

INHALTSVERZEICHNIS

1 Diskrete Wahrscheinlichkeitstheorie

1.1 Laplace-Experiment

Ein Modell für zufälliges Experiment mit *endlich vielen gleichberechtigten* möglichen Ergebnissen wird betrachtet. Die Menge Ω allen Ergebnissen heißt *Ergebnismenge*, ihre Elemente $\omega \in \Omega$ *Elementarereignisse* und ihre Teilmengen $A \subseteq \Omega$ *Ereignisse*. Wenn das Resultat des zufälligen Experimentes ω in A liegt, sagt man, dass *Ereignis A eintritt*, sonst, *tritt A nicht ein*. Die (*Laplace-*)*Wahrscheinlichkeit* von Ereignis A wird durch

$$P(A) = \frac{|A|}{|\Omega|}$$

definiert, wobei $|\cdot|$ die Mächtigkeit der Menge ist.

- $P(\Omega) = 1$
- Für disjunkte $A, B \subset \Omega$, $P(A \cup B) = P(A) + P(B)$

1.2 Urnenmodelle

Eine mit N *gleichartigen* (also fühlbar nicht zu unterscheiden) aber *sichtlich unterscheidbaren* Objekten, z.B. numerierten oder verschiedenfarbigen Kugeln, Urne wird betrachtet. Es wird angenommen, dass die Kugeln *gut gemischt* sind. K Kugeln werden *blind* nacheinander aus der Urne gezogen. Wir unterscheiden die folgenden Verfahren:

- *Ziehen ohne Zurücklegen, mit Reihenfolge*: Die Reihenfolge, in der die Kugeln ausgezogen sind, wird berücksichtigt.

$$\Omega_{o,m} = \{(w_1, \dots, w_K) : w_i \in \{1, \dots, N\}, w_i \neq w_j \text{ für } i \neq j\}, \quad |\Omega_{o,m}| = \frac{N!}{(N-K)!}$$

- *Ziehen ohne Zurücklegen, ohne Reihenfolge*: Die Reihenfolge, in der die Kugeln ausgezogen sind, spielt keine Rolle.

$$\Omega_{o,o} = \{\{w_1, \dots, w_K\} : w_i \in \{1, \dots, N\}, w_i \neq w_j \text{ für } i \neq j\}, \quad |\Omega_{o,o}| = \binom{N}{K}$$

- *Ziehen mit Zurücklegen, mit Reihenfolge*: Nach jedem Zug wird die gezogene Kugel in die Urne zurückgelegt und die Kugeln werden gut gemischt.

$$\Omega_{m,m} = \{(w_1, \dots, w_K) : w_i \in \{1, \dots, N\}\}, \quad |\Omega_{m,m}| = N^K$$

- *Ziehen mit Zurücklegen, ohne Reihenfolge*:

$$\Omega_{m,o} = \{\{w_1, \dots, w_K\} : w_i \in \{1, \dots, N\}\}, \quad |\Omega_{m,m}| = \binom{N+K-1}{K}$$

– Jedes Urnenmodell kann als Verteilungsproblem von K Murmeln auf N Zellen umformuliert werden.

- *Hypergeometrische Verteilung:* In eine Urne befinden sich S schwarze und W weiße Kugeln. n Kugeln werden ohne Zurücklegen gezogen. Für $0 \leq s \leq S$ und $0 \leq w \leq W$ mit $s+w = n$, sei $A_{s,w}$ das Ereignis, dass es genau s schwarzen und w weißen Kugeln gezogen wird. Dann

$$P(A_{s,w}) = \frac{\binom{S}{s} \binom{W}{w}}{\binom{S+W}{s+w}}$$

- *Verallgemeinerte Hypergeometrische Verteilung:* In eine Urne befinden sich N Kugeln, von denen N_i Kugeln die Farbe i haben, $1 \leq i \leq k$. n Kugeln werden ohne Zurücklegen gezogen. Für jede $n_1, \dots, n_k \geq 0$ mit $n_1 + \dots + n_k = n$, sei $A_{n_1, \dots, n_k}$ das Ereignis, dass es genau n_i Kugeln der Farbe i gezogen wird. Dann

$$P(A_{n_1, \dots, n_k}) = \frac{\binom{N_1}{n_1} \cdots \binom{N_k}{n_k}}{\binom{N}{n}}$$

- *Bertand's ballot Theorem:* In eine Urne befinden sich S schwarze und W weiße Kugeln. Die Kugeln werden nacheinander aus der Urne gezogen, bis die Urne leer ist. Sei $A_>$ das Ereignis, dass nach *jedem* Zug die Anzahl der gezogenen schwarzen Kugeln größer als die Anzahl der gezogenen weißen Kugeln ist. Sei $A_\geq$ das Ereignis, dass nach jedem Zug die Anzahl der gezogenen schwarzen Kugeln gleich oder größer als die Anzahl der gezogenen weißen Kugeln ist. Dann für $S \geq W$,

$$P(A_>) = \frac{S-W}{S+W}, \quad P(A_\geq) = \frac{S+1-W}{S+1}$$

1.3 Diskreter Wahrscheinlichkeitsraum

Definition: Ein *diskreter Wahrscheinlichkeitsraum* ist ein Tupel (Ω, P) , wobei Ω eine endliche oder höchstens abzählbar Unendliche Menge und $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ mit

1. $P(\Omega) = 1$
2. für alle Folgen $(A_i)_{i \geq 1}$ von paarweise disjunkten Mengen $A_i \subseteq \Omega$ gilt

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

ist. Die Menge Ω nennt man *Ergebnismenge*, ihre Elemente *Elementarereignisse*, ihre Teilmengen *Ereignisse*. Die Abbildung P heißt *Verteilung* oder *Wahrscheinlichkeitsmaß* auf Ω und $P(A)$ die *Wahrscheinlichkeit* von A .

Bemerkung: Jede Verteilung auf Ω wird durch ihre Werte für einelementigen Mengen eindeutig bestimmt. Sei $p(\omega) = P(\{\omega\})$. Dann gelten (1) $p(\omega) \in [0, 1]$, (2) $\sum_{\omega \in \Omega} p(\omega) = 1$, (3) für alle $A \subseteq \Omega$, $P(A) = \sum_{\omega \in A} p(\omega)$. Und umgekehrt, bestimmt jede Funktion $p : \Omega \rightarrow [0, 1]$ mit der Eigenschaft $\sum_{\omega \in \Omega} p(\omega) = 1$ eindeutig eine Verteilung auf Ω durch $P(A) = \sum_{\omega \in A} p(\omega)$. Also, um eine Verteilung zu definieren, genügt es, die entsprechende Einzelwahrscheinlichkeiten $p(\omega) = P(\{\omega\})$ zu bestimmen.

Lemma 1 (Eigenschaften). Sei (Ω, P) ein diskreter W-Raum. Dann

1. $P(\emptyset) = 0$
2. $P(A^c) = 1 - P(A)$
3. für $A \subseteq B$, $P(A) \leq P(B)$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5. $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$
6. (Sylvester-Siebformel / Prinzip von Inklusion und Exklusion)

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{i=1}^n P(A_i) - \sum_{i_1 < i_2} P(A_{i_1} \cap A_{i_2}) + \sum_{i_1 < i_2 < i_3} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ &\quad - \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n). \end{aligned}$$

Beispiel: 1. Bernoulli-Verteilung mit Parameter p , $B(p)$: $\Omega = \{0, 1\}$,

$$p(1) = p, \quad p(0) = 1 - p.$$

2. Binomialverteilung mit Parametern n und p , $B(n, p)$: $\Omega = \{0, 1, \dots, n\}$,

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

- Binomialverteilung beschreibt die Anzahl der Erfolge in ein Glücksspiel mit n Versuchen, die jeweils Erfolgswahrscheinlichkeit p haben (s.g. Bernoulli-Prozess).

3. Multinomialverteilung mit Parametern $n, p_1, \dots, p_k$:

$$p(n_1, \dots, n_k) = \binom{n}{n_1, \dots, n_k} p_1^{n_1} \dots p_k^{n_k}, \quad n_1, \dots, n_k \geq 0, \quad \sum_{i=1}^k n_i = n.$$

4. Hypergeometrische Verteilung mit Parametern N, K, n :

$$p(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad \max(0, n + K - N) \leq k \leq \min(n, K).$$

- Hypergeometrische Verteilung beschreibt die Anzahl von gezogenen schwarzen Kugeln bei einem Zug von n Kugeln aus einer Urne mit K schwarzen und $N - K$ weißen Kugeln.
- Wenn N und K groß sind und $\frac{K}{N} \rightarrow p$, kann die Hypergeometrische Verteilung durch Binomialverteilung mit Parametern n und p approximiert werden.

5. Geometrische Verteilung mit parameter $p \in [0, 1]$, $\text{Geo}(p)$:

$$p(k) = p(1-p)^{k-1}, \quad k \geq 1.$$

- Geometrische Verteilung beschreibt die Wartezeit auf ersten Erfolg in Glücksspiel mit Erfolgswahrscheinlichkeit p .

6. Poisson-Verteilung mit Parameter $\lambda > 0$, $\text{Poi}(\lambda)$:

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \geq 0.$$

1.4 Grenzwertsätze

Satz 2 (Poisson-Approximationssatz). Seien $\lambda > 0$ und $(p_n)_{n \geq 1}$ eine Folge in $[0, 1]$ mit $np_n \rightarrow \lambda$. Dann gilt

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{für alle } k \geq 0.$$

Satz 3 (Satz von DeMoivre-Laplace). Für alle $p \in (0, 1)$, $q = 1 - p$ und $a < b$, gilt

$$\sum_{np+a\sqrt{npq} \leq k \leq np+b\sqrt{npq}} \binom{n}{k} p^k q^{n-k} \rightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

- Stirling-Formel: $m! \sim \sqrt{2\pi m} m^m e^{-m}$ für $m \rightarrow \infty$.

1.5 Bedingte Wahrscheinlichkeit

Definition: Seien (Ω, P) ein diskreter W-Raum und $A, B \subseteq \Omega$ mit $P(B) > 0$. Die *bedingte Wahrscheinlichkeit von A vorausgesetzt B* (oder auch *unter der Bedingung B*) ist definiert durch

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

- Wenn $P(B) = 0$, setzen wir $P(A|B) = 0$.

Lemma 4 (Eigenschaften). Sei (Ω, P) ein diskreter W-Raum.

1. Für alle $B \subseteq \Omega$ mit $P(B) > 0$, ist $P(\cdot|B)$ eine Verteilung auf $\Omega \cap B$.
2. Für $A \supseteq B$, $P(A|B) = 1$.
3. (Multiplikationssatz)

$$P(A_1 \cap \dots \cap A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

- Diese Formel lässt sich anschaulich mit Baumdigrammen darstellen.

4. (Gesetz der totalen Wahrscheinlichkeit) Für jede Partition $B_1, \dots, B_n$ von Ω gilt

$$P(A) = \sum_{i=1}^n P(B_i) P(A|B_i).$$

5. (Satz von Bayes) Sei $P(A) > 0$ und $P(B_i) > 0$, dann gilt

$$P(B_1|A) = \frac{P(B_1) P(A|B_1)}{\sum_{i=1}^n P(B_i) P(A|B_i)}.$$

- Hier nennt man B_i die Hypothese, $P(B_i)$ die Apriori Wahrscheinlichkeit von B_i und $P(B_i|A)$ die Aposteriori Wahrscheinlichkeit von B_i .

1.6 Unabhängigkeit

Definition: Sei (Ω, P) ein diskreter W-Raum.

1. Zwei Ereignisse $A, B \subseteq \Omega$ heißen (stochastisch) *unabhängig*, wenn

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

2. Ereignisse $A_1, \dots, A_n$ heißen *unabhängig*, wenn für jede Teilmenge $I \subseteq \{1, \dots, n\}$ gilt

$$P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i).$$

3. Ereignisse $A_1, \dots, A_n$ heißen *paarweise unabhängig*, wenn für jede $i \neq j$ A_i und A_j unabhängig sind.

Lemma 5 (Eigenschaften). Sei (Ω, P) ein diskreter W-Raum.

1. Wenn A, B unabhängig sind und $P(B) > 0$, dann $P(A|B) = P(A)$.
2. $A_1, \dots, A_n$ unabhängig sind genau dann, wenn $B_1, \dots, B_n$ unabhängig sind für jede Wahl von $B_i \in \{A_i, A_i^c\}$.
 - In der Tat, $A_1, \dots, A_n$ sind unabhängig genau dann, wenn $P(\cap_{i=1}^n B_i) = \prod_{i=1}^n P(B_i)$ für alle $B_i \in \{A_i, A_i^c\}$.

Definition: Seien $(\Omega_1, P_1), \dots, (\Omega_n, P_n)$ diskrete W-Räume. Das *Produkt* von $(\Omega_i, P_i)_{1 \leq i \leq n}$ ist der diskrete W-Raum (Ω, P) mit $\Omega = \Omega_1 \times \dots \times \Omega_n$ und $P = P_1 \otimes \dots \otimes P_n$, d.h. für alle $\omega = (\omega_1, \dots, \omega_n) \in \Omega$, $p(\omega) = \prod_{i=1}^n p_i(\omega_i)$.

Satz 6. Seien (Ω, P) das Produkt von $(\Omega_i, P_i)_{1 \leq i \leq n}$ und $A'_i \subseteq \Omega_i$. Dann sind die Ereignisse $A_i = \{\omega \in \Omega : \omega_i \in A'_i\}$ unabhängig in (Ω, P) .

- Die einzelnen Ergebnisse eines Bernoulli-Prozesses sind unabhängig.

1.7 Zufallsvariablen

Definition: Sei (Ω, P) ein diskreter W-Raum. Jede Abbildung $X : \Omega \rightarrow \mathbb{R}$ heißt eine (reellwertige) *Zufallsvariable*.

Beispiel: 1. *Bernoulli-Prozess:* $\Omega = \{0, 1\}^n$, $X(\omega) = \sum_{i=1}^n \omega_i$ ist die Anzahl der Erfolge.

2. *Indikator auf $A \subseteq \Omega$:* $1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$

Definition: Seien (Ω, P) ein diskreter W-Raum und X eine Zufallsvariable. Die *Verteilung* von X ist die Verteilung P_X auf dem Bild $X(\Omega)$ definiert durch

$$P_X(B) = P(X^{-1}(B)) = \sum_{\omega \in \Omega, X(\omega) \in B} p(\omega).$$

Beispiel: Seien (Ω, P) ein diskreter W-Raum und X eine Zufallsvariable.

1. X ist *bernoulliverteilt* mit Parameter p , wenn $P(X = 1) = 1 - P(X = 0) = p$. Man schreibt dann $X \sim B(p)$.
2. X ist *binomialverteilt* mit Parametern n und p , wenn $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ für $0 \leq k \leq n$. Man schreibt dann $X \sim B(n, p)$.

Definition: Seien X eine Zufallsvariable auf (Ω, P) mit Verteilung P_X und Y eine Zufallsvariable auf (Ω', P') mit Verteilung P_Y . Die Zufallsvariablen X, Y heißen *gleichverteilt*, wenn $P_X = P_Y$. Dies wird als $X \stackrel{d}{=} Y$ bezeichnet.

- Seien $\Omega = \{0, 1\}^n$, $p(\omega) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}$, $X(\omega) = \sum_{i=1}^n \omega_i$ und $\Omega' = \{0, \dots, n\}$, $p'(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $Y(k) = k$. Dann sind X, Y gleichverteilt.
- Sei $\Omega = \{0, 1\}$, $p(0) = p(1) = \frac{1}{2}$, $X(\omega) = \omega$, $Y(\omega) = 1 - \omega$. Dann $X \neq Y$ aber $X \stackrel{d}{=} Y$.

Definition: Zufallsvariablen $X_1, \dots, X_n$ auf einem diskreten W-Raum (Ω, P) heißen *unabhängig*, wenn für alle $B_1, \dots, B_n \subseteq \mathbb{R}$,

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

- $X_1, \dots, X_n$ sind unabhängig genau dann, wenn für alle $x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)$, $P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$.
- Ereignisse $A_1, \dots, A_n$ sind unabhängig genau dann, wenn die Zufallsvariablen $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ unabhängig sind.
- Die gemeinsame Verteilung $P_{X_1, \dots, X_n}$ von Zufallsvariablen $X_1, \dots, X_n$ ist die Verteilung von Zufallsvektor $X = (X_1, \dots, X_n)$. Sie ist eindeutig bestimmt durch die Einzelwahrscheinlichkeiten $p_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$. $X_1, \dots, X_n$ sind unabhängig genau dann, wenn $P_{X_1, \dots, X_n} = P_{X_1} \otimes \dots \otimes P_{X_n}$.

Beispiel (Bernoulli-Prozess): Seien $\Omega = \{0, 1\}^n$, $p(\omega) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}$ und $X_i(\omega) = \omega_i$. Dann sind $X_1, \dots, X_n$ unabhängige bernoulliverteilte Zufallsvariablen. Beachte, dass ihre Summe binomialverteilt ist.

Lemma 7 (Eigenschaften). Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen auf einem diskreten W-Raum (Ω, P) .

1. Für jede $I \subseteq \{1, \dots, n\}$ sind $\{X_i, i \in I\}$ unabhängig.
2. Für jede paarweise disjunkte $I_1, \dots, I_k \subseteq \{1, \dots, n\}$ und $f_i : \mathbb{R}^{|I_i|} \rightarrow \mathbb{R}$, sind $f_1(X_j, j \in I_1), \dots, f_k(X_j, j \in I_k)$ unabhängig.
3. $P(X_1 + X_2 = x) = \sum_y P(X_1 = y) P(X_2 = x - y)$.

Beispiel: 1. Wenn X, Y, Z unabhängig sind, dann sind $X + Y$ und Z^2 unabhängig.

2. Wenn $X_1, \dots, X_n$ unabhängige bernoulliverteilte Zufallsvariablen mit Parameter p auf einem beliebigen diskreten W-Raum sind, dann ist ihre Summe binomialverteilt mit Parametern n und p . Wenn $X \sim B(m, p)$ und $Y \sim B(n, p)$ unabhängig sind, dann $X + Y \sim B(m + n, p)$.
3. Wenn X, Y unabhängige poissonverteilte Zufallsvariablen mit Parametern λ bzw. μ auf einem beliebigen diskreten W-Raum sind, dann $X + Y \sim \text{Poi}(\lambda + \mu)$.

Proposition 8 (Gedächtnislosigkeit der geometrischen Verteilung). *Sei X eine $\mathbb{N}$ -wertige Zufallsvariable auf einem diskreten W -Raum (Ω, P) . Dann ist X geometrischverteilt genau dann, wenn*

$$P(X = k) = P(X = n + k \mid X > n), \quad k, n \in \mathbb{N}.$$

- Einige Autoren definieren die geometrische Verteilung als $p(k) = p(1-p)^k$ für $k \geq 0$. Die auf diese Weise definierte Verteilung besitzt keine Gedächtnislosigkeit.

1.8 Erwartungswert und Varianz

Definition: 1. Eine Zufallsvariable X auf einem diskreten W -Raum (Ω, P) ist *integrierbar*, wenn $\sum_{\omega \in \Omega} |X(\omega)|p(\omega) < \infty$.

2. Für jede integrierbare oder nicht-negative Zufallsvariable X heißt die Summe

$$E[X] = \sum_{\omega \in \Omega} X(\omega)p(\omega)$$

der *Erwartungswert* von X . Die Integrierbarkeit von X wird auch als $E[|X|] < \infty$ bezeichnet.

Lemma 9 (Eigenschaften). *Sei (Ω, P) ein diskreter W -Raum.*

1. (Substitutionsformel) *Eine Zufallsvariable X ist genau dann integrierbar, wenn $\sum_{x \in X(\Omega)} |x|P(X = x) < \infty$. In diesem Fall gilt*

$$E[X] = \sum_{x \in X(\Omega)} x P(X = x) = \sum_{x \in X(\Omega)} x p_X(x).$$

Insbesondere, sind die Integrierbarkeit und der Erwartungswert von X nur von der Verteilung von X bestimmt, d.h. wenn X integrierbar ist und $Y \stackrel{d}{=} X$, dann ist Y auch integrierbar und $E[Y] = E[X]$.

2. Wenn $X \geq 0$, dann $E[X] \geq 0$.

3. (Linearität) Wenn X, Y integrierbar sind, dann ist $aX + bY$ für $a, b \in \mathbb{R}$ auch integrierbar und gilt

$$E[aX + bY] = aE[X] + bE[Y].$$

4. $|E[X]| \leq E[|X|]$

5. (Cauchy-Schwarz-Ungleichung) $|E[XY]| \leq E[X^2]^{\frac{1}{2}} E[Y^2]^{\frac{1}{2}}$

6. Wenn X, Y unabhängig und integrierbar sind, dann ist XY auch integrierbar und $E[XY] = E[X] E[Y]$.

7. $E[\mathbb{1}_A] = P(A)$

8. Wenn $X \geq 0$ und $E[X] = 0$, dann $P(X = 0) = 1$.

9. Wenn $X \in \mathbb{N}$, dann $E[X] = \sum_{n=1}^{\infty} P(X \geq n)$.

Beispiel: 1. $X \sim B(p)$, $E[X] = p$.

2. $X \sim B(n, p)$, $E[X] = np$.
3. $X \sim \text{Geo}(p)$, $E[X] = \frac{1}{p}$.
4. $X \sim \text{Poi}(\lambda)$, $E[X] = \lambda$.

Definition: Sei X eine Zufallsvariable mit $E[X^2] < \infty$. Die *Varianz von X* ist definiert durch

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

- Die Quadratwurzel aus $\text{Var}(X)$ heißt die *Standardabweichung von X* .

Lemma 10 (Eigenschaften). 1. Wenn $X = c$ (Konst), dann $\text{Var}(X) = 0$.

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

3. (Gleichung von Bienaymé) Wenn $X_1, \dots, X_n$ unabhängig und mit endlichen Varianzen sind, dann

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

Beispiel: 1. $X \sim B(p)$, $\text{Var}(X) = p(1 - p)$.

2. $X \sim B(n, p)$, $\text{Var}(X) = np(1 - p)$.

3. $X \sim \text{Geo}(p)$, $\text{Var}(X) = \frac{1-p}{p^2}$.

4. $X \sim \text{Poi}(\lambda)$, $\text{Var}(X) = \lambda$.

Definition: Seien X, Y Zufallsvariablen auf einem (demselden) diskreten W-Raum (Ω, P) . Die *Kovarianz von X, Y* ist definiert durch

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

Lemma 11 (Eigenschaften). 1. Wenn X, Y unabhängig sind, dann sind X, Y unkorreliert, d.h. $\text{cov}(X, Y) = 0$, aber nicht umgekehrt.

2. Seien $X_1, \dots, X_n$ Zufallsvariablen, dann

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j).$$

1.9 Erzeugende Funktion

Definition: Die *erzeugende Funktion einer $\mathbb{N}_0$ -wertige Zufallsvariable X* ist definiert durch

$$\varphi_X(z) = \sum_{n=0}^{\infty} z^n P(X = n), \quad |z| < 1.$$

- $\varphi_X(z) = E[z^X]$

Lemma 12 (Eigenschaften). 1. φ_X ist analytisch im Inneren des komplexen Einheitskreises. Insbesondere,

$$P(X = n) = \frac{\varphi_X^{(n)}(0)}{n!}.$$

Somit $\varphi_X = \varphi_Y$ genau dann, wenn $X \stackrel{d}{=} Y$.

2. Wenn X, Y unabhängige $\mathbb{N}_0$ -wertige Zufallsvariablen sind, dann $\varphi_{X+Y} = \varphi_X \varphi_Y$.

Beispiel: 1. $X \sim B(p)$, $\varphi_X(z) = 1 - p + pz$.

2. $X \sim B(n, p)$, $\varphi_X(z) = (1 - p + pz)^n$.

3. $X \sim \text{Poi}(\lambda)$, $\varphi_X(z) = e^{\lambda(z-1)}$.

2 Allgemeine Wahrscheinlichkeitstheorie

Hier werden die bisher eingeführten Begriffe auf allgemein unzählbare Ergebnismengen erweitert.

2.1 Wahrscheinlichkeitsraum

Definition: Ein *Wahrscheinlichkeitsraum* ist ein Maßraum $(\Omega, \mathcal{F}, P)$, dessen Maß P ein Wahrscheinlichkeitsmaß ist, d.h. $P(\Omega) = 1$.

- Die Menge Ω nennt man *Ergebnismenge*, ihre Elemente *Elementarereignisse*, ihre messbaren Teilmengen $A \in \mathcal{F}$ *Ereignisse*. Das Maß P heißt *Verteilung* oder auch *Wahrscheinlichkeitsmaß*, $P(A)$ die *Wahrscheinlichkeit von A*.
- Der Begriff 'fast überall' von Maßtheorie wird in Wahrscheinlichkeitstheorie normalerweise als *fast sicher* bezeichnet.
- Wenn Ω abzählbar ist, dann ist jeder diskreter W-Raum (Ω, P) auch ein W-Raum mit $\mathcal{F} = \mathcal{P}(\Omega)$.

Lemma 13 (Eigenschaften). Seien $(\Omega, \mathcal{F}, P)$ ein W-Raum und $A, B, A_i \in \mathcal{F}$. Dann gilt

1. $P(\emptyset) = 0$
2. $P(A^c) = 1 - P(A)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. (Bonferroni-Ungleichung) $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$
5. (σ -Stetigkeit) Falls $A_i \uparrow A$ (d.h. $A_1 \subseteq A_2 \subseteq \dots$) oder $A_i \downarrow A$, dann $P(A_n) \rightarrow P(A)$.
6. (Sylvester-Siebformel / Prinzip von Inklusion und Exklusion)

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{i=1}^n P(A_i) - \sum_{i_1 < i_2} P(A_{i_1} \cap A_{i_2}) + \sum_{i_1 < i_2 < i_3} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ &\quad - \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n). \end{aligned}$$

Beispiel (Diskrete Verteilungen): Eine Verteilung P auf $(\Omega, \mathcal{F})$ heißt *diskret*, wenn es $N \in \mathbb{N} \cup \{+\infty\}$, $x_i \in \Omega$ und $p_i \in [0, 1]$ für $1 \leq i \leq N$ geben mit $\sum_{i=1}^N p_i = 1$, sodass für jedes $A \in \mathcal{F}$,

$$P(A) = \sum_{i: x_i \in A} p_i.$$

Bernoulli-, Binomial-, Poisson-, Geometrischeverteilung sind Beispiele von diskreten Verteilungen.

Beispiel (Absolut stetige Verteilungen): Eine Verteilung P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißt *absolut stetig*, wenn es eine Borel-messbare $f : \mathbb{R} \rightarrow [0, +\infty)$ gibt mit $\int_{\mathbb{R}} f(x) dx = 1$, sodass für jedes $A \in \mathcal{B}(\mathbb{R})$,

$$P(A) = \int_A f(x) dx.$$

Die Funktion f nennt man die *Dichte* der Verteilung P .

- *Gleichverteilung auf $[a, b]$, $U[a, b]$:*

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

- *Normalverteilung mit Parametern m und σ^2 , $N(m, \sigma^2)$:*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Im Fall von $m = 0$, $\sigma^2 = 1$, spricht man von *Standardnormalverteilung* $N(0, 1)$.

- *Exponentialverteilung mit Parameter $\lambda > 0$, $\text{Exp}(\lambda)$:*

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

- *Standard-Cauchy-Verteilung:*

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Beispiel (Gleichverteilung auf $\Lambda \subseteq \mathbb{R}^d$): $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{L}(\mathbb{R}^d)$, $\Lambda \in \mathcal{F}$,

$$P(A) = \frac{\lambda(A)}{\lambda(\Lambda)}, \quad A \in \mathcal{F}, A \subseteq \Lambda.$$

2.2 Eindeutigkeit und Existenz von Wahrscheinlichkeitsmaßen

Proposition 14. Sei $\Omega \neq \emptyset$ und $\mathcal{G} \subseteq \mathcal{P}(\Omega)$. Es gibt die kleinste σ -Algebra die $\mathcal{G}$ enthält. Sie heißt die von $\mathcal{G}$ erzeugte σ -Algebra und ist mit $\sigma(\mathcal{G})$ bezeichnet.

Definition: Die Familie $\mathcal{G}$ heißt ein *Erzeuger* von σ -Algebra $\mathcal{F}$, wenn $\mathcal{F} = \sigma(\mathcal{G})$.

Beispiel: 1. Sei Ω abzählbar, dann $\mathcal{P}(\Omega) = \sigma(\{\omega\}, \omega \in \Omega)$.

2. $\mathcal{B}(\mathbb{R}^d) = \sigma(O \subseteq \mathbb{R}^d, O \text{ offen})$ heißt die *Borelsche σ -Algebra auf $\mathbb{R}^d$* .

- $\mathcal{B}(\mathbb{R}^d) = \sigma([a_1, b_1] \times \dots \times [a_d, b_d], a_i < b_i, a_i, b_i \in \mathbb{Q}), \mathcal{B}(\mathbb{R}) = \sigma((-\infty, x], x \in \mathbb{R})$.

3. Sei $\Lambda \in \mathcal{B}(\mathbb{R}^d)$. $\mathcal{B}(\Lambda) = \{A \subseteq \Lambda, A \in \mathcal{B}(\mathbb{R}^d)\}$ heißt die Borelsche σ -Algebra auf Λ .

4. Für $A \subseteq \Omega$ gilt $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$.

5. Seien $(\Omega_i, \mathcal{F}_i)_{i \in I}$ messbare Räume. Seien $\Omega = \prod_{i \in I} \Omega_i$ und $X_i : \Omega \rightarrow \Omega_i$ gegeben durch $X_i(\omega) = \omega_i$ (die Projektion auf i -te Koordinate). Die *Produkt- σ -Algebra der $\mathcal{F}_i$ auf Ω* , bezeichnet als $\bigotimes_{i \in I} \mathcal{F}_i$ ist die von

$$\{X_i^{-1}(A_i), i \in I, A_i \in \mathcal{F}_i\}$$

erzeugte σ -Algebra.

- Falls I abzählbar ist, dann gilt

$$\bigotimes_{i \in I} \mathcal{F}_i = \sigma\left(\prod_{i \in I} A_i, A_i \in \mathcal{F}_i\right).$$

- $\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n)$.

Definition: 1. Eine Familie $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ heißt *π -System auf Ω* , wenn für $C, C' \in \mathcal{C}$, auch $C \cap C' \in \mathcal{C}$.

2. Eine Familie $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ heißt *Dynkin-System auf Ω* (oder auch *λ -System*), wenn

- (a) $\Omega \in \mathcal{D}$,
- (b) wenn $A \in \mathcal{D}$, dann $A^c \in \mathcal{D}$,
- (c) wenn $A_1, A_2, \dots \in \mathcal{D}$ paarweise disjunkt sind, dann $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}$.

Proposition 15. Sei $\Omega \neq \emptyset$ und $\mathcal{G} \subseteq \mathcal{P}(\Omega)$. Es gibt das kleinste Dynkin-System, das $\mathcal{G}$ enthält. Es heißt das von $\mathcal{G}$ erzeugte Dynkin-System und wird mit $d(\mathcal{G})$ bezeichnet.

Lemma 16 (Lemma von Dynkin). Sei $\mathcal{G}$ ein π -System auf Ω . Dann $d(\mathcal{G}) = \sigma(\mathcal{G})$. Insbesondere, wenn $\mathcal{D}$ ein Dynkin-System, das $\mathcal{G}$ enthält, dann enthält $\mathcal{D}$ auch $\sigma(\mathcal{G})$.

Beispiel: Seien P, Q Verteilungen auf $(\Omega, \mathcal{F})$. Dann ist

$$\mathcal{D} = \{A \in \mathcal{F} : P(A) = Q(A)\}$$

ein Dynkin-System.

Obwohl in diskreten W-Räumen jede Verteilung durch ihre Werte für einelementigen Mengen eindeutig bestimmt wird, trifft dies im Allgemeinen nicht zu. Der nächste Satz ist ein nützlicher Ersatz. Er besagt, dass jede Verteilung eindeutig durch ihre Werte für Elementen eines erzeugenden π -Systems bestimmt wird.

Satz 17 (Eindeutigkeit). Seien $(\Omega, \mathcal{F})$ ein Messraum und $\mathcal{G}$ ein π -System auf Ω mit $\sigma(\mathcal{G}) = \mathcal{F}$. Seien P, Q Verteilungen auf $(\Omega, \mathcal{F})$ mit $P(A) = Q(A)$ für alle $A \in \mathcal{G}$. Dann $P = Q$.

Der nächste Satz erlaubt die Verteilungen durch Angabe ihren Werten für Elementen einer erzeugenden Algebra definieren.

Definition: Ein nichtleeres Mengensystem $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt eine *Algebra auf Ω* , wenn (1) $\Omega \in \mathcal{A}$, (2) wenn $A \in \mathcal{A}$, dann $A^c \in \mathcal{A}$, (3) wenn $A, B \in \mathcal{A}$, dann $A \cup B \in \mathcal{A}$.

Satz 18 (Existenz / Erweiterungssatz von Carathéodory). *Seien $(\Omega, \mathcal{F})$ ein Messraum, $\mathcal{A}$ eine Algebra auf Ω mit $\sigma(\mathcal{A}) = \mathcal{F}$ und $P_0 : \mathcal{A} \rightarrow [0, 1]$ ein Prämaß, d.h.*

1. $P_0(\Omega) = 1$
2. für alle $A_1, \dots \in \mathcal{A}$ paarweise disjunkt mit $\cup_{i \geq 1} A_i \in \mathcal{A}$ gilt $P_0(\cup_{i \geq 1} A_i) = \sum_{i \geq 1} P_0(A_i)$.

Dann gibt es genau eine Verteilung P auf $(\Omega, \mathcal{F})$ die P_0 erweitert, d.h. $P(A) = P_0(A)$ für alle $A \in \mathcal{A}$.

Proposition 19. *Seien $(\Omega_i, \mathcal{F}_i, P_i)_{1 \leq i \leq n}$ W-Räume, $\Omega = \prod_{i=1}^n \Omega_i$ und $\mathcal{F} = \bigotimes_{i=1}^n \mathcal{F}_i$. Dann existiert eine eindeutige Verteilung P auf $(\Omega, \mathcal{F})$, sodass*

$$P(A_1 \times \dots \times A_n) = P_1(A_1) \dots P_n(A_n) \quad \text{für alle } A_i \in \mathcal{F}_i.$$

Sie heißt das Produktmaß von $P_1, \dots, P_n$ und wird mit $P = \bigotimes_{i=1}^n P_i$ bezeichnet.

2.3 Zufallsvariablen

In diskreten W-Räumen nennt man jede reellwertige Abbildung eine Zufallsvariable. Um die Theorie auf den allgemeinen Fall zu erweitern, sollten wir uns auf *messbare* Abbildungen beschränken.

Definition: Sei $(\Omega, \mathcal{F})$ ein Messraum. Jede messbare Abbildung von $(\Omega, \mathcal{F})$ nach

1. $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißt eine *Zufallsvariable auf $(\Omega, \mathcal{F})$* .
2. $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ heißt ein *Zufallsvektor auf $(\Omega, \mathcal{F})$* .
3. $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ heißt eine *numerische Zufallsvariable auf $(\Omega, \mathcal{F})$* .

Bemerkung: Die folgenden Abkürzungen sind in der Wahrscheinlichkeitstheorie üblich:

$$\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}, \quad \{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\}.$$

Der folgende Satz ist ein nützliches Kriterium für die Messbarkeit einer Abbildung.

Proposition 20. 1. Sei $f : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ und $\mathcal{F}' = \sigma(\mathcal{G}')$. Dann gilt

$$f \text{ ist messbar} \iff X^{-1}(A') \in \mathcal{F} \text{ für alle } A' \in \mathcal{G}'.$$

2. X ist eine Zufallsvariable auf $(\Omega, \mathcal{F})$ genau dann, wenn $\{X \leq x\} \in \mathcal{F}$ für alle $x \in \mathbb{R}$.
3. X ist eine numerische Zufallsvariable auf $(\Omega, \mathcal{F})$ genau dann, wenn $\{X \leq x\} \in \mathcal{F}$ für alle $x \in \mathbb{R}$.

- Lemma 21** (Eigenschaften). 1. Wenn $f : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ und $g : (\Omega', \mathcal{F}') \rightarrow (\Omega'', \mathcal{F}'')$ messbar sind, dann ist $g \circ f : (\Omega, \mathcal{F}) \rightarrow (\Omega'', \mathcal{F}'')$ auch messbar.
2. Wenn $X_1, \dots, X_n$ Zufallsvariablen sind und $f : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ messbar ist, dann ist $f(X_1, \dots, X_n)$ eine Zufallsvariable.
3. Wenn $X_1, X_2, \dots$ numerische Zufallsvariablen sind, dann sind $\inf_n X_n$, $\sup_n X_n$, $\liminf_n X_n$ und $\limsup_n X_n$ auch numerische Zufallsvariablen.
- Wenn $X_1, \dots, X_n$ Zufallsvariablen sind, dann ist $X_1 + \dots + X_n$ auch eine Zufallsvariable.

Jetzt führen wir den Begriff der Verteilung einer Zufallsvariable ein.

Proposition 22 (Bild von P unter φ). Seien $\varphi : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ eine messbare Abbildung und P eine Verteilung auf $(\Omega, \mathcal{F})$. Dann ist $\varphi \circ P$, definiert durch

$$\varphi \circ P(A') = P(\varphi^{-1}(A')), \quad A' \in \mathcal{F}',$$

eine Verteilung auf $(\Omega', \mathcal{F}')$, die das Bild von P unter φ heißt.

- In Literatur wird oft auch die Notation $P \circ \varphi^{-1}$ benutzt.

Definition: 1. Sei X eine (numerische) Zufallsvariable auf $(\Omega, \mathcal{F}, P)$. Das Bild von P unter X heißt *die Verteilung von X* und wird mit P_X bezeichnet:

$$P_X = X \circ P = P(X^{-1}(\cdot)).$$

2. Seien X eine Zufallsvariable auf $(\Omega, \mathcal{F}, P)$ und Y eine Zufallsvariable auf $(\Omega', \mathcal{F}', P')$. Dann heißen X und Y *gleichverteilt*, wenn $P_X = P_Y$. Dies wird mit $X \stackrel{d}{=} Y$ bezeichnet.

Bemerkung: Da $\mathcal{G} = \{(-\infty, x], x \in \mathbb{R}\}$ ein π -System ist, das $\mathcal{B}(\mathbb{R})$ erzeugt, ist die Verteilung von jede Zufallsvariable X eindeutig durch die Werte von $P(X \leq x)$, $x \in \mathbb{R}$, bestimmt. Dies motiviert die folgende Definition.

Definition: Sei X eine Zufallsvariable auf $(\Omega, \mathcal{F}, P)$. Die Funktion

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}$$

heißt *die Verteilungsfunktion von X* .

Proposition 23. Sei F die Verteilungsfunktion einer Zufallsvariable. Dann gilt

1. F ist monoton steigend,
2. $\lim_{x \rightarrow +\infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$,
3. F ist rechtsstetig.

Interessanter ist jedoch die folgende Umkehrung, die eine “praktische” Beschreibung aller möglichen Verteilungen von Zufallsvariablen bietet.

Proposition 24 (Satz von Lebesgue-Stieltjes). Sei $F : \mathbb{R} \rightarrow [0, 1]$. Dann ist F die Verteilungsfunktion einer Zufallsvariable genau dann, wenn sie die Bedingungen 1-3 des obigen Satzes erfüllt.

Beispiel: 1. Diskrete Zufallsvariable X erfüllt $P(X = x_i) = p_i$, wobei $\sum_{i=1}^N p_i = 1$ und $N \in \mathbb{N} \cup \{+\infty\}$. Ihre Verteilungsfunktion

$$F_X(x) = \sum_{i: x_i \leq x} p_i$$

ist eine stückweise konstante Funktion.

2. Absolut stetige Zufallsvariable X erfüllt $P(X \in A) = \int_A f_X(x) dx$, wobei f_X die Dichte der Verteilung von X ist. Ihre Verteilungsfunktion

$$F_X(x) = \int_{-\infty}^x f_X(y) dy$$

ist eine differenzierbare Funktion.

2.4 Erwartungswert und Varianz

Sei $(\Omega, \mathcal{F}, P)$ ein W-Raum.

Definition: 1. Eine (numerische) Zufallsvariable ist *integrierbar*, wenn das Lebesgue-Integral $\int_{\Omega} |X| dP$ endlich ist.

2. Für jede nicht-negative oder integrierbare (numerische) Zufallsvariable X , der *Erwartungswert von X* wird definiert als Lebesgue-Integral

$$E[X] = \int_{\Omega} X dP.$$

- Es sei daran erinnert, dass das Lebesgue-Integral einer messbaren Funktion f in drei Schritten definiert wird: (1) Wenn f einfach ist, d.h. $f(\omega) = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ für $c_i \in \mathbb{R}$ und $A_i \in \mathcal{F}$, dann $\int_{\Omega} f dP = \sum_{i=1}^n c_i P(A_i)$, (2) Wenn $f \geq 0$ ist, dann definiert man $\int_{\Omega} f dP = \sup\{\int_{\Omega} g dP, g \text{ ist einfach und } g \leq f\}$ oder äquivalent als $\int_{\Omega} f dP = \lim_{n \rightarrow \infty} \int_{\Omega} f_n dP$ für einfache Funktionen $f_n \uparrow f$, (3) Für beliebige f mit $\int_{\Omega} |f| dP < \infty$, definiert man $\int_{\Omega} f dP = \int_{\Omega} f_+ dP - \int_{\Omega} f_- dP$, wobei $f_+ = \max(f, 0)$ der Positivteil von f und $f_- = \max(-f, 0)$ der Negativteil von f sind. Dieselbe Schritte nutzt man oft, um verschiedene Eigenschaften vom Integral zu beweisen.

Definition: Sei X eine Zufallsvariable mit $E[X^2] < \infty$. Die *Varianz von X* wird definiert durch

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

Lemma 25 (Eigenschaften). Seien X, Y integrierbare (numerische) Zufallsvariablen.

1. Für alle $a, b \in \mathbb{R}$ ist $aX + bY$ integrierbar und $E[aX + bY] = aE[X] + bE[Y]$.
2. Wenn $X \leq Y$, dann $E[X] \leq E[Y]$.
3. Wenn $X \geq 0$ P -fast sicher und $E[X] = 0$, dann $X = 0$ P -fast sicher.
4. Wenn $E[X^2] < \infty$, dann

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

5. (Jensensche Ungleichung) Sei X integrierbare Zufallsvariable und $\varphi : \mathbb{R} \rightarrow [0, +\infty)$ eine konvexe Funktion. Dann gilt

$$E[\varphi(X)] \geq \varphi(E[X]).$$

6. (Höldersche Ungleichung) Seien $p, q \in [1, +\infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt

$$E[|XY|] \leq \|X\|_p \|Y\|_q,$$

wobei $\|X\|_r = E[|X|^r]^{\frac{1}{r}}$ für $r < \infty$ und $\|X\|_\infty = \inf\{M : P(|X| > M) = 0\}$.

7. (Markov-Ungleichung) Sei $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ monoton steigend und $X \geq 0$ eine Zufallsvariable. Dann gilt

$$P(X \geq x) \leq \frac{E[f(X)]}{f(x)}.$$

8. (Chebyshev-Ungleichung) Sei X eine Zufallsvariable mit $E[X^2] < \infty$, dann für $x > 0$

$$P(|X - E[X]| \geq x) \leq \frac{\text{Var}(X)}{x^2}.$$

9. (Momente einer nichtnegativen Zufallsvariable) Sei X eine nichtnegative Zufallsvariable. Dann für jedes $p > 0$,

$$E[X^p] = \int_0^\infty px^{p-1} P(X > x) dx.$$

Satz 26 (Konvergenzsätze). 1. (Lemma von Fatou) Seien $X_n \geq 0$ (numerische) Zufallsvariablen. Dann gilt

$$E \left[\liminf_{n \rightarrow \infty} X_n \right] \leq \liminf_{n \rightarrow \infty} E[X_n].$$

2. (Satz von den monotonen Konvergenz) Seien X und $X_n \geq 0$ (numerische) Zufallsvariablen mit $X_n \uparrow X$ f.s. Dann gilt

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

3. (Satz von den dominierten Konvergenz / Lebesgue) Seien X, Y und X_n (numerische) Zufallsvariablen mit $X_n \rightarrow X$ f.s., $|X_n| \leq Y$ f.s. für alle $n \in \mathbb{N}$, und $E[Y] < \infty$. Dann gilt

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

Satz 27. (Substitutionssatz) Sei $\varphi : (\Omega, \mathcal{F}, P) \rightarrow (\Omega', \mathcal{F}')$ eine messbare Abbildung und Y eine Zufallsvariable auf $(\Omega', \mathcal{F}')$. Dann ist Y integrierbar bezüglich $\varphi \circ P$ genau dann, wenn $Y \circ \varphi$ integrierbar bezüglich P ist und es gilt

$$E_{\varphi \circ P}[Y] = E_P[Y \circ \varphi].$$

Insbesondere, für jede Zufallsvariable X auf $(\Omega, \mathcal{F}, P)$,

$$E[X] = \int_{\mathbb{R}} x dP_X.$$

Beispiel: 1. Wenn X eine diskrete Zufallsvariable ist, dann stellen wir die zuvor erhaltene Formel für $E[X]$ wieder her:

$$E[X] = \sum_x x p_X(x) = \sum_x x P(X = x).$$

Insbesondere stimmen der Erwartungswert und die Varianz von bernoulli-, binomial-, poisson- und geometrischverteilten Zufallsvariablen mit den zuvor berechneten überein.

2. Wenn X eine absolut stetige Zufallsvariable mit Dichte f , dann gilt

$$E[X] = \int_{\mathbb{R}} x f(x) dx, \quad \text{Var}(X) = \int_{\mathbb{R}} x^2 f(x) dx - \left(\int_{\mathbb{R}} x f(x) dx \right)^2.$$

3. $X \sim U[a, b]$, $E[X] = \frac{b+a}{2}$.

4. $X \sim \text{Exp}(\lambda)$: $E[X^k] = \frac{k!}{\lambda^k}$, $\text{Var}(X) = \frac{1}{\lambda^2}$.

5. $X \sim N(m, \sigma^2)$: $E[X] = m$, $\text{Var}(X) = \sigma^2$.

6. Eine standard-Cauchy-verteilte Zufallsvariable ist nicht integrierbar.

2.5 Unabhängigkeit

Sei $(\Omega, \mathcal{F}, P)$ ein W-Raum.

Definition: 1. Ereignisse $A, B \in \mathcal{F}$ heißen (stochastisch) unabhängig, wenn

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

2. Teil- σ -Algebren $\mathcal{A}, \mathcal{B}$ von $\mathcal{F}$ heißen unabhängig, wenn alle $A \in \mathcal{A}$ und $B \in \mathcal{B}$ unabhängig sind.

3. Zufallsvariablen X, Y heißen unabhängig, wenn die von ihnen erzeugten σ -Algebren $\sigma(X), \sigma(Y)$ unabhängig sind.

- Die σ -Algebra $\sigma(X_1, \dots, X_n)$ erzeugt von Zufallsvariablen $X_1, \dots, X_n$ ist die kleinste Teil- σ -Algebra $\mathcal{G}$ von $\mathcal{F}$, sodass alle X_i auch Zufallsvariablen auf $(\Omega, \mathcal{G})$ sind. Sie ist erzeugt von $\{X_i^{-1}(B), B \in \mathcal{B}(\mathbb{R}), 1 \leq i \leq n\}$.

Mit anderen Worten,

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B) \quad \text{für alle } A, B \in \mathcal{B}(\mathbb{R}).$$

Definition: 1. Ereignisse $A_1, \dots, A_n \in \mathcal{F}$ heißen (*stochastisch*) *unabhängig*, wenn für alle $I \subseteq \{1, \dots, n\}$,

$$P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i).$$

2. Teil- σ -Algebren $\mathcal{A}_1, \dots, \mathcal{A}_n$ von $\mathcal{F}$ heißen *unabhängig*, wenn

$$P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i), \quad \text{für alle } A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n.$$

3. Zufallsvariablen $X_1, \dots, X_n$ heißen *unabhängig*, wenn die von ihnen erzeugten σ -Algebren $\sigma(X_1), \dots, \sigma(X_n)$ unabhängig sind. Mit anderen Worten,

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n) \quad \text{für alle } B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}).$$

Überprüfung der Bedingungen für die Unabhängigkeit ist im Allgemeinen schwierig. Hier ist ein nützliches Kriterium, das die Überprüfung auf erzeugende π -Systeme reduziert.

Proposition 28. Seien $\mathcal{G}_1, \dots, \mathcal{G}_n$ π -Systeme mit $\Omega \in \mathcal{G}_i$ für alle i und

$$P(\cap_{i=1}^n C_i) = \prod_{i=1}^n P(C_i) \quad \text{für alle } C_1 \in \mathcal{G}_1, \dots, C_n \in \mathcal{G}_n.$$

Dann sind die σ -Algebren $\sigma(\mathcal{G}_1), \dots, \sigma(\mathcal{G}_n)$ unabhängig.

Korollar 29. 1. (Numerische) Zufallsvariablen $X_1, \dots, X_n$ sind unabhängig genau dann, wenn

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \dots P(X_n \leq x_n), \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

2. Seien $\mathcal{F}_{i,j}$, $1 \leq i \leq n$, $1 \leq j \leq n_i$, unabhängige σ -Algebren. Dann sind die σ -Algebren $\mathcal{A}_i = \sigma(\cup_{j=1}^{n_i} \mathcal{F}_{i,j})$, $1 \leq i \leq n$ unabhängig.

3. Seien $X_{i,j}$, $1 \leq i \leq n$, $1 \leq j \leq n_i$, unabhängige Zufallsvariablen und $f_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ Borel-messbare Funktionen, dann sind die Zufallsvariablen $Y_i = f_i(X_{i,1}, \dots, X_{i,n_i})$, $1 \leq i \leq n$, unabhängig.

- Jede Teilfamilie von unabhängigen Objekten ist unabhängig.
- Ereignisse $A_1, \dots, A_n \in \mathcal{F}$ sind unabhängig $\iff \sigma$ -Algebren $\sigma(A_1), \dots, \sigma(A_n)$ sind unabhängig $\iff$ Zufallsvariablen $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ sind unabhängig.

Proposition 30 (Produktmaß). Seien $(\Omega_i, \mathcal{F}_i, P_i)_{1 \leq i \leq n}$ W -Räume. Dann gibt es eine eindeutige Verteilung P auf $(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{F}_i)$, sodass

$$P(A_1 \times \dots \times A_n) = P_1(A_1) \dots P_n(A_n), \quad \text{für alle } A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n.$$

Sie heißt das Produktmaß von $P_1, \dots, P_n$ und wird mit $\bigotimes_{i=1}^n P_i$ bezeichnet. Der W -Raum $(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{F}_i, \bigotimes_{i=1}^n P_i)$ heißt das Produktraum von $(\Omega_i, \mathcal{F}_i, P_i)_{1 \leq i \leq n}$.

Satz 31 (Satz von Fubini). Seien $(\Omega_1, \mathcal{F}_1, P_1)$ und $(\Omega_2, \mathcal{F}_2, P_2)$ W-Räume und $(\Omega, \mathcal{F}, P)$ ihr Produktraum. Für jede nichtnegative oder P -integrierbare Zufallsvariable X auf $(\Omega, \mathcal{F}, P)$ gilt

$$\int_{\Omega} X dP = \int_{\Omega_1} \left[\int_{\Omega_2} X(x, y) P_2(dy) \right] P_1(dx) = \int_{\Omega_2} \left[\int_{\Omega_1} X(x, y) P_1(dx) \right] P_2(dy).$$

Korollar 32. Seien X und Y unabhängige Zufallsvariablen. Dann gilt:

1. Falls X und Y integrierbar sind, dann ist das Produkt XY auch integrierbar und

$$E[XY] = E[X] \cdot E[Y].$$

2. $F_{X+Y} = F_X * F_Y$, d.h.

$$F_{X+Y}(z) = \int_{\mathbb{R}} F_X(z - y) dF_Y(y).$$

3. Falls X, Y besitzen Dichten, dann ist die Dichte von $X + Y$ gleich

$$f_{X+Y} = f_X * f_Y.$$

Beispiel: Falls X_1 und X_2 unabhängige Zufallsvariablen mit $X_i \sim N(m_i, \sigma_i^2)$ sind, dann $X_1 + X_2 \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Proposition 33. (Gleichung von Bienaymé) Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit $E[X_i^2] < \infty$ für alle $1 \leq i \leq n$. Dann gilt

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

3 (Unendliche) Folgen von Zufallsvariablen

Definition: Zufallsvariablen $X_1, X_2, \dots$ auf $(\Omega, \mathcal{F}, P)$ heißen *unabhängig*, wenn $X_1, \dots, X_n$ unabhängig sind für jedes n .

In diesem Abschnitt werden wir uns hauptsächlich mit Fragen zu unendlichen Folgen unabhängiger Zufallsvariablen befassen. Die erste natürliche Frage ist, ob solche Folgen überhaupt in einem Wahrscheinlichkeitsraum definiert werden können.

3.1 Existenz

Definition: Seien $(\Omega, \mathcal{F}, P)$ ein W-Raum und $X_1, \dots, X_n$ Zufallsvariablen. Die *gemeinsame Verteilung* von $X_1, \dots, X_n$, bezeichnet mit $P_{X_1, \dots, X_n}$, ist die Verteilung von Zufallsvektor $(X_1, \dots, X_n)$ in $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, d.h.

$$P_{X_1, \dots, X_n}(B) = P((X_1, \dots, X_n) \in B) \quad B \in \mathcal{B}(\mathbb{R}^n).$$

- Seien $X_1, X_2, \dots$ Zufallsvariablen und P_n die gemeinsame Verteilung von $X_1, \dots, X_n$, dann $P_n(B) = P_{n+1}(B \times \mathbb{R})$ für alle $B \in \mathcal{B}(\mathbb{R}^n)$. Wichtiger ist aber die folgende Umkehrung.

Definition: Eine Folge von Verteilungen P_n auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $n \geq 1$, heißt *konsistent*, wenn

$$P_n(B) = P_{n+1}(B \times \mathbb{R}) \quad \text{für alle } B \in \mathcal{B}(\mathbb{R}^n).$$

Satz 34 (Erweiterungssatz von Kolmogorov). *Sei P_n eine konsistente Folge von Verteilungen auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Dann gibt es eindeutige Verteilung P auf $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, sodass*

$$P(\omega : (\omega_1, \dots, \omega_n) \in B) = P_n(B) \quad \text{für alle } n \in \mathbb{N} \text{ und } B \in \mathcal{B}(\mathbb{R}^n).$$

Mit anderen Worten, wenn $\pi_n : \mathbb{R}^\infty \rightarrow \mathbb{R}^n$ die Projektion auf ersten n Koordinaten ist, dann gilt $\pi_n \circ P = P_n$ für alle n .

Korollar 35. *Seien $Q_1, Q_2, \dots$ Verteilungen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Es gibt eine eindeutige Verteilung P auf $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, sodass $\pi_n \circ P = Q_1 \otimes \dots \otimes Q_n$ für alle n .*

Insbesondere, wenn $X_i : \mathbb{R}^\infty \rightarrow \mathbb{R}$ die Projektion auf i -te Koordinate ist, d.h. $X_i(\omega) = \omega_i$, dann sind die Zufallsvariablen $X_1, X_2, \dots$ auf dem W-Raum $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), P)$ unabhängig und für alle i gilt $P_{X_i} = Q_i$.

- Erweiterungssatz von Kolmogorov kann auf den Produktraum $(\prod_{i=1}^\infty \Omega_i, \otimes_{i=1}^\infty \mathcal{F}_i)$ erweitert werden, wobei Ω_i vollständige separable metrische Räume und $\mathcal{F}_i = \mathcal{B}(\Omega_i)$ sind.
- Um den Fall von allgemeinen messbaren Räumen $(\Omega_i, \mathcal{F}_i)$ zu behandeln, müssen die Voraussetzungen auf die Folge P_n verstärkt werden. Zum Beispiel gilt die Existenz von Produktmaß $\otimes_{i=1}^\infty Q_i$ auf $(\prod_{i=1}^\infty \Omega_i, \otimes_{i=1}^\infty \mathcal{F}_i)$.

3.2 0 – 1 Gesetze

Definition: Seien $(\Omega, \mathcal{F}, P)$ ein W-Raum und $A_1, A_2, \dots \in \mathcal{F}$. Das Ereignis, dass unendlich viele von A_n eintreten, wird mit $\limsup_n A_n$, oder auch $\{A_n \text{ u.o.}\}$, bezeichnet,

$$\limsup_n A_n = \{A_n \text{ u.o.}\} = \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m.$$

Das Ereignis, dass alle A_n mit der Ausnahme von endlich vielen treten ein, wird mit $\liminf_n A_n$ bezeichnet,

$$\liminf_n A_n = \bigcup_{n \geq 1} \bigcap_{m \geq n} A_m.$$

- Diese Notationen sind durch die Beziehungen

$$\mathbb{1}_{\limsup_n A_n}(\omega) = \limsup_n \mathbb{1}_{A_n}(\omega), \quad \mathbb{1}_{\liminf_n A_n}(\omega) = \liminf_n \mathbb{1}_{A_n}(\omega)$$

motiviert.

Lemma 36 (Lemma von Borel-Cantelli). *Seien $(\Omega, \mathcal{F}, P)$ ein W-Raum und $A_1, \dots \in \mathcal{F}$.*

1. *Wenn $\sum_{n \geq 1} P(A_n) < \infty$, dann $P(\limsup_n A_n) = 0$.*

- Mit anderen Worten, treten mit P -Wahrscheinlichkeit 1 nur endlich viele von $\{A_n\}_{n \geq 1}$.
- Die Umkehrung ist im Allgemeinen falsch.

2. *Wenn $A_1, \dots$ unabhängig sind und $\sum_{n \geq 1} P(A_n) = \infty$, dann $P(\limsup_n A_n) = 1$.*

- Mit anderen Worten, treten mit P -Wahrscheinlichkeit 1 unendlich viele von $\{A_n\}_{n \geq 1}$.

Definition: Seien $(\Omega, \mathcal{F}, P)$ ein W-Raum und $X_1, X_2, \dots$ Zufallsvariablen. Die asymptotische σ -Algebra der Folge X_n ist definiert durch

$$\mathcal{T} = \bigcap_{n \geq 1} \sigma(X_n, X_{n+1}, \dots).$$

- $\{\exists \lim_n X_n\}, \{\sum_k X_k \text{ konvergiert}\} \in \mathcal{T}$.

Satz 37 (0 – 1 Gesetz von Kolmogorov). Seien $(\Omega, \mathcal{F}, P)$ ein W-Raum und $X_1, X_2, \dots$ unabhängige Zufallsvariablen. Dann ist die asymptotische σ -Algebra $\mathcal{T}$ von X_n trivial, d.h.

$$P(A) \in \{0, 1\} \quad \text{für alle } A \in \mathcal{T}.$$

3.3 Konvergenzbegriffe

Definition: Seien $(\Omega, \mathcal{F}, P)$ ein W-Raum und $X, X_1, X_2, \dots$ Zufallsvariablen.

1. X_n konvergiert fast sicher gegen X , $X_n \xrightarrow{\text{f.s.}} X$, wenn

$$P(\omega : \lim_n X_n(\omega) = X(\omega)) = 1$$

2. X_n konvergiert in Wahrscheinlichkeit gegen X , $X_n \xrightarrow{P} X$, wenn

$$P(|X_n - X| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für jedes } \varepsilon > 0.$$

3. Sei $p \geq 1$. X_n konvergiert in p -ten Mittel (oder auch in L^p) gegen X , $X_n \xrightarrow{L^p} X$, wenn

$$E[|X_n|^p] < \infty, E[|X|^p] < \infty \text{ und } E[|X_n - X|^p] \xrightarrow{n \rightarrow \infty} 0.$$

Proposition 38 (Zusammenhänge). Seien $X, X_1, X_2, \dots$ Zufallsvariablen.

1. Wenn $X_n \xrightarrow{\text{f.s.}} X$, dann $X_n \xrightarrow{P} X$.
2. Wenn $X_n \xrightarrow{L^p} X$, dann $X_n \xrightarrow{P} X$.
3. Wenn $X_n \xrightarrow{P} X$, dann existiert eine Teilfolge $n(k)$, sodass $X_{n(k)} \xrightarrow{\text{f.s.}} X$.
4. $X_n \xrightarrow{P} X$ genau dann, wenn für jede Teilfolge $n(k)$ eine Teilfolge $k(m)$ existiert, sodass $X_{n(k(m))} \xrightarrow{\text{f.s.}} X$.

Lemma 39 (Eigenschaften). Seien X, Y, X_n, Y_n Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$.

1. Falls $X_n \rightarrow X$ und $X_n \rightarrow Y$ (f.s. oder in Wahrscheinlichkeit oder in p -Mittel), dann $X = Y$ f.s.
2. Falls $X_n \rightarrow X$ und $Y_n \rightarrow Y$ (f.s. oder in Wahrscheinlichkeit oder in p -Mittel), dann $aX_n + bY_n \rightarrow aX + bY$ (f.s. oder in Wahrscheinlichkeit oder in p -Mittel).

3. Falls $X_n \rightarrow X$ und $Y_n \rightarrow Y$ (f.s. oder in Wahrscheinlichkeit), dann $X_n Y_n \rightarrow XY$ (f.s. oder in Wahrscheinlichkeit).

Satz 40 (Schwaches Gesetz von der großen Zahlen). Seien $X_1, X_2, \dots$ unabhängige identisch verteilte integrierbare Zufallsvariablen. Dann

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} E[X_1].$$

Satz 41 (Starkes Gesetz von der großen Zahlen). Seien $X_1, X_2, \dots$ paarweise unabhängige identisch verteilte integrierbare Zufallsvariablen. Dann

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{f.s.} E[X_1].$$

- Wenn $X_1, X_2, \dots$ unabhängig und identisch verteilt sind mit $E[|X_1|] = \infty$, dann $P(\exists \lim_n \frac{1}{n} \sum_{i=1}^n X_i \in \mathbb{R}) = 0$.

3.4 Konvergenz stochastischer Reihe

Proposition 42 (Maximale Ungleichung von Kolmogorov). Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit $E[X_i^2] < \infty$ und $E[X_i] = 0$ für alle $1 \leq i \leq n$. Dann gilt für alle $x > 0$,

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq x\right) \leq \frac{\text{Var}(S_n)}{x^2},$$

wobei $S_k = X_1 + \dots + X_k$.

Satz 43 (Zweireihensatz von Kolmogorov). Seien X_n unabhängige Zufallsvariablen. Wenn

1. $\sum_{n=1}^{\infty} \text{Var}(X_n) < \infty$
2. $\sum_{n=1}^{\infty} E[X_n]$ konvergiert,

dann konvergiert die Reihe $\sum_{n=1}^{\infty} X_n$ fast sicher.

- (Lemma von Kronecker) Seien x_n und $a_n \uparrow \infty$ Folgen von reellen Zahlen. Falls die Reihe $\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ konvergiert, dann $\frac{1}{a_n} \sum_{k=1}^n x_k \rightarrow 0$.

Satz 44. (Dreireihensatz von Kolmogorov) Seien X_n unabhängige Zufallsvariablen, dann sind äquivalent:

1. Die Reihe $\sum_{n=1}^{\infty} X_n$ konvergiert fast sicher.
2. Es gibt $A > 0$ so dass für $Y_n = X_n \mathbb{1}_{|X_n| \leq A}$, konvergieren die folgende Reihe:
 - (a) $\sum_{n=1}^{\infty} P(|X_n| > A),$
 - (b) $\sum_{n=1}^{\infty} E[Y_n],$
 - (c) $\sum_{n=1}^{\infty} \text{Var}(Y_n).$
3. Die zweite Bedingung gilt für alle $A > 0$.

3.5 Schwache Konvergenz und Konvergenz in Verteilung

Definition: 1. Seien μ, μ_n Verteilungen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktionen F bzw. F_n .

- $F(x) = \mu((-\infty, x]), F_n(x) = \mu_n((-\infty, x])$.

μ_n konvergiert schwach gegen μ , $\mu_n \xrightarrow{s} \mu$ oder auch $\mu_n \Rightarrow \mu$, wenn

$$F_n(x) \longrightarrow F(x) \quad \text{für alle Stetigkeitspunkte } x \text{ von } F.$$

2. Seien X, X_n Zufallsvariablen mit Verteilungen μ bzw. μ_n (auf möglicherweise verschiedenen W-Räumen $(\Omega_n, \mathcal{F}_n, P_n)$). X_n konvergiert in Verteilung gegen X , $X_n \Rightarrow X$, wenn $\mu_n \xrightarrow{s} \mu$.

Beispiel: 1. (Dirac-Maßen) Seien $\mu_n(dx) = \delta_{\frac{1}{n}}(dx)$, $\mu(dx) = \delta_0(dx)$, dann $\mu_n \xrightarrow{s} \mu$.

2. (Satz von DeMoivre-Laplace) Seien $X_1, X_2, \dots$ unabhängig mit $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$ und $S_n = X_1 + \dots + X_n$. Dann konvergiert $\frac{S_n}{\sqrt{n}}$ in Verteilung gegen eine standardnormalverteilte Zufallsvariable.
3. (Poisson-Approximationssatz) Seien $X_n \sim B(n, p_n)$ und $np_n \rightarrow \lambda > 0$, dann konvergiert X_n in Verteilung gegen eine poissonverteilte Zufallsvariable mit Parameter λ .
4. Sei $X_p \sim \text{Geo}(p)$, dann konvergiert X_p in Verteilung gegen eine exponentialverteilte Zufallsvariable mit Parameter 1 für $p \rightarrow 0$.

Proposition 45 (Äquivalente Bedingungen der schwachen Konvergenz). *Die folgende Bedingungen sind äquivalent:*

1. $\mu_n \xrightarrow{s} \mu$.
2. Es existieren Zufallsvariablen X und X_n , $n \geq 1$, auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, so dass X_n μ_n -verteilt ist, X μ -verteilt ist, und $X_n \xrightarrow{P-f.s.} X$.
3. Für alle stetige beschränkte Funktionen $f: \mathbb{R} \rightarrow \mathbb{R}$,

$$\int_{\mathbb{R}} f d\mu_n \longrightarrow \int_{\mathbb{R}} f d\mu.$$

- Die dritte Bedingung lässt sich gut auf allgemeine Räume, z.B. $\mathbb{R}^d$, verallgemeinern und wird oft direkt als Definition der schwachen Konvergenz benutzt.

Die folgenden zwei Sätze liefern eine effiziente Methode, um die schwache Konvergenz zu beweisen.

Definition: Eine Folge von Verteilungen μ_n auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißt *straff*, wenn

$$\forall \varepsilon > 0 \quad \exists M > 0 \quad \text{s.d.} \quad \sup_n \mu_n((-\infty, -M] \cup [M, +\infty)) < \varepsilon.$$

Satz 46. Für jede straffe Folge von Verteilungen μ_n auf $\mathbb{R}$, existiert eine Teilfolge $n(k)$ und eine Verteilung μ auf $\mathbb{R}$, sodass $\mu_{n(k)} \xrightarrow{s} \mu$.

- (Lemma von Helly) Sei F_n eine Folge von Verteilungsfunktionen auf $\mathbb{R}$. Dann existiert eine Teilfolge $n(k)$ und eine rechtsstetige monoton steigende $F: \mathbb{R} \rightarrow [0, 1]$, sodass $F_{n(k)}(x) \rightarrow F(x)$ für alle Stetigkeitsstellen x von F .

Satz 47. Seien μ_n und μ Verteilungen auf $\mathbb{R}$. Existiert für jede Teilfolge $n(k)$ eine Teilfolge $k(m)$ mit $\mu_{n(k(m))} \xrightarrow{s} \mu$, dann $\mu_n \xrightarrow{s} \mu$.

3.6 Charakteristische Funktion

Definition: 1. Sei μ eine Verteilung auf $\mathbb{R}$. Die charakteristische Funktion von μ ist die Funktion $\varphi_\mu : \mathbb{R} \rightarrow \mathbb{C}$ definiert durch

$$\varphi_\mu(t) = \int_{\mathbb{R}} e^{itx} \mu(dx) \quad t \in \mathbb{R}.$$

2. Die charakteristische Funktion einer Zufallsvariable X ist die charakteristische Funktion der Verteilung von X :

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} P_X(dx) = E[e^{itX}].$$

- Der Erwartungswert von einer komplexwertige Zufallsvariable $Z = X + iY$ ist $E[Z] = E[X] + iE[Y]$.

Lemma 48 (Eigenschaften). Sei μ eine Verteilung auf $\mathbb{R}$. Dann gelten:

1. $\varphi_\mu(0) = 1$,
2. $|\varphi_\mu(t)| \leq 1$, für alle $t \in \mathbb{R}$,
3. $\varphi_\mu(-t) = \overline{\varphi_\mu(t)}$,
4. φ_μ ist gleichmäßig stetig auf $\mathbb{R}$,
5. falls $\int |x|^k d\mu < \infty$ für $k \in \mathbb{N}$, dann φ_μ k -mal stetig differenzierbar ist und

$$\varphi_\mu^{(l)}(t) = \int_{\mathbb{R}} (ix)^l e^{itx} d\mu, \quad \text{für alle } 0 \leq l \leq k.$$

Proposition 49 (Charakteristische Funktion von Summe unabhängiger Zufallsvariablen). Seien X und Y unabhängige Zufallsvariablen. Dann

$$\varphi_{X+Y} = \varphi_X \cdot \varphi_Y.$$

Beispiel: 1. Sei $\mu(dx) = \frac{1}{2}\delta_{-1}(dx) + \frac{1}{2}\delta_1(dx)$, dann $\varphi_\mu(t) = \cos t$.

2. Sei μ die Poisson-Verteilung mit Parameter λ , dann $\varphi_\mu(t) = e^{\lambda(e^{it}-1)}$.

3. Sei μ die Standardnormalverteilung, dann $\varphi_\mu(t) = e^{-\frac{t^2}{2}}$.

4. Sei μ die Normalverteilung mit Parametern m und σ^2 , dann $\varphi_\mu(t) = e^{itm - \frac{\sigma^2 t^2}{2}}$.

5. Sei μ die Standard-Cauchy-Verteilung, dann $\varphi_\mu(t) = e^{-|t|}$.

6. (Zusammengesetzte-Poisson-Verteilung) Seien $X_1, X_2, \dots$ unabhängige identisch verteilte Zufallsvariablen und $N \sim \text{Poi}(\lambda)$ unabhängig von $\{X_n\}_{n \geq 1}$. Dann heißt die Verteilung der Zufallsvariable

$$Y = \sum_{n=1}^N X_n$$

Zusammengesetzte-Poisson-Verteilung.

- Wenn $X_n = 1$ f.s., dann $Y = N \sim \text{Poi}(\lambda)$.

Es gilt $\varphi_Y(t) = \exp(\lambda(\varphi_{X_1}(t) - 1))$.

Die charakteristische Funktion bestimmt die Verteilung eindeutig:

Satz 50 (Eindeutigkeit). *Seien μ und ν Verteilungen auf $\mathbb{R}$ mit $\varphi_\mu = \varphi_\nu$. Dann $\mu = \nu$.*

- Zufallsvariable X ist symmetrisch genau dann, wenn $\varphi_X(t) \in \mathbb{R}$ für alle t .
- Seien $X \sim \text{Poi}(\lambda)$ und $Y \sim \text{Poi}(\mu)$ unabhängig, dann $X + Y \sim \text{Poi}(\lambda + \mu)$.
- Seien $X \sim N(m_1, \sigma_1^2)$ und $Y \sim N(m_2, \sigma_2^2)$ unabhängig, dann $X + Y \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Der nächste Satz ist das Hauptresultat über die punktweise Konvergenz charakteristischer Funktionen und ihrer Verbindung zur schwachen Konvergenz der jeweiligen Verteilungen. Er gibt eine nützliche Methode, um die schwache Konvergenz zu beweisen.

Satz 51 (Stetigkeitssatz). *Seien μ_n Verteilungen auf $\mathbb{R}$.*

1. Falls $\mu_n \xrightarrow{s} \mu$ für eine Verteilung μ , dann $\varphi_{\mu_n}(t) \rightarrow \varphi_\mu(t)$ für alle $t \in \mathbb{R}$.
2. Falls $\varphi_{\mu_n}(t) \rightarrow \varphi(t)$ für alle $t \in \mathbb{R}$, wobei φ eine stetige in 0 Funktion ist, dann existiert eine Verteilung μ auf $\mathbb{R}$, sodass $\varphi_\mu = \varphi$ und $\mu_n \xrightarrow{s} \mu$.

- Hier wird die folgende Eigenschaft von charakteristischen Funktionen verwendet: für alle $\delta > 0$,

$$\mu\left(\left\{x : |x| \geq \frac{2}{\delta}\right\}\right) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \varphi_\mu(t)) dt,$$

Satz 52. (Zentraler Grenzwertsatz) *Seien X_n unabhängige identisch verteilte Zufallsvariablen mit $E[X_1^2] < \infty$. Dann konvergiert für $n \rightarrow \infty$ die Folge*

$$\frac{S_n - mn}{\sigma\sqrt{n}}, \quad \text{wobei } S_n = X_1 + \dots + X_n, \quad m = E[X_1], \quad \sigma^2 = \text{Var}(X_1),$$

in Verteilung gegen eine standardnormalverteilte Zufallsvariable.