dots, X_n \leq x_n]$$

Für diskrete Zufallsvariablen definiert man die gemeinsame Gewichtsfunktion $p : \mathbb{R}^n \rightarrow [0, 1]$ durch

$$p(x_1, \dots, x_n) := P[X_1 = x_1, \dots, X_n = x_n]$$

Aus der gemeinsamen Gewichtsfunktion p bekommt man die gemeinsame Verteilungsfunktion:

$$F(x_1, \dots, x_n) = \sum_{y_1 \leq x_1, \dots, y_n \leq x_n} p(y_1, \dots, y_n)$$

Haben X und Y die gemeinsame Verteilungsfunktion F , so ist die Funktion $F_X : \mathbb{R} \rightarrow [0, 1]$

$$x \mapsto F_X(x) := P[X \leq x] = P[X \leq x, Y < \infty] = \lim_{y \rightarrow \infty} F(x, y)$$

die Verteilungsfunktion der Randverteilung von X , wobei das Analogon für Y gilt. Für diskrete ZV X und Y ist die Gewichtverteilung der Randverteilung von X gegeben durch: $p_X : \mathcal{W}(X) \rightarrow [0, 1]$ (wobei $p_X(x) = P[X = x]$)

$$p_X(x) = \sum_{y_j \in \mathcal{W}(Y)} P[X = x, Y = y_j] = \sum_{y_j \in \mathcal{W}(Y)} p(x, y_j)$$

Die Zufallsvariablen X_1, \dots, X_n heissen unabhängig, falls gilt:

$$F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

Dies kann im diskreten Fall auch über die Gewichtsfunktion definiert werden. Die diskreten ZV X_1, \dots, X_n heissen unabhängig, genau dann wenn

$$p(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

für alle x_1, \dots, x_n . Eine äquivalente Definition im diskreten Fall ist, dass für beliebige Teilmengen $B_i \subseteq \mathcal{W}(X_i)$, $i = 1, \dots, n$ folgendes gilt:

$$P[X_1 \in B_1, \dots, X_n \in B_n] = \prod_{i=1}^n P[X_i \in B_i]$$

Unabhängigkeit von Zufallsvariablen bleibt unter Transformationen erhalten. Wenn X_1, \dots, X_n diskrete Zufallsvariablen sind und $Y_i = f_i(X_i)$ für irgendwelche Abbildungen $f_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, \dots, n$, so sind Y_1, \dots, Y_n auch unabhängig, wenn X_1, \dots, X_n unabhängig sind.

Funktionen von mehreren Zufallsvariablen

Seien X_1, \dots, X_n diskrete Zufallsvariablen und $Y = g(X_1, \dots, X_n)$ für eine Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$. Dann ist Y wieder eine Zufallsvariable. Wenn X_1, \dots, X_n diskrete Zufallsvariablen mit endlichen Erwartungswerten $E[X_1], \dots, E[X_n]$ sind, so haben wir für $Y = a + \sum_{\ell=1}^n b_{\ell} X_{\ell}$ mit Konstanten a, b_1, \dots, b_n :

$$E[Y] = a + \sum_{\ell=1}^n b_{\ell} E[X_{\ell}]$$

Für die Varianz gilt allgemein:

$$\text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] + 2 \text{Cov}(X_1, X_2)$$

Wobei mit $\text{Cov}(X_1, X_2)$ die Kovarianz von X_1 und X_2 gemeint ist, sie ist folgendermassen definiert:

$$E[X_1 X_2] - E[X_1] E[X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])]$$

Es gilt $\text{Cov}(X, X) = X$. Die obige Formel kann man auf n Zufallsvariablen erweitern und erhält somit die allgemeine Summenformel für Varianzen:

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

Wenn $\text{Cov}(X_1, X_2) = 0$ ist, so heissen die Zufallsvariablen X_1 und X_2 unkorreliert; die Zufallsvariablen X_1, \dots, X_n heissen paarweise unkorreliert, wenn alle Paare X_i, X_j mit $i \neq j$ unkorreliert sind. Für paarweise unkorrelierte X_1, \dots, X_n gilt also:

$$\text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}[X_i]$$

Wenn X_1, \dots, X_n unabhängig sind, so gilt für den Erwartungswert ihres Produktes:

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

Es gelten die folgenden Implikationen bzgl. Unabhängigkeit / Unkorreliertheit, wobei die umgekehrten Implikationen in der Regel falsch sind:

$$\text{unabhängig} \implies \text{paarweise unabhängig} \implies \text{unkorreliert}$$

Seien X und Y diskrete Zufallsvariablen mit gemeinsamer Gewichtsfunktion $p(x, y)$, dann ist auch ihre Summe $Z = X + Y$ diskret. Die Gewichtsfunktion von Z lautet:

$$p_Z(z) = P[Z = z] = \sum_{x_k \in \mathcal{W}(X)} p(x_k, z - x_k)$$

Wobei dies via Symmetrie auch als $\sum_{y_j \in \mathcal{W}(Y)} p(z - y_j, y_j)$ geschrieben werden kann. Wenn X und Y unabhängig sind, folgt somit für $p_Z(z)$ (was auch als Faltung der Gewichtsfunktionen p_X und p_Y bezeichnet wird):

$$\sum_{x_k \in \mathcal{W}(X)} p_X(x_k) p_Y(z - x_k) = \sum_{y_j \in \mathcal{W}(Y)} p_X(z - y_j) p_Y(y_j)$$

Bedingte Verteilungen

Seien X und Y diskrete Zufallsvariablen mit gemeinsamer Gewichtsfunktion $p(x, y)$. Die bedingte Gewichtsfunktion, gegeben dass $Y = y$, ist definiert durch:

$$p_{X|Y}(x|y) := P[X = x|Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{p(x, y)}{p_Y(y)}$$

für $p_Y(y) > 0$ und 0 sonst. Daraus folgt, dass X und Y genau dann unabhängig sind, wenn für alle y mit $p_Y(y) > 0$ gilt: $p_{X|Y}(x|y) = p_X(x)$. Sei X eine diskrete Zufallsvariable und A ein Ereignis mit $P[A] > 0$. Dann erhalten wir allgemeiner, dass die bedingte Verteilung von X gegeben A durch die Gewichtsfunktion definiert ist:

$$p_{X|A}(x) := P[X = x|A] = \frac{P[\{X = x\} \cap A]}{P[A]}$$

Wichtige diskrete Verteilungen

Diskrete Gleichverteilung

Die diskrete Gleichverteilung auf einer endlichen Menge $\mathcal{W} = \{x_1, \dots, x_N\}$ gehört zu einer Zufallsvariablen X mit Wertebereich \mathcal{W} und Gewichtsfunktion

$$p_X(x_k) = P[X = x_k] = \frac{1}{N}$$

für $k = 1, \dots, N$. Der Erwartungswert ist allgemein $\frac{1}{n} \sum_{i=1}^n x_i$, im Falle von $\mathcal{W} = \{a, a+1, \dots, b-1, b\}$ ist er $\frac{a+b}{2}$. Für die Varianz gilt im Allgemeinen:

$$\text{Var}[X] = \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right)$$

Für $\mathcal{W} = \{a, a+1, \dots, b-1, b\}$ ergibt sich $\frac{(b-a+2)(b-a)}{12}$.

Bernoulli-Verteilung

Wird mit X ein einziges 0-1-Experiment modelliert, so hat X eine Bernoulli-Verteilung mit Parameter p . Somit ist $\mathcal{W}(X) = \{0, 1\}$ und die Gewichtsfunktion ist gegeben durch $p_X(1) = P[X = 1] = p$, $p_X(0) = P[X = 0] = 1 - p$, also $p_X(x) = p^x(1-p)^{1-x}$ für $x \in \{0, 1\} = \mathcal{W}(X)$. Man schreibt kurz $X \sim \text{Be}(p)$. Es gilt $E[X] = p$ und $\text{Var}[X] = p(1-p)$.

Binomialverteilung

Die Binomialverteilung mit Parametern n und p (notiert als $X \sim \text{Bin}(n, p)$) beschreibt die Anzahl der Erfolge bei n unabhängigen 0-1-Experimenten mit Erfolgsparameter p , diese Verteilung hat also die Zufallsvariable

$$X = \sum_{i=1}^n I_{A_i} = \sum_{i=1}^n Y_i$$

Es gilt $\mathcal{W}(X) = \{0, 1, 2, \dots, n\}$ und die Gewichtsfunktion ist

$$p_X(k) = P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

für $k = 0, 1, \dots, n$. Es gilt $E[X] = np$ und $\text{Var}[X] = np(1-p)$.

Geometrische Verteilung

Die geometrische Verteilung mit Parameter p (notiert als $X \sim \text{Geom}(p)$) beschreibt die Wartezeit auf den ersten Erfolg bei einer unendlichen Folge von unabhängigen 0-1-Experimenten mit Erfolgsparameter p . Es gilt $\mathcal{W}(X) = N = \{1, 2, 3, \dots\}$ und die Gewichtsfunktion lautet:

$$p_X(k) = P[X = k] = p(1-p)^{k-1}$$

für $k = 1, 2, 3, \dots$. Es gilt $E[X] = \frac{1}{p}$ und $\text{Var}[X] = \frac{1-p}{p^2}$.

Negativbinomiale Verteilung

Die negativbinomiale Verteilung mit Parametern r und p (notiert als $X \sim \text{NB}(r, p)$) beschreibt die Wartezeit auf den r -ten Erfolg (mit $r \in \mathbb{N}$) bei einer unendlichen Folge von unabhängigen 0-1-Experimenten mit Erfolgsparameter p . Es gilt $\mathcal{W}(X) = \{r, r+1, r+2, \dots\}$ und die Gewichtsfunktion lautet:

$$p_X(k) = P[X = k] = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

für $k = r, r+1, \dots$. Es gilt $E[X] = \frac{r}{p}$ und $\text{Var}[X] = \frac{r(1-p)}{p^2}$. Wenn die Zufallsvariablen $X_1, \dots, X_r \sim \text{Geom}(p)$ und unabhängig sind, so ist ihre Summe $X := X_1 + \dots + X_r \sim \text{NB}(r, p)$

Hypergeometrische Verteilung

Die hypergeometrische Verteilung mit Parametern $n \in \mathbb{N}$ und $m, r \in \{1, \dots, n\}$ modelliert folgendes Szenario: In einer Urne sind n Gegenstände, davon r vom Typ 1 und $n-r$ vom Typ 2. Man zieht ohne Zurücklegen m der Gegenstände; die Zufallsvariable X beschreibt die Anzahl der Gegenstände vom Typ 1 in dieser Stichprobe vom Umfang m . Es gilt $\mathcal{W}(X) = \{0, 1, \dots, \min(m, r)\}$ und die Gewichtsfunktion lautet:

$$p_X(k) = \frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}}$$

für $k \in \mathcal{W}(X)$. Es gilt $E[X] = m \frac{r}{n}$

Poisson-Verteilung

Die Poisson-Verteilung mit Parameter $\lambda \in (0, \infty)$ (notiert als $X \sim \mathcal{P}(\lambda)$) ist eine Verteilung auf der Menge $N_0 = \{0, 1, 2, \dots\}$ mit Gewichtsfunktion:

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

für $k = 0, 1, 2, \dots$. Die Poisson-Verteilung ist ein Grenzwert von Binomialverteilungen, wenn $np_n \rightarrow \lambda$ für $n \rightarrow \infty$ gilt. Die Poisson-Verteilung kann zur Modellierung der Anzahl von seltenen Ereignissen in einem Zeitraum verwendet werden, wobei λ anschaulich etwa die durchschnittliche Anzahl im betrachteten Zeitraum ist. Für grosse n und kleine p (Faustregel: $np^2 \leq 0.05$) kann mit der Poisson-Verteilung auch eine Binomialverteilung approximiert werden. Es gilt $E[X] = \lambda$ und $\text{Var}[X] = \lambda$.

Allgemeine Zufallsvariablen

Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum, also Ω ein Grundraum, $\mathcal{F} \subseteq 2^\Omega$ eine σ -Algebra von beobachtbaren Ereignissen und P ein Wahrscheinlichkeitsmass auf \mathcal{F} . Eine (reellwertige) Zufallsvariable auf Ω ist eine messbare Funktion $X: \Omega \rightarrow \mathbb{R}$; d.h., dass die Menge $\{X \leq t\} = \{\omega | X(\omega) \leq t\}$ für jedes t ein (beobachtbares) Ereignis, also in \mathcal{F} , sein muss. Die Verteilungsfunktion von X ist dann die Abbildung:

$$t \mapsto F_X(t) := P[X \leq t] := P[\{\omega | X(\omega) \leq t\}]$$

Jede Verteilungsfunktion $F_X(t)$ hat die folgenden Eigenschaften:

- F_X ist wachsend und rechtsstetig; das bedeutet, dass $F_X(s) \leq F_X(t)$ für $s \leq t$ gilt und $F_X(u) \rightarrow F_X(t)$ für $u \rightarrow t$ mit $u > t$.
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$, $\lim_{t \rightarrow +\infty} F_X(t) = 1$

Umgekehrt ist jede Funktion F mit diesen beiden Eigenschaften die Verteilungsfunktion F_X einer Zufallsvariablen X . Das stochastische Verhalten einer Zufallsvariablen X wird durch ihre Verteilung beschrieben; das ist das Wahrscheinlichkeitsmass μ_X auf \mathbb{R} , definiert durch:

$$\mu_X(B) := P[X \in B]$$

Das Mass $\mu_X(B)$ ist festgelegt, sobald man die Verteilungsfunktion F_X kennt, und umgekehrt ist $F_X(t) = \mu_X((-\infty, t])$, weswegen F_X auch durch μ_X festgelegt ist.

Eine Zufallsvariable X mit Verteilungsfunktion $F_X(t) = P[X \leq t]$ heisst (absolut)stetig mit Dichte(funktion) $f_X: \mathbb{R} \rightarrow [0, \infty)$, falls gilt

$$F_X(t) = \int_{-\infty}^t f_X(s) ds$$

für alle $t \in \mathbb{R}$. Eine Dichtefunktion F_X hat stets folgende Eigenschaften:

- $f_X \geq 0$ und $f_X = 0$ ausserhalb von $\mathcal{W}(X)$.
- $\int_{-\infty}^{\infty} f_X(s) ds = 1$, was aus $\lim_{t \rightarrow +\infty} F_X(t) = 1$ folgt.

Zu einer gegebenen Funktion f mit diesen Eigenschaften kann umgekehrt auch eine stetige Zufallsvariable X konstruiert werden. Für $a < b$ gilt:

$$P[a < X \leq b] = F_X(b) - F_X(a) = \int_a^b f_X(s) ds$$

Und allgemein für (messbare) Mengen $B \subseteq \mathbb{R}$:

$$P[X \in B] = \int_B f_X(s) ds$$

Daraus folgt, dass $P[X = t] = 0$ für jedes $t \in \mathbb{R}$ gilt. Ist f_X stetig an der Stelle t , gilt jedoch $P[X \in (t, t + \varepsilon]] \approx \varepsilon f_X(t)$ für ε klein. An jeder Stelle t , wo f_X stetig ist, gilt weiter:

$$f_X(t) = \frac{d}{dt} F_X(t) = F'_X(t)$$

Wichtige stetige Verteilungen

Gleichverteilung

Die Gleichverteilung auf einem Intervall $[a, b]$ (notiert als $X \sim \mathcal{U}(a, b)$) ist ein Modell für die zufällige Wahl eines Punktes in $[a, b]$. Die zugehörige Zufallsvariable X hat den Wertebereich $\mathcal{W}(X) = [a, b]$, die Dichtefunktion

$$f_X(t) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq t \leq b \\ 0 & \text{sonst} \end{cases}$$

und die Verteilungsfunktion

$$F_X(t) = \begin{cases} 0 & \text{für } t < a \\ \frac{t-a}{b-a} & \text{für } a \leq t \leq b \\ 1 & \text{für } t > b \end{cases}$$

Es gilt $E[X] = \frac{a+b}{2}$ und $Var[X] = \frac{1}{12}(b-a)^2$.

Exponentialverteilung

Die Exponentialverteilung mit Parameter $\lambda > 0$ (notiert als $X \sim \text{Exp}(\lambda)$) ist das stetige Analogon der geometrischen Verteilung. Die zugehörige Zufallsvariable X hat $\mathcal{W}(X) = [0, \infty)$, Dichte

$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

und Verteilungsfunktion

$$F_X(t) = \int_{-\infty}^t f_X(s) ds = \begin{cases} 1 - e^{-\lambda t} & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

Wie die geometrische Verteilung ist sie "gedächtnislos", d.h.:

$$P[X > t + s | X > s] = P[X > t]$$

Es gilt $E[X] = \frac{1}{\lambda}$ und $Var[X] = \frac{1}{\lambda^2}$.

Normalverteilung

Die Normalverteilung hat zwei Parameter $\mu \in \mathbb{R}, \sigma^2 > 0$ (notiert als $X \sim \mathcal{N}(\mu, \sigma^2)$). Die dazugehörige Zufallsvariable X hat den Wertebereich $\mathcal{W}(X) = \mathbb{R}$ und die Dichtefunktion

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad \text{für } t \in \mathbb{R}$$

Die zugehörige Dichte der Standard-Normalverteilung ($\mu = 0, \sigma^2 = 1$, also $\mathcal{N}(0, 1)$) wird mit $\varphi(t)$ und die Verteilungsfunktion mit $\Phi(t)$ bezeichnet. Für F_X / Φ gibt es keinen geschlossenen Ausdruck, jedoch ist folgendes Integral tabelliert:

$$\Phi(t) = \int_{-\infty}^t \varphi(s) ds = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}s^2} ds$$

Es gilt $E[X] = \mu$, $Var[X] = \sigma^2$.

Wenn $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$, also:

$$F_X(t) = P[X \leq t] = P\left[\frac{X-\mu}{\sigma} \leq \frac{t-\mu}{\sigma}\right] = \Phi\left(\frac{t-\mu}{\sigma}\right)$$

Cauchy-Verteilung

Eine Cauchy-verteilte Zufallsvariable X hat den Wertebereich $\mathcal{W}(X) = \mathbb{R}$ und die Dichtefunktion

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

Die Verteilungsfunktion lautet

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

Sind Y und Z unabhängige $\mathcal{N}(0, 1)$ -verteilte Zufallsvariablen, so ist der Quotient $X := Y/Z$ Cauchy-verteilt. Die Cauchy-Verteilung hat keinen Erwartungswert und keine Varianz.

Gamma-Verteilung

Die Gamma-Verteilung ist eine stetige Verteilung mit der Dichtefunktion

$$f(z) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha z^{\alpha-1} e^{-\lambda z} \quad \text{für } z \geq 0$$

mit Parametern $\alpha > 0, \lambda > 0$. Dabei ist

$$\Gamma(\alpha) := \int_0^\infty u^{\alpha-1} e^{-u} du \quad \text{für } \alpha > 0$$

die Gamma-Funktion, wobei $\Gamma(n) = (n-1)!$ für $n \in \mathbb{N}$. Eine Summe von n unabhängigen $\text{Exp}(\lambda)$ -verteilten Zufallsvariablen ist $Ga(n, \lambda)$ verteilt.

Chiquadrat-Verteilung

Die Chiquadrat-Verteilung ist eine stetige Verteilung mit dem Freiheitsgrad n als Parameter. Die Dichtefunktion f_n der χ_n^2 -Verteilung ist:

$$f_n(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad \text{für } x > 0$$

Somit ist sie ein Spezialfall einer Gamma-Verteilung mit $\alpha = \frac{n}{2}$ und $\lambda = \frac{1}{2}$. Hat man n unabhängige, standardnormalverteilte Zufallsvariablen Z_i , so ist die Summe der Quadrate $Z_1^2 + \dots + Z_n^2$ Chiquadrat-verteilt mit n Freiheitsgraden.

t-Verteilung

Die t -Verteilung mit n Freiheitsgraden gehört zu einer stetigen Zufallsvariablen Z mit Dichtefunktion

$$f_Z(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}} \quad \text{für } z \in \mathbb{R}$$

Für $n = 1$ ist das eine Cauchy-Verteilung und für $n \rightarrow \infty$ erhält man eine $\mathcal{N}(0, 1)$ -Verteilung. Die t -Verteilung ist symmetrisch um 0, jedoch langschwänziger (umso mehr, je kleiner n ist). Sind X und Y unabhängig mit $X \sim \mathcal{N}(0, 1)$ und $Y \sim \chi_n^2$, so ist der Quotient $Z := \frac{X}{\sqrt{\frac{1}{n}Y}}$ t -verteilt mit n

Freiheitsgraden.

Erwartungswerte

Beliebige reellwertige Zufallsvariablen X können immer durch eine Folge von diskreten Zufallsvariablen approximiert werden, für $X \geq 0$ bspw.:

$$X_n := \sum_{k=1}^{n2^n} (k-1)2^{-n} I_{\{(k-1)2^{-n} \leq X < k2^{-n}\}} + n I_{\{X \geq n\}}$$

Dann erhält man den Erwartungswert von $X \geq 0$ als:

$$E[X] := \lim_{n \rightarrow \infty} E[X_n]$$

Beliebige Zufallsvariablen X zerlegt man als $X = X^+ - X^- := \max(X, 0) - \max(-X, 0)$ mit $X^+ \geq 0, X^- \geq 0$ und setzt $E[X] := E[X^+] - E[X^-]$, sofern beide diese Werte endlich sind. Die allgemeinen Eigenschaften und Resultate für Erwartungswerte gelten im allgemeinen Fall genau gleich wie für diskrete Zufallsvariablen. Ist X stetig mit Dichte $f_X(x)$, gilt

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

sofern das Integral absolut konvergiert, d.h. falls $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$. Wenn das Integral nicht absolut konvergent ist, existiert der Erwartungswert (in \mathbb{R}) nicht. Für Funktionen von Zufallsvariablen gilt: Sei X eine Zufallsvariable und $Y = g(X)$ eine weitere Zufallsvariable. Ist X stetig mit Dichte $f_X(x)$, so ist

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

sofern das Integral absolut konvergiert.

Gemeinsame Verteilungen, unabhängige Zufallsvariablen

Die gemeinsame Verteilungsfunktion von n Zufallsvariablen X_1, \dots, X_n ist die Abbildung $F: \mathbb{R}^n \rightarrow [0, 1]$

$$(x_1, \dots, x_n) \mapsto F(x_1, \dots, x_n) := P[X_1 \leq x_1, \dots, X_n \leq x_n]$$

Falls die gemeinsame Verteilungsfunktion F sich schreiben lässt als

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_n \dots dt_1$$

für eine Funktion $f: \mathbb{R}^n \rightarrow [0, \infty)$, so heisst $f(x_1, \dots, x_n)$ die gemeinsame Dichte von X_1, \dots, X_n . Wenn eine gemeinsame Dichte existiert, gilt analog zum diskreten Fall:

1. $f(x_1, \dots, x_n) \geq 0$ und $= 0$ ausserhalb von $\mathcal{W}(X_1, \dots, X_n)$
2. $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_n \dots dx_1 = 1$
3. $P[(X_1, \dots, X_n) \in A] = \int_{(x_1, \dots, x_n) \in A} f(x_1, \dots, x_n) dx_n \dots dx_1$

Haben X und Y die gemeinsame Verteilungsfunktion F , so ist die Funktion $F_X : \mathbb{R} \rightarrow [0, 1]$

$$x \mapsto F_X(x) := P[X \leq x] = P[X \leq x, Y < \infty] = \lim_{y \rightarrow \infty} F(x, y)$$

die Verteilungsfunktion der Randverteilung von X (was symmetrisch für Y) gilt. Falls X und Y eine gemeinsame Dichte $f(x, y)$ haben, so haben auch die Randverteilungen von X und Y Dichten $f_X : \mathbb{R} \rightarrow [0, \infty)$ bzw. $f_Y : \mathbb{R} \rightarrow [0, \infty)$. Es gilt:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad \text{bzw.} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Wie im diskreten Fall gilt, dass die Zufallsvariablen unabhängig heissen, wenn

$$F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

Wobei dies äquivalent (wenn die Zufallsvariablen Dichten haben) zu

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

ist.

Funktionen und Transformationen von Zufallsvariablen

Seien X und Y stetige Zufallsvariablen mit gemeinsamer Dichtefunktion $f(x, y)$. Für $Z = X + Y$ gilt:

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x) dx = \int_{-\infty}^{\infty} f(z - y, y) dy$$

Wenn X und Y zusätzlich unabhängig sind, gilt somit:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) dy$$

Womit f_Z die Faltung $((f_X * f_Y)(z))$ von f_X und f_Y ist. Sei nun $g : \mathbb{R} \rightarrow \mathbb{R}$ eine (messbare) Funktion und $Y = g(X)$. Der allgemeine Ansatz für die Verteilungsfunktion F_Y und die Dichte (falls existent) f_Y sieht folgendermassen aus:

$$F_Y(t) = P[Y \leq t] = P[g(X) \leq t] = \int_{A_g} f_X(s) ds$$

wobei $A_g := \{s \in \mathbb{R} | g(s) \leq t\}$. Die Dichte f_Y ist dann (sofern existent) die Ableitung von F_Y .

Für affine Transformationen, d.h. $Y = g(X) = aX + b$ mit $a > 0, b \in \mathbb{R}$ folgt $F_Y(t) = F_X\left(\frac{t-b}{a}\right)$ und nach der Kettenregel $f_Y(t) = \frac{d}{dt} F_Y(t) = \frac{1}{a} f_X\left(\frac{t-b}{a}\right)$. Für $X \sim \mathcal{N}(\mu, \sigma^2)$ und $Y = aX + b$, folgt daraus $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. Für $g(x) = x^2$, d.h. $Y = X^2$ gilt:

$$F_Y(t) = P[-\sqrt{t} \leq X \leq \sqrt{t}] = F_X(\sqrt{t}) - F_X(-\sqrt{t})$$

Für $X \sim \mathcal{N}(0, 1)$ gilt somit $F_Y(t) = \Phi(\sqrt{t}) - \Phi(-\sqrt{t})$ und damit nach der Kettenregel $f_Y(t) = \frac{1}{\sqrt{t}} \varphi(\sqrt{t}) = \frac{1}{\sqrt{2\pi}} t^{-\frac{1}{2}} e^{-\frac{1}{2}t}$, wobei es sich um die Dichte einer χ^2 -Verteilung, d.h. einer Chiquadrat-Verteilung mit einem Freiheitsgrad handelt.

Für $g(x) = \frac{1}{x}$, d.h. $Y = \frac{1}{X}$ gilt:

$$F_Y(t) = P[Y \leq t] = P\left[\frac{1}{X} \leq t\right] = P\left[X \geq \frac{1}{t}\right] = 1 - F_X\left(\frac{1}{t}\right)$$

Und somit folgt $f_Y(t) = \frac{1}{t^2} f_X\left(\frac{1}{t}\right)$
Sei F eine stetige und streng monoton wachsende Verteilungsfunktion, mit Umkehrfunktion F^{-1} . Ist $X \sim \mathcal{U}(0, 1)$ und $Y = F^{-1}(X)$, so hat Y gerade die Verteilungsfunktion F . Denn:
 $F_Y(t) = P[Y \leq t] = P[F^{-1}(X) \leq t] = P[X \leq F(t)] = F(t)$
Dies liefert Simulationsalgorithmen nach der Inversionsmethode, da man mit einem Computer einfach $F^{-1}(X)$ mit $X \sim \mathcal{U}(0, 1)$ berechnen kann und somit eine Zufallsvariable mit Verteilungsfunktion $F(t)$ erhält.

Ungleichungen und Grenzwertsätze

Ungleichungen

Markov-Ungleichung

Sei X eine Zufallsvariable und ferner $g : \mathcal{W}(X) \rightarrow [0, \infty)$ eine wachsende Funktion. Für jedes $c \in \mathbb{R}$ mit $g(c) > 0$ gilt dann:

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

Beweis: $g(c)I_{\{X \geq c\}} = g(c)I_{\{g(X) \geq g(c)\}} \leq g(X)$, womit aus der Monotonie des Erwartungswertes folgt:
 $E[g(X)] \geq g(c)E[I_{\{X \geq c\}}] = g(c)P[X \geq c]$

Chebyshev-Ungleichung

Sei Y eine Zufallsvariable mit endlicher Varianz. Für jedes $b > 0$ gilt dann:

$$P[|Y - E[Y]| \geq b] \leq \frac{\text{Var}[Y]}{b^2}$$

Beweis: Folgt aus der Markov-Ungleichung, wenn man $X := |Y - E[Y]|$ und $g(x) := x^2$ wählt.

Das Gesetz der grossen Zahlen

Schwaches Gesetz der grossen Zahlen

Sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen (*Anmerkung: Statt Unabhängigkeit reicht auch paarweise Unkorreliertheit, d.h. $\text{Cov}(X_i, X_k) = 0$ für $i \neq k$*), die alle den gleichen Erwartungswert $E[X_i] = \mu$ und die gleiche Varianz $\text{Var}[X_i] = \sigma^2$ haben. Sei $\bar{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$. Dann konvergiert \bar{X}_n für $n \rightarrow \infty$ in Wahrscheinlichkeit/stochastisch gegen $\mu = E[X_i]$, d.h.:

$$P[|\bar{X}_n - \mu| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0 \quad \text{für jedes } \varepsilon > 0$$

Beweis: Es gilt $E[\bar{X}_n] = \mu$ und wegen der Summenformel für Varianzen $\text{Var}[\bar{X}_n] = \frac{1}{n} \sigma^2$. Also liefert die Chebyshev-Ungleichung für alle $\varepsilon > 0$:

$$P[|\bar{X}_n - \mu| > \varepsilon] \leq \frac{\text{Var}[\bar{X}_n]}{\varepsilon^2} = \frac{\frac{1}{n} \sigma^2}{\varepsilon^2} = \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

Das schwache GGZ kann bspw. für die Monte Carlo-Integration verwendet werden. Gesucht sei ein Integral über $[0, 1]^d$ einer Funktion $h : [0, 1]^d \rightarrow \mathbb{R}$. Dieser Integral kann nun als Erwartungswert aufgefasst werden, z.B. gilt für $d = 1$ und U gleichverteilt auf $[0, 1]$:

$$E[h(U)] = \int_{-\infty}^{\infty} h(x) f_U(x) dx = \int_0^1 h(x) dx$$

Aus dem schwachen GGZ folgt dann, dass $\overline{h(U_n)}$ in Wahrscheinlichkeit gegen $E[h(U_1)]$ konvergiert.

Starkes Gesetz der grossen Zahlen

Sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen, die alle dieselbe Verteilung haben, und ihr Erwartungswert $\mu = E[X_i]$ sei endlich. Für $\bar{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$ gilt dann:

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{} \mu \quad P\text{-fastsicher } (P - f.s.)$$

d.h.

$$P[\{\omega \in \Omega | \bar{X}_n(\omega) \rightarrow_{n \rightarrow \infty} \mu\}] = 1$$

Der Zentrale Grenzwertsatz

Sei X_1, X_2, \dots eine Folge von i.i.d Zufallsvariablen mit $E[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$. Für die Summe $S_n = \sum_{i=1}^n X_i$ gilt dann

$$\lim_{n \rightarrow \infty} P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right] = \Phi(x) \quad \text{für alle } x \in \mathbb{R}$$

S_n hat Erwartungswert $E[S_n] = n\mu$ und Varianz $\text{Var}[S_n] = n\sigma^2$, somit ist

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}}$$

standardnormalverteilt. Es gilt also $S_n^* \stackrel{\text{approx.}}{\sim} \mathcal{N}(0, 1)$, $S_n \stackrel{\text{approx.}}{\sim} \mathcal{N}(n\mu, n\sigma^2)$ und $\bar{X}_n \stackrel{\text{approx.}}{\sim} \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$, wobei $\bar{X}_n = \frac{1}{n} S_n$.

Normalapproximation für die Binomialverteilung

Die Normalverteilung wird oft zur Approximation der Binomialverteilung verwendet. Sei S_n eine binomialverteilte Zufallsvariable ($S_n \sim \text{Bin}(n, p)$), dann gilt:

$$P(S_n \leq x) \approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

Die Approximation wird mittels der Kontinuitätskorrektur genauer. Dabei wird folgendermassen korrigiert:

$$P[a < S_n \leq b] \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Für Wahrscheinlichkeiten der Form $S_n \leq a$ wird nur $+\frac{1}{2}$ addiert, aus $P[S_n = a]$ wird $P[a - \frac{1}{2} \leq S_n \leq a + \frac{1}{2}]$.

Grosse Abweichungen und Chernoff-Schranken

Die momenterzeugende Funktion einer Zufallsvariable X ist

$$M_X(t) := E \left[e^{tX} \right] \quad \text{für } t \in \mathbb{R}$$

Dies ist immer wohldefiniert in $[0, \infty]$, kann aber $+\infty$ werden. Seien X_1, \dots, X_n i.i.d. Zufallsvariablen, für welche die momenterzeugende Funktion $M_X(t)$ für alle $t \in \mathbb{R}$ endlich ist. Für jedes $b \in \mathbb{R}$ gilt dann

$$P[S_n \geq b] \leq \exp \left(\inf_{t \in \mathbb{R}} (n \log M_X(t) - tb) \right)$$

Dies folgt aus der Markov-Ungleichung mit $g(x) = e^{tx}$. Daraus folgt: Seien X_1, \dots, X_n unabhängig mit $X_i \sim Be(p_i)$ und $S_n = \sum_{i=1}^n X_i$. Sei ferner $\mu_n := E[S_n] = \sum_{i=1}^n p_i$ und $\delta > 0$. Dann gilt

$$P[S_n \geq (1 + \delta)\mu_n] \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\mu_n}$$

Statistik

Statistische Grundlagen

Die Grundidee der induktiven Statistik ist, dass man die Daten x_1, \dots, x_n als Realisierung $X_1(\omega), \dots, X_n(\omega)$ von Zufallsvariablen X_1, \dots, X_n auffasst und dann (unter geeigneten Zusatzannahmen) Aussagen über die Verteilung von X_1, \dots, X_n sucht. Die Gesamtheit der Beobachtungen x_1, \dots, x_n oder Zufallsvariablen X_1, \dots, X_n nennt man oft eine Stichprobe, die Anzahl n heisst dann der Stichprobenumfang. Ausgangspunkt der Betrachtungen ist in der Regel ein Datensatz x_1, \dots, x_n aus einer Stichprobe X_1, \dots, X_n , für die ein Modell gesucht wird. Dieses ist beschreibbar durch einen (möglicherweise hochdimensionalen) Parameter $\vartheta \in \Theta$. Man betrachtet simultan eine ganze Familie von Wahrscheinlichkeitsräumen; man hat einen festen Grundraum (Ω, \mathcal{F}) und für jeden Parameter ϑ aus dem Parameterraum Θ ein Wahrscheinlichkeitsmass P_ϑ auf (Ω, \mathcal{F}) . In vielen Fällen ist der Parameterraum Θ eine Teilmenge von \mathbb{R}^m . Zu der parametrischen statistischen Analyse gehören die folgenden Schritte:

1. Wahl eines (parametrischen) Modells: Hier wird die Parametermenge Θ und die Familie $(P_\vartheta)_{\vartheta \in \Theta}$ von Modellen, mit denen man arbeiten will, spezifiziert.
2. Schätzung der Parameter: Aufgrund der Daten will man ein möglichst gut passendes Modell wählen. Dazu benutzt man einen Schätzer; die dazugehörige Schätzfunktion ist eine Abbildung, die gegebenen Daten einen Parameter $\vartheta \in \Theta$ zuordnet.
3. Kritische Modellüberprüfung (Anpassungstest): Hier fragt man, ob die Daten zu dem gewählten Parameter ϑ bzw. Modell P_ϑ gut passen; das macht man mit einem geeigneten statistischen Test.

4. Aussagen über Zuverlässigkeit der Schätzungen: Statt eines einzigen Parameterwertes kann man auch versuchen, einen Bereich in Θ so zu spezifizieren, dass die zugehörige Modelle P_ϑ gut zu den Daten passen; man spricht dann von einem Konfidenzbereich.

Schätzer

Sei X_1, \dots, X_n eine Stichprobe, für die wir ein Modell suchen. Wir haben also einen Parameterraum Θ und für jedes $\vartheta \in \Theta$ einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P_\vartheta)$. Meistens ist $\Theta \subseteq \mathbb{R}^m$, und wir suchen dann für die Parameter $\vartheta_1, \dots, \vartheta_m$ Schätzer T_1, \dots, T_m aufgrund unserer Stichprobe. Solche Schätzer sind Zufallsvariablen der Form $T_j = t_j(X_1, \dots, X_n)$, wobei die Schätzfunktionen $t_j: \mathbb{R}^n \rightarrow \mathbb{R}$ geeignet gewählt werden müssen. Einsetzen von Daten $X_i(\omega), i = 1, \dots, n$ liefert dann die Schätzwerte $T_j(\omega) = t_j(x_1, \dots, x_n)$ für $\vartheta_j, j = 1, \dots, m$.

Eigenschaften

1. Ein Schätzer T heisst erwartungstreu für ϑ , falls gilt $E_\vartheta[T] = \vartheta$, im Mittel (über alle denkbaren Realisationen ω) schätzt T also richtig. Allgemein heisst $E_\vartheta[T] - \vartheta$ der Bias (oder erwartete Schätzfehler) von T ; erwartungstreu bedeutet also, dass der Bias null ist. Der mittlere quadratische Schätzfehler (mean-squared error, MSE) ist definiert als $\text{MSE}_\vartheta[T] := E_\vartheta[(T - \vartheta)^2]$. Der MSE kann zerlegt werden in

$$\text{MSE}_\vartheta[T] = E_\vartheta[(T - \vartheta)^2] = \text{Var}_\vartheta[T] + (E_\vartheta[T] - \vartheta)^2$$

Für erwartungstreu Schätzer ist somit die Varianz und der MSE dasselbe.

2. Eine Folge von Schätzern $T^{(n)}, n \in \mathbb{N}$ heisst konsistent für ϑ , falls $T^{(n)}$ für $n \rightarrow \infty$ in P_ϑ -Wahrscheinlichkeit gegen ϑ konvergiert, d.h. für jedes $\vartheta \in \Theta$ gilt

$$\lim_{n \rightarrow \infty} P_\vartheta \left[\left| T^{(n)} - \vartheta \right| > \varepsilon \right] = 0 \quad \text{für jedes } \varepsilon > 0$$

Maximum-Likelihood-Methode (ML-Methode)

Die Likelihood-Funktion ist

$$L(x_1, \dots, x_n; \vartheta) := \begin{cases} p(x_1, \dots, x_n; \vartheta) & \text{im diskreten Fall} \\ f(x_1, \dots, x_n; \vartheta) & \text{im stetigen Fall.} \end{cases}$$

Die Funktion $\log L(x_1, \dots, x_n; \vartheta)$ heisst *log-Likelihood-Funktion*. Sie hat den Vorteil, dass sie im i.i.d-Fall durch eine Summe gegeben ist und damit oft einfacher zum Rechnen ist.

Der Maximum-Likelihood-Schätzer (ML-Schätzer) T für ϑ wird dadurch definiert, dass er

$$\vartheta \mapsto L(X_1, \dots, X_n; \vartheta)$$

als Funktion von ϑ maximiert. *Anmerkung: In den Rechnungen arbeitet man oft mit $L(x_1, \dots, x_n; \vartheta)$, wobei das optimale ϑ^* eine Funktion $t(x_1, \dots, x_n)$ von x_1, \dots, x_n ist. Damit der resultierende Schätzer T von der Stichprobe X_1, \dots, X_n abhängt, muss x_1, \dots, x_n durch X_1, \dots, X_n ersetzt werden.*

Sind X_1, \dots, X_n i.i.d. $\sim Be(p)$, so ist der ML-Schätzer $T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$. Sind X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$, ist der ML-Schätzer für $\vartheta = (\mu, \sigma^2)$:

$$T_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$$

$$T_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2$$

T_2 ist jedoch nicht erwartungstreu. Ein erwartungstreuer Schätzer ist die empirische Stichprobenvarianz S^2 :

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} (\bar{X}_n)^2$$

Momentenmethode

Seien X_1, \dots, X_n i.i.d. unter (jedem) P_ϑ und X eine Zufallsvariable mit der gleichen Verteilung wie jedes der X_i . Für eine Funktion $h: \Theta \rightarrow \mathbb{R}^d$ sollen die d Grössen $h_j(\vartheta), j = 1, \dots, d$ geschätzt werden. Zur Bestimmung des Momentenschätzers für $h(\vartheta)$ wird in der Regel folgendermassen vorgegangen:

1. Für $j = 1, \dots, d$ wird in jedem Modell P_ϑ das j -te Moment $m_j(\vartheta) = E_\vartheta[X^j]$ als Funktion des Parameters ϑ berechnet.
2. Für $j = 1, \dots, d$ wird das j -te empirische Mittel / Stichprobenmittel folgendermassen definiert:

$$\tilde{m}_j(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n x_i^j$$

3. Das System von d Gleichungen $\tilde{m}_j(x_1, \dots, x_n) = m_j(\vartheta), j = 1, \dots, d$ wird betrachtet und als ein Gleichungssystem für die d Unbekannten $h_1(\vartheta), \dots, h_d(\vartheta)$ aufgefasst. Falls eine eindeutige Lösung existiert, so wird diese $t_{\text{MM}}(x_1, \dots, x_n)$ genannt; das gibt also eine Funktion $t_{\text{MM}}: \mathbb{R}^n \rightarrow \mathbb{R}^d$. Der Momentenschätzer für $h(\vartheta)$ ist dann definiert durch $T_{\text{MM}} := t_{\text{MM}}(X_1, \dots, X_n)$.

Verteilungsaussagen

Einen allgemeinen approximativen Zugang für Verteilungsaussagen liefert der Zentrale Grenzwertsatz. Oft ist ein Schätzer T eine Funktion einer Summe $\sum_{i=1}^n Y_i$, wobei die Y_i im Modell P_ϑ i.i.d. sind. Dies ist nach dem ZGS approximativ normalverteilt unter P_ϑ mit Parametern $\mu = nE_\vartheta[Y_i]$ und $\sigma^2 = n\text{Var}_\vartheta[Y_i]$. Für normalverteilte Stichproben hat man exakte Aussagen. Seien also X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt:

1. $\bar{X}_n \sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$ und $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$
2. $\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ ist χ^2 -verteilt mit $n-1$ Freiheitsgraden.
3. \bar{X}_n und S^2 sind unabhängig.

4. Der folgende Quotient ist t -verteilt mit $n - 1$ Freiheitsgraden:

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\frac{S/\sigma}{\sqrt{\frac{1}{n-1} \frac{n-1}{\sigma^2} S^2}}} = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\frac{S/\sigma}{\sqrt{\frac{1}{n-1} \frac{n-1}{\sigma^2} S^2}}}$$

Tests

Es wird wieder eine Familie von Wahrscheinlichkeiten P_ϑ mit $\vartheta \in \Theta$ betrachtet, die die möglichen Modelle beschreiben. Die Vermutung, wo in Θ der richtige (aber unbekannte) Parameter ϑ liegen könnte, soll mit Hilfe der Daten überprüft ("getestet") werden. Das Grundproblem ist somit, eine Entscheidung zwischen zwei konkurrierenden Modellklassen zu treffen - der Hypothese $\Theta_0 \subseteq \Theta$ und der Alternative $\Theta_A \subseteq \Theta$, wobei $\Theta_0 \cap \Theta_A = \emptyset$ ist. Meist schreibt man dies als Hypothese $H_0 : \vartheta \in \Theta_0$ sowie Alternative $H_A : \vartheta \in \Theta_A$. Wenn keine explizite Alternative spezifiziert ist, so gilt $\Theta_A = \Theta_0^c = \Theta \setminus \Theta_0$. Hypothesen und / oder Alternativen heissen einfach, falls Θ_0 bzw. Θ_A aus einem einzelnen Wert, ϑ_0 bzw. ϑ_A bestehen, also z.B. $\Theta_0 = \{\vartheta_0\}$ ist; sonst heissen sie zusammengesetzt. Explizit formuliert ist die Hypothese also: $H_0 : \text{"der wahre (aber unbekannte) Parameter } \vartheta \text{ liegt in der Menge } \Theta_0 \text{ und die Alternative: } H_A : \text{"der wahre Parameter liegt in } \Theta_A \text{"}$. Ein Test ist nun allgemein eine Entscheidungsregel, die zu gegebenen Daten x_1, \dots, x_n einen von zwei Werten liefert (0 oder 1), wobei der Wert 1 als die Entscheidung, die Hypothese H_0 bzw. Θ_0 abzulehnen, interpretiert wird. Konkreter sieht dies meist so aus: Man hat eine (messbare) Abbildung $t : \mathbb{R}^n \rightarrow \mathbb{R}$, $(x_1, \dots, x_n) \mapsto t(x_1, \dots, x_n)$ und einen kritischen Bereich oder Verwerfungsbereich $K \subseteq \mathbb{R}$. Die Zufallsvariable $T = t(X_1, \dots, X_n)$ heisst dann Teststatistik und die Entscheidungsregel ist $I_{\{t(x_1, \dots, x_n) \in K\}}$. Man verwirft die Hypothese also genau dann, wenn der realisierte Wert $t(x_1, \dots, x_n) = t(X_1(\omega), \dots, X_n(\omega)) = T(\omega)$ im Verwerfungsbereich K liegt. Die Entscheidung bei einem Test kann auf zwei verschiedene Arten falsch herauskommen:

1. Fehler 1. Art: Die Hypothese wird zu Unrecht abgelehnt, d.h. obwohl sie richtig ist. Das passiert für $\vartheta \in \Theta_0$ und $T \in K$, deshalb heisst $P_\vartheta[T \in K]$ für $\vartheta \in \Theta_0$ die Wahrscheinlichkeit für einen Fehler 1. Art.
2. Fehler 2. Art: Die Hypothese wird zu Unrecht nicht verworfen, d.h. man akzeptiert die Hypothese, obwohl sie falsch ist. Das passiert für $\vartheta \in \Theta_A$ und $T \notin K$ und deshalb heisst $P_\vartheta[T \notin K] = 1 - P_\vartheta[T \in K]$ für $\vartheta \in \Theta_A$ die Wahrscheinlichkeit für einen Fehler 2. Art.

Man möchte grundsätzlich einen Test immer so konstruieren (d.h. T und K wählen), dass beide Fehler möglichst klein werden. Das heisst, man möchte die Funktion $\vartheta \mapsto P_\vartheta[T \in K]$ auf Θ_0 möglichst klein und auf Θ_A möglichst gross haben. Da diese Funktion in der Regel jedoch stetig ist und $\Theta = \Theta_0 \cup \Theta_A$, so dass Θ_0 und Θ_A direkt nebeneinander liegen, ist ein gleichzeitiges Minimieren / Maximieren in der Regel grundsätzlich nicht möglich. Deswegen hat sich das folgende Vorgehen durchgesetzt:

1. Man wählt zuerst ein Signifikanzniveau $\alpha \in (0, 1)$ und sorgt zunächst für

$$\sup_{\vartheta \in \Theta_0} P_\vartheta[T \in K] \leq \alpha$$

d.h. man kontrolliert die Wahrscheinlichkeit für einen Fehler 1. Art durch α

2. Anschliessend versucht man, die Macht des Tests, also die Funktion

$$\beta : \Theta_A \rightarrow [0, 1], \quad \vartheta \mapsto \beta(\vartheta) := P_\vartheta[T \in K]$$

möglichst gross zu bekommen. Dies bedeutet also, dass man die Grösse $1 - \beta(\vartheta) = P_\vartheta[T \notin K]$ für $\vartheta \in \Theta_A$, also die Wahrscheinlichkeit für einen Fehler 2. Art, möglichst klein bekommen will.

Das asymmetrische Verfahren macht es schwieriger, die Hypothese zu verwerfen als sie beizubehalten. Deswegen sollte als Hypothese immer die Negation der eigentlich gewünschten Aussage benutzt werden. Es kann passieren, dass die gleiche inhaltliche Frage zu unterschiedlichen Antworten führt, wenn bei ihrem Test Hypothese und Alternative vertauscht werden.

Konstruktion von Tests

Sei $L(x_1, \dots, x_n; \vartheta)$ die Likelihood-Funktion. Für $\vartheta_0 \in \Theta_0$ und $\vartheta_A \in \Theta_A$ betrachten wir den Likelihood-Quotienten:

$$R(x_1, \dots, x_n; \vartheta_0, \vartheta_A) := \frac{L(x_1, \dots, x_n; \vartheta_0)}{L(x_1, \dots, x_n; \vartheta_A)}$$

Ist dieser Quotient klein, so sind die Beobachtungen x_1, \dots, x_n als Resultate im Modell P_{ϑ_A} deutlich wahrscheinlicher als im Modell P_{ϑ_0} , die Daten sprechen also gegen ϑ_0 im Vergleich zu ϑ_A . Als Teststatistik kann somit $T := R(X_1, \dots, X_n; \vartheta_0, \vartheta_A)$ gewählt werden und als kritischen Bereich $K := [0, c]$. Das Neyman-Pearson-Lemma besagt: Sei $\Theta_0 = \{\vartheta_0\}$ und $\Theta_A = \{\vartheta_A\}$, T und K wie oben gewählt und $\alpha^* := P_{\vartheta_0}[T \in K] = P_{\vartheta_0}[T < c]$. Dieser Test ist dann folgendermassen optimal: Jeder andere Test mit Signifikanzniveau $\alpha \leq \alpha^*$ hat kleinere Macht (also eine grössere Wahrscheinlichkeit für einen Fehler 2. Art). Bei zusammengesetzten Hypothesen und Alternativen wird der verallgemeinerte Likelihood-Quotient betrachtet:

$$R(x_1, \dots, x_n) := \frac{\sup_{\vartheta \in \Theta_0} L(x_1, \dots, x_n; \vartheta)}{\sup_{\vartheta \in \Theta_A} L(x_1, \dots, x_n; \vartheta)}$$

Oder auch:

$$\tilde{R}(x_1, \dots, x_n) := \frac{\sup_{\vartheta \in \Theta_0} L(x_1, \dots, x_n; \vartheta)}{\sup_{\vartheta \in \Theta_A \cup \Theta_0} L(x_1, \dots, x_n; \vartheta)}$$

Und wählt als Teststatistik $T_0 := R(X_1, \dots, X_n)$ bzw. $\tilde{T} := \tilde{R}(X_1, \dots, X_n)$ mit kritischem Bereich $K_0 := [0, c_0]$. Durch Umformen (Beobachten, wenn der Quotient klein wird; diesen Term als Teststatistik verwenden) erhält man damit oft einfachere Tests.

Wichtige Tests

Normalverteilung, Test für Erwartungswert bei bekannter Varianz (z-Test): Hier sind X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(\vartheta, \sigma^2)$ unter P_ϑ mit bekannter Varianz σ^2 und die Hypothese $H_0 : \vartheta = \vartheta_0$ soll getestet werden. Mögliche Alternative H_A sind $\vartheta > \vartheta_0$ oder $\vartheta < \vartheta_0$ (einseitig) bzw. $\vartheta \neq \vartheta_0$ (zweiseitig). Die Teststatistik ist

$$T := \frac{\bar{X}_n - \vartheta_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{unter } P_{\vartheta_0}$$

Der kritische Bereich ist von der Form $(c_>, \infty)$ für $H_A : \vartheta > \vartheta_0$, bzw. $(-\infty, c_<)$ oder $(-\infty, -c_\neq) \cup (c_\neq, +\infty)$. Im zweiseitigen Fall verwirft man H_0 also, wenn $|T| > c_\neq$. Für die Konstanten gilt: $c_> = \Phi^{-1}(1 - \alpha) =: z_{1-\alpha}$, $c_< = z_\alpha = -z_{1-\alpha}$ und $c_\neq = z_{1-\frac{\alpha}{2}}$.

Normalverteilung, Test für Erwartungswert bei unbekannter Varianz (t-Test): Hier sind X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(\vartheta, \sigma^2)$ unter P_ϑ , wobei $\tilde{\vartheta} = (\mu, \sigma^2)$. Bzw. explizit:

$$\Theta_0 = \{\mu_0\} \times (0, \infty) = \{\tilde{v} = (\mu, \sigma^2) \mid \mu = \mu_0\}$$

Die Teststatistik lautet:

$$T := \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad \text{unter } P_{\vartheta_0}$$

wobei $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Der kritische Bereich hat (je nach Alternative) die gleichen Formen wie beim z-Test, die kritischen Werte sind: $c_> = t_{n-1, 1-\alpha}$, $c_< = t_{n-1, \alpha} = -t_{n-1, 1-\alpha}$ oder $c_\neq = t_{n-1, 1-\frac{\alpha}{2}}$. Bei Einstichproben-Tests hat man nur Daten aus einer Stichprobe, während man bei Zweistichproben-Test von Zufallsvariablen X_1, \dots, X_n und Y_1, \dots, Y_m ausgeht, die unter P_ϑ alle unabhängig sind; ausserdem sind die X_i und die Y_j unter P_ϑ jeweils für sich betrachtet i.i.d.

Gepaarter Zweistichproben-Test bei

Normalverteilung: Hier sind X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(\mu_X, \sigma^2)$ und Y_1, \dots, Y_n i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma^2)$ unter P_ϑ . Insbesondere ist $m = n$ und die Varianz bei beiden Stichproben dieselbe. Dies tritt bei Experimenten mit natürlicher Paarbildung zwischen X_i und Y_i (bspw. Person, die zwei Dinge ausprobiert). Tests über den Vergleich von μ_X und μ_Y können auf den Fall mit nur einer Stichprobe zurückgeführt werden, die Differenzen $Z_i := X_i - Y_i$ sind nämlich unter P_ϑ i.i.d. $\sim \mathcal{N}(\mu_X - \mu_Y, 2\sigma^2)$. Somit können die bisherigen Tests benutzt werden.

Ungepaarter Zweistichproben-Test bei

Normalverteilung: Hier sind unter P_ϑ X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(\mu_X, \sigma^2)$ und Y_1, \dots, Y_m i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma^2)$, wobei die Varianz identisch ist, aber $m \neq n$ sein kann. Dieser Test wird auch verwendet, wenn zufällig $m = n$ ist, aber die Daten nicht natürlich gepaart sind.

1. Ist σ^2 bekannt, so ist die Teststatistik

$$T := \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0, 1)$$

unter jedem P_ϑ . Die kritischen Werte für den Verwerfungsbereich sind wie oben geeignete Quantile der Normalverteilung, je nach Alternative. Bei diesem Test handelt es sich um den ungepaarten Zweistichproben-z-Test.

2. Ist σ^2 unbekannt, so brauchen wir zuerst die beiden empirischen Varianzen $S_X^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ und $S_Y^2 := \frac{1}{m-1} \sum_{j=1}^m (Y_j - \bar{Y}_m)^2$. Mit

$$S^2 := \frac{1}{m+n-2} ((n-1)S_X^2 + (m-1)S_Y^2)$$

ist dann die Teststatistik

$$T := \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

unter jedem P_ϑ . Der Rest erfolgt analog wie oben. Dieser Test heisst ungepaarter Zweistichproben-t-Test.

Konfidenzbereiche

Ein Konfidenzbereich für ϑ zu Daten x_1, \dots, x_n ist eine Menge $C(x_1, \dots, x_n) \subseteq \Theta$; in den meisten Fällen ist das ein Intervall, dessen Endpunkte von x_1, \dots, x_n abhängen. $C(X_1, \dots, X_n)$ ist also eine zufällige Teilmenge von Θ mit Realisierung $C(\omega) = C(X_1(\omega), \dots, X_n(\omega))$ bei einem festen ω . Ein solches C heisst ein Konfidenzbereich zum Niveau $1 - \alpha$, wenn gilt

$$P_\vartheta [C(X_1, \dots, X_n) \ni \vartheta] \geq 1 - \alpha \quad \text{für alle } \vartheta \in \Theta$$

Wichtige Konfidenzbereiche

Seien X_1, \dots, X_n unter P_ϑ i.i.d. $\mathcal{N}(\mu, \sigma^2)$. Das Konfidenzintervall für μ zum Niveau $1 - \alpha$ lautet:

$$C(X_1, \dots, X_n) = \left[\bar{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$$

Das Konfidenzintervall für σ^2 zum Niveau $1 - \alpha$ lautet:

$$C(X_1, \dots, X_n) = \left[\frac{(n-1)S^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)S^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right]$$

Zusammenhang mit Tests

Sei $C(X_1, \dots, X_n)$ ein Konfidenzbereich für ϑ zum Niveau $1 - \alpha$. Um die Hypothese $H_0 : \vartheta = \vartheta_0$ zu testen, wird durch $I_{\{\vartheta_0 \notin C(X_1, \dots, X_n)\}}$ ein Test definiert, d.h. H_0 wird genau dann abgelehnt, wenn ϑ_0 nicht in $C(X_1, \dots, X_n)$ liegt. Dann gilt wegen $\Theta_0 = \{\vartheta_0\}$ für jedes $\vartheta \in \Theta_0$:

$$P_\vartheta [\vartheta_0 \notin C(X_1, \dots, X_n)] = 1 - P_{\vartheta_0} [C(X_1, \dots, X_n) \ni \vartheta_0] \leq \alpha$$

Aus einem Konfidenzberiech für ϑ erhält man somit eine ganze Familie von Tests, nämlich einen für jede einfache Hypothese $\Theta_0 = \{\vartheta_0\}$ mit einem $\vartheta_0 \in \Theta$.

Sei umgekehrt für jede einfache Hypothese $\Theta_0 = \{\vartheta_0\}$ ein Test zum Niveau α gegeben; für jedes ϑ_0 haben wir also einen kritischen Bereich K_{ϑ_0} , so dass $H_0 : \vartheta = \vartheta_0$ genau dann

abgelehnt wird, wenn $(X_1, \dots, X_n) \in K_{\vartheta_0}$ ist. Zudem gilt wegen Niveau α

$$P_{\vartheta_0} [(X_1, \dots, X_n) \in K_{\vartheta_0}] \leq \alpha \quad \text{für jedes } \vartheta_0 \in \Theta$$

Nun wird eine Teilmenge $C(X_1, \dots, X_n)$ von Θ durch folgende Bedingung definiert:

$$\vartheta \in C(X_1, \dots, X_n) \quad :\Longleftrightarrow \quad (X_1, \dots, X_n) \notin K_\vartheta$$

Dies ist ein Konfidenzbereich zum Niveau $1 - \alpha$, denn für jedes $\vartheta \in \Theta$ gilt

$$\begin{aligned} P_\vartheta [C(X_1, \dots, X_n) \ni \vartheta] &= P_\vartheta [(X_1, \dots, X_n) \notin K_\vartheta] \\ &= 1 - P_\vartheta [(X_1, \dots, X_n) \in K_\vartheta] \geq 1 - \alpha \end{aligned}$$

Kombinatorik

1. Anzahl Möglichkeiten, n Objekte (z.B. nebeneinander) anzuordnen: $n!$
2. Anzahl Möglichkeiten, k aus den n Objekten auszuwählen (mit $k \leq n$, ohne Zurücklegen):

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

3. Sequenzen der Länge m mit n Symbolen (Variationen mit Wiederholung): n^m .

Analysis

Differenzialrechnung

Ableitungsregeln Sind $f, g : \Omega \subset \mathbb{R} \rightarrow \mathbb{R}$ an der Stelle $x_0 \in \Omega$ differenzierbar, so sind es auch $f + g$, $f * g$ und $\frac{f}{g}$ (falls g keine Nullstellen hat). Die Ableitungen lauten:

$$(f + g)'(x_0) = f'(x_0) + g'(x_0) \quad (\text{Summenregel})$$

$$(f \cdot g)(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0) \quad (\text{Produktregel})$$

$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g^2(x_0)} \quad (\text{Quotientenregel})$$

Seien $f : \Omega \subset \mathbb{R} \rightarrow \mathbb{R}$ und $g : U \rightarrow \mathbb{R}$ mit $U \subset f(\Omega)$ an den Stellen x_0 bzw. $f(x_0)$ differenzierbar. Dann ist die Komposition $g \circ f : \Omega \rightarrow \mathbb{R}$ an der Stelle x_0 differenzierbar und die Ableitung lautet:

$$(g \circ f)'(x_0) = (g(f(x_0)))' = g'(f(x_0)) \cdot f'(x_0)$$

Integralrechnung

Eine Stammfunktion F zu einer gegebenen stetigen Funktion f ist eine Funktion, so dass $F'(x) = f(x)$ für alle x im Definitionsbereich von f gilt. D.h.

$$F(x) = \int f(x)dx = \int F'(x)dx$$

Ist F eine Stammfunktion, ist auch $F + C$ (mit $C \in \mathbb{R}$) eine. Differenzierbare Funktionen sind stetig und nach dem Hauptsatz integrierbar, die Umkehrungen gelten nicht (es gibt integrierbare Funktionen, welche nicht stetig sind, bspw. stückweise stetige Funktionen).

Elementare Integrale Siehe Anhang, können durch Umkehrung der Ableitung direkt angegeben werden.

Direkte Integrale Es gilt:

$$\int f(g(x))g'(x)dx = F(g(x))$$

D.h. wenn der Integrand die Form $f(g(x)) \cdot g'(x)$ hat, wird die Stammfunktion von f in $g(x)$ ausgewertet.

Partielle Integration

$$\int f' \cdot g dx = f \cdot g - \int f \cdot g' dx$$

Der Integrand wird also als Produkt von zwei Funktionen geschrieben, wobei die Funktion f' integriert (\uparrow) und g

abgeleitet (\downarrow) wird. Als Merkhilfe kann die folgende Tabelle dienen:

\uparrow	1 (falls arc-Funktion oder Logarithmus vorkommt), $x^n, \frac{1}{1-x^2}, \frac{1}{1+x^2}, \dots$
\downarrow	$x^n, \log(x), \arcsin(x), \arccos(x), \arctan(x), \operatorname{arcsinh}(x), \operatorname{arcosh}(x), \operatorname{arctanh}(x), \dots$
”egal”	$e^x, \sin(x), \cos(x), \sinh(x), \cosh(x), \dots$

Integrale rationaler Funktionen Integrale der Form

$\int \frac{p(x)}{q(x)} dx$ wobei $p(x)$ und $q(x)$ zwei Polynome sind.

- Falls $\text{Grad}(p) \geq \text{Grad}(q) \Rightarrow$ führe die Polynomdivision $p(x) : q(x)$ durch.
- Falls $\text{Grad}(p) < \text{Grad}(q) \Rightarrow$ führe die Partialbruchzerlegung (PBZ) durch (siehe Anhang für Anleitung).

Substitutionsregel Die Idee ist, eine

Variablentransformation $y = g(x) \Leftrightarrow x = g^{-1}(y)$ durchzuführen. Ist f stetig und g wie oben, gilt dann:

$$\int f(g(x))g'(x)dx = \int f(y)dy$$

Wobei $dy = g'(x)dx$. Der Ansatz lautet also: Um $\int f(y)dy$ zu berechnen, substituiere $y = g(x)$ im Integrand und ersetze das dy mit $g'(x)dx$ mit dem Ziel, das neue Integral $\int f(g(x)) \cdot g'(x)dx$ leichter zu bestimmen.

Substitutionsansätze

- Integrale von Funktionen, die $e^x, \sinh(x), \cosh(x), \dots$ enthalten: Diese Integrale werden oft mit der Substitution $e^x = t$ ($dx = \frac{1}{t} dt$) gelöst.
- Integrale von Funktionen, die $\log x$ enthalten: Diese werden oft mit der Substitution $\log x = t$ ($dx = e^x dt$) gelöst.
- Integrale von Funktionen, die $\sqrt[n]{Ax+B}$ enthalten. Diese werden oft mit der Substitution $\sqrt[n]{Ax+B} = t$ gelöst.
- Integrale von Funktionen, die $\cos x, \sin x$ in geraden Potenzen oder $\tan x$ enthalten. Diese Integrale werden oft mit der Substitution $\tan x = t$ ($dx = \frac{1}{1+t^2} dt$) gelöst. Bei dieser Substitution wird $\sin^2 x$ zu $\frac{t^2}{1+t^2}$ und $\cos^2 x$ zu $\frac{1}{1+t^2}$

- Integrale von Funktionen, die $\cos x, \sin x$ in ungeraden Potenzen enthalten. Diese werden oft mit der Substitution $\tan\left(\frac{x}{2}\right) = t \quad \left(dx = \frac{2}{1+t^2} dt\right)$ gelöst.
- Integrale mit $\sqrt{Ax^2 + Bx + C}$ im Nenner. Diese werden mithilfe der quadratischen Ergänzung auf eine der folgenden Fälle gebracht: $\int \frac{1}{\sqrt{1-x^2}} dx = \arcsin x + C$,
 $\int \frac{1}{\sqrt{x^2-1}} dx = \operatorname{arccosh} x + C$ oder
 $\int \frac{1}{\sqrt{1+x^2}} dx = \operatorname{arcsinh} x + C$.
- Integrale mit $\sqrt{Ax^2 + Bx + C}$ im Zähler. Diese werden mithilfe der quadratischen Ergänzung auf eine der folgenden Formen gebracht und dann mit der Substitution gelöst:

$$\int \sqrt{1-x^2} dx \quad \text{Substitution } x = \sin t$$

$$\int \sqrt{x^2-1} dx \quad \text{Substitution } x = \cosh t$$

$$\int \sqrt{1+x^2} dx \quad \text{Substitution } x = \sinh t$$

Bestimmte Integrale

Das bestimmte Integral von f von a bis b ($a < b$) ist:

$$\int_a^b f(x) dx := F(b) - F(a) = [F(x)]_a^b$$

Für $a > b$ setzt man: $\int_a^b f(x) dx = -\int_b^a f(x) dx$. Ausserdem definiert man $\int_a^a f(x) dx = 0$. Das bestimmte Integral hat die folgenden Eigenschaften (1 - Linearität, 2 - Gebietsadditivität, 3 - Positivität, 4 - Monotonie und 5 - Dreiecksungleichung):

$$\int_a^b (\alpha f(x) + \beta g(x)) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx \quad (1)$$

$$\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx \quad (2)$$

$$f(x) \geq 0 \forall x \in [a, b], \text{ so ist } \int_a^b f(x) dx \geq 0 \quad (3)$$

$$f(x) \geq g(x) \forall x \in [a, b], \text{ so ist } \int_a^b f(x) dx \geq \int_a^b g(x) dx \quad (4)$$

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx \quad (5)$$

Ist $f(x)$ ungerade, so gilt für alle um den Ursprung symmetrischen Integrale:

$$\int_{-a}^a f(x) dx = 0$$

Aus dem Hauptsatz der Integralrechnung folgt:

$$\frac{d}{dx} \int_a^x f(t) dt = f(x)$$

Integralrechnung in \mathbb{R}^n

Das Riemann-Integral in \mathbb{R}^n

Für alle beschränkten, geschlossenen Teilmengen $X \subset \mathbb{R}^n$ und alle stetigen Funktionen $f: X \rightarrow \mathbb{R}$, kann man das Integral von f über X definieren, notiert mit:

$$\int_X f(x) dx$$

Dabei handelt es sich um eine reelle Zahl und es erfüllt die folgenden Bedingungen:

1. **Compatibility:** Für $n = 1$ und $X = [a, b]$ (mit $a \leq b$) ist das Integral von f über X das Riemann-Integral von f :

$$\int_{[a,b]} f(x) dx = \int_a^b f(x) dx$$

2. **Linearity:** Wenn f und g stetig auf X sind und a, b reelle Zahlen sind, dann:

$$\int_X (af_1(x) + bf_2(x)) dx = a \int_X f_1(x) dx + b \int_X f_2(x) dx$$

3. **Positivity:** Wenn $f \leq g$, dann:

$$\int_X f(x) dx \leq \int_X g(x) dx$$

Wenn $f \geq 0$ somit:

$$\int_X f(x) dx \geq 0$$

Wenn $Y \subset X$ kompakt ist und $f \geq 0$:

$$\int_Y f(x) dx \leq \int_X f(x) dx$$

4. **Upper Bound / triangle inequality:**

$$\left| \int_X f(x) dx \right| \leq \int_X |f(x)| dx$$

$$\left| \int_X (f(x) + g(x)) dx \right| \leq \int_X |f(x)| dx + \int_X |g(x)| dx$$

5. **Volume:** Für $f = 1$ ist das Integral von f das Volumen in \mathbb{R}^n der Menge X und wenn $f \geq 0$, das Integral von f ist das Volumen der Menge:

$$\{(x, y) \in X \times \mathbb{R} | 0 \leq y \leq f(x)\} \subset \mathbb{R}^{n+1}$$

Wenn X ein beschränktes Rechteck ist, also $X = [a_1, b_1] \times \dots \times [a_n, b_n] \subset \mathbb{R}^n$ und $f = 1$, dann gilt:

$$\int_X dx = (b_n - a_n) \dots (b_1 - a_1)$$

6. **Multiple integral / Fubini's Theorem** Wenn n_1, n_2 natürliche Zahlen ≥ 1 sind so dass $n = n_1 + n_2$, definiere für $x_1 \in \mathbb{R}^{n_1}$:

$$Y_{x_1} = \{x_2 \in \mathbb{R}^{n_2} | (x_1, x_2) \in X\} \subset \mathbb{R}^{n_2}$$

Sei X_1 die Menge der $x_1 \in \mathbb{R}^{n_1}$ so dass Y_{x_1} nicht leer ist. X_1 ist kompakt in \mathbb{R}^{n_1} und Y_{x_1} ist kompakt in \mathbb{R}^{n_2} für alle $x_1 \in X_1$. Wenn die Funktion auf X_1

$$g(x_1) = \int_{Y_{x_1}} f(x_1, x_2) dx_2$$

stetig ist, dann:

$$\int_X f(x_1, x_2) dx = \int_{X_1} g(x_1) dx_1 = \int_{X_1} \left(\int_{Y_{x_1}} f(x_1, x_2) dx_2 \right) dx_1$$

Tauscht man die Rolle von x_1 und x_2 aus, erhält man:

$$\int_X f(x_1, x_2) dx = \int_{X_2} \left(\int_{Z_{x_2}} f(x_1, x_2) dx_1 \right) dx_2$$

wobei $Z_{x_2} = \{x_1 | (x_1, x_2) \in X\}$.

Eine Teilmenge von \mathbb{R}^2 heisst y -Normalbereich, wenn sie sich wie folgt darstellen lässt:

$\Omega = \{(x, y) \in \mathbb{R}^2 | a \leq x \leq b, f(x) \leq y \leq g(x)\}$ Beim x -Normalbereich sind die Rollen vertauscht: $\Omega = \{(x, y) \in \mathbb{R}^2 | c \leq y \leq d, h(y) \leq x \leq k(y)\}$ Für einen y -Normalbereich gilt mit Fubini's Theorem:

$$\int_{\Omega} f(x, y) dx = \int_a^b \int_{f(x)}^{g(x)} f(x, y) dy dx$$

Und für einen x -Normalbereich:

$$\int_{\Omega} f(x, y) dx = \int_c^d \int_{h(y)}^{k(y)} f(x, y) dx dy$$

Das Konzept lässt sich auf \mathbb{R}^n erweitern, wobei die Grenzen immer von den vorherigen Variablen abhängen müssen.

7. **Domain additivity:** Wenn X_1 und X_2 kompakte Teilmengen von \mathbb{R}^n sind und f stetig auf $X_1 \cup X_2$ ist, dann gilt:

$$\int_{X_1 \cup X_2} f(x) dx + \int_{X_1 \cap X_2} f(x) dx = \int_{X_1} f(x) dx + \int_{X_2} f(x) dx$$

Insbesondere also wenn $X_1 \cap X_2$ leer ist:

$$\int_{X_1 \cup X_2} f(x) dx = \int_{X_1} f(x) dx + \int_{X_2} f(x) dx$$

Uneigentliche Integrale Sei $I \subset \mathbb{R}$ ein beschränkter Intervall und $J = [a, +\infty[$ für ein $a \in \mathbb{R}$. Sei f eine stetige Funktion auf $X = J \times I$. f ist Riemann-integrierbar auf X wenn der folgende Limes existiert:

$$\lim_{x \rightarrow +\infty} \int_{[a,x] \times I} f(x, y) dx dy = \lim_{x \rightarrow +\infty} \int_a^x \left(\int_I f(x, y) dy \right) dx$$

$$= \lim_{x \rightarrow +\infty} \int_I \left(\int_a^x f(x, y) dx \right) dy$$

Sei f stetig auf \mathbb{R}^2 und $f \geq 0$. Man sagt, dass f auf \mathbb{R}^2 Riemann-integrierbar ist, wenn der Limes existiert:

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \lim_{R \rightarrow +\infty} \int_{[-R, R]^2} f(x, y) dx dy$$

Mit Fubinis Formel erhalten wir für den obigen Integral auch:

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{+\infty} f(x, y) dy \right) dx = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x, y) dx \right) dy$$

Wenn $|f| \leq g$ und man weiss, dass das unbestimmte Integral von g existiert, folgt daraus die Existenz des unbestimmten Integrals von f .

Change of variable Seien $\bar{X} \subset \mathbb{R}^n$ und $\bar{Y} \subset \mathbb{R}^n$ kompakte Teilmengen, $\varphi: \bar{X} \rightarrow \bar{Y}$ eine stetige Abbildung. Wir nehmen an, dass wir $\bar{X} = X \cup B$ und $\bar{Y} = Y \cup C$ schreiben können, wobei X und Y offen sind, B und C vernachlässigbar sind und

φ auf X eine C^1 bijektive Abbildung von X nach Y ist. Dann gilt:

$$\int_{\bar{X}} f(\varphi(x)) |\det(J_\varphi(x))| dx = \int_{\bar{Y}} f(y) dy$$

Eine wichtige Anwendung sind Polarkoordinaten mit

$$\bar{Y}_R = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq R^2\}$$

Man definiert $\varphi: \bar{X}_R \rightarrow \bar{Y}_R$ mit $\bar{X}_R = [0, R] \times [-\pi, \pi]$ und

$$\varphi(r, \theta) = (r \cos(\theta), r \sin(\theta))$$

Die Jacobi-Matrix ist

$$J_\varphi(r, \theta) = \begin{pmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{pmatrix}$$

und $\det(J_\varphi(r, \theta)) = r$. Somit ergibt sich:

$$\int_{\Delta} f(x, y) dx dy = \int_0^R \int_a^b f(r \cos \theta, r \sin \theta) r dr d\theta$$

Eine weitere wichtige Anwendung sind Kugelkoordinaten (r, θ, φ) in \mathbb{R}^3 , wobei $r \geq 0, 0 \leq \theta \leq 2\pi$ und $0 \leq \varphi \leq \pi$. Die Determinante ist $-r^2 \sin(\varphi)$, somit ergibt sich:

$$\int_B f(x, y, z) dx dy dz =$$

$$\int_0^R \int_0^{2\pi} \int_0^\pi f(r \cos \theta \sin \varphi, r \sin \theta \sin \varphi, r \cos \varphi) r^2 \sin(\varphi) d\varphi d\theta dr$$

Ausserdem existieren in \mathbb{R}^2 elliptische Koordinaten $(0 \leq r < \infty, 0 \leq \varphi < 2\pi)$ mit $x = r a \cos \varphi, y = r b \sin \varphi, dx dy = a b r dr d\varphi$. Und in \mathbb{R}^3 Zylinderkoordinaten $(0 \leq r < \infty, 0 \leq \theta < 2\pi, -\infty < z < \infty)$ mit $x = r \cos \varphi, y = r \sin \varphi, z = z$ und $dx dy dz = r dr d\varphi dz$.

Anhang

Elementare Integrale

$f'(x)$	$f(x)$	$F(x)$
0	c	$cx + C$
$r \cdot x^{r-1}$	x^r	$\frac{x^{r+1}}{r+1} + C$
$-\frac{1}{x^2} = -x^{-2}$	$\frac{1}{x} = x^{-1}$	$\ln x + C$
$\frac{1}{2\sqrt{x}} = \frac{1}{2}x^{-\frac{1}{2}}$	$\sqrt{x} = x^{\frac{1}{2}}$	$\frac{2}{3}x^{\frac{3}{2}} + C$
$\cos(x)$	$\sin(x)$	$-\cos(x) + C$
$-\sin(x)$	$\cos(x)$	$\sin(x) + C$
$2\sin(x)\cos(x)$	$\sin^2(x)$	$\frac{1}{2}(x - \sin(x)\cos(x)) + C$
$-2\sin(x)\cos(x)$	$\cos^2(x)$	$\frac{1}{2}(x + \sin(x)\cos(x)) + C$
$\cos^2(x) - \sin^2(x)$	$\sin(x)\cos(x)$	$-\frac{1}{2}(\cos^2(x)) + C$
$1 + \tan^2(x) = \frac{1}{\cos^2(x)}$	$\tan(x)$	$-\ln \cos(x) + C$
e^x	e^x	$e^x + C$
$c \cdot e^{cx}$	e^{cx}	$\frac{1}{c} \cdot e^{cx} + C$
$\ln(c) \cdot c^x$	c^x	$\frac{c^x}{\ln(c)} + C$
$\frac{1}{x}$	$\ln x $	$x(\ln x - 1) + C$
$\frac{1}{\ln(a) \cdot x}$	$\log_a x $	$\frac{x}{\ln(a)}(\ln x - 1) + C$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x)$	$x \cdot \arcsin(x) + \sqrt{1-x^2} + C$
$-\frac{1}{\sqrt{1-x^2}}$	$\arccos(x)$	$x \cdot \arccos(x) - \sqrt{1-x^2} + C$
$\frac{1}{1+x^2}$	$\arctan(x)$	$x \cdot \arctan(x) - \frac{1}{2} \ln(1+x^2) + C$
$\cosh(x)$	$\sinh(x) = \frac{e^x - e^{-x}}{2}$	$\cosh(x) + C$
$\sinh(x)$	$\cosh(x) = \frac{e^x + e^{-x}}{2}$	$\sinh(x) + C$
$\frac{1}{\cosh^2(x)}$	$\tanh(x)$	$\log(\cosh(x)) + C$
$\frac{1}{\sqrt{1+x^2}}$	$\operatorname{arcsinh}(x)$	$x \cdot \operatorname{arcsinh}(x) - \sqrt{x^2 + 1} + C$
$\frac{1}{\sqrt{x^2 - 1}}$	$\operatorname{arccosh}(x)$	$x \cdot \operatorname{arccosh}(x) - \sqrt{x^2 - 1} + C$
$\frac{1}{1-x^2}$	$\operatorname{arctanh}(x)$	$x \cdot \operatorname{arctanh}(x) + \frac{1}{2} \ln(1-x^2) + C$

Trigonometrische Substitution

Ausdruck	Substitution	Identität
$\sqrt{a^2 - x^2}$	$x = a \sin \theta, -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$	$1 - \sin^2 \theta = \cos^2 \theta$
$\sqrt{a^2 + x^2}$	$x = a \tan \theta, -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$	$1 + \tan^2 \theta = \sec^2 \theta$
$\sqrt{x^2 - a^2}$	$x = a \sec \theta, 0 \leq \theta < \frac{\pi}{2} \text{ oder } \pi \leq \theta < \frac{3\pi}{2}$	$\sec^2 \theta - 1 = \tan^2 \theta$

Anleitung Partialbruchzerlegung

- Nullstellen des Nenners berechnen
- Zuordnung eines Partialbruchs zur Nullstelle:
Reelle Nullstellen: Wenn x_1 eine einfache Nullstelle ist $\rightarrow \frac{A}{x-x_1}$. Wenn es eine zweifache Nullstelle ist $\rightarrow \frac{A}{x-x_1} + \frac{B}{(x-x_1)^2}$. Bei einer r-fachen Nullstelle $\rightarrow \frac{A_1}{x-x_1} + \frac{A_2}{(x-x_1)^2} + \dots + \frac{A_r}{(x-x_1)^r}$
Nichtreelle Nullstellen: Aus quadratischem Term $x^2 + px + q$ wird $\frac{Ax+B}{x^2+px+q}$
- Ansatz zur Partialbruchzerlegung aufstellen, Koeffizienten $((A, B, C \dots))$ berechnen.
Bspw. $\frac{A}{x-x_1} + \frac{B}{x-x_2} + \frac{C}{x-x_3} = \frac{A(x-x_2)(x-x_3) + B(x-x_1)(x-x_3) + C(x-x_1)(x-x_2)}{(x-x_1)(x-x_2)(x-x_3)}$

Anleitung Integralsubstitution

Bei bestimmten Integralen kann die Grenze mitsubstituiert werden oder die Stammfunktion ermittelt werden und am Schluss eingesetzt werden.

- Eine geeignete Substitution $u = g(x)$ finden (oftmals Terme in Wurzeln / Potenzen / Brüchen).
- $\frac{dx}{du}$ (wofür zuerst $x = f(u)$ umgeformt werden muss) berechnen, nach dx umformen und einsetzen.

Beispiele

X, Y unabhängig $\sim \text{Exp}(\lambda)$. Berechne Dichte- / Verteilungsfunktion für $V := X + Y: v > 0$,
 $P[V \leq v] = \lambda^2 \int_0^\infty e^{-\lambda x} \left(\int_0^\infty 1_{\{x+y \leq v\}} e^{-\lambda y} dy \right) dx$
 $= \lambda^2 \int_0^\infty e^{-\lambda x} 1_{\{x \leq v\}} \left(\int_0^\infty 1_{\{y \leq v-x\}} e^{-\lambda y} dy \right) dx = \lambda \int_0^v e^{-\lambda x} \left(\int_0^{v-x} \lambda e^{-\lambda y} dy \right) dx$
 $= \lambda \int_0^v e^{-\lambda x} (1 - e^{-\lambda(v-x)}) dx = 1 - e^{-\lambda v} - \lambda v e^{-\lambda v}$