Skript zur Vorlesung

WAHRSCHEINLICHKEITS-THEORIE I

Jürgen Gärtner

unter Mitwirkung von
C. Klatt, J. Tams
und **S. Sturm**

Version Oktober 2007

Inhaltsverzeichnis

Vorbemerkungen

Dieser Abschnitt soll einige Vorbemerkungen beziehungsweise einleitende Worte enthalten, die am Schluß noch eingefügt werden. Im Moment werden diese Seiten noch frei gelassen.

Kapitel 1

Mathematische Modellierung von Zufallsexperimenten

1.1. Ereignisse und Wahrscheinlichkeiten. Axiomatischer Aufbau der Wahrscheinlichkeitstheorie

Ziel dieses Abschnitts ist die axiomatische Einführung der Begriffe „Ereignis“ und „Wahrscheinlichkeit“. Dieser axiomatische Aufbau der Wahrscheinlichkeitstheorie geht auf A.N. Kolmogorov (1933) zurück.

Die mehr philosophische Frage, was Zufall ist und ob es ihn überhaupt gibt, soll hier nicht erörtert werden. Außerdem wird die historische Entwicklung dieser mathematischen Disziplin über die letzten Jahrhunderte bis 1933 kaum Erwähnung finden können. Als Anregung in dieser Richtung sei die folgende Literatur angeführt:

- A. Rényi, *Briefe über die Wahrscheinlichkeit.* Birkhäuser 1969.
- I. Schneider (Hrsg.), *Die Entwicklung der Wahrscheinlichkeit von den Anfängen bis 1933.* Wiss. Buchges., Darmstadt 1988.
- G. Gigerenzer et al., *Das Reich des Zufalls, Wissen zwischen Wahrscheinlichkeiten, Häufigkeiten und Unschärfen.* Spektrum Akad. Verlag 1999.

Zu empfehlen ist auch die Originalarbeit von Kolmogorov:

- A.N. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Ergebnisse d. Math. 2, Heft 3. Springer-Verlag, Berlin 1933.

Unsere erste Aufgabe soll in der Erstellung eines mathematischen Modells zur Beschreibung von Zufallsexperimenten bestehen. Unter einem *Zufallsexperiment* kann man sich einen (gedanklichen oder tatsächlichen) Versuch vorstellen, der unter gleichen Bedingungen beliebig oft wiederholbar ist, dessen Ausgang sich aber nicht exakt vorhersagen läßt. D.h. bei Wiederholung des Versuchs gelangt man jedesmal zu einem anderen konkreten Ergebnis. Man denke dabei etwa an das Werfen von Münzen, das Würfeln oder die Anzahl der Fahrzeuge, die eine Kreuzung zwischen 12 und 13 Uhr befahren. Weitere Beispiele sind Anzahl und Größe der Datenpakete, die an einem Knoten in einem Computernetzwerk anstehen, das Ticken eines Geigerzählers bei radioaktiver Strahlung oder die irreguläre Anordnung der Siliziummoleküle in einem dotierten Halbleiter.

Die endlich oder unendlich vielen möglichen Ausgänge unseres Experiments nennen wir *Elementarereignisse*. Diese fassen wir zu einer Menge Ω zusammen, die wir *Menge der Elementarereignisse* nennen. Ein *Ereignis* setzt sich aus Elementarereignissen zusammen, ist also eine Teilmenge von Ω. Später werden wir sehen, dass es manchmal sinnvoll ist, nicht alle Teilmengen von Ω als Ereignisse anzusehen. Angenommen, die folgenden Objekte sind gegeben:

$A \subseteq \Omega$	Ereignis
$\omega \in \Omega$	Konkreter Ausgang des Experiments. (Bei Wiederholung des Versuchs kann dies jedesmal ein anderes Element von Ω sein.)

Diese Situation wird dann folgendermaßen interpretiert:

$\omega \in A$	Ereignis A ist eingetreten;
$\omega \notin A$	Ereignis A ist nicht eingetreten.

Sind $A, B, \ldots$ Ereignisse, so benutzt man die folgenden Sprechweisen für deren mengenalgebraische Verknüpfungen:

$\emptyset$	unmögliches Ereignis
Ω	sicheres Ereignis
$A^c := \Omega \setminus A$	Komplementärereignis zu A; tritt genau dann ein, wenn A nicht eintritt.
$A \cup B$	Ereignis, dass mindestens eines der beiden Ereignisse eintritt („A oder B“).
$A \cap B$	Ereignis, dass die Ereignisse A und B (gleichzeitig) eintreten. („A und B“).
$A \cap B = \emptyset$	A und B treten nicht gleichzeitig ein (sind unvereinbar, disjunkt).
$A \triangle B$	Ereignis, dass genau eines der beiden Ereignisse A und B eintritt.
$A \subseteq B$	A impliziert B.

Dabei bezeichnet $A \triangle B := (A \setminus B) \cup (B \setminus A)$ die symmetrische Differenz der Mengen A und B.

Um den Begriff der Wahrscheinlichkeit eines Ereignisses zu motivieren, beginnt man am Besten mit der relativen Häufigkeit eines Ereignisses. Hierzu betrachten wir N sich nicht beeinflussende Wiederholungen unseres Experiments mit den konkreten Versuchsausgängen $\omega_1, \ldots, \omega_N \in \Omega$. Die *relative Häufigkeit*

$h_N(A)$ eines Ereignisses $A \subseteq \Omega$ in diesen N Versuchen ist dann definiert durch

$$\begin{aligned} h_N(A) &:= h_N(A;\omega_1,\dots,\omega_N) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_A(\omega_i) \\ &= \frac{\text{Anzahl der günstigen Versuchsausgänge}}{\text{Gesamtzahl der Versuche}}. \end{aligned}$$

Dabei bezeichnet $\mathbb{1}_A$ die *Indikatorfunktion* des Ereignisses A:

$$\mathbb{1}_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A, \\ 0, & \text{falls } \omega \notin A. \end{cases}$$

Bei nochmaliger Ausführung der N Teilexperimente gelangt man im Allgemeinen zu anderen Ergebnissen $(\tilde{\omega}_1,\dots,\tilde{\omega}_n) \neq (\omega_1,\dots,\omega_N)$ und damit zu einem anderen Wert für die relative Häufigkeit $h_N(A)$. Die Erfahrung lehrt uns jedoch, dass für große N (große Zahl der Versuchswiederholungen) die Werte von $h_N(A)$ „kleine Fluktuationen" um eine Zahl $P(A)$ vollführen, die weder von N noch von den konkreten Versuchsausgängen abhängen:

$$h_N(A) \underset{N\to\infty}{\approx} P(A).$$

Diese Zahl ist die *Wahrscheinlichkeit des Ereignisses* A. Dies ist natürlich keine mathematische Definition und die Bestimmung von $P(A)$ etwa aus den dem Experiment zu Grunde liegenden Naturgesetzen ist auch nicht in erster Linie Aufgabe der Mathematik. Ausgehend von den Eigenschaften relativer Häufigkeiten wollen wir Regeln für Wahrscheinlichkeiten aufstellen, die wir dann als axiomatischen Ausgangspunkt für die Entwicklung unserer mathematischen Theorie nehmen.

Die relativen Häufigkeiten besitzen die folgenden Eigenschaften ($A, B \subseteq \Omega$ beliebige Ereignisse):

(i) $0 \le h_N(A) \le 1$;

(ii) $h_N(\emptyset) = 0$, $h_N(\Omega) = 1$;

(iii) $A \cap B = \emptyset \implies h_N(A \cup B) = h_N(A) + h_N(B)$.

Um die Gültigkeit von (iii) nachzuprüfen, bemerken wir zunächst, dass wegen $A \cap B = \emptyset$

$$\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B$$

gilt. Damit erhält man

$$\begin{aligned} h_N(A\cup B) &= \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{A\cup B}(\omega_i) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_A(\omega_i) + \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_B(\omega_i) \\ &= h_N(A) + h_N(B). \end{aligned}$$

Deshalb ist es naheliegend, von den Wahrscheinlichkeiten die analogen Eigenschaften zu fordern:

(i) $0 \leq P(A) \leq 1$;

(ii) $P(\emptyset) = 0$, $P(\Omega) = 1$;

(iii) $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B) \qquad$ (*Additivität*).

Folgende Sprechweisen haben sich eingebürgert:

$P(A) = 0$ Ereignis A tritt fast nie ein;
$P(A) = 1$ Ereignis A tritt fast sicher ein.

($P(A) = 0$ bzw. $P(A) = 1$ kann auch für gewisse Ereignisse $A \neq \emptyset$ bzw. $A \neq \Omega$ gelten.)

Beispiel 1.1. (Idealer Würfel)
Mögliche Ausgänge des Experiments „Würfeln" sind die Augenzahlen $1, \ldots, 6$, die die Elementarereignisse darstellen. Deshalb ist

$$\Omega = \{1, 2, \ldots, 6\}.$$

Alle Mengen $A \subseteq \Omega$ sind als Ereignisse zugelassen. Für einen *idealen* Würfel erwartet man, dass

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{6\})$$

ist. Auf Grund der Eigenschaften (ii) und (iii) von P muss

$$\sum_{i=1}^{6} P(\{i\}) = P(\Omega) = 1$$

sein. Zusammen ergibt dies

$$P(\{i\}) = \frac{1}{6} \qquad \text{für } i \in \Omega.$$

Nochmalige Anwendung der Additivität liefert schließlich

$$P(A) = \frac{|A|}{6}, \qquad A \subseteq \Omega.$$

Dabei steht $|A|$ für die Anzahl der Elemente der Menge A. Man prüft leicht nach, dass die so eingeführten Wahrscheinlichkeiten die obigen Forderungen (i)–(iii) tatsächlich erfüllen. Das Ereignis, eine ungerade Augenzahl zu würfeln, ist $A = \{1, 3, 5\}$ und hat die Wahrscheinlichkeit $P(A) = |A|/6 = 1/2$.

Aufbauend auf diesen heuristischen Vorbetrachtungen wollen wir nun die präzisen mathematischen Definitionen einführen. Gegeben sei eine nichtleere Menge Ω, die wir *Raum der Elementarereignisse* nennen. Mit $\mathfrak{P}(\Omega)$ bezeichnen wir die Potenzmenge von Ω, d.h. das System aller Teilmengen von Ω (einschließlich $\emptyset$ und Ω). Die Ereignisse bilden ein System $\mathfrak{F}$ von Teilmengen von Ω, $\mathfrak{F} \subseteq \mathfrak{P}(\Omega)$. Unsere erste axiomatische Forderung besteht darin, dass endliche *und abzählbar unendliche* mengenalgebraische Operationen mit Ereignissen wieder zu Ereignissen führen, d.h. dass das Mengensystem $\mathfrak{F}$ bezüglich solcher Operationen „abgeschlossen" ist. Dies führt uns auf den Begriff der σ-Algebra.

Definition 1.2. Ein Mengensystem $\mathfrak{F} \subseteq \mathfrak{P}(\Omega)$ heißt *σ-Algebra* (von Ereignissen in Ω), falls folgendes gilt:

(i) $\emptyset \in \mathfrak{F}$;

(ii) $A \in \mathfrak{F} \Longrightarrow A^c \in \mathfrak{F}$;

(iii) $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{F} \Longrightarrow \bigcup_n A_n \in \mathfrak{F}$.

Aus (i) und (ii) folgt insbesondere $\Omega \in \mathfrak{F}$. Das folgende Lemma zeigt, dass eine σ-Algebra tatsächlich bezüglich aller denkbaren endlichen und abzählbar unendlichen Mengenoperationen abgeschlossen ist.

Lemma 1.3. *Ist $\mathfrak{F}$ eine σ-Algebra von Ereignissen in Ω, so gilt*

a) $A, B \in \mathfrak{F} \Longrightarrow A \cup B, A \cap B, A \setminus B, A \triangle B \in \mathfrak{F}$;

b) $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{F} \Longrightarrow \bigcap_n A_n \in \mathfrak{F}$.

Beweis. Aus den Axiomen (i) und (iii) einer σ-Algebra folgt für $A_1 := A$, $A_2 := B$, $A_n := \emptyset$ für $n \geq 3$:

$$A \cup B = \bigcup_n A_n \in \mathfrak{F}, \qquad \text{falls } A, B \in \mathfrak{F}.$$

Also ist $\mathfrak{F}$ bezüglich der Vereinigung zweier Mengen abgeschlossen. Die Abgeschlossenheit bezüglich des Durchschnitts ergibt sich durch Komplementbildung (Eigenschaft (ii) einer σ-Algebra) aus der gerade bewiesenen Abgeschlossenheit in Bezug auf Vereinigungen, denn

$$A \cap B = (A^c \cup B^c)^c \,.$$

Restliche Aussagen: *Übungsaufgabe.* □

Bemerkung 1.4. Ist $\mathfrak{F}$ eine σ-Algebra in Ω, so nennt man das Paar $(\Omega, \mathfrak{F})$ auch *meßbaren Raum*.

Jedem Ereignis soll nun eine Wahrscheinlichkeit zugeordnet werden. Von den Wahrscheinlichkeiten fordern wir nicht nur die endliche Additivität, sondern die Additivität bezüglich abzählbar vieler Ereignisse. Ohne diese erweiterte Forderung würde man nur eine sehr „armselige“ Theorie erhalten.

Definition 1.5. Sei $(\Omega, \mathfrak{F})$ ein meßbarer Raum. Eine Abbildung $P\colon \mathfrak{F} \to [0,1]$ heißt *Wahrscheinlichkeitsmaß* auf $\mathfrak{F}$, falls

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

(ii) $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{F}$, $A_k \cap A_l = \emptyset$ für $k \neq l \Longrightarrow P(\bigcup_n A_n) = \sum_n P(A_n)$ (*σ-Additivität*).

Folgerung 1.6.

a) $P(\emptyset) = 0$;

b) $A, B \in \mathfrak{F}$ *und* $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$;

c) $P(A^c) = 1 - P(A)$ *für alle* $A \in \mathfrak{F}$;

d) $A, B \in \mathfrak{F}$ *und* $A \subseteq B \implies P(A) \leq P(B)$;

e) $A \in \mathfrak{F}$, $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{F}$ *(nicht notwendigerweise paarweise disjunkt),*
$A \subseteq \bigcup_n A_n \implies P(A) \leq \sum_n P(A_n)$
(Subadditivität).

Beweis. a) Wegen $\emptyset = \bigcup_n \emptyset$ folgt aus der σ-Additivität von P, dass

$$P(\emptyset) = \sum_n P(\emptyset)$$

ist. Dies ist aber nur für $P(\emptyset) = 0$ erfüllt.

b) Die Behauptung folgt aus der σ-Additivität von P und a), wenn man

$$A_1 := A,\ A_2 := B,\ A_n = \emptyset \text{ für } n \geq 3$$

setzt.

c) Wegen $\Omega = A \cup A^c$ und $A \cap A^c = \emptyset$ erhalten wir mit b) und Axiom (i) von P:

$$1 = P(\Omega) = P(A) + P(A^c).$$

d), e): *Übungsaufgabe.* □

Die Axiome einer σ-Algebra und eines Wahrscheinlichkeitsmaßes bilden den Ausgangspunkt für alle wahrscheinlichkeitstheoretischen Betrachtungen. Diese grundlegenden Objekte fassen wir im Begriff des Wahrscheinlichkeitsraumes zusammen.

Definition 1.7. Sei Ω eine nichtleere Menge, $\mathfrak{F}$ eine σ-Algebra in Ω und P ein Wahrscheinlichkeitsmaß auf $\mathfrak{F}$. Dann heißt das Tripel $(\Omega, \mathfrak{F}, P)$ *Wahrscheinlichkeitsraum.*

Wir wollen nun noch einige weitere Eigenschaften von Wahrscheinlichkeitsmaßen herleiten.

Lemma 1.8. *Sei* $(\Omega, \mathfrak{F}, P)$ *ein Wahrscheinlichkeitsraum.*

a) Für beliebige $A, B \in \mathfrak{F}$ *gilt*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

b) Für beliebige $n \geq 2$ und $A_1, \ldots, A_n \in \mathfrak{F}$ ist

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) \tag{1.1}$$

(Formel von Sylvester).

Beweis. Wir bemerken zunächst, dass für beliebige $A, B \in \mathfrak{F}$ mit $A \subseteq B$ die Beziehung

$$P(B) = P(A) + P(B \setminus A)$$

gilt. Wegen $B = A \cup (B \setminus A)$ und $A \cap (B \setminus A) = \emptyset$ folgt dies aus der Additivität von P.

a) Wir stellen $A \cup B$ als disjunkte Vereinigung von Ereignissen dar:

$$A \cup B = A \cup [B \setminus (A \cap B)].$$

Deshalb können wir wieder die Additivität von P ausnutzen. Wegen $A \cap B \subseteq B$ können wir außerdem auf das Ereignis in den eckigen Klammern unsere obige Bemerkung anwenden. Wir erhalten

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \setminus (A \cap B)) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

b) Wir benutzen Induktion nach n. Für $n = 2$ stimmt unsere Aussage mit der gerade bewiesenen Behauptung a) überein. Angenommen, die Formel (1.1) gilt für ein $n \geq 2$ (und alle $A_1, \ldots, A_n \in \mathfrak{F}$). Dann erhalten wir unter Benutzung von a)

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right).$$

Wenden wir die Induktionsvoraussetzung auf die erste und dritte Wahrscheinlichkeit auf der rechten Seite an, so können wir wie folgt fortfahren:

$$\begin{aligned} &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) \\ &\quad + P(A_{n+1}) \\ &\quad - \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P\left(A_{i_1} \cap \cdots \cap A_{i_k} \cap A_{n+1}\right) \\ &= \sum_{k=1}^{n+1} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n+1} P\left(A_{i_1} \cap \cdots \cap A_{i_k}\right). \end{aligned}$$

Die Gültigkeit der letzten Gleichheit erkennt man, wenn man die letzte Summe in zwei Teilsummen aufspaltet, indem man einmal nur über $1 \leq i_1 < \cdots < i_k \leq n$ summiert und zum anderen die Summe über alle Indices mit $1 \leq i_1 < \cdots < i_{k-1} < i_k = n+1$ erstreckt und anschließend im zweiten Teil eine Indexverschiebung von k um 1 durchführt. $\square$

Wir werden später noch folgendes Ergebnis aus der Maßtheorie benötigen, das wir ohne Beweis angeben.

Satz 1.9. *(Stetigkeit von Wahrscheinlichkeitsmaßen)*
Gegeben seien ein Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, P)$ *und eine Folge* (A_n) *von Ereignissen aus* $\mathfrak{F}$.

a) Falls $A_n \uparrow A$ *(d.h.* $A_1 \subseteq A_2 \subseteq \ldots$ *und* $\bigcup_n A_n = A$*), so folgt*

$$P(A_n) \uparrow P(A).$$

b) Falls $A_n \downarrow A$ *(d.h.* $A_1 \supseteq A_2 \supseteq \ldots$ *und* $\bigcap_n A_n = A$*), so folgt*

$$P(A_n) \downarrow P(A).$$

1.2. Diskrete und stetige Wahrscheinlichkeitsräume

Definition 1.10. Ist Ω endlich oder abzählbar, $\mathfrak{F} = \mathfrak{P}(\Omega)$ und P ein Wahrscheinlichkeitsmaß auf $\mathfrak{F}$, so heißt $(\Omega, \mathfrak{F}, P)$ *diskreter Wahrscheinlichkeitsraum* (Kurzbezeichnung: (Ω, P)). P heißt dann *diskretes Wahrscheinlichkeitsmaß.*

Ist Ω höchstens abzählbar, so ist $\mathfrak{P}(\Omega)$ die einzige σ-Algebra in Ω, die alle einpunktigen Mengen enthält. Deshalb ist es natürlich, in der obigen Definition als σ-Algebra die Potenzmenge von Ω zu nehmen.

Unter den *Einzelwahrscheinlichkeiten* $p(\omega)$ versteht man die Wahrscheinlichkeiten der Elementarereignisse ω:

$$p(\omega) := P(\{\omega\}), \qquad \omega \in \Omega.$$

Diese besitzen die folgenden Eigenschaften:

(i) $p(\omega) \geq 0$ für alle $\omega \in \Omega$;

(ii) $\sum_{\omega \in \Omega} p(\omega) = 1$.

Wir merken an, dass die letzte Summe nicht von der Summationsreihenfolge abhängt, da alle Summanden nichtnegativ sind. Die Eigenschaften (i) und (ii) charakterisieren ein diskretes Wahrscheinlichkeitsmaß vollständig.

Proposition 1.11. *Ist Ω höchstens abzählbar und $p\colon \Omega \to \mathbb{R}$ eine Funktion mit den Eigenschaften (i) und (ii), so wird durch*

$$P(A) := \sum_{\omega \in A} p(\omega), \qquad A \subseteq \Omega,$$

ein Wahrscheinlichkeitsmaß P auf $\mathfrak{P}(\Omega)$ definiert.

Beweis. Dass $0 \leq P(A) \leq 1$ und $P(\Omega) = 1$ ist, folgt unmittelbar aus (i) und (ii). Um die σ-Additivität von P nachzuweisen, sei eine beliebige Folge (A_n) paarweise disjunkter Ereignisse gegeben. Dann ist

$$\mathbb{1}_{\bigcup_n A_n} = \sum_n \mathbb{1}_{A_n}.$$

Deshalb erhalten wir

$$\begin{aligned}
P\left(\bigcup_n A_n\right) &= \sum_{\omega \in \bigcup_n A_n} p(\omega) = \sum_\omega p(\omega) \mathbb{1}_{\bigcup_n A_n}(\omega) \\
&= \sum_\omega p(\omega) \sum_n \mathbb{1}_{A_n}(\omega) \\
&= \sum_n \sum_\omega p(\omega) \mathbb{1}_{A_n}(\omega) = \sum_n \sum_{\omega \in A_n} p(\omega) \\
&= \sum_n P(A_n).
\end{aligned}$$

Dabei wurde entscheidend ausgenutzt, dass für nichtnegative Summanden die Summationsreihenfolge vertauscht werden kann. □

Wir betrachten nun einen wichtigen Spezialfall.

Definition 1.12. Ein *Laplacescher Wahrscheinlichkeitsraum* ist ein diskreter Wahrscheinlichkeitsraum (Ω, P) mit $|\Omega| =: n < \infty$ und identischen Einzelwahrscheinlichkeiten

$$p(\omega) = \frac{1}{n}, \qquad \omega \in \Omega.$$

In diesem Falle nennt man P *Gleichverteilung* auf Ω.

Eine Gleichverteilung P ist somit durch

$$P(A) = \frac{|A|}{|\Omega|}, \qquad A \in \mathfrak{P}(\Omega),$$

gegeben.

Beispiel 1.13. (Mehrmaliges Würfeln)
Wie groß ist die Wahrscheinlichkeit, dass beim n-ten Wurf zum ersten Mal eine "6" auftritt? Als Raum der Elementarereignisse kann man

$$\Omega := \{1, \dots, 6\}^n = \{(i_1, \dots, i_n) \colon 1 \le i_1, \dots, i_n \le 6\}$$

nehmen. Da alle Versuchsausgänge $(i_1, \dots, i_n)$ gleichberechtigt auftreten, handelt es sich um einen Laplaceschen Wahrscheinlichkeitsraum (Ω, P) mit

$$|\Omega| = 6^n \quad \text{und} \quad p(\omega) \equiv 6^{-n}.$$

Das Ereignis, dass beim n-ten Wurf zum ersten Mal eine "6" auftritt, hat die Gestalt

$$A = \{(i_1, \dots, i_{n-1}, 6) \colon 1 \le i_1, \dots, i_{n-1} \le 5\}\,.$$

Da $|A| = 5^{n-1}$ ist, erhält man

$$P(A) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}.$$

Man könnte das Experiment auch anders gestalten, indem man solange würfelt, bis zum ersten Mal eine "6" geworfen wurde. Nimmt man als Elementarereignisse die Zahl der erforderlichen Würfe, so erhält man einen diskreten Wahrscheinlichkeitsraum (Ω, P) mit $\Omega = \mathbb{N} = \{1, 2, \dots\}$ und Einzelwahrscheinlichkeiten

$$p(n) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}, \qquad n \in \mathbb{N}.$$

Man sieht sofort, dass dies tatsächlich Einzelwahrscheinlichkeiten sind. (A priori war das nicht klar, da wir ja nicht von vornherein den Fall ausschließen konnten, dass in einer Folge von Würfen nie eine "6" geworfen wird.)

Beispiel 1.14. (Zusammentreffen von Geburtstagen)
Angenommen, n Personen treffen sich „auf gut Glück". Wie groß ist die Wahrscheinlichkeit, dass mindestens zwei dieser Personen am gleichen Tag Geburtstag haben? Die Modellierung geschieht wieder mit einem Laplaceschen Wahrscheinlichkeitsraum (Ω, P), wobei

$$\Omega = \{1, \dots, 365\}^n, \quad |\Omega| = 365^n \quad \text{und} \quad p(\omega) \equiv \frac{1}{365^n}$$

ist. Mit A bezeichnen wir das Ereignis, dass zwei der n Personen am gleichen Tag Geburtstag haben. Es ist einfacher, zunächst die Wahrscheinlichkeit des Komplementärereignisses

$$A^c = \{(i_1, \dots, i_n) \in \Omega \colon i_k \neq i_l \text{ für } k \neq l\}$$

zu berechnen. Dies ist das Ereignis, dass die Geburtstage alle auf verschiedene Tage fallen. Wir erhalten

$$\begin{aligned} |A^c| &= 365 \cdot 364 \dots (365 - n + 1) \\ &= \frac{365!}{(365-n)!} = \binom{365}{n} n!. \end{aligned}$$

(Für die erste Person gibt es 365 Möglichkeiten. Dann bleiben noch 364 Möglichkeiten für die zweite Person, usw.) Damit ist

$$P(A^c) = \frac{\binom{365}{n} n!}{365^n},$$

und folglich

$$P(A) = 1 - \frac{\binom{365}{n} n!}{365^n}.$$

Man erhält die folgende Tabelle:

n	10	20	22	23	30	40	50	60
P(A)	0,12	0,41	0,48	0,51	0,71	0,89	0,97	0,99

Geometrische Wahrscheinlichkeiten und Lebesgue-Maß

Wir haben bisher nur diskrete Wahrscheinlichkeitsräume (d.h. mit höchstens abzählbar vielen Zuständen) betrachtet. Um zum kontinuierlichen Fall überzugehen, stellen wir uns folgendes *Nadelexperiment* vor: Wirft man eine Nadel mit der Spitze auf gut Glück auf das Einheitsintervall $[0,1]$, so setzt man $\Omega = [0,1]$ und fordert

$$P([a,b]) = b - a \qquad \text{für } [a,b] \subseteq \Omega. \tag{1.2}$$

Als σ-Algebra der Ereignisse nimmt man die kleinste σ-Algebra in $[0,1]$, die alle Intervalle $[a,b]$ enthält. Diese σ-Algebra heißt *Borel-σ-Algebra* und wird mit $\mathfrak{B}_{[0,1]}$ bezeichnet. In der Maßtheorie wird gezeigt, dass genau ein Wahrscheinlichkeitsmaß P auf $\mathfrak{B}_{[0,1]}$ existiert, das (1.2) erfüllt.

Bemerkung 1.15. (ohne Beweis)
1. $\mathfrak{B}_{[0,1]}$ enthält alle offenen und abgeschlossenen Teilmengen von $[0,1]$.
2. P lässt sich nicht zu einem Wahrscheinlichkeitsmaß auf $\mathfrak{P}(\Omega)$ fortsetzen.

Allgemeiner bezeichnet man mit $\mathfrak{B}^d$ die *σ-Algebra der Borelmengen des* $\mathbb{R}^d$. Darunter versteht man die kleinste σ-Algebra, die alle achsenparallelen Quader (und damit auch alle offenen und abgeschlossenen Teilmengen) des $\mathbb{R}^d$ enthält. Weiterhin sei λ das d-dimensionale Lebesgue-Maß, d.h. diejenige Mengenfunktion $\lambda\colon \mathfrak{B}^d \longrightarrow [0,\infty]$, die durch folgende Eigenschaften charakterisiert wird:

(i) $\lambda(\emptyset) = 0$;

(ii) für jede Folge (A_n) paarweise disjunkter Mengen aus $\mathfrak{B}^d$ gilt

$$\lambda\left(\bigcup_n A_n\right) = \sum_n \lambda(A_n) \qquad (\sigma\text{-Additivität});$$

(iii)
$$\lambda([a_1, b_1] \times \cdots \times [a_d, b_d]) = (b_1 - a_1) \cdots (b_d - a_d).$$

(Anschaulich: λ ordnet allen achsenparallelen Quadern und wegen (ii) auch allen „hinreichend regulären“ Mengen ihr Volumen zu.)

Für jedes $\Omega \in \mathfrak{B}^d$ mit $\lambda(\Omega) < \infty$ kann man wie folgt einen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{B}_\Omega, P)$ definieren:

$$\begin{aligned} \mathfrak{B}_\Omega &:= \{A \in \mathfrak{B}^d \colon A \subseteq \Omega\}, \\ P_\Omega(A) &:= \frac{\lambda(A)}{\lambda(\Omega)}, \qquad A \in \mathfrak{B}_\Omega. \end{aligned}$$

Man prüft leicht nach, dass $\mathfrak{B}_\Omega$ tatsächlich eine σ-Algebra und P ein Wahrscheinlichkeitsmaß auf $\mathfrak{B}_\Omega$ (die sogenannte *„Gleichverteilung auf Ω“*) ist.
$(\Omega, \mathfrak{B}_\Omega, P_\Omega)$ dient als mathematisches Modell für das Experiment, dass man einen Punkt „auf gut Glück“ aus Ω auswählt.

Beispiel 1.16. (Bertrandsches Paradoxon)
Wie groß ist die Wahrscheinlichkeit, dass die Länge einer Kreissehne größer als die Länge eines einbeschriebenen gleichseitigen Dreiecks ist? (Radius 1, Seitenlänge $\sqrt{3}$, Länge der Sehne L.)

Erstes Experiment:
Durchmesser auf gut Glück durch Kreismittelpunkt legen, Punkt auf Durchmesser auf gut Glück auswählen, Sehne durch diesen Punkt senkrecht zum Durchmesser zeichnen:

$$\begin{aligned} \Omega &= \{(\varphi, x) \colon 0 \le \varphi < \pi, -1 \le x \le 1\} \\ &= [0, \pi) \times [-1, 1], \\ P &= \text{Gleichverteilung auf } \Omega. \end{aligned}$$

Länge der Sehne bei Versuchsausgang (φ, x):

$$\begin{aligned} L &= L(\varphi, x) = 2\sqrt{1 - x^2}, \\ A &= \{(\varphi, x) \in \Omega : L(\varphi, x) > \sqrt{3}\} \\ &= \{(\varphi, x) \in \Omega : |x| < 1/2\} \qquad \text{Ereignis, dass } L > \sqrt{3}, \\ P(A) &= \frac{\lambda(A)}{\lambda(\Omega)} = \frac{\pi}{\pi \cdot 2} = \frac{1}{2}. \end{aligned}$$

Dieses Experiment dient uns nur zur Illustration der Modellbildung bei exakter Berücksichtigung des Modells. Da die Länge L der zufälligen Sehne nur vom Punkt x auf dem Durchmesser und nicht vom Winkel φ abhängt, hätte man auch o.B.d.A. einen Durchmesser (z.B. die x-Achse) fixieren und für P die

Gleichverteilung auf $[-1, 1]$ wählen können. Als Ergebnis hätten wir die gleiche Wahrscheinlichkeit erhalten.

Zweites Experiment:
Zwei Punkte auf Kreisumfang auf gut Glück auswählen und durch eine Sehne verbinden:

$$\begin{aligned}
\Omega &= \{(\varphi_1, \varphi_2)\colon 0 \leq \varphi_1, \varphi_2 < 2\pi\}, \\
&= [0, 2\pi)^2, \\
\frac{L}{2} &= \left|\sin \frac{\varphi_1 - \varphi_2}{2}\right|, \\
A &= \{(\varphi_1, \varphi_2) \in \Omega\colon L(\varphi_1, \varphi_2) > \sqrt{3}\}, \\
&= \left\{(\varphi_1, \varphi_2) \in \Omega\colon \frac{2}{6}\pi < \left|\frac{\varphi_2 - \varphi_1}{2}\right| < \frac{4}{6}\pi\right\}, \\
P &= \text{Gleichverteilung auf } \Omega, \\
P(A) &= \frac{\lambda(A)}{\lambda(\Omega)} = \frac{\frac{2}{3}\pi \cdot 2\pi}{(2\pi)^2} = \frac{1}{3}.
\end{aligned}$$

Auch hier kann man erwarten, dass eine Fixierung des ersten Punktes auf dem Kreisumfang (z.B. $(1, 0)$) und das Bestimmen des zweiten Punktes auf gut Glück zur gleichen Wahrscheinlichkeit führt. Dies ist auch anschaulich-geometrisch plausibel. Bemerkenswerter ist jedoch, dass die Wahrscheinlichkeit des Ereignisses A im ersten und zweiten Experiment *verschieden* sind. Die ursprüngliche Aufgabe, eine Sehne „zufällig“ auszuwählen, war also nicht korrekt gestellt. Man muss präzisieren, *wie* dies geschehen soll. In beiden Experimenten wird aber auf ganz unterschiedliche Weise vorgegangen!

Kapitel 2

Bedingte Wahrscheinlichkeiten und Unabhängigkeit von Ereignissen

Sei (Ω, P) ein Laplacescher Wahrscheinlichkeitsraum. Sind A und B beliebige Ereignisse, so interessieren wir uns für die Wahrscheinlichkeit von A, wenn wir bereits wissen, dass B eingetreten ist. Eine Skizze suggeriert hierfür folgenden Ansatz:

$$\frac{|A \cap B|}{|B|} = \frac{|A \cap B|}{|\Omega|} \Big/ \frac{|B|}{|\Omega|} = \frac{P(A \cap B)}{P(B)}.$$

Dies führt auf folgende Definition (durch Abstraktion auf beliebige Wahrscheinlichkeitsräume).

Definition 2.1. Seien $(\Omega, \mathfrak{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum und $B \in \mathfrak{F}$ mit $P(B) > 0$. Dann heißt

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

bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung (Hypothese) B.

Beispiel 2.2. Zweimal würfeln:
Wie groß ist die Wahrscheinlichkeit, dass beim zweiten Wurf eine „3“ geworfen wurde, wenn bereits bekannt ist, dass die Augensumme „5“ ist?

$$\begin{aligned}
|\Omega| &= 36, \\
A &= \{(1,3),(2,3),\ldots,(6,3)\}, && \text{"3" beim 2. Wurf,} \\
B &= \{(1,4),(2,3),(3,2),(4,1)\}, && \text{Augensumme 5,} \\
A\cap B &= \{(2,3)\}, \\
P(A|B) &= \frac{P(A\cap B)}{P(B)} = \frac{1/36}{4/36} = \frac{1}{4}.
\end{aligned}$$

Proposition 2.3. *Seien $(\Omega, \mathfrak{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum, $B \in \mathfrak{F}$ und $P(B) > 0$. Dann ist $P(\cdot\,|B)$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{F})$.*

Beweis. Für ein beliebiges Ereignis $A \in \mathfrak{F}$ ist

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

wohldefiniert, und es gilt $0 \le P(A|B) \le 1$, da wegen $A\cap B \subseteq B$ auch $P(A\cap B) \le P(B)$ ist. Um die σ-Additivität zu zeigen, sei $(A_n)_{n\in\mathbb{N}}$ eine beliebige Folge von Ereignissen aus $\mathfrak{F}$ mit $A_k \cap A_l = \emptyset$ für $k \neq l$. Dann ist

$$\begin{aligned}
P\left(\bigcup_n A_n \middle| B\right) &= \frac{P\left(\bigcup_n (A_n \cap B)\right)}{P(B)} \\
&= \frac{\sum_n P(A_n \cap B)}{P(B)} \\
&= \sum_n P(A_n|B),
\end{aligned}$$

wobei in der vorletzten Zeile die σ-Additivität von P ausgenutzt wurde. □

Weitere Rechenregeln:

$$\begin{aligned}
A\cap B = \emptyset &\;\Rightarrow\; P(A|B) = 0, \\
A \supseteq B &\;\Rightarrow\; P(A|B) = 1.
\end{aligned}$$

Satz 2.4. *(Formel von der totalen Wahrscheinlichkeit)*
Sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum. Weiterhin sei $(B_n) \subset \mathfrak{F}$ eine höchstens abzählbare Zerlegung von Ω in paarweise disjunkte Ereignisse, d.h. $\Omega = \bigcup_n B_n$ und $B_k \cap B_l = \emptyset$ für $k \neq l$. Dann gilt für alle $A \in \mathfrak{F}$:

$$P(A) = \sum_n P(A|B_n)P(B_n).$$

(Ist $P(B_n) = 0$, so wird der Summand $P(A|B_n)P(B_n)$ gleich Null gesetzt.)

Beweis. (Falls $P(B_n) > 0$ für alle n.)
Das Ereignis A lässt sich wie folgt als disjunkte Vereinigung darstellen:

$$A = \bigcup_n (A \cap B_n).$$

Deshalb gilt

$$\begin{aligned} P(A) &= \sum_n P(A \cap B_n) \\ &= \sum_n \frac{P(A \cap B_n)}{P(B_n)} P(B_n) \\ &= \sum_n P(A|B_n) P(B_n). \end{aligned}$$

□

Beispiel 2.5. (Binäre Signalübertragung mit Fehler)
Es wird eines der beiden Signale “0“ und “1“ gesendet und geschaut, welches dieser Signale beim Empfänger ankommt. Für die Modellierung nehmen wir

$$\Omega = \{(0,0),(0,1),(1,0),(1,1)\},$$

wobei die erste Komponente das jeweils gesendete und die zweite das empfangene Signal darstellt.

Bekannt sei:
1/3 der gesendeten Signale sind “1“-en.
Die “1“ wird mit Wahrscheinlichkeit 1/10 falsch übertragen.
Die “0“ wird mit Wahrscheinlichkeit 1/5 falsch übertragen.
Gesucht: Fehlerwahrscheinlichkeit der Übertragung.

$$\begin{aligned} F &= \{(1,0),(0,1)\} \quad \text{fehlerhafte Übertragung,} \\ S_0 &= \{(0,0),(0,1)\} \quad \text{“0“ wird gesendet,} \\ S_1 &= \{(1,0),(1,1)\} \quad \text{“1“ wird gesendet,} \\ S_0 \cup S_1 &= \Omega, \\ S_0 \cap S_1 &= \emptyset, \\ P(S_0) &= \frac{2}{3}, \quad P(S_1) = \frac{1}{3}, \\ P(F|S_0) &= \frac{1}{5}, \quad P(F|S_1) = \frac{1}{10}. \end{aligned}$$

Damit erhalten wir

$$\begin{aligned} P(F) &= P(F|S_0)P(S_0) + P(F|S_1)P(S_1). \\ &= \frac{1}{5} \cdot \frac{2}{3} + \frac{1}{10} \cdot \frac{1}{3} = \frac{5}{30} = \frac{1}{6}. \end{aligned}$$

Satz 2.6. *(Bayessche Formel)*
Die Voraussetzungen des vorangegangenen Satzes seien erfüllt. Zusätzlich gelte $P(A) > 0$. Dann ist

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{\sum\limits_j P(A|B_j)P(B_j)}$$

für alle n.

Beweis. Wegen

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

folgt die Behauptung, wenn man $P(A)$ im Nenner mit Hilfe der Formel der totalen Wahrscheinlichkeit ausdrückt. □

Beispiel 2.7. Signalübertragung (Fortsetzung)
Es wurde eine "1" empfangen. Wie groß ist die Wahrscheinlichkeit, dass eine "1" gesendet wurde?

$$E_1 = \{(0,1),(1,1)\} \qquad \text{"1" empfangen,}$$

Man beachte: $E_1 \cap S_1 = F^c \cap S_1$ und $E_1 \cap S_0 = F \cap S_0$. Damit folgt:

$$\begin{aligned} P(S_1|E_1) &= \frac{P(E_1|S_1)P(S_1)}{P(E_1|S_1)P(S_1) + P(E_1|S_0)P(S_0)} \\ &= \frac{[1 - P(F|S_1)]P(S_1)}{[1 - P(F|S_1)]P(S_1) + P(F|S_0)P(S_0)} \\ &= \frac{\frac{9}{10}\cdot\frac{1}{3}}{\frac{9}{10}\cdot\frac{1}{3} + \frac{1}{5}\cdot\frac{2}{3}} = \frac{9}{13}. \end{aligned}$$

Ein Ereignis A wird man sinnvollerweise unabhängig von B nennen, wenn die Zusatzinformation über das Eintreten von B die Wahrscheinlichkeit von A nicht verändert:

$$P(A|B) = P(A), \text{ d.h. } P(A \cap B) = P(A)P(B).$$

Die Gleichung auf der rechten Seite ist symmetrisch in A und B und auch für $P(B) = 0$ sinnvoll. Dies führt auf die folgende für die Wahrscheinlichkeitstheorie fundamentale Definition.

Definition 2.8. Sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum.

a) Zwei Ereignisse $A, B \in \mathfrak{F}$ heißen (stochastisch) *unabhängig*, falls

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

gilt.

b) Eine beliebige Familie $(A_j)_{j \in J} \subseteq \mathfrak{F}$ von Ereignissen heißt (in ihrer Gesamtheit) *unabhängig*, falls für je endlich viele paarweise verschiedene $j_1 \dots, j_n \in J$

$$P(A_{j_1} \cap \ldots \cap A_{j_n}) = P(A_{j_1}) \ldots P(A_{j_n})$$

gilt.

Bemerkung 2.9.

1. Aus $A_1, \ldots, A_n \in \mathfrak{F} \quad (n \geq 3)$ und $P(A_1 \cap \ldots \cap A_n) = P(A_1) \ldots P(A_n)$ folgt i.a. *nicht* die Unabhängigkeit von $A_1, \ldots, A_n$: Man wähle

$$A_1 = A_2 = \ldots = A_{n-1} = A \text{ mit } 0 < P(A) < 1, \qquad A_n = \emptyset.$$

Dann gilt

$$P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdots P(A_n) = 0,$$

aber

$$P(A_1 \cap A_2) = P(A) \neq P(A)^2 = P(A_1)P(A_2).$$

2. Aus der paarweisen Unabhängigkeit von Ereignissen folgt i.a. *nicht* deren Unabhängigkeit (Übungsaufgabe).

Unabhängigkeit von Teilexperimenten und diskrete Produkträume

$(\Omega_1, P_1), \ldots, (\Omega_n, P_n)$ seien diskrete Wahrscheinlichkeitsräume, die n Zufallsexperimente modellieren. Beeinflussen sich die n Experimente nicht gegenseitig, so führt deren Zusammenfassung zu einem Gesamtexperiment auf die Betrachtung des folgenden diskreten Wahrscheinlichkeitsraumes:

$$\Omega = \Omega_1 \times \cdots \times \Omega_n = \{(\omega_1, \ldots, \omega_n) \colon \omega_1 \in \Omega_1, \ldots, \omega_n \in \Omega_n\}.$$

Einzelwahrscheinlichkeiten:

$$p(\omega_1, \ldots, \omega_n) \colon = p_1(\omega_1) \ldots p_n(\omega_n),$$

wobei $p_1(\cdot), \ldots, p_n(\cdot)$ die Einzelwahrscheinlichkeiten zu $P_1, \ldots, P_n$ sind. Wegen $p(\omega_1, \ldots, \omega_n) \geq 0$ und

$$\sum_{(\omega_1, \ldots, \omega_n) \in \Omega} p(\omega_1, \ldots, \omega_n) = \left(\sum_{\omega_1 \in \Omega_1} p_1(\omega_1) \right) \ldots \left(\sum_{\omega_n \in \Omega_n} p_n(\omega_n) \right) = 1$$

sind dies tatsächlich Einzelwahrscheinlichkeiten und definieren somit ein Wahrscheinlichkeitsmaß P auf Ω (*Produktmaß*, Bezeichnung: $P = P_1 \otimes \dots \otimes P_n$).
(Ω, P) ist der gesuchte diskrete Wahrscheinlichkeitsraum. Es gilt

$$P(A_1 \times \dots \times A_n) = P_1(A_n) \dots P_n(A_n) \qquad (A_1 \subseteq \Omega_1, \dots, A_n \subseteq \Omega_n). \quad (2.1)$$

Tatsächlich,

$$\begin{aligned} P(A_1 \times \dots \times A_n) &= \sum_{(\omega_1,\dots,\omega_n) \in A_1 \times \dots \times A_n} p(\omega_1, \dots, \omega_n) \\ &= \sum_{\omega_1 \in A_1, \dots, \omega_n \in A_n} p_1(\omega_1) \dots p_n(\omega_n) \\ &= \sum_{\omega_1 \in A_1} p_1(\omega_1) \dots \sum_{\omega_n \in A_n} p_n(\omega_n) \\ &= P_1(A_1) \cdots P_n(A_n). \end{aligned}$$

Das Produktmaß $P = P_1 \otimes \dots \otimes P_n$ wird durch (2.1) vollständig charakterisiert. Ist $A_i \subseteq \Omega_i$ ein beliebiges Ereignis im i-ten Teilexperiment, so wird dieses im Gesamtexperiment durch die Menge

$$\hat{A}_i := \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n = \{(\omega_1, \dots, \omega_n) \in \Omega : \omega_i \in A_i\}$$

beschrieben ($i = 1, \dots, n$). Die Ereignisse $\hat{A}_1, \dots, \hat{A}_n$ sind unabhängig: Wegen

$$\hat{A}_1 \cap \ldots \cap \hat{A}_n = A_1 \times \dots \times A_n$$

folgt unter Benutzung von (2.1)

$$\begin{aligned} P(\hat{A}_1 \cap \ldots \cap \hat{A}_n) &= P(A_1 \times \dots \times A_n) = P_1(A_1) \cdots P_n(A_n) \\ &= P(\hat{A}_1) \cdots P(\hat{A}_n). \end{aligned}$$

Die entsprechenden Aussagen für beliebig $\hat{A}_{j_1}, \dots, \hat{A}_{j_m}$ $(1 \leq m \leq n, \quad 1 \leq j_1 < \ldots < j_m \leq n)$ folgen, wenn man oben A_i durch Ω_i (d.h. $\hat{A}_i$ durch Ω) ersetzt für $i \notin \{j_1, \dots, j_m\}$.
Also sind Ereignisse verschiedener Teilexperimente unabhängig, was die Richtigkeit der Modellierung des Gesamtexperiments über das Produktmaß unterstreicht.

Bemerkung 2.10. Sind $(\Omega_1, \mathfrak{F}_1, P_1), \dots, (\Omega_n, \mathfrak{F}_n, P_n)$ beliebige Wahrscheinlichkeitsräume, so kann man einen *Produktraum* $(\Omega, \mathfrak{F}, P)$ wie folgt konstruieren:

$\Omega = \Omega_1 \times \dots \times \Omega_n$,	
$\mathfrak{F} = \mathfrak{F}_1 \otimes \dots \otimes \mathfrak{F}_\mathfrak{n}$:	σ-Algebra, die von Mengen der Gestalt $A_1 \times \dots \times A_n \quad (A_1 \in \mathfrak{F}_1, \dots, A_n \in \mathfrak{F}_\mathfrak{n})$ erzeugt wird (kleinste σ-Algebra, die diese Mengen enthält),
$P = P_1 \otimes \dots \otimes P_n$:	dasjeniger Wahrscheinlichkeitsmaß auf $\mathfrak{F}$, für das $P(A_1 \times \dots \times A_n) = P_1(A_1) \dots P_n(A_n)$ $(A_1 \in \mathfrak{F}_1, \dots, A_n \in \mathfrak{F}_\mathfrak{n})$ gilt (siehe Maßtheorie).

Beispiel 2.11.

a) Das Produkt Laplacesche Wahrscheinlichkeitsräume ist ein Laplacescher Wahrscheinlichkeitsraum: Seien $(\Omega_1, P_1), \ldots, (\Omega_n, P_n)$ Laplacesche Wahrscheinlichkeitsräume mit den Einzelwahrscheinlichkeiten

$$p_1(\omega_1) \equiv \frac{1}{|\Omega_1|}, \cdots, p_n(\omega_n) \equiv \frac{1}{|\Omega_n|}.$$

Dann sind die Einzelwahrscheinlichkeiten des Produktraumes $\Omega = \Omega_1 \times \cdots \times \Omega_n$ gleich

$$p(\omega_1, \ldots, \omega_n) \equiv \frac{1}{|\Omega_1|} \cdots \frac{1}{|\Omega_n|} = \frac{1}{|\Omega_1 \times \ldots \times \Omega_n|} = \frac{1}{|\Omega|}$$

und beschreiben ebenfalls eine Gleichverteilung.

b) Beim n-maligen Werfen eines Würfels nimmt man
$\Omega = \Omega_0^n$ mit $\Omega_0 = \{1, \ldots, 6\}$,
$P = P_0^{\otimes n}$ mit P_0 Gleichverteilung auf Ω_0.

Kapitel 3

Zufallsvariable und deren Eigenschaften

3.1. Definition und erste Eigenschaften

Versuchsausgängen $\omega \in \Omega$ werden häufig Zahlenwerte $X(\omega)$ $(\in \mathbb{N}, \mathbb{Z}, \mathbb{R}, \dots)$ zugeordnet (Temperatur, Kurs einer Aktie, Datendurchsatz an einem Internet-Router ...).

Beispiel 3.1.

(i) n-maliges Würfeln: $\Omega = \{1, \dots, 6\}^n$
Anzahl der geworfenen „6"-en: $X(\omega_1, \dots, \omega_n) = \Sigma_{i=1}^n \mathbb{1}_{\{6\}}(\omega_i)$
Anzahl der Würfe vor erster „6": $Y(\omega_1, \dots, \omega_n) = \max\{i \leq n : \omega_1, \dots, \omega_i \neq 6\}$

(ii) Länge einer zufälligen Kreissehne (2. Experiment)
$\Omega = [0, 2\pi)^2$
$L(\varphi_1, \varphi_2) = 2|\sin \frac{\varphi_2 - \varphi_1}{2}|, \ \ (\varphi_1, \varphi_2) \in \Omega.$

Man möchte, dass die Mengen

$$\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\}$$

für eine große Klasse von „Zahlenmengen" B Ereignisse sind (und somit deren Wahrscheinlichkeiten angegeben werden können).

Definition 3.2. Sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $(E, \mathfrak{E})$ ein messbarer Raum (das heisst $\mathfrak{E}$ ist eine σ-Algebra in E). Eine Abbildung $X : \Omega \to E$ heisst *Zufallsvariable*, falls

$$\{X \in B\} \in \mathfrak{F} \ \text{ für alle } \ B \in \mathfrak{E}.$$

Dabei ist $\{X \in B\}$ das Ergebnis, dass die Zufallsvariable X Werte in der Menge B annimmt:

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

$X(\omega)$ heißt *Realisierung* der Zufallsvariablen X (zum Elementarereignis $\omega \in \Omega$).

Wichtige Spezialfälle:

(i) *Diskrete Zufallgrößen:* Sei E höchstens abzählbar, $\mathfrak{E} = \mathfrak{P}(E)$ (zum Beispiel E=$\mathbb{N}$, E=$\mathbb{Z}$)

(ii) *Reelle Zufallsgrößen:* $(E, \mathfrak{E})$=$(\mathbb{R}, \mathfrak{B})$ ($\mathfrak{B}$: *Borel-σ-Algebra*)

(iii) *Zufallsvektoren* (n-dimensionale Zufallsgrößen): $(E, \mathfrak{E})$=$(\mathbb{R}^n, \mathfrak{B}^n)$

Bemerkungen:

(i) Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum, $(E, \mathfrak{E})$ eine beliebiger messbarer Raum. Dann ist jede Abbildung $X : \Omega \to E$ Zufallsvariable:

$$\{X \in B\} \in \mathfrak{P}(\Omega), \ \ \forall \ B \in \mathfrak{E}$$

(ii) Sind X und Y reelle Zufallsgrößen, so sind zum Beispiel auch

$$\alpha X \ \ (\alpha \in \mathbb{R}), \ \ X + Y, \ \ X \cdot Y$$

reelle Zufallsgrößen (o.B.)

(iii) Sei $(\Omega, \mathfrak{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum. Dann sind die Indikatorfunktionen $\mathbb{1}_A$ von Ereignissen $A \in \mathfrak{F}$ reelle Zufallsgrößen:

$$\{\mathbb{1}_A \in B\} = \begin{cases} \emptyset, & \text{für } 0 \notin B, 1 \notin B \\ A, & \text{für } 0 \notin B, 1 \in B \\ A^c, & \text{für } 0 \in B, 1 \notin B \\ \Omega, & \text{für } 0 \in B, 1 \in B \end{cases}$$

Die Behauptung ist wegen $\emptyset, A, A^c, \Omega \in \mathfrak{F}$ offensichtlich.

Wir betrachten nun die Wahrscheinlichkeit

$$P(\{X \in B\}) \ \ (\text{kurz} : P(X \in B)),$$

dass die Zufallsvariable X Werte in $B \in \mathfrak{E}$ annimmt. Hierdurch wird ein Wahrscheinlichkeitsmaß auf $(E, \mathfrak{E})$ definiert.

Satz 3.3. *Seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum, $(E, \mathfrak{E})$ ein messbarer Raum und $X : \Omega \to E$ eine Zufallsvariable. Dann wird durch*

$$P_X(B) := P(X \in B), \ \ B \in \mathfrak{E},$$

ein Wahrscheinlichkeitsmaß P_X auf $(E, \mathfrak{E})$ definiert (oft auch mit $P \circ X^{-1}$ bezeichnet).

Beweis. Die Mengenfunktion P_X ist wohldefiniert, da nach Voraussetzung $\{X \in B\} \in \mathfrak{F}$ für alle $B \in \mathfrak{E}$.

$$P_X(B) \geq 0, \;\; P_X(E) = P(X \in E) = P(\Omega) = 1$$

σ-Additivität von P_X:
Sei $(B_n) \subset \mathfrak{E}$ eine Folge paarweise disjunkter Mengen, dann ist

$$\{X \in \bigcup_n B_n\} = \bigcup_n \{X \in B_n\}$$

eine disjunkte Vereinigung.
Somit folgt mittels σ-Additivität

$$\begin{aligned} & P_X(\bigcup_n B_n) = P(X \in \bigcup_n B_n) \\ = & \sum_n P(X \in B_n) = \sum_n P_X(B_n). \end{aligned}$$

□

Definition 3.4. Das Wahrscheinlichkeitsmaß P_X heißt *Verteilung* der Zufallsvariable X.

Anschaulich beschreibt P_X mit welcher Wahrscheinlichkeit die Zufallsvariable X welche Werte aus E annimmt. Diese Wahrscheinlichkeiten sind die eigentlich wichtigen Charakteristika einer Zufallsvariable. Auf welchem konkreten Wahrscheinlichkeitsraum X definiert ist und wie die Abbildung X im Detail aussieht ist aus wahrscheinlichkeitstheoretischer Sicht meist zweitrangig. Wichtig ist die Verteilung von X.

In einigen Fällen ist eine einfachere Charakterisierung der Verteilung von Zufallsgrößen möglich:
Diskrete Zufallsgrößen: Sei (E, P_X) ein diskreter Wahrscheinlichkeitsraum. Die vollständige Charakterisierung von P_X erfolgt durch die *Einzelwahrscheinlichkeiten*

$$p_X(e) := P_X(\{e\}) = P(X = e), \;\; e \in E. \tag{3.1}$$

Oft nennt man auch reelle Zufallsgröße X, die fast sicher Werte aus einer endlichen oder abzählbaren Menge $E \subset \mathbb{R}$ annehmen (das heisst $P(X \in E) = 1$) diskret. Deren Verteilung P_X auf $(\mathbb{R}, \mathfrak{B})$ ist dann auch durch die Einzelwahrscheinlichkeiten (3.1) vollständig charakterisiert, da $P_X(E) = 1$.

Beispiel 3.5.
Wir betrachten n-maliges Würfeln. Sei X die Zahl der Würfe vor erster „6“ mit den Einzelwahrscheinlichkeiten: $p_k = P(X = k)$ $(k = 0, 1, \ldots, n)$,

$$p_k = \begin{cases} \left(\frac{5}{6}\right)^k \frac{1}{6}, & \text{für } \; k = 0, \ldots, n-1 \\ \left(\frac{5}{6}\right)^n, & \text{für } \; k = n. \end{cases}$$

Reelle Zufallsgrößen:
Charakterisierung durch ihre *Verteilungsfunktion*

$$F_X(x) := P_X\left((-\infty, x]\right) = P(X \le x), \quad x \in \mathbb{R}.$$

Beispiel 3.6.

(i) *Einpunktverteilung:* $P(X = c) = 1$ (deterministische, entartete Zufallsgröße)

$$\Rightarrow \quad F_X(x) = \begin{cases} 0, & x < c \\ 1, & x \ge c \end{cases}$$

(ii) *diskrete Verteilung:* Seien $P(X \in E) = 1$, $E \subset \mathbb{R}$ höchstens abzählbar und $p_X(e) = P(X = e)$, $e \in E$ die zugehörigen Einzelwahrscheinlichkeiten, dann ist

$$F_X(x) = \sum_{e \le x} p_X(e).$$

Zum Beispiel sei $x_1 < x_2 < \ldots < x_n$, $p_k := P(X = x_k)$ $(k = 1, \ldots, n)$ und $\sum_{k=1}^n p_k = 1$.

(iii) *Gleichverteilung* auf dem Intervall $[a, b]$: Punkte aus dem Intervall $[a, b]$ „auf gut Glück" auswählen. Sei X die Zufallsgröße, die die zufällige Lage dieses Punktes beschreibt.

(zum Beispiel $\Omega = [a, b]$, $\mathfrak{F} = \mathfrak{B}_{[a,b]}$, $P(A) = \frac{\lambda(A)}{b-a}$ $(A \in \mathfrak{B}_{[a,b]})$, $X(\omega) = \omega$, $\omega \in [a, b]$)

$$F_X(x) = P(X \le x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b. \end{cases}$$

Satz 3.7. *Die Verteilungsfunktion $F = F_X$ einer reellen Zufallsgröße X besitzt folgende Eigenschaften:*

(i) F nicht fallend;

(ii) F rechtsstetig;

(iii) $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to \infty} F(x) = 1$.

Beweis.

(i) $x < y \Rightarrow (-\infty, x] \subset (-\infty, y]$ und auf Grund der Monotonie von P_X folgt $P_X((-\infty, x]) \le P_X((-\infty, y])$, das heisst $F(x) \le F(y)$.

(ii) $x_n \downarrow x \Rightarrow (-\infty, x_n] \downarrow (-\infty, x]$ (Die Abgeschlossenheit der Intervalle wird benötigt) und auf Grund der Stetigkeit von Wahrscheinlichkeitsmaßen bezüglich monotoner Mengenfolgen folgt $P_X((-\infty, x_n]) \to P_X((-\infty, x])$, das heißt
$F(x_n) \to F(x)$.

(iii) Analog zu (ii) unter Benutzung von

$(-\infty, -n] \downarrow \emptyset$ und $(-\infty, n] \uparrow \mathbb{R}$ für $n \uparrow \infty$.

□

Bemerkung 3.8. (ohne Beweis)

(i) Jede reelle Funktion F mit den im Satz genannten Eigenschaften (i)-(iii) ist Verteilungsfunktion einer Zufallsgröße: Es existiert ein Wahrscheinlichkeitsraum und darauf eine reelle Zufallsgröße X derart, dass die vorgegebene Funktion F mit der Verteilungfunktion F_X von X übereinstimmt.

(ii) Eine Verteilung ist durch Angabe ihrer Verteilungsfunktion vollständig charakterisiert: Für beliebige reelle Zufallsgrößen X und Y gilt:

$F_X = F_Y \Rightarrow P_X = P_Y$.

Stetige Zufallsgrößen:
Existiert eine (Lebesgue-) integrierbare Funktion $f_X : \mathbb{R} \to \mathbb{R}_+$ mit

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,\mathrm{d}y, \quad x \in \mathbb{R},$$

so nennt man X *stetige* Zufallsgröße. Die Verteilungsfunktion F_X heißt dann *(absolut-) stetig* mit *Dichte* f_X.

Bemerkung 3.9.

(i) Ist f Dichte einer Zufallsgröße, so gilt $f \geq 0$ und $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$.

(ii) Ist eine Verteilungsfunktion F stetig und stückweise stetig differenzierbar, so ist F absolut stetig mit Dichte $f = F'$. (f ist nicht eindeutig; zum Beispiel kann man f an den Stellen, wo F' nicht existiert beliebig wählen.)

(iii) $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(y)\,\mathrm{d}y$

Beispiel 3.10.

(i) Gleichverteilung auf [a,b] mit der Dichte:

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

(ii) Verteilung der Länge einer zufälligen Kreissehne L:

$$L(\phi_1, \phi_2) = 2 \left| \sin \frac{\phi_2 - \phi_1}{2} \right|, \quad (\phi_1, \phi_2) \in \Omega = [0, 2\pi)^2$$

und P Gleichverteiltung auf Ω.

Ist $x < 0$ so ist die Menge $\{L \leq x\} = \emptyset$ und somit $F_L(x) = 0$. Ist $x \geq 2$, so ist die Menge $\{L \leq x\} = \Omega$ und somit $F_L(x) = 1$.

Für $0 \leq x < 2$:

$$\{L \leq x\} = \left\{ (\phi_1, \phi_2) \in \Omega : \quad \left| \sin \frac{\phi_2 - \phi_1}{2} \right| \leq \frac{x}{2} \right\},$$

somit ist

$$F_L = P(L \leq x) = \frac{\lambda(L \leq x)}{\lambda(\Omega)} = \frac{8\pi \arcsin \frac{x}{2}}{(2\pi)^2}.$$

Insgesamt erhalten wir:

$$F_L(x) = \begin{cases} \frac{2}{\pi} \arcsin \frac{x}{2}, & 0 \leq x < 2; \\ 1, & x \geq 2. \end{cases}$$

Somit ist F_L absolut stetig mit Dichte:

$$f_L(x) = \begin{cases} 0, & x < 0; \\ \frac{1}{\pi\sqrt{1-(x/2)^2}}, & 0 \leq x < 2; \\ 0, & x \geq 2. \end{cases}$$

3.2. Einige häufig auftretende diskrete Verteilungen

Eine Reihe von diskreten Verteilungen tritt im Zusammenhang mit dem Bernoullischen Versuchsschema (Bernoulli-Experiment) auf. Wir werden später sehen, dass sich viele Aussagen der Wahrscheinlichkeitstheorie anhand dieses einfachen stochastischen Modells illustrieren lassen.

Bernoullisches Versuchsschema:
Eine Folge unabhängiger identischer Teilversuche. Bei jedem Teilversuch wird beobachtet, ob ein vorgegebenes Ereignis eingetreten ist (zum Beispiel: „6“ gewürfelt oder „Zahl“ geworfen).
Gegeben: Eine Folge (A_n) unabhängiger Ereignisse mit gleicher Erfolgswahrscheinlichkeit p, $0 < p < 1$:
$P(A_n) = p$ für alle $n \in \mathbb{N}$.

$$X_n := \mathbb{1}_{A_n} = \begin{cases} 1, & \text{falls } (A_n) \text{ eingetreten („Erfolg“);} \\ 0, & \text{falls } (A_n) \text{ nicht eingetreten („Mißerfolg“).} \end{cases}$$

$P(X_n = 1) = p$, $P(X_n = 0) = q := 1 - p$
Sei $S_n := \sum_{i=1}^n X_i$ die Anzahl der erfolgreichen Versuchsausgänge in Serie von n Versuchen.
Betrachte die Verteilung von S_n:

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \ldots, n$$

Wobei $\binom{n}{k}$ die Anzahl der Möglichkeiten dafür ist, dass genau k der n Versuchsausgänge erfolgreich sind und $p^k(1-p)^{n-k}$ die Wahrscheinlichkeit jeder dieser Möglichkeiten ist (Unabhängigkeit der Teilversuche).
Binomialverteilung mit Parametern n und p ($n \in \mathbb{N}, \;\; 0 < p < 1$). Die Binomialverteilung ist diskret auf $\{0, 1, \ldots, n\}$ mit Einzelwahrscheinlichkeiten

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \;\; k = 0, 1, \ldots, n$$

Auf Grund der binomischen Formel gilt tatsächlich

$$\sum_{k=0}^{n} b(k; n, p) = [p + (1-p)]^n = 1.$$

Sei $Y := \sup\{k \geq 0 : X_1 = \ldots = X_k = 0\}$ die Anzahl der Misserfolge vor dem ersten Erfolg:

$$P(Y = k) = P(X_1 = 0, \ldots, X_k = 0, X_{k+1} = 1)$$

Den Ausdruck in der letzten Klammer kann man auch wie folgt schreiben:

$$\{X_1 = 0\} \cap \ldots \cap \{X_k = 0\} \cap \{X_{k+1} = 1\}.$$

Somit gilt:

$$P(A_1^c \cap \ldots \cap A_k^c \cap A_{k+1}) = (1-p)^k p, \ \ k = 0, 1, 2 \ldots$$

Geometrische Verteilung mit Parameter $p \in (0,1)$:
Diskrete Verteilung auf $\mathbb{N}_0 = 0, 1, 2, \ldots$ mit Einzelwahrscheinlichkeiten

$$p_k = (1-p)^k p, \ \ k \in \mathbb{N}_0$$

Verallgemeinerung (siehe Übung):
Sei Z_n die Anzahl der Misserfolge vorm n-ten Erfolg mit

$$\begin{aligned}
P(Z_n = k) &= \binom{n+k-1}{k} p^n q^k \\
&= \frac{(n+k-1)(n+k-2)\cdots(n+1)n}{1 \cdot 2 \ldots (k-1)k} p^n q^k \\
&= \frac{(-n)(-n-1)\ldots(-n-k+2)(-n-k+1)}{1 \cdot 2 \ldots (k-1)k} p^n (-q)^k \\
&= \binom{-n}{k} p^n (-q)^k, \ \ k \in \mathbb{N}_0.
\end{aligned}$$

Dies ist die negative Binomialverteilung mit Parameter $n \in \mathbb{N}$ und $p \in (0,1)$:
Diese Verteilung wird diskret verteilt auf $\mathbb{N}_0$ mit Einzelwahrscheinlichkeiten

$$p(k; n, p) = \binom{-n}{k} p^n (-q)^k, \ \ k \in \mathbb{N}_0.$$

Die geometrische Verteilung ist ein Spezialfall der eg. Binomialverteilung für $n = 1$.

$$\sum_{k=0}^{\infty} f(k; n, p) = p^n \sum_{k=0}^{n} \binom{-n}{k} (-q)^k = p^n (1-q)^{-n} = 1$$

Bei der vorletzten Gleichung ging die „Binomische Formel“ ein (Taylorreihe für $(1-x)^{-n}$).

Die Poissonverteilung mit Parameter $\lambda > 0$ ist eine diskrete Verteilung auf $\mathbb{N}_0$ mit Einzelwahrscheinlichkeiten:

$$\pi_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \ \ k \in \mathbb{N}_0.$$

Die Poissonverteilung tritt insbesondere als Grenzverteilung der Binomialverteilung auf.

Satz 3.11. (Poissonscher Grenzwertsatz) *Für* $\lambda > 0$ *und* $np_n \to \lambda$ *gilt*

$$\lim_{n\to\infty} b(k; n, p_n) = \pi_\lambda(k), \quad k \in \mathbb{N}_0.$$

Beweis. Sei $\lambda_n := np_n \to \lambda, \; k \in \mathbb{N}_0$ beliebig fixiert.

$$\begin{aligned}
b(k; n, p_n) &= b\left(k; n, \frac{\lambda_n}{n}\right) \\
&= \binom{n}{k}\left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!}\frac{\lambda_n^k}{n^k}\frac{(1-\lambda_n/n)^n}{(1-\lambda_n/n)^k} \\
&= \frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^n\left(1-\frac{\lambda_n}{n}\right)^{-k},
\end{aligned}$$

wobei das Produkt der ersten $k-1$ Terme gegen 1, der vorletzte gegen $e^{-\lambda}$ und der letzte wieder gegen 1 konvergiert. Somit konvergiert der gesamte letzte Ausdruck für $n \to \infty$ gegen $(\lambda^k/k!)e^{-\lambda} = \pi_\lambda(k)$.
Dass $\lim\limits_{n\to\infty}\left(1 - \frac{\lambda_n}{n}\right)^n = e^{-\lambda}$ ist, sieht man wie folgt nach Logarithmieren:

$$\log\left(1-\frac{\lambda_n}{n}\right)^n = n\log\left(1-\frac{\lambda_n}{n}\right) = \lambda_n\frac{\log(1-\lambda_n/n)}{\lambda_n/n},$$

für $n \to \infty$ folgt:

$$\lambda\frac{\mathrm{d}}{\mathrm{d}x}\log(1-x)|_{x=0} = -\lambda.$$

□

Interpretation: Für große Versuchsreihen n und kleine Erfolgswahrscheinlichkeiten p („seltene Ereignisse") im Bernoulli-Experimernt ist näherungsweise $b(k; n, p) \approx \pi_\lambda(k)$ mit $\lambda = np$.

Sei X geometrisch verteilt mit Parameter $p \in (0, 1)$. Die Verteilung einer solchen Zufallsgröße besitzt die folgende spezielle Eigenschaft:

$$P(X \geq n + i | X \geq n) = P(X \geq i) \;\; (n, i \in \mathbb{N}_0). \tag{3.2}$$

Tatsächlich ist

$$P(X \geq n) = \sum_{k=n}^{\infty} p(1-p)^k = (1-p)^n,$$

und folglich

$$\begin{aligned} P(X \geq n+i | X \geq n) &= \frac{P(X \geq n+i)}{P(X \geq n)} \\ &= \frac{(1-p)^{n+i}}{(1-p)^n} \\ &= (1-p)^i \\ &= P(X \geq i). \end{aligned}$$

Man sagt, dass die Verteilung von X *nicht altert (gedächtnislos ist).* Man überlegt sich leicht, dass auch die Umkehrung gilt: Ist X eine $\mathbb{N}_0$-wertig Zufallsgröße, $P(X \geq n) > 0$ für alle $n \in \mathbb{N}_0$ und gilt (3.2), so ist X geometrisch verteilt (mit einem gewissen Parameter $p \in (0,1)$).

Urnenmodelle:
Eine Urne (Schachtel) enthalte r rote und s schwarze Kugeln. Es werden nacheinander n Kugeln entnommen. Wie groß ist Wahrscheinlichkeit $p_k^{(n)}$, genau k mal eine rote Kugel zu ziehen?

(i) *Ziehung mit Zurücklegen:*

In jeder der n Ziehungen werden rote Kugeln mit der Wahrscheinlichkeit $\frac{r}{r+s}$ gezogen. Somit handelt es sich um eine Binomialverteilung.

$p_k^{(n)} = b\left(k; n, \frac{r}{r+s}\right) = \binom{n}{k}\left(\frac{r}{r+s}\right)^k \left(\frac{s}{r+s}\right)^{n-k}$, $k = 0, 1, \ldots, n$.

(ii) *Ziehung ohne Zurücklegen:* $(1 \leq n \leq r+s)$

Sei $\binom{r+s}{n}$ die Gesamtanzahl der Möglichkeiten, mit n zu ziehenden Kugeln. Dann ist $\binom{r}{k}\binom{s}{n-k}$ die Anzahl der günstigsten Möglichkeiten, wobei k die Anzahl der roten und $n-k$ die der schwarzen Kugeln sind.

$p_k^{(n)} = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}$, $0 \leq k \leq r$, $0 \leq n-k \leq s$.

(hypergeometrische Verteilung)

(iii) *Verallgemeinerung:*

Man legt diegezogene Kugel zurück und fügt der Urne i Kugeln der gezogenen Farbe und j Kugeln der anderen Farbe hinzu ($i, j \in \mathbb{Z}$). Dies geschieht aber nur solange, wie noch Kugeln in der Urne sind.

$i = 0$, $j = 0$: Ziehung *mit* Zurücklegen

$i = -1$, $j = 0$: Ziehung *ohne* Zurücklegen.

$i > 0$, $j = 0$: Pólya-Modell für Wachstum zweier konkurrierenden Populationen mit Selbstverstärkungseffekt

$i = -1$, $j = 1$: Ehrenfest-Modell: Diffusion von Gasmolekülen durch eine Membran.

Spezialfall:

$i \in \mathbb{Z}$, $j = 0$:

R_l bei der l-ten Ziehung rot.

S_l bei der l-ten Ziehung schwarz.

Wir benutzen folgende Formel für die bedingte Wahrscheinlichkeit:

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

(falls $P(A_1 \cap \ldots \cap A_{n-1}) > 0$).

Die Wahrscheinlichkeit, dass die ersten k Ziehungen „rote“ und die restlichen $n-k$ Ziehungen „schwarze“ Kugeln liefern:

$$\begin{aligned} P(R_1 \cap \ldots \cap R_k \cap S_{k+1} \cap \ldots \cap S_n) &= P(R_1)P(R_2|R_1)\cdots \\ &\quad P(R_k|R_1 \cap \ldots \cap R_{k-1}) \\ &\quad P(S_{k+1}|R_1 \cap \ldots \cap R_k)\cdots \\ &\quad P(S_{k+2}|R_1 \cap \ldots \cap R_k \cap S_{k+1})\cdots \\ &\quad P(S_n|R_1 \cap \ldots \cap R_k \cap S_{k+1} \cap \ldots \cap S_{n-1}) \\ &= \frac{r}{r+s}\frac{r+i}{r+s+i}\cdots\frac{r+(k-1)i}{r+s+(k-1)i} \\ &\quad \frac{s}{r+s+ki}\frac{s+i}{r+s+(k+1)i} \\ &\quad \cdots\frac{s+(n-k-1)i}{r+s+(n-1)i} \end{aligned}$$

Die Wahrscheinlichkeit hängt nicht von der Reihenfolge von „rot“ und „schwarz“ ab. Es geht nur jeweils ein, wieviel „rote“ bzw. „schwarze“ schon gezogen wurden.

Die Anzahl der möglichen Anordnungen ist also: $\binom{n}{k}$

Insgesamt ist für $r+s+(n-1)i > 0$:

$$p_k^{(n)} = \binom{n}{k}\frac{\prod_{l=0}^{k-1}(r+li)\prod_{l=0}^{n-k-1}(s+li)}{\prod_{l=0}^{n-1}(r+s+li)}, \quad (k = 0, 1, \ldots, n),$$

die *(Pólya-Verteilung)*.

3.3. Gemeinsame Verteilung und Unabhängigkeit von Zufallsvariablen

Definition 3.12. Sei $(\Omega, \mathfrak{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum und X_i eine Zufallsvariable mit Werten im messbarem Raum $(E_i, \mathfrak{E}_i)$, $i = 1, \ldots, n$.
Die Zufallsvariablen $X_1, \ldots, X_n$ heißen *unabhängig*, falls für beliebige $B_1 \in \mathfrak{E}_1, \ldots, B_n \in \mathfrak{E}_n$ die Ereignisse $\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$ unabhängig sind.

Bemerkung 3.13.

(i)

$$\begin{aligned} X_1, \ldots, X_n \text{ sind unabhängig} \quad \Leftrightarrow \quad & P(X_1 \in B_1, \ldots, X_n \in B_n) \\ & = P(X_1 \in B_1) \cdots P(X_n \in B_n) \\ & \text{für beliebige } B_1 \in \mathfrak{E}_1, \ldots, B_n \in \mathfrak{E}_n \end{aligned}$$

Setzt man $B_i = E_i$ für gewisse i liefert die rechte Seite für Teilmengen der betreffenden Ereignisse.

(ii) Eine beliebige Familie von Zufallsvariablen auf demselben Wahrscheinlichkeitsraum heißt unabhängig, wenn jede endliche Teilfamilie unabhängig ist.

Diskrete Zufallsgrößen auf $(\Omega, \mathfrak{F}, P)$
Sei X_i diskret mit Werten in einer höchstens abzählbaren Menge E_i $(i = 1, \ldots, n)$, dann ist $\mathfrak{X} = (X_1, \ldots, X_n)$ diskret mit Werten in $E := E_1 \times \ldots \times E_n$ (auch höchstens abzählbares Produkt). Die Verteilung von $\mathfrak{X}$ heißt *gemeinsame Verteilung* der Zufallsgrößen $X_1, \ldots, X_n$.
Charakterisierung durch Einzelwahrscheinlichkeiten:

$$\begin{aligned} p_{X_i}(e_i) &:= P(X_i = e_i), \ e_i \in E_i, \\ p_{\mathfrak{X}}(e) &:= P(\mathfrak{X} = e) = P(X_1 = e_1, \ldots, X_n = e_n), \ e = (e_1, \ldots, e_n) \in E. \end{aligned}$$

Proposition 3.14. *Unter den obigen Voraussetzungen gilt*

$$\begin{aligned} X_1, \ldots, X_n \textit{unabhängig} \quad \Leftrightarrow \quad & P(X_1 = e_1, \ldots, X_n = e_n) = P(X_1 = e_1) \\ & \cdots P(X_n = e_n) \\ & \textit{für alle } (e_1, \ldots, e_n) \in E_1 \times \ldots \times E_n \\ \Leftrightarrow \quad & p_{\mathfrak{X}}(e_1, \ldots, e_n) = p_{X_1}(e_1) \cdots p_{X_n}(e_n) \\ & \textit{für alle } (e_1, \ldots, e_n) \in E_1 \times \ldots \times E_n \\ & (\textit{kurz: } p_{\mathfrak{X}} = p_{X_1} \otimes \ldots \otimes p_{X_n} \ \ \textit{Tensorprodukt}) \end{aligned}$$

Beweis. „⇒“ offensichtlich nach der Definition der Unabhängigkeit.

„$\Leftarrow$" Sei $B_i \subseteq E_i$ beliebig $(i = 1, \ldots, n)$

$$\begin{aligned}
P(X_1 \in B_1, \ldots, X_n \in B_n) &= \sum_{(e_1,\ldots,e_n)\in B_1\times\ldots\times B_n} P(X_1 = e_1, \ldots, X_n = e_n) \\
&= \sum_{(e_1,\ldots,e_n)\in B_1\times\ldots\times B_n} P(X_1 = e_1) \cdots P(X_n = e_n) \\
&= \sum_{e_1\in B_1} P(X_1 = e_1) \cdots \sum_{e_n \in B_n} P(X_n = e_n)
\end{aligned}$$

Wobei jede Summe gleich $P(X_i \in B_i)$, $i = (1, \ldots, n)$ ist. $\square$

Beispiel 3.15. Bernoullisches Versuchsschema.
Sei (A_n) eine Folge unabhängiger Ereignisse mit $P(A_n) = p$ für alle n und $X_n = \mathbb{1}_{A_n}$ der Ausgang des n-ten Versuchs. Dann ist (X_n) eine Folge *unabhängiger* (diskreter) Zufallsgrößen mit Werten in $\{0{,}1\}$.

$$\begin{aligned}
P(X_1 = i_1, \ldots, X_n = i_n) &= P(A_1^{(i_1)} \cap \ldots \cap A_n^{(i_n)}) \\
&= P(A_1^{(i_1)}) \cdots P(A_n^{(i_n)}) \\
&= P(X_1 = i_1) \cdots P(X_n = i_n)
\end{aligned}$$

mit $(i_1, \ldots, i_n \in \{0, 1\})$ und

$$A^{(i)} := \begin{cases} A, & i = 1; \\ A^c, & i = 0. \end{cases}$$

Hierbei wurde benutzt, dass mit $A_1, \ldots, A_n$ auch die Ereignisse $A_1^{(i_1)}, \ldots, A_n^{(i_n)}$ unabhängig sind (Übungsaufgabe).
Dann ist $\mathfrak{X}_n = (X_1, \ldots, X_n)$ mit den Einzelwahrscheinlichkeiten:
$p_{\mathfrak{X}_n}(i_1, \ldots, i_n) = p^{\sum_{k=1}^n i_k}(1-p)^{\sum_{k=1}^n (1-i_k)}$.

Reelle Zufallsgrößen auf $(\Omega, \mathfrak{F}, P)$

$$X_1, \ldots, X_n \text{ reell} \quad \Leftrightarrow \quad \mathfrak{X} = (X_1, \ldots, X_n) \ n\text{-dimensionaler Zufallsvektor (das heißt } (\mathbb{R}^n, \mathfrak{B}^n)\text{-wertig).}$$

Gemeinsame Verteilungsfunktion:
Sei $F_{\mathfrak{X}}(x_1, \ldots, x_n) := P(X_1 \leq x_1, \ldots, X_n \leq x_n)$, mit $x_1, \ldots, x_n \in \mathbb{R}$. Da $X_1, \ldots, X_n$ unabhängig sind, folgt, dass die Ereignisse
$\{X_1 \leq x_1\}, \ldots, \{X_n \leq x_n\}$ unabhängig für beliebige $x_1, \ldots x_n \in \mathbb{R}$ sind.
Somit gilt $F_{\mathfrak{X}}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$ $(x_1, \ldots, x_n \in \mathbb{R})$.

Satz 3.16. (o.B.) *Seien $X_1, \ldots, X_n$ reelle Zufallsgrößen. Dann gilt:*

$$X_1, \ldots, X_n \quad \text{unabhängig} \quad \Leftrightarrow \quad F_{\mathfrak{X}}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \ \forall x_1, \ldots, x_n \in \mathbb{R}$$
$$(\textit{kurz: } F_{\mathfrak{X}} = F_{X_1} \otimes \ldots \otimes F_{n})$$

„$\Leftarrow$" erfordert maßtheoretische Hilfsmittel.

Gemeinsame Dichte: Man sagt, $X_1, \ldots, X_n$ besitzen eine *gemeinsame Dichte* $f_{\mathfrak{X}} : \mathbb{R}^n \to \mathbb{R}$, falls $f_{\mathfrak{X}} \geq 0$ (Lebesgue-)integrierbar ist und

$$F_{\mathfrak{X}}(x_1, \ldots, x_n) \quad = \quad \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathfrak{X}}(y_1, \ldots, y_n) \, \mathrm{d}y_n \ldots \mathrm{d}y_1$$

für alle $x_1, \ldots, x_n \in \mathbb{R}$.

Bemerkung 3.17. Sei $\mathfrak{X} = (X_1, \ldots, X_n)$

(i) Da $\{X_1 \leq x_1, \ldots, X_n \leq x_n\}$ gegen Ω konvergiert für $x_1 \uparrow \infty, \ldots, x_n \uparrow \infty$, so folgt auf Grund der Stetigkeit des Maßes P

$$F_{\mathfrak{X}}(x_1, \ldots, x_n) \to 1 \text{ für } x_1 \to \infty, \ldots, x_n \to \infty.$$

Dann ist

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathfrak{X}}(y_1, \ldots, y_n) \, \mathrm{d}y_n \ldots \mathrm{d}y_1 = 1.$$

(ii) $P(\mathfrak{X} \in B) = \int \cdots \int_B f_{\mathfrak{X}}(y_1, \ldots, y_n) \, \mathrm{d}y_n \ldots \mathrm{d}y_1$ für alle $B \in \mathfrak{B}^n$ (ohne Beweis).

Proposition 3.18. *Besitzen die Zufallsgrößen $X_1, \ldots, X_n$ eine gemeinsame Dichte $f_{\mathfrak{X}}$, so besitzen die Zufallsgrößen X_i ebenfalls Dichten f_{X_i} und*

$$f_{X_i}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathfrak{X}}(y_1, \ldots, y_{i-1}, x, y_{i+1}, \ldots, y_n) \prod_{i \neq j} \mathrm{d}y_j \quad (i = 1, \ldots, n).$$

Beweis. Für
$\{X_1 \leq x_1, \ldots, X_{i-1} \leq x_{i-1}, X_i \leq x, X_{i+1} \leq x_{i+1}, \ldots, X_n \leq x_n\} \uparrow$
$\{X_i \leq x\}$ für beliebige Folgen von $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ mit
$x_1 \uparrow \infty, \ldots, x_{i-1} \uparrow \infty, x_{i+1} \uparrow \infty, \ldots, x_n \uparrow \infty$ folgt

$$\begin{aligned}
F_{X_i}(x) &= \lim_{x_j \uparrow \infty \ (j \neq i)} F_{\mathfrak{X}}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) \\
&= \lim_{x_j \uparrow \infty \ (j \neq i)} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{i-1}} \int_{-\infty}^{x} \int_{-\infty}^{x_{i+1}} \cdots \int_{-\infty}^{x_n} f_{\mathfrak{X}}(y_1, \ldots, y_{i-1}, x, y_{i+1}, \ldots, y_n) \\
&\quad \mathrm{d}y_n \ldots \mathrm{d}y_1 \qquad (3.3) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x} \cdots \int_{-\infty}^{\infty} f_{\mathfrak{X}}(y_1, \ldots, y_{i-1}, x, y_{i+1}, \ldots, y_n) \\
&\quad \mathrm{d}y_n \ldots \mathrm{d}y_1 \qquad (3.4) \\
&= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathfrak{X}}(y_1, \ldots, y_{i-1}, x, y_{i+1}, \ldots, y_n) \prod_{i \neq j} \mathrm{d}y_j \right) \\
&\quad \mathrm{d}y_i \qquad (3.5)
\end{aligned}$$

In (3.3) wurde $\lim_{x_j \uparrow \infty \ (j \neq i)}$verwendet, in (3.4) die monotone Konvergenz der Intergrale und in (3.5) die Vertauschung der Integralreihenfolge nach Fubini. Der letzte Klammerausdruck ist gleichbedeutend mit $f_{X_i}(y_i)$ und ist die Dichte von F_{X_i}. □

Satz 3.19. *Angenommen $X_1, \ldots, X_n$ besitzen eine gemeinsame Dichte $f_{\mathfrak{X}}$. Dann gilt*

$$X_1, \ldots, X_n \quad \Leftrightarrow \quad f_{\mathfrak{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n) \text{ für } \lambda\text{-fast alle } (x_1, \ldots, x_n) \in \mathbb{R}^n$$
$$(\text{kurz } f_{\mathfrak{X}} = f_{X_1} \otimes \ldots \otimes f_{X_n})$$

Beweis.

$$\begin{aligned}
F_{\mathfrak{X}}(x_1, \ldots, x_n) &= \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{\mathfrak{X}}(y_1, \ldots, y_n)\, \mathrm{d}y_n \ldots \mathrm{d}y_1 \\
F_{X_1}(x_1) \cdots F_{X_n}(x_n) &= \int_{-\infty}^{x_1} f_{X_1}(y_1)\, \mathrm{d}y_1 \ldots \int_{-\infty}^{x_n} f_{X_n}(y_n)\, \mathrm{d}y_n \\
&= \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{X_1}(y_1) \ldots f_{X_n}(y_n)\, \mathrm{d}y_n \ldots \mathrm{d}y_1
\end{aligned}$$

In die letzte Gleichung ist der Satz von Fubini eingegangen, damit folgt:

$$X_1, \ldots, X_n \text{ unabhängig} \Leftrightarrow F_{\mathfrak{X}} = F_{X_1} \otimes \ldots \otimes F_{X_n} \Leftrightarrow f_{\mathfrak{X}} = f_{X_1} \otimes \ldots \otimes f_{X_n}.$$

□

Bemerkung 3.20. Aus dem Beweis ist ersichtlich: Sind $X_1, \ldots, X_n$ *unabhängig* mit Dichten $f_{X_1}, \ldots, f_{X_n}$, so besitzen sie eine gemeinsame Dichte f_X und $f_{\mathfrak{X}} = f_{X_1} \otimes \ldots \otimes f_{X_n}$.

Summen unabhängiger Zufallsgrößen und Faltung von Verteilungen
Seien X und Y *unabhängige* reelle Zufallsgrößen. Wie ist $Z := X + Y$ verteilt?

1. X, Y diskret mit ganzzahligen Werten
Seien $p_X(i) = P(X = i)$ und $p_Y(j) = P(Y = j)$ mit $(i, j \in \mathbb{Z})$ Einzelwahrscheinlichkeiten. Dann ist

$$p_Z(n) = P(Z = n) = \sum_{i \in \mathbb{Z}} P(X = i, Y = n - i).$$

Auf Grund der Unabhängigkeit von X und Y ist der letzte Ausdruck gleich

$$\sum_{i \in \mathbb{Z}} p_X(i) p_Y(n - i).$$

Somit gilt:

$$p_{X+Y}(n) = \sum_{i \in \mathbb{Z}} p_X(i) p_Y(n - i), \quad n \in \mathbb{Z}$$

ist die *Faltung* der Einzelwahrscheinlichkeiten. Kurzschreibweise: $p_{X+Y} = p_X * p_Y$
Die Verteilung von $X+Y$ hängt nur von den Verteilungen von X und Y ab. Es gilt die Kommutativität der Faltung: $p_X * p_Y = p_Y * p_X$

Beispiel 3.21.

(i) Seien X, Y unabhängig binomialverteilt mit dem Parameter (m, p) bzw. (n, p). Dann ist auch $X+Y$ binomialverteilt mit dem Parameter $(m+n, p)$.

Dies ist klar, da wir ein Bernoullisches Versuchsschema (X_n) mit Erfolgswahrscheinlichkeit p betrachten können. Dann sind $X := X_1 + \ldots + X_m$ und $Y := X_{m+1} + \ldots + X_{m+n}$ $b_{m,p}$ bzw. $b_{n,p}$-verteilt und unabhängig (Beweis). Das Ereignis ergibt sich auch durch Faltung der Einzelwahrscheinlichkeiten:

$$\begin{aligned}
p_{X+Y}(k) &= \sum_{i=0}^{k} b(i; m, p) b(k-i; n, p) \\
&= \sum_{i=0}^{k} \binom{m}{i} p^i (1-p)^{m-i} \binom{n}{k-i} p^{k-i} (1-p)^{n-k+i} \\
&= \sum_{i=0}^{k} \binom{m}{i} \binom{n}{k-i} p^k (1-p)^{m+n-k} \\
&= b(k; m+n, p) \text{ für } k = 0, 1 \ldots, m+n.
\end{aligned}$$

Beweis. (der benutzten Formel für die Binomialkoeffizienten)
Einerseits ist

$$(x+1)^{m+n} = \sum_{k=0}^{m+n} \binom{m+n}{k} x^k,$$

andererseits gilt

$$\begin{aligned}
(x+1)^{m+n} &= (x+1)^m (x+1)^n \\
&= \sum_{i=0}^{m} \binom{m}{i} x^i \sum_{j=0}^{n} \binom{n}{j} x^j \\
&= \sum_{i=0}^{m} \sum_{j=0}^{n} \binom{m}{i} \binom{n}{j} x^{i+j} \\
&= \sum_{i=0}^{m} \sum_{k=i}^{i+n} \binom{m}{i} \binom{n}{k-i} x^k \\
&= \sum_{k=0}^{m+n} \sum_{i=0}^{k} \binom{m}{i} \binom{n}{k-i} x^k.
\end{aligned}$$

In der letzten Gleichung wurden die Summationszeichen vertauscht. Es wurde außerdem $k = i + j$ verwendet. Ein Koeffizientenvergleich liefert die Behauptung $\square$

(ii) Seien X und Y unabhängig Poisson-verteilt mit dem Parameter λ bzw. μ $(\lambda, \mu > 0)$.

$$
\begin{aligned}
p_{X+Y}(n) &= \sum_{i=0}^{n} p_X(i) p_Y(n-i) \\
&= \sum_{i=0}^{n} \frac{\lambda^i}{i!} e^{-\lambda} \frac{\mu^{n-i}}{(n-i)!} e^{-\mu} \\
&= \frac{1}{n!} e^{-(\lambda+\mu)} \sum_{i=0}^{n} \binom{n}{i} \lambda^i \mu^{n-i} \\
&= \frac{(\lambda+\mu)^n}{n!} e^{-(\lambda+\mu)}
\end{aligned}
$$

Das heißt $X + Y$ sind Poisson-verteilt mit Parameter $\lambda + \mu$.

2. Seien X, Y beliebig unabhängige reellwertige Zufallsgrößen mit Dichten f_X, f_Y.

Sei $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$ die gemeinsame Dichte von X, Y.

$$
\begin{aligned}
F_{X+Y}(z) &= P(X + Y \leq z) \\
&= P((X, Y) \in B), \;\; B = \{(x, y) \in \mathbb{R}^2 : x + y \leq z\} \\
&= \iint_B f_X(x) f_Y(y) \, \mathrm{d}y \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y) \, \mathrm{d}x \, \mathrm{d}y
\end{aligned}
$$

Setzt man $y = \tilde{y} - x$, so erhält man für das innere Integral: $\int_{-\infty}^{z} f_Y(\tilde{y} - x) \, \mathrm{d}\tilde{y}$. Bei Vertauchung der Integrationsreihenfolge erhält man:

$$
\int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_X(x) f_Y(y - x) \, \mathrm{d}x \right) \mathrm{d}y \;\; \text{für alle } z.
$$

Das innere Integral ist gleich $f_{X+Y}(y)$. Somit besitzt F_{X+Y} eine Dichte und es gilt:

$$
f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_y(z - y) \, \mathrm{d}x, \;\; z \in \mathbb{R}
$$

ist die *Faltung* der Dichten f_X und f_Y. Kurzschreibweise: $f_{X+Y} = f_X * f_Y$
Die Faltung ist kommutativ: $f * g = g * f$

Beispiel 3.22. Seien X, Y unabhängig exponentialverteilt mit Parameter λ bzw. μ $(\lambda, \mu > 0)$ und $f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x)$ bzw. $f_Y(y) = \mu e^{-\mu y} \mathbb{1}_{[0,\infty)}(y)$ die zugehörigen Dichten. Dann ist

$$\begin{aligned} f_{X+Y}(z) &= \int_0^z \lambda e^{-\lambda x} \mu e^{-\mu(z-x)}\, \mathrm{d}x = \lambda\mu e^{-\mu z} \int_0^z e^{(\mu-\lambda)x}\, \mathrm{d}x, \quad z \geq 0; \\ f_{X+Y}(z) &= \begin{cases} \lambda\mu \frac{e^{-\lambda x} - e^{-\mu z}}{\mu - \lambda} \mathbb{1}_{[0,\infty)}(z), & \text{für } \lambda \neq \mu; \\ \lambda^2 z e^{-\lambda z} \mathbb{1}_{[0,\infty)}(z), & \text{für } \lambda = \mu. \end{cases} \end{aligned}$$

3.4. Erwartungswert und Varianz reeller Zufallsgrößen

Seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine *reelle* Zufallsgröße.
Sei zunächst X *diskret* mit reellen Werten x_n, $n \in \mathbb{N}$. Dann wird der *Erwartungswert* EX der Zufallsgröße X wie folgt definiert:

(i) Ist $X \geq 0$, das heißt $x_n \geq 0$ für alle n, so setzt man

$$EX := \sum_n x_n P(X = x_n) \in [0, \infty] \tag{3.6}$$

(also $EX = +\infty$ zugelassen).

(ii) Man sagt, X besitzt *endlichen Erwartungswert (der Erwartungswert von X existiert)*, falls $E|X| < \infty$ ist. In diesem Fall definiert man EX ebenfalls über (3.6).

Achtung: $E|X| = \sum_n |x_n| P(X = x_n)$, so dass $E|X| < \infty$ gleichbedeutend mit der absoluten Konvergenz der Reihe in (3.6) ist.

Interpretation des Erwartungswertes EX:
„Mittelwert“ der Zufallsgröße X ist das Mittel der Werte x_n gewichtet mit der Wahrscheinlichkeit $P(X = x_n)$ ihres Auftretens. Die Zufallsgröße X beschreibt die zufälligen „Abweichungen“ von diesem „Mittel“.

Beispiel 3.23. triviale Beispiele:

(i) Münzwurf: $P(X = 0) = P(X = 1) = \frac{1}{2} \Rightarrow EX = 0 \cdot P(X = 0)$
$+1 \cdot P(X = 1) = \frac{1}{2}$.

(ii) Würfeln: $P(X = i) = \frac{1}{6}$ $(i = 1, \ldots, 6) \Rightarrow EX = \sum_{i=1}^{6} iP(X = i) = \frac{1}{6} \sum_{i=1}^{6} i = 3,5$.

(iii) Sei $A \in \mathfrak{F}$, dann ist $X = \mathbb{1}_A$ eine Zufallsgröße mit $P(X = 1) = P(A)$ und $P(X = 0) = 1 - P(A)$. Daraus folgt für den Erwartungwert:

$EX = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = P(A)$, das heißt $E\mathbb{1}_A = P(A)$, $A \in \mathfrak{F}$

(iv) Sei $\phi : \mathbb{R} \to \mathbb{R}$ beliebig. Dann ist $\phi(X) = \phi \circ X$ eine diskrete Zufallsgröße mit dem Wertebereich $\{\phi(x_n) : n \in \mathbb{N}\} = \{y_k : k \in I\}$ für eine gewisse Indexmenge $I \subseteq \mathbb{N}$ und paarweise verschiedenen y_k.

(a) Sei $\phi \geq 0$:

$$\begin{aligned} E\phi(X) &= \sum_{k \in I} y_k P(\phi(X) = y_k) = \sum_{k \in I} y_k \sum_{n : \phi(x_n) = y_k} P(X = x_n) \\ &= \sum_{k \in I} \sum_{n : \phi(x_n) = y_k} \phi(x_n) P(X = x_n) \end{aligned}$$

Wobei die Doppelsumme gleich $\sum_n$ ist, also

$$E\phi(X) = \sum_n \phi(x_n)P(X = x_n). \tag{3.7}$$

(b) Ist $E|\phi(X)| = \sum_n |\phi(x_n)|P(X = x_n) < \infty$, so existiert $E\phi(X)$ und ist ebenfalls durch (3.7) gegeben.

Die Formel (3.7) wurde bereits vorher zur Berechnung von $E|X|$ herangezogen (mit $\phi(x) = |x|$).

Allgemeine Eigenschaften des Erwartungswertes: Seien X und Y beliebige (diskrete) Zufallsgrößen mit endlicher Erwartung.

(i) Seien $\alpha, \beta \in \mathbb{R}$. Dann besitzt $\alpha X + \beta Y$ ebenfalls einen endlichen Erwartungswert und $E(\alpha X + \beta Y) = \alpha EX + \beta EY$; (Linearität)

(ii) Sei $X \geq 0$, dann ist $EX \geq 0$; (Positivität)
Sei $X \geq Y$, dann ist auch $EX \geq EY$;

(iii) Es gilt $|EX| \leq E|X|$. (Dreiecksungleichung)

Beweis.

(i) Nehme X Werte x_m und Y Werte y_n an.
Wertebereich von $\alpha X + \beta Y$: $\{\alpha x_m + \beta y_n : m, n \in \mathbb{N}\} = \{z_k : k \in I\}$, $I \subseteq \mathbb{N}$, z_k paarweise verschieden. Mit $|\alpha x_m + \beta y_n| \leq |\alpha||x_m| + |\beta||y_n|$, $\sum_n P(X = x_m, Y = y_n) = P(X = x_m)$ und $\sum_m P(X = x_m, Y = y_n) = P(Y = y_n)$ folgt

$$\begin{aligned}
E|\alpha X + \beta Y| &= \sum_k |z_k| P(\alpha X + \beta Y = z_k) \\
&= \sum_k |z_k| \sum_{m,n:\alpha x_m + \beta y_n = z_k} P(X = x_m, Y = y_n) \\
&= \sum_k \sum_{m,n:\alpha x_m + \beta y_n = z_k} |\alpha x_m + \beta y_n| P(X = x_m, Y = y_n) \\
&= \sum_{m,n} |\alpha x_m + \beta y_n| P(X = x_m, Y = y_n) \\
&\leq |\alpha| \sum_m |x_m| \sum_n P(X = x_m, Y = y_n) \\
&\quad + |\beta| \sum_n |y_n| \sum_m P(X = x_m, Y = y_n) \\
&= |\alpha| E|X| + |\beta| E|Y| < \infty.
\end{aligned}$$

Dies sichert absolute Reihenkonvergenzen (und insbesondere Vertauschbarkeit der Summationsreihenfolgen) in analogen Ausdrücken für $E(\alpha X + \beta Y)$, und wir erhalten

$$E(\alpha X + \beta Y) = \alpha EX + \beta EY.$$

(ii) Sei $X \geq 0$, dann ist für alle $x_n \geq 0$ und $EX = \sum_n x_n P(X = x_n) \geq 0$.
Sei $X \geq Y$, dann ist $(X - Y) \geq 0$ und $EX - EY = E(X - Y) \geq 0$. Daraus folgt $EX \geq EY$.

(iii) $|EX| = |\sum_n x_n P(X = x_n)| \leq \sum_n |x_n| P(X = x_n) = E|X|$.

$\square$

Verallgemeinerung auf beliebige reelle Zufallsgrößen:

$$EX := \int_\Omega X \,\mathrm{d}P, \quad \text{falls} \quad \int_\Omega |X| \,\mathrm{d}P < \infty.$$

Dabei wird das Integral in der Maß- und Integrationstheorie wie folgt definiert:

(i) Sei $X \geq 0$ diskret:
$X = \sum_n x_n \mathbb{1}_{A_n}$ $(x_n \geq 0, A_n \in \mathfrak{F}$ für alle $n)$, wobei $A_n = \{X = x_n\}$
Dann ist:

$$\int X \,\mathrm{d}P := \sum_n x_n P(A_n).$$

(ii) Sei $X \geq 0$ beliebig, man approximiert X monoton durch diskrete Zufallsgrößen X_n:
$0 \leq X_n(\omega) \uparrow X(\omega)$ für alle $\omega \in \Omega$.
Somit ist:

$$\int X \,\mathrm{d}P := \lim_{n\to\infty} \int X_n \,\mathrm{d}P.$$

(iii) Sei X eine beliebige Zufallsgröße mit $\int |X| \,\mathrm{d}P < \infty$, sei $X(\omega) = X^+(\omega) - X^-(\omega)$ die Zerlegung in Positiv- und Negativteil.
Dann ist:

$$\int X \,\mathrm{d}P := \int X^+ \,\mathrm{d}P - \int X^- \,\mathrm{d}P.$$

Die Eigenschaften (i)-(iii) des Erwartungswertes übertragen sich auf den allgemeinen Fall.

Wir betrachten nun noch den wichtigen Spezialfall, dass X eine reelle Zufallsgröße mit Dichte f ist. Die folgenden Formeln werden ohne Beweis angeführt. Sie sind analog zum diskreten Fall, wobei man Summen durch Integrale und Einzelwahrscheinlichkeiten durch Dichten ersetzt:

$$EX = \int_{-\infty}^{\infty} x f(x) \,\mathrm{d}x, \quad \text{falls} \quad E|X| = \int_{-\infty}^{\infty} |x| f(x) \,\mathrm{d}x < \infty$$

(das heißt, falls das Integral absolut konvergiert).

Ist $\phi : \mathbb{R} \to \mathbb{R}$ Borel-messbar, so gilt:

$$E\phi(X) = \int_{-\infty}^{\infty} \phi(x) f(x)\,\mathrm{d}x, \text{ falls } \int_{-\infty}^{\infty} |\phi(x)| f(x)\,\mathrm{d}x < \infty.$$

Sprechweise: EX^p heißt *p-tes Moment* der Zufallsgröße X $(p = 1, 2 \ldots)$.

Satz 3.24. (Cauchy-Schwarzsche Ungleichung): *Seien X und Y beliebige Zufallsgrößen mit $EX^2 < \infty$ bzw. $EY^2 < \infty$. Dann ist auch $|EXY| < \infty$ und*

$$(E(XY))^2 \leq EX^2 EY^2.$$

Beweis. Idee: Sei $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$. Dann folgt auf Grund der Monotonie und Linearität des Erwartungswertes, dass $E|XY| \leq \frac{1}{2}EX^2 + \frac{1}{2}EY^2 < \infty$.
Dann ist $\mathfrak{L}^2 := \{X : X \text{ Zufallsgröße mit } EX^2 < \infty\}$ ein Vektorraum über $\mathbb{R}$ und ist $\langle X, Y\rangle := E(XY)$ ein Pseudoskalarprodukt in $\mathfrak{L}^2$:

(i) $\langle X, X\rangle \geq 0$ (aber $\langle X, X\rangle = 0 \not\Rightarrow X \equiv 0$)

(ii) $\langle X, Y\rangle = \langle Y, X\rangle$

(iii) $\langle X, \alpha Y + \beta Z\rangle = \alpha\langle X, Y\rangle + \beta\langle X, Z\rangle$.

Der Rest folgt wie in der linearen Algebra. $\square$

Der Spezialfall $Y \equiv 1$ liefert:

$$EX^2 < \infty \Rightarrow E|X| < \infty \text{ und } (EX)^2 \leq EX^2.$$

Definition 3.25. Sei X eine Zufallsgröße mit $E|X| < \infty$. Dann heißt

$$V(X) := E(X - EX)^2 \in [0, \infty]$$

Varianz (oder auch *Streuung*) der Zufallsgröße X.

Die Varianz misst die „mittlere quadratische Abweichung" einer Zufallsgröße von ihrem Erwartungswert.

Es gilt:

$$\begin{aligned} V(X) &\geq 0,\ V(\alpha X) = \alpha^2 V(X) \\ V(X) &< \infty \Leftrightarrow EX^2 < \infty \end{aligned} \tag{3.8}$$

denn:

$$\begin{aligned} X^2 &\leq 2(X - EX)^2 + 2(EX)^2, \\ (X - EX)^2 &\leq 2X^2 + 2(EX)^2 \end{aligned}$$

Nun kann man auf beiden Seiten den Erwartungswert anwenden.

Satz 3.26. (Steinerscher Satz) *Ist $E|X| < \infty$, so gilt*

$$V(X) = EX^2 - (EX)^2.$$

Beweis. Wegen (3.8) sei o.B.d.A $EX^2 < \infty$. Dann gilt:

$$\begin{aligned} V(X) &= E[X^2 - 2X \cdot EX + (EX)^2] \\ &= EX^2 - 2EX \cdot EX + (EX)^2 = EX^2 - (EX)^2. \end{aligned}$$

In die vorletzte Gleichung ging die Linearität des Erwartungswertes ein. □

Definition 3.27. Seien X und Y Zufallsgrößen, dann heißt

$$\begin{aligned} \operatorname{cov}(X,Y) &:= E(XY) - EX \cdot EY \\ &= E(X - EX)(Y - EY) \end{aligned}$$

Kovarianz, falls alle Erwartungswerte existieren.

Definition 3.28. Die Zufallsgrößen X und Y heissen *unkorreliert*, falls

$$\begin{aligned} E|X| < \infty,\ E|Y| &< \infty,\ E|XY| < \infty \text{ und} \\ \operatorname{cov}(X,Y) &= 0. \end{aligned}$$

Satz 3.29. *Seien $E|X| < \infty$, $E|Y| < \infty$ und X sowie Y unabhängig. Dann ist $E|XY| < \infty$ und $E(XY) = EX \cdot EY$*
(das heißt aus der Unabhängigkeit folgt die Unkorreliertheit).

Beweis. in Spezialfällen:

(i) Seien X, Y diskret mit Werten x_i bzw. y_j. Dann ist

$$\begin{aligned} E|XY| &= \sum_{i,j} |x_i y_j| P(X = x_i, Y = y_j) \\ &= \sum_{i,j} |x_i y_j| P(X = x_i) P(Y = y_j) \\ &= \sum_i |x_i| P(X = x_i) \sum_j |y_j| P(Y = y_j) \\ &= E|X| E|Y| < \infty \end{aligned}$$

Damit folgt analog:

$$E(XY) = EX \cdot EY$$

(ii) Seien X, Y stetig mit Dichten f_X und f_Y:

Da X, Y unabhängig sind, besitzen X und Y eine gemeinsame Dichte $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. Deshalb ist

$$\begin{aligned} E|XY| &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |xy| f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} |x| f_X(x)\,\mathrm{d}x \int_{-\infty}^{\infty} |y| f_Y(y)\,\mathrm{d}y \\ &= E|X|E|Y| < \infty \end{aligned}$$

usw.

(iii) allgemeiner Fall mit maßtheoretischen Hilfsmitteln.

□

Aus der Unkorreliertheit folgt im Allgemeinen *nicht* die Unabhängigkeit.

Seien $A_1, A_2 \subset A_3$ Ereignisse mit $A_1 \cap A_2 = \emptyset$ und $0 < P(A_1) = P(A_2) < P(A_3) < 1$ und $X := 1\!\!1_{A_1} - 1\!\!1_{A_2}$, $Y := 1\!\!1_{A_3}$ ihre Zufallsgrößen. Dann ist

$$\begin{aligned} XY &= 1\!\!1_{A_1 \cap A} - 1\!\!1_{A_2 \cap A_3} \\ &= 1\!\!1_{A_1} - 1\!\!1_{A_2} \\ &= X, \\ EX &= P(A_1) - P(A_2) \\ &= 0, \end{aligned}$$

und somit

$$\begin{aligned} E(XY) &= EX \cdot EY \\ &= 0. \end{aligned}$$

Aber X, Y sind nicht unabhängig:

$$\begin{aligned} P(X = 1, Y = 1) &= P(A_1 \cap A_3) \\ &= P(A_1) \\ &\neq P(A_1)P(A_3) \\ &= P(X = 1)P(Y = 1). \end{aligned}$$

Satz 3.30. (von Bienaymé) *Seien X_i paarweise unkorrelierte Zufallsgrößen $EX_i^2 < \infty$ $(i = 1, \ldots, n)$. Dann gilt:*

$$X_1, \ldots, X_n \Rightarrow V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i).$$

Beweis.

$$\begin{aligned}
V(\sum_{i=1}^n X_i) &= E\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)^2 \\
&= E\left(\sum_{i=1}^n (X_i - EX_i)\right)^2 \\
&= E\left[\sum_{i=1}^n (X_i - EX_i)^2 + \sum_{i\neq j}^n (X_i - EX_i)(X_j - EX_j)\right] \\
&= \sum_{i=1}^n V(X_i) + \sum_{i\neq j}^n E(X_i - EX_i)(X_j - EX_j)
\end{aligned}$$

Der letzte Summand ist gleich $\operatorname{cov}(X_i, Y_i) = 0$. Somit folgt die Behauptung. □

Für beliebige $a_1, \ldots, a_n \in \mathbb{R}$ folgt unter den Voraussetzungen des Satzes die paarweise Unkorreliertheit von $a_1X_1, \ldots, a_nX_n$ und damit

$$V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i).$$

Beispiel 3.31.

(i) Sei X eine Bernoulli-Variable, $P(X = 1) = p$ und $P(X = 0) = 1 - p$. Dann ist $EX = p$ und $V(X) = EX^2 - (EX)^2 = p - p^2 = p(1 - p)$.

(ii) Sei X binomialverteilt mit Parametern (n, p) und $q := 1 - p$.

$$\begin{aligned}
EX &= \sum_{k=1}^n k\binom{n}{k} p^k q^{n-k} \\
&= p\sum_{k=1}^n \binom{n}{k} kp^{k-1} q^{n-k} \\
&= p\frac{\partial}{\partial p}\left(\sum_{k=0}^n \binom{n}{k} p^k q^{n-k}\right) \\
&= p\frac{\partial}{\partial p}(p+q)^n \\
&= np(p+q)^{n-1}
\end{aligned}$$

In den letzten beiden Zeilen fasst man vorübergehend die Funktion als eine zweier Variablen

$p, q \in \mathbb{R}$ auf.

$$\begin{aligned} EX &= np \\ EX^2 &= \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} \\ &= p^2 \sum_{k=2}^{n} \binom{n}{k} k(k-1) p^{k-2} q^{n-k} + \sum_{k=1}^{n} \binom{n}{k} k p^k q^{n-k} \\ &= n(n-1)p^2 + np \end{aligned}$$

Nach dem Satz von Steiner folgt daraus

$$\begin{aligned} V(X) &= EX^2 - (EX)^2 \\ &= np(1-p). \end{aligned}$$

Einfacher fasst man X als Summe $\sum_{i=1}^{n} Y_i$ auf, wobei $Y_1, \ldots, Y_n$ unabhängig Bernoulli-verteilt mit Parameter p.

$$\begin{aligned} EX &= \sum_{i=1}^{n} EY_i = np, \\ V(X) &= \sum_{i=1}^{n} V(X_i) = np(1-p). \end{aligned}$$

(iii) Sei X normalverteilt mit Dichte $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (siehe Übung).

$$\begin{aligned} EX &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} \,\mathrm{d}x = 0 \\ EX^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} \,\mathrm{d}x \end{aligned}$$

Der Ausdruck im Integral ist gleich $-x \frac{\mathrm{d}}{\mathrm{d}x} e^{-x^2/2}$. Nach partieller Integration folgt:

$$EX^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \,\mathrm{d}x = 1$$

und somit

$$V(X) = 1.$$

(iv) Empirischer Mittelwert und emiprische Varianz:

Sei (X_n) eine Folge u.i.v. (unabhängig identisch verteilter) Zufallsgrößen mit $EX_1^2 < \infty$, $m := EX_1$ und $\sigma^2 := V(X_1)$. Seien außerdem:

$M_n := \frac{1}{n} \sum_{i=1}^{n} X_i$ das empirische Mittel von $X_1, \ldots, X_n$ und

$S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - M_n)^2$ die empirische Varianz von $X_1, \ldots, X_n$.

Dann ist $EM_n = m$ und nach Bienaymé $V(M_n) = \frac{1}{n^2}\sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n}$.

Dieser Term kovergiert für $n \to \infty$ gegen 0.

Anschaulich: Für $n \to \infty$ geht die mittlere quadratische Abweichung von M_n von m gegen 0, das heisst M_n kann als *„Schätzer“* für den Erwartungswert m von X_1 genommen werden.

$$
\begin{aligned}
\sum_{i=1}^n (X_i - M_n)^2 &= \sum_{i=1} n \left(X_i - \frac{1}{n}\sum_{j=1}^n X_j \right)^2 \\
&= \sum_{i=1}^n X_i^2 - \frac{2}{n}\sum_{i,j=1}^n X_i X_j + \frac{1}{n}\sum_{i,j=1} n X_i X_j \\
&= (1 - \frac{1}{n})\sum_{i=1} n X_i^2 - \frac{1}{n}\sum_{i,j=1; i \neq j} X_i X_j
\end{aligned}
$$

Dann folgt mit dem Satz von Steiner:

$$
\begin{aligned}
E\sum_{i=1}^n (X_i - M_n)^2 &= (n-1)EX_i^2 - (n-1)(EX_1) \\
&= (n-1)V(X_1),
\end{aligned}
$$

das heißt

$$ES_n^2 = \sigma^2.$$

Ist $EX_1^4 < \infty$, so lässt sich zeigen:

$$V(S_n^2) \leq \frac{c}{n}.$$

Dieser Term geht für $n \to \infty$ gegen 0, wobei c eine positive Konstante bezeichnet.

Kapitel 4

Erzeugende und charakteristische Funktionen

4.1. Erzeugende Funktionen

Sei X eine diskrete Zufallsgröße mit Werten in $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ und $p_n := P(X = n), \quad n \in \mathbb{N}$ die zugehörigen Einzelwahrscheinlichkeiten.

Definition 4.1.

(i) Die Funktion

$$G_X(s) := Es^X = \sum_{n=0}^{\infty} p_n s^n$$

heißt *erzeugende Funktion* von X. Der Definitionsbereich von G_X ist der Konvergenzbereich der rechtsstehenden Potenzreihe.

(ii) Ist $P = (p_n)_{n \in \mathbb{N}_0}$ eine beliebige Verteilung auf $\mathbb{N}_0$, so heißt

$$G(s) := \sum_{n=0}^{\infty} p_n s^n \tag{4.1}$$

entsprechend *erzeugende Funktion* der Verteilung P.

Es gilt

$$\sum_{n=0}^{\infty} |p_n s^n| \leq \sum_{n=0}^{\infty} p_n |s|^n \leq \sum_{n=0}^{\infty} p_n < \infty \text{ für } |s| \leq 1.$$

Also konvergiert (4.1) für $|s| \leq 1$ und ist für $|s| < 1$ beliebig oft gliedweise differenzierbar. Insbesondere ist

$$p_n = \frac{G^{(n)}(0)}{n!},$$

das heisst *jede Verteilung auf* $\mathbb{N}_0$ *ist durch ihre erzeugende Funktion eindeutig bestimmt.*

Beispiel 4.2.

(i) Binomialverteilung mit Parametern (n, p): $q := 1 - p$, $p_k = \binom{n}{k} p^k q^{n-k}$ $(k = 0, 1, \ldots, n)$

$$\begin{aligned} G(s) &= \sum_{k=0}^{n} \binom{n}{k} (ps)^k q^{n-k} \\ &= (ps + q)^n, \quad s \in \mathbb{R}. \end{aligned}$$

(ii) Poissonverteilung mit dem Parameter λ: $p_n = \frac{\lambda^n}{n!} e^{-\lambda}$, $n \in \mathbb{N}_0$.

$$\begin{aligned} G(s) &= \sum_{n=0}^{\infty} \frac{(\lambda s)^n}{n!} e^{-\lambda} \\ &= e^{\lambda(s-1)}, \quad s \in \mathbb{R}. \end{aligned}$$

(iii) Sei $p_n = c/n^2$ $(n = 1, 2, \ldots)$, wobei c so gewählt ist, dass $\sum_{n=1}^{\infty} p_n = 1$ $(c = 6/\pi^2)$. Dann ist $G(s) = c \sum_{n=1}^{\infty} s^n/n^2$ mit Konvergenzradius 1. (Insbesondere existiert $G'(1)$ nicht!)

Satz 4.3. *(Berechnung von Momenten) Sei G die erzeugende Funktion von X.*

(i) $EX = G'(1-)$ $(= \lim_{s \uparrow 1} G'(s))$. *Hierbei sei* $+\infty$ *zugelassen.*

(ii) Ist $G'(1-) < \infty$, *so gilt*

$$V(X) = G''(1-) + G'(1-)(1 - G'(1-)).$$

Insbesondere ist der Erwartungswert genau dann endlich, wenn $G'(1-) < \infty$, *und die Varianz ist endlich, wenn zusätzlich* $G''(1-) < \infty$ *ist.*

Beweis.

(i) Für $|s| < 1$ lässt sich G gliedweise differenzieren:

$$G'(s) \quad = \quad \sum_{n=1}^{\infty} np_n s^{n-1}.$$

Diese Reihe ist nach dem Satz von Abel monoton konvergent gegen

$$\sum_{n=1}^{\infty} np_n \quad = \quad EX \text{ für } s \uparrow 1.$$

(ii) Sei $G'(1-) = \Sigma_n np_n < \infty$. Für $|s| < 1$ erhält man

$$G''(s) \quad = \quad \sum_{n=2}^{\infty} n(n-1)p_n s^{n-2},$$

Dies ist wieder nach dem Satz von Abel für $s \uparrow 1$ konvergent gegen

$$\begin{aligned} \sum_{n=1}^{\infty} n(n-1)p_n \quad &= \quad \sum_{n=1}^{\infty} n^2 p_n - \sum_{n=1}^{\infty} np_n \\ &= \quad EX^2 - EX \end{aligned}$$

Nach dem Satz von Steiner ist

$$G''(1-) = V(X) + (EX)^2 - EX = V(X) + G'(1-)^2 - G'(1-).$$

□

Beispiel 4.4. Sei X binomialverteilt mit dem Parameter (n, p) und $q := 1 - p$, $G_X(s) = (ps + q)^n$ ist auf ganz $\mathbb{R}$ definiert. Deshalb ist

$$G'_X(1-) = G'_X(1) = np$$

und

$$G''_X(1-) = G''_X(1) = n(n-1)p^2.$$

Folglich erhalten wir mit dem obigen Satz

$$EX = np$$

und

$$V(X) = n(n-1)p^2 + np(1 - np) = np(1 - p).$$

Satz 4.5. *Seien $X_1, \ldots, X_n$ unabhängige $\mathbb{N}_0$-wertige Zufallsgrößen. Dann gilt*

$$G_{\sum_{i=1}^n X_i}(s) = \prod_{i=1}^n G_{Xi}(s), \ |s| \leq 1.$$

Beweis. Nur für $n = 2$. Der allgemeine Fall wird durch vollständige Induktion bewiesen; $X_1, \ldots, X_n$ unabhängig $\Rightarrow$ X_1 und $(X_2, \ldots, X_n)$ unabhängig.
Da X, Y unabhängig sind, sind auch s^X, s^Y für beliebige s unabhängig und es folgt

$$G_{X+Y}(s) = Es^{X+Y} \underset{4.2}{=} Es^X \cdot Es^Y = G_X(s) \cdot G_Y(s) \ \ (|s| \leq 1). \tag{4.2}$$

In (4.2) wurde verwendet, dass aus der Unabhängigkeit die Unkorreliertheit folgt.

Alternativbeweis des Satzes mittels der Faltungsformel $p_{X+Y} = p_X * p_Y$.

$$\begin{aligned}
G_{X+Y} &= \sum_{n=0}^{\infty} p_X * p_Y(n) s^n \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{n} p_X(k) p_Y(n-k) s^n \\
&= \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} (p_X(k) s^k)(p_Y(n-k) s^{n-k}) \\
&\underset{(*)}{=} \sum_{k=0}^{\infty} p_X(k) s^k \sum_{l=0}^{\infty} p_Y(l) s^l \\
&= G_X(s) G_Y(s).
\end{aligned}$$

In die vorletzte Gleichung (*) wurde diese Substitution durchgeführt $n - k = l$. Es bleibt noch die absolute Konvergenz der Reihe zu zeigen. □

Beispiel 4.6.

(i) Seien $Y_1, \ldots, Y_n$ unabhängig Bernoulli-verteilt mit den Erfolgswahrscheinlichkeiten p, $q := 1 - p$. Dann ist $X = Y_1 + \ldots + Y_n$ binomialverteilt mit den Parametern (n, p).
Da $G_{Y_i}(s) = qs^0 + ps^1 = ps + q$, ist $G_X(s) = (G_{Y_i}(s))^n = (ps + q)^n$.

(ii) Verteilung von Summen mit zufälliger Summandenzahl. $N, X_1, X_2, \ldots$ seien unabhängige $\mathbb{N}_0$-wertige Zufallsgrößen.

$X_1, X_2, \ldots$ seien identisch verteilt, $Y := \Sigma_{i=1}^N X_i$ und G_N, G_X, G_Y sind erzeugende Funktionen von N, X_i bzw. Y.

Die erzeugende Funktion von Y lässt sich wie folgt berechnen:

$$G_Y(s) = Es^{\sum_{i=1}^N X_i} = E \sum_{n=0}^{\infty} \mathbb{1}_{\{N=n\}} s^{\sum_{i=1}^n X_i} = \sum_{n=0}^{\infty} E \mathbb{1}_{\{N=n\}} s^{\sum_{i=1}^n X_i},$$

wobei die Faktoren unabhängig sind,

$$\sum_{n=0}^{\infty} E\mathbb{1}_{\{N=n\}} \cdot Es^{\sum_{i=1}^{n} X_i} = \sum_{n=0}^{\infty} P(N=n)G_X(s)^n = G_N(G_X(s)), \quad |s| \le 1.$$

Also ist

$$G_Y = G_N \circ G_X.$$

Bei der Herleitung wurde (insbesondere zur Vertauschung der Summationsreihenfolge) die absolute Konvergenz benutzt. Die Reihe ist tatsächlich absolut konvergent, da

$$E\sum_{n=0}^{\infty} |\mathbb{1}_{\{N=n\}} s^{\sum_{i=1}^{n} X_i}| \le E\sum_{n=0}^{\infty} \mathbb{1}_{\{N=n\}} = 1 \text{ für } |s| \le 1,$$

also

$$|G_X(s)| \le 1 \text{ für } |s| \le 1.$$

(iii) Verzweigungsprozesse (Übung).

Satz 4.7. *Seien $P_n = (p_k^{(n)})$, $k \in \mathbb{N}_0$, $n \in \mathbb{N}_0$ eine Folge von Verteilungen auf $\mathbb{N}_0$ und G_n die zugehörige erzeugende Funktion. Dann gilt:*
Der Limes $p_k := \lim\limits_{n\to\infty} p_k^{(n)}$ existiert für alle $k \in \mathbb{N}_0$, genau dann wenn der Limes $G(s) := \lim\limits_{n\to\infty} G_n(s)$ existiert für $|s| < 1$.
Ist eine der beiden Aussagen (und damit auch die andere) erfüllt, so gilt $p_k \ge 0$ ($k \in \mathbb{N}_0$), $\sum_{k=0}^{\infty} p_k \le 1$ und $G(s) = \sum_{k=0}^{\infty} p_k s^k$ für $|s| < 1$.

Bemerkung 4.8.

(i) Für

$$p_0^{(0)} = 1, \; p_k^{(n)} := \begin{cases} \frac{1}{2}, & k=0, \; k=n; \\ 0, & \text{sonst} \end{cases}$$

ist

$$\lim_{n\to\infty} p_k^{(n)} =: p_k = \begin{cases} \frac{1}{2}, & k=0; \\ 0, & \text{sonst} \end{cases}$$

und damit $\sum_{k=0}^{\infty} p_k < 1$.

(ii) Im Falle $\sum_{k=0}^{\infty} p_k = 1$ kann man auf der rechten Seite der Gleichung des Satzes $|s| < 1$ durch $|s| \le 1$ ersetzen.

Beweis. „$\Rightarrow$“: Angenommen, $p_k := \lim\limits_{n\to\infty} p_k^{(n)}$ existiert für alle k. Dann gilt $p_k \geq 0$ $(k \in \mathbb{N}_0)$ und $\sum_{k=0}^{\infty} p_k \leq 1$. $\left(\sum_{k=0}^{m} p_k = \lim\limits_{n\to\infty} \sum_{k=0}^{m} p_k^{(n)} \leq 1 \text{ für alle } m\right)$
Sei $G(s) := \sum_{k=0}^{\infty} p_k s^k$ für $|s| < 1$. Dann erhält man mit der Dreiecksungleichung

$$\begin{aligned} |G_n(s) - G(s)| &\leq \sum_{k=0}^{\infty} |p_k^{(n)} - p_k||s|^k \\ &\leq \sum_{k=0}^{m} |p_k^{(n)} - p_k| + \sum_{k=m+1}^{\infty} |s|^k \end{aligned}$$

Wobei der erste Summand in der letzen Ungleichung für beliebige m und $n \to \infty$ gegen Null geht. Der zweite Summand geht gegen Null für $m \to \infty$.
Man kann also zuerst die zweite Summe auf der rechten Seite beliebig klein machen, indem man m hinreichend groß wählt. Für dieses m wird dann auch die erste Summe beliebig klein für große n. Also ist

$$\lim_{n\to\infty} G_n(s) = G(s) \text{ für } |s| < 1.$$

„$\Leftarrow$“: Der Grenzwert $\lim\limits_{n\to\infty} G_n(s) =: G(s)$ existiert für $|s| < 1$. Für beliebige $m \in \mathbb{N}$ ist
$G_n(s) = \sum_{k=0}^{m-1} p_k^{(n)} s^k + p_k^{(n)} s^m + \sum_{k=m+1}^{\infty} p_k^{(n)} s^k$ und folglich

$$\left| \frac{G_n(s) - \sum_{k=0}^{m-1} p_k^{(n)} s^k}{s^m} - p_m^{(n)} \right| \leq \frac{|s|}{1-|s|}, \quad 0 < |s| < 1. \tag{4.3}$$

Wir zeigen mit Induktion nach k, dass

$$\lim_{n\to\infty} p_k^{(n)} =: p_k \text{ existiert.} \tag{4.4}$$

Für $k = 0$ gilt diese Aussage:

$$p_0^{(n)} = G_n(0) \underset{n\to\infty}{\longrightarrow} G(0)$$

Angenommen (4.4) gilt für $k = 0, 1, \ldots, m-1$. Wir zeigen, dass dies auch für $k = m$ richtig ist. Durch Grenz“ubergang für $n \to \infty$ folgt aus 4.3

$$\begin{aligned} \frac{G_n(s) - \sum_{k=0}^{m-1} p_k^{(n)} s^k}{s^m} - \frac{|s|}{1-|s|} &\leq \varliminf_{n\to\infty} p_m^{(n)} \\ &\leq \varlimsup_{n\to\infty} p_m^{(n)} \\ &\leq \frac{G_n(s) - \sum_{k=0}^{m-1} p_k^{(n)} s^k}{s^m} + \frac{|s|}{1-|s|} \end{aligned}$$

Die Differenz zwischen den Ausdrücken der linken und der rechten Seite konvergiert für $s \to \infty$ gegen Null. Deshalb ist

$$\underline{\lim}_{n\to\infty} p_m^{(n)} \leq \overline{\lim}_{n\to\infty} p_m^{(n)},$$

das heißt der Grenzwert (4.3) existiert auch für $k = m$. $\square$

Auwendung: Alternativer Beweis des Poissonschen Grenzwertsatzes. Seien X_n binomialverteilte Zufallsgrößen mit Parameter (n, p_n), wobei $\lim_{n\to\infty} np_n =: \lambda > 0$ gelte.
Zu zeigen ist

$$\lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0.$$

Seien $G_n(s) = (1 + p_n(s-1))^n$ die erzeugende Funktion von X_n und $G(s) = e^{\lambda(s-1)}$ die erzeugende Funktion einer Poissonverteilung mit Parameter λ. Nach unserem Konvergenzsatz genügt es zu beweisen, dass

$$\lim_{n\to\infty} G_n(s) = G(s), \quad \text{für } |s| < 1 \text{ ist.}$$

Wir erhalten mit $\lambda_n := np_n$

$$G_n(s) = \left[(1 + p_n(s-1))^{1/(p_n(s-1))}\right]^{\lambda_n(s-1)} \underset{n\to\infty}{\to} e^{\lambda(s-1)}, \quad s \in \mathbb{R}.$$

Hierbei wurde benutzt, dass $\lim_{x\to 0}(1 + x)^{1/x} = e$ und $\lim_{n\to\infty} \lambda_n = \lambda$ ist. (Aus letzterem folgt insbesondere $p_n \to 0$.)

4.2. Charakteristische Funktion

Vorbetrachtungen: Erzeugende Funktionen sind nur für $\mathbb{N}_0$-wertige Zufallsgrößen X definiert. Eine solche Funktion

$$G(s) = \sum_{n=0}^{\infty} p_n s^n = Es^X$$

ist analytisch im offenen Einheitskreis ($|s| < 1$) der *komplexen* Ebene und stetig auf dem Rand $|s| = 1$ fortsetzbar. Aus der komplexen Funktionentheorie ist bekannt, dass eine solche Funktion durch ihre Werte auf dem Rand eindeutig festgelegt ist. Diese lassen sich in der Form $s = e^{it}$, $t \in \mathbb{R}$, schreiben. Deshalb können wir alternativ zur erzeugenden Funktion die Funktion

$$\chi(t) = Ee^{itX}, \ \ t \in \mathbb{R},$$

betrachten. Der Vorteil ist, dass eine solche Funktion für beliebige Zufallsgrößen X definiert ist.

Definition 4.9. Sei X eine beliebige reelle Zufallsgröße. Dann heißt die komplexwertige Funktion

$$\chi_X(t) := Ee^{itX}, \ \ t \in \mathbb{R},$$

charakteristische Funktion der Zufallsgröße X.

Es gilt $\chi_X(0) = 1$, $|\chi_X(t)| \leq E|e^{itX}| = 1$ $(t \in \mathbb{R})$. Sei X weiterhin diskret, $\sum_n P(X = x_n) = 1$, dann ist $\chi_X(t) = \sum_n P(X = x_n)e^{itx_n}$. Sei X stetig mit der Dichte f_X, dann ist $\chi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\,\mathrm{d}x$ (Fourier-Transformierte von f_X; aber i.a. $f_X \notin L^2$)

Eigenschaften der charakteristischen Funktionen

(i) Die charakteristische Funktion bestimmt die Verteilung einer Zufallsgröße eindeutig:

$$F_X = F_Y \Leftrightarrow \chi_X = \chi_Y$$

(ii) Seien $X_1, \ldots, X_n$ unabhängige Zufallsgrößen daraus folgt

$$\chi_{\sum_{i=1}^n X_i} = \prod_{i=1}^n \chi_{X_i}$$

(iii) *Konvergenzsatz:* Seien $X, X_1, X_2, \ldots$ Zufallsgrößen. Dann ist $\lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x)$ für alle Stetigkeitpunkte x von F, genau dann wenn $\lim\limits_{n\to\infty} \chi_{X_n}(t) = \chi_X(t)$ für alle $t \in \mathbb{R}$.

F ist rechtsstetig, jedoch nicht unbedingt stetig: Für $X_n \equiv \frac{1}{n}$ und $X \equiv 0$ gilt $F_{X_n}(x) \to F_X(x)$, für $x = 0$ ist jedoch $\chi_{X_n}(t) = e^{it/n} \to 1 = \chi_X(t)$ und $F_{Xn}(0) = 0$, $F_X(0) = 1$.

(iv) *Momente:* Besitzt X ein n-tes Moment (d.h. $E|X|^n < \infty$), so folgt

$$EX^n = \frac{1}{i^n} \chi_X^{(n)}(0), \quad n \in \mathbb{N}.$$

Die Umkehrung ist unter gewissen Einschränkungen ebenfalls richtig.

Kapitel 5

Gesetze der großen Zahlen

Satz 5.1. Chebyshevsche Ungleichung

(i) Für eine beliebige Zufallsgröße X und beliebiges $a > 0$ gilt

$$P(|X| \geq a) \leq \frac{1}{a} E|X|.$$

(ii) Für eine beliebige Zufallsgröße X mit endlichem Erwartungswert und beliebigem $a > 0$ gilt

$$P(|X - EX| \geq a) \leq \frac{1}{a^2} V(X).$$

Beweis. Da $\mathbb{1}_{\{|X| \geq a\}} \leq \frac{1}{a}|X|$ folgt

(i)

$$\begin{aligned} P(|X| \geq a) &= E\mathbb{1}_{\{|X| \geq a\}} \\ &\leq E\frac{1}{a}|X| \leq E\frac{|X|}{a} \\ &= \frac{1}{a}E|X| \end{aligned}$$

(ii) Man wende (i) auf die Zufallsgröße $(X - EX)^2$ an:

$$\begin{aligned} P(|X - EX| \geq a) &= P(|X - EX|^2 \geq a^2) \\ &\leq \frac{1}{a^2} E|X - EX|^2 \\ &= \frac{1}{a^2} V(X). \end{aligned}$$

□

Definition 5.2. Seien $X, X_1, X_2, \ldots$ beliebige (reelle) Zufallsgrößen. Man sagt, die Folge (X_n) konvergiert *in Wahrscheinlichkeit* (bzw. *stochastisch*) gegen X, falls

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \text{ für alle } \varepsilon > 0$$

gilt.
Schreibweise: $X_n \xrightarrow{P} X$, $X = P\text{-}\lim\limits_{n\to\infty} X_n$

Satz 5.3. (Schwaches Gesetz der großen Zahlen) *Sei (X_n) eine Folge unkorrelierter, identisch verteilter Zufallsgrößen mit endlicher Varianz. Dann gilt:*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} EX_1 \text{ für } n \to \infty.$$

Beweis. Seien $m := EX_1$ und $\varepsilon > 0$ beliebig. Dann ist auf Grund der identischen Verteilung $m = E\frac{1}{n}\sum_{i=1}^{n} X_i$ und

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2} V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \\ &= \frac{1}{\varepsilon^2 n} V(X_1). \end{aligned}$$

Die erste Ungleichung folgt aufgrund der Chebyshevschen Ungleichung. Die erste Gleichheit folgt nach Bienaymé und die zweite Gleichheit gilt wegen der identischen Verteilung von X_i. Für n gegen unendlich konvergiert der letzte Term gegen 0. □

Beispiel 5.4. Bernoullisches Versuchsschema mit Einzelwahrscheinlichkeit p. Sei (X_n) u.i. Bernoulli-verteilt mit dem Parameter p. Dann sind die (X_n) auch paarweise unkorreliert. Sei weiter $S_n := \sum_{i=1}^{n} X_i$ die zufällige Anzahl der Erfolge in n Versuchen. Das schwache Gesetz der Großen Zahlen liefert $\frac{S_n}{n} \xrightarrow{P} p$.

Im folgenden wollen wir einen weiteren Konvergenzbegriff für Folgen von Zufallsgrößen einführen, der sich als stärker als die Konvergenz in Wahrscheinlichkeit erweisen wird.
Nachtrag zum schwachen Gesetz der großen Zahlen:
Abrücken von der Forderung der identischen Verteilung der Zufallsgrößen.

Verallgemeinerung: Für jedes $n \in \mathbb{N}$ seien $X_1^{(n)}, \ldots, X_n^{(n)}$ paarweise unkorrelierte Zufallsgrößen mit endlicher Varianz. Gilt

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i^{(n)}) = 0,$$

so genügen die $X_i^{(n)}$ dem folgenden schwachen Gesetz der großen Zahlen:

$$\frac{1}{n}\sum_{i=1}^{n}(X_i^{(n)} - EX_i^{(n)}) \xrightarrow{P} 0 \text{ für } n \to \infty.$$

Beweis.

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(n)} - EX_i^{(n)})\right| > \varepsilon\right) &\le \frac{1}{\varepsilon^2}V\left(\frac{1}{n}\sum_{i=1}^{n}X_i^{(n)}\right) \\ &= \frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n}V(X_i^{(n)}) \end{aligned}$$

Dies konvergiert für $n \to \infty$ gegen Null. □

Definition 5.5. Seien $X, X_1, X_2, \ldots$ beliebige Zufallsgrößen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, P)$. Man sagt, die Folge (X_n) *konvergiert fast sicher* (f.s.) gegen X, falls

$$P(X_n \to X) = 1.$$

Schreibweise: $X_n \xrightarrow{f.s.} X$, $X = \lim\limits_{n\to\infty} X_n$ f.s.

Konvergenz fast sicher ist gleichbedeutend mit

$$P(X_n \nrightarrow X) = 0.$$

Um diesen Konvergenzbegriff mit der Konvergenz in Wahrscheinlichkeit zu vergleichen, benötigen wir eine geeignete Darstellung für das Ereignis

$$\{X_n \to X\} = \{\omega \in \Omega : X_n(\omega) \to X\}$$

bzw. für das Komplementärereignis $\{X_n \nrightarrow X\}$.
Wegen

$$X_n(\omega) \to X(\omega) \Leftrightarrow \forall m \ \exists k \ \forall l \ge k : |X_l(\omega) - X(\omega)| \le \frac{1}{m}$$

ist

$$\{X_n \to X\} = \bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{l\ge k}\{|X_l - X| \le \frac{1}{m}\}$$

und deshalb

$$\begin{aligned} \{X_n \nrightarrow X\} &= \bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{l\ge k}\{|X_l - X| > \frac{1}{m}\} \\ &= \bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\{\sup_{l\ge k}|X_l - X| > \frac{1}{m}\} \end{aligned}$$

Die Ereignisse in den geschweiften Klammern sind monoton fallend in k. Deshalb lässt sich auf diese der Stetigkeitssatz für Wahrscheinlichkeitmaße anwenden. Wir erhalten

$$\begin{aligned} X_n \overset{f.s.}{\to} X \quad &\Leftrightarrow \quad P(X_n \nrightarrow X) = 0 \\ &\Leftrightarrow \quad P(\bigcap_{k=1}^{\infty} \{\sup_{l\geq k} |X_l - X| > \frac{1}{m}\}) = 0, \;\; \forall m \\ &\Leftrightarrow \quad \lim_{k\to\infty} P(\sup_{l\geq k} |X_l - X| > \frac{1}{m}) = 0, \;\; \forall m \\ &\Leftrightarrow \quad \sup_{l\geq k} |X_l - X| \overset{P}{\to} 0 \;\text{ für }\; k \to \infty. \end{aligned}$$

Die zweite Äquivalenz folgt auf Grund der Subadditivität von P und die dritte Äquivalenz aufgrund der Stetigkeit bezüglich monotoner Mengenfolgen.
Damit haben wir folgende Aussage erhalten.
Behauptung:

$$X_n \overset{f.s.}{\to} X \Leftrightarrow \sup_{j\geq n} |X_j - X| \overset{P}{\to} 0 \;\text{ für }\; n \to \infty.$$

Nun ist aber

$$\sup_{j\geq n} |X_j - X| \geq |X_n - X|$$

und damit

$$\begin{aligned} \{\sup_{j\geq n} |X_j - X| > \varepsilon\} \quad &\supseteq \quad \{|X_n - X| > \varepsilon\}, \\ P(\sup_{j\geq n} |X_j - X| > \varepsilon) \quad &\geq \quad P(|X_n - X| > \varepsilon), \;\; \forall \varepsilon > 0. \end{aligned}$$

Kombinieren wir dieses Resultat mit der obigen Behauptung, so ergibt sich der Zusammenhang zwischen unseren beiden Konvergenzarten.

Behauptung:

$$X_n \overset{f.s.}{\to} X \Rightarrow X_n \overset{P}{\to} X.$$

Aus der fast sicheren Konvergenz folgt die Konvergenz in Wahrscheinlichkeit.

Satz 5.6. (Starkes Gesetz der großen Zahlen) *(Kolmogorov 1930): Sei (X_n) eine Folge unabhängiger, identisch verteilter Zufallsgrößen mit endlichem Erwartungswert. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \overset{f.s.}{\to} EX_1 \;\; \textit{für} \;\; n \to \infty.$$

Bemerkung 5.7.

(i) Unter den obigen Voraussetzungen gilt auch ein schwaches Gesetz der großen Zahlen:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} EX_1$$

(ii) Es existieren Verallgemeinerungen auf Folgen unabhängiger, aber nicht identisch verteilter Zufallsgrößen (siehe Literatur).

Beweis. (für $E|X_1|^4 < \infty$): Die Zufallsgrößen $Y_n := X_n - EX_n$ sind ebenfalls unabhängig, identisch verteilt und $EY_n = 0$.
Zu zeigen ist:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{f.s.} 0 \text{ für } n \to \infty.$$

Dies ist gleichbedeutend mit

$$\lim_{m\to\infty} P\left(\sup_{n\geq m}\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon\right) = 0 \ \forall \varepsilon > 0.$$

Das Argument von P ist gleich

$$\bigcup_{n\geq m}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon\right\}$$

Wegen

$$P\left(\sup_{n\geq m}\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon\right) \leq \sum_{n=m}^{\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon\right)$$

bleibt nur

$$\sum_{n=1}^{\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon\right) < \infty, \ \forall \ \varepsilon > 0 \tag{5.1}$$

zu zeigen.
Mit der Chebyshevschen Ungleichung erhält man, dass der Erwartungswert $E(Y_iY_jY_kY_l)$ Null ist, falls ein Index von allen anderen verschieden ist (Un-

abhägigkeit, $EY_i = 0$).

$$\begin{aligned}
P\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i\right| > \varepsilon\right) &= P\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i\right|^4 > \varepsilon^4\right) \\
&\leq \frac{1}{\varepsilon^4} E\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)^4 \\
&= \frac{1}{\varepsilon^4 n^4} \sum_{i,j,k,l=1}^n E(Y_i Y_j Y_k Y_l) \\
&= \frac{1}{\varepsilon^4 n^4} \left[\sum_{i=1}^n EY_i^4 + 3 \sum_{i,j=1; i\neq j} E(Y_i^2 Y_j^2)\right] \\
&= \frac{1}{\varepsilon^4 n^4} \left[nEY_1^4 + 3n(n-1)(EY_1^2)^2\right] \\
&\leq \frac{const.}{n^2}.
\end{aligned}$$

In die vorletzte Gleichung ging ein, dass aufgrund der Unabhängigkeit von Y_i, $E(Y_i^2 Y_j^2) = EY_i^2 EY_j^2$. Da $\sum_{n=1}^\infty \frac{1}{n^2} < \infty$, folgt (5.1). □

Sei (A_n) eine Folge von Ereignissen und $A := \bigcap_m \bigcup_{n\geq m} A_n = \limsup A_n$ das Ereignis, dass unendlich viele Ereignisse A_n eintreten ($\omega \in A \Leftrightarrow \omega \in A_n$ für unendlich viele n).

Lemma 5.8. *Borel-Cantelli:*

(i)

$$\sum_n P(A_n) < \infty \Rightarrow P(A) = 0$$

(das heißt fast sicher treten nur endlich viele der Ereignisse A_n ein)

(ii)

$$\sum_n P(A_n) = \infty \ \textit{und}\ A_n \ \textit{unabhängig} \Leftrightarrow P(A) = 1$$

(das heißt fast sicher treten unendlich viele der Ereignisse A_n ein).

Beweis.

(i) Aufgrund der Stetigkeit bezüglich monotoner Mengenfolgen, folgt aus

$$\bigcup_{n\geq m} A_n \downarrow A \text{ für } m \uparrow \infty,$$

dass mit der Subadditivität

$$\begin{aligned} P(A) &= \lim_{m\to\infty} P\left(\bigcup_{n\geq m} A_n\right) \\ &\leq \lim_{m\to\infty} \sum_{n=m}^{\infty} P(A_n) = 0, \end{aligned}$$

da $\sum_n P(A_n) < \infty$.

(ii) Sei $A^c = \bigcup_m \bigcap_{n\geq m} A_n^c$. Zu zeigen: $P(A^c) = 0$.
Es genügt zu zeigen, dass für jedes m

$$P\left(\bigcap_{n\geq m} A_n^c\right) = 0$$

ist. Da die Ereignisse A_n unabhängig sind, sind auch die Komplementäreignisse A_n^c unabhängig. Somit ist wegen $1 - x \leq e^{-x}$

$$\begin{aligned} P\left(\bigcap_{n=m}^{M} A_n^c\right) &= \prod_{n=m}^{M} P(A_n^c) \\ &= \prod_{n=m}^{M} [1 - P(A_n)] \\ &\leq \prod_{n=m}^{M} e^{-P(A_n)} \\ &= \exp\left\{-\sum_{n=m}^{M} P(A_n)\right\} \end{aligned}$$

Durch Grenzübergang für $M \to \infty$ folgt hieraus mit dem Stetigkeitssatz für Wahrscheinlichkeitsmaße und wegen $\sum_{n=m}^{\infty} P(A_n) = \infty$:

$$P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 0$$

□

Die Konvergenz $X_n \stackrel{f.s.}{\to} X$ hängt wie folgt mit dem Lemma von Borel-Cantelli zusammen:

Wir definieren die Mengen

$$\begin{aligned} A_n^\varepsilon &:= \{|X_n - X| > \varepsilon\} \\ A^\varepsilon &:= \bigcap_k \bigcup_{l\geq k} A_l^\varepsilon \text{ (unendlich viele der Ereignisse } A_n^\varepsilon \text{ treten auf).} \end{aligned}$$

Wir wissen bereits

$$\{X_n \nrightarrow X\} = \bigcup_m A^{1/m},$$

das heisst

$$X_n \overset{f.s.}{\to} X \Leftrightarrow P(A^{1/m}) = 0, \ \forall m \in \mathbb{N} \quad \Leftrightarrow \quad P(A^{\varepsilon}) = 0 \ \forall \ \varepsilon > 0.$$

Hinreichend hierfür ist nach Borel-Cantelli

$$\sum_n P(A_n^{\varepsilon}) < \infty, \ \forall \varepsilon > 0,$$

das heisst

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty, \ \forall \varepsilon > 0.$$

Genau das wurde bei unserem Beweis des starken Gesetzes der großen Zahlen benutzt:

$$\sum_{n=1}^{\infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} Y_i \right| > \varepsilon \right) < \infty, \ \forall \varepsilon > 0.$$

Kapitel 6

Zentraler Grenzwertsatz

Definition 6.1. Es seien X_n und X Zufallsgrößen mit Verteilungsfunktionen F_n bzw. F ($n \in \mathbb{N}$). Man sagt, die Folge (F_n) (bzw. die Folge (X_n)) konvergiert *schwach* (bzw. *in Verteilung*) gegen F (bzw. X), falls $\lim_{n\to\infty} F_n(x) = F(x)$ für alle Punkte $x \in \mathbb{R}$, in denen F stetig ist. (Konvergenz der Verteilungsfunktionen in allen Stetigkeitspunkten der Grenzverteilungsfunktion.)

Schreibweisen: $F_n \Rightarrow F$, $F_n \xrightarrow{w} F$, $X_n \Rightarrow X$, $X_n \xrightarrow{d} X, \ldots$

Proposition 6.2. *Aus der Konvergenz in Wahrscheinlichkeit folgt die Konvergenz in Verteilung.*

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$$

Beweis. Angenommen $X_n \xrightarrow{P} X$ und x sei ein Stetigkeitspunkt von F. Dann gilt für beliebige $\varepsilon > 0$:

$$\begin{aligned} F_n(x) &= P(X_n \le x) \\ &= P(X_n \le x, |X_n - X| \le \varepsilon) + P(X_n \le x, |X_n - X| > \varepsilon) \\ &\le P(X \le x + \varepsilon) + P(|X_n - X| > \varepsilon) \\ &= F(x+\varepsilon) + P(|X_n - X| > \varepsilon), \end{aligned}$$

und da $P(|X_n - X| > \varepsilon)$ für $n \to \infty$ gegen 0 konvergiert, erhalten wir

$$\limsup_{n\to\infty} F_n(x) \le F(x+\varepsilon).$$

Wegen der Rechtsstetigkeit von F folgt für $\varepsilon \downarrow 0$

$$\limsup_{n\to\infty} F_n(x) \le F(x). \tag{6.1}$$

Entsprechend erhält man durch Vertauschen von X_n und X und Übergang zu $x - \varepsilon$

$$F(x-\varepsilon) \le F_n(x) + P(|X - X_n| > \varepsilon)$$

und damit

$$\liminf_{n\to\infty} F_n(x) \geq F(x-\varepsilon).$$

Da F als stetig im Punkt x vorausgesetzt wurde, folgt für $\varepsilon \downarrow 0$

$$\liminf_{n\to\infty} F_n(x) \geq F(x). \tag{6.2}$$

Aus (6.1) und (6.2) folgt

$$\lim_{n\to\infty} F_n(x) = F(x).$$

□

Sei (X_n) eine Folge u.i.v. Zufallsgrößen mit endlicher Varianz,

$$\begin{aligned} m &:= EX_1, \qquad 0 < \sigma^2 := V(X_n) < \infty, \\ S_n &:= \sum_{i=1}^{n} X_i \qquad \text{Partialsummen.} \end{aligned}$$

Der Übergang zu *standardisierten* (das heißt *zentrierten* und *normierten*) Summen erfolgt durch

$$Z_n := \frac{S_n - nm}{\sqrt{n}\sigma},$$

somit ist

$$EZ_n = 0, V(Z_n) = 1 \qquad \text{(Bienaymé).}$$

Z_n beschreibt die geeignet normierten zufälligen Abweichungen der Summe S_n von ihrem Erwartungswert.

Wie sind diese „zufälligen Fluktuationen“ asymptotisch verteilt? Wir definieren die Verteilungsfunktion der standardisierten Normalverteilung durch

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\,\mathrm{d}y.$$

Satz 6.3. Zentraler Grenzwertsatz
Sei (X_n) eine Folge u.i.v. Zufallsgrößen, $m := EX_1$, $0 < \sigma^2 := V(X_1) < \infty$, $S_n := \sum_{i=1}^{n} X_i$. Dann gilt

$$\lim_{n\to\infty} P\left(\frac{S_n - nm}{\sqrt{n}\sigma} \leq x\right) = \Phi(x) \tag{6.3}$$

für alle $x \in \mathbb{R}$, das heißt $\frac{S_n-nm}{\sqrt{n}\sigma}$ konvergiert in Verteilung gegen die standardisierte Normalverteilung.

Aus (6.3) folgt

$$\begin{aligned}\lim_{n\to\infty} P\left(a \le \frac{S_n - nm}{\sqrt{n}\sigma} \le b\right) &= \Phi(b) - \Phi(a)\\ &= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,\mathrm{d}x\end{aligned}$$

für beliebige $a < b$.

Folgerung 6.4. *(Version eines Satzes von Moivre-Laplace)*
Sei (X_n) eine Folge unabhängiger Bernoulli-verteilter Zufallsgrößen mit Parameter p, $0 < p < 1$ und $S_n := \sum_{i=1}^n X_i$ binomialverteilt mit Parameter (n, p). Dann gilt

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,\mathrm{d}x$$

für beliebige $a < b$ mit $q := 1 - p$.

Beweis. Für die Bernoulli-Variable X_1 ist $m = EX_1 = p$, $\sigma^2 = V(X_1) = p\,q$. □

Universalität des Zentralen Grenzwertsatzes: Die Grenzverteilung (Normalverteilung) hängt nicht von der konkreten Gestalt der Verteilung von X_1 ab, sondern nur von zwei Parametern: Erwartungswert m und Varianz σ^2.

Beweis. Idee für den Zentralen Grenzwertsatz. Sei $\chi_n(t) := E\left[\exp\left\{it\frac{S_n - nm}{\sqrt{n}\sigma}\right\}\right]$ die charakteristische Funktion von $\frac{S_n - nm}{\sqrt{n}\sigma}$, man zeigt, dass $\chi(t) := e^{-t^2/2}$ die charakteristische Funktion der Standard-Normalverteilung ist.

Auf Grund des Konvergenzsatzes für charakteristische Funktionen genügt es zu zeigen, dass

$$\lim_{n\to\infty} \chi_n(t) = \chi(t) \qquad \text{für alle } t \in \mathbb{R}.$$

Da S_n Summe u.i.v. Zufallsgrößen $\frac{X_k - m}{\sqrt{n}\sigma}$ ist, erhalten wir

$$\chi_n(t) = \left(Ee^{it\frac{X_1 - m}{\sqrt{n}\sigma}}\right)^n,$$

wobei $\frac{X_1 - m}{\sqrt{n}\sigma}$ „klein“ ist für große n.
Eine Taylorentwicklung liefert

$$\begin{aligned}Ee^{it\frac{X_1 - m}{\sqrt{n}\sigma}} &\approx E\left[1 + it\frac{X_1 - m}{\sqrt{n}\sigma} - \frac{t^2}{2n}\left(\frac{X_1 - m}{\sigma}\right)^2\right] \qquad (6.4)\\ &= 1 - \frac{t^2}{2n}.\end{aligned}$$

Hierbei wurde benutzt, dass

$$E\left(\frac{X_1 - m}{\sqrt{n}\sigma}\right)^2 = \frac{1}{n\sigma^2}V(X_1) = \frac{1}{n}.$$

Damit erhält man (falls die Restglieder der Taylorapproximation tatsächlich vernachlässigbar sind)

$$\begin{aligned}\lim_{n\to\infty} \chi_n(t) &= \lim_{n\to\infty}\left(1 - \frac{t^2}{2n}\right)^n \\ &= e^{-t^2/2} \\ &= \chi(t).\end{aligned}$$

Durch eine geeignete (etwas diffizile) Restgliedabschätzung in (6.4) kann diese Idee zu einem vollständigen Beweis ausgebaut werden. Dies ist einfach, falls $P(|X_1| \leq C) = 1$ für eine gewisse Konstante $C < \infty$, aber komplizierter, falls nur $EX_1^2 < \infty$ vorausgesetzt wird. □

Genügt eine Folge (X_n) einem Zentralen Grenzwertsatz, so genügt sie auch dem schwachen Gesetz der großen Zahlen. Genauer, mit $S_n = \sum_{i=1}^n X_i$ und ξ standard-normalverteilt gilt:

$$\frac{S_n - nm}{\sqrt{n}\sigma} \xrightarrow{d} \xi \qquad \Rightarrow \qquad \frac{S_n}{n} \xrightarrow{P} m.$$

Beweis. Skizze

$$\begin{aligned}\frac{S_n - nm}{\sqrt{n}\sigma} \xrightarrow{d} \xi \quad &\Rightarrow \quad \frac{S_n - nm}{n} \xrightarrow{d} 0 \\ &\overset{(*)}{\Rightarrow} \quad \frac{S_n - nm}{n} \xrightarrow{P} 0 \\ &\Leftrightarrow \quad \frac{S_n}{n} \xrightarrow{P} m.\end{aligned}$$

Dabei wurde in Schritt $(*)$ benutzt, dass aus der Verteilungskonvergenz gegen eine Konstante die Konvergenz in Wahrscheinlichkeit gegen diese Konstante folgt. □

Kapitel 7

Normalverteilung (Gauß-Verteilung)

Normalverteilungen bilden zusammen mit den Poisson-Verteilungen die beiden zentralen Verteilungen in der Wahrscheinlichkeitstheorie. Dies beruht auf ihrer Universalität als Limesverteilungen im Zentralen Grenzwertsatz und Poissonschen Grenzwertsatz.

Zunächst sollen nochmal eindimensionale Normalverteilungen rekapituliert werden, bevor wir zu mehrdimensionalen Normalverteilungen übergehen.

Standard-Normalverteilung: Verteilung einer Zufallsgröße mit Dichte

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

Eigenschaften:

a) Die Dichte φ ist symmetrisch, besitzt genau ein Maximum bei 0 und genau zwei Wendepunkte bei ± 1.

b) Die Funktion φ ist tatsächlich eine Wahrscheinlichkeitsdichte, das heißt zu Eins normiert: Sei

$$J := \int_{-\infty}^{\infty} \varphi(x)\ \mathrm{d}x.$$

Daraus folgt bei Übergang zu Polarkoordinaten

$$\begin{aligned} J^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2+y^2}{2}\right\} \mathrm{d}x\, \mathrm{d}y \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r\, \mathrm{d}r\, \mathrm{d}\varphi = 1. \end{aligned}$$

Da $\int_0^\infty e^{-r^2/2} r \, \mathrm{d}r = 1$ ist $J = 1$.

c) Für die zugehörige Verteilungsfunktion Φ gilt:

$$\Phi(x) = \int_{-\infty}^{x} \varphi(y) \, \mathrm{d}y = \int_{-x}^{\infty} \varphi(y) \, \mathrm{d}y = 1 - \int_{-\infty}^{-x} \varphi(y) \, \mathrm{d}y,$$

das heißt, $\Phi(x) = 1 - \Phi(-x)$ bzw. $\Phi(-x) = 1 - \Phi(x)$.

d) Eine standard-normalverteilte Zufallsgröße X ist tatsächlich „standardisiert", das heißt $EX = 0$, $V(X) = EX^2 = 1$:

$$EX = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \, e^{-x^2/2} \, \mathrm{d}x = 0 \qquad \text{aus „Symmetriegründen".}$$

Mit Hilfe partieller Integration erhält man

$$\begin{aligned} EX^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} \, \mathrm{d}x = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left(\frac{\mathrm{d}}{\mathrm{d}x} e^{-x^2/2} \right) \mathrm{d}x \\ &= -\frac{1}{\sqrt{2\pi}} x \, e^{-x^2/2} \Big|_{x=-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \, \mathrm{d}x = 0 + 1. \end{aligned}$$

Man spricht deshalb auch von einer $\mathcal{N}(0,1)$-*verteilten* Zufallsgröße (siehe später).

Die charakteristische Funktion $\chi(t) = Ee^{itX}$ einer $\mathcal{N}(0,1)$-verteilten Zufallsgröße X hat die Gestalt

$$\chi(t) = e^{-t^2/2}, t \in \mathbb{R} \qquad \text{(ohne Beweis).}$$

Ausgehend von standard-normalverteilten Zufallsgrößen sollen nun allgemeinere normalverteilte Zufallsgrößen durch „lineare Transformation" eingeführt werden: Die Zufallsgröße X sei $\mathcal{N}(0,1)$-verteilt und $Y := \sigma X + \mu$ ($\mu \in \mathbb{R}$, $\sigma > 0$), dann gilt: $EY = \mu$, $V(Y) = \sigma^2$. Benutzt man die Variablensubstitution $u = \frac{v-\mu}{\sigma}$, so erhält man für Y folgende Verteilungsfunktion:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\left(X \leq \frac{y-\mu}{\sigma} \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y-\mu}{\sigma}} e^{-u^2/2} \, \mathrm{d}u \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{y} e^{-(v-\mu)^2/2\sigma^2} \, \mathrm{d}v. \end{aligned}$$

Also besitzt Y eine Dichte und diese hat die Gestalt

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(v-\mu)^2}{2\sigma^2} \right\}. \tag{7.1}$$

Eine Zufallsgröße mit dieser Dichte heißt *normalverteilt* mit Erwartungswert (Mittelwert) μ und Varianz $\sigma^2 > 0$, kurz $\mathcal{N}(\mu, \sigma^2)$*-verteilt*.

Die Klasse der (eindimensionalen) Normalverteilungen wird also durch eine zweiparametrige Schar von Dichten der Gestalt (7.1) beschrieben, wobei der erste Parameter der Erwartungswert und der zweite die Varianz ist.

Wir berechnen die charakteristische Funktion einer $\mathcal{N}(\mu, \sigma^2)$-Verteilung:

$$\begin{aligned} \chi_{\mu,\sigma^2}(t) &= Ee^{itY} = Ee^{it(\sigma X+\mu)} \\ &= e^{it\mu} Ee^{i(t\sigma)X} = e^{it\mu}\chi(\sigma t), \end{aligned}$$

das heißt

$$\chi_{\mu,\sigma^2}(t) = \exp\left\{ it\mu - \frac{\sigma^2}{2}t^2 \right\}, \qquad t \in \mathbb{R}.$$

Die Dichte hat ein Maximum bei μ und zwei Wendepunkte bei $\mu \pm \sigma$. Man überzeugt sich leicht davon, dass obige Konstruktion auch umgekehrt werden kann: Die Zufallsgröße X sei $\mathcal{N}(\mu, \sigma^2)$-verteilt, daraus folgt $X^* := \frac{X-\mu}{\sigma^2}$ ist $\mathcal{N}(0,1)$-verteilt (X^* entsteht aus X durch Standardisierung). Zum Beweis kann man entweder die Dichte oder die charakteristische Funktion von X^* ausrechnen, denn die charakteristische Funktion bestimmt ja ebenfalls die Verteilung einer Zufallsgröße eindeutig.

Behauptung 7.1. *Die Zufallsgröße X sei $\mathcal{N}(\mu, \sigma^2)$-verteilt und Y sei $\mathcal{N}(\nu, \rho^2)$-verteilt. Sind X und Y unabhängig, so folgt $X+Y$ ist $\mathcal{N}(\mu+\nu, \sigma^2+\rho^2)$-verteilt.*

Beweis. Dass $Z := X + Y$ Erwartungswert $\mu + \nu$ und Varianz $\sigma^2 + \rho^2$ hat, ist sofort ersichtlich. Es ist aber noch zu zeigen, dass Z normalverteilt ist. Am einfachsten geht das, wenn man mit Hilfe der Unabhängigkeit von X und Y zeigt, dass die charakteristische Funktion von Z die richtige Gestalt hat:

$$\begin{aligned} \chi_Z(t) &= \chi_{X+Y}(t) = \chi_X(t)\,\chi_Y(t) \\ &= \exp\left\{ it\mu - \frac{\sigma^2}{2}t^2 \right\} \exp\left\{ it\nu - \frac{\rho^2}{2}t^2 \right\} \\ &= \exp\left\{ it(\mu+\nu) - \frac{(\sigma^2+\rho^2)}{2}t^2 \right\}. \end{aligned}$$

□

Auf die gleiche Weise beweist man folgenden allgemeineren Satz:

Satz 7.2. *Die Zufallsgrößen X_k seien $\mathcal{N}(\mu_k, \sigma_k^2)$-verteilt (k=1,...,n) und unabhängig. Für beliebige „Gewichte“ $a_k \in \mathbb{R}$ $(k = 1, \ldots, n)$ sind die Linearkombinationen $Z := \sum_{k=1}^n a_k X_k$ dann $\mathcal{N}(\mu, \sigma^2)$-verteilt mit $\mu = \sum_{k=1}^n a_k \mu_k$ und $\sigma^2 = \sum_{k=1}^n a_k^2 \sigma_k^2$.*

(Linearkombinationen *unabhängiger* normalverteilter Zufallsgrößen sind stets normalverteilt.)

Mehrdimensionale Normalverteilung Wichtige Kenngrößen (eindimensionaler) Zufallsgrößen sind Erwartungswert und Varianz. Bei (n-dimensionalen) Zufallsvektoren $X = (X_1, \ldots, X_n)^{\mathrm{t}}$ interessiert man sich entsprechend für den *Erwartungswertvektor* (Mittelwertvektor)

$$EX := \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix},$$

dessen Komponenten die Erwartungswerte der einzelnen Zufallsgrößen $X_1, \ldots, X_n$ sind, sowie für die *Kovarianzmatrix*

$$\operatorname{cov}(X, X) := E[(X - EX)(X - EX)^{\mathrm{t}}].$$

Der Erwartungswert der Matrix $(X - EX)(X - EX)^{\mathrm{t}}$ ist dabei ebenfalls komponentenweise definiert, das heißt

$$\operatorname{cov}(X, X) = \{\operatorname{cov}(X_i, X_j)\}_{i,j=1}^{n},$$

wobei

$$\operatorname{cov}(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)] = E(X_i X_j) - EX_i EX_j$$

die früher eingeführte Kovarianz zwischen den Zufallsgrößen X_i und X_j bezeichnet.

Kovarianzmatrizen sind (offensichtlich) symmetrisch und positiv semidefinit. Letzteres sieht man wie folgt: Für alle Vektoren $\lambda = (\lambda_1, \ldots, \lambda_n)^{\mathrm{t}} \in \mathbb{R}^n$ gilt:

$$\begin{aligned} \lambda^{\mathrm{t}} \operatorname{cov}(X, X) \lambda &= \sum_{i,j=1}^{n} \lambda_i \operatorname{cov}(X_i, X_j) \lambda_j \\ &\overset{(*)}{=} E\Big(\sum_{i=1}^{n} \lambda_i (X_i - EX_i)\Big)^2 \geqq 0. \end{aligned}$$

Die Gültigkeit von $(*)$ sieht man, wenn man das Quadrat der Summe auf der rechten Seite als Doppelsumme schreibt und die Linearität des Erwartungswertes ausnutzt.

Eine wichtige Rolle spielt auch die *charakteristische Funktion* eines Zufallsvektors, die dessen Verteilung vollständig charakterisiert und auch sonst Eigenschaften besitzt, die zu denen charakteristischer Funktionen reeller Zufallsgrößen ähnlich sind (ohne Beweis).

Die charakteristische Funktion χ_X eines n-dimensionalen Zufallsvektors X wird wie folgt definiert:

$$\chi_X(\lambda) := E e^{i\langle \lambda, X \rangle}, \qquad \lambda = (\lambda_1, \ldots, \lambda_n)^t \in \mathbb{R}^n.$$

Dabei bezeichnet $\langle ., . \rangle$ das Standard-Skalarprodukt im $\mathbb{R}^n$.

Die n-dimensionale Standard-Normalverteilung: ist die Verteilung eines Zufallvektors $X = (X_1, \ldots, X_n)^{\mathrm{t}}$, dessen Komponenten $X_1, \ldots, X_n$ unabhängige standard-normalverteilte Zufallsgrößen sind. Insbesondere besitzt ein solcher Zufallsvektor X eine Dichte φ, nämlich das Tensorprodukt der Dichten von $X_1, \ldots, X_n$:

$$\varphi(x) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} \right) = (2\pi)^{-n/2} e^{-|x|^2/2}, \qquad x = (x_1, \ldots, x_n)^{\mathrm{t}} \in \mathbb{R}^n.$$

Offenbar ist $EX = 0$ (Nullvektor) und $\operatorname{cov}(X, X) = I$ (I ist die n-dimensionale Einheitsmatrix), da aus der Unabhängigkeit der Zufallsgrößen $X_1, \ldots, X_n$ deren paarweise Unkorreliertheit folgt. Wegen der Unabhängigkeit der Komponenten $X_1, \ldots, X_n$ ist die charakteristische Funktion χ von X das Tensorprodukt der charakteristischen Funktionen von $X_1, \ldots, X_n$ (Beweis?) und somit

$$\chi(\lambda) = e^{-\lambda^2/2}, \qquad \lambda \in \mathbb{R}^n.$$

Wir wollen nun allgemeine mehrdimensionale Normalverteilungen einführen, indem wir *lineare Transformationen* standard-normalverteilter Zufallsvektoren betrachten:

$$Y = AX + \mu. \tag{7.2}$$

Dabei bezeichne X einen m-dimensionalen standard-normalverteilten Zufallsvektor, A eine $(n \times m)$-Matrix und μ einen Vektor aus $\mathbb{R}^n$. Also ist Y ein n-dimensionaler Zufallsvektor. Unter Benutzung der Linearität des Erwartungswertes erhält man

$$EY = A\,EX + \mu = \mu,$$

da $EX = 0$ und

$$\begin{aligned} \operatorname{cov}(Y, Y) &= E[(Y - EY)(Y - EY)^{\mathrm{t}}] \\ &= E[AX(AX)^{\mathrm{t}}] \\ &= A\,E(XX^{\mathrm{t}})\,A^{\mathrm{t}} \\ &= AA^{\mathrm{t}} =: \Sigma^2, \end{aligned}$$

da $E(XX^{\mathrm{t}}) = \operatorname{cov}(X, X) = I$. Die Matrix AA^{t} ist als Kovarianzmatrix symmetrisch und positiv semidefinit und besitzt deshalb eine symmetrische und positive semidefinite „Wurzel“, die wir mit Σ bezeichnet haben.
die charakteristische Funktion χ_Y von Y hat die Gestalt:

$$\begin{aligned} \chi_Y(\lambda) &= Ee^{i\langle \lambda, Y \rangle} = Ee^{i\langle \lambda, AX+\mu \rangle} \\ &= e^{i\langle \lambda, \mu \rangle} Ee^{i\langle A^{\mathrm{t}}\lambda, X \rangle} \\ &= e^{i\langle \lambda, \mu \rangle} e^{-1/2(A^{\mathrm{t}}\lambda)^2} \\ &= \exp\{i\langle \lambda, \mu \rangle - 1/2\lambda^{\mathrm{t}} AA^{\mathrm{t}} \lambda\}, \end{aligned}$$

das heißt,

$$\chi_Y(\lambda) = \exp\{i\langle\lambda,\mu\rangle - 1/2\lambda^{\mathrm{t}}\Sigma^2\lambda\}, \qquad \lambda \in \mathbb{R}^n. \tag{7.3}$$

Da die charakteristische Funktion von Y deren Verteilung eindeutig bestimmt (ohne Beweis), sehen wir, dass die Verteilung von Y nur vom Erwartungswertvektor μ und der Kovarianzmatrix Σ^2 abhängt (und nicht explizit von A; AA^{t} kann für verschiedene Matrizen A gleich sein).

Eine Verteilung im $\mathbb{R}^n$ mit charakteristischer Funktion (7.3) heißt $\mathcal{N}(\mu,\Sigma^2)$-*Verteilung* oder auch *Normalverteilung* mit Erwartungswertvektor μ und Kovarianzmatrix Σ^2. Y ist somit $\mathcal{N}(\mu,\Sigma^2)$-verteilt. Wann besitzt Y eine Dichte und wie sieht diese aus? Der durch (7.2) definierte Zufallsvektor nimmt nur Werte im affinen Unterraum $\Lambda := \{Ax + \mu : x \in \mathbb{R}^m\}$ des $\mathbb{R}^n$ an. Seine Dimension ist gleich dem Rang der Matrix A, die wiederum mit dem Rang von $\Sigma^2 = AA^{\mathrm{t}}$ übereinstimmt (siehe Lineare Algebra). Ist die Dimension dieses Unterraums kleiner als n, das heißt $\det\Sigma^2 = 0$, so kann Y keine Dichte besitzen (das Integral einer solchen Funktion wäre nicht 1, sondern 0, da sie auf einer Menge Λ konzentriert wäre, deren n-dimensionales Lebesgue-Maß null ist).

Wir zeigen, dass bei nichtentarteter Kovarianzmatrix Σ^2 (das heißt $\det\Sigma^2 \neq 0$) der Zufallsvektor Y eine Dichte besitzt, die wir explizit berechnen. Hierzu benutzen wir folgenden, auch in anderen Situationen nützlichen, Transformationssatz:
Sei X ein n-dimensionaler Zufallsvektor mit Dichte $f : \mathbb{R}^n \to \mathbb{R}^n$, $Y := \Phi(X)$, mit $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ bijektiv, so dass Φ, Φ^{-1} (Umkehrabbildung) beide stetig differenzierbar sind.

Die Funktion $\Phi^{-1} = (\Phi_1^{-1},\ldots,\Phi_n^{-1})^{\mathrm{t}}$ besitzt die Jordanmatrix $J(y) = \{\partial\Phi_i^{-1}(y)/\partial y_j\}_{i,j=1}^n$ und die zugehörige Funktionaldeterminante $\det J(y)$. Dann gilt für eine beliebige Borelmenge $B \in \mathfrak{B}^n$ (und speziell für $B = (-\infty,b_1]\times\cdots\times(-\infty,b_n])$:

$$\begin{aligned} P(Y\in B) &= P(X\in\Phi^{-1}(B)) \\ &= \int_{\Phi^{-1}(B)} f(x)\,\mathrm{d}x \\ &= \int_B f(\Phi^{-1}(y))|\det J(y)|\,\mathrm{d}y \end{aligned}$$

nach dem Transformationssatz für Integrale, das heißt Y besitzt die Dichte $f(\Phi^{-1}(y))|\det J(y)|$.

Um dies auf den Gauß-Vektor (7.2) anzuwenden, können wir o.B.d.A. annehmen, dass Y durch $Y = \Sigma\, X + \mu$ gegeben ist mit einem n-dimensionalen standardnormalverteilten Vektor X. Man beachte, dass dieser Zufallsvektor und der durch (7.2) gegebene Zufallsvektor die gleiche Kovarianzmatrix Σ^2 besitzen, damit auch die gleiche charakteristische Funktion und die gleiche Verteilung.
In unserem Fall ist

$$\Phi(x) = \Sigma x + \mu, \qquad \Phi^{-1}(y) = \Sigma^{-1}(y-\mu)$$

$\det \Sigma^2 = (\det \Sigma)^2 \neq 0$, $J(y) = \Sigma^{-1}$, $|\det J(y)| = |\det \Sigma^{-1}| = 1/\sqrt{\det \Sigma^2}$. Unter Benutzung der Dichte von X erhalten wir deshalb für die Dichte φ_{μ,Σ^2} von Y:

$$\varphi_{\mu,\Sigma^2}(y) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma^2}} \exp\{-\frac{1}{2}(y-\mu)^{\mathrm{t}}(\Sigma^2)^{-1}(y-\mu)\}, \qquad y \in \mathbb{R}^n,$$

da $|\Sigma^{-1}y|^2 = y^{\mathrm{t}}(\Sigma^{-1})^{\mathrm{t}}\Sigma^{-1}y = y^{\mathrm{t}}(\Sigma^{-1})^2 y = y^{\mathrm{t}}(\Sigma^2)^{-1}y$.

Ausgehend von unserer Definition normalverteilter Vektoren sieht man leicht, dass *lineare Transformationen normalverteilter Vektoren* wieder normalverteilte Vektoren liefern: Die Zufallsgröße X sei $\mathcal{N}(\mu, \Sigma^2)$-verteilt; $\mu \in \mathbb{R}^m, \Sigma^2$ eine positive semidefinite $(m \times m)$-Matrix und $Y = BX + \nu$ mit $\nu \in \mathbb{R}^n$ und B eine $(n \times m)$-Matrix.

Da uns nur die Verteilung von Y interessiert, können wir o.B.d.A. annehmen, dass X die Gestalt $X = \Sigma Z + \mu$ mit einer m-dimensionalen $\mathcal{N}(0, I)$-verteilten Zufallsgrößen Z hat. Durch Einsetzen erhält man $Y = (B\Sigma)Z + (B\mu + \nu)$. Also ist Y normalverteilt mit Erwartungswertvektor $B\mu + \nu$ und Kovarianzmatrix $B\Sigma(B\Sigma)^{\mathrm{t}} = B\Sigma^2 B^{\mathrm{t}}$: Y ist $\mathcal{N}(B\mu + \nu, B\Sigma^2 B^{\mathrm{t}})$-verteilt.

Spezialfälle:

1. Sei $\nu = 0, B = (b_1, \ldots, b_m)$ ein Zeilenvektor, daraus folgt $Y = \sum_{k=1}^m b_k X_k$, das heißt die Linearkombination der Komponenten $X_1, \ldots, X_m$ eines normalverteilten Vektors ist normalverteilt. Hierzu reicht es nicht aus vorauszusetzen, dass die Zufallsgrößen $X_1, \ldots, X_m$ normalverteilt sind. Diese Größen müssen als gemeinsame Verteilung eine Normalverteilung besitzen (was insbesondere gilt, wenn $X_1, \ldots, X_m$ unabhängig sind)!

2. Sei

$$B = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & \vdots & 0 \end{pmatrix}$$

eine $(r \times m)$-Matrix mit $1 \leq r \leq m$, $\nu = 0$.
Dann ist

$$Y = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix},$$

das heißt, Teilvektoren normalverteilter Vektoren sind wieder normalverteilt.

Kapitel 8

Markov-Ketten

Sei I eine endliche oder abzählbare Menge und $(X_n)_{n\in\mathbb{N}_0}$ eine Folge I-wertiger Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, P)$.

Definition 8.1.

a) Die Folge (X_n) heißt *Markov-Kette* mit Zustandsraum I, falls für alle $n \in \mathbb{N}_0$ und alle $i_0, i_1, \ldots, i_{n+1} \in I$ mit $P(X_0 = i_0, \ldots, X_n = i_n) > 0$

$$\begin{aligned} P(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i_n) \\ = P(X_{n+1} = i_{n+1} | X_n = i_n) \end{aligned}$$

(*Markov-Eigenschaft*) gilt.

b) Die Markov-Kette (X_n) heißt *homogen* , falls die Wahrscheinlichkeiten

$$P(X_{n+1} = j | X_n = i), \qquad i, j \in I,$$

nicht von n abhängen (für die $i \in I$ und $n \in \mathbb{N}_0$, für die $P(X_n = i) > 0$).

Interpretation: Die Markov-Kette (X_n) wird als zufällige Bewegung auf I betrachtet, wobei X_n der zufällige Zustand zum Zeitpunkt n ist. Die Zufallsgrößen X_n sind im Allgemeinen nicht unabhängig.
Die Verteilung des Zustandes X_{n+1} in der „Zukunft“ hängt bei bekannter „Gegenwart“ $X_n = i_n$ nicht von den in der „Vergangenheit“ durchlaufenen Zuständen $i_0, \ldots, i_{n-1}$ ab.

Homogenität: Die stochastischen Regeln für den Übergang vom Zustand i zum Zustand j ändern sich nicht mit der Zeit. Im Folgenden nehmen wir an, dass alle Markov-Ketten homogen sind, ohne dass die Homogenität jedesmal explizit erwähnt wird.

Definition 8.2. Sei (X_n) eine (homogene) Markov-Kette mit Zustandsraum I. Dann heißen $p_{ij} := P(X_{n+1} = j | X_n = i)$, $i, j \in I$, *Übergangswahrscheinlichkeiten.* Die Matrix $P = (p_{ij})_{i,j\in I}$ heißt *Übergangsmatrix* der Markov-Kette.

Für eine homogene Markov-Kette gilt somit:

$$P(X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) = p_{i_n i_{n+1}}.$$

Behauptung 8.3.

$$p_{ij} \geq 0 \text{ für alle } i, j \in I \text{ und} \sum_{j \in I} p_{ij} = 1 \text{ für alle } i. \tag{8.1}$$

Bemerkung 8.4.

(i) Die Matrix $P = (p_{ij})_{i,j \in I}$ kann abzählbar unendlich viele Zeilen und Spalten besitzen.

(ii) Ist $P(X_n = i) = 0$ für alle n, so setzt man für ein solches i die p_{ij} beliebig, aber so, dass $p_{ij} \geq 0$ und $\sum_{j \in I} p_{ij} = 1$.

(iii) Eine Matrix P mit den Eigenschaften (8.1) heißt *stochastische Matrix.*

Definition 8.5. Sei (X_n) eine Markov-Kette. Die Verteilung von X_0 heißt *Anfangsverteilung* der Markov-Kette. Die gemeinsamen Verteilungen von $X_0, X_1, \ldots, X_n$ $(n \in N_0)$ heißen *endlichdimensionale Verteilungen.*

Die Anfangsverteilung ist durch ihre Einzelwahrscheinlichkeiten

$$p_i^0 := P(X_0 = i), \qquad i \in I,$$

gegeben. Die endlichdimensionalen Verteilungen enthalten alle stochastisch relevanten Informationen über die Markov-Kette.

Satz 8.6. *Sei (X_n) eine Markov-Kette mit Anfangsverteilung (p_i^0) und Übergangswahrscheinlichkeiten (p_{ij}). Dann gilt*

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = p_{i_0}^0 p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$$

$(n \in \mathbb{N}_0;\, i_0, \ldots, i_n \in I)$.

Beweis. Aus der Markov-Eigenschaft und Homogenität folgt

$$\begin{aligned}
& P(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i_n) \\
= \;& P(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) P(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
= \;& \cdots = p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \cdots p_{i_0 i_1} P(X_0 = i_0) \\
= \;& p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \cdots p_{i_0 i_1} p_{i_0}^0.
\end{aligned}$$

□

Eine Markov-Kette ist stochastisch, das heißt im Sinne der endlichdimensionalen Verteilungen, durch Ausgabe von Anfangsverteilungen und Übergangsmatrix vollständig charakterisiert.

Beispiel 8.7.

(i) Sei (X_n) eine Folge u.i.v. I-wertiger Zufallsgrößen (I höchstens abzählbar) mit Einzelwahrscheinlichkeiten $p_i := P(X_0 = i)$, $i \in I$. Aus der Unabhängigkeit folgt

$$\begin{aligned} P(X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) &= P(X_{n+1} = i_{n+1}) = p_{i_{n+1}} \\ P(X_{n+1} = i_{n+1} | X_n = i_n) &= p_{i_{n+1}}. \end{aligned}$$

Somit ist (X_n) eine Markov-Kette mit Anfangsverteilung $p_i^0 = p_i$ und Übergangswahrscheinlichkeit $p_{ij} = p_j$.

(ii) *Einfache Irrfahrt auf dem eindimensionalen Gitter*

Sei (ξ_n) eine Folge u.i.v. $\{-1, 1\}$-wertiger Zufallsgrößen mit
$P(\xi_n = 1) =: p$ und $P(\xi_n = -1) =: q$ wobei $p + q = 1$. Nun sei
$X_n := \sum_{i=1}^n \xi_i$ mit $X_0 = 0$.

$$\begin{aligned} & P(X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) \\ = \; & P(\xi_{n+1} = i_{n+1} - i_n | X_n = i_n, \ldots, X_0 = i_0) \\ = \; & P(\xi_{n+1} = i_{n+1} - i_n), \end{aligned}$$

da $X_n = i_n, \ldots, X_0 = i_0$ nur von $\xi_1, \ldots, \xi_n$ abhängt. Analog zeigt man $P(X_{n+1} = i_{n+1} | X_n - i_n) = P(\xi_{n+1} = i_{n+1} - i_n)$.
Damit ist (X_n) eine Markov-Kette mit Zustandsraum $\mathbb{Z}$ und Anfangsverteilung $p_i^0 = \delta_{0i}$, $(i \in \mathbb{Z})$ und

Übergangswahrscheinlichkeit $p_{ij} = \begin{cases} p, & \text{j=i+1;} \\ q, & \text{j=i-1;} \\ 0, & \text{sonst.} \end{cases}$

(iii) *Einfache symmetrische Irrfahrt auf* $\mathbb{Z}^d$:

Sei $X_n := \sum_{k=1}^n \xi_k$, wobei (ξ_n) unabhängig und gleichverteilt auf
$\{i \in \mathbb{Z}^d : |i| = 1\}$. Sei (X_n) eine Markov-Kette mit Zustandsraum $\mathbb{Z}^d$ und Anfangsverteilung $p_i^0 = \delta_{0i}$, $(i \in \mathbb{Z}^d)$ mit Übergangswahrscheinlichkeiten

$$p_{ij} = \begin{cases} \frac{1}{2d}, & \text{i-j=1;} \\ 0, & \text{sonst.} \end{cases}$$

Somit beschreibt die Markov-Kette gleichwahrscheinliche Sprünge in Nachbarpunkte.

(iv) *Irrfahrt mit Absorption*

Wir betrachten den Zustandsraum $I = \{0, 1, \ldots, N\}$ mit den Wahrscheinlichkeiten $p, q \geq 0$, $p + q = 1$
$p_{i,i+1} = p, \qquad p_{i,i-1} = q \qquad$ für $i = 1, \ldots, N-1$

und Absorption an den Rändern $p_{00} = 1$, $p_{NN} = 1$.
Die Übergangsmatrix hat folgende Gestalt

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & q & 0 & p \\ 0 & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Deutung: 2 Spieler mit Anfangskapital m, $0 \leq m \leq N$, bzw. $N - m$ (das heißt $p_i^0 = \delta_{mi}$).
Bei jedem Spiel gewinnt der 1. Spieler eine Einheit mit Wahrscheinlichleit p und verliert eine Einheit mit Wahrscheinlichkeit q. Das Spiel wird abgebrochen, sobald einer der Spieler bankrott ist (Absorption).
Das Kapital des 1. Spielers zum Zeitpunkt k, das heißt nach k Spielen, ist X_k.

(v) *Irrfahrt mit Reflexion*

Wie im 4. Beispiel mit dem Zustandsraum $I = \{0, 1, \ldots, N\}$ und den Wahrscheinlichkeiten

$p_{i,i-1} = q$, $p_{i,i+1} = p$ für $i = 1, \ldots, N-1$.

Aber diesmal mit folgenden Reflexionsbedingungen: $p_{0,1} = 1$, $p_{N,N-1} = 1$

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & q & 0 & p \\ 0 & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}$$

Sei nun $P = (p_{ij})_{i,j \in I}$ eine stochastische Matrix. Stochastische Matrizen lassen sich problemlos potenzieren: Für $n \in \mathbb{N}_0$ definiert man $P^n = (p_{ij}^n)_{i,j \in I}$ rekursiv wie folgt: Sei $p_{ij}^{(0)} := \delta_{ij}$ die Einheitsmatrix und

$$p_{ij}^{n+1} := \sum_{k \in I} p_{ik}^{(n)} p_{kj} \qquad (i, j \in I). \tag{8.2}$$

Für jedes n ist P^n eine stochastische Matrix: Das ist klar für $n = 0$ und $n = 1$.

Induktionsschluss $n \curvearrowright n+1$:

$$\begin{aligned}\sum_{j\in I} p_{ij}^{(n+1)} &= \sum_{k\in I} p_{ik}^{(n)} \underbrace{\sum_{j\in I} p_{kj}}_{=1} \\ &= \sum_{k\in I} p_{ik}^{(n)} \\ &= 1\end{aligned}$$

Insbesondere konvergiert die Reihe auf der rechten Seite von (8.2) für alle $i \in I$. Es gelten die aus der linearen Algebra bekannten Rechenregeln für Matrizen. Insbesondere gilt

$$P^{m+n} = P^m P^n,$$

das heißt

$$p_{ij}^{(m+n)} = \sum_{k\in I} p_{ik}^{(m)} p_{kj}^{(n)}, \qquad (i,j \in I).$$

Diese Gleichung heißt *Chapman-Kolmogorov-Gleichung.*

Behauptung 8.8. *Sei (X_n) eine Markov-Kette mit Übergangsmatrix $P = (p_{ij})$. Dann gilt für beliebige $m, n \geq 0$ und alle $i, j \in I$ mit $P(X_m = i) > 0$:*

$$P(X_{m+n} = j | X_m = i) = p_{ij}^{(n)}.$$

Deshalb heißt $P^n = (p_{ij}^{(n)})$ *Übergangsmatrix in n Schritten* und $p_{ij}^{(n)}$ ist die *Übergangswahrscheinlichkeiten in n Schritten.*

Beweis. Die Behauptung gilt für $n = 0$ und $n = 1$ nach Definition.
Sei $n \geq 2$.

$$P(X_{m+n} = j | X_m = i) = \frac{P(X_m = i, X_{m+n} = j)}{P(X_m = i)} \tag{8.3}$$

Verwendet man die Formel für endlichdimensionale Verteilungen, so folgt

$$\begin{aligned}P(X_m = i, X_{m+n} = j) &= \sum_{i_0,\dots,i_{m-1}} \sum_{j_1,\dots,j_{n-1}} P(X_0 = i_0, \dots, X_{m-1} = i_{m-1}, \\ &\qquad X_m = i, X_{m+1} = j_1, \dots, X_{m+n-1} = j_{n-1}, X_{m+n} = j) \\ &= \sum_{i_0,\dots,i_{m-1}} \sum_{j_1,\dots,j_{n-1}} p_{i_0}^0 p_{i_0 i_1} \cdots p_{i_{m-1} i} p_{i j_1} \cdots p_{j_{n-1} j} \\ &= \sum_{i_0} p_{i_0}^0 \sum_{i_1,\dots,i_{m-1}} p_{i_0 i_1} \cdots p_{i_{m-1} i} \sum_{j_1,\dots,j_{n-1}} p_{i j_1} \cdots p_{j_{n-1} j} \\ &= \sum_{i_0} p_{i_0}^0 p_{i_0 i}^{(m)} p_{ij}^{(n)}\end{aligned}$$

Der letzte Schritt folgt aus der Definition von $p_{ij}^{(n)}$ und Matrizenprodukten.

Wir haben folgende Gleichung erhalten:

$$P(X_m = i, X_{m+n=j}) = \left(\sum_{i_0} p_{i_0}^0 p_{i_0 i}^{(m)}\right) p_{ij}^{(n)} \tag{8.4}$$

Die Summation beider Seiten bezüglich j liefert

$$P(X_m = i) = \sum_{i_0} p_{i_0}^0 p_{i_0 i}^{(m)}. \tag{8.5}$$

Setzt man (8.4) und (8.5) in (8.3) ein, so erhält man die Behauptung

$$P(X_{m+n} = j | X_m = i) = p_{ij}^{(n)}.$$

□

Die Gleichung (8.5) liefert die folgende Aussage.

Behauptung 8.9. *Sei (X_n) eine Markov-Kette mit Anfangsverteilung (p_i^0) und Übergangswahrscheinlichkeiten (p_{ij}). Dann gilt für einen beliebigen Zeitpunkt $n \geq 0$:*

$$P(X_n = j) = \sum_{i \in I} p_i^0 p_{ij}^{(n)} \qquad (j \in I). \tag{8.6}$$

Startet die Markov-Kette in einem festen Zustand $i \in I$, das heißt gilt

$$p_k^0 = \delta_{ik} = \begin{cases} 1, & k{=}i; \\ 0, & k \neq i; \end{cases}$$

so gilt $P(X_n = j) = p_{ij}^{(n)}$.

Die Formel (8.6) beschreibt die Verteilung der Zufallsgrößen X_n, deren Einzelwahrscheinlichkeiten wir mit $p_i^{(n)}$ bezeichnen wollen:

$$p_i^{(n)} := P(X_n = i), \qquad i \in I, n \in \mathbb{N}_0.$$

Dies steht im Einklang mit der vorher eingeführten Bezeichnung p_i^0 für die Einzelwahrscheinlichkeiten der Anfangsverteilungen.

Bezeichnen wir mit p^0 den *Zeilen*vektor $(p_i^0)_{i\in I}$ der Anfangsverteilung und mit $p^{(n)}$ den *Zeilen*vektor $(p_i^{(n)})_{i\in I}$, so läßt sich (8.6) in „Matrixform“ schreiben:

$$p^{(n)} = p^0 P^n, \qquad n \in \mathbb{N}_0.$$

Bei gegebenem Anfangsvektor p^0 und gegebener Übergangsmatrix P lassen sich somit die Verteilungen der Zufallsgrößen X_n einfach berechnen.

Im weiteren wollen wir die Frage untersuchen, ob sich die Verteilung einer Markovkette X_n asymptotisch (für die Zeit $n \to \infty$) einer „Gleichgewichtsverteilung“ nähert. Hierzu wollen wir zunächst alle möglichen „Gleichgewichte“ untersuchen, die wir stationäre oder invariante Verteilungen nennen wollen.

Definition 8.10. Eine Markov-Kette (X_n) heißt *stationär*, falls für beliebige $m \in \mathbb{N}$ die Ketten $(X_n)_{n\in\mathbb{N}_0}$ und $(X_{m+n})_{n\in\mathbb{N}_0}$ die gleichen endlichdimensionalen Verteilungen besitzen, das heißt falls für beliebige $m \in \mathbb{N}$ und $r \in \mathbb{N}_0$ die Vektoren $(X_0, \dots, X_r)$ und $(X_m, X_{m+1}, \dots, X_{m+r})$ die gleiche Verteilung haben.

Für eine stationäre Markov-Kette (X_n) hängen insbesondere die eindimendionalen Verteilungen nicht von n ab, das heißt $X_0, X_1, X_2, \dots$ sind identisch verteilt. Kern der folgenden Aussage ist, dass auch die Umkehrung gilt.

Behauptung 8.11. *Sei (X_n) eine Markov-Kette mit Anfangsverteilung (p_i^0) und Übergangswahrscheinlichkeiten (p_{ij}). Dann gilt:*

$$(X_n)\,\text{stationär} \Leftrightarrow \sum_{i\in I} p_i^0 p_{ij} = p_j^0 \textit{ für alle } j \in I.$$

Beweis. „$\Rightarrow$“: Sei (X_n) stationär. Dann besitzen insbesondere X_0 und X_1 die gleiche Verteilung, das heißt

$$\underbrace{P(X_1 = j)}_{\sum_{i\in I} p_i^0 p_{ij}} = \underbrace{P(X_0 = j)}_{p_j^0}, \qquad \forall\, j \in I.$$

„$\Leftarrow$“: Angenommen es gilt

$$\sum_{i\in I} p_i^0 p_{ij} = p_j^0, \qquad \forall\, j \in I. \tag{8.7}$$

Es ist zu zeigen, dass für alle $m \in \mathbb{N}$, $r \in \mathbb{N}_0$ und $i_0, i_1, \dots, i_r \in I$

$$P(X_0 = i_0, \dots, X_r = i_r) = P(X_m = i_0, \dots, X_{m+r} = i_r)$$

gilt. Wir drücken beide Wahrscheinlichkeiten mit Hilfe der Formel für die endlichdimensionalen Verteilungen aus:

$$\begin{aligned}
P(X_0 = i_0, \dots, X_r = i_r) &= p_{i_0}^0 p_{i_0 i_1} \cdots p_{i_{r-1} i_r} \\
P(X_m = i_0, \dots, X_{m+r} = i_r) &= \sum_{j_0,\dots,j_{m-1}\in I} P(X_0 = j_0, \dots, X_{m-1} = j_{m-1}, \\
&\qquad X_m = i_0, \dots, X_{m+r} = i_r) \\
&= p_{i_0 i_1} \cdots p_{i_{r-1} i_r} \sum_{j_{m-1}\in I} \left(\cdots \left(\sum_{j_1\in I} \left(\sum_{j_0\in I} p_{j_0}^0 p_{j_0 j_1} \right) p_{j_1 j_2} \right) \cdots \right) \\
\sum_{j_0\in I} p_{j_0}^0 p_{j_0 j_1} \quad &\underset{(8.7)}{=} \quad p_{j_1}^0, \sum_{j_1\in I} p_{j_1}^0 p_{j_1 j_2} \underset{(8.7)}{=} p_{j_2}^0 \text{ usw.} \\
&= p_{i_0 i_1} \cdots p_{i_{r-1} i_r} \underbrace{\sum_{j_{m-2}\in I} p_{j_{m-1}}^0 p_{j_{m-1} i_0}}_{=_{(8.7)} p_{i_0}^0} \\
&= p_{i_0}^0 p_{i_0 i_1} \cdots p_{i_{r-1} i_r}
\end{aligned}$$

Also stimmen beide Wahrscheinlichkeiten überein. $\square$

Die letzte Behauptung führt uns auf die folgende Definition.

Definition 8.12. Gegeben sei eine Markov-Kette mit Zustandsraum I und Übergangsmatrix $P = (p_{ij})_{i,j\in I}$. Eine Wahrscheinlichkeitsverteilung $p = (p_i)_{i\in I}$ heißt *stationäre Anfangsverteilung* (oder auch *invariante Verteilung*) der Markov Kette, falls

$$\sum_{i\in I} p_i p_{ij} = p_j, \qquad \forall\, j \in I,$$

gilt.

Betrachtet man $p = (p_i)_{i\in I}$ als Zeilenvektor, so sind die stationären Anfangsverteilungen diejenigen Lösungen p der „Matrixgleichung“

$$pP = p,$$

für die *zusätzlich* $p_i \geq 0$ und $\sum_{k\in I} p_k = 1$ gilt.

Beispiel 8.13.
1) Irrfahrt auf $I = \{0, 1, \ldots, N\}$ mit Absorption in den Randpunkten mit $0 < p < 1$, $p + q = 1$.
Das Gleichungssystem für die invarianten Verteilungen hat die Gestalt

$$\begin{aligned}
p_0 + qp_1 &= p_0 \\
qp_2 &= p_1 \\
pp_{i-1} + qp_{i+1} &= p_i \qquad i = 2, \ldots, N-2 \\
pp_{N-2} &= p_{N-1} \\
pp_{N-1} + p_N &= p_N
\end{aligned}$$

Hieraus ist sofort ersichtlich, dass

$$p_1 = p_2 = \cdots = p_{N-1} = 0$$

ist und p_0, p_N „beliebig“ sein können. Das heißt wir erhalten eine einparametrige Schar stationärer Anfangsverteilungen $p = (p_i)$:

$$p_0 = \gamma,\ p_1 = \cdots = p_{N-1} = 0,\ p_N = 1 - \gamma \qquad (0 \leq \gamma \leq 1).$$

Dieses Ergebnis wird sofort plausibel, wenn man berücksichtigt, dass 0 und N absorbierende Zustände sind. Startet die Kette in 0 ($\gamma = 0$) oder in N ($\gamma = 1$), so bleibt sie für immer in diesem Zustand.

2) Irrfahrt auf $\mathbb{Z}_+ = \{0, 1, \ldots\}$ mit Reflexion in 0 und den Wahrscheinlichkeiten $0 < p < 1$, $p + q = 1$.
Das Gleichungssystem für die invariante Verteilung $p = (p_i)_{i\in\mathbb{Z}_+}$ hat die Gestalt

$$\begin{aligned}
qp_1 &= p_0 \\
p_0 + qp_2 &= p_1 \\
pp_{i-1} + qp_{i+1} &= p_i \qquad \text{für } i \geq 2
\end{aligned}$$

Die Lösung der homogenen Differenzengleichung

$$pp_{i-1} - p_i + qp_{i+1} = 0$$

bestimmt man wie folgt. Der Ansatz $p_i = \lambda^i$ führt auf die charakteristische Gleichung

$$p - \lambda + q\lambda^2 = 0$$

mit den beiden Lösungen

$$\lambda_1 = 1, \qquad \lambda_2 = p/q.$$

Damit ist analog zu den Formeln für lineare homogene Differentialgleichungen 2.Ordnung

$$\begin{aligned} p_i &= C + Di \qquad \text{für } p = q = \frac{1}{2} \\ p_i &= C + D\left(\frac{p}{q}\right)^i \qquad \text{für } p \neq q \end{aligned}$$

wobei C und D beliebige Konstanten sind.
Für $p \geq q$ existiert daher *keine* stationäre Anfangsverteilung, da entweder $C = D = 0$ (daraus folgt $p_i = 0$ für $i \geq 1$ daraus folgt $p_0 = 0$)
oder $\sum p_i = \infty$.
Sei nun $p < q$. Damit $p_i \geq 0$ und $\sum_i p_i = 1$, muss $C = 0$ und $D \geq 0$ sein.
Sei also

$$p_i = D\left(\frac{p}{q}\right)^i, \qquad i \geq 1.$$

Setzen wir dies in die ersten beiden Gleichungen ein, so ergibt sich

$$p_0 = Dp \Leftrightarrow \begin{cases} qD\frac{p}{q} = p_0 \\ p_0 + qD(\frac{p}{q})^2 = D\frac{p}{q} \end{cases}$$

Die Variable D ist so zu wählen, dass

$$\begin{aligned} 1 &= \sum_{i=0}^{\infty} p_i \\ &= Dp + D\sum_{i=1}^{\infty}(\frac{p}{q})^i \\ &= D\left[p + \frac{p}{q}\frac{1}{1 - p/q}\right] \\ &= D\frac{2p\,q}{q-p} \\ \Leftrightarrow D &= \frac{q-p}{2p\,q} \end{aligned}$$

Damit existiert für $p < q$ *genau eine* stationäre Anfangsverteilung, nämlich

$$\begin{cases} p_0 = \frac{q-p}{2q} \\ p_i = \frac{q-p}{2p\,q}\left(\frac{p}{q}\right)^i & \text{für } i \geq 1. \end{cases}$$

Wir haben gesehen, dass im Allgemeinen die stationäre Anfangsverteilung weder zu existieren braucht, noch eindeutig sein muss. Für Markovketten mit endlichem Zustandsraum ist jedoch auf Grund eines Kompaktheitsarguments zumindest die Existenz gesichert.

Satz 8.14. *Jede Markovkette mit* endlichem *Zustandsraum besitzt mindestens eine stationäre Anfangsverteilung.*

Beweis. Seien $I = \{1, \ldots, N\}$ der endliche Zustandsraum, $P = (p_{ij})$ die Übergangsmatrix der Markovkette und

$$S_N := \left\{ (p_1, \ldots, p_N) \in \mathbb{R}^N : \qquad p_j \geq 0, \sum_{j=1}^{N} p_j = 1 \right\}$$

das Simplex der Verteilungen (Wahrscheinlichkeitsvektoren) auf I. Weiterhin sei $p^0 \in S_N$ eine beliebige Anfangsverteilung. Wir setzen

$$p^{(n)} := p^0 P^n \qquad (n \in \mathbb{N}_0).$$

Dann sind die Zeilenvektoren

$$\pi^{(n)} := \frac{1}{n} \sum_{i=0}^{n-1} p^{(i)} \qquad (n \in \mathbb{N})$$

ebenfalls Wahrscheinlichkeitsvektoren, das heißt Elemente von S_N.

Da S_N kompakt ist, finden wir eine Teilfolge $\left(\pi^{(n_k)}\right)$, die gegen einen Wahrscheinlichkeitsvektor $\pi \in S_N$ konvergiert.
Nun ist

$$\begin{aligned} \pi^{(n_k)} P &= \frac{1}{n_k} \sum_{i=0}^{n_k - 1} \underbrace{p^{(i)} P}_{=p^{(i+1)}} \\ &= \frac{1}{n_k} \sum_{i=1}^{n_k} p^{(i)}, \end{aligned}$$

das heißt

$$\pi^{(n_k)} P = \pi^{(n_k)} + \frac{1}{n_k} \left(p^{(n_k)} - p^0 \right).$$

Durch Grenzübergang für $k \to \infty$ erhält man

$$\pi P = \pi,$$

das heißt π ist eine stationäre Anfangsverteilung. $\square$

Wir formulieren nun eines der Hauptergebnisse zur Asymptotik endlicher Markovketten.

Theorem 8.15. (Konvergenzsatz für endliche Markovketten)
Sei (X_n) eine Markovkette mit endlichem *Zustandsraum I und der Übergangsmatrix $P = (p_{ij})$. Angenommen, es existiert eine natürliche Zahl r, so dass P^r mindestens eine Spalte besitzt, deren Elemente* alle *positiv sind. Dann gelten folgende Aussagen:*

a) Für beliebige $i, j \in I$ existieren die von i unabhängigen Grenzwerte

$$\lim_{n\to\infty} p_{ij}^{(n)} =: p_j.$$

b) Die einzige stationäre Anfangsverteilung der Markovkette ist $p = (p_j)_{j\in I}$.

Beweis. a) Sei

$$m_j^{(n)} := \min_i p_{ij}^{(n)} \text{ und } M_j^{(n)} := \max_i p_{ij}^{(n)}.$$

Mit der Chapman-Kolmogorov-Gleichung erhält man

$$\begin{aligned} m_j^{(n)} &= \sum_k p_{ik} m_j^{(n)} \\ &\leq p_{ij}^{(n+1)} \\ &= \sum_k p_{ik} p_{kj}^{(n)} \\ &\leq \sum_k p_{ik} M_j^{(n)} \\ &= M_j^{(n)} \end{aligned}$$

und damit

$$m_j^{(n)} \leq m_j^{(n+1)} \leq M_j^{(n+1)} \leq M_j^{(n)}.$$

Deshalb existieren die (endlichen) Grenzwerte von $m_j^{(n)}$ und $M_j^{(n)}$ für $n \to \infty$, und es gilt

$$\lim_{n\to\infty} m_j^{(n)} \leq \lim_{n\to\infty} M_j^{(n)}, \qquad j \in I.$$

Wegen

$$m_j^{(n)} \leq p_{ij}^{(n)} \leq M_j^{(n)}, \qquad i, j \in I,$$

folgt die Behauptung a), wenn wir zeigen, dass $M_j^{(n)} - m_j^{(n)} \to 0$ für $n \to \infty$ und alle $j \in I$.

Nach Voraussetzung existieren $r \in \mathbb{N}$, $k_0 \in I$ und $\delta > 0$ derart, dass

$$p_{ik_0}^{(r)} \geq \delta \qquad \text{für alle } i \in I. \qquad \text{(I endlich)} \tag{8.8}$$

Die Chapman-Kolmogorov-Gleichung liefert zunächst

$$p_{ij}^{(n+r)} = \sum_k p_{ik}^{(r)} p_{kj}^{(n)},$$

und wir erhalten

$$\begin{aligned} M_j^{(r+n)} - m_j^{(r+n)} &= \max_i \sum_k p_{ik}^{(r)} p_{kj}^{(n)} - \min_m \sum_k p_{mk}^{(r)} p_{kj}^{(n)} \\ &= \max_{i,m} \sum_k \left(p_{ik}^{(r)} - p_{mk}^{(r)} \right) p_{kj}^{(n)}. \end{aligned} \tag{8.9}$$

Wir wollen die letzte Summe in zwei Teile aufspalten, indem wir jeweils nur über die Indices k summieren, für die $p_{ik}^{(r)} - p_{mk}^{(r)}$ positiv bzw. negativ ist.

Hierzu führen wir die Indexmengen

$$I_{i,m,r}^+ := \{k \in I : \qquad p_{ik}^{(r)} - p_{mk}^{(r)} \geq 0\}$$

und

$$I_{i,m,r}^- := \{k \in I : \qquad p_{ik}^{(r)} - p_{mk}^{(r)} < 0\}$$

sowie die Summen

$$S_{i,m,r}^+ := \sum_{k \in I_{i,m,r}^+} \left(p_{ik}^{(r)} - p_{mk}^{(r)} \right)$$

und

$$S_{i,m,r}^- := \sum_{k \in I_{i,m,r}^-} \left(p_{ik}^{(r)} - p_{mk}^{(r)} \right)$$

ein. Offenbar ist

$$S_{i,m,r}^+ - S_{i,m,r}^- = \underbrace{\sum_k p_{ik}^{(r)}}_{=1} - \underbrace{\sum_k p_{mk}^{(r)}}_{=1} = 0$$

Damit folgt aus (8.9)

$$\begin{aligned} M_j^{(n+r)} - m_j^{(n+r)} &\leq \max_{i,m} \left[S_{i,m,r}^+ M_j^{(n)} + S_{i,m,r}^- m_j^{(n)} \right] \\ &= \left(\max_{im} S_{i,m,r}^+ \right) \left(M_j^{(n)} - m_j^{(n)} \right) \end{aligned} \tag{8.10}$$

Durch Iteration dieser Ungleichung erhalten wir

$$M_j^{(lr+s)} - m_j^{(lr+s)} \leq \left(\max_{i,m} S^+_{i,m,r} \right)^l \left(M_j^{(s)} - m_j^{(s)} \right),$$

$l \geq 0$, $0 \leq s < r$.
Hieraus folgt $M_j^{(n)} - m_j^{(n)} \to 0$ für $n \to \infty$, falls

$$\max_{i,m} S^+_{i,m,r} < 1$$

ist. Um dies zu beweisen, benutzen wir die Voraussetzung (8.8).
Für $k_0 \in I^+_{i,m,r}$ gilt

$$S^+_{i,m,r} \leq \sum_{k \in I} p_{ik}^{(r)} - p_{mk_0}^{(r)} \leq 1 - \delta,$$

und für $k_o \notin I^+_{i,m,r}$ ist

$$\begin{aligned} S^+_{i,m,r} &\leq \sum_{k \neq k_0} p_{ik}^{(r)} \\ &= 1 - p_{ik_0}^{(r)} \\ &\leq 1 - \delta. \end{aligned}$$

Also gilt

$$\max_{i,m} S^+_{i,m,r} \leq 1 - \delta,$$

und wir sind fertig.

Im Hinblick auf den Beweis der Behauptung b) merken wir noch an, dass wegen $\sum_{j \in I} p_{ij}^{(n)} = 1$ für alle $i \in I$ und der Endlichkeit der Indexmenge I auch für die Grenzwerte $p_j = \lim_{n \to \infty} p_{ij}^{(n)}$ die Summe $\sum_{j \in I} p_j$ gleich Eins ist.

b) Die stationäre Anfangsverteilung der Markovkette ist $(p_j)_{j \in I}$. Tatsächlich, aus der Chapman-Kolmogorov-Gleichung

$$p_{ij}^{(n+1)} = \sum_k p_{ik}^{(n)} p_{kj}, \qquad i, j \in I,$$

folgt mit a) durch Grenzübergang für $n \to \infty$

$$p_j = \sum_k p_k p_{kj}, \qquad j \in I.$$

Sei nun $(\pi_i)_{i \in I}$ eine beliebige stationäre Anfangsverteilung. Dann ist

$$\pi_j = \sum_i \pi_i p_{ij}^{(n)}, \qquad j \in I,$$

das heißt $\pi = \pi P$ daraus folgt $\pi = \pi P^n$ für beliebige $n \geq 1$.
Für $n \to \infty$ erhalten wir

$$\pi_j = \sum_i \pi_i p_j = p_j, \qquad j \in I.$$

Dies liefert die Eindeutigkeit. □

Folgerung 8.16. *Die Voraussetzungen des Theorems seien erfüllt. Dann gilt:*
a) Konvergenz der eindimensionalen Verteilungen gegen die stationäre Anfangsverteilung:

$$\lim_{n\to\infty} P(X_n = j) = p_j, \qquad j \in I.$$

b) Konvergenz der endlich dimensionalen Verteilungen gegen die entsprechenden Verteilungen zugehörigen stationären Markovketten:
Für beliebige $r \in \mathbb{N}_0$ und $i_0, \ldots, i_r \in I$ ist

$$\lim_{n\to\infty} P(X_n = i_0, X_{n+1} = i_1, \ldots, X_{n+r} = i_r) = p_{i_0} p_{i_0 i_1} \ldots p_{i_{r-1} i_r}.$$

Beweis.
a) Sei (p_i^0) die Anfangsverteilung der Markovkette. Dann ist

$$P(X_n = j) = \sum_i p_i^0 p_{ij}^{(n)} \underset{n\to\infty}{\to} \sum_i p_i^0 p_j = p_j.$$

b) Analog erhält man leicht

$$P(X_n = i_0, X_{n+1} = i_1, \ldots, X_{n+r} = i_r) = \sum_i p_{i_0}^0 p_{ii_0}^{(n)} p_{i_0 i_1} \ldots p_{i_{r-1} i_r}.$$

Für $n \to \infty$ folgt hieraus $p_{ii_0}^{(n)} \to p_{i_0}$ und $\sum_i p_i^0 = 1$ die Behauptung. □

Im Konverganzsatz für Markovketten bleibt zunächst offen, was die Voraussetzungen der Positivität einer Spalte von P^r bedeutet und wie diese Bedingung überprüft werden kann. Eine Antwort hierauf werden wir am Ende des folgenden Teilabschnitts erhalten.

8.1. Klassifizierung der Zustände

Im weiteren wollen wir eine geeignete *Klassifizierung der Zustände* Markovscher Ketten einführen. Diese Klassifizierung soll hier nur soweit entwickelt werden, wie sie für Markovketten mit *endlichem* Zustandsraum erforderlich ist.

Unser Hauptziel dabei ist die Beantwortung der Frage nach asymptotischen Verhalten der Übergangswahrscheinlichkeiten $p_{ij}^{(n)}$ für $n \to \infty$ für *beliebige* endliche Markovketten. Die Klassifizierung selbst ist nicht an den endlichen Zustandsraum gebunden. Sie ist nur für den abzählbar unendlichen Zustandsraum nicht ausreichend, um p_{ij} für $n \to \infty$ zu studieren.

Im folgenden sei (X_n) eine Markovkette mit Zustandsraum I und Übergangsmatrix $P = (p_{ij})_{i,j \in I}$. Die Übergangswahrscheinlichkeiten in n Schritten bezeichnen wir mit $p_{ij}^{(n)}$.

Definition 8.17. Seien $i, j \in I$ beliebige Zustände.

a) Man sagt, der Zustand j ist vom Zustand i aus *erreichbar*, falls ein $n \in \mathbb{N}_0$ existiert, für das $p_{ij}^{(n)} > 0$ ist.
Schreibweise: $i \rightsquigarrow j$

b) Man sagt, die Zustände i und j sind *äquivalent*, falls $i \rightsquigarrow j$ und $j \rightsquigarrow i$ gilt.
Schreibweise: $i \leftrightsquigarrow j$

Behauptung 8.18. *Die Relation „$\leftrightsquigarrow$" ist eine Äquivalenzrelation auf dem Zustandsraum I.*

Beweis.

(i) Reflexivität: $i \leftrightsquigarrow i$, da nach Vereinbarung $p_{ii}^{(0)} = 1 > 0$;

(ii) Symmetrie: $i \leftrightsquigarrow j$ daraus folgt $j \leftrightsquigarrow i$ nach Definition;

(iii) Transitivität: $i \leftrightsquigarrow j$, $j \leftrightsquigarrow k$ daraus folgt $i \leftrightsquigarrow k$
Hierzu genügt es zu zeigen, dass aus $i \rightsquigarrow j$ und $j \rightsquigarrow k$ die Relation $i \rightsquigarrow k$ folgt.
$i \rightsquigarrow j$ bedeutet, dass ein $m \in \mathbb{N}_0$ existiert mit $p_{ij}^{(m)} > 0$
$j \rightsquigarrow k$ bedeutet, dass ein $n \in \mathbb{N}_0$ existiert mit $p_{jk}^{(n)} > 0$.
Unter Benutzung der Chapman-Kolmogorov-Gleichung erhalten wir

$$\begin{aligned} p_{ik}^{(m+n)} &= \sum_{l \in I} p_{il}^{(m)} p_{lk}^{(n)} \\ &\geq p_{ij}^{(m)} p_{jk}^{(n)} > 0, \end{aligned}$$

das heißt $i \rightsquigarrow k$.

□

Die Äquivalenzrelation „$\leftrightsquigarrow$" induziert eine Zerlegung des Zustandsraumes I in *Äquivalenzklassen* (*Klassen äquivalenter Zustände*).

Definition 8.19. Eine Markovkette heißt *irreduzibel*, falls sie nur aus einer Äquivalenzklasse besteht.

Definition 8.20. Ein Zustand $i \in I$ heißt *wesentlich*, falls aus $i \rightsquigarrow j$ stets $j \rightsquigarrow i$ folgt. Andernfalls heißt i *unwesentlich.*

„Wesentlich" und „unwesentlich" sind Klasseneigenschaften:

Behauptung 8.21. *Sei $i \in I$ beliebig gewählt. Dann gilt:*
a) i wesentlich $\Rightarrow$ j wesentlich für alle $j \leftrightsquigarrow i$
b) i unwesentlich $\Rightarrow$ j unwesentlich für alle $j \leftrightsquigarrow i$.

Beweis.

a) Sei i wesentlich und $j \leftrightsquigarrow i$. Angenommen $j \rightsquigarrow k$. Es ist zu zeigen, dass dann auch $k \rightsquigarrow j$ gilt.

$$\begin{aligned} i \rightsquigarrow j, j \rightsquigarrow k \quad &\underset{\text{„Transitivität“}}{\Rightarrow} \quad i \rightsquigarrow k \underset{i \text{ wesentlich}}{\Rightarrow} k \rightsquigarrow i \\ k \rightsquigarrow i, i \rightsquigarrow j \quad &\underset{\text{„Transitivität“}}{\Rightarrow} \quad k \rightsquigarrow j. \end{aligned}$$

b) folgt aus der Negation aus a). □

Man kann also von *wesentlichen* und *unwesentlichen Klassen* sprechen. Eine wesentliche Klasse kann nicht verlassen werden, eine unwesentliche kann man verlassen.

Beispiel 8.22. a) Die symmetrische Irrfahrt auf $\mathbb{Z}^d$ ist irreduzibel:
Für beliebige $i, j \in \mathbb{Z}^d$ ist $i \rightsquigarrow j$ zu zeigen.
Hierzu wählen wir einen beliebigen Pfad

$$i = i_0 \to i_1 \to \cdots \to i_n = j \qquad \text{mit } |i_k - i_{k-1}| = 1 \text{ für } k = 1, \ldots, n.$$

Auf Grund der Chapman-Kolmogorov-Gleichung erhält man

$$p_{ij}^{(n)} \geq p_{ii_1} p_{i_1 i_2} \ldots p_{i_{n-1} j} > 0,$$

das heißt $i \rightsquigarrow j$.

b) Irrfahrt mit Absorption am Rand und Zustandsraum $I = \{0, 1, \ldots, N-1, N\}$, wobei 0 und N absorbierend sind.
Es gibt 3 Äquivalenzklassen

(i) $K_1 = \{0\}$ wesentlich

(ii) $K_2 = \{1, 2, \ldots, N-1\}$ unwesentlich ($1 \rightsquigarrow 0$, aber $0 \not\rightsquigarrow 1$)

(iii) $K_3 = \{N\}$ wesentlich

Die Einschränkung einer Übergangsmatrix $P = (p_{ij})_{i,j \in I}$ auf eine *wesentliche* Klasse $W \subseteq I$ ist wieder eine stochstische Matrix:
Für $P^W := (p_{ij})_{i,j \in W}$ gilt

$$\sum_{j \in W} p_{ij} = 1, \qquad \forall i \in W.$$

(Wäre $p_{ij} > 0$ für ein $i \in W$ und ein $j \notin W$, so müsste $i \rightsquigarrow j$, aber $j \not\rightsquigarrow i$ gelten, das heißt i wäre unwesentlich. Also für $i \in W$ ist $1 = \sum_{j \in I} p_{ij} = \sum_{j \in W} p_{ij}$.)
Damit kann man die Einschränkung (X_n^W) der Markov-Kette (X_n) auf den Zustandsraum W betrachten. Die Kette ist irreduzibel.
Sei I endlich dann kann man I in

(i) wesentliche Klassen $K_1, \ldots, K_w$ und

(ii) unwesentliche Klassen $K_{w+1}, \ldots, K_{w+n}$

zerlegen. Bei entsprechender Numerierung der Zustände hat die Übergangsmatrix die folgrnde Blockgestalt

$$P = \begin{pmatrix} P_{11} & & 0 & & | & & & \\ & P_{22} & & & | & & & \\ 0 & & \ddots & & | & & 0 & \\ & & & P_{ww} & | & & & \\ - & - & - & - & - & - & - & - \\ & & & & | & P_{w+1,w+1} & & * \\ & * & & & | & & \ddots & \\ & & & & | & * & & P_{w+u,w+u} \end{pmatrix}$$

Dabei sind $P_{ll} = (p_{ij})_{i,j \in K_l}$ mit $l = 1, \ldots, w$, die zu den wesentlichen Klassen K_l gehörenden stochastischen Teilmatrizen. Die Matrizen P_{ll} mit $l = w+1, \ldots, w+n$, zu den unwesentlichen Klassen sind nicht stochastisch, weshalb die „$*$“-Bereiche von Null verschiedene Elemente enthalten können. An dieser Stelle ahnt man bereits, weshalb eine solche Klassifizierung der Zustände so nützlich ist. Sei nämlich W eine wesentliche Klasse. Um für $i \in W$ das Verhalten von $p_{ij}^{(n)}$ für $n \to \infty$ zu studieren, hat man zwei Fälle zu betrachten. Ist $j \in I \setminus W$, so gilt $p_{ij}^{(n)} = 0$ für alle n. Ist dagegen $j \in I$, so reduziert sich die Frage auf die Betrachtung der irreduziblen Kette mit Zustandsraum W und Übergangsmatrix $P^W = (p_{ij})_{i,j \in W}$.

Allerdings reicht die Irreduzibilität der Kette noch nicht aus, um unseren Konvergenzsatz für Markov-Ketten anwenden zu können:

Die Kette mit Zustandsraum $I = \{1, 2\}$ und Übergangswahrscheinlichkeiten $p_{12} = p_{21} = 1$ (und $p_{11} = p_{22} = 0$) ist irreduzibel. Jedoch ist

$$P^n = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ für } n \text{ ungerade}$$

und

$$P^n = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ für } n \text{ gerade.}$$

Es sind also weder die Voraussetzungen noch die Aussage des Konvergenzsatzes erfüllt. Dies hat offenbar etwas mit der „Periodizität“ (Periode 2) dieser (deterministischen) Kette zu tun. Wir wollen deshalb „periodisches“ Verhalten Markovscher Ketten etwas eingehender untersuchen.

Definition 8.23. a) die Zahl

$$d_i := g.g.T.\{n \in \mathbb{N} : p_{ii}^{(n)} > 0\} \qquad \text{(„größter gemeinsamer Teiler“)}$$

heißt *Periode* des Zustandes $i \in I$.

b) Ein Zustand $i \in I$ heißt *aperiodisch*, falls $d_i = 1$.

Die Periode ist eine Klasseneigenschaft:

Behauptung 8.24. *Seien $i, j \in I$ beliebige Zustände. Dann gilt*

$$i \leftrightsquigarrow j \text{ daraus folgt } d_i = d_j.$$

Beweis. Sei $i \leftrightsquigarrow j$. Dann existieren $k, l \in \mathbb{N}_0$ mit

$$p_{ij}^{(k)} > 0 \text{ und } p_{ji}^{(l)} > 0.$$

$$\Lambda_i := \{m \in \mathbb{N} : p_{ii}^{(m)} > 0\},\ \Lambda_j := \{n \in \mathbb{N} : p_{jj}^{(n)} > 0\}$$

Wir erhalten

$$p_{ii}^{(n+d_i(k+l))} \geq p_{ij}^{(k)} p_{jj}^{(n)} p_{ji}^{(l)} \underbrace{p_{ij}^{(k)} p_{ji}^{(l)} \dots p_{ij}^{(k)} p_{ji}^{(l)}}_{(d_i-1)-mal} > 0 \text{ für } n \in \Lambda_j,$$

das heißt

$$\begin{aligned} n \in \Lambda_j \quad &\Rightarrow \quad n + d_i(k+l) \in \Lambda_i \\ &\Rightarrow \quad d_i | n \qquad (\text{„}d_i \text{ Teiler von n“}) \\ &\Rightarrow \quad d_i | d_j \end{aligned}$$

Analog folgt $d_j | d_i$, das heißt $d_i = d_j$. $\square$

Beispiel 8.25.

a) Symmetrische Irrfahrt auf $\mathbb{Z}^d$ mit $p_{ii}^{(n)} = 0$ für n ungerade. Die Irrfahrt ist irreduzibel $p_{ii}^{(2)} > 0$. Daraus folgt Irrfahrt besitzt die Periode 2.

b) Irrfahrt mit Absorption am Rand $I = \{0, 1, \dots, N-1, N\}$.
Es gibt 3 Klassen

(i) $W_1 = \{0\}$, wesentlich, aperiodisch

(ii) $W_2 = \{1, 2, \dots, N-1\}$, unwesentlich, Periode 2

(iii) $W_3 = \{N\}$, wesentlich, aperiodisch

c) Irrfahrt mit *periodischer Randbedingung*
Sei $I = \{1, \dots, N\}$. Wir betrachten Sprünge mit positiver Wahrscheinlichkeit in Nachbarpunkte, wobei N und 1 benachbart sind. Die Irrfahrt ist

$$\text{irreduzibel und} \begin{cases} \text{aperiodisch,} & \text{für N ungerade;} \\ \text{Periode 2,} & \text{für N gerade.} \end{cases}$$

Da

$$\begin{aligned} p_{ii}^{(2)} > 0, p_{ii}^{(N)} > 0 \quad &\Rightarrow \quad d = 1 \text{ für N ungerade} \\ p_{ii}^{(2)} > 0, p_{ii}^{(n)} = 0 \text{ für n ungerade} \quad &\Rightarrow \quad d = 2 \text{ für N gerade.} \end{aligned}$$

Lemma 8.26. *Sei Λ eine beliebige nichtleere Teilmenge von $\mathbb{N}$ und $d := g.g.T.(\Lambda)$. Dann existieren $m \in \mathbb{N}$ und $\lambda_1, \ldots, \lambda_m \in \Lambda$ sowie $n_0 \in \mathbb{N}$, so dass für jedes $n \geq n_0$ Zahlen $k_1, \ldots, k_m \in \mathbb{N}$ existieren mit*

$$nd = \sum_{j=1}^{m} k_j \lambda_j.$$

Beweis.

$$G := \left\{ \sum_{j=1}^{m} r_j \lambda_j : m \in \mathbb{N}; r_1, \ldots, r_m \in \mathbb{Z}; \lambda_1, \ldots, \lambda_m \in \Lambda \right\}$$

ist die kleinste additive Untergruppe von $\mathbb{Z}$, die Λ enthält. Das kleinste positive Element von G sei d'. Wir wollen zeigen: $d' = d$.
Das Elemnt d teilt alle Elemente von Λ, daraus folgt d teilt alle Elemente von G, daraus folgt $d|d' (\Rightarrow d \leq d')$.

Andererseits: Jedes $n \in G$ besitzt eine Darstellung $n = rd' + s$ mit $0 \leq s < d'$

$\Rightarrow \quad s = n - rd' \in G$
$\Rightarrow \quad s = 0$
$\Rightarrow \quad d'$ teilt alle Elemente von G und damit auch von Λ
$\Rightarrow \quad d'|d$
$\Rightarrow \quad d' \leq d.$

Also ist $d' = d$ und insbesondere $d \in G$, das heißt es existiereren $\lambda_1, \ldots, \lambda_m \in \Lambda$ und $r_1, \ldots, r_m \in \mathbb{Z} \setminus \{0\}$ mit

$$d = \sum_{j=1}^{m} r_j \lambda_j. \tag{8.11}$$

Einige der Zahlen $r_1, \ldots, r_m$ könnten jedoch negativ sein. Nach Voraussetzung ist

$$\lambda_j = b_j d \text{ mit } b_j \in \mathbb{N} \text{ für } j = 1, \ldots, m, \tag{8.12}$$

wir setzen

$$\begin{aligned} b &:= \min\{b_1, \ldots, b_m\} \\ n_0 &:= \sum_{j=1}^{m} b|r_j|b_j. \end{aligned}$$

Jedes $n \geq n_0$ besitzt damit eine Darstellung in der Form

$$n = s_1 b_1 + \cdots + s_m b_m + s \tag{8.13}$$

mit $0 \leq s < b$ und $s_j \geq b|r_j|$, $\forall j$.
Damit folgt

$$\begin{aligned} nd &\underset{(8.13)}{=} \sum_{j=1}^{m} s_j b_j d + sd \\ &\underset{(8.12) und (8.11)}{=} \sum_{j=1}^{m} s_j \lambda_j + s \sum r_j \lambda_j \\ &= \sum_{j=1}^{m} (s_j + sr_j)\lambda_j \end{aligned}$$

Da nach Konstruktion $s_j + sr_j \geq 1$. $\square$

Folgerung 8.27. *Sei* (X_n) *eine Markov-Kette mit* endlichem *Zustandsraum* I *und Übergangsmatrix* P. *Ist die Kette irreduzibel und aperiodisch, so existiert ein* $r \in \mathbb{N}$, *so dass* alle *Elemente der Matrix* P^r *positiv sind. Insbesondere sind die Voraussetzungen unseres Konverganzsatzes erfüllt.*

Beweis. Sei $j \in I$ beliebig fixiert. Da I endlich ist, genügt es zu zeigen, dass ein $n_j \in \mathbb{N}$ existiert, so dass

$$p_{ij}^{(n)} > 0 \text{ für alle } i \in I \text{ und alle } n \geq n_j \tag{8.14}$$

gilt.

Sei $\Lambda_j := \{n \in \mathbb{N} : p_{jj}^{(n)} > 0\}$. Wegen der Aperiodizität ist g.g.T.$(\Lambda_j) = 1$ und das vorangegangene Lemma besagt, dass Zahlen $n_0, m \in \mathbb{N}$ und $\lambda_1, \ldots, \lambda_m \in \Lambda_j$ existieren mit folgender Eigenschaft: alle natürlichen Zahlen $n \geq n_0$ besitzen eine Darstellung der Gestalt

$$n = \sum_{\alpha=1}^{m} k_\alpha \lambda_\alpha \qquad \text{für gewisse } k_1, \ldots, k_m \in \mathbb{N}. \tag{8.15}$$

(Die Zahlen $m, n_0, \lambda_1, \ldots, \lambda_m$ hängen von j ab, was wir aber in der Notation nicht explizit zum Ausdruck gebracht haben.) Da die Kette irreduzibel ist, also alle Zustände äquivalent sind, existieren zu beliebigem $i \in I$ natürliche Zahlen n_{ij} mit

$$p_{ij}^{(n_{ij})} > 0 \qquad (\forall\, i \in I). \tag{8.16}$$

Wir zeigen, dass die Aussage (8.14) für $n_j := n_0 + \max_{i \in I} n_{ij}$ gilt. Hierzu fixieren wir $i \in I$ beliebig. Wegen (8.15) besitzt jedes $n \geq n_j$ eine Darstellung der Gestalt

$$n = n_{ij} + \sum_{\alpha=1}^{m} k_\alpha \lambda_\alpha \text{ mit } k_1, \ldots, k_m \in \mathbb{N}$$

($n - n_{ij} \geq n_0$; $k_1, \ldots, k_m$ hängen dabei selbstverständlich noch von i ab). Unter Benutzung von „Chapman-Kolmogorov“ erhält man deshalb für solche n:

$$p_{ij}^{(n)} \geq p_{ij}^{(n_{ij})} \prod_{\alpha=1}^{m} \left(p_{jj}^{(\lambda_\alpha)}\right)^{k_\alpha} > 0.$$

Da nach Voraussetzung die $p_{ij}^{(n_{ij})} > 0$ und auch die $p_{jj}^{(\lambda_\alpha)} > 0\,(\alpha = 1, \ldots, m)$ sind, da $\lambda_1, \ldots, \lambda_m \in \Lambda_j$. $\square$

Stichwortverzeichnis