Mitschrieb der Vorlesung

Stochastik für Informatiker und Bioinformatiker

Prof. Dr. Manfred Wolff

Sommersemseter 2008*

Mitschrieb in LaTeX von
Rouven Walter

*Letzte Änderung: 8. März 2011

Lizenz

Vorwort

Dieser Mitschrieb entstand während meiner Nachbearbeitung zur Stochastik Vorlesung im Sommersemester 2008 bei Prof. Dr. Manfred Wolff an der Eberhard-Karls-Universität Tübingen.

Ich erhebe keinen Anspruch auf Vollständigkeit oder Richtigkeit. Bei Verständnisschwierigkeiten zum Inhalt empfehle ich daher ausdrücklich, sich an die jeweiligen Dozenten/Tutoren zu wenden.

Wer Fehler findet, Verbesserungsvorschläge hat oder sonstige Anregungen mitteilen möchte, kann mir gerne eine E-Mail an folgende Adresse schicken:
`rouvenwalter@web.de` oder
`kontakt@upenon.de`

Danksagung

Ich möchte mich ganz herzlich bei Prof. Dr. Manfred Wolff bedanken. Er reichte viele aufwendige Beweise nach und machte Ergänzungen, so dass das Skript durch ihn erst vollständig wurde.

Mein Dank geht auch an Steffen Just, der half einige Fehler im Skript ausfindig zu machen.

Inhaltsverzeichnis

Teil I.

Diskrete Warscheinlichkeitsräume

1. Beispiele und Grundlegende Definitionen

1.1. Beispiele

Münzwurf: $\{Z, W\} = \Omega,\ P(Z) = P(W) = \frac{1}{2},\ (\Omega, P)$.

Allgemeiner:
Erfolgs-Misserfolgs-Experiment:
Erfolg wird durch 1 kodiert, Misserfolg durch 0.
$\Omega = \{0, 1\},\ P(1) = \frac{1}{6} =: p,\ P(0) = \frac{5}{6} = 1 - p =: q$

1.2. Beispiel

Würfeln: $\Omega = \{1, 2, 3, 4, 5, 6\},\ P(\omega) = \frac{1}{6}$

Frage: Was ist die Wahrscheinlichkeit für einen Wurf mit der Augenzahl ≥ 5 ?

$A = \{5, 6\}$

$$\begin{aligned} P(A) &= \frac{1}{3} \\ &= P(5) + P(6) \\ &= \sum_{\omega \in A} P(\omega) \\ &= \frac{|A|}{|\Omega|} \end{aligned}$$

in unserem Fall (wird es ein Laplac'escher Wahrscheinlichtkeits-Raum).

1.3. Beispiel

Zahlen-Lotto: $\Omega = \{\omega : \omega \subset \{1, \ldots, 49\},\ |\omega| = 6\}$

$$P(\omega) = \frac{1}{\binom{49}{6}}$$

$A = \{\omega : \omega$ enthält 5 Richtige und die Zusatzzahl$\}$

Special Leading Case

Ergebnis: (1, 2, 3, 4, 5, 6), Zusatzzahl = 7
$|A| = 6$

$$\begin{aligned} P(A) &= \frac{|A|}{|\Omega|} \\ &= \frac{\binom{6}{5}}{\binom{49}{6}} \end{aligned}$$

1.4. Beispiel

$\Omega = \mathbb{N} = \{0, 1, 2, \ldots\}, \ \lambda > 0, \ \lambda \in \mathbb{R}$

$$\begin{aligned} P(k) &= e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ \sum_{k=0}^{\infty} P(k) &= e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \cdot e^{\lambda} \\ &= 1 \end{aligned}$$

Dies ist die sogenannte *Poisson-Verteilung* auf $\mathbb{N}_0$.

$\lambda = 1:$

$$\begin{aligned} P(\{0, 1, 2, 3\}) &= e^{-1} \cdot \left(1 + 1 + \frac{1}{2} + \frac{1}{6}\right) \\ &= 0{,}981 \end{aligned}$$

1.5. Definition

Sei $\emptyset \neq \Omega$ eine endliche oder abzählbare Menge. Sei $P : \Omega \to [0, 1]$, $\omega \in \Omega \mapsto P(\omega)$ und es gelte $\sum_{\omega \in \Omega} P(\omega) = 1$.
Wir setzen für $A \subset \Omega$

$$P(A) := \sum_{\omega \in A} P(\omega)$$

also $P : \mathcal{P}(\Omega) \to [0, 1]$. Insbesondere $P(\{\omega\}) := P(\omega)$.
(Ω, P) heißt *diskreter Wahrscheinlichkeitsraum*.
$\omega \in \Omega$ heißt *Elementarereignis*, $A \subset \Omega$ heißt *Ereignis*.
$\emptyset$ heißt *unmögliches Ereignis*, Ω *sicheres Ereignis*.
$P : \mathcal{P}(\Omega) \to [0, 1]$ heißt *Wahrscheinlichkeitsmaß*.

1.6. Satz

Sei (Ω, P) ein diskreter Warscheinlichkeits-Raum. Dann gilt

a) $0 \leq P(A) \leq 1$ für alle $A \subset \Omega$

b) $P(\emptyset) = 0,\ P(\Omega) = 1$ (Beweis s. Definiton 1.5)

c) $(A \subset B) \Rightarrow P(A) \leq P(B)$ (Beweis s. Definiton 1.5)

d) Sei (A_k) eine Folge von Ereignissen mit $A_i \cap A_k = \emptyset$ für $i \neq k$. Dann gilt

$$P\left(\bigcup_k A_k\right) = \sum_k P(A_k)$$

Wenn $A_i \cap A_k = \emptyset$, so sagt man A_i und A_k schließen sich gegenseitig aus.

Beweis:

Zu d):

(i) Seien A, B mit $A \cap B = \emptyset$, $A \cup B =: C$.

$$\begin{aligned} P(C) &= \sum_{\omega \in C} P(\omega) \\ &= \sum_{w \in A} P(A) + \sum_{\omega \in B} P(B) \\ &= P(A) + P(B) \end{aligned}$$

(ii) $A_1, \ldots, A_{n+1}$ mit $A_i \cap A_k = \emptyset$ für $i \neq k$,
$A = \bigcup_{k=1}^n A_k$, $B = A_{n+1}$, $A \cap B = \emptyset$.
Also

$$\begin{aligned} P\left(\bigcup_{k=1}^{n+1} A_k\right) &= P(A \cup B) \\ &\underset{(i)}{=} P(A) + P(B) \\ &\underset{\text{IV}}{=} \sum_{k=1}^{n} P(A_k) + P(A_{n+1}) \\ &= \sum_{k=1}^{n+1} P(A_k) \end{aligned}$$

Alternativer Beweis zu d):

Man hat also die endliche Additivität. Sei nun $(A_n)_n$ eine abzählbare Folge von sich paarweise ausschließenden Ereignissen und $A = \bigcup_{n \in \mathbb{N}} A_n$. Dann gilt $\bigcup_{k=1}^n A_k \subseteq A$, also nach c) $\sum_{k=1}^{n} P(A_k) = P(\bigcup_{k=1}^{n} A_k) \leq P(A)$. Da n beliebig war, folgt $\sum_{k=1}^{\infty} P(A_k) \leq P(A)$.

Sei nun $\varepsilon > 0$ beliebig. Dann gibt es eine endliche Teilmenge $M \subseteq A$ mit $P(A) - \varepsilon < \sum_{\omega \in M} P(\omega) = P(M) \leq P(A)$., aber dann ist $A_j \cap M \neq \emptyset$ nur für endlich viele j wahr und für diese ist $M = \bigcup_{A_j \cap M \neq \emptyset} A_j \cap M$, also erhält man in der nächsten Ungleichung nach dem Gleichheitszeichen in Wahrheit nur eine endliche Summe

$$P(A) - \varepsilon < P(M) = \sum_{k=1}^{\infty} P(A_k \cap M) \leq \sum_{k=1}^{\infty} P(A_k) \leq P(A)$$

und die Behauptung folgt, weil $\varepsilon > 0$ beliebig war.

1.7. Laplac'escher Warscheinlichkeits-Raum

Sei Ω endlich, $P(\omega) = \frac{1}{|\Omega|}$, $A \subseteq \Omega$,

$$\begin{aligned} P(A) &= \sum_{\omega \in A} \frac{1}{|\Omega|} \\ &= \frac{|A|}{|\Omega|} \\ &= \frac{\text{Anzahl günstiger Ereignisse}}{\text{Anzahl möglicher Ereignisse}} \end{aligned}$$

1.8. Wiederholung von Experimenten

Sei (Ω_0, P_0) ein diskreter Warscheinlichkeits-Raum.
Die r-malige Wiederholung dieses „Experiments" wird modelliert durch

$$\begin{aligned} \Omega &= \Omega_0^r \\ &= \{(\omega_1, \ldots, \omega_r) : \omega_k \in \Omega_0\} \end{aligned}$$

$$\begin{aligned} P(\vec{\omega}) &= P((\omega_1, \ldots, \omega_r)) \\ &= P_0(\omega_1) \cdot P_0(\omega_2) \cdot \ldots \cdot P_0(\omega_r) \\ &= \prod_{i=1}^{r} P_0(\omega_i) \end{aligned}$$

1.9. Beispiele

a) $\Omega_0 = \{1, \ldots, 6\}$, $r = 2$.

$$\begin{aligned} \Omega &= \{1, \ldots, 6\}^2 \\ &= \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,6)\} \end{aligned}$$

$$\begin{aligned} P_0(k) &= \frac{1}{6} \\ P((i,k)) &= P_0(i) \cdot P_0(k) \\ &= \frac{1}{6} \cdot \frac{1}{6} \\ &= \frac{1}{36} \end{aligned}$$

b) $r = 5$, $\Omega_0 = \{0, 1\}$

$$\begin{aligned} P_0(1) &= p \\ P_0(0) &= q \\ &= 1 - p \end{aligned}$$

$$\Omega = \Omega_0^5$$

Beispiel:

$$P((0,1,1,0,0)) = p^2 \cdot q^3$$

1.10. Anwendung des Erfolgs-Misserfolgs-Experiement auf den Vorzeichentest

(Bartz-Lienert-Boehnke: Verteilungsfreie Methoden in der Biostatistik, S.256 ff)

Ergebnis:

	1	2	3	4	5	6	7	8	9	10
Roh	61	60	56	63	56	63	59	56	44	61
geröstet	55	54	47	59	51	61	57	54	62	58
Differenz	6	6	9	4	5	2	2	2	18	3
Vorzeichen	+	+	+	+	+	+	+	+	−	+

H_0 reiner Zufall,

$$\begin{aligned} P_{10}(\text{Anzahl}+ \geq 9) &= \left(\frac{1}{2}\right)^{10} + \frac{10}{2^{10}} \\ &= \frac{11}{1024} \\ &\approx 0{,}0107 \end{aligned}$$

Signifikanzniveau $\alpha = 0{,}05$. Ist $P(H_0) \leq 0{,}05$, dann H_0 abgelehnt.

1.11. Beispiel

$\{A, C, G, T\}$ Sequenz der Länge 10.

$$P(\text{Übereinstimmung}) = \frac{1}{4}$$

In einer Sequenz von Paaren der Länge 10 haben wir 8 Übereinstimmungen.
Paare: (A, A), (A, C), (C, A), ...

$$\begin{aligned} P(\text{acht und mehr Übereinstimmungen}) &= \binom{10}{8} \cdot \frac{1}{4^8} \cdot \frac{3^2}{4^2} + \binom{10}{9} \cdot \frac{1}{4^9} \cdot \frac{3}{4} \\ &\quad + \binom{10}{10} \cdot \frac{1}{4^{10}} \\ &= \frac{1}{2^{20}} \cdot (45 \cdot 9 + 10 \cdot 3 + 1) \end{aligned}$$

2. Zufallsvariable

2.1. Beispiele

a) Würfel.

$$\begin{aligned}\Omega &= \{\text{Augenzahl } 1, \text{Augenzahl } 2, \text{Augenzahl } 3,\\ &\quad \text{Augenzahl } 4, \text{Augenzahl } 5, \text{Augenzahl } 6\}\end{aligned}$$

$$X : \Omega \to \mathbb{R}$$

wobei

$$\begin{aligned}X(\text{Augenzahl } 1) &\mapsto 1\\ X(\text{Augenzahl } 2) &\mapsto 2\\ X(\text{Augenzahl } 3) &\mapsto 3\\ X(\text{Augenzahl } 4) &\mapsto 4\\ X(\text{Augenzahl } 5) &\mapsto 5\\ X(\text{Augenzahl } 6) &\mapsto 6\end{aligned}$$

b)

$$\begin{aligned}\Omega_0 &= \{0,1\}\\ P(0) &= q\\ &= 1-p\\ P(1) &= p\end{aligned}$$

$$\begin{aligned}\Omega &= \Omega_0^n\\ P((\omega_1,\ldots,\omega_n)) &= p^{\text{Erfolgsanzahl}} \cdot (1-p)^{\text{Misserfolgsanzahl}}\\ &= p(\omega_1)\cdot\ldots\cdot p(\omega_n)\end{aligned}$$

$$X((\omega_1,\ldots,\omega_n)) = \omega_1 + \ldots + \omega_n$$

2.2. Definition

Gegeben (Ω, P). Eine Funktion

$$X : \Omega \to \begin{cases}\mathbb{R}\\ \mathbb{C}\end{cases}$$

heißt *Zufallsvariable* (ZV).

2.3. Definition

a) Sei $X : \Omega \to \mathbb{C}$ eine Zufallsvariable und

$$\begin{aligned} \Omega' &= X(\Omega) \\ &= \{X(\omega) : \omega \in \Omega\} \end{aligned}$$

$z \in \Omega'$

$$P_X(z) = P(\{\omega : X(\omega) = z\})$$

(Ω', P_X) heißt *Verteilung* von X.

b) $X : \Omega \to \mathbb{R}$, $F_X : \mathbb{R} \to \mathbb{R}$

$$F_X(t) = P(\{\omega : X(\omega) \leq t\})$$

heißt *Verteilungsfunktion*.

2.4. Beispiel

a) Ω wie in Beispiel 2.1 a),
$X : \Omega \to \mathbb{R}$, $X(\text{Augenzahl } x) = \text{Augenzahl}$,

$$\begin{aligned} X(\Omega) &= \Omega' \\ &= \{1, 2, 3, 4, 5, 6\} \\ P(X = i) &= \frac{1}{6} \qquad \text{für } i \in \Omega' \end{aligned}$$

$$F_X(t) = \begin{cases} 0 & t < 1 \\ \frac{1}{6} & 1 \leq t < 2 \\ \frac{1}{3} & 2 \leq t < 3 \\ \frac{1}{2} & 3 \leq t < 4 \\ \frac{2}{3} & 4 \leq t < 5 \\ \frac{5}{6} & 5 \leq t < 6 \\ 1 & 6 \leq t \end{cases}$$

b) *Binomialverteilung*.

Ω siehe 2.1 b).

$$\begin{aligned} X((\omega_1,\dots,\omega_n)) &= \sum_{j=1}^{n} \omega_j \\ X(\Omega) &= \Omega' \\ &= \{0,1,2,\dots,n\} \\ P_X(0) &= (1-p)^n \\ &= q^n \\ P_X(1) &= \binom{n}{1} \cdot p \cdot q^{n-1} \\ P_X(k) &= \binom{n}{k} \cdot p^k \cdot q^{n-k} \end{aligned}$$

$\{0,1,\dots,n\}$ *Binomialverteilung* $B(n,p)$

c) $\Omega \subset \mathbb{R}$, $X : \Omega \to \mathbb{R}$, $X(\omega) = \omega$, $P_X = P$.

Z.B. $\Omega = \mathbb{N}_0$, $P(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$

d) $\Omega = \{\omega \in \{0,1\}^{\mathbb{N}} :\ \exists k[\omega_l = 0\ \forall l \geq k]\}$

$$\begin{aligned} L(\omega) &= \min\{k : \omega_l = 0 \text{ für alle } l \geq k+1\} \\ S(\omega) &= \sum_{j=1}^{\infty} \omega_j \\ &= \sum_{j=1}^{L(\omega)} \omega_j \\ P(\omega) &= p^{S(\omega)} \cdot q^{L(\omega)-S(\omega)} \end{aligned}$$

Wobei $0 < p < 1$ und $q = 1 - p$.

$$Y(\omega) = \min\{k :\ \omega_k = 0\}$$

$$\begin{aligned} Y((0,0,0,\dots,0)) &= 1 \\ Y((1,1,0,1,1,1,0,\dots,0)) &= 3 \end{aligned}$$

$$\begin{aligned} P(Y(\omega) = k) &= P_Y(k) \\ &= q \cdot p^{k-1} \end{aligned}$$

$$\begin{aligned} \sum_{k=1}^{\infty} q \cdot p^{k-1} &= \frac{q}{1-p} \\ &= 1 \end{aligned}$$

$\Omega' = \{1,2,\dots,\} = \mathbb{N}$, P_Y heißt *Geometrische Verteilung*.

2.5. Definition (Erwartungswert, Varianz, Streuung)

a) (Ω, P) endlich und $X : \Omega \to \mathbb{C}$ Zufallsvariable.
Dann heißt

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega) \cdot P(\omega)$$

Erwartungswert von X.

b) Sei Ω abzählbar und nicht endlich.
Es konvergiere

$$\sum_{\omega\in\Omega} |X(\omega)| \cdot P(\omega) \qquad (< \infty)$$

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega) \cdot P(\omega)$$

c) Wir setzen

$$\begin{aligned} \overline{X} &= X - \mathbb{E}(X) \cdot 1_\Omega \\ &= X - \mathbb{E}(X) \end{aligned}$$

$$V(X) = \mathbb{E}(|\overline{X}|^2)$$

heißt *Varianz*.

$$\sigma(X) = \sqrt{V(X)}$$

heißt *Streuung*.

2.6. Beispiele

a) $\Omega \subset \mathbb{R}$, $\Omega = |n|$, Laplace'scher Warscheinlichkeitsraum, also $P(\omega) = \frac{1}{|\Omega|}$

$$\begin{aligned} X(\omega) &= \omega \\ \mathbb{E}(X) &= \sum_{\omega\in\Omega} X(\omega) \cdot P(\omega) \\ &= \sum_{\omega\in\Omega} \omega \cdot \frac{1}{|\Omega|} \\ &= \frac{1}{|\Omega|} \cdot \sum_{\omega\in\Omega} \omega \\ &= \frac{1}{n} \cdot \sum_{\omega\in\Omega} \omega \end{aligned}$$

arithmetisches Mittel

$$V(X) = \frac{1}{n} \cdot ((\omega - \mathbb{E}(X))^2)$$

mittlere quadratische Abweichung

b) $\Omega = \{0, 1, \ldots, n\}$

$$\begin{aligned} P(k) &= \binom{n}{k} \cdot p^k \cdot q^{n-k} \\ X(k) &= k \\ \mathbb{E}(X) &= ? \end{aligned}$$

2.7. Erzeugendenfunktion

Sei $\Omega \subset \mathbb{N}_0$

$$\begin{aligned} g(z) &= \sum_{k \in \Omega} z^k \cdot P(k) \\ \sum_{k \in \Omega} P(k) &= 1 \end{aligned}$$

$$\begin{aligned} g'(z) &= \sum_{k \in \Omega} k \cdot z^{k-1} \cdot P(k) \\ g'(1) &= \mathbb{E}(X) \qquad X(k) = k \end{aligned}$$

$$\begin{aligned} g''(z) &= \sum_{k \in \Omega} k \cdot (k-1) \cdot z^{k-2} \cdot P(k) \\ g''(1) &= \mathbb{E}(X^2) - \mathbb{E}(X) \end{aligned}$$

2.8. Beispiele

a) $B(n, p)$

$$\begin{aligned} g(z) &= \sum_{k=0}^{n} z^k \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ &= (z \cdot p + (1-p))^n \end{aligned}$$

b) *Poisson-Verteilung*

$$\begin{aligned} g(z) &= \sum_{k=0}^{\infty} z^k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \\ &= e^{-\lambda} \cdot e^{z \cdot \lambda} \end{aligned}$$

c) Geometrische Verteilung

$$\begin{aligned} g(z) &= \sum_{k=1}^{\infty} q \cdot z^k \cdot p^{k-1} \\ &= q \cdot z \cdot \frac{1}{1 - z \cdot p} \qquad |z| \leq 1 \end{aligned}$$

Bemerkung:
$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega), \quad X(\Omega) = \Omega' \subset \mathbb{C}$

$$\begin{aligned} P_X(t) &= P(X = t) \\ &= P(\{\omega : X(\omega) = t\}) \\ \{\omega : X(\omega) = t\} &=: X^{-1}(\{t\}) \\ &= X^{-1}(t) \end{aligned}$$

$$P_X(t) = P\left(X^{-1}(t)\right)$$

2.9. Satz

Es gilt

$$\mathbb{E}(X) = \sum_{t \in \Omega'} t \cdot P_X(t)$$

Beweis:

$$\begin{aligned} \Omega' &= \{t_1, \dots, t_r\} \\ &= X(\Omega) \end{aligned}$$

Daraus folgt:

$$\Omega = \biguplus_{j=1}^{r} X^{-1}(t_j)$$

Weil $t_j \neq t_k$ ist $X^{-1}(t_j) \cap X^{-1}(t_k) = \emptyset$ für $j \neq k$.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{\omega\in\Omega} X(\omega)\cdot P(\omega) \\
&= \sum_{\omega\in X^{-1}(t_1)} X(\omega)\cdot P(\omega) + \ldots + \sum_{\omega\in X^{-1}(t_r)} X(\omega)\cdot P(\omega) \\
&= t_1 \cdot \underbrace{\sum_{\omega\in X^{-1}(t_1)} P(\omega)}_{=P_X(t_1)} + \ldots + t_r \cdot \underbrace{\sum_{\omega\in X^{-1}(t_r)} P(\omega)}_{=P_X(t_r)} \\
&= \sum_{j=1}^{r} t_j \cdot P_X(t_j) \\
&= \sum_{t\in\Omega'} t \cdot P_X(t)
\end{aligned}$$

2.10. Satz (Eigenschaften des Erwartungswertes)

a) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
   $\mathbb{E}(\alpha \cdot X) = \alpha \cdot \mathbb{E}(X)$
   Linearität des Erwartungswertes

b) $X \leq Y \quad \Rightarrow \quad \mathbb{E}(X) \leq \mathbb{E}(Y)$
   Monotonie des Erwartungswertes

c) $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$

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

e) $\sigma(X+Y) \leq \sigma(X) + \sigma(Y)$

Beweis:

a) Klar.

b) Klar.

c) Für $z \in \mathbb{C}$, $z \neq 0$, gibt es α mit $z = e^{i\cdot\alpha} \cdot |z|$
   Also (O.B.d.A. $\mathbb{E} = z \neq 0$)

$$\begin{aligned}
\mathbb{E}(X) &= e^{i\cdot\alpha} \cdot |z| \\
&= e^{i\cdot\alpha} \cdot |\mathbb{E}(X)|
\end{aligned}$$

   Daraus folgt:

$$\begin{aligned}
|\mathbb{E}(X)| &= e^{i\cdot\alpha} \cdot \mathbb{E}(X) \\
&= \sum_{\omega\in\Omega} e^{-i\cdot\alpha} \cdot X(\omega) \cdot P(\omega)
\end{aligned}$$

Daraus folgt:

$$\begin{aligned}
|\mathbb{E}(X)| &= ||\mathbb{E}(X)|| \\
&= \left|\sum_{\omega\in\Omega} e^{-i\cdot\alpha}\cdot X(\omega)\cdot P(\omega)\right| \\
&\leq \sum_{\omega\in\Omega}\left|e^{-i\cdot\alpha}\cdot X(\omega)\right|\cdot P(\omega) \\
&= \sum_{\omega\in\Omega}|X(\omega)|\cdot P(\omega) \\
&= \mathbb{E}(|X|)
\end{aligned}$$

d)

$$\begin{aligned}
1_A(\omega) &= \begin{cases}1 & \omega\in A\\ 0 & \omega\notin A\end{cases} \\
V(X) &= \mathbb{E}((X-\mathbb{E}(X)\cdot 1_\Omega)^2) \\
&= \mathbb{E}(X^2-2\cdot\mathbb{E}(X)\cdot X+\mathbb{E}(X)^2\cdot 1_\Omega) \\
&\underset{2.10\text{ a)}}{=} \mathbb{E}(X^2)+\mathbb{E}(-2\cdot\mathbb{E}(X)\cdot X)+\mathbb{E}(\mathbb{E}(X)^2\cdot 1_\Omega) \\
&\underset{2.10\text{ a)}}{=} \mathbb{E}(X^2)-2\cdot\mathbb{E}(X)\cdot\mathbb{E}(X)+\mathbb{E}(X)^2\cdot\underbrace{\mathbb{E}(1_\Omega)}_{=1} \\
&= \mathbb{E}(X^2)-\mathbb{E}(X)^2
\end{aligned}$$

e) Vorbemerkungen:

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(Y) \\
&= 0 \\
V(X) &= \mathbb{E}(X^2) \\
&= \sigma(X)^2 \\
\sigma(X) &= \sqrt{V(X)} \\
\mathbb{E}(X^2) &= \sum_{\omega\in\Omega} X(\omega)^2\cdot P(\omega) \\
\|X\|_2^2 &= \sum_{\omega\in\Omega} X_j^2 \\
\|X\|_2 &= \sqrt{\sum_{\omega\in\Omega} X_j^2} \\
\|X+Y\|_2 &\leq \|X\|_2+\|Y\|_2
\end{aligned}$$

Start des Beweises:

(I) Cauchy-Schwarzsche-Ungleichung (CSU):

$$\begin{aligned} \mathbb{E}(X \cdot Y)^2 &\leq \mathbb{E}(X^2) \cdot \mathbb{E}(Y^2) \\ \alpha &= \mathbb{E}(Y^2) \\ &\neq 0 \end{aligned}$$

Dann

$$0 \leq \mathbb{E}\left(\left(X - \frac{1}{\alpha}Y\right)^2\right)$$

ausrechnen.

(II)

$$\begin{aligned} \overline{X} &= X - \mathbb{E}(X) \\ \overline{Y} &= Y - \mathbb{E}(Y) \end{aligned}$$

$$\begin{aligned} V(X+Y) &= \mathbb{E}\left(((X+Y) - \mathbb{E}(X+Y))^2\right) \\ &= \mathbb{E}((\overline{X}+\overline{Y})^2) \\ &= \mathbb{E}(\overline{X}^2) + \mathbb{E}(\overline{Y}^2) + 2 \cdot \mathbb{E}(\overline{X}) \cdot \mathbb{E}(\overline{Y}) \\ &\leq \mathbb{E}(\overline{X}^2) + \mathbb{E}(\overline{Y}^2) + 2 \cdot \sqrt{\mathbb{E}(\overline{X}) \cdot \mathbb{E}(\overline{Y})} \\ &= \sigma(X)^2 + \sigma(Y)^2 + 2 \cdot \sigma(X) \cdot \sigma(Y) \\ &= (\sigma(X) + \sigma(Y))^2 \end{aligned}$$

$\Rightarrow$ Behauptung wegen

$$V(X+Y) = \sigma(X+Y)^2$$

2.11. Weitere Parameter von reellen Zufallsvariablen

$$\begin{aligned} M(X) &:= \inf\left\{t : P(\{\omega : X(\omega) \leq t\}) \geq \frac{1}{2}\right\} \quad \text{für unendliche W-räume} \\ &:= \min\left\{t : P(\{\omega : X(\omega) \leq t\}) \geq \frac{1}{2}\right\} \quad \text{für endliche W-räume} \end{aligned}$$

heißt *Median* von X.

$$Q(X)_{\min} := \inf\left\{t : P(\{\omega : X(\omega) \leq t\}) \geq \frac{1}{4}\right\}$$

heißt *unteres Quartil* von X.

$$Q(X)_{\max} \quad := \quad \inf \left\{ t : P(\{\omega : X(\omega) \leq t\}) \geq \frac{3}{4} \right\}$$

heißt *oberes Quartil* von X.

3. Bedingte Warscheinlichkeiten, Stochastische Unabhängigkeit

3.1. Beispiel

Urnen U_1 und U_2.
U_1 enthält 30 rote und 70 schwarze Kugeln,
U_2 enthält 70 rote und 30 schwarze Kugeln.

In U_1 wird mit 80% Warscheinlichkeit gegriffen, in U_2 mit 20% Warscheinlichkeit.

Rote Kugel wird gezogen, $P(U_1|rot) =?$

3.2. Definition

Seien A, B Ereignisse eines Warscheinlichkeitsraums (Ω, P).
Dann heißt die Zahl

$$P(A|B) \quad := \quad \begin{cases} 0 & P(B) = 0 \\ \frac{P(A \cap B)}{P(B)} & \text{sonst} \end{cases}$$

die *bedingte Warscheinlichkeit von A unter der Bedigung B.*

Gleichwertig:

$$P(A \cap B) \quad = \quad P(A|B) \cdot P(B)$$

Folgerung:

$$\begin{aligned} P(A|B) \cdot P(B) &= P(A \cap B) \\ &= P(B \cap A) \\ &= P(B|A) \cdot P(A) \end{aligned}$$

Beispiel 3.1 aufgreifen:

$$\begin{aligned} P(rot|U_1) &= 0,3 \\ P(rot|U_2) &= 0,7 \\ P(U_1) &= 0,8 \\ P(U_2) &= 0,2 \end{aligned}$$

3.3. Satz

Seien $A_1, \dots, A_n$ Ereignisse. Dann gilt

$$\begin{aligned} P(A_1 \cap \ldots \cap A_n) &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \\ &\quad \cdot P(A_4|A_1 \cap A_2 \cap A_3) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_{n-1}) \end{aligned}$$

Beweis:
Induktion über n.
IA: $n = 2$

$$P(A_1 \cap A_2) \underset{\text{Def.}}{=} P(A_1) \cdot P(A_2|A_1)$$

IS: $n \to n + 1$

$$\begin{aligned} P(A_1 \cap \ldots \cap A_{n+1}) &= P(\underbrace{(A_1 \cap \ldots \cap A_n)}_{=:A} \cap \underbrace{A_{n+1}}_{=:B}) \\ &= P(A) \cdot P(B|A) \\ &= P(A_1 \cap \ldots \cap A_n) \cdot P(A_{n+1}|A_1 \cap \ldots \cap A_n) \\ &\underset{\text{IV}}{=} P(A_1) \cdot P(A_2|A_1) \cdot \ldots \cdot P(A_{n+1}|A_1 \cap \ldots \cap A_n) \end{aligned}$$

3.4. Theorem (Satz von Bayes über a posteriori Warscheinlichkeiten)

Seien $A_1, \dots, A_n$ paarweise disjunkte Ereignisse mit

(i) $P(A_k) \neq 0$

(ii) $\biguplus_{k=1}^{n} A_k = \Omega$

(wobei $\biguplus$ die disjunkte Vereinigung kennzeichnet)

Sei B ein beliebigens Ereignis. Dann gilt für jedes k, $1 \leq k \leq n$,

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \ldots + P(A_n) \cdot P(B|A_n)}$$

Beispiel 3.1 aufgreifen:

$$\begin{aligned} P(U_1) &= 0,8 \\ P(U_2) &= 0,2 \end{aligned}$$

B: Ereignis rote Kugel.

$$\begin{aligned} P(B|U_1) &= 0,3 \\ P(B|U_2) &= 0,7 \end{aligned}$$

Nach dem Satz von Bayes gilt

$$\begin{aligned} P(U_1|B) &= \frac{P(U_1)\cdot P(B|U_1)}{P(U_1)\cdot P(B|U_1)+P(U_2)\cdot P(B|U_2)} \\ &= \frac{0,8\cdot 0,3}{0,8\cdot 0,3+0,2\cdot 0,7} \\ &= \frac{0,24}{0,38} \\ &\approx 0,632 \end{aligned}$$

Beweis (Satz von Bayes):

$$P(A_k|B) = \frac{P(A_k)\cdot P(B|A_k)}{P(A_1)\cdot P(B|A_1)+\ldots+P(A_n)\cdot P(B|A_n)}$$

Zähler:

$$P(A_k)\cdot P(B|A_k) \underset{3.2}{=} P(A_k\cap B)$$

Nenner:

$$\underbrace{P(A_1)\cdot P(B|A_1)}_{\underset{3.2}{=}P(A_1\cap B)} + \ldots + \underbrace{P(A_n)\cdot P(B|A_n)}_{\underset{3.2}{=}P(A_n\cap B)}$$

Wir hatten an Voraussetzungen:

(i) $A_k \cap A_l = \emptyset$ für $k \neq l$
$\Rightarrow (A_k\cap B)\cap(A_l\cap B) = \emptyset$ für $k\neq l$
$\Rightarrow$ Nenner $= P\left(\bigcup_{l=1}^n (A_l\cap B)\right) \underset{\text{(ii)}}{=} P(B)$

(ii)

$$\bigcup_{l=1}^n A_l = \Omega$$

$\Rightarrow$

$$\begin{aligned} \bigcup_{l=1}^n (A_l\cap B) &= B\cap\bigcup_{l=1}^n A_l \\ &= B\cap\Omega \\ &= B \end{aligned}$$

Also

$$\begin{aligned} P(A_k|B) &= \frac{P(A_k) \cdot P(B|A_k)}{P(A_1) \cdot P(B|A_1) + \ldots + P(A_n) \cdot P(B|A_n)} \\ &= \frac{P(A_k \cap B)}{P(B)} \end{aligned}$$

Nach 3.3 folgt die Behauptung.

3.5. Beispiel

$\Omega_0 = \{1, \ldots, 6\}$, $\Omega = \Omega_0^2$,
$P((\omega_1, \omega_2)) = \frac{1}{36} = \frac{1}{|\Omega|}$,
$A = \{\omega : \omega_1 = 6\}$, $B = \{\omega : \omega_2 = 6\}$

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P((6,6))}{P(A)} \\ &= \frac{1 \cdot 6}{36} \\ &= \frac{1}{6} \\ &= P(B) \end{aligned}$$

$\Rightarrow$

$$P(A \cap B) = P(A) \cdot P(B)$$

3.6. Definition

a) Die Ereignisse $A_1, \ldots, A_n$ heißen *(stochastisch) unabhängig*, wenn für $2 \leq k \leq n$ und jede Indexfolge $1 \leq i_1 < \ldots < i_k \leq n$ stets gilt

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_k})$$

Es gibt endliche Warscheinlichkeitsräume Ω und A_1, A_2, A_3, so dass A_i und A_j unabhängig für $i < j$ (paarweise unabhängig), aber A_1, A_2, A_3 stoachstisch abhängig.

b) Die abzählbaren vielen Ereignisse $A_1, A_2, \ldots$ heißen *(stochastisch) unabhängig*, wenn jede endliche Teilmenge $\{A_{k_1}, \ldots, A_{k_r}\} \subset \{A_1, \ldots\}$ (stochastisch) unabhängig gemäß a) ist.

c) Die Zufallsvariablen $X_1, \ldots, X_n$ heißen *stochastisch unabhängig*, wenn die Urbilder $X_1^{-1}(]s_1, t_1]), X_2^{-1}(]s_2, t_2]), \ldots, X_n^{-1}(]s_n, t_n])$ stochastisch unabhängig sind für alle $s_i < t_i$, $i = 1, \ldots, n$.

3.7. Satz

Es sei (Ω, P) ein diskreter Warscheinlichkeitsraum und $X_k(\Omega)$ seien endlich.
$X_1, \ldots, X_n$ sind unabhängig genau dann, wenn

$$P([X_1 = s_1] \cap \ldots \cap [X_n = s_n]) = P([X_1 = s_1]) \cdot \ldots \cdot P([X_n = s_n])$$

für $s_j \in X_j(\Omega)$.

Bezeichnung:

$$[X = t] = \{\omega : X(w) = t\}$$

Analog

$$[s < X \leq t] = \{\omega : s < X(\omega) \leq t\}$$

usw.

Beispiel:
$\Omega_0 = \{1, \ldots, 6\}$, $\Omega = \Omega_0^3$,

$$\begin{aligned} P((\omega_1, \omega_2, \omega_3)) &= \frac{1}{6^3} \\ &= \frac{1}{216} \\ &= P(\omega_1) \cdot P(\omega_2) \cdot P(\omega_3) \end{aligned}$$

$X_1(\omega) = \omega_1$, $X_2(\omega) = \omega_2$, $X_3(\omega) = \omega_3$ sind unabhängig, denn z.B.

$$\begin{aligned} P([X_1 = 6] \cap [X_2 = 3] \cap [X_3 = 5]) &= P((6, 3, 5)) \\ &= \frac{1}{6^3} \\ &= P([X_1 = 6]) \cdot P([X_2 = 3]) \cdot P([X_3 = 5]) \end{aligned}$$

Beweis:
Offensichtlich.

3.8. Beispiel:

$\Omega_0 = \{0, 1\}$, $0 < p < 1$, $q = 1 - p$,
$\Omega = \Omega_0^n$, $X_j(\omega) = \omega_j$, $S_n(\omega) = \sum_{j=1}^n \omega_j = \sum_{j=1}^n X_j(\omega)$

$$\begin{aligned} P(\omega) &= P(\omega_1) \cdot \ldots \cdot P(\omega_n) \\ &= p^{S_n(\omega)} \cdot q^{n - S_n(\omega)} \end{aligned}$$

$\Rightarrow X_1, \ldots, X_n$ sind stochastisch unabhängig.
Denn

$$\begin{aligned} P(\underbrace{[X_1 = \omega_1] \cap \ldots \cap [X_n = \omega_n]}_{=\{(\omega_1,\ldots,\omega_n)\}}) &= P((\omega_1, \ldots, \omega_n)) \\ &= P(\omega_1) \cdot \ldots \cdot P(\omega_n) \\ &= P([X_1 = \omega_1]) \cdot \ldots \cdot P([X_n = \omega_n]) \end{aligned}$$

3.9. Das Gesetz der seltenen Ereignisse

$\Omega_0 = \{0, 1\}$, $\Omega = \Omega_0^n$,
$S_n(\omega) = \sum_{j=1}^n \omega_j$,
Verteilung S_n: $B(n, p_n)$:

$$P([S_n = k]) = \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k}$$

p_n so gewählt, dass

$$\begin{aligned} \lim_{n\to\infty} n \cdot p_n &= \lim_{n\to\infty} \mathbb{E}(S_n) \\ &= \lambda \end{aligned}$$

wobei $n \cdot p_n = \mathbb{S}_{\Bbbk}$.

Satz

$$\begin{aligned} \lim_{n\to\infty} P([S_n = k]) &= e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= P_{Poission(\lambda)}(k) \end{aligned}$$

Beweis:
Buch WHK, S.491.

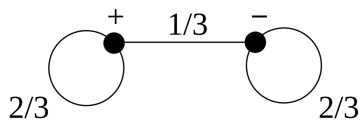
Teil II.

Markoff-Ketten auf endlichen Zustandsräumen

4. Definition und einfache Eigenschaften

4.1. Beispiele

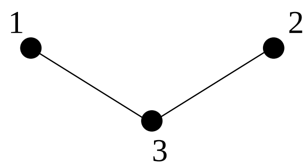
a) System mit zwei Zuständen: + und −



$$\begin{aligned} P(X_1 = +|X_0 = +) &= \frac{2}{3} \\ P(X_1 = -|X_0 = +) &= \frac{1}{3} \\ P(X_1 = +|X_0 = -) &= \frac{1}{3} \\ P(X_1 = -|X_0 = -) &= \frac{2}{3} \end{aligned}$$

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

b) Baum mit 3 Ecken:



Zustände: Die Knoten 1, 2, 3

$$
\begin{aligned}
P(X_1 = 1 | X_0 = 1) &= 0 \\
P(X_1 = 2 | X_0 = 1) &= 0 \\
P(X_1 = 3 | X_0 = 1) &= 0 \\
P(X_1 = 1 | X_0 = 2) &= 1 \\
P(X_1 = 2 | X_0 = 2) &= 0 \\
P(X_1 = 3 | X_0 = 2) &= 1 \\
P(X_1 = 1 | X_0 = 3) &= \frac{1}{2} \\
P(X_1 = 2 | X_0 = 3) &= \frac{1}{2} \\
P(X_1 = 3 | X_0 = 3) &= 0
\end{aligned}
$$

$$
P = \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ 1 & 1 & 0 \end{pmatrix}
$$

4.2. Präzisierung und Verallgemeinerung

Gegeben:

(i) $\mathcal{Z} = \{z_1, \ldots, z_r\}$

(ii) $X_0, X_1, X_2, \ldots : \Omega \to \mathcal{Z}$ mit:
für jedes n ist

$$
\Omega = \biguplus_{k=1}^{r} X_n^{-1}(z_k)
$$

Dann folgt

$$
\sum_{k=1}^{r} P([X_n = z_k]) = 1
$$

Für jedes n hat man dann eine Warscheinlichkeitsverteilung P_{X_n} auf $\mathcal{Z}$, gegeben durch

$$
\begin{aligned}
P_{X_n}(z_k) &= P([X_n = z_k]) \\
&= P(\{\omega : X_n(\omega) = z_k\})
\end{aligned}
$$

Zwei Anforderungen:

(I)

$$P(X_{n+1} = z_k | X_0 = z_{k_0}, X_1 = z_{k_1}, \ldots, X_n = z_{k_n}) \quad = \quad P(X_{n+1} = z_k | X_n = z_{k_n})$$

(Abhängigkeit von der Gegenwart *allein*)

(II)

$$P(X_{n+1} = z_k | X_n = z_j) \quad = \quad P(X_1 = z_k | X_0 = z_j)$$

(Zeitliche Konstistenz der Übergangswarscheinlichkeiten)

4.3. Definition

Eine Folge $(X_n)_{n \geq 0}$ von Zufallsvariablen X_n mit Werten im Zustandsraum $\mathcal{Z} = \{z_1, \ldots, z_r\}$ mit den Eigenschaften (I) und (II) heißt *Markoff-Kette auf* $\mathcal{Z}$.

Die Matrix P der Übergangswarscheinlichkeiten

$$p_{ij} \quad := \quad P(X_1 = z_i | X_0 = z_j)$$

heißt *Übergangsmatrix*.

Zu untersuchende Probleme:
Problem 1:
Gibt es eine Anfangsverteilung

$$\begin{aligned} p^{(0)} &= \begin{pmatrix} p_1^{(0)} \\ \vdots \\ p_r^{(0)} \end{pmatrix} \\ &= \begin{pmatrix} P(X_0 = z_1) \\ \vdots \\ P(X_0 = z_r) \end{pmatrix} \end{aligned}$$

so dass für alle n

$$\begin{aligned} P_{X_n} &= \begin{pmatrix} P(X_n = z_1) \\ \vdots \\ P(X_n = z_r) \end{pmatrix} \\ &= p^{(0)} \end{aligned}$$

Dann heißt die Markoff-Kette *stationär*.

Problem 2:
Gibt es eine Warscheinlichkeitsverteilung $p^{(\infty)}$ auf $\mathcal{Z}$ mit

$$\begin{aligned} \lim_{n \to \infty} P_{X_n} &= \lim_{n \to \infty} \begin{pmatrix} P(X_n = z_1) \\ \vdots \\ P(X_n = z_r) \end{pmatrix} \\ &= p^{(\infty)} \end{aligned}$$

4.4. Theorem

Sei $p^{(0)}$ eine Anfangsverteilung und P Übergangsmatrix. Dann ist

$$\begin{aligned} P_{X_n} &= P^n \cdot p^{(0)} \\ &= \underbrace{P \cdot \ldots \cdot P}_{n\text{-mal}} \cdot p^{(0)} \end{aligned}$$

Beweis:
Beweis durch Induktion über n:

$n = 1$: Es ist

$$\begin{aligned} P([X_1 = z_j]) &= \sum_{k=1}^{r} P([X_1 = z_j] \cap [X_0 = z_k]) \\ &= \sum_{k=1}^{r} P([X_1 = z_j | X_0 = z_k]) \cdot P([X_0 = z_k]) \\ &= Pp^{(0)} \end{aligned}$$

Angenommen es gilt $P([X_n = z_j]) = (P^n p^{(0)})_j$. Dann ist

$$\begin{aligned} P([X_{n+1} = z_j]) &= \sum_{k=1}^{r} P([X_{n+1} = z_j] \cap [X_n = z_k]) \\ &= \sum_{k=1}^{r} P([X_{n+1} = z_j | X_n = z_k]) \cdot P([X_n = z_k]) \\ &\underset{\text{Eigenschaft (II)}}{=} \sum_{k=1}^{r} P([X_1 = z_j]|[X_0 = z_k]) \cdot P([X_n = z_k]) \\ &\underset{\text{IV}}{=} (P \cdot P^n p^{(0)})_j \\ &= (P^{n+1} p^{(0)})_j \end{aligned}$$

Damit ist das Theorem bewiesen.

Beispiel:

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

$$\begin{aligned} p^{(0)} &= \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \\ &= \frac{1}{2} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} P(X_0 = 1) \\ P(X_0 = 2) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} P \cdot p^{(0)} &= \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \\ &= p^{(0)} \end{aligned}$$

$\Rightarrow$

$$P^n \cdot p^{(0)} = p^{(0)}$$

$\Rightarrow$

$$P_{X_n} = p^{(0)}$$

4.5. Korollar

Es mögen die Potenzen P^n konvergieren

$$\lim_{n\to\infty} P^n = Q$$

Dann gilt für jede Anfangsverteilung $p^{(0)}$ stets

$$\begin{aligned} \lim_{n\to\infty} P^n \cdot p^{(0)} &= \lim P_{X_n} \\ &= Q \cdot p^{(0)} \end{aligned}$$

4.6. Jukes-Cantor Modell

$$P = \begin{pmatrix} 1-3\cdot\alpha & \alpha & \alpha & \alpha \\ \alpha & 1-3\cdot\alpha & \alpha & \alpha \\ \alpha & \alpha & 1-3\cdot\alpha & \alpha \\ \alpha & \alpha & \alpha & 1-3\cdot\alpha \end{pmatrix}$$

$0 < \alpha \leq \frac{1}{3}$

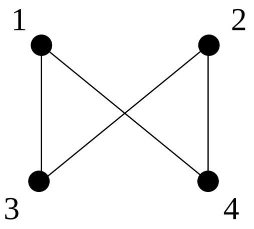
$$\begin{aligned} P \cdot \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix} &= \frac{1}{4} \cdot P \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix} \end{aligned}$$

$\overset{!}{\Rightarrow}$

$$\lim_{n\to\infty} P^n = \frac{1}{4}\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\end{pmatrix}$$

4.7. Beispiel

Graph mit 4 Ecken.



$$P^{2n+1} = \begin{pmatrix}0 & 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\end{pmatrix}$$

$$P^{2n} = \begin{pmatrix}\frac{1}{2} & \frac{1}{2} & 0 & 0\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\end{pmatrix}$$

Dies ist ein Beispiel, bei dem die Potenzen $(P^n)_n$ *nicht* konvergieren.

5. Stochastische Matrizen und Konvergenzsätze

5.1. Definition

Seien

$$\begin{aligned} A &= (a_{ij})_{i=1,\dots,r\ j=1,\dots,s} \\ B &= (b_{ij})_{i=1,\dots,r\ j=1,\dots,s} \end{aligned}$$

zwei $r \times s$ Matrizen reeller Zahlen. Wir schreiben $A \leq B$, falls für alle Indizes $a_{ij} \leq b_{ij}$ gilt.

5.2. Satz

Seien A, B, C $r \times s$-Matrizen. Es gilt

a) Ist $A \leq B$, so gilt

$$A + C \leq B + C$$

für alle $r \times s$-Matrizen C.

b) Ist D eine $s \times t$-Matrix mit $D \geq 0$ und ist $A \leq B$, so gilt

$$A \cdot D \leq B \cdot D$$

5.3. Definition

A heißt *positiv* (*nicht negativ*), falls alle $a_{ij} \geq 0$. Schreibweise: $A \geq 0$.
A heißt *strikt positiv*, falls alle $a_{ij} > 0$. Schreibweise: $A \gg 0$.

Beispiel:
Matrix

$$A = \frac{1}{r} \cdot \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix}$$

ist strikt positiv für alle $r \in \mathbb{R}^+$.

Matrix

$$B = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}$$

ist strikt positiv.

5.4. Satz

a) Sei A strikt positive $r \times s$-Matrix und $B \geq 0$ $s \times t$-Matrix, in der alle Spalten ungleich Nullvektor sind, so ist AB strikt positiv.

b) Sei A strikt positiv und $B \geq C$.
Ist

$$AB = AC$$

so ist

$$B = C$$

Beweis:

a) Durch einfaches Ausrechnen. Ist nämlich $AB = (c_{ij})$ so ist $c_{ij} = \sum_{k=1}^{s} a_{ik} b_{kj} > 0$ weil mindestens ein Summand > 0 ist.

b) Es ist $D := B - C \geq 0$ Wäre $D \neq 0$, so gäbe es eine Spalte $d_j^{\downarrow} \neq 0$, also wäre nach a) $Ad_j^{\downarrow} \gg 0$ und damit $AD = A(B-C) \neq 0$ im Gegensatz zur Voraussetzung $AB = AC$, also $AB - AC = A(B - C) = 0$.

5.5. Satz

Sei $P = (p_{ij})_{i,j=1,\dots,r}$ Übergangsmatrix.

Es gilt

$$p_{1k} + p_{2k} + \ldots + p_{rk} = 1$$

für alle $k = 1, \ldots, r$

Beweis:
Zunächst ist $p_{jk} = P([X_1 = j] | [X_0 = k])$. Nach Voraussetzung über Markoffsche Ketten ist $\Omega = \biguplus_{j=1}^{r} [X_1 = j]$. Also ist

$$[X_0 = k] = \biguplus_{j=1}^{r} [X_0 = k] \cap [X_1 = j]$$

Damit ist

$$P([X_0 = k]) = \sum_{k=1}^{r} P([X_1 = j] \cap [X_0 = k])$$

Division durch $P([X_0 = k])$ liefert

$$\begin{aligned} 1 &= \sum_{j=1}^{r} \frac{P([X_1 = j] \cap [X_0 = k])}{P([X_0 = k])} \\ &= \sum_{j=1}^{r} P([X_1 = j] | [X_0 = k]) \\ &= \sum_{j=1}^{r} p_{jk} \end{aligned}$$

Definition:
Eine quadratische Matrix $P \geq 0$ in der alle Spaltensumen gleich 1 sind, heißt *stochastisch*.
$\Rightarrow$

$$P^t \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$\Rightarrow$
1 ist Eigenwert für jede stochastische Matrix.

5.6. Definition

a) Eine stochastische Matrix P heißt *primitiv*, wenn es n gibt, so dass

$$P^n$$

strikt positiv ist.

b) Sie heißt *irreduzibel*, wenn es ein n gibt, so dass

$$\sum_{j=1}^{n} P^j$$

stritkt positiv ist.

Es gilt: primitiv $\Rightarrow$ irreduzibel

5.7. Theorem

Sei S eine stochastische irreduzible $r \times r$-Matrix. Dann gilt

a) 1 ist Eigenwert von S und der zugehörige Eigenraum $F(S) = \{x : Sx = x\}$ ist eindimensional und enthält genau eine Warscheinlichkeitsverteilung

$$\begin{aligned} p &= \begin{pmatrix} p_1 \\ \vdots \\ p_r \end{pmatrix} \\ F(S) &= \mathbb{R} \cdot p \\ &= \{\alpha \cdot p : \alpha \in \mathbb{R}\} \end{aligned}$$

b)

$$Q = \underbrace{(p, \dots, p)}_{r \text{ Spalten}}$$

ist eine Projektion auf den Eigenraum $F(S)$ mit

$$\begin{aligned} SQ &= QS \\ &= Q \end{aligned}$$

Q ist stochastisch.

c) Es gilt

$$\mathbb{R}^r = F(S) \oplus (I - S)\mathbb{R}^r$$

wobei

$$(I - S)\mathbb{R}^r = \{(I - S)y : y \in \mathbb{R}^r\}$$

und es gilt

$$S(I - S)\mathbb{R}^r \subset (I - S)\mathbb{R}^r$$

Beweis:
Bezeichnungen:

Für einen Vektor $x = \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}$ sei x_j die j. Koordinate und $|x| = \begin{pmatrix} |x_1| \\ \vdots \\ |x_r| \end{pmatrix}$.

a) (I) Behauptung: Sei $0 < p \leq Sp$. Dann gilt $p = Sp$ und p ist strikt positiv.
Beweis: Der Zeilenvektor $\vec{1} = (1, 1, \dots, 1)$ ist strikt positiv und es gilt $\vec{1}S = \vec{1}$, das bedeutet ja gerade, dass S stochastisch ist. Damit folgt

$$\begin{aligned} \vec{1}(Sp) &= (\vec{1}S)p \\ &= \vec{1}p \end{aligned}$$

also $Sp = p$ nach 5.4 b). Da S irreduzibel ist, gibt es ein n, so dass $T := \sum_{k=1} S^k$ strikt positiv ist. Wegen $0 < p$ ist dann

$$\begin{aligned} Tp &= \sum_{k=1}^{n} \underbrace{S^k p}_{=p} \\ &= np \end{aligned}$$

strikt positiv nach 5.4 a).

(II) Behauptung: Es gilt stets $|Sx| \leq S|x|$.
Beweis: Wir betrachten die j. Koordinate. Es ist

$$\begin{aligned} |Sx|_j &= |\sum_{k=1}^{r} s_{jk} x_k| \\ &\underset{\text{Dreiecksugl.}}{\leq} \sum_{k=1}^{r} |s_{jk} x_k| \\ &\underset{s_{jk} \geq 0}{=} \sum_{k=1}^{r} s_{jk} |x_k| = (S|x|)_j \end{aligned}$$

(III) Behauptung: Ist $Sx = x$, so ist $S|x| = |x|$, insbesondere gibt es zu 1 einen strikt positiven Eigenvektor.
Beweis: Aus $Sx = x$ folgt nach (II) $|x| = |Sx| \leq S|x|$ und damit nach (I) die Behauptung $S|x| = |x|$. Nach Voraussetzung ist S stochastisch also 1 ein Eigenwert ($\vec{1}S = \vec{1}$), also gibt es ein $x \neq 0$ mit $Sx = x$, also $S|x| = |x|$ und dies $|x|$ muss nach (I) strikt positiv sein.

(IV) Behautpung: Seien $p, q > 0$ mit $Sp = p$, $Sq = q$. Dann gibt es ein $\lambda > 0$ mit $q = \lambda p$.
Beweis: p ist (wie q) strikt positiv nach (I). Also ist die Größe $\lambda := \max\{\frac{q_j}{p_j} : 1 \leq j \leq r\}$ wohl definiert (im Nenner steht niemals 0), und es gibt (mindestens) einen Index j_0 mit $\lambda = \frac{q_{j_0}}{p_{j_0}}$. Die j. Koordinate von λp ist $\lambda p_j \geq \frac{q_j p_j}{p_j} = q_j$, als ist $\lambda p \geq q$, und für j_0 gilt $\lambda p_{j_0} = q_{j_0}$. Damit ist $0 \leq v := \lambda p - q$ *nicht* strikt positiv, aber es gilt

$$\begin{aligned} v &= \lambda Sp - Sq \\ &= \lambda p - q = v \end{aligned}$$

Aus (I) folgt $v = 0$.

(V) Sei q ein nach (III) existierender strikt positiver Eigenvektor. Wir setzen $p = \frac{q}{\sum_{j=1}^r q_j}$ und erhalten $\sum_{j=1}^r p_j = 1$, p ist also eine strikt positive Wahrscheinlichkeitsverteilung mit $Sp = p$.
Sei $Sv = v \neq 0$ ein beliebiger Eigenvektor von S zum Eigenwert 1.
Behauptung: Es gibt $\mu \neq 0$ mit $v = \mu p$.
Beweis: Nach (III) ist $S|v| = |v|$, also gilt nach (IV) $|v| = \lambda p$ für ein $\lambda > 0$. Sei $w = \lambda^{-1} v$. Dann ist $Sw = w$ und $|w| = p$. Wegen $|w| \geq w$ ist $p - w \geq p - |w| = 0$. Wir unterscheiden zwei Fälle:

a) Ist $p = \pm w$ so ist $v = \pm \lambda p$, die Behauptung also bewiesen.

b) Liegt a) nicht vor, so gibt es einen Index j mit $w_j = p_j$ und einen Index k mit $w_k = -p_k$. Dann gilt $(p-w)_j = 0$ und $(p-w)_k = 2p_k \neq 0$. Also ist $0 < S(p-w) = Sp - Sw = p - w$. Dann wäre aber $p - w$ nach (I) strikt positiv, im Widerspruch zu $(p-w)_j = 0$. Als kann b) nicht gelten, das heißt, es gilt a) und Teil a) des Theorems ist bewiesen.

b) Es ist

$$\begin{aligned} Qx &= \sum_{k=1}^r x_k p \\ &= \vec{1}x \cdot p \in F(S) \end{aligned}$$

Daraus folgt insbesondere $Qp = \vec{1}p \cdot p = p$, weil p eine Wahrscheinlichkeitsverteilung ist. Also erhält man $Q^2 = Q(p, \dots p) = (Qp, \dots, Qp) = Q$, Q ist also eine Projektion.
Es ist $SQ = (Sp, \dots, Sp) = (p, \dots, p) = Q$ und $QSx = \vec{1}Sx = \vec{1}x = Qx$, wegen $\vec{1}S = \vec{1}$, also $QS = Q$. Damit ist b) bewiesen.

c) Es ist $(I-Q)\mathbb{R}^r = \ker(Q)$. Da Q eine Projektion auf den Raum $F(S)$ ist, erhält man $\mathbb{R}^r = F(S) \oplus (I-Q)\mathbb{R}^r$. Aus $QS = Q$ folgt $Q(I-S) = 0$, also $(I-S)\mathbb{R}^r \subseteq \ker(Q)$. Andererseits ist $F(S) = \ker(I-S)$, also folgt aus der Dimensionsformel $r - \dim(F(S)) = \dim(I-S)\mathbb{R}^r$ und ebenso $r - \dim(F(S)) = \dim(\ker(Q))$. Damit ist $\dim(\ker(Q)) = \dim((I-S)\mathbb{R}^r)$ und damit folgt $(I-S)\mathbb{R}^r = \ker(Q)$ und c) ist bewiesen.

5.8. Korollar

Sei S irreduzibel stochastisch und

$$M_n = \frac{1}{n} \sum_{k=0}^{n-1} S^k$$

Es gilt

$$\lim_{n \to \infty} M_n = Q$$

D.h. ist $q = p^{(0)}$ eine Startwarscheinlichkeit, dann ist

$$\frac{1}{n} M_n \cdot q \to p$$

Es ist $S^k \cdot q = P_{X_k}$, also

$$M_n \cdot q \quad = \quad \text{arithmetisches Mittel der Verteilungen } P_{X_k} \ (P_{X_0} = q)$$

Beweis:

(I) Wir benutzen für die Konvergenz die 1-Norm:

$$\begin{aligned} \|x\| &= |x_1| + \dots + |x_r| \\ &= \vec{1}|x| \qquad \text{(Zeile mal Spalte)} \end{aligned}$$

Für sie gilt:

$$\begin{aligned} |y| &\leq |x| \\ &\Rightarrow \\ \|y\| &\leq \|x\| \end{aligned}$$

insbesondere gilt wegen $|Sx| \leq S|x|$

$$\begin{aligned} \|Sx\| &\leq \|S|x|\| \\ &= \vec{1}S|x| \\ &\underset{S \text{ stochastisch}}{=} \vec{1}|x| \\ &= \|x\| \end{aligned}$$

also auch

$$\|S^n x\| \leq \|x\|$$

(II) Wegen $SQ = Q$ ist $M_n Q = Q$, also

$$\begin{aligned} Qx &= M_n Qx \\ &= \lim_{k \to \infty} M_k Qx \end{aligned}$$

(III) Nach 5.7c) gibt es zu $(I - Q)x$ ein y mit $(I - Q)x = (I - S)y$. Nun ist $M_n(I - S) = \frac{1}{n}(I - S^n)$ (einfaches Ausrechnen), also ist

$$\begin{aligned} M_n(I - Q)x &= M_n(I - S)y \\ &= \frac{1}{n}(y - S^n y) \end{aligned}$$

und damit

$$\begin{aligned}\|M_n(I-Q)x\| &= \frac{1}{n}\|y - S^n y\| \\ &\leq \frac{1}{n}(\|y\| + \underbrace{\|S^n y\|}_{\leq\|y\|}) \\ &\leq \frac{2}{n}\|y\| \\ &\to 0 \qquad \text{für } n \to \infty\end{aligned}$$

(IV) Es ist $x = Qx + (I-Q)x$. Aus (II) und (III) folgt

$$\begin{aligned}\lim_{n\to\infty} M_n x &= Qx + 0 \\ &= Qx\end{aligned}$$

5.9. Theorem

Sei S eine primitive stochastische Matrix mit stationärer Verteilung p, d.h. es gilt

$$S \cdot p = p$$

Sei $Q : x \to (1|x) \cdot p$ die Projetkion auf $R \cdot p = F(S)$. Dann gilt für alle x

$$\lim_{n\to\infty} S^n \cdot x = Q \cdot x$$

Beweis:

(I) Wir haben $S^n Q = Q$, weil S als primitive Matrix ja irreduzibel ist. Also müssen wir nur $\lim_{n\to\infty} S^n(I-Q)x = 0$ zeigen.

(II) Zu x existiert nach 5.7c) ein y mit $(I-Q)x = (I-S)y =: u$.
Sei nun zunächst einmal $z := (I-S)v \in (I-S)\mathbb{R}^r$ beliebig. Aus Ungleichung (I) folgt

$$\begin{aligned}\|S^{n+1}z\| &= \|S(S^n z)\| \\ &\leq \|S^n z\|\end{aligned}$$

die Folge $(\|S^n z\|)_n$ ist also monoton fallend und durch 0 nach unten beschränkt, also konvergent. Wir müssen nur zeigen, dass sie gegen 0 konvergiert.

(III) Wir behandeln erst den Spezialfall, dass S strikt positiv ist. Dann ist

$$\begin{aligned}1 &> \min\{s_{ij} : i,j = 1\ldots r\} \\ &=: a \\ &> 0\end{aligned}$$

Sei $T = (1^{\downarrow}, \ldots, 1^{\downarrow})$. Dann ist $T \gg Q$, und nach Definition von a ist $S - aT \geq 0$, also $U := S - aQ \gg S - aT \geq 0$. $G = \frac{1}{1-a}U$ ist stochastisch (leichte Rechnung) und es gilt $S = (1-a)G + aQ$. Daraus ergibt sich wegen $Q(I-S) = 0$ und $z = (I-S)v$

$$\begin{aligned} S(z) &= (1-a)Gz + aQz \\ &= (1-a)Gz \end{aligned}$$

G ist stochastisch, also ist nach (I) $\|Gz\| \leq \|z\|$ und damit

$$\begin{aligned} \|Sz\| &= (1-a)\|Gz\| \\ &\leq (1-a)\|z\| \end{aligned}$$

Durch einfache Induktion folgt hieraus

$$\begin{aligned} \|S^{n+1}z\| &= \|S(S^n z)\| \\ &\leq (1-a)\|S^n z\| \\ &\leq \underset{\text{Induktion}}{\ldots} \\ &\leq (1-a)^{n+1}\|z\| \end{aligned}$$

also ist wegen $0 < 1-a < 1$ $\lim_{n\to\infty} \|S^n z\| = 0$. Insbesondere gilt dies für $z = (I-S)y = (I-Q)x$.

(IV) Sei nun S eine beliebige primitive Matrix. Dann gibt es ein k, so dass $S^k \gg 0$. Nach (III), angewandt auf S^k, ist dann $\lim_{n\to\infty} \|S^{kn}(I-Q)x\| = 0$, eine Teilfolge der monoton fallenden Folge $(\|S^n(I-Q)x\|)_n$ konvergiert also gegen 0. Damit konvergiert die Folge selbst gegen 0 und das Theorem ist bewiesen.

6. Anwendung auf Markoff-Ketten

Sei $P = (p_{ij})$ und $p_{ij} = P(X_1 = i|X_0 = j)$ die Übergangsmatrix einer Markoffkette mit Zustandsraum $Z = \{1, \ldots, r\}$.

$$\begin{aligned} P_{X_0} &= \begin{pmatrix} p_1 \\ \vdots \\ p_r \end{pmatrix} \\ &= p^{(0)} \end{aligned}$$

Dann

$$P_{X_n} = P^n \cdot p^{(0)}$$

Für

$$p^{(0)} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

ist

$$\begin{aligned} P^n \cdot p^{(0)} &= \begin{pmatrix} p_{11}^{(n)} \\ p_{21}^{(n)} \\ \vdots \\ p_{r1}^{(n)} \end{pmatrix} \\ &= \begin{pmatrix} P(X_n = 1) \\ \vdots \\ P(X_n = r) \end{pmatrix} \\ &= \begin{pmatrix} P(X_n = 1|X_0 = 1) \\ \vdots \\ P(X_n = r|X_0 = 1) \end{pmatrix} \end{aligned}$$

Allgemein

$$P^n = \begin{pmatrix} P(X_n = 1|X_0 = 1) & \ldots & P(X_n = 1|X_0 = r) \\ \vdots & & \vdots \\ P(X_n = r|X_0 = 1) & \ldots & P(X_n = r|X_0 = r) \end{pmatrix}$$

6.1. Irreduzibel und primitiv

Irreduzibel:
Zu je zwei Zuständen (i, j) gibt es ein n mit

$$\begin{aligned} P(X_n = i | X_0 = j) &= p_{ij}^{(n)} \\ &> 0 \end{aligned}$$

Primitiv:
Es gibt ein (gemeinsames) n, so dass für alle i, j gilt

$$\begin{aligned} P(X_n = i | X_0 = j) &= p_{ij}^{(n)} \\ &> 0 \end{aligned}$$

6.2. Satz

Es ist

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = P(X_0 = i_0) \cdot p_{i_1 i_0} \cdot p_{i_2 i_1} \cdot \ldots \cdot p_{i_n i_{n-1}}$$

Insbesondere:
Ist $P(X_0 = i_0) = 1$, so

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = p_{i_1 i_0} \cdot p_{i_2 i_1} \cdot \ldots \cdot p_{i_n i_{n-1}}$$

Beweis:
Nach Satz 3.3 erhalten wir

$$\begin{aligned} & P(X_0 = i_0, \cdots, X_n = i_n) \\ \underset{(1)}{=} \; & P(X_0 = i_0) P(X_1 = i_1 | X_0 = i_0) P(X_2 = i_2 | X_1 = i_1, X_0 = i_0) \\ & \cdots P(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \end{aligned}$$

Aufgrund der beiden Eigenschaften (I) und (II) für Markoffketten gilt nun aber

$$\begin{aligned} P(X_k = i_k | X_{k-1} = i_{k-1}, \ldots, X_0 = i_0) &= P(X_k = i_k | X_{k-1} = i_{k-1}) \\ &= P(X_1 = i_k | X_0 = i_{k-1}) \\ &= p_{i_k i_{k-1}} \end{aligned}$$

Einsetzen in (1) liefert die Behauptung.

6.3. Satz

Eine Markoffkette ist genau dann irreduzibel, wenn es zu jedem Paar (i, j) von Zuständen ein n gibt und eine Kette $(k_0, k_1, \ldots, k_n)$ von Zuständen mit $k_0 = j$, $k_n = i$ und

$$p_{k_{l+1}k_l} \neq 0 \qquad \text{für } l = 0, \ldots, n-1$$

Beweis:
Durch Induktion zeigt man

$$p_{ij}^{(n)} = \sum_{l_1=1}^{r} \sum_{l_2=1}^{r} \cdots \sum_{l_{n-1}=1}^{r} p_{il_1} p_{l_1 l_2} \cdots p_{l_{n-1}j}$$

Da alle $p_{mn} \geq 0$, ist diese Riesensumme genau dann $\neq 0$, wenn mindestens ein Summand $\neq 0$ ist. Nach 6.1 folgt der Satz.

Teil III.

Allgemeine Warscheinlichkeits-Theorie

7. Einführung

7.1. Definition

$\Omega = \langle a, b\rangle$, Ereignisse: Endliche Vereinigungen von Teilintervallen.
Sei $f : \Omega \to \mathbb{R}_+$ eine stetige Funktion mit

$$\begin{aligned}\int_a^b f(x)\,\mathrm{d}x &= \int_\Omega f(x)\,\mathrm{d}x \\ &= 1\end{aligned}$$

Dann wird durch

$$\begin{aligned}P : \langle u, v\rangle &\to [0,1] \\ P(\langle u, v\rangle) &= \int_u^v f(x)\,\mathrm{d}x\end{aligned}$$

und allgemein

$$P(A) = \int_A f(x)\,\mathrm{d}x$$

ein Warscheinlichkeits-Maß der Menge der endlichen Vereinigungen von Teilintervallen erklärt.
f heißt die *Dichte* zu P.

Beispiele:

a) $\Omega = \mathbb{R}$,

$$\begin{aligned}P(|x| > 1) &= P(]-\infty, -1[\cup]1, \infty[) \\ &= \int_{-\infty}^{-1} f(t)\,\mathrm{d}t + \int_1^\infty f(t)\,\mathrm{d}t\end{aligned}$$

b) $\Omega = [0,1]$, $f(t) = 1$

$$P([u,v]) = v - u$$

c) $\Omega = [a, b]$, $b > a$

$$f(t) = \frac{1}{b-a}$$

7.2. Beispiel

a) $\Omega = [a, b]$ und $f(x) = \frac{1}{b-a}$, $P(\langle u, v\rangle) = \frac{v-u}{b-a}$.
(Stetige) Gleichverteilung auf $[a, b]$.

$$P(\{u\}) = 0$$

Allgemein:

$$P([x, x + \mathrm{d}x]) = f(x)\,\mathrm{d}x$$

als Vorstellung.

b) Standard-Normalverteilung, $\Omega = \mathbb{R}$.

Gauß'sche Glockenkurve :

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}$$

Verteilungsfunktion:

$$\begin{aligned}
\phi(u) &= \int_{-\infty}^{u} f(x)\,\mathrm{d}x \\
&= P(]-\infty, u]) \\
&= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{u} e^{-\frac{x^2}{2}}\,\mathrm{d}x
\end{aligned}$$

Es gilt

$$f(x) = f(-x)$$

ferner

$$\begin{aligned}
\phi(0) &= \frac{1}{2} \\
\phi(u) &= 1 - \phi(-u)
\end{aligned}$$

$u < 0$:

$$\begin{aligned}
\phi(u) &= \int_{-\infty}^{u} f(x)\,\mathrm{d}x \\
\frac{1}{2} &= \phi(0) \\
&= \phi(u) + \int_{u}^{0} f(x)\,\mathrm{d}x \\
&= \phi(u) + \int_{0}^{-u} f(x)\,\mathrm{d}x \\
&= \phi(u) - \frac{1}{2} + \underbrace{\left(\int_{0}^{-u} f(x)\,\mathrm{d}x + \frac{1}{2}\right)}_{=\phi(-u)}
\end{aligned}$$

Standard-Normalverteilung $N(0,1)$ **:**

$$\begin{aligned}
\frac{1}{\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\,\mathrm{d}x &= 1\\
\frac{1}{\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{\infty} x\cdot e^{\frac{-x^2}{2}}\,\mathrm{d}x &= 0\\
&= \mathbb{E}(X)\\
\frac{1}{\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{\infty} x^2\cdot e^{\frac{-x^2}{2}}\,\mathrm{d}x &= 1\\
&= V(X)
\end{aligned}$$

Allgemeine Normalverteilung $N(\mu,\sigma)$ **:**

$$\begin{aligned}
f(x) &= \frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot e^{\frac{-(x-\mu)^2}{2\cdot\sigma^2}}\\
\mathbb{E}(X) &= \mu\\
V(X) &= \sigma^2\\
\mathbb{E}(X) &= \int_\Omega x\cdot f(x)\,\mathrm{d}x\\
V(X) &= \int_\Omega (x-\mathbb{E}(X))^2\cdot f(x)\,\mathrm{d}x
\end{aligned}$$

c) $\Omega=[0,\infty]=\mathbb{R}_+$,

$$\begin{aligned}
f(x) &= \lambda\cdot e^{-\lambda\cdot x}\\
P([T\leq t]) &= \lambda\cdot\int_0^t e^{-\lambda\cdot s}\,\mathrm{d}s
\end{aligned}$$

Fragen nach

$$\begin{aligned}
P([T<t+\mathrm{d}t]|[T\geq t]) &\underset{\text{Def. bed. Warscheinlichkeit}}{=} \frac{P([T<t+\mathrm{d}t]\cap[T\geq t])}{P([T\geq t])}\\
&= \frac{\int_t^{t+\mathrm{d}t} f(s)\,\mathrm{d}s}{P([T\geq t])}\\
&= \frac{f(t+\Theta t\mathrm{d}t)\cdot\mathrm{d}t}{P([T\geq t])} \qquad 0\leq\Theta\leq 1
\end{aligned}$$

Dann Division durch $\mathrm{d}t$:

$$\begin{aligned}
\frac{P([T<t+\mathrm{d}t]|[T\geq t])}{\mathrm{d}t} &= \frac{f(t+\Theta t\mathrm{d}t)}{P([T\geq t])}\\
&= \frac{f(t)}{P([T\geq t])}
\end{aligned}$$

7.3. Theorem (de Moivre-Laplace)

Sei $\Omega_0 = \{0,1\}$, $P(1) = p$, $0 < p < 1$, $P(0) = 1 - p$, $n \in \mathbb{N}$, $\Omega_n = \{0,1\}^n$.

$$\begin{aligned}
S_n(\omega) &= \sum_{k=1}^{n} \omega_k \\
P(\omega) &= p^{S_n(\omega)} \cdot (1-p)^{n-S_n(\omega)} \\
\mathbb{E}(S_n) &= n \cdot p \\
V(S_n) &= n \cdot p \cdot (1-p) \\
S_n^*(\omega) &= \frac{S_n(\omega) - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}
\end{aligned}$$

Es gilt

$$\begin{aligned}
\lim_{n\to\infty} P(\{\omega : a \leq S_n^*(\omega) \leq b\}) &= \frac{1}{\sqrt{2\pi}} \cdot \int_a^b e^{\frac{-x^2}{2}}\,\mathrm{d}x \\
&= \phi(b) - \phi(a)
\end{aligned}$$

Beispiel:
$n = 36$, $p = \frac{1}{2} = q$,

$$\begin{aligned}
n \cdot p &= 18 \\
\sqrt{n \cdot p \cdot q} &= 3
\end{aligned}$$

$$P(S_{36} \leq 15) = ?$$

$$\begin{aligned}
S_n \leq 15 &\Leftrightarrow S_n - n \cdot p \leq 15 - n \cdot p \\
&\Leftrightarrow S_n - n \cdot p \leq -3 \\
&\Leftrightarrow S_n^* = \frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot q}} \leq \frac{-3}{\sqrt{n \cdot p \cdot q}} \\
&\Leftrightarrow S_n^* = \frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot q}} \leq -1
\end{aligned}$$

$$\begin{aligned}
P(S_{36} \leq 15) &= P(S_{36}^* \leq -1) \\
&\underset{7.3}{\approx} \phi(-1) - \phi(-\infty) \\
&= \phi(-1) \\
&\underset{7.2\text{ b}}{=} 1 - \phi(1) \\
&\underset{\text{Tabelle}}{\approx} 1 - 0{,}84134 \\
&= 0{,}15866
\end{aligned}$$

8. Allgemeine Warscheinlichkeits-Räume

Einführung:
$\Omega = \mathbb{R}$,

$$\begin{aligned}
\varepsilon_0 &= \{A \subset \Omega : \ A \text{ ist endliche Vereinigung von Intervallen}\} \\
P(A) &= \int_A f(x)\,\mathrm{d}x \\
\{0\} &= \bigcap_{n=1}^{\infty} \left[-\frac{1}{n}, \frac{1}{n}\right]
\end{aligned}$$

8.1. Definition

a) $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt *Boolsche Unteralgebra*, falls gilt
   (i) $A \in \mathcal{A} \quad \Rightarrow \quad A^c = \Omega \setminus A \in \mathcal{A}$
   (ii) $A, B \in \mathcal{A} \quad \Rightarrow \quad A \cap B \in \mathcal{A}$
   (iii) $A, B \in \mathcal{A} \quad \Rightarrow \quad A \cup B \in \mathcal{A}$

b) $\varepsilon \subset \mathcal{P}(\Omega)$ heißt *σ-Algebra*, falls gilt
   (i) $\Omega \in \varepsilon$
   (ii) $A \in \varepsilon \quad \Rightarrow \quad A^c = \Omega \setminus A \in \varepsilon$
   (iii) $(A_n) \subset \varepsilon \quad \Rightarrow \quad (\bigcap_{n=1}^{\infty} A_n) \in \varepsilon$

   Daraus folgt unmittelbar mit De Morgan:

   $(A_n) \subset \varepsilon \quad \Rightarrow \quad (\bigcup_{n=1}^{\infty} A_n) \in \varepsilon$

8.2. Bemerkung

(a) Sei $\mathcal{A} \subset \mathcal{P}(\Omega)$ eine Boolsche Algebra, so gilt

$$\emptyset \in \mathcal{A}$$
$$\Omega \in \mathcal{A}$$

Beweis:

$$\begin{aligned} A \in \mathcal{A} &\underset{\text{(i)}}{\Rightarrow} A^c \in \mathcal{A} \\ &\underset{\text{(ii)}}{\Rightarrow} \emptyset = A \cap A^c \in \mathcal{A} \\ &\underset{\text{(i)}}{\Rightarrow} \emptyset^c = \Omega \in \mathcal{A} \end{aligned}$$

(b) Sei $(\mathcal{A}_\alpha)_\alpha$ Familie von σ-Algebren. Dann ist

$$\bigcap_\alpha \mathcal{A}_\alpha$$

wieder eine σ-Algebra.

(c) Anwendung von (2) :

Sei $\mathcal{F} \subset \mathcal{P}(\Omega)$,

$$\mathbb{A} \quad := \quad \{\mathcal{A} : \ \mathcal{A} \text{ ist } \sigma\text{-Algebra und } \mathcal{F} \in \mathcal{A}\}$$

Die Menge

$$\varepsilon(\mathcal{F}) \quad = \quad \bigcap_{\mathcal{A} \in \mathbb{A}} \mathcal{A}$$

heißt *die von $\mathcal{F}$ erzeugte σ-Algebra.*

8.3. Konkret

a) $\Omega = \mathbb{R}$, $\mathcal{F} =$ Menge aller Intervalle,
die Menge

$$\varepsilon(\mathcal{F}) = \mathcal{B}$$

heißt σ-Algebra der *Borelmengen.*

Jede abgeschlossene und jede offene Menge ist Borelmenge.

Bemerkung:

$$\int_{\mathcal{B}} f(x)\,\mathrm{d}x$$

erklärbar für alle Dichtefunktionen f und alle Borelmengen.

b) $\Omega = \mathbb{R}^r$, $\mathcal{F} =$ Menge aller r-dimensionalen Quader

$$\varepsilon(\mathcal{F}) = \mathcal{B}(\mathbb{R}^r)$$

heißt σ-Algebra der Borelmengen in $\mathbb{R}^r$.

Jede abgeschlossene, jede offene Menge ist Borelmenge.
Z.B. in $\mathbb{R}^2$ ist $G = \{(x,x) : x \in \mathbb{R}\}$ auch eine Borelmenge.

c) $\Omega = \{1,\dots,r\}^{\mathbb{N}}$, $Z \subset \Omega$ heißt *Zylindermenge*, wenn es ein n gibt und $A \subset \{1,\dots,r\}^n = \Omega_0^n$, so dass

$$Z = A \times \{1,\dots,r\}^{\mathbb{N}\setminus\{1,\dots,n\}}$$

Z.B.:

$$\begin{aligned} Z &= \{(1,1,x_3,x_4,x_5,\dots) : x_k \in \Omega_0\} \\ A &= \{(1,1)\} \\ &\subset \{1,\dots,r\}^2 \end{aligned}$$

$$\mathcal{Z} = \text{die von Zylindermengen erzeugte } \sigma\text{-Algebra}$$

8.4. ???

8.5. Definition

Sei $\emptyset \neq \Omega$, $\varepsilon \subset \mathcal{P}(\Omega)$ eine σ-Algebra und $P : \varepsilon \to [0,1]$ eine Funktion mit

(i) $P(\Omega) = 1$

(ii) Ist $(A_n) \subset \varepsilon$ eine Folge paarweise disjunkter Mengen, so ist

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Dann heißt (Ω, ε, P) *Warscheinlichkeits-Raum*.
ε heißt *Ereignisalgebra*.
P heißt *Warscheinlichkeits-Maß/-Verteilung*.

Beispiel:

(1) Ω endlich oder abzählbar, $\varepsilon = \mathcal{P}(\Omega)$, P wie bisher.

(2) $\Omega = [a,b]$, ε Menge der Borelmengen in Ω

$$P(B) = \frac{1}{b-a} \cdot \int_B 1 \,\mathrm{d}x$$

(3) $\Omega = \mathbb{R}$, ε: Borelmengenalgebra $\mathcal{B}$.

$$P(B) = \frac{1}{\sqrt{2 \cdot \pi}} \int_B e^{-\frac{x^2}{2}} \,\mathrm{d}x$$

(4) $\Omega = \mathbb{R}^r$, $\varepsilon = \mathcal{B}(\mathbb{R}^r)$,

$$\begin{aligned} f(x) &= (2\pi)^{-\frac{r}{2}} \cdot e^{-\frac{(x_1^2+\ldots+x_r^2)}{2}} \\ P(B) &= \int_B f(x) \,\mathrm{d}x \end{aligned}$$

(5) $\Omega = \{0,1\}^{\mathbb{N}}$, $\mathcal{Z}$ von Zylindermengen erzeugte σ-Algebra.
P auf den Zylindern

$$\begin{aligned} Z &= A \times \Omega^{\mathbb{N}\setminus\{1,\ldots,n\}} \\ P(Z) &= P_n(A) \end{aligned}$$

8.6. Satz (Einfache Eigenschaften)

a) $P(\emptyset) = 0$

b) $A \subseteq B \quad \Rightarrow \quad P(A) \leq P(B)$

c) $P(A^c) = 1 - P(A)$

d) (i) $(A_n)_{n\geq 1}$ Folge mit $A_n \subseteq A_{n+1}$.
Dann gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

(ii) $(A_n)_{n\geq 1}$ Folge mit $A_n \supseteq A_{n+1}$.
Dann gilt

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

Beweis:

a) $\Omega = \Omega \cup \emptyset$ ergibt nach 8.5 sofort

$$\begin{aligned} 1 &= P(\Omega \cup \emptyset) \\ &= P(\Omega) + P(\emptyset) \\ &= 1 + P(\emptyset) \end{aligned}$$

woraus die Behauptung folgt.

b) $A \subseteq B$ impliziert $B = A \uplus B \setminus A$, also

$$\begin{aligned} P(B) &= P(A) + P(B \setminus A) \\ &\geq P(A) \end{aligned}$$

c) Folgt wegen $\Omega = A \uplus A^c$, also $1 = P(\Omega) = P(A) + P(A^c)$.

d) (i) Sei $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $\ldots B_n = A_n \setminus A_{n-1}$. Dann ist $A_n = \biguplus_{k=1}^n B_k$ und $A = \bigcup A_n = \biguplus_{k=1}^\infty B_k$. Nach 8.5 (ii) ist also

$$P(A_n) = \sum_{k=1}^n P(B_k)$$

und

$$\begin{aligned} P(A) &= \sum_{k=1}^\infty P(B_k) \\ &= \lim_{n\to\infty} \sum_{k=1}^n P(B_k) \\ &= \lim_{n\to\infty} P(A_n) \end{aligned}$$

(ii) Sei $A = \bigcap_{n=1}^\infty A_n$. Es ist $A_n^c \subseteq A_{n+1}^c$, also nach (i) $P(A^c) = \lim_{n\to\infty} P(A_n^c)$. Mit c) folgt die Behauptung.

8.7. Beispiel

(i) $\Omega = \mathbb{R}$, $f : \Omega \to \mathbb{R}_+$ stetig mit

$$\int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1$$

$$P(\langle a, b\rangle) = \int_a^b f(t)\,\mathrm{d}t$$

Es gilt

$$P(\{t\}) = 0$$

Denn

$$\{t\} = \bigcap_{n=1}^\infty \underbrace{[t, t + \frac{1}{n}]}_{=A_n}$$

$$A_1 \supset A_2 \supset \ldots$$

Dann gilt nach 8.6:

$$\begin{aligned} P(\{t\}) &= \lim_{n\to\infty} P(A_n) \\ &= \lim_{n\to\infty} \int_t^{t+\frac{1}{n}} f(s)\,\mathrm{d}s \\ &= \lim_{n\to\infty} \frac{1}{n} \cdot f(S_n) \end{aligned}$$

Sei

$$\gamma = \sup\{f(s) : t \leq s \leq t+1\}$$

$\Rightarrow$

$$\begin{aligned} P(t) &\leq \lim_{n\to\infty} \frac{1}{n} \cdot \gamma \\ &= 0 \end{aligned}$$

(ii) $\Omega = \mathbb{R}^n$,

$$f(t_1,\ldots,t_n) = \frac{1}{(\sqrt{2\pi})^n} \cdot e^{-\frac{t_1^2+\ldots+t_n^2}{2}}$$

$$P([a_1,b_1]\times\ldots\times[a_n,b_n]) = \frac{1}{(\sqrt{2\pi})^n} \cdot \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} e^{-\frac{t_1^2+\ldots+t_n^2}{2}}\,\mathrm{d}t_1 \ldots \mathrm{d}t_n$$

8.8. Definition

Sei (Ω, ε, P) ein Warscheinlichkeits-Raum.
$X : \Omega \to \mathbb{R}$ heißt (reelle) Zufallsvariable, falls

$$X^{-1}(]a,b]) \in \varepsilon$$

für alle Intervalle $]a,b]$.

$$\begin{aligned} X^{-1}(]a,b]) &= \{\omega : a < X(\omega) \leq b\} \\ &= [a < X \leq b] \end{aligned}$$

8.9. Satz

a) X ist genau dann eine Zufallsvariable, wenn für jede Borelmenge $B \subseteq \mathbb{R}$ das Urbild $X^{-1}(B) \in \varepsilon$.

b) Summe, Produkt und Absolutbetrag von Zufallsvariablen, sowie $\max(X,Y)$ und $\min(X,Y)$ sind wieder Zufallsvariablen.

c) Sei (X_n) eine Folge von Zufallsvariablen und

$$X(\omega) \quad = \quad \lim_{n\to\infty} X_n(\omega)$$

für alle ω, so ist X Zufallsvariable.

8.10. Definition

Sei (Ω, ε, P) ein Warscheinlichkeits-Raum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.
Dann ist durch

$$P_X(B) \quad := \quad P(\{\omega : X(\omega) \in B\})$$

auf der σ-Algebra $\mathcal{B}(\mathbb{R})$ der Borelmengen ein Warscheinlichkeits-Maß gegeben, die Verteilung von X auf $\mathbb{R}$.

8.11. Bemerkung

a) Oft ist (Ω, ε, P) gar nicht wichtig, sondern nur P_X auf $\mathbb{R}$ (Normalverteilung, Exponentialverteilung usw.).

b) Sei $f : \mathbb{R}^r \to \mathbb{R}_+$ eine stetige Dichte und

$$P(B) \quad = \quad \int_B f(x)\,\mathrm{d}x$$

Dann ist

$$(\mathbb{R}^r, \mathcal{B}(\mathbb{R}^r), P)$$

ein Warscheinlichkeits-Raum und

$$X_j(x) = x_j$$

sind Zufallsvariablen.
$X(t) = t$ ist Zufallvariable.

8.12. Satz

Wesentliche Ergebnisse aus Teil 1 und 2 gelten auch im allgemeinen Warscheinlichkeits-Raum.

Bedingte Warscheinlichkeit:

$$P(A|B) \quad = \quad \begin{cases} 0 & P(B) = 0 \\ \frac{P(A\cap B)}{P(B)} & P(B) \neq 0 \end{cases}$$

Es gilt

$$P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_n)$$

Es gilt auch das Theorem von Bayes:

$A_1, \ldots, A_n$ paarweise disjunkte Ereignisse und $\bigcup_{k=1}^n A_k = \Omega$. Sei B ein Ereignis, so gilt für alle $k = 1, \ldots, n$:

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{P(A_1) \cdot P(B|A_1) + \ldots + P(A_n) \cdot P(B|A_n)}$$

Unabhängigkeit:

$A_2, A_2 \in \varepsilon$ heißen stochastisch unabhängig, wenn

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

$A_1, \ldots, A_n$ heißen stochastisch unabhängig, wenn für alle $k \leq n$ und Indizes $i_1 < i_2 < \ldots < i_k \leq n$ stets

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j})$$

Zufallsvariablen $X_1, \ldots, X_n$ heißen stochastisch unabhängig, wenn für alle Intervalle $J_1, \ldots, J_n$ die Urbilder $X_1^{-1}(J_1), \ldots, X_n^{-1}(J_n)$ stochastisch unabhängig sind.

$(A_n)_{n \geq 1}$ heißt stochastisch unabhängig, wenn jede endliche Teilfolge stochastisch unabhängig ist.
Nicht übertragbar auf den allgemeinen Fall:
$X, Y : \Omega \to \mathbb{R}$ stochastisch unabhängig, wenn $[X = a]$, $[Y = b]$ $(a, b \in \mathbb{R})$.

Beispiel

1) $\Omega = [0,1]^2$, Dichte $f(x,y) = 1$, $X_j((x_1, x_2)) = x_j$

2) $\Omega = \mathbb{R}^2$, $f(x,y) = g(x) \cdot h(y)$, g, h stetige Dichten auf $\mathbb{R}$.

$$\begin{aligned} X_j(x_1, x_2) &= x_j \\ J_1 &= [0,1] \\ J_2 &= [4,5] \\ X_1^{-1}(J_1) &= [0,1] \times \mathbb{R} \\ X_2^{-1}(J_2) &= \mathbb{R} \times [4,5] \\ X_1^{-1}(J_1) \cap X_2^{-1}(J_2) &= [0,1] \times [4,5] \end{aligned}$$

$$\begin{aligned} P([0,1]\times[4,5]) &= \int_0^1\int_4^5 f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^1\int_4^5 g(x)\cdot h(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^1 g(x)\,\mathrm{d}x\cdot\int_4^5 h(y)\,\mathrm{d}y \\ &= P\left(X_1^{-1}(J_1)\right)\cdot P\left(X_2^{-1}(J_2)\right) \end{aligned}$$

8.13. Definition (Erwartungswert)

(Ω,ε,P) Warscheinlichkeits-Raum.

a)

$$X = \sum_{j=1}^n \alpha_j\cdot 1_{A_j}$$

mit $\alpha_j\in\mathbb{R}$ und $A_j\in\varepsilon$ heißt *elementare Zufallsvariable*.

$$\mathbb{E}(X) = \sum_{j=1}^n \alpha_j\cdot P(A_j)$$

heißt Erwartungswert von X.

b) $X:\Omega\to\mathbb{R}_+$ sei Zufallsvariable.
Erwartungswert

$$\mathbb{E}(X) = \sup\{\mathbb{E}(Y): 0\le Y\le X \text{ und } Y \text{ ist elementar}\}$$

Wenn $\mathbb{E}(X)<\infty$, dann heißt X integrierbar.

c) Sei $X:\Omega\to\mathbb{R}$,

$$\begin{aligned} X^+(\omega) &= \max\{X(\omega),0\} \\ X^-(\omega) &= \max\{-X(\omega),0\} \end{aligned}$$

$$\begin{aligned} X^+ - X^- &= X \\ X^+ - X^- &= |X| \end{aligned}$$

X heißt integrierbar, wenn X^+ und X^- integrierbar sind,

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

8.14. Beispiel

1) Ω endlich, dann nichts neues.

2) Ω abzählbar, dann alt = neu, siehe Übungsaufgabe 36.

3) $\Omega = \langle a, b\rangle$, $f : \Omega \to \mathbb{R}$ stetige Dichte, P üblich

$$X(t) \quad = \quad t$$

$$X \text{ integrierbar} \quad \Leftrightarrow \quad \int_a^b |t| \cdot f(t)\,\mathrm{d}t \text{ existiert}$$

und dann

$$\mathbb{E}(X) \quad = \quad \int_a^b t \cdot f(t)\,\mathrm{d}t$$

8.15. Satz (Eigenschaften des Erwartungswertes)

a) X und Y integrierbar, $\alpha, \beta \in \mathbb{R}$
$\Rightarrow$
$\alpha \cdot X + \beta \cdot Y$ integrierbar und

$$\mathbb{E}(\alpha X + \beta Y) \quad = \quad \alpha \cdot \mathbb{E}(X) + \beta \cdot \mathbb{E}(X)$$

b) X, Y integrierbar und $X \leq Y$
$\Rightarrow$

$$\mathbb{E}(X) \quad \leq \quad \mathbb{E}(Y)$$

Insbesondere

$$|\mathbb{E}(X)| \quad \leq \quad \mathbb{E}(|X|)$$

c) X integrierbar, Y Zufallsvariable mit

$$|Y| \quad \leq \quad |X|$$

$\Rightarrow$
Y integrierbar und

$$|\mathbb{E}(Y)| \quad \leq \quad \mathbb{E}(|X|)$$

8.16. Theorem

a) **Konvergenzsatz von Levi:**

Sei (X_n) eine monotone Folge integrierbarer Funktionen und die Folge $(\mathbb{E}(X_n))$ sei beschränkt.
Sei

$$X(\omega) \quad = \quad \lim_{n\to\infty} X_n(\omega) \quad \in \mathbb{R} \quad \forall\omega$$

Dann ist X integrierbar und

$$\mathbb{E}(X) \quad = \quad \lim_{n\to\infty} \mathbb{E}(X_n)$$

b) **Konvergenzsatz von Lebesgue:**

Sei (X_n) eine monotone Folge integrierbarer Funktionen und

$$|X_n| \quad \leq \quad Y$$

wo Y integrierbar. Ferner sei

$$X(\omega) \quad = \quad \lim_{n\to\infty} X_n(\omega)$$

Dann ist X integrierbar und es gilt

$$\mathbb{E}(X) \quad = \quad \lim_{n\to\infty} \mathbb{E}(X_n)$$

Regel:

$$\mathbb{E}(\lim) \quad = \quad \lim \mathbb{E}$$

8.17. Satz und Definition

Sei $X : (\Omega, \varepsilon, P) \to \mathbb{R}$ eine Zufallsvariable.
X heißt *quadratisch integrierbar*, wenn X^2 integrierbar ist.
Dann ist auch X integrierbar. Das Integral

$$\mathbb{E}((X - \mathbb{E}(X))^2) \quad =: \quad V(X)$$

heißt *Varianz* von X.

$$\sigma(X) \quad := \quad \sqrt{V(X)}$$

heißt *Streuung*.

Es gilt

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

(Übung von diskreten zu allgemeinen Warscheinlichkeits-Räumen: Statt Summe Integral)

Beweis:

$$|t| \quad \leq \quad 1 + t^2$$

also

$$|X| \quad \leq \quad 1_\Omega + X^2$$

8.18. Definition und Satz

Seien X und Y quadratisch integrierbar.
Dann ist $X \cdot Y$ integrierbar.

Die Größe

$$\mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))) \quad = \quad \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

heißt *Covarianz* $C(X, Y)$ und

$$\frac{C(X, Y)}{\sigma(X) \cdot \sigma(Y)}$$

heißt *Korrelationskoeffizient*.

Ist dieser gleich 0, so heißen X und Y unkorreliert.

Beweis:
Es ist für 2 reelle Zahlen a, b stets

$$\begin{aligned} a^2 - 2ab + b^2 &= (a - b)^2 \\ &\geq 0 \end{aligned}$$

also $|ab| \leq 2|ab| \leq a^2 + b^2$. Damit ist $|XY| \leq X^2 + Y^2$. Da X^2 und Y^2 integrierbar sind, ist XY nach 8.15 c) integrierbar. Der Rest ist reine Rechnerei.

8.19. Satz

Sei $X : (\Omega, \varepsilon, P) \to \mathbb{R}$ eine Zufallsvariable und $P_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$ die Verteilung von X gegeben durch

$$P_X(B) \quad = \quad P(X^{-1}(B))$$

insbesondere

$$P_X(]a,b]) \quad = \quad P([a < X \leq b])$$

Dann gilt

$$\begin{aligned}
\mathbb{E}(X) &= \int_{\mathbb{R}} t \cdot P_X \,\mathrm{d}t \\
&= \lim_{n\to\infty} \sum_{k=2^{2n}}^{2^{2n}} \frac{k}{2^n} P_X\left(\left]\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right) \\
&= \lim_{n\to\infty} \sum_{k=2^{2n}}^{2^{2n}} \frac{k}{2^n} P\left(\left]\frac{k}{2^n} < X \leq \frac{k+1}{2^n}\right]\right)
\end{aligned}$$

$$\mathbb{E}(X^2) \quad = \quad \int_{\mathbb{R}} t^2 P_X \,\mathrm{d}t$$

$$\begin{aligned}
P_X(]a,b]) &= \int_a^b f(t)\,\mathrm{d}t \\
&\Rightarrow \\
\mathbb{E}(X) &= \int t \cdot f(t)\,\mathrm{d}t \\
\mathbb{E}(X^2) &= \int t^2 \cdot f(t)\,\mathrm{d}t
\end{aligned}$$

Beispiel:
Glechverteilung

$$f(t) \quad = \quad \begin{cases} 1 & 0 \leq t \leq 1 \\ 0 & \text{sonst} \end{cases}$$

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^1 t \cdot f(t)\,\mathrm{d}t \\
&= \frac{1}{2} \\
\mathbb{E}(X^2) &= \int_0^1 t^2 \cdot f(t)\,\mathrm{d}t \\
&= \frac{1}{3}
\end{aligned}$$

8.20. Satz

Seien X und Y Zufallsvariablen. Dann wird durch

$$P_{X,Y}(]a,b]\times]c,d]) \quad = \quad P([a < X \leq b] \cap [c < Y \leq d])$$

ein Warscheinlichkeitsmaß auf $\mathbb{R}^2$ definiert, die *gemeinsame* Verteilung von X und Y. Mit ihm ist

$$Co(X,Y) \quad = \quad \int\int (X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) P_{X,Y} \,\mathrm{d}x\,\mathrm{d}y$$

$P_{X,Y}$ gegeben durch $h(x,y)$ damit

$$Co(X,Y) \quad = \quad \int\int (X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) h(x,y) \,\mathrm{d}x\,\mathrm{d}y$$

9. Grenzwertsätze

9.1. Einführung

(Ω, ε, P) Warscheinlichkeits-Raum.

(I) Erinnerung: (A_n) Folge von Ereignissen.

Ist $A_n \subseteq A_{n+1}$ für alle n, so gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

Ist $A_{n+1} \subseteq A_n$ für alle n, so gilt

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

(II) $(A_n)_n$ Folge von Ereignissen.

a)

$$\bigcap_{n=1}^{\infty}\left(\bigcup_{k=n}^{\infty} A_k\right) =: C$$

$$\omega \in C \quad \Leftrightarrow \quad \forall n \exists k \geq n : \omega \in A_k$$

b)

$$\bigcup_{n=1}^{\infty}\left(\bigcap_{k=n}^{\infty} A_k\right) =: D$$

$$\omega \in D \quad \Leftrightarrow \quad \exists n \forall k \geq n : \omega \in A_k$$

Einschub:
Tastatur mit 50 Tasten. $\Omega_0 = \{a_1, \dots, a_{50}\}$, $\Omega = \Omega_0^{\mathbb{N}}$,

$$P_0(a_k) = \frac{1}{50}$$

$$P\left((\omega_1, \dots, \omega_n) \times \Omega^{\mathbb{N}\setminus\{1,\dots,n\}}\right) = \frac{1}{50^n}$$

$k \in \mathbb{N}$,
$\omega_k = b$, $\omega_{k+1} = a$, $\omega_{k+2} = n$, $\omega_{k+3} = a$, $\omega_{k+4} = n$, $\omega_{k+5} = e$.

$$\begin{aligned} B_k &= \{\omega : (\omega_k, \ldots, \omega_{k+5}) = (b, a, n, a, n, e)\} \\ A_n &= B_{6n+1} \\ C &= \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k \end{aligned}$$

9.2. Hilfssätze

a)

$$P\left(\bigcup_{j=1}^{n} C_j\right) \leq \sum_{j=1}^{n} P(C_j)$$

b) $0 \leq x \leq 1$, so $1 - x \leq e^{-x}$

Beweis:

a) Induktion über n.
Es ist

$$C_1 \cup C_2 = C_1 \uplus (C_2 \setminus C_1 \cap C_2)$$

also

$$\begin{aligned} P(C_1 \cup C_2) &= P(C_1) + \underbrace{P((C_2 \setminus C_1 \cap C_2))}_{\leq P(C_2)} \\ &\leq P(C_1) + P(C_2) \end{aligned}$$

Wende diesen Schluss nun an auf $\bigcup_{k=1}^{n} C_k \cup C_{n+1}$.

b) Die Funktion $f(x) = e^{-x} - (1-x)$ erfüllt $f(0) = 0$ und $f'(x) = -e^{-x} + 1 = 1 - \frac{1}{e^x} \geq 0$ auf $[0, 1]$. f ist dort also monoton wachsend, also wegen $f(0) = 0$ immer ≥ 0.

9.3. Lemma von Borel-Cantelli

Sei (A_n) eine beliebige Folge aus ε.

a) Ist

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

dann ist

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\right) = 0$$

b) Ist

$$\sum_{n=1}^{\infty} P(A_n) = \infty$$

und die A_n unabhängig, dann ist

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\right) = 1$$

Beweis:

a) Wir setzen $B_n := \bigcup_{k \geq n} A_k$ und erhalten $B := \bigcap_{n=1}^{\infty} B_n$, sowie $B_{n+1} \subseteq B_n$. Also gilt nach 8.6 d) (ii)

$$P(B) = \lim_{n \to \infty} P(B_n)$$

Aber

$$P(B_n) \leq \sum_{k=n}^{\infty} P(A_k)$$

nach dem Hilfssatz, Teil a). Da die Summe $\sum_{k=1}^{\infty} P(A_k)$ konvergiert, ist die Folge $(\sum_{k=n}^{\infty} P(A_k))_n$ eine Nullfolge, und die Behauptung folgt.

b) Wir zeigen (mit den Bezeichnungen von Teil a) des Beweises) $P(B^c) = 0$. Wegen $B_{n+1}^c \supseteq B_n^c$ ist $P(B^c) = \lim_{n \to \infty} P(B_n^c)$ (s. 8.6.d) (i)). Wieder mit 8.6 d) erhalten wir (unter Anwendung der deMorganschen Regeln für das Berechnen des Komplements der Vereinigung und des Durchschnitts)

$$\begin{aligned} P(B_n^c) &= P(\bigcap_{k=n}^{\infty} A_k^c) \\ &= \lim_{l \to \infty} P(\bigcap_{k=n}^{n+l} A_k^c) \end{aligned}$$

Nun sind die A_k stochastisch unabhängig, also auch die A_k^c. Damit gilt

$$\begin{aligned} P(\bigcap_{k=n}^{n+l} A_k^c) &= \prod_{k=n}^{l+n} P(A_k^c) \\ &= \prod_{k=n}^{n+l} (1 - P(A_k)) \\ &\leq \prod_{k=n}^{n+l} e^{-P(A_k)} \\ &= e^{-\sum_{k=n}^{n+l} P(A_k)} \end{aligned}$$

Da die Reihe $\sum_{k=1}^{\infty} P(A_k)$ divergiert, gilt auch für alle n stets $\lim_{l\to\infty} \sum_{k=n}^{l+n} P(A_k) = \infty$, also $\lim_{l\to\infty} e^{-\sum_{k=n}^{l+n} P(A_k)} = 0$. Damit ist $P(B_n^c) = 0$, und daraus folgt die Behauptung.

9.4. Beispiel

Aufgriff des vorherigen Beispiels mit der Tastatur.

$$\begin{aligned} A_n &= B_{6n+1} \\ P(A_n) &= \frac{1}{50^6} \\ \sum_{n=1}^{\infty} P(A_n) &= \sum_{n=1}^{\infty} \frac{1}{50^6} \\ &= \infty \end{aligned}$$

Nach dem Satz 9.3 b) ist also $P(C) = 1$, das bedeutet: Mit Wahrscheinlichkeit 1 tritt das Wort *Banane* unendlich oft auf. Analog behandelt man das Beispiel, dass die Bibel mit Wahrscheinlichkeit 1 unendlich oft auftritt, wenn man zufällig auf der Tastatur eines PC umhertippt.

9.5. Satz (Ungleichung von Kolmogorow)

Seien $Z_1, \dots, Z_n$ unabhängige Zufallsvariablen mit $\mathbb{E}(Z_k) = 0$, $V(Z_k) < \infty$.
Sei

$$\begin{aligned} S_k(\omega) &= \sum_{j=1}^{k} Z_j(\omega) \\ Y_n(\omega) &= \max\{|S_k(\omega)| : 1 \leq k \leq n\} \end{aligned}$$

$\eta > 0$ beliebig.

$$\begin{aligned} P(\{\omega : Y(\omega) \geq \eta\}) &\leq \frac{\sum_{j=1}^{n} V(Z_j)}{\eta^2} \\ &= \frac{1}{\eta^2} V(S_n) \end{aligned}$$

Beweis:
Sei

$$A_1 = \{\omega : |S_1(\omega)| \geq \eta\}$$

und für $k > 1$ sei

$$A_k = \{\omega : |S_l(\omega)| < \eta,\ \text{für } l < k, |S_k(\omega)| \geq \eta\}$$

Sei $A = [Y \geq \eta]$. Dann gilt $A = \bigcup_{k=1}^{n} A_k$ und $A_k \cap A_l = \emptyset$ für $k \neq l$. Also ist $P(A) = \sum_{k=1}^{n} P(A_k)$.

Nun ist $P(A_k) = \mathbb{E}(1_{A_k})$ und ist $\omega \in A_k$, so ist $1 \leq \frac{|S_k(\omega)|}{\eta} \leq \frac{S_k(\omega)^2}{\eta^2}$. Also gilt $1_{A_k} \leq \frac{S_k^2}{\eta^2} \cdot 1_{A_k}$ und damit

$$\begin{aligned} P(A_k) &= \mathbb{E}(1_{A_k}) \\ &\leq \frac{1}{\eta^2}\mathbb{E}(S_k^2 \cdot 1_{A_k}) \\ &\leq \frac{1}{\eta^2}\mathbb{E}(S_n^2 \cdot 1_{A_k}) \end{aligned}$$

Hieraus folgt nun wegen $1_A = \sum_{k=1}^{n} 1_{A_k}$

$$\begin{aligned} P(A) &\underset{(1)}{=} \sum_{k=1}^{n} P(A_k) \\ &= \sum_{k=1}^{n} \mathbb{E}(1_{A_k}) \\ &\leq \sum_{k=1}^{n} \frac{1}{\eta^2}\mathbb{E}(S_n^2 \cdot 1_{A_k}) \\ &= \frac{1}{\eta^2}\mathbb{E}(S_n^2 1_A) \\ &\leq \frac{1}{\eta^2}\mathbb{E}(S_n^2) \end{aligned}$$

Nun ist $S_n^2 = \sum_{k=1}^{n} Z_k^2 + 2\sum_{j<k} Z_j Z_k$. Da die Z_j unabhängig sind und den Mittelwert 0 haben, ist $\mathbb{E}(Z_j Z_k) = 0$ (siehe die Definition der Kovarianz) und $\mathbb{E}(Z_k^2) = V(Z_k)$, also wegen $\mathbb{E}(S_n) = 0$ schließlich

$$\begin{aligned} V(S_n) &= \mathbb{E}(S_n^2) \\ &= \sum_{k=1}^{n} V(Z_k) \end{aligned}$$

Einsetzen in Ungleichung (1) liefert die Behauptung.

9.6. (Ursprünglich 9.9) Lemma

Sei (X_n) eine Folge von unabhängiger, quadratische integrierbarer Zufallsvariablen und es gelte

(i) $\mathbb{E}(X_n) = \mu$ für alle n

(ii) $\forall n : V(X_n) \leq \beta$ für ein $\beta > 0$

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Sei $\eta > 0$ beliebig und $N \in \mathbb{N}$.

Behauptung:

$$\begin{aligned} P(\{\omega : \sup_{n\geq N}|\overline{X}_n - \mu| > \eta\}) &= P\left(\left\{\omega : \sup_{n\geq N}\left|\frac{1}{n}\sum_{k=1}^{n} X_k(\omega) - \mu\right| > \eta\right\}\right) \\ &\leq \frac{4\cdot\beta}{N\cdot\eta^2} \end{aligned}$$

Beweis:

(I) Damit wir die Ungleichung von Kolmogoroff anwenden können, setzen wir $Z_n = X_n - \mu$, ferner $S_n = \sum_{k=1}^{n} Z_k$ und erhalten, dass die Z_n stochastisch unabhängig sind, den Mittelwert 0 und die Varianz $V(Z_n) = V(X_n)$ haben.

Es ist

$$\begin{aligned} \bar{X}_n - \mu &= \frac{1}{n}\sum_{k=1}^{n} X_k - \mu \\ &= \frac{1}{n}\left(\sum_{k=1}^{n}(X_k - \mu)\right) \\ &= \frac{1}{n}\sum_{k=1}^{n} Z_k \\ &= \frac{1}{n}S_n \\ &=: \bar{Z}_n \end{aligned}$$

(II) Für beliebiges $l \geq 0$ gilt

$$\begin{aligned} \left[\max_{N2^l\leq n<N2^{l+1}} |\bar{Z}_n| \geq \eta\right] &\underset{(1)}{=} \left[\max_{N2^l\leq n<N2^{l+1}} |S_n| \geq \eta n\right] \\ &\subseteq \left[\max_{N2^l\leq n<N2^{l+1}} |S_n| \geq \eta N2^l\right] \\ &\subseteq \left[\max_{n<N2^{l+1}} |S_n| \geq \eta N2^l\right]. \end{aligned}$$

Ferner gilt

$$\left[\sup_{n\geq N} |\bar{Z}_n| > \eta\right] \subseteq \biguplus_{l=0}^{\infty}\left[\max_{N2^l\leq n<N2^{l+1}} |\bar{Z}_n| \geq \eta\right]$$

Daraus folgt

$$\begin{aligned} P\left(\left[\sup_{n\geq N}|\bar{Z}_n|>\eta\right]\right) &\underset{(2)}{\leq} \sum_{l=0}^{\infty} P\left(\left[\max_{N2^l\leq n<N2^{l+1}}|\bar{Z}_n|\geq\eta\right]\right) \\ &\underset{(2)}{le} \sum_{l=0}^{\infty} P\left(\left[\max_{n<N2^{l+1}}|S_n|\geq\eta N2^l\right]\right) \end{aligned}$$

Die einzelnen Summanden schätzen wir mit der Kolmogoroffschen Ungleichung ab. Dabei beachten wir die Voraussetzung $V(X_k)=V(Z_k)\leq\beta$ und erhalten

$$P\left(\left[\max_{n<N2^{l+1}}|S_n|\geq\eta N2^l\right]\right) \quad\leq\quad \frac{1}{\eta^2N^22^{2l}}\cdot N2^{l+1}\beta$$

Einsetzen in (2) liefert wegen $\bar{Z}_n=\bar{X}_n-\mu$

$$P\left(\left[\sup_{n\geq N}|\bar{X}_n-\mu|>\eta\right]\right)\leq\frac{2\beta}{\eta^2N}\sum_{l=0}^{\infty}2^{-l}=\frac{4\beta}{\eta^2\cdot N}$$

9.7. Theorem (Starkes Gesetz der großen Zahlen)

Seien (X_n) unabhängige Zufallsvariablen mit $\mathbb{E}(X_n)=\mu$, $V(X_n)\leq\beta$ für alle n.
Dann gilt

$$P\left(\left\{\omega:\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k(\omega)=\mu\right\}\right) \quad=\quad 1$$

Beweis:

(I) Wir setzen wieder $\bar{X}_n-\mu=\bar{Z}_n$ mit $Z_k=X_k-\mu$ (vergl. den Beweis von 9.6, Teil (II)) und müssen zeigen:

$$P\left(\left[\lim_{n\to\infty}\bar{Z}_n=0\right]\right) \quad=\quad 1$$

Äquivalent dazu ist die Aussage

$$P\left(\{\omega:(|\bar{Z}_n(\omega)|\nrightarrow 0\}\right) \quad=\quad 0$$

Wir setzen der Bequemlichkeit halber

$$U_n \quad=\quad |\bar{Z}_n|$$

(II) Sei $A=\{\omega:(U_n(\omega)|\nrightarrow 0\}$ und für r, $n\in\mathbb{N}$ sei $A_{n,r}=[\sup_{k\geq n}U_k>1/r]$. Dann ist

$$A \quad=\quad \bigcup_{r=1}^{\infty}\bigcap_{n=1}^{\infty}A_{n,r}$$

Dies ergibt sich aus Folgendem:

Es gilt

$$\begin{gathered}\lim_{n\to\infty} U_n(\omega) = 0\\ \Leftrightarrow\\ \forall r \exists n \forall k \geq n (U_k(\omega) \leq 1/r)\\ \Leftrightarrow\\ \forall r \exists n \left(\sup_{k\geq n} U_k(\omega) \leq 1/r\right)\end{gathered}$$

Damit gilt

$$\begin{gathered}\omega \in A\\ \Leftrightarrow\\ (U_n(\omega)) \nrightarrow 0\\ \Leftrightarrow\\ \exists r \forall n \left(\sup_{k\geq n} U_k(\omega) > 1/r\right)\\ \Leftrightarrow\\ \omega \in \bigcup_{r=1}^{\infty} \bigcap_{n=1}^{\infty} A_{n,r}\end{gathered}$$

(III) Für jedes feste $n \in \mathbb{N}$ und $r \in \mathbb{N}$ ist nun aber nach dem Lemma

$$P\left(\left[\sup_{k\geq n} |\bar{Z}_k| > \frac{1}{r}\right]\right) = P(A_{n,r}) \leq \frac{4\beta r^2}{n}$$

Also ist wegen $A_{n+1,r} \subseteq A_{n,r}$

$$\begin{aligned} P(\bigcap_{n=1}^{\infty} A_{n,r}) &= \lim_{n\to\infty} P(A_{n,r})\\ &\leq \lim_{n\to\infty} \frac{4\beta r^2}{n}\\ &= 0\end{aligned}$$

Daraus folgt

$$\begin{aligned} P(A) &\leq \sum_{r=1}^{\infty} P(\bigcap_{n=1}^{\infty} A_{n,r})\\ &= 0\end{aligned}$$

9.8. Beispiel

Bernoulli-Experiment,
$\Omega = \{0,1\}^{\mathbb{N}}$, $X_n(\omega) = \omega_n$, $\mathbb{E}(X_n) = p$, $V(X_n) = p(1-p)$

$$P([\lim_{n\to\infty} (\frac{1}{n}\sum_{k=1}^{n} \omega_k = p]) = 1$$

9.9. Theorem

Sei (X_n) eine Folge von unabhängigen Zufallsvariablen mit der gleichen Verteilung, d.h. es ist $P_{X_n} = P_{X_m}$ für alle n, m und es möge die Varianz existieren.
Sei

$$S_n^*(\omega) = \frac{1}{\sigma \cdot \sqrt{n}} \sum_{k=1}^{n} (X_k(\omega) - \mu)$$

Dann gilt

$$\begin{aligned} \lim_{n\to\infty} P([a \leq S_n^* < b]) &= \phi(b) - \phi(a) \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{X^2}{2}} \,\mathrm{d}x \end{aligned}$$

Außerdem

$$P_{S_n^*} \approx N(0,1)$$

Beispiel zum zentralen Grenzwertsatz:
(Y_n) Folge von Zufallsvariablen, unabhängig.
Poissonverteilung mit λ.

$$\begin{aligned} \overline{Y}_n &= \frac{1}{n} \sum_{k=1}^{n} Y_k \\ S_n^* &= \frac{1}{\sqrt{\lambda n}} \sum_{k=1}^{n} (Y_k - \lambda) \end{aligned}$$

Es gilt

$$P(a < S_n^* \leq b) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \,\mathrm{d}x$$

Gefragt ist

$$P\left(k < \sum_{j=1}^{n} Y_j \leq l\right) \approx ?$$

Es gilt

$$\begin{aligned} k < \sum_{j=1}^{n} Y_j \leq l &\Leftrightarrow k - n \cdot \lambda < \sum_{j=1}^{n} (Y_j - \lambda) \leq l - n \cdot \lambda \\ &\Leftrightarrow \frac{k - n \cdot \lambda}{\sqrt{n \cdot \lambda}} < S_n^* \leq \frac{l - n \cdot \lambda}{\sqrt{n \cdot \lambda}} \end{aligned}$$

Daraus folgt

$$\begin{aligned} P\left(k < \sum_{j=1}^{n} Y_j \leq l\right) &= P\left(\frac{k - n \cdot \lambda}{\sqrt{n \cdot \lambda}} < S_n^* \leq \frac{l - n \cdot \lambda}{\sqrt{n \cdot \lambda}}\right) \\ &= \phi\left(\frac{l - n \cdot \lambda}{\sqrt{n \cdot \lambda}}\right) - \phi\left(\frac{k - n \cdot \lambda}{\sqrt{n \cdot \lambda}}\right) \end{aligned}$$

Sei $\lambda = 1$, $n = 36$, $k = 30$, $l = 40$.

$$\begin{aligned} \frac{l - n \cdot \lambda}{\sqrt{n \cdot \lambda}} &= \frac{40 - 36 \cdot 1}{\sqrt{36 \cdot 1}} \\ &\approx 0,67 \\ \frac{k - n \cdot \lambda}{\sqrt{n \cdot \lambda}} &= \frac{30 - 36 \cdot 1}{\sqrt{36 \cdot 1}} \\ &= -1 \end{aligned}$$

$$\begin{aligned} P\left(30 < \sum_{j=1}^{n} Y_j \leq 40\right) &= \phi(0,67) - \phi(-1) \\ &= 0,786 - 1 + 0,84 \\ &= 0,786 - 0,16 \\ &= 0,726 \end{aligned}$$

9.10. Satz

Sei $\mathbb{E}(Z_j) = 0$, so gilt

$$P\left(\left[\max_{k \leq n} \left|\sum_{j=1}^{k} Z_k\right| \geq \eta\right]\right) < \frac{n \cdot \beta}{\eta^2}$$

Kolmogorow'sche Ungleichung.

Ist $n = 1$, so erhält man die Tschebyscheff'sche Ungleichung:

$$P\left([|X - \mathbb{E}(X)| \geq \eta]\right) < \frac{\sigma^2}{\eta^2}$$

9.11. Satz (Schwaches Gesetz der großen Zahlen)

Seien $(X_n)_n$ paarweise unkorreliert mit $\mathbb{E}(X_n) = \mu\ \forall n$ und $V(X_n) = \sigma^2\ \forall n$.

Dann gilt:
Zu jedem $\eta > 0$ und $\epsilon > 0$ gibt es ein $n(\eta, \epsilon)$, so dass für alle $n \geq n(\eta, \epsilon)$ gilt

$$P\left(\left\{\omega : \left|\overline{X}_n - \mu\right| \geq \eta\right\}\right) \leq \epsilon$$

Äquivalenz dazu ist folgende Aussage

$$\lim_{n\to\infty} P\left(\left[\left|\overline{X}_n - \mu\right| \geq \eta\right]\right) \quad = \quad 0$$

Beweis:
TODO