Diskrete Wahrscheinlichkeitstheorie

Grundlagen

Diskreter Wahrscheinlichkeitsraum

diskreter W'keitsraum	$\Omega = \{\omega_1, \omega_2, \dots\}$
Elementarereignis	ω_i
Elementarw'keit	$\Pr[\omega_i]$
	$0 \leq \Pr[\omega_i] \leq 1$
	$\sum_{\omega \in \Omega} \Pr[\omega] = 1$
endlicher W'keitsraum	$\Omega = \{\omega_1, \dots, \omega_n\}$
unendlicher W'keitsraum	$\Omega = \{\omega_1, \omega_2, \dots\}$

Ereignis

Eine Menge $E \subseteq \Omega$ heißt Ereignis mit Wahrscheinlichkeit $\Pr[E] = \sum_{\omega \in E} \Pr[\omega]$

Sind A und B Ereignisse $\implies$ $A \cup B$, $A \cap B$, $A \backslash B$, etc. sind Ereignisse
Ist $A \subseteq B \implies \Pr[A] \leq \Pr[B]$
$\Pr[\emptyset] = 0$, $\Pr[\Omega] = 1$
$0 \leq \Pr[E] \leq 1$

Komplementäres Ereignis

$\bar{E}$ heißt komplementäres Ereignis zu E
$\Pr[\bar{E}] = 1 - \Pr[E]$

Disjunkte/Unvereinbare Ereignisse

Zwei Ereignisse A und B sind disjunkt/unvereinbar, falls $A \cap B = \emptyset$
Ereignisse sind paarweise disjunkt, falls für alle Paare $i \neq j$ gilt: $A_i \cap A_j = \emptyset$

relative Häufigkeit

relative Häufigkeit von $E := \dfrac{\text{absolute Häufigkeit von } E}{\text{Anzahl aller Beobachtungen}}$
$= \dfrac{\text{Anzahl Eintreten von } E}{\text{Anzahl aller Beobachtungen}}$

Sei X_i Bernoulli-verteilt mit Erfolgswahrscheinlichkeit $p \implies$
$Z = \frac{1}{n}(X_1 + \dots + Z_n)$ gibt die relative Häufigkeit an, mit der $X = 1$ bei n Wiederholungen des Versuches eintritt

relative Abweichung

$$|\frac{1}{n} \sum_i X_i - p|$$

absolute Abweichung

$$|\sum_i X_i - np|$$

Siebformel, Prinzip der Inklusion/Exklusion

Seien $A_1, \dots, A_n$ Ereignisse mit $n \geq 2 \implies$

$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \Pr[A_i] - \sum_{1 \leq i_1 \leq i_2 \leq n} \Pr[A_{i_1} \cap A_{i_2}] + - \dots$$
$$+ (-1)^{l-1} \sum_{1 \leq i_1 < \dots < i_l \leq n} \Pr[A_{i_1} \cap \dots \cap A_{i_l}] + - \dots$$
$$+ (-1)^{n-1} \cdot \Pr[A_1 \cap \dots \cap A_n]$$

Additionssatz

Seien $A_1, \dots, A_n$ paarweise disjunkte Ereignisse $\implies$
$\Pr\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \Pr[A_i]$
$\Pr[A \cup B] = \Pr[A] + \Pr[B]$

Boolesche Ungleichung

Seien $A_1, \dots, A_n$ Ereignisse $\implies$
$\Pr\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \Pr[A_i]$

Prinzip von Laplace

Sei E ein Ereignis und alle Elementarereignisse aus E gleich wahrscheinlich $\implies \Pr[E] = \dfrac{|E|}{|\Omega|}$

Bedingte Wahrscheinlichkeit

Seien A und B Ereignisse mit $\Pr[B] > 0 \implies$
Die bedingte Wahrscheinlichkeit $\Pr[A|B]$ von A gegeben B ist:
$\Pr[A|B] := \dfrac{\Pr[A \cap B]}{\Pr[B]}$

$\Pr[B|B] = 1$
$\Pr[A|\Omega] = \Pr[A]$
für festes B ist $\Pr[A|B]$ proportional zu $\Pr[A \cap B]$
$\Pr[\emptyset|B] = 0$
$\Pr[\bar{A}|B] = 1 - \Pr[A|B]$
$\Pr[A \cap B] = Pr[B|A] \cdot \Pr[A] = \Pr[A|B] \cdot \Pr[B]$

Multiplikationssatz

Seien $A_1, \dots, A_n$ Ereignisse und $\Pr[A_1 \cap \dots \cap A_n] > 0 \implies$
$\Pr[A_1 \cap \dots \cap A_n] = \Pr[A_1] \cdot \Pr[A_2|A_1] \cdot \Pr[A_3|A_1 \cap A_2] \cdot \dots \cdot \Pr[A_n|A_1 \cap \dots \cap A_{n-1}]$

Satz von der totalen Wahrscheinlichkeit

Seien $A_1, \dots, A_n$ paarweise disjunkte Ereignisse und $B \subseteq A_1 \cup \dots \cup A_n \implies \Pr[B] = \sum_{i=1}^{n} \Pr[B|A_i] \cdot \Pr[A_i]$

Satz von Bayes

Seien $A_1, \dots, A_n$ paarweise disjunkte Ereignisse mit $\Pr[A_j] > 0 \; \forall j$ und $B \subseteq A_1 \cup \dots \cup A_n$ ein Ereignis mit $\Pr[B] > 0 \implies$
Für beliebige $i = 1, \dots, n$ gilt:
$\Pr[A_i|B] = \dfrac{\Pr[A_i \cap B]}{\Pr[B]} = \dfrac{\Pr[B|A_i] \cdot \Pr[A_i]}{\sum_{j=1}^{n} \Pr[B|A_j] \cdot \Pr[A_j]}$

Unabhängigkeit

Sei $\Pr[A|B] \implies$ Die zwei Ereignisse A und B sind unabhängig, falls B keinen Einfluss auf die Wahrscheinlichkeit des Eintretens von A hat. Es gilt:
$\Pr[A|B] = \Pr[A]$
$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$

Seien $A_1, \dots, A_n$ paarweise verschiedene Ereignisse $\implies A_1, \dots, A_n$ sind unabhängig, falls für alle Teilmengen $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ mit $i_1 < i_2 < \dots < i_k$ gilt, dass
$\Pr[A_{i_1} \cap \dots \cap A_{i_k}] = \Pr[A_{i_1}] \cdot \ldots \cdot \Pr[A_{i_k}]$

Die paarweise verschiedenen Ereignisse $A_1, \dots, A_n$ sind genau dann unabhängig, wenn für alle $(s_1, \dots, s_n) \in \{0,1\}^n$ gilt, dass
$\Pr[A_1^{s_1} \cap \dots \cap A_n^{s_n}] = \Pr[A_1^{s_1}] \cdot \ldots \cdot \Pr[A_n^{s_n}]$,
wobei $A_i^0 = \bar{A}_i$ und $A_i^1 = A_i$

Seien A und B zwei unabhängige Ereignisse $\implies$ $\bar{A}$ und B, A und $\bar{B}$, $\bar{A}$ und $\bar{B}$ sind unabhängig.

Seien A, B und C unabhängige Ereignisse $\implies$ $A \cap B$ und C bzw. $A \cup B$ und C sind unabhängig.

Zufallsvariablen

Sei ein Wahrscheinlichkeitsraum auf der Ergebnismenge Ω gegeben. Eine Abbildung
$X : \Omega \to \mathbb{R}$
heißt (numerische) Zufallsvariable

Zugehörige Ereignis:
$A_i := \{\omega \in \Omega; X(\omega) = x_i\} = X^{-1}(x_i)$

Sei Ω endlich oder abzählbar unendlich $\implies$ die Zufallsvariable X ist diskret und ihr Wertebereich ebenfalls endlich oder abzählbar unendlich

Wertebereich

Sei X eine Zufallsvariable $\implies$
$W_X := X(\Omega) = \{x \in \mathbb{R}; \exists \omega \in \Omega \text{ mit } X(\omega) = x\}$

Dichtefunktion

$f_X : \mathbb{R} \ni x \mapsto \Pr[X = x] \in [0,1]$
$(\Pr[X^{-1}(x_i)] = \Pr[\text{“}X = x_i\text{”}] = \Pr[X = x_i])$

Verteilungsfunktion

$F_X : \mathbb{R} \ni x \mapsto \Pr[X \leq x] = \sum_{x' \in W_X : x' \leq x} \Pr[X = x'] \in [0,1]$
(Stammfunktion der Dichtefunktion)

Erwartungswert

$\mathbb{E}[X] := \sum_{x \in W_X} x \cdot \Pr[X = x] = \sum_{x \in W_X} x \cdot f_X(x) = \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega]$,
falls $\sum_{x \in W_X} |x| \cdot \Pr[X = x]$ konvergiert bzw.
$\sum_{\omega \in \Omega} |X(\Omega)| \cdot \Pr[\omega]$ existiert

Sei X Zufallsvariable mit $W_X \subseteq \mathbb{N}_0 \implies$
$\mathbb{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$

Sei X eine Zufallsvariable und $A_1, \dots, A_n$ paarweise disjunkte Ereignisse mit $A_1 \cup \dots \cup A_n = \Omega$ und $\Pr[A_1], \dots, \Pr[A_n] > 0 \implies$
$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i] \cdot \Pr[A_i]$,
falls alle Erwartungswerte auf der rechten Seite existieren und $\sum_{i=1}^{\infty} |\mathbb{E}[X|A_i]| \cdot \Pr[A_i]$ konvergiert

Monotonie des Erwartungswerts

Seien X, Y Zufallsvariablen über dem W'keitsraum $\Omega \implies$
Ist $X(\omega) \leq Y(\omega)\ \forall \omega \in \Omega \implies \mathbb{E}[X] \leq \mathbb{E}[Y]$
Ist $a \leq X(\omega) \leq b\ \forall \omega \in \Omega \implies a \leq \mathbb{E}[X] \leq b$

Rechenregeln für den Erwartungswert

Sei $Y := f(X) = f \circ X$ mit $f : \mathcal{D} \to \mathbb{R}$ eine beliebige Funktion mit $W_X \subseteq \mathcal{D} \subseteq \mathbb{R} \implies$
$\mathbb{E}[f(X)] = \mathbb{E}[Y] = \sum_{x \in W_X} f(x) \cdot \Pr[X = x] = \sum_{\omega \in \Omega} f(X(\omega)) \cdot \Pr[\omega]$

Linearität des Erwartungswerts

Sei X eine Zufallsvariable und $a, b \in \mathbb{R} \implies$
$\mathbb{E}[a \cdot X + b] = a \cdot \mathbb{E}[X] + b$

Varianz

Sei X eine Zufallsvariable und $\mu = \mathbb{E}[X] \implies$
$\mathrm{Var}[X] := \mathbb{E}[(X - \mu)^2] = \sum_{x \in W_X} (x - \mu)^2 \cdot \Pr[X = x]$
$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
$\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X]$

Standardabweichung

$\sigma := \sqrt{\mathrm{Var}[X]}$ heißt Standardabweichung von X

Momente

Sei X eine Zufallsvariable $\implies$
$\mathbb{E}[X^k]$ ist das k-te Moment
$\mathbb{E}[(X - \mathbb{E}[X])^k]$ ist das k-te zentrale Moment
Der Erwartungswert ist erstes Moment, die Varianz ist zweites zentrales Moment

Bedingte Zufallsvariable

Sei X eine Zufallsvariable und A ein Ereignis mit $\Pr[A] > 0 \implies$
Bedingte Zufallsvariable ist $X|A$ mit Dichte
$f_{X|A}(x) := \Pr[X = x|A] = \frac{\Pr[\text{”}X = x\text{”} \cap A]}{\Pr[A]}$
und Erwartungswert
$\mathbb{E}[X|A] = \sum_{x \in W_X} x \cdot f_{X|A}(x)$

Mehrere Zufallsvariablen

Gemeinsame Dichte

$f_{X,Y}(x, y) := \Pr[X = x, Y = y] = \Pr[\{\omega; X(\omega) = x, Y(\omega) = y\}]$
heißt gemeinsame Dichte der Zufallsvariablen X und Y

Randdichte

$f_X(x) = \Pr[X = x] = \sum_{y \in W_Y} f_{X,Y}(x, y)$ bzw.
$f_Y(y) = \Pr[Y = y] = \sum_{x \in W_X} f_{X,Y}(x, y)$

gemeinsame Verteilung

Seien X, Y Zufallsvariablen. Deren gemeinsame Verteilung ist:
$F_{X,Y}(x,y) = \Pr[X \leq x, Y \leq y] = \Pr[\{\omega; X(\omega) \leq x, Y(\omega) \leq y\}]$
$= \sum_{x' \leq x} \sum_{y' \leq y} f_{X,Y}(x', y')$

Randverteilung

$$F_X(x) = \sum_{x' \leq x} f_X(x') = \sum_{x' \leq x} \sum_{y \in W_Y} f_{X,Y}(x', y)$$
$$F_Y(y) = \sum_{y' \leq y} f_Y(y') = \sum_{y' \leq y} \sum_{x \in W_X} f_{X,Y}(x, y')$$

Unabhängigkeit

Seien $X_1, \ldots, X_n$ Zufallsvariablen.
$X_1, \ldots, X_n$ sind unabhängig, falls für alle
$(x_1, \ldots, x_n) \in W_{X_1} \times \cdots \times W_{X_n}$ gilt:
$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \Pr[X_1 = x_1] \cdot \ldots \cdot \Pr[X_n = x_n]$
bzw.
$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen und $S_1, \ldots, S_n$ beliebige Mengen mit $S_i \subseteq W_{X_i} \implies$
Die Ereignisse "$X_1 \in S_1$", …, "$X_n \in S_n$" sind unabhängig.

Seien $f_1, \ldots, f_n$ reellwertige Funktionen ($f_i : \mathbb{R} \to \mathbb{R}$):
Sind $X_1, \ldots, X_n$ unabhängig $\implies f_1(X_1), \ldots, f_n(X_n)$ sind unabhängig

Zusammengesetzte Zufallsvariablen

Seien X und Y unabhängige Zufallsvariablen und $Z := X + Y \implies$
$f_Z(z) = \sum_{x \in W_X} f_X(x) \cdot f_Y(z - x)$

Linearität des Erwartungswerts

Seien $X_1, \ldots, X_n$ Zufallsvariablen und $X := a_1 X_1 + \cdots + a_n X_n$ mit $a_1, \ldots, a_n \in \mathbb{R} \implies$
$\mathbb{E}[X] = a_1 \mathbb{E}[X_1] + \cdots + a_n \mathbb{E}[X_n]$

Multiplikativität des Erwartungswerts

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen $\implies$
$\mathbb{E}[X_1 \cdot \ldots \cdot X_n] = \mathbb{E}[X_1] \cdot \ldots \cdot \mathbb{E}[X_n]$

Indikatorvariable

Sei A ein Ereignis. Die Zufallsvariable
$I_A := \begin{cases} 1 & \text{falls } A \text{ eintritt} \\ 0 & \text{sonst} \end{cases}$
heißt Indikatorvariable des Ereignisses A

$\mathbb{E}[I_A] = \Pr[A]$
$\mathbb{E}[I_{A_1} \cdot \ldots \cdot I_{A_n}] = \Pr[A_1 \cap \cdots \cap A_n]$

Varianz

Seien $X_1, \ldots, X_n$ unabh. Zufallsvariablen und $X := X_1 + \cdots + X_n$
$\implies \text{Var}[X] = \text{Var}[X_1] + \cdots + \text{Var}[X_n]$

Wichtige diskrete Verteilungen

Bernoulli-Verteilung

Sei X eine Zufallsvariable mit $W_X = \{0, 1\}$
und Dichte $f_X(x) = \begin{cases} p & \text{für } x = 1 \\ 1 - p & \text{für } x = 0 \end{cases}$
$\implies$ X ist Bernoulli-verteilt
p ist Erfolgswahrscheinlichkeit

$\mathbb{E}[X] = p$, $\text{Var}[X] = pq$, $q := 1 - p$

Binomialverteilung

Sei $X := X_1 + \cdots + X_n$ eine Summe aus n unabhängigen, Bernoulli-verteilten Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p
$\implies$ X ist binomialverteilt mit den Parametern n und p:
$X \sim \text{Bin}(n, p)$

$W_X = \{0, \ldots, n\}$
$f_X(x) := b(x; n, p) = \binom{n}{x} p^x q^{n-x}$, $q := 1 - p$

$\mathbb{E}[X] = np$, $\text{Var}[X] = npq$

Seien $X \sim \text{Bin}(n_x, p)$ und $Y \sim \text{Bin}(n_y, p)$ unabh. Zufallsvariablen und $Z := X + Y \implies$
$Z \sim \text{Bin}(n_x + n_y, p)$

Geometrische Verteilung

Man betrachte ein Experiment, das so lange wiederholt wird, bis Erfolg eintritt. Gelingt ein einzelner Versuch mit Wahrscheinlichkeit p, so ist die Anzahl der Versuche bis zum Erfolg geometrisch verteilt.

Sei X eine geometrisch verteilte Zufallsvariable mit Parameter (Erfolgswahrscheinlichkeit) $p \in (0, 1]$ und $q := 1 - p \implies$
$f_X(i) = pq^{i-1}$ für $i \in \mathbb{N}$
$\mathbb{E}[X] = \frac{1}{p}$, $\quad \text{Var}[X] = \frac{q}{p^2}$

Negative Binomialverteilung

Seien $X_1, \ldots, X_n$ unabhänige, geometrisch verteilte Zufallsvariablen mit Parameter p und $Z := X_1 + \cdots + X_n$ (Z $\equiv$ Anzahl der Versuche bis zum n-ten erfolgreichen Experiment) $\implies$
Z ist negativ binomialverteil mit Ordnung n und
$f_Z(z) = \binom{z-1}{n-1} \cdot p^n (1 - p)^{z-n}$

Coupon-Collector-Problem

Poisson-Verteilung

Wird verwendet um die Anzahl von Ereignissen zu modellieren, welche mit konstanter Rate und unabhängig voneinander in einem Zeitintervall auftreten.
$X \sim \text{Po}(\lambda)$

Sei X eine Poisson-verteilte Zufallsvariable mit $\lambda \geq 0 \implies$
$W_X = \mathbb{N}_0$

$f_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}$ für $i \in \mathbb{N}_0$

$\mathbb{E}[X] = \lambda$, $\text{Var}[X] = \lambda$

Poisson-Verteilung als Grenzwert der Binomialverteilung

$\lim_{n\to\infty} b(k; n, \lambda/n) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$
$X \sim \text{Bin}(n, \lambda/n) \xrightarrow{n\to\infty} X \sim \text{Po}(\lambda)$

Faustregel: $n \geq 30$ und $p \leq 0.05$

Gesetz seltener Ereignisse

Ist $n >> \lambda \implies X \sim \text{Bin}(n, \lambda/n) \approx X \sim \text{Po}(\lambda)$,
falls folgende Voraussetzungen erfüllt sind:

- Die Ereignisse treten nie zur gleichen Zeit auf
- Die Wahrscheinlichkeit, dass ein Ereignis in einem (kleinen) Zeitintervall auftritt, ist proportional zur Länge des Intervalls
- Die Anzahl der Ereignisse in einem festen Zeitintervall hängt nur von dessen Länge ab, nicht aber von der Lage auf der Zeitachse
- Wenn man zwei disjunkte Zeitintervalle betrachtet, so sind die Anzahlen der Ereignisse in diesen Zeiträumen voneinander unabhängig

Summe von Poisson-verteilten Zufallsvariablen

Seien X, Y unabhängige Zufallsvariablen mit $X \sim \text{Po}(\lambda)$ und $Y \sim \text{Po}(\mu) \implies$
$Z := X + Y \sim \text{Po}(\lambda + \mu)$

Abschätzen von Wahrscheinlichkeiten

Markov-Ungleichung

Sei X eine Zufallsvariable, die nur nicht-negative Werte annimmt und $t \in \mathbb{R}^+$
$\implies \Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t} \iff \Pr[X \geq t \cdot \mathbb{E}[X]] \leq \frac{1}{t}$

Chebyshev-Ungleichung

Sei X eine Zufallsvariable und $t \in \mathbb{R}^+ \implies$
$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\text{Var}[X]}{t^2}$
$\iff \Pr[|X - \mathbb{E}[X]| \geq t\sqrt{\text{Var}[X]}] \leq \frac{1}{t^2}$

Anw: $\Pr[X \geq k] = \Pr[X - \mathbb{E}[X] \geq k - \mathbb{E}[X]] \leq \Pr[|X - \mathbb{E}[X] \geq k - \mathbb{E}[X]] \leq \frac{\text{Var}[X]}{(k - \mathbb{E}[X])^2}$

Gesetz der großen Zahlen

Sei X eine Zufallsvariable und $\varepsilon, \delta > 0$ beliebig aber fest
Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit derselben Verteilung wie X und $Z := \frac{X_1 + \cdots + X_n}{n}$
Für alle $n \geq \frac{\text{Var}[X]}{\varepsilon\delta^2}$ gilt:
$\Pr[|Z - \mathbb{E}[X]| \geq \delta] \leq \varepsilon$

Chernoff-Schranken

Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen (nicht zwingend gleichverteilt) mit Erfolgswahrscheinlichkeit p_i und $X := \sum_{i=1}^n X_i$ und $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$ und jedes $\delta > 0 \implies$

$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$ für alle $\delta > 0$

$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu$ für alle $0 < \delta < 1$

$\Pr[X \geq (1+\delta)\mu] \leq e^{\frac{-\mu\delta^2}{3}}$ für alle $0 < \delta \leq 1$

$\Pr[X \leq (1-\delta)\mu] \leq e^{\frac{-\mu\delta^2}{2}}$ für alle $0 < \delta \leq 1$

$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{\frac{-\mu\delta^2}{3}}$ für alle $0 < \delta \leq 1$

$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}$

$\Pr[X \geq t] \leq 2^{-t}$ für $t \geq 2e\mu$

Lemma

für $0 \leq \delta < 1$ gilt
$(1-\delta)^{1-\delta} \geq e^{-\delta+\delta^2/2}$ und $(1+\delta)^{1+\delta} \geq e^{\delta+\delta^2/3}$

Erzeugende Funktionen

Eine wahrscheinlichkeitserzeugende Funktion ist die (gewöhnliche) erzeugende Funktion der Folge $(f_i)_{i\in\mathbb{N}_0}$ mit $f_i := \Pr[X = i]$

Für eine Zufallsvariable X mit $W_X \subseteq \mathbb{N}_0$ ist die (wahrscheinlichkeits-)erzeugende Funktion definiert durch
$G_X(s) := \sum_{k=0}^{\infty} \Pr[X = k] \cdot s^k = \mathbb{E}[x^X]$

Sei $Y := X + t$ mit $t \in \mathbb{N}_0 \implies$
$G_Y(s) = s^t \cdot G_X(s)$

$G_X^{(i)}(0) = \Pr[X = i] \cdot i!$
$\mathbb{E}[X] = G'_X(1)$
$\text{Var}[X] = G''_X(1) + G'_X(1) - (G'_X(1))^2$

Eindeutigkeit der wahrscheinlichkeitserzeugenden Funktion

Die Dichte und die Verteilung einer Zufallsvariable X mit $W_X \subseteq \mathbb{N}$ sind durch ihre wahrscheinlichkeitserzeugende Funktion eindeutig bestimmt

Bernoulli-Verteilung

Sei X eine Bernoulli-verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit $p \implies$
$G_X(s) = 1 - p + ps$

Gleichverteilung auf {0, ..., n}

Sei X auf $\{0, \ldots, n\}$ gleichverteilt, d.h. für $0 \leq k \leq n$ ist $\Pr[X = k] = \frac{1}{n+1} \implies$
$G_X(s) = \sum_{k=0}^{n} \frac{1}{n+1} \cdot s^k = \frac{s^{n+1} - 1}{(n+1)(s-1)}$

Binomialverteilung

Sei $X \sim \text{Bin}(n, p) \implies$
$G_X(s) = (1 - p + ps)^n$

$G'_X(s) = n \cdot (1 - p + ps)^{n-1} \cdot p$

Geometrische Verteilung

Sei X eine geometrisch verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit $p \implies$
$G_X(s) = \dfrac{ps}{1 - (1 - p)s}$

Poisson-Verteilung

Sei $X \sim \text{Po}(\lambda) \implies$
$G_X(s) = e^{\lambda(s-1)}$

Momenterzeugende Funktionen

Sei X eine Zufallsvariable. Die zugehörige momenterzeugende Funktion ist
$M_X(s) := \mathbb{E}[e^{Xs}] = \sum_{i=0}^{\infty} \dfrac{\mathbb{E}[X^i]}{i!} \cdot s^i$

Sei $W_X \subseteq \mathbb{N}_0 \implies$
$M_X(s) = G_X(e^s)$

Erzeugende Funktion einer Summe

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen und $Z := X_1 + \cdots + X_n \implies$
$G_Z(s) = G_{X_1}(s) \cdot \ldots \cdot G_{X_n}(s)$ und
$M_Z(s) = M_{X_1}(s) \cdot \ldots \cdot M_{X_n}(s)$

Zufällige Summen

Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit der wahrscheinlichkeitserzeugenden Funktion $G_X(s)$ und N eine unabhängige Zufallsvariable mit der wahrscheinlichkeitserzeugenden Funktion $G_N(s)$ und $Z := X_1 + \cdots + X_N \implies$
$G_Z(s) = G_N(G_X(s))$

Kontinuierliche Wahrscheinlichkeitsräume

Grundlagen

Ereignis

Eine Menge $A \subseteq \mathbb{R}$, die durch Vereinigung $A = \bigcup_k I_k$ abzählbar vieler paarweise disjunkter Intervalle beliebiger Art (offen, geschlossen, halboffen, einseitig unendlich) gebildet werden kann, heißt Ereignis. Ein Ereignis A tritt ein, wenn X einen Wert aus A annimmt.
$\Pr[A] = \int_A f_X(x)dx = \sum_k \int_{I_k} f_X(x)dx$

Kolmogorov-Axiome und σ-Algebren

σ-Algebren

Sei Ω eine Menge. Eine Menge $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt σ-Algebra über Ω, falls gilt

- $\Omega \in \mathcal{A}$
- Wenn $A \in \mathcal{A}$, dann folgt $\bar{A} \in \mathcal{A}$
- Für $n \in \mathbb{N}$ sei $A_n \in \mathcal{A}$. Dann gilt auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

Für jede (endliche) Menge Ω stellt die Menge $\mathcal{P}(\Omega)$ eine σ-Algebra dar.

Für $\Omega = \mathbb{R}$ ist die Klasse der Borel'schen Mengen, die aus allen Mengen $A \subseteq \mathbb{R}$ besteht, welche sich durch abzählbare Vereinigungen und Schnitte von Intervallen (offen, halboffen oder geschlossen) darstellen lassen, eine σ-Algebra

Kolmogorov-Axiome

Sei Ω eine beliebige Menge und $\mathcal{A}$ eine σ-Algebra über Ω. Eine Abbildung $\Pr[.] : \mathcal{A} \to [0, 1]$ heißt Wahrscheinlichkeitsmaß auf $\mathcal{A}$, falls

- $\Pr[\Omega] = 1$
- $A_1, A_2, \ldots$ seien paarweise disjunkte Ereignisse $\implies$
  $\Pr\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \Pr[A_i]$

Für ein Ereignis $A \in \mathcal{A}$ heißt $\Pr[A]$ Wahrscheinlichkeit von A. Ein Wahrscheinlichkeitsraum ist definiert durch das Tupel $(\Omega, \mathcal{A}, \Pr)$

Sei $(\Omega, \mathcal{A}, \Pr)$ ein Wahrscheinlichkeitsraum und $A, B, A_1, A_2, \ldots$ Ereignisse $\implies$

- $\Pr[\emptyset] = 0, \Pr[\Omega] = 1$
- $0 \leq \Pr[A] \leq 1$
- $\Pr[\bar{A}] = 1 - \Pr[A]$
- Wenn $A \subseteq B$, so folgt $\Pr[A] \leq \Pr[B]$

Lebesgue-Integrale

messbare Funktionen

Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt messbar, falls das Urbild jeder Borel'schen Menge ebenfalls eine Borel'sche Menge ist.

Für jede Borel'sche Menge A ist die zugehörige Indikatorfunktion messbar.
Jede stetige Funktion ist messbar.
Summen und Produkte von messbaren Funktionen sind wiederum

messbar.

Jeder messbaren Funktion kann man ein Integral, das so genannte Lebesgue-Integral $\int f d\lambda$ zuordnen

Sei $f : \mathbb{R} \to \mathbb{R}_0^+$ eine messbare Funktion $\implies$
$\Pr : A \to \int f \cdot I_A d\lambda$ ist eine Abbildung auf den Borel'schen Mengen, die die zweite Eigenschaft der Kolmogorov-Axiome erfüllt
Gilt zusätzlich $\Pr[\mathbb{R}] = 1 \implies$
f definiert auf natürliche Weise einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \Pr)$ mit $\Omega = \mathbb{R}$ und $\mathcal{A}$ ist die Menge der Borel'schen Mengen.

Kontinuierliche Zufallsvariablen

Dichtefunktion

Sei X eine kontinuierliche oder auch stetige Zufallsvariable $\implies$
Dichtefunktion: $f_X : \mathbb{R} \to \mathbb{R}_0^+$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Verteilungsfunktion

$$F_X(x) := \Pr[X \leq x] = \Pr[\{t \in \mathbb{R} | t \leq x\}] = \int_{-\infty}^{x} f_X(t)dt$$

F_X ist monoton steigend
F_X ist stetig
$\lim_{x \to -\infty} F_X(x) = 0$ und $\lim_{x \to \infty} F_X(x) = 1$

Für jede differenzierbare Funktion F, welche die zuvor genannten Eigenschaften erfüllt $\implies$
$f(x) = F'(x)$ und $\Pr[a < X \leq b] = F_X(b) - F_X(a)$

Für $a < X \leq b$, $a \leq X \leq b$, $a \leq X < b$, $a < X < b$ gilt:

$$\int_{[a,b]} f(t)dt = \int_{]a,b]} f(t)dt = \int_{[a,b[} f(t)dt = \int_{]a,b[} f(t)dt$$

Kontinuierliche Zufallsvariablen als Grenzwerte diskreter Zufallsvariablen

Sei X eine kontinuierliche Zufallsvariable
Defniert man für ein festes $\delta > 0$
$X_\delta = n\delta \iff X \in [n\delta, (n+1)\delta[$ für $n \in \mathbb{Z} \implies$
$\Pr[X_\delta = n\delta] = F_X((n+1)\delta) - F_X(n\delta)$

Für $\delta \to 0$ nähert sich die Verteilung von X_δ der Verteilung von X imm mehr an

Erwartungswert

Sei X eine Zufallvariable $\implies$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t \cdot f_X(t)dt,$$

sofern das Integral $\int_{-\infty}^{\infty} |t| \cdot f_X(t)dt$ existiert

Sei X, Y Zufallsvariablen mit $Y := g(X) \implies$

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} g(t) \cdot f_X(t)dt$$

Varianz

Sei X eine Zufallsvariable $\implies$

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (t - \mathbb{E}[X])^2 \cdot f_X(t)dt,$$

sofern $\mathbb{E}[(X - \mathbb{E}[X])^2]$ existiert

Wichtige stetige Verteilungen

Gleichverteilung

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in [a,b] \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 1 & \text{für } x > b \end{cases}$$

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

Normalverteilung

Sei X eine Zufallsvariable mit $W_X = \mathbb{R}$ und Parametern $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}^+$ X ist normalverteilt, falls

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) = \varphi(x; \mu, \sigma)$$

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \int_{-\infty}^{x} \exp\Big(-\frac{(t-\mu)^2}{2\sigma^2}\Big)dt =: \Phi(x; \mu, \sigma)$$

$\mathbb{E}[X] = \mu$, $\text{Var}[X] = \sigma^2$

$X \sim \mathcal{N}(\mu, \sigma^2)$

Lineare Tranformation der Normalverteilung

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ und $Y = aX + b$ mit beliebigen $a \in \mathbb{R}\backslash 0$ und $b \in \mathbb{R} \implies$
$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

Sei $Y := \frac{X - \mu}{\sigma} \implies Y \sim \mathcal{N}(0,1)$. Y heißt auch normiert.

Additivität der Normalverteilung

Seien $X_1, \ldots, X_n$ unabhängige und normalverteilte Zufallsvariablen mit Parameter $\mu_i, \sigma_i (1 \leq i \leq n)$ und $Z := a_1X_1 + \cdots + a_nX_n \implies$
Z ist normalverteilt mit Erwartungswert $\mu = a_1\mu_1 + \cdots + a_n\mu_n$ und Varianz $\sigma^2 = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$

Standardnormalverteilung

X ist standardnormalverteilt, falls $X \sim \mathcal{N}(0,1)$
Die zugehörige Dichte $\varphi(x; 0, 1)$ kürzen wir durch $\varphi(x)$ ab
$\mathbb{E}[X] = 0$, $\text{Var}[X] = 1$

Exponentialverteilung

Die Exponentialverteilung ist das kontinuierliche Analogon zur geometrischen Verteilung. Sie ist gedächtnislos und spielt vor allem bei der Modellierung von Wartezeiten eine große Rolle.

Sei X eine Zufallsvariable mit Parameter $\lambda, \lambda > 0$
X ist exponentialverteilt, falls

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

$\mathbb{E}[X] = \frac{1}{\lambda}$, Var$[X] = \frac{1}{\lambda^2}$

Eigenschaften der Exponentialverteilung

Sei X exponentialverteilt mit Parameter λ und $Y := aX$ mit $a > 0$
$\implies$ Y ist exponentialverteilt mit Parameter λ/a

Gedächtnislosigkeit

Eine (positive) kontinuierliche Zufallsvariable X mit $W_X = \mathbb{R}^+$ ist genau dann exponentialverteilt, wenn für alle $x, y > 0$ gilt, dass
$\Pr[X > x + y | X > y] = \Pr[X > x]$

Anwendung bei Zerfallsraten

$\mathbb{E}[X] \equiv$ erwartete Zeit bis Zerfall
$\lambda = \frac{1}{\mathbb{E}[X]}$

Exponentialverteilung als Grenzwert der geometrischen Verteilung

Sei X_n eine Folge geometrisch verteilter Zufallsvariablen mit Parameter $p_n = \lambda/n$
Sei $Y_n := \frac{1}{n} X_n \implies$ Die Folge Y_n geht für $n \to \infty$ in eine exponentialverteilte Zufallsvariable mit Parameter λ über.

Mehrere kontinuierliche Zufallsvariablen

Mehrdimensionale Dichten

Seien X, Y kontinuierliche Zufallsvariablen $\implies$
Die Dichte ist beschrieben durch: $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_0^+$ mit

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$$

Für ein Ereignis $A \subseteq \mathbb{R}^2$ gilt:
$\Pr[A] = \int_A f_{X,Y}(x,y) dx dy$

Mehrdimensionale Verteilung

Sei $f_{X,Y}$ die gemeinsame Dichte der Zufallsvariablen $X, Y \implies$
Die gemeinsame Verteilung ist definiert als: $F_{X,Y} : \mathbb{R}^2 \to [0,1]$
$F_{X,Y}(x,y) = \Pr[X \leq x, Y \leq y] = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v) du dv$

Randdichte

Sei $f_{X,Y}$ die gemeinsame Dichte der Zufallsvariablen X und Y. Die Randdichte von X ist:
$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,v) dv$

Randverteilung

Sei $f_{X,Y}$ die gemeinsame Dichte der Zufallsvariablen X und Y. Die Randverteilung von X ist:
$F_X(x) = \Pr[X \leq x] = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f_{X,Y}(u,v) dv \right] du$

Unabhängigkeit

Seien $X_1, \ldots, X_n$ Zufallsvariablen $\implies$
$X_1, \ldots, X_n$ sind genau dann unabhängig, wenn
$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$ bzw.
$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$ bzw.

Warteprobleme mit der Exponentialverteilung - Warten auf mehrere Ereignisse

Seien $X_1, \ldots, X_n$ unabhängige und exponentialverteilte Zufallsvariablen mit den Parametern $\lambda_1, \ldots, \lambda_n \implies$
$X := min\{X_1, \ldots, X_n\}$ ist exponentialverteilt mit dem Parameter $\lambda_1 + \cdots + \lambda_n$

Erläuterung:
Wartet man auf das Eintreffen eines Ereignisses aus mehreren unabhängigen Ereignissen, so addieren sich die Raten.

Poisson-Prozess

Seien $T_1, T_2, \ldots$ unabhängige exponentialverteilte Zufallsvariablen mit Parameter λ und sei $X(t) := max\{n \in \mathbb{N} | T_1 + \cdots + T_n \leq t\}$ für $t > 0 \implies X(t)$ ist Poisson-verteilt mit Parameter $t\lambda$

Erläuterung: Wenn man Ereignisse zählt, deren zeitlicher Abstand exonentialverteilt ist, so ist die Anzahl dieser Ereignisse in einer festen Zeitspanne Poisson-verteilt.

Summen von Zufallsvariablen

Seien X und Y unahängige kontinuierliche Zufallsvariablen und $Z := X + Y \implies$
$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) dx$

Momenterzeugende Funktionen für kontinuierliche Zufallsvariablen

Sei X eine Zufallsvariable. Die zugehörige momenterzeugede Funktion ist
$M_X(s) = \mathbb{E}[e^{Xs}]$

Gleichverteilung

Sei U eine gleichverteilte Zufallsvariable auf $[a,b] \implies$
$M_U(t) = \mathbb{E}[e^{tX}] = \frac{e^{tb} - e^{ta}}{t(b-a)}$

Standardnormalverteilung

Sei $N \sim \mathcal{N}(0,1) \implies$
$M_N(t) = e^{\frac{t^2}{2}}$

Normalverteilung

Sei $Y \sim \mathcal{N}(\mu, \sigma^2) \implies$
$M_Y(t) = e^{t\mu + (t\sigma)^2/2}$

Zentraler Grenzwertsatz

Seien $X_1, \dots, X_n$ unabhängige und gleich verteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 ($\sigma^2 > 0$) für X_i. Und sei $Y_n := X_1 + \dots + X_n$ für $n \geq 1 \implies$
$Z_n := \dfrac{Y_n - n\mu}{\sigma\sqrt{n}}$ ist asymptotisch standardnormalverteilt,
also $Z_n \sim \mathcal{N}(0,1)$ für $n \to \infty$
Die Verteilung von Z_n konvergiert gegen die Standardnormalverteilung für $n \to \infty$

Aussage:
Wenn eine Zufallsgröße durch lineare Kombination vieler unabhängiger identisch verteilter Zufallsgrößen entsteht, so erhält man näherungsweise eine Normalverteilung.

Grenzwertsatz von de Moivre

Seien $X_1, \dots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p und $H_n := X_1 + \dots + X_n$ für $n \geq 1 \implies$
$H_n^* := \dfrac{H_n - np}{\sqrt{np(1-p)}}$ konvergiert für $n \to \infty$ gegen die Standardnormalverteilung

Normalverteilung als Grenzwert der Binomialverteilung

Sei $H_n \sim \text{Bin}(n,p)$ eine binomialverteilte Zufallsvariable. $\implies$
$\dfrac{H_n}{n}$ konvergiert für $n \to \infty$ gegen $\mathcal{N}\Big(p, \dfrac{p(1-p)}{n}\Big)$

Verschiedene Approximationen der Binomialverteilung

Sei $H_n \sim$ Bin(n,p) mit Verteilungsfunktion F_n. Für $n \to \infty$ gilt:
$F_n(t) = \Pr[\frac{H_n}{n} \leq \frac{t}{n}] \to \Phi\left(\dfrac{\frac{t}{n} - p}{\sqrt{\frac{p(1-p)}{n}}}\right) = \Phi\left(\dfrac{t - np}{\sqrt{p(1-p)n}}\right)$

Faustregel: $np \geq 5$ und $n(1-p) \geq 5$

Stetigkeitskorrektur

Verwende $\Pr[X \leq x] \approx \Phi\left(\dfrac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$
anstatt $\Pr[X \leq x] \approx \Phi\left(\dfrac{x - np}{\sqrt{np(1-p)}}\right)$

Induktive Statistik

Ziel:
Aus gemessenen Zufallsgrößen auf die zugrunde liegenden Gesetzmäßigkeiten schließen.

Schätzvariablen

Man führe n Messungen durch. Jede Messung wird durch eine Zufallsvariable X_i dargestellt. $\implies$
Die n Messungen heißen Stichproben, und die Variablen X_i nennt man Stichprobenvariablen.

Grundprinzip statistischer Verfahren

Wenn man ein Experiment genügen oft wiederholt, so nähert sich der Durchschnitt der Versuchsergebnisse immer mehr dem Verhalten an, das man im Mittel erwarten würde. Auf diesem Grundprinzip beruhen alle statistischen Verfahren.

arithmetische Mittel

Seien X_i Zufallsvariablen. Das arithmetische Mittel $\bar{X}$ ist:
$\bar{X} := \dfrac{1}{n} \sum_{i=1}^{n} X_i$

Ist $\bar{X}$ ein erwartungstreuer Schätzer, dann gilt:
$\mathbb{E}[\bar{X}] = \dfrac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \dfrac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X] = \mathbb{E}[X]$

erwartungstreue Schätzvariable

Sei X eine Zufallsvariable mit Dichte $f(x;\theta)$
Eine Schätzvariable/Schätzer für den Parameter θ der Dichte X ist eine Zufallsvariable, die aus mehreren (meist unabhänigen und identisch verteilten) Stichprobenvariablen zusammengesetzt ist.
Ein Schätzer U heißt erwartungstreu, wenn gilt
$\mathbb{E}[U] = \theta$

Bias der Schätzvariablen

$\mathbb{E}[U - \theta]$ nennt man Bias der Schätzvariablen U.
Bei erwartungstreuen Schätzvariablen ist der Bias gleich null.

Mean squared error (MSE) - mittlere quadratische Abweichung

Sei U eine Schätzvariable $\implies$
$MSE := \mathbb{E}[(U - \theta)^2]$

Sei U erwartungstreu $\implies$
$MSE = \mathbb{E}[(U - \mathbb{E}[U])^2] = \text{Var}[U]$

effizientere Schätzvariablen

Seien A, B Schätzvariablen und MSE von A kleiner, als die von B $\implies$
A ist effizienter als B

Konsistenz im quadratischen Mittel

Eine Schätzvariable heißt konsistent im quadratischen Mittel, wenn $MSE \to 0$ für $n \to \infty$ (n sei der Umfang der Stichprobe)

Bei jeder Verteilung mit endlicher Varianz folgt: Der Schätzer $\bar{X}$ ist konsistent.

schwache Konsistent

Sei $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ und $\varepsilon > 0$ beliebig, aber fest $\implies$
$\Pr[|\bar{X} - \theta| \geq \varepsilon] = \Pr[|\bar{X} - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\mathsf{Var}[\bar{X}]}{\varepsilon^2} \to 0$ für $n \to \infty$
Also: Für genügend große n liegen als die Werte von $\bar{X}$ beliebig nahe am gesuchten Wert $\theta = \mathbb{E}[X]$. Diese Eigenschaft nennt man auch schwache Konsistenz, da sie aus der Konsistenz im quadratischen Mittel folgt.

Stichprobenmittel

Seien $X_1, \dots, X_n$ Stichproben und $\bar{X}$ ein erwartungstreuer Schätzer für den Erwartungswert $\implies$
$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$
heißt Stichprobenmittel der Stichprobe $X_1, \dots, X_n$

Stichprobenvarianz

Seien $X_1, \dots, X_n$ Stichproben und $\bar{X}$ und S^2 erwartungstreure Schätzer für den Erwartungswert bzw. die Varianz $\implies$
$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
heißt Stichprobenvarianz der Stichprobe $X_1, \dots, X_n$

Maximum-Likelihood-Prinzip (MLE) zur Konstruktion von Schätzvariablen

Likelihood-Funktion

Sei $\vec{X} = (X_1, \dots, X_n)$, wobei $X_1, \dots, X_n$ unabhängige Kopien der Zufallsvariablen X mit der Dichte $f(x;\theta) = \Pr_\theta[X = x]$ ist. Hierbei sei θ der gesuchte Parameter der Verteilung. Außerdem sei $\vec{x} = (x_1, \dots, x_n)$, wobei eine Stichprobe für jede Variable X_i den Wert x_i liefert.

$$L(\vec{x};\theta) := \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \Pr_\theta[X_i = x_i]$$
$$\overset{unabh.}{=} \Pr_\theta[X_1 = x_1, \dots, X_n = x_n]$$
$$\iff \ln(L(\vec{x};\theta)) = \sum_{i=1}^{n} \ln(f(x_i;\theta))$$
entspricht der Wahrscheinlichkeit, dass wir die Stichprobe $\vec{x}$ erhalten, wenn wir den Parameter mit dem Wert θ belegen.
L ist die Likelihood-Funktion der Stichprobe $\vec{x}$

Maximum-Likelihood-Schätzwert

Ein Schätzwert $\hat{\theta}$ für den Parameter einer Verteilung $f(x;\theta)$ heißt Maximum-Likelihood-Schätzwert (ML-Schätzwert) für eine Stichprobe $\vec{x}$, wenn gilt:
$L(\vec{x};\theta) \leq L(\vec{x};\hat{\theta})\ \forall\theta$

Maximum-Likelihood-Funktion

Die Maximum-Likelihood-Funktion der Stichprobe $\vec{x}$ ist:
$\hat{\theta}_{MLE} = \arg\max_\theta L(\vec{x};\theta)$

Konfidenzintervalle

Konfidenzniveau

Sei θ der gesuchte Parameter und U_1, U_2 zwei Schätzvariablen und es gilt:
$\Pr[U_1 \leq \theta \leq U_2] \geq 1 - \alpha$
$\implies$ Die Wahrscheinlichkeit $1 - \alpha$ heißt Konfidenzniveau.

Wenn wir für eine konkrete Stichprobe die Schätzer U_1 und U_2 berechnen und davon ausgehen, dass $\theta \in [U_1, U_2]$ ist, so ziehen wir höchstens mit Wahrscheinlichkeit α einen falschen Schluss.

Konfidenzintervall

Sei θ der gesuchte Parameter und U_1, U_2 zwei Schätzvariablen und es gilt:
$\theta \in [U_1, U_2]$
$\implies$ $[U_1, U_2]$ heißt Konfidenzintervall.

In vielen Fällen verwendet man nur eine Schätzvariable U und konstruiert ein symmetrisches Konfidenzintervall: $[U - \delta, U + \delta]$

γ-Quantil

Sei X eine stetige Zufallsvariable mit Verteilung F_X. Eine Zahl x_γ mit $F_X(x_\gamma) = \gamma$
heißt γ-Quantil von X bzw. der Verteilung F_X

Für die Standardnormalverteilung bezeichnet z_γ das γ-Quantil

Testen von Hypothesen

Definition eines Tests

Seien $X_1, \dots, X_n$ n unabhängige Stichprobenvariablen mit derselben Verteilung wie X und $\vec{x}$ der zugehörige Stichprobenvektor. $\implies$
Für $\vec{x}$ muss nun die Frage beantwortet werden, ob wir für diesen Versuchsausgang die Hypothese annehmen oder ablehnen.

Ablehnungsbereich - kritischer Bereich

Sei $K := \{\vec{x} \in \mathbb{R}^n; \vec{x}$ führt zur Ablehnung der Hypothese$\}$ $\implies$
K ist der Ablehnungsbereich bzw. kritischer Bereich des Tests.

Konstruktion des Ablehnungsbereiches

Gewöhnlich wird der Ablehnungsbereich K konstruiert, indem man die Zufallsvariablen $X_1, \dots, X_n$ zu einer neuen Variablen T, der so genannten Testgröße, zusammenfasst. Dann unterteilt man den Wertebereich $\mathbb{R}$ von T in mehrere Bereiche, die entweder zur Ablehnung der Hypothese führen sollen oder nicht. Dabei betrachtet man meist ein einzelnes halboffenes oder abgeschlossenes Intervall und spricht dann von einem einseitigen bzw. von einem zweiseitigen Test.

Die Menge $\tilde{K} \subseteq \mathbb{R}$ enthalte die Werte von T, die zur Ablehnung der Hypothese führen sollen.

$\tilde{K} \subseteq \mathbb{R}$ entspricht direkt dem Ablehnungsbereich $K = T^{-1}(\tilde{K} \subseteq \mathbb{R}^n)$

Testgröße

Die Testgröße erhält man, indem man die Zufallsvariablen $X_1, \ldots, X_n$ zu einer neuen Variablen T zusammenfasst.
T ist dann die Testgröße.

Hypothese und Alternative

Nullhypothese (H_0):	die zu überprüfende Hypothese
Alternative (H_1):	Eine zweite Hypothese
triviale Alternative:	H_0 gilt nicht

Fehler bei statistischen Tests

Bei jedem Test können mit einer gewissen Wahrscheinlichkeit falsche Schlüsse gezogen werden.

Fehler 1. Art:	H_0 gilt, wird aber abgelehnt (das Ergebnis $\vec{x}$ der Stichprobe liegt in K)
Fehler 2. Art:	H_0 gilt nicht, wird aber angenommen (das Ergebnis $\vec{x}$ der Stichprobe liegt nicht in K)

Signifikanzniveau

Die Wahrscheinlichkeit für den Fehler 1. Art wird mit α bezeichnet. Der Fehler 1. Art wird auch α-Fehler genannt.
α ist das Signifikanzniveau des Tests

Güterfunktion

Die Güterfunktion $g(\cdot, p)$ gibt allgemein die Wahrscheinlichkeit an, mit der ein Test die Nullhypothese verwirft.

Allgemeines Vorgehen bei statistischen Tests

1. Formulierung von Annahmen (bzgl. Verteilung der Stichprobenvariablen und deren Unabhängigkeit)
2. Formulierung der Nullhypothese
3. Auswahl des Testverfahrens
4. Durchführung des Tests und Entscheidung

Berechnung der Fehlerwahrscheinlichkeiten

Fehlerw'keit 1. Art $= \sup_{p \in H_0} \Pr_p[T \in K] = \sup_{p \in H_0} \Pr_p[T \le k]$
Fehlerw'keit 2. Art $= \sup_{p \in H_1} \Pr_p[T \notin K] = \sup_{p \in H_1} \Pr_p[T > k]$

Das richtige Testverfahren finden

Anzahl der beteiligten Zufallsgrößen

Ein-Stichproben-Test:	eine Zufallsgröße
Zwei-Stichproben-Test:	zwei Zufallsgrößen mit potentiell unterschiedlichen Verteilungen, für die jeweils eine Stichprobe erzeugt wird

Formulierung der Nullhypothese

Welche Größe dient zur Definition der Nullhypothese? Hierbei werden in erster Linie Tests unterschieden, die Aussagen über verschiedene so genannte Lageparameter treffen, wie z.B. den Erwartungswert oder die Varianz der zugrunde liegenden Verteilungen.
Gelegentlich wird zur Formulierung der Nullhypothese auch der so genannte Median betrachtet. Der Median einer Verteilung entspricht dem (kleinsten) Wert x mit $F(x) = \frac{1}{2}$

Annahmen über die Zufallsgrößen

Was ist über die Verteilung der untersuchten Größe(n) bekannt? Bei entsprechenden Annahmen könnte es sich z.B. um die Art der Verteilung, den Erwartungswert oder die Varianz handeln.

Ein-Stichproben-Tests für Lageparameter

- Gaußtest
- t-Test
- Wilcoxon-Test
- χ^2-Varianztest

Zwei-Stichproben-Tests für Lageparameter

- Zwei-Stichproben-t-Test
- Zwei-Stichproben-Wilcoxon-Test

Nicht an Lageparametern orientierte Tests

- χ^2-Anpassungstest

Tests

Approximativer Binomialtest

Annahmen:
$X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $\Pr[X_i = 1] = p$ und $\Pr[X_i = 0] = 1 - p$, wobei p unbekannt sei. n sei hinreichend groß, so dass die Approximation nach Grenzwertsatz von de Moivre brauchbare Ergebnisse liefert.

Hypothesen	Ablehnungskriterium für H_0
$H_0: p = p_0$ gegen $H_1: p \neq p_0$	$\lvert Z\rvert > z_{1-\alpha/2}$
$H_0: p \ge p_0$ gegen $H_1: p < p_0$	$Z < z_\alpha$
$H_0: p \le p_0$ gegen $H_1: p > p_0$	$Z > z_{1-\alpha}$

Testgröße:
$Z := \frac{h - np_0}{\sqrt{np_0(1-p_0)}}$
wobei $h := X_1 + \cdots + X_n$ die Häufigkeit bezeichnet, mit der die Ereignisse $X_i = 1$ aufgetreten sind.

Gaußtest

Annahmen:
$X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$, wobei σ^2 bekannt ist. Alternativ gelte $\mathbb{E}[X_i] = \mu$ und Var$[X_i] = \sigma^2$, und n sei groß genug.

Hypothesen	Ablehnungskriterium für H_0
$H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$	$\|Z\| > z_{1-\alpha/2}$
$H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$	$Z < z_\alpha$
$H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$	$Z > z_{1-\alpha}$

Testgröße:
$$Z := \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}$$

t-Test

Annahmen:
$X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$. S^2 sei die Stichprobenvarianz. Alternativ gelte $\mathbb{E}[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$, und n sei groß genug.

Hypothesen	Ablehnungskriterium für H_0
$H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$	$\|T\| > t_{n-1,1-\alpha/2}$
$H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$	$T < t_{n-1,\alpha}$
$H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$	$T > t_{n-1,1-\alpha}$

Testgröße:
$$T := \frac{\bar{X} - \mu_0}{S}\sqrt{n}$$

Zwei-Stichproben-t-Test

Annahmen:
$X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ seien unabhängig und jeweils identisch verteilt, wobei $X_i \sim \mathcal{N}(\mu_X, \sigma_X^2)$ und $Y_i \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ gelte. Die Varianzen seien identisch, also $\sigma_X^2 = \sigma_Y^2$

Hypothesen	Ablehnungskriterium für H_0
$H_0 : \mu_X = \mu_Y$ gegen $H_1 : \mu_X \neq \mu_Y$	$\|T\| > t_{m+n-2,1-\alpha/2}$
$H_0 : \mu_X \geq \mu_Y$ gegen $H_1 : \mu_X < \mu_Y$	$T < t_{m+n-2,\alpha}$
$H_0 : \mu_X \leq \mu_Y$ gegen $H_1 : \mu_X > \mu_Y$	$T > t_{m+n-2,1-\alpha}$

Testgröße:
$$T := \sqrt{\frac{n+m-2}{\frac{1}{m}+\frac{1}{n}}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{(m-1) \cdot S_X^2 + (n-1) \cdot S_Y^2}}$$

χ^2-Anpassungstest

Annahmen:
$X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $W_{X_i} = \{1, \ldots, k\}$.

Hypothesen	
$H_0 : \Pr[X = i] = p_i$	für $i = 1, \ldots, k$
$H_1 : \Pr[X = i] \neq p_i$	für mindestens ein $i \in \{1, \ldots, k\}$

Ablehnungskriterium für H_0 bei Signifikanzniveau α:
$T > \chi^2_{k-1,1-\alpha}$,
dabei sollte gelten, dass $np_i \geq 1$ für alle i und $np_i \geq 5$ für mindestens 80% der Werte $i = 1, \ldots, k$

Testgröße:
$$T = \sum_{i=1}^{k} \frac{(h_i - np_i)^2}{np_i},$$
wobei h_i die Häufigkeit angibt, mit der $X_1, \ldots, X_n$ den Wert i angenommen haben.

Stochastische Prozesse

Wir betrachten zeitliche Folgen von Zufallsexperimenten. Mathematisch beschreibt man diese durch einen so genannten stochastischen Prozess. Darunter versteht man eine Folge von Zufallsvariablen $(X_t)_{t\in T}$, die das Verhalten des Systems zu verschiedenen Zeitpunkten t angeben.

Sei $T = \mathbb{N}_0 \implies$ stochastische Prozesse mit diskreter Zeit.
Sei $T = \mathbb{R}_0^+ \implies$ stochastische Prozesse mit kontinuierlicher Zeit.

Eine besonders einfache Art von stochastischen Prozessen sind so genannte Markov-Ketten. Diese haben die Eigenschaft, dass der nächste Zustand des Prozesses zwar vom aktuellen Zustand abhängen darf, nicht aber von der Historie, d.h. davon, wie der aktuelle Zustand erreicht wurde.

Markov-Kette

Eine (endliche) Markov-Kette (mit diskreter Zeit) über der Zustandsmenge $S = \{0, \ldots n-1\}$
besteht aus einer unendlichen Folge von Zufallsvariablen $(X_t)_{t\in\mathbb{N}_0}$ mit
Wertemenge S
sowie der Starterteilung q_0 mit $q_0^T \in \mathbb{R}^n$.
Die Komponenten von q_0 sind ≥ 0 und addieren sich zu 1.
Für jede Indexmenge $I \subseteq \{0, \ldots, t-1\}$ und beliebige Zustände $i, j, s_k (k \in I)$ gilt:
$\Pr[X_{t+1} = j | X_t = i, \ \forall k \in I : X_k = s_k] = \Pr[X_{t+1} = j | X_t = i]$

Startverteilung

$q_0 = (p_{\varepsilon 0}, p_{\varepsilon 1}, \ldots, p_{\varepsilon n})$ mit $S = 0, \ldots, n$, wobei $p_{\varepsilon i}$ die Wahrscheinlichkeit angibt im Zustand i zu starten.

Zeithomogene Markov-Kette

Sind die Werte $p_{ij} := \Pr[X_{t+1} = j | X_t = i]$ von t unabhängig, so nennt man die Markov-Kette (zeit)homogen.

Übergangsmatrix

Sei die Markov-Kette homogen. $\implies$
Die Übergangsmatrix ist:
$$P = (p_{ij})_{0\leq i,j<n} = \begin{bmatrix} p_{00} & \cdots & p_{0n} \\ \vdots & \ddots & \vdots \\ p_{n0} & \cdots & p_{nn} \end{bmatrix}$$
Knoten $\mathrel{\hat{=}}$ Zustände S
Kanten $\mathrel{\hat{=}}$ Übergangswahrscheinlichkeit

Unendliche Markov-Kette

Sei $S = \mathbb{N}_0 \implies$
die Markov-Kette ist unendlich.

Wahrscheinlichkeitsraum einer Markov-Kette

Man betrachte die Kette von der Zeit 0 bis t_0. Sei $\vec{x} = (x_0, x_1, \ldots, x_{t_0})$ die Folge von Zuständen, die von der Kette in dieser Zeit durchlaufen wurde und $\Omega \subseteq S^{t_0+1}$ die Menge möglicher Zustandsfolgen.
Zu einer beliebigen Folge $\omega := (x_0, x_1, \ldots, x_{t_0}) \in \Omega$ ist der diskrete

Wahrscheinlichkeitsraum:
$\Pr[\omega] = (q_0)_{x_0} \cdot \prod_{i=1}^{t_0} \Pr[X_i = x_i | X_{i-1} = x_{i-1}]$

Berechnung von Übergangswahrscheinlichkeiten

Die Situation zum Zeitpunkt t wird durch den Zustandsvektor q_t (Zeilenvektor) beschrieben.
$(q_t)_i$ bezeichnet die Wahrscheinlichkeit, mit der sich die Kette nach t Schritten im Zustand i aufhält.
Es gilt:
$\Pr[X_{t+1} = k] = \sum_{i=0}^{n-1} \Pr[X_{t+1} = k | X_t = i] \cdot \Pr[X_t = i]$
also $(q_{t+1})_k = \sum_{i=0}^{n-1} p_{ik} \cdot (q_t)_i$
bzw. $q_{t+1} = q_t \cdot P$

$q_t = q_0 \cdot P^t$ und $q_{t+k} = q_t \cdot P^k$

Die Einträge P^k geben an, mit welcher Wahrscheinlichkeit ein Übergang vom Zustand i zum Zustand j in genau k Schritten erfolgt:

$p_{ij}^{(k)} := \Pr[X_{t+k} = j | X_t = i] = (P^k)_{ij}$

Exponentiation von Matrizen

Sei P diagonalisierbar $\implies$
$P^k = B \cdot D^k \cdot B^{-1}$

Übergangszeit

Die Anzahl der Schritte, die von der Markov-Kette für den Weg von Zustand i nach Zustand j benötigt werden, wird als Übergangszeit bezeichnet und ist definiert durch:
$T_{ij} := \min\{n \geq 0 | X_n = j, \text{wenn} X_0 = i\}$

Wenn j nie erreicht wird gilt: $T_{ij} = \infty$

Rückkehrzeit

Die Anzahl der Schritte, die von der Markov-Kette benötigt wird, um von Zustand i zum Zustand i zurückzukehren, wird als Rückkehrzeit bezeichnet und ist definiert durch:
$T_i := T_{ii} = \min\{n \geq 1 | X_n = i, \text{wenn} X_0 = i\}$

Ankunftswahrscheinlichkeit

Die Wahrscheinlichkeit in beliebig vielen Schritten vom Zustand i in den Zustand j zu gelangen, wird als Ankunftswahrscheinlichkeit bezeichnet und ist definiert durch:
$f_{ij} := \Pr[T_{ij} < \infty] = p_{ij} + \sum_{k \neq j} p_{ik} \cdot f_{kj}$ für alle $i, j \in S, i \neq j$

Rückkehrwahrscheinlichkeit

Die Wahrscheinlichkeit in beliebig vielen Schritten vom Zustand j in den Zustand j zurückzukehren, wird als Rückkehrwahrscheinlichkeit bezeichnet und ist definiert durch:
$f_j := f_{jj} = \Pr[T_j < \infty] = p_{jj} + \sum_{k \neq j} p_{jk} \cdot f_{kj}$

erwartete Übergangszeit

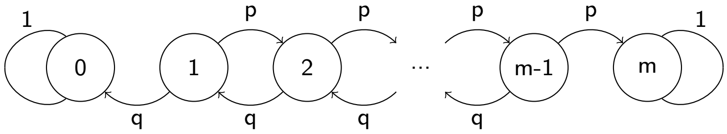
$h_{ij} := \mathbb{E}[T_{ij}] = 1 + \sum_{k \neq j} p_{ik} \cdot h_{kj}$ für alle $i, j \in S, i \neq j$
sofern die Erwartungswerte h_{ij} und h_{kj} existieren.

erwartete Rückkehrzeit

$h_j := h_{jj} = \mathbb{E}[T_j] = 1 + \sum_{k \neq j} p_{jk} \cdot h_{kj}$
sofern der Erwartungswert h_{kj} existiert.

Gambler's Ruin Problem

Sei folgende Markov-Kette gegeben mit $q = 1 - p$:



$$\implies f_{i,m} = \frac{1 - \left(\frac{1-p}{p}\right)^j}{1 - \left(\frac{1-p}{p}\right)^m}$$

Stationäre Verteilung

Sei P die Übergangsmatrix einer Markov-Kette. Als stationäre Verteilung dieser Markov-Kette wird der Zustandsvektor π genannt, wenn gilt:
$\pi = \pi \cdot P$

Bedeutung:
Wenn die Markov-Kette einmal den Zustandsvektor π angenommen hat, so bleibt dieser bei allen weiteren Übergängen erhalten.

Nicht alle Markov-Ketten erfüllen diese Eigenschaft.

absorbierend, transient, rekurrent

absorbierend

Ein Zustand i heißt absorbierend, wenn aus ihm keine Übergänge herausführen, d.h. $p_{ij} = 0$ für alle $j \neq i$ und folglich $p_{ii} = 1$

transient

Ein Zustand i heißt transient, wenn $f_i < 1$, d.h. mit positiver Wahrscheinlichkeit $1 - f_i > 0$ kehrt der Prozess nach einem Besuch von i nie mehr dorthin zurück.

rekurrent

Ein Zustand i mit $f_i = 1$ heißt rekurrent.

Irreduzibilität

Eine Markov-Kette heißt irreduzibel, wenn es für alle Zustandspaare $i, j \in S$ eine Zahl $n \in \mathbb{N}$ gibt, so dass $p_{ij}^{(n)} > 0$

Bedeutung:
Jeder Zustand kann von jedem anderen Zustand aus mit positiver

Wahrscheinlichkeit erreicht werden, wenn nur genügend viele Schritte durchgeführt werden. $\implies$ Graph ist stark zusammenhängend.

Für irreduzible endliche Markov-Ketten gilt: $f_{ij} = \Pr[T_{ij} < \infty] = 1$ für alle Zustände $i, j \in S$ und alle $h_{ij} = \mathbb{E}[T_{ij}]$ existieren.

Eine irreduzible endliche Markov-Kette besitzt eine eindeutige stationäre Verteilung π und es gilt $\pi_j = \dfrac{1}{h_{jj}}$ für alle $j \in S$

Aperiodizität

Periode

Die Periode eines Zustands j ist definiert als die größte Zahl $\xi \in \mathbb{N}$, so dass gilt:
$\{n \in \mathbb{N}_0 | p_{jj}^{(n)} > 0\} \subseteq \{i \cdot \xi | i \in \mathbb{N}_0\}$

Ein Zustand mit Periode $\xi = 1$ heißt aperiodisch.

Ein Zustand $i \in S$ ist genau dann aperiodisch, falls gilt:
Es gibt ein $n_0 \in \mathbb{N}$, sodass $p_{ii}^{(n)} > 0$ für alle $n \in \mathbb{N}, n \geq n_0$

Sind alle Zustände aperiodisch $\implies$ die Markov-Kette ist aperiodisch.

Ein Zustand $i \in S$ einer endlichen Markov-Kette ist sicherlich dann aperiodisch, wenn er im Übergangsdiagramm

- eine Schleife besitzt (also $p_{ii} > 0$) oder
- auf mindestens zwei geschlossenen Wegen W_1 und W_2 liegt, deren Längen l_1 und l_2 teilerfremd sind (für die also ggT$(l_1, l2) = 1$ gilt).

Ergodizität

Irreduzible aperiodische Markov-Ketten nennt man ergodisch.

Für ergodische endliche Markov-Ketten gilt:
Es gibt ein $t \in \mathbb{N}$, sodass unabhängig vom Startzustand $(q_t)_i > 0$ für alle $i \in S$

Für jede ergodische endliche Markov-Kette $(X_t)_{t \in \mathbb{N}_0}$ gilt unabhängig vom Startzustand:
$\lim_{n \to \infty} q_n = \pi$,
wobei π die eindeutige stationäre Verteilung der Kette bezeichnet.

Doppeltstochastische Matrizen

stochastische Matrizen

Eine $n \times n$ Matrix $P = (p_{ij})_{0 \leq i,j < n}$ heißt stochastisch, falls alle Einträge p_{ij} nichtnegativ und alle Zeilensummen gleich 1 sind. Also:
$\sum_{j=0}^{n-1} p_{ij} = 1$ für alle $i = 0, \dots, n-1$

Die Übergangsamtrix einer Markov-Kette ist immer stochastisch und umgekehrt.

doppeltstochastische Matrizen

Eine $n \times n$ Matrix $P = (p_{ij})_{0 \leq i,j < n}$ heißt doppeltstochastisch, falls alle Einträge p_{ij} nichtnegativ und alle Zeilensummen und alle Spaltensummen gleich 1 sind. Also:
$\sum_{j=0}^{n-1} p_{ij} = 1$ für alle $i = 0, \dots, n-1$ und
$\sum_{i=0}^{n-1} p_{ij} = 1$ für alle $j = 0, \dots, n-1$

Sei P eine doppeltstochastische $n \times n$ Matrix $\implies$
$\pi = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)$ ist ein Eigenvektor zum Eigenwert 1 bezüglich der Multiplikation von links: $\pi = \pi \cdot P$

Für jede ergodische endliche Markov-Kette $(X_t)_{t \in \mathbb{N}_0}$ mit doppeltstochastischer Übergangsmatrix gilt unabhängig vom Startzustand:
$\lim_{t \to \infty} q_t = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)$, wobei n die Kardinalität der Zustandsmenge bezeichne.

Mathematische Grundlagen

Bekannte Reihen

geometrische Reihe	$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}$
	$\sum_{k=0}^{n} p^k = \begin{cases} \frac{1-p^{n+1}}{1-p} & p \neq 1 \\ n+1 & p = 1 \end{cases}$
Summe der natürlichen Zahlen	$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$
harmonische Reihe	$\sum_{k=1}^{\infty} \frac{1}{k} = \infty$
alternierende harmonische Reihe	$\sum_{k=1}^{\infty} (-1)^{k+1} \cdot \frac{1}{k} = \ln(2)$
	$\sum_{k=1}^{\infty} \frac{(sp)^k}{k} = \ln\left(\frac{1}{1-sp}\right)$, falls $0 \leq sp < 1$
Exponentialreihe	$\sum_{k=0}^{\infty} \frac{x^k}{k!} = \exp(x)$

Bekannte Integrale

$I := \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$

Integralrechnung

Substitutionsregel

$$\int_a^b f(g(x)) \cdot g'(x) dx = \int_{g(a)}^{g(b)} f(u) du$$

Merkregel:
$u = g(x),\ du = g'(x)dx$
$a \Rightarrow g(a),\ b \Rightarrow g(b)$

Beispiel:
$\int_0^1 \sin(2x) \cdot 2dx,\ g(x) = 2x,\ f(g(x)) = \sin(2x),\ g'(x) = 2$
$u := 2x,\ du = 2dx,\ g(0) = 0,\ g(1) = 2$
$\int_0^1 \sin(2x) \cdot 2dx = \int_0^2 \sin(u) du = [-\cos(u)]_0^2 = [-\cos(2x)]_0^1$

Partielle Integration

$$\int_a^b f(x) g'(x) dx = [f(x) \cdot g(x)]_a^b - \int_a^b f'(x) \cdot g(x) dx$$

Kombinatorik

Anzahl an Möglichkeiten k Elemente aus einer n-elementigen Menge zu ziehen:

	geordnet	ungeordnet
mit zurücklegen	n^k	$\binom{n+k-1}{n-1}$
ohne zurücklegen	$n^{\underline{k}} = \frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Hypergeometrische Verteilung

Eine Zufallsvariable X mit der Dichte
$\Pr[X = x] = \frac{\binom{b}{x}\binom{a}{r-x}}{\binom{a+b}{r}}$
hypergeometrisch verteilt.
Es beschreibt das Ziehen von r Elementen ohne Zurücklegen aus einer Grundmenge der Mächtigkeit $a + b$ mit b besonders ausgezeichneten Elementen.
z.B: Ziehen von x Buben aus einem Set von $32 = a + b$ Karten mit $4 = b$ besonders ausgezeichneten Elementen (Anzahl an Buben im Kartenset)

Sonstiges

Sei $X \sim \mathcal{N}(\mu, \sigma^2) \implies$
$\Pr[-x \leq X \leq x] = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$

Anmerkungen

Dies ist eine Zusammenfassung der Vorlesung Diskrete Wahrscheinlichkeitstheorie an der Technischen Universität München. Gehalten wurde diese Vorlesung durch Albers S. im Sommersemester 2018. Ersteller dieser Zusammenfassung ist Gaida B. Alle Angaben sind ohne Gewähr.