Wahrscheinlichkeitstheorie I

Vorlesungsskript

Peter Mörters

Universität zu Köln
Mathematisches Institut

Sommersemester 2018

Inhaltsverzeichnis

Kapitel 1

Maß und Integrationstheorie

1.1 Existenz und Eindeutigkeit von Maßen

1.1.1 Modellierung von Zufallsexperimenten

Ein Zufallsexperiment wird durch einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und beliebig viele Zufallsvariable $X\colon \Omega \to \Omega'$ modelliert. Dabei ist

- Ω eine Menge, die *Ergebnismenge* genannt wird. Sie ist die Menge aller möglichen Ausgänge des zugrundeliegenden Experiments. Die Elemente von Ω heißen *Ausgänge, Ergebnisse, Stichproben, Realisierungen* oder *Elementarereignisse*.

- $\mathcal{A}$ ein Mengensystem auf Ω, das heißt eine Menge von Teilmengen von Ω. Die Elemente von $\mathcal{A}$ heißen *messbare Mengen*. Das Mengensystem $\mathcal{A}$ heißt auch *Ereignissystem*. Es besteht aus allen Ereignissen, über deren Auftreten ein Beobachter nach Ablauf des Experiments entscheiden kann, und erfüllt daher die Axiome einer σ-Algebra. Anders als in der *Einführung in die Stochastik* werden wir in dieser Vorlesung mitunter $\mathcal{A}$ variieren um unterschiedliche Kenntnisstände verschiedener Beobachter (oder eines Beobachters zu verschiedenen Zeiten) zu modellieren.

  Ein Mengensystem $\mathcal{A}$ auf Ω heißt *Algebra*, wenn gilt

  1. $\Omega \in \mathcal{A}$;
  2. ist $A \in \mathcal{A}$, so ist auch das Komplement $A^c = \Omega \setminus A \in \mathcal{A}$;
  3. sind $A_1, A_2 \in \mathcal{A}$, so ist auch die Vereinigung $A_1 \cup A_2 \in \mathcal{A}$.

  Es heißt *σ-Algebra* wenn zusätzlich gilt

  4. sind $A_1, A_2, \ldots \in \mathcal{A}$, so ist auch die Vereinigung $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

- Ein Paar $(\Omega, \mathcal{A})$ aus eine Menge Ω und einer σ–Algebra $\mathcal{A}$ auf Ω heißt *messbarer Raum* oder *Stichprobenraum*.

- Eine Zufallsvariable ist eine Abbildung $X\colon \Omega \to \Omega'$ zwischen Ergebnismengen. Typischerweise trägt Ω die Struktur eines messbaren Raumes $(\Omega, \mathcal{A})$ und dies induziert eine σ-Algebra $\mathcal{A}_X$ auf Ω' durch

$$\mathcal{A}_X = \{A \subseteq \Omega'\colon X^{-1}(A) \in \mathcal{A}\}.$$

  Wenn Ω' auch die Struktur eines messbaren Raums $(\Omega', \mathcal{A}')$ trägt, so heißt X *messbar*, wenn $\mathcal{A}' \subseteq \mathcal{A}_X$. In diesem Fall fordern wir von einer Zufallsvariablen stets, dass sie messbar sein soll. Dies ist vor allem dann der Fall, wenn die Bildmenge Ω' gleich $\mathbb{R}$ oder $\mathbb{R}^d$ ist und die messbare Struktur auf Ω' damit durch die Borel σ-Algebra $\mathcal{A}'$ gegeben ist.

- Sei $(\Omega, \mathcal{A})$ ein messbarer Raum. Den messbaren Mengen $A \in \mathcal{A}$ ordnet man nun Wahrscheinlichkeiten $P(A)$ zu, so dass die Axiome eines Wahrscheinlichkeitsmaßes erfüllt sind.

  Eine Abbildung $P : \mathcal{A} \to [0, \infty]$ heißt *Maß* auf $(\Omega, \mathcal{A})$, wenn gilt:

  1. $P(\emptyset) = 0$.
  2. Ist eine Folge $A_1, A_2, A_3, \ldots$ in $\mathcal{A}$ *paarweise unvereinbar*, gilt also $A_i \cap A_j = \emptyset$ für $i \neq j$, so gilt

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i)\,.$$

     Diese Eigenschaft heißt *σ–Additivität von P*.

  Gilt zusätzlich $P(\Omega) = 1$ so heißt P *Wahrscheinlichkeitsverteilung* oder ein *Wahrscheinlichkeitsmaß*. Dann heißt das Tripel $(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum*.

Wichtige Eigenschaften eines Maßes P sind

- **Monotonie:** Sind $A_1, A_2 \in \mathcal{A}$ mit $A_1 \subseteq A_2$ so gilt $P(A_1) \leq P(A_2)$.

- **Subadditivität:** Sind $A_1, A_2, \ldots \in \mathcal{A}$ so gilt

$$P\Big(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n).$$

- **Stetigkeit:** Sind $A_1, A_2, \ldots \in \mathcal{A}$ mit $A_n \uparrow A$, das heißt $A_n \subset A_{n+1}$ und $\bigcup_n A_n = A$, so gilt

$$\lim_{n\to\infty} P(A_n) = P(A).$$

1.1.2 Ein Beispiel: Wurf einer zufälligen Münze

Beschreibung des Experiments: Eine auf $(0,1)$ uniform verteilte Zufallsvariable θ wird generiert und anschließend eine Münze geprägt, die mit Wahrscheinlichkeit θ Kopf und mit Wahrscheinlichkeit $1-\theta$ Zahl erzeugt. Die Münze wird anschliessend n mal geworfen.

Aufgabe: Modellieren Sie das Experiment und Zufallsvariablen, die den Ausgang der n Würfe anzeigen. Wählen Sie verschiedene σ-Algebren und erklären Sie ihre Bedeutung.

Antwort: Für $n \in \mathbb{N}$ wählen wir

$$\Omega := (0,1) \times \{\text{Kopf}, \text{Zahl}\}^n$$

und bezeichnen die Elemente $\omega \in \Omega$ als $\omega = (x, a_1, \ldots, a_n)$. Wir definieren

$$X_i \colon \Omega \to \{\text{Kopf}, \text{Zahl}\}, X_i(x, a_1, \ldots, a_n) = a_i,$$

die Zufallsvariable, die den Ausgang des $i$ten Münzwurf angibt. Außerdem sei

$$Y \colon \Omega \to (0,1), X_i(x, a_1, \ldots, a_n) = x,$$

die Zufallsvariable, die die Prägung der Münze festlegt.

Wählt man als σ-Algebra

$$\mathcal{A} = \mathcal{B}((0,1)) \otimes \mathcal{P}\big(\{\text{Kopf}, \text{Zahl}\}^n\big),$$

also die kleinste σ-Algebra, die alle Mengen $A \times B$ mit $A \subset (0,1)$ offen und $B \subset \{\text{Kopf}, \text{Zahl}\}^n$ enthält, so sind die Zufallsvariablen X_i und Y messbar, man kann also sowohl die Prägung der Münze als auch die Münzwürfe selbst beobachten. Allerdings hat man das Problem ein Wahrscheinlichkeitsmaß $P \colon \mathcal{A} \to [0,1]$ zu definieren, wobei das Riemannsche Integral nicht ausreicht um $P(A \times B)$ für alle Borelmengen $A \subset (0,1)$ zu definieren. Dieses Problem werden wir in Kapitel 1.1.4 lösen.

Um die Konstruktion eines Wahrscheinlichkeitsmaßes zu vereinfachen, kann man eine kleinere σ-Algebra wählen, zum Beispiel

$$\mathcal{A}' = \big\{(0,1) \times B \colon B \subseteq \{\text{Kopf}, \text{Zahl}\}^n\big\}.$$

Mit dieser Wahl sind nur die Ausgänge der Münzwürfe beobachtbar, nicht aber die Prägung der Münze. Mit anderen Worten, die Zufallsvariablen X_i sind messbar und wir können zum Beispiel $P(X_i = \text{Kopf})$ berechnen, aber die Zufallsvariable Y ist nicht messbar, da $\{Y \in A\}$ nur für die trivialen Fälle $A = (0,1)$ oder $A = \emptyset$ messbar ist. Wir erhalten also keine direkte Information über die Prägung der Münze.

Wir können ein Wahrscheinlichkeitsmaß P auf $\mathcal{A}'$ erklären durch

$$P\big((0,1)\times\{(a_1,\ldots,a_n)\}\big) = P(X_1 = a_1,\ldots,X_n = a_n) := \int_0^1 \theta^{n(a)}(1-\theta)^{n-n(a)}\,d\theta,$$

wobei $n(a)$ die Anzahl der Köpfe in der Familie $a = (a_1,\ldots,a_n)$. Dies ist ein herkömmliches Riemann Integral.

Eine wichtige Frage ist, ob man auch eine unbeschränkte Zahl von Münzwürfen modellieren kann, also

$$\Omega' := \{\text{Kopf}, \text{Zahl}\}^{\mathbb{N}}$$

so dass für alle Zufallsvariablen

$$X_i\colon \Omega' \to \{\text{Kopf}, \text{Zahl}\}, X_i(a_1,a_2,\ldots,) = a_i,$$

messbar sind. Das zugrundeliegende Maß P sollte dann wie zuvor erfüllen, dass

$$P(X_1 = a_1,\ldots,X_n = a_n) = \int_0^1 \theta^{n(a)}(1-\theta)^{n-n(a)}\,d\theta.$$

In Kapitel 1.1.3 werden wir sehen, dass diese Vorschrift P eindeutig bestimmt. Die Prägung der Münze lässt sich dann aus dem Langzeitverhalten der Münzwürfe im nachhinein rekonstruieren. Ein Konvergenzsatz, den wir in Kapitel 2 beweisen werden, liefert nämlich

$$P\Big\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1\{X_i = \text{Kopf}\} \text{ existiert}\Big\} = 1.$$

Dann kann man eine Zufallsvariable Y so definieren, dass

$$Y = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1\{X_i = \text{Kopf}\}$$

mit Wahrscheinlichkeit eins (also fast sicher). Dieses Y spielt dann die Rolle der Prägung der Münze. Es stellt sich damit heraus, dass die Konstruktion von P auf dem unendlichen Folgenraum Ω' und die Konstruktion der Gleichverteilung, die für die Definition von P auf dem messbaren Raum $(\Omega,\mathcal{A})$ erforderlich ist, im wesentlich äquivalent sind.

1.1.3 Der Eindeutigkeitssatz

Gegeben seien zwei Wahrscheinlichkeitsmaße P und Q auf dem messbaren Raum $(\Omega,\mathcal{A})$. Wir suchen ein möglichst kleines Mengensystem $\mathcal{B} \subseteq \mathcal{A}$ so, dass

$$P(B) = Q(B) \text{ für alle } B \in \mathcal{B} \Longrightarrow P = Q.$$

Beobachtung: Es genügt nicht, wenn $\mathcal{B}$ die σ-Algebra $\mathcal{A}$ erzeugt.

Beispiel: Sei $\Omega = \{1,2,3,4\}$ und $\mathcal{B} = \{\{2,3\},\{3,4\}\}$. Dann erzeugt $\mathcal{B}$ die Potenzmenge $\mathcal{P}(\Omega)$, da die einelementigen Mengen

$$\{1\} = (\{2,3\} \cup \{3,4\})^{\mathrm{c}}, \{2\} = \{2,3\} \cap \{3,4\}^{\mathrm{c}},$$

$$\{3\} = \{2,3\} \cap \{3,4\}, \{4\} = \{3,4\} \cap \{2,3\}^{\mathrm{c}}$$

und all ihre endlichen Vereinigungen in jeder σ-Algebra, die $\mathcal{B}$ entält, enthalten sein müssen. Definiert man nun zwei Wahrscheinlichkeitsmaße P und P' durch

$$P'(\{1\}) = \tfrac{1}{6}, P'(\{2\}) = \tfrac{1}{6}, P'(\{3\}) = \tfrac{1}{2}, P'(\{4\}) = \tfrac{1}{6}$$

und

$$P(\{1\}) = 0, P(\{2\}) = \tfrac{1}{3}, P'(\{3\}) = \tfrac{1}{3}, P'(\{4\}) = \tfrac{1}{3},$$

so gilt

$$P'(\{2,3\}) = \tfrac{2}{3} = P(\{2,3\}), P'(\{3,4\}) = \tfrac{2}{3} = P(\{3,4\}),$$

und somit stimmen P und P' auf $\mathcal{B}$ überein, sind aber offensichtlich verschiedene Wahrscheinlichkeitsmaße.

Ein Mengensystem $\mathcal{B}$ heißt *schnittstabil*, wenn aus $A, B \in \mathcal{B}$ folgt, dass $A \cap B \in \mathcal{B}$. Mit Hilfe dieser Definition können wir nun den nützlichen Eindeutigkeitssatz formulieren.

Satz 1. *Stimmen zwei Wahrscheinlichkeitsmaße auf einem schnittstabilen Erzeuger der zugrundeliegenden σ-Algebra überein, so sind sie gleich.*

Zum Beweis nennen wir ein Mengensystem $\mathcal{D}$ aus Teilmengen von Ω ein *Dynkinsystem*, wenn gilt

- $\Omega \in \mathcal{D}$;
- sind $A, B \in \mathcal{D}$ und $A \subseteq B$, so ist $B \setminus A \in \mathcal{D}$;
- wenn $A_n \in \mathcal{D}$ und $A_n \uparrow A$, so ist $A \in \mathcal{D}$.

In den Übungen beweisen wir die folgenden beiden Lemmata.

Lemma 1.1. *Ein Mengensystem $\mathcal{A}$ aus Teilmengen von Ω ist genau dann eine σ-Algebra, wenn $\mathcal{A}$ ein schnittstabiles Dynkinsystem ist.*

Lemma 1.2 (Dynkin's Lemma)**.** *Ist $\mathcal{B}$ ein schnittstabiles Mengensystem aus Teilmengen von Ω, so enthält jedes Dynkinsystem, das $\mathcal{B}$ enthält, auch die von $\mathcal{B}$ erzeugte σ-Algebra.*

Nun beweisen wir den Eindeutigkeitssatz. Dazu bezeichnen wir mit $\mathcal{B}$ den schnittstabilen Erzeuger der σ-Algebra $\mathcal{A}$ und nehmen an, dass P_1 und P_2 auf $\mathcal{B}$ übereinstimmen. Wir betrachten

$$\mathcal{D} = \{B \in \mathcal{A} \colon P_1(B) = P_2(B)\}.$$

Es ist leicht zu sehen, dass $\mathcal{D}$ ein Dynkinsystem ist. [Wir haben $\Omega \in \mathcal{D}$, da P_1, P_2 Wahrscheinlichkeitsmaße sind; wenn $A \subseteq B$ in $\mathcal{D}$ liegt, so ist $P_1(B \setminus A) = P_1(B) - P_1(A) = P_2(B) - P_2(A) = P_2(B \setminus A)$ und somit $B \setminus A \in \mathcal{D}$; wenn $A_n \uparrow A$ mit $A_n \in \mathcal{D}$ gilt $P_1(A) = \lim P_1(A_n) = \lim P_2(A_n) = P_2(A)$ wegen der Stetigkeit von Wahrscheinlichkeitsmaßen, und somit $A \in \mathcal{D}$.]
Da $\mathcal{B} \subseteq \mathcal{D}$ nach Annahme folgt aus Dynkins Lemma, dass $\mathcal{D}$ auch $\mathcal{A}$ enthält, was zu beweisen war.

Eine nützliche Anwendung des Eindeutigkeitssatzes ist der folgende Satz.

Satz 2 (Satz über monotone Klassen)**.** *Sei $\mathcal{H}$ eine Menge beschränkter messbarer Funktionen $X \colon (\Omega, \mathcal{A}) \to \mathbb{R}$ mit den Eigenschaften*

- *$\mathcal{H}$ ist ein Vektorraum,*
- *$\mathcal{H}$ entält die konstante Funktion* 1*,*
- *Ist X_n eine Folge nichtnegativer Funktionen in $\mathcal{H}$ und $X_n \uparrow X$ für eine beschränkte Funktion X, so gilt $X \in \mathcal{H}$.*

Ein solches $\mathcal{H}$ heißt monotone Klasse*. Wenn $\mathcal{H}$ alle Indikatorfunktionen eines schnittstabilen Erzeugers von $\mathcal{A}$ enthält, so enthält $\mathcal{H}$ auch alle beschränkten messbaren Funktionen.*

Beweis: Das Mengensystem $\mathcal{D}$ aller Mengen A mit $1_A \in \mathcal{H}$ ist ein Dynkinsystem, das einen schnittstabilen Erzeuger und somit alle Elemente von $\mathcal{A}$ enthält. Ist X nichtnegativ, messbar und beschränkt, etwa durch $K > 0$, dann definieren wir $A(n,i) := \{\omega \colon i2^{-n} \leq X(\omega) < (i+1)2^{-n}\}$ und

$$X_n := \sum_{i=0}^{K2^n} i2^{-n} 1_{A(n,i)}.$$

Dann ist $X_n \in \mathcal{H}$ und $0 \leq X_n \uparrow X$, so dass $X \in \mathcal{H}$. Eine beschränkte messbare Funktion schließlich ist die Differenz zweier beschränkter nichtnegativer messbarer Funktionen und damit auch in $\mathcal{H}$.

Beispiel: Auf $\Omega = \{\text{Kopf}, \text{Zahl}\}^{\mathbb{N}}$ ist

$$\Big\{\{(\omega_n)_{n\in\mathbb{N}} \colon \omega_1 = a_1, \dots, \omega_n = a_n\} \colon \quad n \in \mathbb{N}, a_i \in \{\text{Kopf}, \text{Zahl}\}\Big\} \cup \{\emptyset\}$$

ein schnittstabiles Mengensystem und somit kann es *höchstens ein* Wahrscheinlichkeitsmaß P geben mit

$$P\big(\{(\omega_n)_{n\in\mathbb{N}} \colon \omega_1 = a_1, \dots, \omega_n = a_n\}\big) = \int_0^1 \theta^{n(a)}(1-\theta)^{n(a)}\, d\theta.$$

1.1.4 Der Existenzsatz

Der Existenzsatz für Maße, das Hauptresultat dieses Kapitels, erlaubt es uns Maße zu konstruieren, ohne ihre Werte auf allen Elementen der σ-Algebra explizit vorgeben zu müssen.

Satz 3 (Existenzsatz von Carathéodory)**.** *Sei $\mathcal{A}_0$ eine Algebra auf Ω und $\mathcal{A}$ die von $\mathcal{A}_0$ erzeugte σ-Algebra. Sei $P_0 : \mathcal{A}_0 \to [0, \infty]$* σ-additiv*, das heißt $P_0(\emptyset) = 0$ und wenn $A_1, A_2, \dots$ paarweise disjunkt mit $A_n \in \mathcal{A}_0$ für alle n und $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_0$, so gilt*

$$P_0\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P_0(A_n).$$

Dann existiert ein Maß P auf $(\Omega, \mathcal{A})$ mit $P = P_0$ auf $\mathcal{A}_0$. Ist zusätzlich $P_0(\Omega) = 1$, so ist P ein Wahrscheinlichkeitsmaß und eindeutig bestimmt.

Die Eindeutigkeit im Falle von Wahrscheinlichkeitsmaßen folgt aus dem Eindeutigkeitssatz, da die Algebra $\mathcal{A}_0$ schnittstabiler Erzeuger von $\mathcal{A}$ ist. Der Beweis des Existenzsatzes benutzt das folgende Lemma, das wir als Übungsaufgabe beweisen.

Lemma 1.3. *Sei $\mathcal{A}_0$ eine Algebra auf Ω und*

$$P_0 \colon \mathcal{A}_0 \to [0, \infty] \ \textit{mit}\ P_0(\emptyset) = 0.$$

*Eine Menge $A \in \mathcal{A}_0$ heißt P_0-*messbar*, wenn für alle $B \in \mathcal{A}_0$ gilt*

$$P_0(A \cap B) + P_0(A^{\mathrm{c}} \cap B) = P_0(B).$$

Dann ist das Mengensystem $\mathcal{A}_1$ der P_0-messbaren Mengen eine Algebra und P_0 erfüllt

$$P_0\Big(\bigcup_{k=1}^{n} (A_k \cap B)\Big) = \sum_{k=1}^{n} P_0(A_k \cap B),$$

wenn $A_1, \dots, A_n \in \mathcal{A}_1$ paarweise disjunkt sind und $B \in \mathcal{A}_0$.

Eine Abbildung $P_0\colon \mathcal{A}_0 \to [0,\infty]$ heißt *äusseres Maß*, wenn gilt

- $P_0(\emptyset) = 0$,
- wenn $A_1 \subseteq A_2$ in $\mathcal{A}_0$ liegen, so gilt $P_0(A_1) \le P_0(A_2)$;
- wenn $A_1, A_2, \ldots$ in $\mathcal{A}_0$ liegen, so gilt

$$P_0\Big(\bigcup_{k=1}^{\infty} A_k\Big) \le \sum_{k=1}^{\infty} P_0(A_k).$$

Lemma 1.4 (Carathéodory's Lemma)**.** *Sei P ein äusseres Maß auf dem messbaren Raum $(\Omega, \mathcal{A})$. Dann bilden die P-messbaren Mengen eine σ-Algebra $\mathcal{A}'$ auf der P σ-additiv ist, so dass P sogar ein Maß auf $(\Omega, \mathcal{A}')$ ist.*

Das Lemma wird mit Hilfe von Lemma 2.1 als Übungsaufgabe bewiesen. Der Beweis des Existenzsatzes erfolgt nun mit Hilfe von Carathéodory's Lemma in vier Schritten.

Schritt 1: Zunächst betrachten wir die σ-Algebra $\mathcal{P}(\Omega)$ aller Teilmengen von Ω und definieren $P\colon \mathcal{P}(\Omega) \to [0,\infty]$ durch

$$P(A) = \inf \sum_{n=1}^{\infty} P_0(B_n),$$

wobei das Infimum über alle Folgen (B_n) in $\mathcal{A}_0$ mit $A \subseteq \bigcup_n B_n$ gebildet wird. Wir zeigen, dass P ein äusseres Maß ist.

[Die ersten beiden Eigenschaften sind offensichtlich. Seien A_n so, dass $P_0(A_n) < \infty$ für alle n. Für jedes $\varepsilon > 0$ existieren $B_{n,k} \in \mathcal{A}_0$ mit $A_n \subseteq \bigcup_n B_{n,k}$ and $\sum_k P_0(B_{n,k}) < P(A_n) + \varepsilon 2^{-n}$. Dann folgt $\bigcup A_n \subseteq \bigcup_{n,k} B_{n,k}$ und

$$P\Big(\bigcup_n A_n\Big) \le \sum_{n,k} P_0\big(B_{n,k}\big) \le \sum_n \big(P(A_n) + \varepsilon 2^{-n}\big) \le \sum_n P(A_n) + \varepsilon,$$

und die dritte Eigenschaft (Subadditivität) folgt, da $\varepsilon > 0$ beliebig war.]

Schritt 2: Nach Carathéodory's Lemma ist nun P ein Maß auf der σ-Algebra der P-messbaren Mengen. Wenn wir zeigen, dass

- $P = P_0$ auf $\mathcal{A}_0$,
- jedes Element von $\mathcal{A}_0$ auch P-messbar ist,

so ist auch jedes Element der von $\mathcal{A}_0$ erzeugten σ-Algebra $\mathcal{A}$ P-messbar und wir erhalten das gewünschte Maß durch Einschränkung von P auf $\mathcal{A}$.

Schritt 3: Wir zeigen. dass $P = P_0$ auf $\mathcal{A}_0$.

[Sei $A \in \mathcal{A}_0$, dann gilt offensichtlich $P(A) \leq P_0(A)$. Sei nun $A \subseteq \bigcup_n B_n$ mit $B_n \in \mathcal{A}_0$. Wir definieren eine disjunkte Folge $B_n' \in \mathcal{A}_0$ durch $B_1' := B_1$,

$$B_n' := B_n \cap \Big(\bigcup_{k<n} B_k \Big)^{\mathrm{c}},$$

so dass $B_n' \subseteq B_n$, $\bigcup_n B_n' = \bigcup_n B_n$. Die σ-Additivität liefert dann

$$P_0(A) = P_0\Big(\bigcup_n (A \cap B_n') \Big) = \sum_n P_0(A \cap B_n') \leq \sum_n P_0(B_n).$$

Folglich ist $P_0(A) \leq P(A)$ wie gefordert.]

Schritt 4: Wir zeigen. dass jedes Element von $\mathcal{A}_0$ auch P-messbar ist.

[Sei $A \in \mathcal{A}_0$ und $B \in \mathcal{P}(\Omega)$. Für jedes $\varepsilon > 0$ existieren $B_n \in \mathcal{A}_0$ mit $B \subseteq \bigcup_n B_n$ und $\sum_n P_0(B_n) < P(B) + \varepsilon$. Es folgt

$$\begin{aligned} \sum_n P_0(B_n) &= \sum_n P_0(A \cap B_n) + \sum_n P_0(A^{\mathrm{c}} \cap B_n) \\ &\geq P(A \cap B) + P(A^{\mathrm{c}} \cap B). \end{aligned}$$

Da $\varepsilon > 0$ beliebig war folgt, dass $P(B) \geq P(A \cap B) + P(A^{\mathrm{c}} \cap B)$. Die umgekehrte Ungleichung folgt aus der Subadditivität von P, womit gezeigt ist, dass A P-messbar ist.]

Als wichtigste Anwendung des Existenzsatzes konstruieren wir nun das Lebesguemaß oder Volumenmaß auf $(0,1]^d$ (und damit auch auf $\mathbb{R}^d$). Wir betrachten die Algebra $\mathcal{A}_0$ der endlichen Vereinigung von Rechtecken

$$Q = (a_1, b_1] \times \cdots (a_d, b_d],$$

wobei $0 < a_i \leq b_i \leq 1$. Die Algebra $\mathcal{A}_0$ erzeugt die Borelsche σ-Algebra. Jedes Element von $\mathcal{A}_0$ kann als Vereingung von endlich vielen *paarweise disjunkten* derartigen Rechtecken

$$A = Q_1 \cup \cdots \cup Q_n$$

geschrieben werden, wobei wir die Rechtecke schreiben als

$$Q_k = (a_1^{(k)}, b_1^{(k)}] \times \cdots \times (a_d^{(k)}, b_d^{(k)}].$$

Obwohl eine solche Darstellung nicht eindeutig ist, ist es leicht zu sehen, dass die Abbildung

$$P_0\colon \mathcal{A}_0 \to [0,\infty], \quad P_0(A) = \sum_{k=1}^{n} \big(b_1^{(k)} - a_1^{(k)}\big) \cdots (b_d^{(k)} - a_d^{(k)}\big)$$

wohldefiniert ist. Wir zeigen nun:

Lemma 1.5. *P_0 ist σ-additiv.*

Beweis: $P_0(\emptyset) = 0$ nach Definition. Seien nun $A_n \in \mathcal{A}_0$ paarweise disjunkt und $A = \bigcup_n A_n \in \mathcal{A}_0$. Dann erfüllt $A_n' = \bigcup_{k=1}^n A_k$ dass

$$P_0(A_n') = \sum_{k=1}^{n} P_0(A_k). \tag{1.1}$$

Wir zeigen, dass $P_0(A_n') \uparrow P_0(A)$, so dass die σ-Additivität durch Grenzübergang in (1.1) folgt. Dazu setzen wir $H_n := A \setminus A_n'$, so dass $H_n \in \mathcal{A}_0$ und $H_n \downarrow \emptyset$. Wenn $P_0(H_n) \downarrow 0$ folgt aus $P(H_n) = P(A) - P_0(A_n')$ dass $P_0(A_n') \uparrow P(A)$, wie gefordert.

Wir zeigen $P_0(H_n) \downarrow 0$ durch Widerspruch. Nehmen wir also an, dass $P_0(H_n) \geq 2\varepsilon$ für ein festes $\varepsilon > 0$. Aus der Definition von $\mathcal{A}_0$ folgt, dass es $J_k \in \mathcal{A}_0$ gibt mit $\bar{J}_k \subseteq H_k$ und $P_0\big(H_k \setminus J_k\big) \leq \varepsilon 2^{-k}$. Da H_n fallend ist folgt

$$P_0\Big(H_n \setminus \bigcap_{k\leq n} J_k\Big) \leq P_0\Big(\bigcup_{k\leq n} H_k \setminus J_k\Big) \leq \sum_{k\leq n} \varepsilon 2^{-k} < \varepsilon.$$

Da $P_0(H_n) \geq 2\varepsilon$ folgt, für jedes n, dass

$$P_0\Big(\bigcap_{k\leq n} J_k\Big) > \varepsilon.$$

Daraus folgt, dass $\bigcap_{k\leq n} J_k \neq \emptyset$. Somit ist auch

$$K_n := \bigcap_{k\leq n} \bar{J}_k \neq \emptyset$$

und da K_n kompakt und fallend ist folgt

$$\bigcap_{k=1}^{\infty} \bar{J}_k = \bigcap_{n=1}^{\infty} K_n \neq \emptyset.$$

Damit ist aber auch $\bigcap_k H_k \neq \emptyset$ im Widerspruch zu unseren Annahmen.

Da P_0 also σ-additiv ist, existiert nach dem Existenzsatz eine eindeutige Fortsetzung P von P_0 auf $\mathcal{A}$, die ein Wahrscheinlichkeitsmaß ist. Dieses Maß ist das *Lebesguemaß* auf $(0,1]^d$. Durch Verschiebung

$$P^{(n)}(A) := P(\{x \in (0,1]^d \colon x + n \in A\})$$

lässt sich dieses Maß auch auf die Kuben $\prod_{i=1}^d (n_i, n_i + 1]$ übertragen und durch Summation über alle Vektoren $n = (n_1, \ldots, n_d) \in \mathbb{Z}^d$ erhält man schließlich das *Lebesguemaß* oder *Volumenmaß* Vol$=\sum_{n \in \mathbb{Z}^d} P^{(n)}$ auf $\mathbb{R}^d$.

Bemerkung: Der Beweis des Existenzsatzes beinhaltet eine explizite Definition des Lebesguemaßes als äusseres Maß, nämlich

$$\mathrm{Vol}(A) = \inf \sum_{k=1}^{\infty} \left(b_1^{(k)} - a_1^{(k)}\right) \cdots \left(b_d^{(k)} - a_d^{(k)}\right),$$

wobei das Infimum über alle Folgen von Rechtecken (Q_k) mit $A \subseteq \bigcup_k Q_k$ geht. In der Einführung in die Stochastik haben wir diese Formel benutzt, ohne zu beweisen, dass die so definierte Abbildung Vol ein Maß auf der Borelschen σ-Algebra darstellt. Dies ist hiermit nachgeholt.

1.2 Lebesgue Integrationstheorie

Wenn ein Maß P auf einem messbaren Raum $(\Omega, \mathcal{A})$ gegeben ist, so wollen wir für messbare, reellwertige Funktionen $X \colon \Omega \to \mathbb{R}$ auch ein Integral $\int X \, dP$ definieren. Im wichtigen Fall, dass P ein Wahrscheinlichkeitsmaß ist entspricht dies dem mittleren Wert der Funktion X oder in anderen Worten dem Erwartungswert der Zufallsvariablen X.

In der Einführung in die Stochastik haben wir solche Integrale auf den Riemannschen Integrationsbegriff zurückgeführt. Dieser Ansatz erlaubt aber keinen direkten Zugriff auf Resultate über die Vertauschbarkeit von Integration und Limesbildung, die für die Wahrscheinlichkeitstheorie von zentraler Bedeutung sind. Wir verfolgen deshalb hier einen anderen Ansatz und konstruieren den Integralbegriff direkt und ohne Rückgriff auf die Riemannsche Integrationstheorie.

1.2.1 Integration nichtnegativer Funktionen

Wir definieren das Lebesguesche Integral in mehreren Schritten. Wir betrachten zunächst *einfache* Funktionen X, das sind solche, die nur eindlich viele nichtnegative Werte annehmen, dh. es existieren $a_k \in [0,\infty]$ und $A_k \in \mathcal{A}$, $1 \le k \le m$, so dass

$$X = \sum_{k=1}^{m} a_k 1_{A_k},$$

wobei wir (immer) die Konvention $0 \cdot \infty = \infty \cdot 0 = 0$ verwenden wollen. Wir definieren dann

$$\int X\,dP := \sum_{k=1}^{m} a_k P(A_k),$$

und beobachten, dass das Integral dadurch wohldefiniert ist (dh. unterschiedliche Darstellungen derselben Funktion liefern denselben Integralwert). Man prüft leicht nach, dass das Integral für einfache Funktionen die folgenden Eigenschaften hat

- **Linearität:** $\int X+Y\,dP = \int X\,dP + \int Y\,dP$ und $\int cX\,dP = c\int X\,dP$.
- **Monotonie:** Aus $X \le Y$ folgt $\int X\,dP \le \int Y\,dP$.
- **Stabilität:** Wenn $P(X \ne Y) = 0$ so folgt $\int X\,dP = \int Y\,dP$.

Jedes messbare, nichtnegative Funktionen $X\colon \Omega \to [0,\infty]$ kann durch einfache Funktionen von unten approximiert werden. Wir setzen

$$X_n^* := \begin{cases} 0 & \text{wenn } X = 0, \\ (i-1)2^{-n} & \text{wenn } (i-1)2^{-n} < X \le i2^{-n} \le n, \\ n & \text{wenn } X > n. \end{cases} \tag{1.2}$$

Dann gilt $X_n^* \uparrow X$. Dies motiviert die Approximation von $\int X\,dP$ durch die Integrale minorisierender einfacher Funktionen. Wir setzen also für messbare, nichtnegative Funktionen $X\colon \Omega \to [0,\infty]$

$$\int X\,dP := \sup\Big\{ \int X'\,dP \colon X' \le X \text{ einfach } \Big\},$$

und folgern aus der Monontonieeigenschaft, dass dies eine Fortsetzung des Integralbegriffs für einfache Funktionen ist.

1.2.2 Vertauschungssätze für Integrale und Limiten

Wir beweisen nun den ersten (und wichtigsten) Vertauschungssatz.

Satz 4 (Satz von der monotonen Konvergenz)**.** *Ist $X_n\colon \Omega \to [0,\infty]$ messbar so, dass $X_n \uparrow X$, das heisst die Folge X_n konvergiert wachsend gegen X, so konvergiert auch die Folge der Integrale $\int X_n\,dP$ wachsend gegen $\int X\,dP$.*

Wir beweisen zunächst eine Version für einfache Funktionen.

Lemma 1.6. *Sei X eine einfache Funktion und X_n einfache Funktionen mit $X_n \uparrow X$. Dann gilt $\int X_n \, dP \uparrow \int X \, dP$.*

Beweis: Aufgrund der Linearität des Integrals genügt es den Fall $X = 1_A$ zu betrachten. Aus $X_n \leq X$ folgt $\int X_n \, dP \leq P(A)$ und es bleibt nur zu zeigen, dass

$$\liminf \int X_n \, dP \geq P(A).$$

Sei $\varepsilon > 0$ und $A_n := \{\omega \in A \colon X_n(\omega) > 1 - \varepsilon\}$. Dann gilt $A_n \uparrow A$ und aufgrund der Stetigkeit von Maßen folgt $P(A_n) \uparrow P(A)$. Da $(1-\varepsilon)1_{A_n} \leq X_n$ folgt $(1-\varepsilon)P(A_n) \leq \int X_n \, dP$ und die Aussage folgt durch Kombination dieser beiden Fakten da $\varepsilon > 0$ beliebig war.

Für den nächsten Schritt brauchen wir nun ein Lemma, dessen einfachen Beweis wir als Übungsaufgabe geben.

Lemma 1.7 (Doppelfolgenlemma)**.** *Sei $(a_{k,n} \colon k, n \in \mathbb{N})$ eine Doppelfolge mit Werten in $[0, \infty]$, so dass für jedes feste k die Folge $(a_{k,n} \colon n \in \mathbb{N})$ und für jedes feste n die Folge $(a_{k,n} \colon k \in \mathbb{N})$ monoton wachsend ist. Dann gilt*

$$\lim_k \lim_n a_{k,n} = \lim_n \lim_k a_{k,n},$$

und alle involvierten Limiten sind monoton.

Lemma 1.8 (Vertauschungslemma)**.** *Seien X_n und Y_n, $n \in \mathbb{N}$, einfache Funktionen mit $X_n \uparrow X$ und $Y_n \uparrow X$. Dann gilt*

$$\lim \int X_n \, dP = \lim \int Y_n \, dP.$$

Beweis: Setze $X_{n,k} := X_n \wedge Y_k$. Dann gilt $X_{n,k} \uparrow X_n$ für $k \to \infty$ und $X_{n,k} \uparrow Y_k$ für $n \to \infty$. Nach Lemma 1.6 folgt

$$\int X_{n,k} \, dP \uparrow \int Y_k \, dP \text{ für } n \to \infty, \qquad \int X_{n,k} \, dP \uparrow \int X_n \, dP \text{ für } k \to \infty.$$

Nach dem Doppelfolgenlemma folgt die Behauptung.

Wir können nun Lemma 1.6 auf den Fall erweitern, dass die Grenzfunktion nicht notwendig einfach ist.

Lemma 1.9. *Seien X_n einfache Funktionen mit $X_n \uparrow X$. Dann gilt*

$$\int X_n \, dP \uparrow \int X \, dP.$$

Beweis: Nach Konstruktion des Integrals gibt es eine Folge einfacher Funktionen X_n' mit $X_n' \leq X$ und $\int X_n' \, dP \uparrow \int X \, dP$. Ausserdem haben wir eine Folge einfacher Funktionen X_n^* mit $X_n^* \uparrow X$ konstruiert. Setzen wir nun

$$Y_n := \max\{X_n^*, X_1', \dots, X_n'\},$$

so gilt $Y_n \uparrow X$ und $\int Y_n \, dP \uparrow \int X \, dP$. Nach dem Vertauschungslemma gilt dann auch $\int X_n \, dP \uparrow \int X \, dP$, wie gefordert.

Um den Beweis des Satzes von der monotonen Konvergenz zu komplettieren betrachten wir die einfachen Funktionen $X_{n,m}^*$ und X_m^* mit $X_{n,m}^* \uparrow X_n$ und $X_m^* \uparrow X$, die wir in (1.2) konstruiert haben. Nach Konstruktion gilt auch $X_{n,m}^* \uparrow X_m^*$ für $n \to \infty$. Nach Lemma 1.6 folgt

$$\int X_{n,m}^* \, dP \uparrow \int X_m^* \, dP \qquad \text{für } n \to \infty.$$

Nach Lemma 1.9 folgt

$$\int X_{n,m}^* \, dP \uparrow \int X_n \, dP \qquad \text{und} \quad \int X_m^* \, dP \uparrow \int X \, dP \qquad \text{für } m \to \infty.$$

Das Doppelfolgenlemma liefert somit

$$\begin{aligned} \lim_n \int X_n \, dP &= \lim_n \lim_m \int X_{n,m}^* \, dP = \lim_m \lim_n \int X_{n,m}^* \, dP \\ &= \lim_m \int X_m^* \, dP = \int X \, dP, \end{aligned}$$

was behauptet war.

Mit Hilfe des Satzes von der monotonen Konvergenz kann man nun recht einfach die Eigenschaften Linearität, Mononomie und Stabilität von einfachen auf nichtnegative, messbare Zufallsvariable ausdehnen.

Eine weitere nützliche Folgerung aus dem Satz von der monotonen Konvergenz ist das Lemma von Fatou.

Satz 5 (Lemma von Fatou)**.** *Sind $X_n \colon \Omega \to [0, \infty]$ messbar, so gilt*

$$\int \liminf X_n \, dP \leq \liminf \int X_n \, dP.$$

Beweis: Sei $Y_k := \inf_{n \geq k} X_n$. Dann ist (Y_n) monoton wachsend und $Y_k \leq X_n$ für alle $n \geq k$, woraus folgt, dass

$$\int Y_k \, dP \leq \inf_{n \geq k} \int X_n \, dP.$$

Da $\lim Y_k = \liminf X_n$ folgt aus dem Satz von der monotonen Konvergenz

$$\begin{aligned}\int \liminf_n X_n \, dP &= \int \lim_k Y_k \, dP = \lim_k \int Y_k \, dP \\ &\leq \lim_k \inf_{n \geq k} \int X_n \, dP = \liminf_n \int X_n \, dP.\end{aligned}$$

Beispiel: Auf $\Omega = (0,1) \times \{\text{Kopf, Zahl}\}^n$ betrachten wir die Algebra $\mathcal{A}_0$ der endlichen Vereinigungen von Rechtecken $A \times B$ mit $A \subseteq (0,1)$ Borel und $B \subseteq \{\text{Kopf, Zahl}\}^n$ beliebig. Definiere P_0 auf $\mathcal{A}_0$ durch

$$P_0(A \times B) = \sum_{a \in B} \int 1_A(\theta) \theta^{n(a)} (1-\theta)^{n-n(a)} \, d\theta,$$

wobei $d\theta$ hier Lebesgueintegration bezüglich des Volumenmaßes auf $(0,1)$ bezeichnet. P_0 ist nach dem Satz von den monotonen Konvergenz σ-additiv. Folglich existiert ein eindeutiges Wahrscheinlichkeitsmaß P, das P_0 auf die von $\mathcal{A}_0$ erzeugte σ-Algebra fortsetzt.

1.2.3 Integrabilität

Wir betrachten reellwertige (messbare) Zufallsvariablen $X \colon \Omega \to \mathbb{R}$ indem wir ihren Positivteil $X^+ = \max\{X, 0\}$ und Negativteil $X^- = -\min\{X, 0\}$ betrachten. Positivteil und Negativteil einer Funktion sind nichtnegativ mit

$$X = X^+ - X^- \text{ und } |X| = X^+ + X^-.$$

Wenn $\int |X| \, dP < \infty$, so heißt die Funktion X *integrierbar*. Wir definieren

$$\int X \, dP := \int X^+ \, dP - \int X^- \, dP,$$

und beobachten, dass mit dieser Definition die *Dreiecksungleichung*

$$\Big| \int X \, dP \Big| \leq \int |X| \, dP$$

gilt. Zudem übertragen sich Linearität, Monotonie und Stabilität des Integrals von nichtnegativen auf integrierbare Funktionen.

Wir beschließen diesen Abschnit mit einem weiteren "Vertauschungssatz".

Satz 6 (Satz von der dominierten Konvergenz)**.** *Seien $X_n, X \colon \Omega \to \mathbb{R}$ messbare Funktionen mit $\lim X_n = X$ und Y eine integrierbare Funktion mit*

$$|X_n| \leq Y \text{ für alle } n.$$

Dann gilt

$$\lim_{n\to\infty} \int \big|X_n - X\big|\, dP = 0,$$

und insbesondere folgt daraus auch $\lim \int X_n\, dP = \int X\, dP$.

Beobachten wir zunächst, dass die Vertauschbarkeit von Limes und Integral keine Selbstverständlichkeit ist. Ist zum Beispiel $\Omega = (0,1)$ und P das Lebesguemaß auf der Borelschen σ-Algebra, so gilt für die Funktionen $X_n = n1_{(0,\frac{1}{n})}$ dass $X_n \to 0$ aber $\int X_n\, dP = 1$ für alle n. In diesem Fall können die X_n also nicht durch eine integrierbare Funktion Y dominiert werden.

Beweis: Wir haben $|X_n - X| \le 2Y$ und wenden Fatou's Lemma auf $Z_n := 2Y - |X_n - X| \ge 0$ an. Wir erhalten

$$\begin{aligned} \limsup \int |X_n - X|\, dP &= \int 2Y\, dP - \liminf \int Z_n\, dP \\ &\le \int 2Y\, dP - \int \liminf Z_n\, dP \\ &= \int \limsup |X_n - X|\, dP = 0. \end{aligned}$$

Die weitere Aussage folgt aus der Dreiecksungleichung.

1.3 Produktmaße und Fubinis Theorem

1.3.1 Produktmaße und Unabhängigkeit

Gegeben zwei Wahrscheinlichkeitsräume $(\Omega_1, \mathcal{A}_1, P_1)$ und $(\Omega_2, \mathcal{A}_2, P_2)$ wollen wir nun einen Produktraum

$$(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$$

konstruieren, der die *unabhängige* Durchführung der durch die beiden Wahrscheinlichkeitsräume gegebenen Experimente modellieren soll. Hierbei ist $\Omega_1 \times \Omega_2$ das herkömmliche kartesische Produkt und $\mathcal{A}_1 \otimes \mathcal{A}_2$ die von den Produktmengen $A_1 \times A_2$ mit $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$ erzeugte σ-Algebra.

Die Menge der beschränkten messbaren Funktionen $X\colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ mit

- für alle $\omega_1 \in \Omega_1$ ist $X(\omega_1, \cdot)\colon \omega_2 \mapsto X(\omega_1, \omega_2)$ meßbar,
- für alle $\omega_2 \in \Omega_2$ ist $X(\cdot, \omega_2)\colon \omega_1 \mapsto X(\omega_1, \omega_2)$ meßbar,

ist eine monotone Klasse, die alle Indikatorfunktionen der Mengen des schnittstabilen Erzeugers

$$\mathcal{I} := \{A_1 \times A_2 \colon A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$$

enthält. Damit enthält sie alle beschränkten, bezüglich $\mathcal{A}_1 \otimes \mathcal{A}_2$ messbaren Funktionen $X \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ und wir können für jedes solche X die Integrale

$$\int X(\omega_1, \cdot)\, dP_2 \text{ und } \int X(\cdot, \omega_2)\, dP_1$$

definieren. Die Menge aller beschränkten messbaren Funktionen $X \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ so, dass

- die Funktion $\int X\, dP_2 \colon \omega_1 \mapsto \int X(\omega_1, \cdot)\, dP_2$ ist beschränkt und messbar,
- die Funktion $\int X\, dP_1 \colon \omega_2 \mapsto \int X(\cdot, \omega_2)\, dP_1$ ist beschränkt und messbar,
- es gilt $\int \big(\int X\, dP_2 \big)\, dP_1 = \int \big(\int X\, dP_1 \big)\, dP_2$,

ist eine monotone Klasse, die alle Indikatorfunktionen der Mengen des schnittstabilen Erzeugers $\mathcal{I}$ enthält. Damit folgt dann der folgende Satz.

Satz 7 (Satz von Fubini)**.** *Es existiert genau ein Wahrschinlichkeitsmaß* $P = P_1 \otimes P_2$ *auf* $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ *so, dass*

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2) \quad \text{für alle } A_1, \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

Ausserdem gilt für alle beschränkten messbaren Zufallsvariablen X*,*

$$\int X\, dP_1 \otimes P_2 = \int \Big(\int X\, dP_2 \Big)\, dP_1 = \int \Big(\int X\, dP_1 \Big)\, dP_2.$$

Beweis: Das Wahrscheinlichkeitsmaß P kann durch

$$P(A) = \int \Big(\int 1_A\, dP_2 \Big)\, dP_1 = \int \Big(\int 1_A\, dP_1 \Big)\, dP_2 \text{ für } A \in \mathcal{A}$$

definiert werden. Die Axiome eines Wahrscheinlichkeitsmasses folgen dann aus der Linearität des Integrals und dem Satz von der monontonen Konvergenz. Die Eindeutigkeit folgt aus dem Eindeutigkeitssatz und die zusätzliche Aussage aus dem Satz von den monotonen Klassen.

Bemerkungen:

(a) Es ist nun leicht zu sehen, dass die Zufallsvariablen

$$X_1 \colon \Omega_1 \times \Omega_2, (\omega_1, \omega_2) \mapsto \omega_1, \qquad X_2 \colon \Omega_1 \times \Omega_2, (\omega_1, \omega_2) \mapsto \omega_2$$

auf dem Wahrscheinlichkeitsraum $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$ unabhängig sind und die Verteilung P_1 bzw. P_2 haben.

(b) Eine Erweiterung des Satzes auf Produkträume mit mehr als zwei (aber endlich vielen) Faktoren ist offensichtlich.

1.3.2 Unendliche Produkte

Wir wollen nun zeigen, dass zu jeder Folge von Wahrscheinlichkeitsmaßen (P_n) auf $\mathbb{R}$ ein Wahrscheinlichkeitsraum konstruiert werden kann auf dem eine Folge $X_1, X_2, \ldots$ von unabhängigen Zufallsvariablen mit Verteilungen P_1, $P_2, \ldots$ existiert.

Satz 8. *Sei (P_n) eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$, Ω die Menge aller reellen Folgen $(\omega_n)_{n\in\mathbb{N}}$ und*

$$X_n\colon \Omega \to \mathbb{R}, \qquad X_n\big((\omega_n)_{n\in\mathbb{N}}\big) = \omega_n.$$

Sei $\mathcal{A}$ die kleinste σ-Algebra, die alle Zufallsvariablen X_n messbar macht. Dann gibt es genau ein Wahrscheinlichkeitsmaß P auf dem messbaren Raum $(\Omega, \mathcal{A})$ mit der Eigenschaft

$$P\big(\big\{(\omega_n)\colon \omega_1 \in B_1, \ldots, \omega_k \in B_k\big\}\big) = \prod_{n=1}^{k} P_n(B_n)$$

für alle $B_1, \ldots, B_k \in \mathcal{B}$. Ausserdem sind auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ die Zufallsvariablen $X_1, X_2, \ldots$ unabhängig und für jedes $n \in \mathbb{N}$ hat X_n die Verteilung P_n.

Der Rest dieses Abschnitts ist dem Beweis dieses Satzes gewidmet. Zunächst fixieren wir $k \in \mathbb{N}$ und betrachten die von

$$\big\{(\omega_n)\colon \omega_1 \in B_1, \ldots, \omega_k \in B_k\big\}, \qquad \text{für } B_1, \ldots, B_k \in \mathcal{B}(\mathbb{R}),$$

erzeugte σ-Algebra $\mathcal{A}_k$. Die Elemente von $\mathcal{A}_k$ können geschrieben werden als

$$A_k = \{(\omega_n)\colon (\omega_1, \ldots, \omega_k) \in \tilde{B}_k\},$$

für $\tilde{B}_k \in \mathcal{B}(\mathbb{R}^k)$. Auf der Algebra

$$\mathcal{A}_0 := \bigcup_k \mathcal{A}_k$$

definieren wir $P_0\colon \mathcal{A}_0 \to [0,1]$ durch

$$P_0\big(\{(\omega_n)\colon (\omega_1, \ldots, \omega_k) \in \tilde{B}_k\}\big) = P_1 \otimes \cdots \otimes P_k(\tilde{B}_k).$$

P_0 ist endlich additiv auf $\mathcal{A}_0$. Außerdem ist $(\Omega, \mathcal{A}_k, P_0)$ ein Wahrscheinlichkeitsraum auf dem $X_1, \ldots, X_k$ unabhängig sind so, dass X_n Verteilung P_n hat. Um den Existenzsatz von Carathéodory anzuwenden brauchen wir nun das folgende Lemma.

Lemma 1.10. *Ist (H_n) eine fallende Folge von Mengen in $\mathcal{A}_0$ und existiert $\varepsilon > 0$ so, dass $P_0(H_n) \geq \varepsilon$ für all n, so ist $\bigcap_n H_n \neq \emptyset$.*

Der Beweis dieses Lemmas ist wieder eine Übungsaufgabe. Betrachten wir hier, wie die Behauptung aus diesem Lemma folgt. Wir zeigen zunächst, dass P_0 σ-additiv auf $\mathcal{A}_0$ ist. Seien also A_n paarweise disjunkt und $\bigcup_k A_k =: A \in \mathcal{A}_0$. Dann erfüllt $A'_n = \bigcup_{k=1}^n A_k$ dass

$$P_0(A'_n) = \sum_{k=1}^{n} P_0(A_k). \tag{1.3}$$

Wir zeigen, dass $P_0(A'_n) \uparrow P(A)$, so dass die σ-Additivität durch Grenzübergang in (1.3) folgt. Dazu setzen wir $H_n := A \setminus A'_n$, so dass $H_n \in \mathcal{A}_0$ und $H_n \downarrow \emptyset$. Nach dem Lemma folgt $P_0(H_n) \downarrow 0$ und da $P(H_n) = P(A) - P_0(A'_n)$ folgt $P_0(A'_n) \uparrow P(A)$, wie gefordert.

Nach dem Existenzsatz existiert also genau ein Wahrscheinlichkeitsmaß P auf $(\Omega, \mathcal{A})$ so,dass $P = P_0$ auf $\mathcal{A}_k$ für jedes k. Daraus folgt auch die Unabhängigkeit der X_n auf diesem Wahrscheinlichkeitsraum und die Verteilungsaussage für P_n.

Kapitel 2

Martingaltheorie und Anwendungen

2.1 Erwartungen und bedingte Erwartungen

2.1.1 Erwartungswerte als optimale Approximation

Für eine integrierbare oder nichtnegative Zufallsvariable $X\colon (\Omega, \mathcal{A}, P) \to \mathbb{R}$ definieren wir den *Erwartungswert* als

$$\mathbb{E}X := \int X \, dP.$$

Der Erwartungswert ist der bezüglich P gemittelte Wert der Zufallsvariablen. Er ist auch die beste Approximation der Zufallsvariablen X durch eine Konstante in dem folgenden Sinn.

Proposition 2.1. *Sei X eine integrierbare Zufallsvariable und $a \in \mathbb{R}$, so gilt*

$$\mathbb{E}\big[(X-a)^2\big] = \mathrm{Var}(X) + \big(\mathbb{E}X - a\big)^2.$$

Insbesondere wird die mittlere quadratische Abweichung von X und einer konstanten Zufallsvariable a durch die Wahl $a = \mathbb{E}X$ minimiert.

Zum Beweis rechnen wir

$$\begin{aligned}\mathbb{E}[(X-a)^2] = \mathbb{E}\big[X^2 - 2aX + a^2\big] &= \mathbb{E}[X^2] - \big(\mathbb{E}X\big)^2 + \big(\mathbb{E}X\big)^2 - 2a\mathbb{E}X + a^2 \\ &= \mathrm{Var}(X) + \big(\mathbb{E}X - a\big)^2.\end{aligned}$$

Zur Berechnung des Erwartungswertes erinnern wir an zwei wichtige Fälle.

- Falls $X : \Omega \to \mathbb{R}$ nur endlich oder abzählbar viele Werte $a_1, a_2, \dots$ annimmt, gilt
$$\mathbb{E}X == \sum_n a_n P(X = a_n).$$
Dies gilt nach Definition, wenn X nur endlich viele nichtnegative Werte annimmt (d.h. X ist eine einfache Funktion) und lässt sich durch den Satz von der monotonen Konvergenz auf den Fall abzählbar vieler Werte erweitern. Der allgemeine Fall folgt dann durch Zerlegung in Positiv- und Negativteil.

- Falls die Verteilung von X eine Dichte f hat, das heißt
$$P(X \in A) = \int 1_A(x) f(x)\, dx,$$
wobei dx für die Integration bezüglich des Lebesguemaßes auf $\mathbb{R}$ steht, so gilt
$$\mathbb{E}X = \int x \cdot f(x)\, dx.$$
Dies folgt weil $\mathbb{E}h(X) = \int h(x) \cdot f(x)\, dx$ für Indikatorfunktionen gilt und mit Hilfe von Linearität und des Satzes von der monotonen Konvergenz auf nichtnegative Funktionen h ausgedehnt werden kann. Wir wenden dies dann auf $h^+(x) = x \vee 0$ und $h^-(x) = -(x \wedge 0)$ an und nutzen aus, dass $x = h^+(x) - h^-(x)$.

2.1.2 σ-Algebren und Filtrationen

Wir fragen nun wie *zusätzliche Informationen unsere Erwartungen in den Ausgang eines Zufallsexperiments verändern.*

Zusätzliche Informationen zu haben heißt, dass wir den Ausgang von bestimmten Ereignissen kennen. Das Mengensystem $\mathcal{F} \subseteq \mathcal{A}$ all dieser Ereignisse muss logischerweise den Axiomen einer σ-Algebra genügen, denn

- wir wissen immer, dass das sichere Ereginis Ω stattgefunden hat,

- wenn wir wissen, ob A stattgefunden hat, wissen wir auch ob A^c stattgefunden hat, und

- wenn wir von jedem Ereignis $A_1, A_2, \dots$ wissen ob sie stattgefunden haben, wissen wir auch ob *mindestens eines* der Eregnisse stattgefunden hat, d.h. ofb $\bigcup_i A_i$ stattgefunden hat.

In anderen Worten, Information über den Ausgang eines Zufallsexperiments wird durch eine σ-Algebra $\mathcal{F} \subseteq \mathcal{A}$ modelliert.

Beispiele (1) Sei $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}$ gegeben durch $\mathbb{P}\{i\} = 1/6$ für alle $i \in \Omega$, also unser Modell für einen Münzwurf. Die Information ob *eine gerade Zahl* geworfen wurde entspricht der σ-Algebra $\mathcal{F} = \{\emptyset, \{2, 4, 6\}, \{1, 3, 5\}, \Omega\}$.

(2) Seien $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ und $Y : (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$ Zufallsvariablen auf demselben Wahrscheinlichkeitsraum. Wenn wir die Werte von Y kennen, wissen wir ob die Ereignisse $\{Y \in A\}$, $A \in \mathcal{A}'$, stattgefunden haben und diese Information wird durch die von Y induzierte σ-Algebra $\sigma(Y) := \{\{Y \in A\} : A \in \mathcal{A}'\}$ beschrieben.

(3) Wenn $X = (X_n : n \in \mathbb{N})$ ein stochastischer Prozess ist, also eine Folge von Zufallsvariablen auf demselben Wahrscheinlichkeitsraum, wobei wir n als Zeitparameter interpretieren, so haben wir zur Zeit n Kenntnis über die Werte der Zufallsvariablen $X_1, X_2, \ldots, X_n$. Wir setzen $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$, die kleinste σ-Algebra, die die Zufallsvariablen $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ messbar macht. Diese σ-Algebra stellt die zur Zeit n über den Prozess verfügbare Information dar.

Im letzten Beispiel haben wir gesehen, dass ein stochastischer Prozess eine wachsende Folge von σ-Algebren $\mathcal{F}_n$ erzeugt, die den Zuwachs an Information in der Zeit modelliert. Diese Situation ist so wichtig, dass sie eine eigene Definition erfodert.

Definition Eine Folge $(\mathcal{F}_n : n \in \mathbb{N})$ von σ-Algebren mit $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{A}$, für alle $n \in \mathbb{N}$, heißt *Filtration*. Wie gesehen erzeugt ein stochastischer Prozess X stets eine Filtration $(\mathcal{F}_n : n \in \mathbb{N})$, die sogenannte *natürliche Filtration* des Prozesses.

2.1.3 Existenz der bedingten Erwartung

Zusätzliche Informationen verändern unsere Erwartungen in den Ausgang eines Zufallsexperiments. In unserem ersten Beispiel ist intuitiv der neue Erwartungswert des Würfelwurfes,

- wenn eine gerade Zahl gewürfelt wurde gleich dem Mittelwert über $2, 4, 6$ mit gleicher Wahrscheinlichkeit, also gleich vier,

- wenn eine ungerade Zahl gewürfelt wurde gleich dem Mittelwert über $1, 3, 5$ mit gleicher Wahrscheinlichkeit, also gleich drei.

Der *bedingte Erwartungswert* ist also abhängig von den erhaltenen Informationen und damit selbst eine Zufallsvariable, die bezüglich der diese Informationen beschreibenden σ-Algebra messbar ist.

Die Approximationseigenschaft des Erwartungswertes legt nun eine ähnlich Definition des bedingten Erwartungswertes nahe, die im Fall quadratisch integrierbarer Zufallsvariablen funktioniert.

Satz 9. *Sei $X\colon (\Omega, \mathcal{A}, P) \to \mathbb{R}$ eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$ (d.h. X ist quadratisch integrierbar) und $\mathcal{F} \subseteq \mathcal{A}$ eine σ-Algebra.*

(a) Es existiert eine messbare Zufallsvariable $Y\colon (\Omega, \mathcal{F}, P) \to \mathbb{R}$, die quadratisch integrierbar ist und die mittlere quadratische Abweichung $\mathbb{E}[(X-Y)^2]$ minimiert.

(b) Diese Zufallsvariable Y ist eindeutig in dem Sinne, dass zwei auf $(\Omega, \mathcal{F})$ messbare, die mittlere quadratische Abweichung minimierende Zufallsvariablen fast sicher übereinstimmen.

(c) Für alle quadratisch integrierbaren, messbaren Zufallsvariablen $Z\colon (\Omega, \mathcal{F}, P) \to \mathbb{R}$ gilt $\mathbb{E}[(X-Y)Z] = 0$.

Beweis: Dieser Satz ist eine Folgerung aus allgemeinen Eigenschaften von Hilberträumen. Der zugrundeliegende Hilbertraum ist der Raum

$$L^2(\Omega, \mathcal{A}, P) = \big\{X : (\Omega, \mathcal{A}) \to \mathbb{R} \text{ messbar mit } \mathbb{E}[X^2] < \infty\big\}$$

quadratisch integrierbarer Zufallsvariablen, wobei zwei Elemente X und Y dieses Raumes identifiziert werden wenn $P\{X = Y\} = 1$. Das Skalarproduct auf diesem Raum ist

$$\langle X, Y\rangle = \mathbb{E}\big[XY\big]$$

und daher ist die Norm $\|X\| = \sqrt{\mathbb{E}[X^2]}$. Als Hilbertraum ist $L^2(\Omega, \mathcal{A}, P)$ vollständig, das heißt das jede Cauchyfolge konvergiert. Das heißt explizit, dass wenn (X_n) eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ ist mit $\mathbb{E}[X_n^2] < \infty$ für alle n und $\lim_{k\to\infty} \sup_{n,m\geq k} \mathbb{E}[(X_n - X_m)^2] = 0$, so gibt es eine quadratisch integrierbare Zufallsvariable X auf $(\Omega, \mathcal{A}, P)$ so dass

$$\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0\,.$$

Das beweisen wir als Übungsaufgabe. Die mathematische Struktur des Hilbertraumes erlaubt die Definition von Projektionen auf Teilräume, wie das folgende Lemma zeigt.

Lemma 2.1. *Sei $V \subseteq L^2(\Omega, \mathcal{A}, P)$ ein vollständiger linearer Teilraum und $X \in L^2(\Omega, \mathcal{A}, P)$. Dann gibt es ein $Y \in V$, die* Projektion von X auf V *so dass*

- $\|X-Y\| = I := \inf\{\|X-W\| : W \in V\}$,
- $\langle X-Y, Z\rangle = 0$ *für alle* $Z \in V$.

Bemerkung: Sind $Y, Z \in V$ Projektionen von X auf V, so gilt

$$\|Y-Z\|^2 = \langle Y, Y-Z\rangle - \langle Z, Y-Z\rangle = \langle X, Y-Z\rangle - \langle X, Y-Z\rangle = 0,$$

also gilt $Y = Z$ fast sicher.

Beweis des Lemmas: Wir wählen eine Folge (Y_n) in V so dass

$$\|X - Y_n\| \to I.$$

Dann gilt

$$\begin{aligned}
\|X-Y_n\|^2 + \|X-Y_m\|^2 - 2\|X - \tfrac{1}{2}(Y_n+Y_m)\|^2 \\
&= \langle X-Y_n, X-Y_n\rangle + \langle X-Y_m, X-Y_m\rangle \\
&\qquad - 2\langle X - \tfrac{1}{2}(Y_n+Y_m), X - \tfrac{1}{2}(Y_n+Y_m)\rangle \\
&= \|Y_n\|^2 + \|Y_m\|^2 - \tfrac{1}{2}\|Y_n+Y_m\|^2 \\
&= 2\|\tfrac{1}{2}(Y_n-Y_m)\|^2 .
\end{aligned}$$

Da $\frac{1}{2}(Y_n + Y_m) \in V$ ist die linke Seite asymptotisch nichtpositiv und somit ist (Y_n) eine Cauchyfolge und somit aufgrund der Vollständigkeit von V konvergent gegen eine Zufallsvariable $Y \in V$. Nach der Dreiecksungleichung folgt

$$I \le \|X-Y\| \le \|X-Y_n\| + \|Y-Y_n\| \to I\,,$$

und somit $\|X-Y\| = I$. Für alle $Z \in V$ gilt $Y + tZ \in V$ für $t \in \mathbb{R}$ und damit auch

$$\|X-Y-tZ\|^2 \ge \|X-Y\|^2\,,$$

woraus folgt, dass $-2t\langle Z, X-Y\rangle + t^2\|Z\|^2 \ge 0$. Als Funktion von t hat dieser Ausdruck ein Minimum in $t = 0$ und somit verschwindende Ableitung, wie im Lemma behauptet.

Um den Beweis von Satz 9 abzuschliessen projizieren wir nun die Zufallsvariable $X \in L^2(\Omega, \mathcal{A}, P)$ auf den vollständigen Teilraum

$$V := L^2(\Omega, \mathcal{F}, P) \subseteq L^2(\Omega, \mathcal{A}, P)\,,$$

und erhalten Y. Da $Y \in V$ ist Y messbar auf $(\Omega, \mathcal{F}, P)$ und quadratisch integrierbar. Die Eigenschaften in (a) und (c) folgen direkt aus Lemma 2.1. Die in (b) behauptete Eindeutigkeit folgt aus der Eindeutigkeit der Projektion, in deren Beweis wir nur die Charakterisierung der Projektion als Minimum genutzt haben.

Die im Satz konstruierte Zufallvariable Y ist die *bedingte Erwartung von X gegeben $\mathcal{F}$*. Wir erweitern die Definition nun auf den Fall integrierbarer Zufallsvariablen X (beachte, dass jede quadratisch integrierbare Zufallsvariable integrierbar ist, aber nicht umgekehrt). Dabei ersetzen wir die Approximationseigenschaft in der vorherigen Definition durch eine Integralbedingung.

Satz 10. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subseteq \mathcal{A}$ eine σ-Algebra und $X: (\Omega, \mathcal{A}, P) \to \mathbb{R}$ eine integrierbare Zufallsvariable. Dann existiert eine integrierbare Zufallsvariable Y mit den folgenden Eigenschaften*

(1) $Y: (\Omega, \mathcal{F}, P) \to \mathbb{R}$ ist messbar,

(2) für alle Ereignisse $F \in \mathcal{F}$ gilt $\displaystyle\int 1_F Y \, dP = \int 1_F X \, dP$.

Diese Zufallsvariable Y ist eindeutig in dem Sinne, dass zwei Zufallsvariablen, die diese Bedingungen erfüllen, fast sicher übereinstimmen. Wir schreiben $Y = \mathbb{E}[X|\mathcal{F}]$ und nennen Y bedingte Erwartung von X gegeben $\mathcal{F}$.

Beweis: Wir beweisen den Satz in drei Schritten. Zunächst zeigen wir die Eindeutigkeit, dann prüfen wir, dass im Falle quadratisch integrierbarer Zufallsvariablen die Zufallsvariable Y aus Satz 9 unsere Bedingung erfüllt, und zuletzt erweitern wir die Existenzaussage durch Approximation vom Fall quadratisch integrierbarer auf den Fall integrierbarer Zufallsvariablen.

Für den Beweis der **Eindeutigkeit**, seien Z_1 und Z_2 Zufallsvariablen, die die Definition der bedingten Erwartung erfüllen. Wir nehmen an, dass $P\{Z_1 > Z_2\} > 0$. Da

$$0 < P\{Z_1 - Z_2 > 0\} = \lim_{n\to\infty} P\{Z_1 - Z_2 > 1/n\},$$

existiert $n \in \mathbb{N}$ so dass $P\{Z_1 - Z_2 > 1/n\} > 1/n$. Nach (1) liegt das Ereignis $F = \{Z_1 - Z_2 > 1/n\}$ in $\mathcal{F}$. Wir folgern, unter Benutzung von (2) im vorletzten Schritt, dass

$$\begin{aligned}\frac{1}{n^2} &\leq \frac{1}{n} P\{Z_1 - Z_2 > 1/n\} \leq \int 1_F (Z_1 - Z_2) \, dP \\ &= \int 1_F Z_1 \, dP - \int 1_F Z_2 \, dP = \int 1_F X \, dP - \int 1_F X \, dP = 0\,,\end{aligned}$$

ein Widerspruch. Folglich gilt $P\{Z_1 > Z_2\} = 0$ und durch Vertauschung von Z_1 und Z_2 auch $P\{Z_1 < Z_2\} = 0$, und somit $\mathbb{P}\{Z_1 = Z_2\} = 1$.

Im Fall **quadratisch integrierbarer** X wählen wir Y wie in Satz 9 und erinnern uns, dass

$$\langle X - Y, Z \rangle = 0 \text{ für alle } Z \in L^2(\Omega, \mathcal{F}, P).$$

Für $F \in \mathcal{F}$ wählen wir $Z = 1_F$ und erhalten

$$\int 1_F X \, dP = \langle X, 1_F \rangle = \langle Y, 1_F \rangle = \int 1_F Y \, dP \,.$$

Folglich ist Y eine bedingte Erwartung von X gegeben $\mathcal{F}$.

Es bleibt nun, dies auf **integrierbare Zufallsvariablen** X zu erweitern. Es genügt dabei $X \geq 0$ zu betrachten. Wähle $X_n = X \wedge n$ und beobachte, dass $X_n \in L^2(\Omega, \mathcal{A}, P)$ und dass es somit bedingte Erwartungen Y_n gibt. Aus Satz 9(c) angewandt auf $Z = 1_{Y_n < 0}$ sieht man, dass $Y_n \geq 0$ fast sicher. Zudem sieht man auf diese Weise auch leicht, dass Y_n wachsend ist. Somit existiert Y mit

$$Y = \lim_{n \to \infty} Y_n. \text{ fast sicher.}$$

$Y \colon (\Omega, \mathcal{F}, P) \to \mathbb{R}$ ist messbar und, da $X_n \uparrow X$ schliessen wir aus dem Satz von der monotonen Konvergenz, dass Y eine bedingte Erwartung von X gegeben $\mathcal{F}$ ist.

Berechnung des bedingten Erwartungswerts (Beispiele):

- Sei $\mathcal{F}$ eine σ-Algebra, die von einer endlichen oder abzählbaren Partition $A_1, A_2, \ldots \subseteq \Omega$ erzeugt wird und X eine Zufallsvariable so dass

  $$P(\{X \in A\} \cap A_k) = \int_A f_k(x) \, dx \qquad \text{für alle } A \in \mathcal{B}.$$

  Dann gilt

  $$\mathbb{E}[X|\mathcal{F}] = \sum_{k=1}^{\infty} \frac{1_{A_k}}{P(A_k)} \int x f(x) \, dx.$$

- Wir betrachten Zufallsvariable X und Y, die endlich viele verschiedene Werte $\{x_1, \ldots, x_n\}$ bzw. $\{y_1, \ldots, y_m\}$ mit positiver Wahrscheinlichkeit annehmen. Die gemeinsame Verteilung sei gegeben durch

  $$P(X = x_i, Y = y_j) = p_{ij} \text{ für } i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}.$$

  Betrachten wir die von Y erzeugte σ-Algebra $\mathcal{F}$, so gilt

  $$\mathbb{E}[X|\mathcal{F}] = \frac{\sum_{i=1}^n x_i p_{ij}}{\sum_{i=1}^n p_{ij}} \text{ auf der Menge } \{Y = y_j\}.$$

Beispiel: Wir betrachten als Beispiel wieder unsere Folge von Münzwürfen mit der zur uniformen Zufallsvariablen θ geprägten Münze. Sei Y_n die Anzahl der Köpfe unter den ersten n Würfen. Was ist der bedingte Erwartungswert von θ, wenn wir den Wert von Y_n beobachtet haben?

Antwort: Wir betrachten die von der Partition $\{Y_n = k\}$, $k = 0, \dots, n$, erzeugte σ-Algebra $\mathcal{F}$. Dann gilt auf der Menge $\{Y_n = k\}$, dass

$$\mathbb{E}[\theta|\mathcal{F}] = \frac{\binom{n}{k} \int \theta^{k+1}(1-\theta)^{n-k}\, d\theta}{\binom{n}{k} \int \theta^{k}(1-\theta)^{n-k}\, d\theta}.$$

Es gilt

$$\int \theta^{k+1}(1-\theta)^{n-k} d\theta = \frac{k+1}{n-k+1} \int \theta^{k}(1-\theta)^{n-k+1} d\theta$$

$$= \frac{k+1}{n-k+1} \int \theta^{k}(1-\theta)^{n-k} d\theta - \frac{k+1}{n-k+1} \int \theta^{k+1}(1-\theta)^{n-k} d\theta$$

woraus folgt

$$\int \theta^{k+1}(1-\theta)^{n-k} d\theta = \frac{k+1}{n+2} \int \theta^{k}(1-\theta)^{n-k} d\theta.$$

Damit ergibt sich auf $\{Y_n = k\}$ dass

$$\mathbb{E}[\theta|\mathcal{F}] = \frac{k+1}{n+2},$$

beziehungsweise im allgemeinen

$$\mathbb{E}[\theta|\mathcal{F}] = \frac{Y_n+1}{n+2}.$$

2.1.4 Eigenschaften der bedingten Erwartung

Wir beobachten zunächst zwei extreme Fälle von bedingten Erwartungen:

- **[Keine Information]**
  Wenn $X\colon (\Omega, \mathcal{A}, P) \to \mathbb{R}$ und $\mathcal{F} = \{\emptyset, \Omega\}$ so gilt $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}X$.

- **[Vollständige Information]**
  Wenn $X\colon (\Omega, \mathcal{A}, P) \to \mathbb{R}$ und $\mathcal{F} = \mathcal{A}$ so gilt $\mathbb{E}[X|\mathcal{F}] = X$.

Gegeben ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine σ-Algebra $\mathcal{F} \subseteq \mathcal{A}$ definieren wir eine Abbildung $P[\cdot|\mathcal{F}]\colon \mathcal{A} \to [0,1]$ durch $P[A|\mathcal{F}] = \mathbb{E}[1_A|\mathcal{F}]$. Die Axiome eines Wahrscheinlichkeitsmasses gelten dann für jede Wahl im $\forall$ Quantor jeweils fast sicher. Da es aber in der Regel überabzählbar viele solche Wahlen gibt, können wir nicht folgern, dass $P[\cdot|\mathcal{F}]$ fast sicher ein Wahrscheinlichkeitsmaß ist. Dennoch lassen sich viele Eigenschaften des Erwartungswertes auch auf bedingte Erwartungen übertragen.

Proposition 2.2. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subseteq \mathcal{A}$ eine σ-Algebra und X, Y integrierbare Zufallsvariable. Dann hat die bedingte Erwartung die folgenden Eigenschaften:*

Linearität *Für $a, b \in \mathbb{R}$ fast sicher*

$$\mathbb{E}\big[aX + bY \mid \mathcal{F}\big] = a\mathbb{E}[X \mid \mathcal{F}] + b\mathbb{E}[Y \mid \mathcal{F}]\,.$$

Positivität *Wenn $X \geq 0$, so gilt $\mathbb{E}[X|\mathcal{F}] \geq 0$ fast sicher.*

Monotone Konvergenz *Wenn $0 \leq X_n \uparrow X$, so folgt*

$$\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}] \textit{ fast sicher.}$$

Lemma von Fatou *Wenn $0 \leq X_n$ und $\mathbb{E}[X_n|\mathcal{F}] < \infty$, so gilt*

$$\mathbb{E}[\liminf_{n\to\infty} X_n \mid \mathcal{F}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{F}] \textit{ fast sicher.}$$

Dominierte Konvergenz *Existiert eine integrierbare Zufallsvariable Z mit $|X_n| \leq Z$ für alle n, und gilt $X_n \to X$ fast sicher, so folgt*

$$\mathbb{E}[X_n|\mathcal{F}] \to \mathbb{E}[X|\mathcal{F}] \textit{ fast sicher.}$$

Den Beweis einiger dieser Aussagen empfehle ich als Übungsaufgabe. Wir betrachten hier die folgenden Eigenschaften der bedingten Erwartung.

Proposition 2.3. *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subseteq \mathcal{A}$ eine σ-Algebra und X eine integrierbare Zufallsvariable. Dann hat die bedingte Erwartung $\mathbb{E}[X|\mathcal{F}]$ die folgenden Eigenschaften:*

Turmeigenschaft *Sei $\mathcal{G} \subseteq \mathcal{F}$ eine σ-Algebra, so gilt*

$$\mathbb{E}\Big[\mathbb{E}\big[X|\mathcal{F}\big]\Big|\mathcal{G}\Big] = \mathbb{E}[X|\mathcal{G}] \textit{ fast sicher.}$$

Bekanntes Rausziehen *Wenn $Z\colon (\Omega, \mathcal{F}, P) \to \mathbb{R}$ messbar und beschränkt ist, so gilt*

$$\mathbb{E}[ZX|\mathcal{F}] = Z\mathbb{E}[X|\mathcal{F}] \textit{ fast sicher.}$$

Insbesondere gilt, wenn X selbst $(\Omega, \mathcal{F})$ messbar ist, so ist $\mathbb{E}[X|\mathcal{F}] = X$ fast sicher.

Unabhängigkeit *Ist X unabhängig von den Ereignissen in $\mathcal{F}$, so gilt*

$$\mathbb{E}\big[X \bigm| \mathcal{F}\big] = \mathbb{E}[X] \textit{ fast sicher.}$$

Beweis: Wir dürfen annehmen, dass $X \geq 0$ in allen Beweisteilen.

Für die **Turmeigenschaft** beachte, dass die linke Seite der Gleichung bezüglich $(\Omega, \mathcal{G})$ messbar ist und, für alle $G \in \mathcal{G} \subseteq \mathcal{F}$,

$$\int 1_G \, \mathbb{E}\Big[\mathbb{E}\big[X|\mathcal{F}\big]\Big|\mathcal{G}\Big] \, dP = \int 1_G \, \mathbb{E}\big[X|\mathcal{F}\big] \, dP = \int 1_G \, X \, dP \, .$$

Also erfüllt der Integrand der linken Seite alle Bedingungen für die bedingte Erwartung $\mathbb{E}[X|\mathcal{G}]$.

Für **Bekanntes Rausziehen** beobachte, dass $Z\mathbb{E}[X|\mathcal{F}]$ messbar und integrierbar auf $(\Omega, \mathcal{F}, P)$ ist. Es bleibt zu zeigen, dass für $F \in \mathcal{F}$,

$$\int 1_F Z X \, dP = \int 1_F Z \mathbb{E}[X|\mathcal{F}] \, dP \, ,$$

dann folgt, dass $Z\mathbb{E}[X|\mathcal{F}]$ bedingte Erwartung von ZX gegeben $\mathcal{F}$ ist. Die Gleichung gilt für $Z = 1_H$, $H \in \mathcal{F}$ nach Definition und folgt für beschränkte, messbare Z durch Anwendung des Satzes von den monotonen Klassen.

Der Spezialfall gilt, da unter den gegebenen Voraussetzungen X selbst die Bedingungen einer bedingten Erwartung von X gegeben $\mathcal{F}$ erfüllt.

Für die **Unabhängigkeitseigenschaft** beobachte, dass $\mathbb{E}[X]$ konstant und somit messbar bezüglich $\mathcal{F}$ ist. Wenn X unabhängig von den Ereignissen in $\mathcal{F}$ ist, so ist X auch unabhängig von jeder messbaren Zufallsvariablen $Y \colon (\Omega, \mathcal{F}, P) \to \mathbb{R}$, und somit folgt $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Insbesondere gilt

$$\int 1_F X \, dP = \mathbb{E}\big[1_F X\big] = \mathbb{E}\big[1_F\big] \, \mathbb{E}\big[X\big] = P(F)\mathbb{E}[X] = \int 1_F \mathbb{E}[X] \, dP \, .$$

Also ist die konstante Funktion $\mathbb{E}[X]$ bedingte Erwartung von X gegeben $\mathcal{F}$.

2.2 Martingale

2.2.1 Martingale und Supermartingale

Martingale sind stochastische Prozesse, die fairen Spielen entsprechen. Der Begriff ist aus einer Strategie beim Glücksspiel abgeleitet.

Definition: Sei $(X_n : n \geq 0)$ ein stochastischer Prozess, also eine Folge von Zufallsvariablen definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, und $(\mathcal{F}_n : n \geq 0)$ eine Filtration, also eine wachsende Folge

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{A}$$

von σ-Algebren. Wir sagen dass (X_n) an die Filtration **adaptiert** ist, wenn

$$X_n\colon (\Omega, \mathcal{F}_n) \to \mathbb{R} \text{ messbar ist, für alle } n \geq 0.$$

In den meisten Fällen betrachten wir die natürliche Filtration des Prozesses. Offensichtlich ist jeder Prozess an seine natürliche Filtration adaptiert.

Definition: Ein Prozess $(X_n\colon n \geq 0)$ ist ein **Martingal** bezüglich der Filtration $(\mathcal{F}_n\colon n \geq 0)$, wenn

- er an die Filtration adaptiert ist,
- X_n integrierbar ist für alle n, und
- $\mathbb{E}[X_n \,|\, \mathcal{F}_{n-1}] = X_{n-1}$ fast sicher, für alle $n \geq 1$.

Gilt nur $\leq$ in der letzten Gleichung, so ist (X_n) ein **Supermartingal**, gilt $\geq$ so ist (X_n) ein **Submartingal**.

Wir betrachten gleich ein paar interessante Beispiele.

(1) Irrfahrten. Seien $X_1, X_2, \ldots$ unabhängige Zufallsvariable mit $\mathbb{E}|X_n| < \infty$ und $\mathbb{E}X_n = 0$. Setze $S_n = \sum_{k=1}^n X_k$ und $(\mathcal{F}_n\colon n \geq 0)$ die natürliche Filtration des Prozesses $(S_n\colon n \geq 0)$. Dies ist auch die natürliche Filtration der $(X_n\colon n \geq 1)$. Dann gilt für $n \geq 1$,

$$\begin{aligned}
\mathbb{E}[S_n \,|\, \mathcal{F}_{n-1}] &= \mathbb{E}\Big[\sum_{k=1}^{n-1} X_k + X_n \,\Big|\, \mathcal{F}_{n-1}\Big] \\
&= \sum_{k=1}^{n-1} X_k + \mathbb{E}X_n = S_{n-1} \text{ fast sicher,}
\end{aligned}$$

also ist die Irrfahrt $(S_n\colon n \geq 0)$ ein Martingal.

(2) Zufällige Produkte. Seien $X_1, X_2, \ldots$ unabhängige nichnegative Zufallsvariable mit $\mathbb{E}X_n = 1$. Sei

$$M_n = \prod_{k=1}^n X_k$$

und $(\mathcal{F}_n\colon n \geq 0)$ die natürliche Filtration des Prozesses $(M(n)\colon n \geq 0)$. Dann gilt

$$\mathbb{E}\big[M_n \,\big|\, \mathcal{F}_{n-1}\big] = M_{n-1}\mathbb{E}\big[X_n \big|\, \mathcal{F}_{n-1}\big] = M_{n-1}\mathbb{E}X_n = M_{n-1}\,.$$

Also ist $(M(n)\colon n \geq 0)$ ein Martingal.

(3) Akkumulation von Information über eine Zufallsvariable. Sei $(\mathcal{F}_n)$ eine beliebige Filtration und X eine integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Setze $X_n = \mathbb{E}[X|\mathcal{F}_n]$, was wir als beste Approximation an X gegeben die Information zur Zeit n interpretieren. Die Martingaleigenschaft von $(X_n \colon n \geq 0)$ folgt aus der Turmeigenschaft der bedingten Erwartung

$$\mathbb{E}\big[X_n \,\big|\, \mathcal{F}_{n-1}\big] = \mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{F}_n] \,\big|\, \mathcal{F}_{n-1}\big] = \mathbb{E}[X \,|\, \mathcal{F}_{n-1}] = X_{n-1}\,.$$

(4) Galton Watson Prozesse. Sei $p = (p_0, p_1, p_2, \ldots)$ eine Wahrscheinlichkeitsfolge. Der *Galton Watson Prozess* $(X_n \colon n \geq 0)$ mit Nachkommensverteilung p ist ein wie folgt konstruierter Verzweigungsprozess. Seien Y_k^n, $k, n \geq 1$, unabhängige Zufallsvariable mit $P(Y_k^n = i) = p_i$, und definiere X_n rekursiv durch $X_0 = 1$ und

$$X_{n+1} = \sum_{k=1}^{X_n} Y_k^{n+1} \text{ für alle } n \geq 0.$$

Wir nehmen an, dass

$$\mu := \sum_{k=0}^{\infty} k p_k,$$

die erwartete Anzahl der Nachkommen, endlich ist. Setze

$$\mathcal{F}_n = \sigma(Y_k^m \,:\, m \leq n, k \geq 1)$$

und bebachte, dass (X_n) an die Filtration $(\mathcal{F}_n)$ adaptiert ist. Dann gilt

$$\mathbb{E}\big[X_{n+1} \,|\, \mathcal{F}_n\big] = \sum_{k=1}^{\infty} \mathbb{E}\big[1_{k \leq X_n} Y_k^{n+1} \,|\, \mathcal{F}_n\big] = \sum_{k=1}^{\infty} 1_{k \leq X_n}\, \mathbb{E}\big[Y_k^{n+1}\big] = \mu\, X_n.$$

wo wir bekanntes rausgezogen haben und Unabhängigkeit genutzt haben. Folglich ist (X_n) ein Martingal genau dann wenn $\mu = 1$, und ein Supermartingal wenn $\mu \leq 1$.

(5) Polyas Urnenschema. Zur Zeit 0 entält eine Urne eine schwarze und eine weiße Kugel. Zu jedem Zeitpunkt $n = 1, 2, \ldots$ wird eine Kugel zufällig aus der Urne gezogen und zusammen mit einer weiteren Kugel derselben Farbe zurückgelegt. Zur Zeit n sind also $n + 2$ Kugeln in der Urne, davon $B_n + 1$ schwarze, wobei B_n die Zahl der zur Zeit n gezogenen schwarzen Kugeln ist. Sei

$$M_n = \frac{B_n + 1}{n + 2}$$

der Anteil der schwarzen Kugeln in der Urne zur Zeit n. Wir zeigen, dass $(M_n \colon n \geq 0)$ ein Martingal ist bezüglich der natürlichen Filtration $(\mathcal{F}_n \colon n \geq 0)$ von $(B_n \colon n \geq 0)$.

Seien $(X_n\colon n \geq 0)$ Zufallsvariable, wobei X_n den Wert eins annimmt, wenn im $n$ten Schritt eine schwarze Kugel gezogen wird, ansonsten ist $X_n = 0$. Dann ist $B_n = X_1 + \cdots + X_n$. Es gilt $P(X_1 = 1) = P(X_1 = 0) = 1/2$ und

$$P\big(X_n = 1 \,\big|\, B_{n-1} = k\big) = \frac{k+1}{n+1} \text{ for } 0 \leq k \leq n-1, n \geq 2.$$

Für die bedingte Erwartung heißt das, dass

$$\mathbb{E}\big[X_n \,\big|\, \mathcal{F}_{n-1}\big] = \frac{B_{n-1}+1}{n+1} \text{ für } n \geq 2.$$

Für $n \geq 1$ folgt

$$\begin{aligned}
\mathbb{E}\big[M_n \,|\, \mathcal{F}_{n-1}\big] &= \mathbb{E}\Big[\frac{B_{n-1}+X_n+1}{n+2} \,\Big|\, \mathcal{F}_{n-1}\Big] \\
&= \frac{B_{n-1}+1}{n+2} + \mathbb{E}\Big[\frac{X_n}{n+2} \,\Big|\, \mathcal{F}_{n-1}\Big] \\
&= \frac{B_{n-1}+1}{n+2} + \frac{B_{n-1}+1}{(n+1)(n+2)} = \frac{B_{n-1}+1}{n+1} = M_{n-1},
\end{aligned}$$

womit die Martingaleigenschaft gezeigt wäre.

Wir ziehen nun einige Konsequenzen aus der Definition. Für jedes Martingal (X_n) folgt aus der Turmeigenschaft, dass für $m < n$,

$$\mathbb{E}[X_n|\mathcal{F}_m] = \mathbb{E}[\mathbb{E}[X_n|\mathcal{F}_{n-1}] \,|\, \mathcal{F}_m] = \mathbb{E}[X_{n-1}|\mathcal{F}_m] = \ldots = X_m\,.$$

Nimmt man den Erwartungswert, so folgt

$$\mathbb{E}[X_n] = \mathbb{E}[X_0] \text{ für alle } n.$$

Es folgt aus der Definition dass, für jedes Martingal $(X_n\colon n \geq 0)$,

$$\mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] = 0\,. \tag{2.1}$$

Interpretiert man $(X_n\colon n \geq 0)$ als das Kapital eines Spielers zur Zeit n heißt das, dass das Spiel fair ist, denn der erwartete Profit ist stets null. Im Supermartingalfall ist der ewartete Profit negativ, und das Spiel ist unvorteilhaft. Wir verallgemeinern diese Beobachtung jetzt weiter.

Sei $(C_n : n \geq 1)$ der Einsatz im $n$ten Spiel. Ein Spieler wählt seinen Einsatz C_n auf der Basis des bisherigen Speilverlaufs, das heißt C_n muss messbar bezüglich $\mathcal{F}_{n-1}$ sein.

Definition: Ein Prozess $(C_n : n \geq 1)$ heißt *prävisibel* wenn C_n messbar bezüglich $\mathcal{F}_{n-1}$ ist.

Der Gewinn im $n$ten Spiel ist dann $C_n(X_n - X_{n-1})$ und der Gesamtgewinn nach n Spielen ist

$$Y_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}) =: (C \bullet X)_n \,.$$

Der Prozess $((C \bullet X)_n \colon n \geq 0))$ heißt **Martingaltransformation** von (X_n) durch (C_n). Die Frage ist nun: Kann man die Einsätze (C_n) so wählen, dass der erwartete Gesamtgewinn zu einer festen Zeit positiv ist? Leider ist die Antwort auf diese Frage negativ.

Proposition 2.4 (Man kann das System nicht besiegen)**.** *Sei $(C_n \colon n \geq 1)$ ein beschränkter, prävisibler Process, d.h. es existiert $C > 0$ so dass $|C_n(\omega)| \leq C$ für alle $n \geq 1$ und $\omega \in \Omega$. Ist dann $(X_n \colon n \geq 0)$ ein Martingal, so ist auch $((C \bullet X)_n \colon n \geq 0)$ ein Martingal. Ausserdem gilt $(C \bullet X)_0 = 0$ und somit $\mathbb{E}[(C \bullet X)_n] = 0$ für alle $n \geq 0$.*

Beweis: $Y_n := (C \bullet X)_n$ ist integrierbar, da C_n beschränkt ist und nach Definition ist Y_n messbar bezüglich $\mathcal{F}_n$. Wir berechnen

$$\begin{aligned}
\mathbb{E}[Y_n|\mathcal{F}_{n-1}] &= \mathbb{E}\Big[\sum_{k=1}^{n} C_k(X_k - X_{k-1})\Big|\mathcal{F}_{n-1}\Big] \\
&= \sum_{k=1}^{n-1} C_k(X_k - X_{k-1}) + C_n\mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] = Y_{n-1} \,.
\end{aligned}$$

Das komplettiert den Beweis.

Bemerkungen:

- Ist $C_n \geq 0$ und (X_n) ein Supermartingal, so ist auch $((C \bullet X)_n)$ ein Supermartingal.
- Definiert man Martingale in stetiger Zeit, so stellt das *stochastische Integral* ein Analogon zur Martingaltransformation dar. Stochastische Integration ist eine wichtiges Thema in der Wahrscheinlichkeitstheorie, siehe WT2.

Beispiel: Sei Y_n die Anzahl der Male in denen Kopf gewürfelt wurde unter den ersten n Würfen mit der nach einer uniform verteilten Zufallsvariablen θ geprägten Münze. Zeige, dass

$$M_n = \frac{Y_n + 1}{n + 2}$$

ein Martingal definiert.

Antwort: Es gilt auf $\{Y_n = k\}$ dass

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = (k+1)\mathbb{E}[X_{n+1} = \text{Kopf}|\mathcal{F}_n] + k\mathbb{E}[X_{n+1} = \text{Zahl}|\mathcal{F}_n]$$
$$= \frac{\binom{n}{k}\int \theta^k(1-\theta)^{n-k}((k+1)\theta + k(1-\theta))d\theta}{\binom{n}{k}\int \theta^k(1-\theta)^{n-k}d\theta}$$

Es gilt

$$\int \theta^k(1-\theta)^{n-k}((k+1)\theta + k(1-\theta))d\theta = \int \theta^k(1-\theta)^{n-k}(\theta + k)d\theta$$

$$= \int \theta^{k+1}(1-\theta)^{n-k}d\theta + \int k\theta^k(1-\theta)^{n-k}d\theta.$$

und (siehe letztes Kapitel)

$$\int \theta^{k+1}(1-\theta)^{n-k}d\theta = \frac{k+1}{n+2}\int \theta^k(1-\theta)^{n-k}d\theta.$$

Damit ergibt sich

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = \frac{k+1}{n+2} + k$$

und somit auf $\{Y_n = k\} = \{M_n = \frac{k+1}{n+2}\}$

$$\mathbb{E}\big[M_{n+1}\big|\mathcal{F}_n\big] = \frac{k+1}{(n+2)(n+3)} + \frac{k+1}{n+3}$$
$$= \frac{k+1+(k+1)(n+2)}{(n+2)(n+3)} = \frac{k+1}{n+2}.$$

2.2.2 Der Stoppsatz von Doob

Wir betrachten nun eine Klasse von zufälligen Zeiten, die sogenannten Stoppzeiten. Intuitiv heißt eine zufällige Zeit Stoppzeit für einen stochastischen Prozess, wenn Kenntnis des Prozesses zur Zeit n ausreicht, um zu entscheiden, ob die Stoppzeit zur Zeit n eingetreten ist, oder nicht. Die formale Definition benutzt eine Filtration, um die fortschreitenden Erkenntnisse über den Verlauf eines Prozesses zu modellieren.

Definition: Sei $(\mathcal{F}_n \colon n \in \mathbb{N}_0)$ eine Filtration eines Wahrscheinlichkeitsraumes $(\Omega, \mathcal{A}, P)$. Eine Zufallsvariable $T : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ heißt *Stoppzeit* wenn

$$\{T \le n\} = \{\omega \,:\, T(\omega) \le n\} \in \mathcal{F}_n \text{ für alle } n \in \mathbb{N}_0\,.$$

Eine äquivalente Bedingung ist $\{T = n\} \in \mathcal{F}_n$ für alle $n \in \mathbb{N}_0$. Dies folgt, da die erste Definition impliziert, dass $\{T = n\} = \{T \le n\} \setminus \{T \le n-1\} \in \mathcal{F}_n$, und die zweite Definition impliziert, dass $\{T \le n\} = \bigcup_{k=0}^n \{T = k\} \in \mathcal{F}_n$.

Interpretiert man das Martingal (X_n) als faires Spiel, so sind die Stoppzeiten diejenigen zufälligen Zeiten, zu denen man ein Spiel beenden kann (und den Gewinn oder Verlust erhalten kann). Wenn wir die durch eine Stoppzeit T gegebene Strategie verfolgt, und bis zur Stoppzeit eine Geldeinheit pro Martingaleinheit einsetzt, so ist der Einsatz

$$C_n = 1_{\{n \le T\}} = 1 - 1_{\{T \le n-1\}}$$

messbar bezüglich $\mathcal{F}_{n-1}$ und somit (C_n) prävisibel. Die Gewinne werden durch den Prozess

$$(C \bullet X)_n = \sum_{k=1}^n C_k (X_k - X_{k-1}) = X_{T \wedge n} - X_0$$

gegeben. Definiere $(X_n^T \colon n \in \mathbb{N}_0)$, den *zur Zeit T gestoppten Prozess* (X_n), durch

$$X_n^T(\omega) = X_{T(\omega) \wedge n}(\omega)\,.$$

Dann ist $(X_n^T - X_0)$ die Martingaltransformation von (X_n) durch den Prozess C. Proposition 2.4 kann angewandt werden und ergibt:

Proposition 2.5 (Elementarer Stoppsatz)**.** *Ist (X_n) ein Martingal und T eine Stoppzeit, dann ist auch (X_n^T) ein Martingal. Insbesondere gilt*

$$\mathbb{E}[X_{T \wedge n}] = \mathbb{E}[X_0] \text{ für alle } n\,.$$

Ist X ein Supermartingal, so ist auch (X_n^T) eines und es gilt daher

$$\mathbb{E}[X_{T \wedge n}] \le \mathbb{E}[X_0] \text{ für alle } n \in \mathbb{N}_0\,.$$

Beispiel: Wir betrachten eine *einfache symmetrische Irrfahrt*. Dazu seien $Y_1, Y_2, \ldots$ unabhängige Zufallsvariable mit Verteilung

$$P\{Y_n = 1\} = P\{Y_n = -1\} = \tfrac{1}{2}\,,$$

und

$$X_n = \sum_{k=1}^n Y_k\,.$$

Wir haben bereits gesehen, dass (X_n) ein Martingal (bezüglich der natürlichen Filtration) ist. Sei

$$T = \inf\{n \ge 0 \,:\, X_n = 1\}\,.$$

Offensichtlich ist T eine Stoppzeit. Eine Übungsaufgabe zeigt, dass $P\{T < \infty\} = 1$. Unsere Proposition sagt, dass

$$\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0] \text{ für alle } n.$$

Da aber $T < \infty$ fast sicher, gilt $X_T = 1$ fast sicher und somit

$$1 = \mathbb{E}X_T \neq \mathbb{E}X_0 = 0\,.$$

Man kann also den Grenzwert $n \to \infty$ auf der linken Seite der vorletzten Gleichung nicht mit dem Integral vertauschen. In gewisser Hinsicht zeigt dieses Besipiel, dass man das System besiegen kann, wenn man einen unendliche Menge Zeit und Kapital hat.

Das Beispiel ermuntert uns, nach Kriterien zu suchen, wann $\mathbb{E}X_T = \mathbb{E}X_0$. Dies ist eine Frage der Vertauschung von Integral und Grenzwert und kann mit Hilfe unserer Vertauschungssätze geklärt werden.

Satz 11 (Stoppsatz von Doob)**.** *Sei T eine Stoppzeit und (X_n) ein Martingal. Dann ist X_T integrierbar und*

$$\mathbb{E}[X_T] = \mathbb{E}[X_0]\,,$$

wenn eine der folgenden Bedingungen gilt:

(1) *T ist beschränkt (d.h. es existiert N so dass $T(\omega) < N$ für alle ω),*

(2) *T ist fast sicher endlich und (X_n) beschränkt, (d.h. es existiert K so dass $|X_n(\omega)| < K$ für alle n und ω),*

(3) *$\mathbb{E}[T] < \infty$ und es existiert $K > 0$ so dass, für alle n und ω, $|X_n(\omega) - X_{n-1}(\omega)| \leq K$.*

Ist (X_n) ein Supermartingal und eine der vorherigen Bedingungen oder die Bedingung

(4) *X is nichtnegativ und T fast sicher endlich.*

gilt, so ist X_T integrierbar und $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.

Beweis: Wir nehmen an, dass (X_n) ein Supermartingal ist. Dann ist auch $(X_{T\wedge n})$ ein Supermartingal und, insbesondere, integrierbar mit

$$\mathbb{E}[X_{T\wedge n} - X_0] \leq 0\,.$$

Für (1) folgt das Ergebnis durch die Wahl $n = N$. Für (2) betrachte $n \to \infty$ und benutze dominierte Konvergenz. Für (3) beobachten wir, dass

$$|X_{T\wedge n} - X_0| = \Big|\sum_{k=1}^{T\wedge n}(X_k - X_{k-1})\Big| \leq KT\,.$$

Nach Annahme ist KT integrierbar und wir benutzen erneut dominierte Konvergenz. Für (4) benutzen wir das Lemma von Fatou wie folgt

$$\mathbb{E}[X_T] = \mathbb{E}[\liminf_{n\to\infty} X_{T\wedge n}] \leq \liminf_{n\to\infty} \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]\,.$$

Die Aussage über Martingale folgt durch Anwendung auf die Supermartingale (X_n) und $(-X_n)$.

Bei der Anwendung auf unser Beispiel müssen alle Bedingungen versagen. Ein Blick auf Bedingung (3) ergibt folgende interessante Aussage.

Korollar 2.1. *Für eine einfache Irrfahrt ist die erwartete erste Treffzeit von Level 1 unendlich.*

2.2.3 Anwendungen des Stoppsatzes

Erstmaliges Auftreten eines Musters in unabhängigen Folgen

Aufgabe: Ein fairer Würfel wird wiederholt geworfen. Was ist der erwartete erste Zeitpunkt an dem drei aufeinanderfolgende Sechsen geworfen wurden?

Lösung: Seien $X_1, X_2, \ldots$ die Augenzahlen der Würfe, das heißt eine Folge von unabhängigen Zufallsvariablen mit $P(X_i = k) = 1/6$ für alle $k \in \{1, \ldots, 6\}$. Setze $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Sei T der erste Zeitpunkt an dem drei aufeinanderfolgende Sechsen beobachtet werden

$$T = \min\{n \geq 3 \colon X_n = X_{n-1} = X_{n-2} = 6\}.$$

T ist eine Stoppzeit und wir suchen $\mathbb{E}[T]$.

Wir argumentieren zunächst, dass $\mathbb{E}[T] < \infty$. Für das Ereignis $\{T = k\}$ muss notwendigerweise in jedem Tupel $(X_{3m+1}, X_{3m+2}, X_{3m+3})$ für $3m+3 < k$ eine von 6 verschiedene Zahl auftreten. Die Wahrscheinlichkeit hierfür ist $1 - 1/6^3 = 215/216$ und die Ereignisse sind unabängig für jedes m. Folglich gilt

$$P(T = k) \leq \left(\frac{215}{216}\right)^{\frac{k-3}{3}-1},$$

und somit ist $\mathbb{E}[T] = \sum_{k=1}^{\infty} kP(T = k)$ konvergent. Dieses Argument liefert nur eine grobe obere Schranke und versetzt uns nicht in die Lage $\mathbb{E}[T]$ auszurechnen.

Wir machen das folgende Gedankenexperiment. Vor jedem Zeitpunkt n erscheint ein Spieler und wettet einen Taler, dass der nächste Wurf eine 6 ergibt. Wenn er verliert geht er wieder, aber wenn er gewinnt erhält er 6 Taler, die er alle auf eine 6 im $(n+1)$ten Wurf setzt. Wenn er verliert geht er wieder, aber wenn er erneut gewinnt erhält er 36 Taler, die er alle auf eine 6 im $(n+2)$ten Wurf setzt, und so weiter. Dieses Spiel ist fair, so dass der erwartete totale Gewinn zu jedem Zeitpunkt $n \geq 3$ gleich dem Einsatz aller Spieler zu diesem Zeitpunkt, also gleich n Taler, ist. Da T eine Stoppzeit ist, die Bedingung (3) im Doobschen Stoppsatz erfüllt, gilt dies auch zur Zeit T und somit gilt

$$\mathbb{E}[T] = 6 + 6^2 + 6^3 = 258.$$

In der Tat, $\mathbb{E}[T]$ ist der erwartete Einsatz aller Spieler zur Zeit T, und zu diesem Zeitpunkt hat der letzte Spieler 6 Taler, der vorletzte 36 Taler und der drittletzte 216 Taler gewonnen. Alle anderen Spieler haben ihren Einsatz verloren.

Um diesen Ansatz rigoros zu machen setzten wir S_n gleich dem Einsatz aller Spieler zur Zeit n (also $S_n = 1 + 6 + \cdots + 6^k$ wenn wir zur Zeit n in einem Lauf von k Sechsen sind) und $M_n = S_n - n$, insbesondere $M_0 = 1$. Dann ist (M_n) ein Martingal, denn

$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = (5/6)(1-(n+1)) + (1/6)(6S_n + 1 - (n+1)) = S_n - n = M_n.$$

Wir betrachten das gestoppte Martingal M^T. Da $\mathbb{E}[T] < \infty$ und $|M_n^T - M_{n-1}^T| \leq 260$, folgt nach dem Doobschen Stoppsatz, Teil (3), dass

$$1 = \mathbb{E}[M_0] = \mathbb{E}[M_T] = 1 + 6 + 6^2 + 6^3 - \mathbb{E}[T],$$

woraus $\mathbb{E}[T] = 258$ folgt.

Bemerkung: Dieses Argument lässt sich natürlich auch auf das Auftreten beliebiger anderer Muster in u.i.v. Folgen übertragen, siehe Übungsaufgaben.

Das Ruinproblem

Seien $X_1, X_2, \ldots$ unabhängig, identisch verteilte Zufallsvariable mit

$$P(X_i = 1) = p, P(X_i = -1) = q, \text{ wobei } 0 < p = 1 - q < 1 \text{ und } p > q.$$

Setze $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ und wähle ganze Zahlen $0 < a < b$. Sei

$$S_n = a + X_1 + \cdots + X_n,$$

was das Kapital eines Spielers darstellt, der ausgehend von einem Startkapital a ein vorteilhaftes Spiel spielt. Er spielt solange, bis er entweder Bankrott geht, oder er das Zielkapital b erreicht. Die entsprechnde Stoppzeit ist also

$$T = \inf\{n \in \mathbb{N} : S_n = 0 \text{ oder } S_n = b\}.$$

Aufgabe: Berechne die Ruinwahrscheinlichkeit $P(S_T = 0)$, die erwartete Spieldauer $\mathbb{E}[T]$ und das erwartetete Endkapital $\mathbb{E}[S_T]$.

Lösung: Wir zeigen zunächst, dass $\mathbb{E}T < \infty$. Das ist einfach, denn $\{T = k\}$ impliziert, dass jedes Tupel $(X_{nb+1}, \ldots, X_{(n+1)b})$ mit $(n+1)b < k$ mindestens einen Schritt nach unten enthält. Also fällt $P(T = k)$ exponentiell schnell ab, und somit ist T integrierbar. Wir zeigen in den Übungen, dass

$$M_n = \left(\frac{q}{p}\right)^{S_n} \text{ und } N_n = S_n - n(p - q)$$

Martingale definieren. Anwendung des Doobschen Stoppsatzes mit Bedingung (3) auf (M_n) liefert, da $|M_n - M_{n+1}| \leq 1$, dass

$$\left(\frac{q}{p}\right)^a = \mathbb{E}[M_0] = \mathbb{E}[M_T] = P(S_T = 0) + \left(\frac{q}{p}\right)^b P(S_T = b).$$

Beachte $P(S_T = b) = 1 - P(S_T = 0)$ und daher

$$P(S_T = 0) = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}.$$

Daraus schliessen wir

$$\mathbb{E}[S_T] = bP(S_T = b) = b \times \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^b}.$$

Anwendung von Satz 11 (3) auf (N_n), unter Beachtung von $|N_n - N_{n+1}| \leq 1 + p + q$, liefert

$$a = \mathbb{E}[N_0] = \mathbb{E}[N_T] = \mathbb{E}[S_T - T(p - q)] = \mathbb{E}[S_T] - \mathbb{E}[T](p - q),$$

woraus folgt, dass

$$\mathbb{E}[T] = \frac{1}{p - q}(\mathbb{E}[S_T] - a) = \frac{b}{p - q} \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^b} - \frac{a}{p - q}.$$

2.3 Martingalkonvergenzsätze

2.3.1 Der Martingalkonvergenzsatz von Doob

Der Martingalkonvergenzsatz gibt eine hinreichende Bedingung für die Konvergenz eines Martingals (oder Supermartingals) gegen eine Grenzvariable. Zur Formulierung dieser Bedingung nennen wir eine Folge (X_n) von Zufallsvariablen *L^p-beschränkt*, wenn $K > 0$ existiert, sodass

$$\mathbb{E}[|X_n|^p] \le K \quad \text{für alle } n.$$

Satz 12 (Doobscher Martingalkonvergenzsatz)**.** *Sei (X_n) ein Supermartingal, das L^1-beschränkt ist. Dann existiert eine reellwertige Zufallsvariable X, sodass*

$$\lim_{n\to\infty} X_n = X \textit{ fast sicher.}$$

Bemerkung: Ist X_n nichtnegativ, so gilt

$$\mathbb{E}|X_n| = \mathbb{E}X_n = \mathbb{E}X_0 =: K$$

und somit ist X_n automatisch L^1-beschränkt und $\lim_{n\to\infty} X_n = X$ existiert.

Die Beweisidee besteht darin, die Anzahl der Überquerungen eines Intervalls zu zählen. Grob gesprochen, wenn (X_n) nicht konvergiert, so oszilliert es und überquert daher ein Intervall $[a, b]$ unendlich oft. Dann kann man die folgende antizyklische Wettstrategie verfolgen: Wenn der Wert von (X_n) unter a liegt, so setzen wir einen Taler bis der Wert wieder über b liegt. Wir gewinnen dabei bei jeder Überquerung mindestens $b - a$ Taler. Aber eine solche Gewinnstrategie kann es nicht geben...

Für den rigorosen Beweis fixiere $a < b$ und setze $I = [a, b]$. Die Anzahl $U_N[a, b]$ der Überquerungen von $[a, b]$ durch (X_n) bis zur Zeit N ist die größte ganze Zahl k so dass Zahlen

$$0 \le s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \le N$$

existieren mit $X_{s_i} < a$ und $X_{t_i} > b$ für alle $1 \le i \le k$. Dann ist $U_N[a, b]$ eine Zufallsvariable die folgende Ungleichung erfüllt.

Lemma 2.2 (Doobsches Überquerungslemma)**.** *Sei (X_n) ein Supermartingal und $a < b$. Dann gilt, für alle N,*

$$(b - a)\, \mathbb{E}U_N[a, b] \le \mathbb{E}[(X_N - a)^-]\,.$$

Beweis: Wir betrachten die folgende Spielstrategie. Setze $C_k = 0$ für $k = 0, \dots, n_1$, wobei $n_1 = \min\{k \geq 0 \colon X_k < a\}$. Dann setzen wir $C_k = 1$ für $k = n_1 + 1, \dots, n_2$ wobei $n_2 = \min\{k \geq n_1 \colon X_k > b\}$. Dann $C_k = 0$ für $k = n_2 + 1, \dots, n_3$ wobei $n_3 = \min\{k \geq n_2 \colon X_k < a\}$, und so fort. Also ist

$$C_1 = 1_{\{X_0 < a\}} \text{ und } C_n = 1_{\{C_{n-1}=1\}} 1_{\{X_{n-1} \leq b\}} + 1_{\{C_{n-1}=0\}} 1_{\{X_{n-1} < a\}} .$$

(C_n) ist prävisibel und wir betrachten die Martingaltransformierte

$$Y_n = (C \bullet X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}) .$$

Nach Wahl unserer Strategie gilt

$$Y_N \geq (b - a) U_N[a, b] - (X_N - a)^- .$$

Jede Überquerung erhöht den Wert von Y_n um mindestens $b - a$ und der letzte Term trägt einer möglichen unvollständigen Überquerung am Ende Rechnung. Da C prävisibel, nichtnegativ und beschränkt ist, ist (Y_n) ein Supermartingal und

$$0 = \mathbb{E}[Y_0] \geq \mathbb{E}[Y_N] \geq (b - a)\mathbb{E}[U_N[a, b]] - \mathbb{E}[(X_N - a)^-],$$

wie behauptet.

Damit folgt, dass die erwartete Anzahl $U_N[a, b]$ der Überquerungen von $[a, b]$ durch (X_n) bis zur Zeit N beschränkt ist, wenn $\mathbb{E}[(X_N - a)^-]$ beschränkt ist.

Lemma 2.3. *Sei (X_n) ein Supermartingal, das L^1-beschränkt ist. Dann gilt für den monotonen Limes*

$$U[a, b] := \lim_{N \to \infty} U_N[a, b]$$

dass

$$(b - a)\mathbb{E}[U[a, b]] \leq |a| + \sup_n \mathbb{E}|X_n| < \infty .$$

Insbesondere, $P\{U[a, b] = \infty\} = 0$.

Beweis: Nach dem Doobschen Überquerungslemma gilt

$$(b - a)\mathbb{E}[U_N[a, b]] \leq \mathbb{E}[(X_N - a)^-] \leq |a| + \mathbb{E}|X_N| \leq |a| + \sup_n \mathbb{E}|X_n| .$$

Durch Grenzübergang $N \to \infty$ mit Hilfe von monotoner Konvergenz folgt die Behauptung.

Um den Beweis des Martingalkonvergenzsatzes abzuschließen betrachten wir nun das Ereignis

$$M[a,b] := \{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\}.$$

Wenn X_n weder konvergiert noch gegen $\pm\infty$ divergiert, so gibt es rationale Zahlen $a < b$, sodass das Ereignis $M[a,b]$ stattfindet. Findet aber $M[a,b]$ statt, so ist $U[a,b] = \infty$. Aber das Ereignis, dass rationale Zahlen a, b mit $U[a,b] = \infty$ existieren hat Wahrscheinlichkeit null. Folglich konvergiert (X_n) gegen eine möglicherweise unendliche Zufallsvariable X. Nach dem Lemma von Fatou aber gilt

$$\mathbb{E}|X| = \mathbb{E}[\liminf_{n\to\infty} |X_n|] \leq \liminf_{n\to\infty} \mathbb{E}|X_n| \leq K < \infty\,,$$

und somit $|X| < \infty$ fast sicher. Damit ist der Martingalkonvergenzsatz bewiesen.

Beispiele. (a) Sei (X_n) eine einfache symmetrische Irrfahrt. Dann ist (X_n) ein Martingal, das nicht konvergiert. Wir wissen nämlich, dass sowohl 0 als auch 1 unendlich oft besucht werden, was der Konvergenz widerspricht. (X_n) ist offensichtlich nicht L^1-beschränkt.

(b) Ist $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots$ eine Filtration und X eine nichtnnegative, integrierbare Zufallsvariable, so ist $\mathbb{E}|\mathbb{E}[X|\mathcal{F}_n]| = \mathbb{E}X < \infty$ und es folgt aus dem Martingalkonvergenzsatz, dass

$$\lim_{n\to\infty} \mathbb{E}[X|\mathcal{F}_n]$$

fast sicher existiert. Wir werden später notwendige und hinreichende Bedingungen formulieren, wann dieser Limes gleich X ist.

(c) Betrachte den Galton-Watson Prozess (X_n) mit Nachkommensverteilung gegeben durch $(p_0, p_1, \ldots)$ und mittlerer Nachkommenszahl

$$\sum_{n=0}^{\infty} np_n = 1\,.$$

Da (X_n) ein nichtnegatives Martingal ist konvergiert es fast sicher gegen eine Zufallsvariable $X \geq 0$. Da (X_n) Werte in den ganzen Zahlen annimmt gilt somit $X_n = X$ für hinreichend große n, fast sicher. Wenn $p_1 < 1$ (im Fall $p_1 = 1$ gilt trivialerweise $X_n = 1$ für alle n), so gilt für alle K und $k > 0$,

$$P\{X_n = k \text{ für alle } n \geq K\} \leq \lim_{M\to\infty} \prod_{n=K+1}^{M} \mathbb{P}(X_n = k \,|\, X_{n-1} = k) = 0\,,$$

und damit $X = 0$ fast sicher. Mit anderen Worten, *der kritische Galton-Watson Prozess stirbt in endlicher Zeit aus.* Beachte aber, dass $\mathbb{E}[X_n] = 1$ für alle n. Die Konvergenz im Martingalkonvergenzsatz schliesst also nicht Konvergenz der Erwartungswerte mit ein.

2.3.2 Uniform integrierbare Martingale

Die Beispiele zeigen, dass Konvergenze der Martingale besonders wertvoll ist, wenn ausser den Martingalen selbst, auch die Erwartungswerte konvergieren. Wir untersuchen deshalb auch eine andere Art von Konvergenz, die L^p-Konvergenz, für $p \geq 1$.

Eine Folge (X_n) *konvergiert in* L^p gegen eine Zufallsvariable X, wenn

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n - X|^p\big] = 0.$$

Offensichtlich folgt aus L^1-Konvergenz, dass auch $\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X$.

Lemma 2.4. *Sei $1 \leq p \leq q$. Wenn die Folge (X_n) in L^q gegen X konvergiert, so konvergiert sie auch in L^p.*

Beweis: Wir benutzen die Jensensche Ungleichung $\varphi(\mathbb{E}Y) \leq \mathbb{E}\varphi(Y)$ für die konvexe Funktion $\varphi(x) = x^{q/p}$ und $Y = |X_n - X|^p$. Damit ist

$$\mathbb{E}\big[|X_n - X|^p\big]^{q/p} \leq \mathbb{E}\big[|X_n - X|^q\big] \to 0,$$

woraus die Behauptung folgt.

Lemma 2.5. *Sei $p \geq 1$ beliebig.*

(a) Wenn (X_n) in L^p gegen X konvergiert, so konvergiert (X_n) auch in Wahrscheinlichkeit gegen X, das heißt für alle $\varepsilon > 0$ gilt

$$\lim_{n\to\infty} P\big\{|X_n - X| \geq \varepsilon\big\} = 0.$$

(b) Wenn (X_n) fast sicher gegen X konvergiert, so konvergiert (X_n) auch in Wahrscheinlichkeit gegen X.

Bemerkung Es gibt Folgen, die fast sicher konvergieren, aber nicht in L^p, siehe Beispiel (c) oder auch das Beispiel in Abschnitt 1.2.3. Es gibt aber auch Folgen, die in L^p konvergieren, aber nicht fast sicher, siehe Übungen.

Beweis: (a) Wir haben nach der Markovschen Ungleichung

$$P\big\{|X_n - X| \geq \varepsilon\big\} = P\big\{|X_n - X|^p \geq \varepsilon^p\big\} \leq \varepsilon^{-p}\mathbb{E}\big[|X_n - X|^p\big] \to 0.$$

(b) Wenn $X_n(\omega) \to X(\omega)$, so gibt es ein $n(\omega)$ mit $|X_n(\omega) - X(\omega)| \leq \varepsilon$ für alle $n \geq n(\omega)$. Dann gilt aufgrund der Stetigkeit von Wahrscheinlichkeitsmaßen

$$P\big\{|X_n - X| \leq \varepsilon\big\} \geq P\{n(\omega) \leq n\} \to P\{n(\omega) < \infty\} = 1.$$

Definition: Eine Folge (X_n) von Zufallsvariablen heißt *uniform integrierbar* wenn für alle $\varepsilon > 0$ ein $K \geq 0$ existiert mit

$$\mathbb{E}\big[1_{\{|X_n|>K\}}|X_n|\big] < \varepsilon \text{ für alle } n\,.$$

In den folgenden Beweisen wird das folgende Lemma nützlich sein.

Lemma 2.6. *Sei X integrierbare Zufallsvariable. Dann gibt es für alle $\varepsilon > 0$ ein $\delta > 0$, sodass für alle $F \in \mathcal{A}$ mit $P(F) < \delta$ gilt*

$$\mathbb{E}\big[1_F|X|\big] < \varepsilon\,.$$

Beweis: Wenn das für ein $\varepsilon > 0$ nicht gilt, so gibt es eine Folge $F(n) \in \mathcal{A}$ von Ereignissen mit $P(F(n)) < 2^{-n}$ und $\mathbb{E}1_{F(n)}|X| \geq \varepsilon$. Dann betrachten wir das Ereignis H dass unendlich viele der Ereignisse $F(n)$ stattfinden. Wir erhalten $P(H) = 0$, da

$$P(H) \leq P\Big(\bigcup_{n\geq m} F(n)\Big) \leq \sum_{n\geq m} \mathbb{P}\big(F(n)\big) \leq \sum_{n\geq m} 2^{-n} = 2^{-m+1} \overset{m\to\infty}{\longrightarrow} 0\,.$$

Gleichzeitig gilt aber auch

$$\mathbb{E}\big[1_H|X|\big] = \int \limsup_{n\to\infty} 1_{F(n)}|X|\,dP \geq \limsup_{n\to\infty} \int 1_{F(n)}|X|\,dP \geq \varepsilon\,,$$

nach dem Lemma von Fatou. Das ist ein Widerspruch.

Proposition 2.6.

(a) Gibt es eine integrierbare Zufallsvariable Y mit $|X_n| \leq Y$ für alle n, so ist die Folge (X_n) uniform integrierbar. Insbesondere ist jede beschränkte Folge von Zufallsvariablen uniform integrierbar.

(b) Ist die Folge (X_n) L^p-beschränkt für ein $p > 1$, so ist sie uniform integrierbar.

(c) Ist die Folge (X_n) konvergent in L^1, so ist sie uniform integrierbar.

(d) Jede uniform integrierbare Folge ist L^1-beschränkt.

(e) Es gibt L^1-beschränkte Folgen, die nicht uniform integrierbar sind.

Beweis: (a) Es gilt

$$\mathbb{E}[1_{\{|X_n|>K\}}|X_n|] \leq \mathbb{E}[1_{\{|Y|>K\}}|Y|]\,.$$

Nach dem Satz von der monotonen Konvergenz gilt

$$\mathbb{E}|Y| = \lim_{K\to\infty} \mathbb{E}\big[1_{\{|Y|\leq K\}}|Y|\big]\,.$$

Damit folgt $\lim_{K\to\infty} \mathbb{E}[1_{\{|Y|>K\}}|Y|] = 0$ und zusammen mit der ersten Ungleichung folgt uniforme Integrierbarkeit der Folge (X_n).

Für (b) nehmen wir an, dass $\sup_n \mathbb{E}[|X_n|^p] < C$ für $p > 1$. Für jedes K gilt

$$\mathbb{E}\big[1_{\{|X_n|>K\}}|X_n|\big] \leq K^{1-p}\mathbb{E}\big[1_{\{|X_n|>K\}}|X_n|^p\big] \leq K^{1-p}C\,.$$

Uniforme Integrierbarkeit folgt durch geeignete Wahl von K.

Für (c) sei X integrierbar. Dann gibt es nach dem Satz von der monotonen Konvergenz ein $K > 1$, sodass

$$\mathbb{E}\big[1_{\{|X|>K\}}|X|\big] < \frac{\varepsilon}{3}.$$

Nach Lemma 2.6 gibt es für jedes $\varepsilon > 0$ ein $\delta > 0$, so dass $P(A) < \delta$ impliziert $\mathbb{E}[1_A|X|] < \varepsilon$. Folglich finden wir bei gegebenem K ein N mit

$$\mathbb{E}\big[1_{\{|X-X_n|>K\}}|X|\big] < \frac{\varepsilon}{3} \text{ für alle } n \geq N.$$

Daher gilt

$$\begin{aligned}\mathbb{E}\big[1_{\{|X_n|>2K\}}|X_n|\big] &\leq \mathbb{E}\big[1_{\{|X_n|>2K\}}|X|\big] + \mathbb{E}\big[|X - X_n|\big]\\ &\leq \mathbb{E}\big[1_{\{|X|>K\}}|X|\big] + \mathbb{E}\big[1_{\{|X-X_n|>K\}}|X|\big] + \frac{\varepsilon}{3} \leq \varepsilon,\end{aligned}$$

für alle $n \geq N$. Die endlich vielen verbleibenden Werte können durch eine weitere Vergrößerung von K mit eingeschlossen werden.

Zum Beweis von (d) wählen wir $\varepsilon = 1$ und finden $K > 0$ mit

$$\mathbb{E}|X_n| \leq \mathbb{E}\big[1_{\{|X_n|>K\}}|X_n|\big] + \mathbb{E}\big[1_{\{|X_n|\leq K\}}|X_n|\big] \leq 1 + K\,.$$

Für (e) sei U uniform verteilt auf $(0,1)$ und

$$X_n = n\mathbb{1}_{\{U \le 1/n\}} \ge 0 .$$

Dann ist $\mathbb{E}|X_n| = 1$ und die Folge somit L^1-beschränkt. Da

$$\mathbb{E}\big[\mathbb{1}_{\{|X_n|>K\}}|X_n|\big] = 1 \text{ für alle } n > K$$

ist die Folge nicht uniform integrierbar.

Der Hauptsatz dieses Abschnitts zeigt unter anderem, dass uniform integrierbare Martingale immer konvergent in L^1 sind.

Satz 13 (Charakterisierung der L^1-Konvergenz)**.** *Sei (X_n) Folge von Zufallsvariablen, die in Wahrscheinlichkeit gegen eine Zufallsvariable X konvergiert. Dann konvergiert sie genau dann in L^1, wenn sie uniform integrierbar ist.*

Beweis: Wir wissen bereits aus der Proposition, dass L^1-Konvergenz der Folge (X_n) uniforme Integrierbarkeit impliziert. Sei die Folge (X_n) nun uniform integrierbar. Nach dem Lemma von Fatou gilt $\mathbb{E}|X| \le \liminf \mathbb{E}|X_n| < \infty$. Für jedes $K \ge 0$ definieren wir die cutoff-Funktion

$$\varphi_K(x) = \begin{cases} K & \text{if } x > K \\ x & \text{if } |x| \le K \\ -K & \text{if } x < -K. \end{cases}$$

Aufgrund der uniformen Integrierbarkeit existiert K, sodass für alle n,

$$\mathbb{E}\big[|\varphi_K(X_n) - X_n|\,\big] < \tfrac{\varepsilon}{3} \text{ und } \mathbb{E}\big[|\varphi_K(X) - X|\big] < \tfrac{\varepsilon}{3} .$$

Da

$$|\varphi_K(X) - \varphi_K(X_n)| \le |X - X_n|$$

folgt, dass $(\varphi_K(X_n))$ in Wahrscheinlichkeit gegen $\varphi_K(X)$ konvergiert. Also existiert N sodass

$$P\big\{|\varphi_K(X_n) - \varphi_K(X)| > \tfrac{\varepsilon}{6}\big\} < \tfrac{\varepsilon}{12K} \text{ für alle } n \ge N .$$

Also gilt, für alle $n \ge N$,

$$\begin{aligned} \mathbb{E}|\varphi_K(X_n) - \varphi_K(X)| &\le \mathbb{E}\big[\mathbb{1}_{\{|\varphi_K(X_n)-\varphi_K(X)|>\varepsilon/6\}}|\varphi_K(X_n) - \varphi_K(X)|\big] \\ &\quad + \mathbb{E}\big[\mathbb{1}_{\{|\varphi_K(X_n)-\varphi_K(X)|\le\varepsilon/6\}}|\varphi_K(X_n) - \varphi_K(X)|\big] \\ &\le 2K\, P\big\{|\varphi_K(X_n) - \varphi_K(X)| > \tfrac{\varepsilon}{6}\big\} + \tfrac{\varepsilon}{6} \le \tfrac{\varepsilon}{3} . \end{aligned}$$

Nach der Dreiecksungleichung gilt, für alle $n \ge N$,

$$\mathbb{E}\big[|X_n - X|\big] \le \mathbb{E}|X_n - \varphi_K(X_n)| + \mathbb{E}|\varphi_K(X_n) - \varphi_K(X)| + \mathbb{E}|X - \varphi_K(X)| \le \varepsilon .$$

Das komplettiert den Beweis.

Bemerkung: Nach Proposition 2.6(b), ist der Konvergenzsatz für uniform integrierbare Folgen eine Verallgemeinerung des Satzes von der dominierten Konvergenz.

Zurück zur Untersuchung von Martingalen. Sei $(\mathcal{F}_n)$ eine Filtration und (X_n) ein Martingal beüglich dieser Filtration. Der nächste Satz zeigt, dass es für jedes uniform integrierbare Martingal eine Zufallsvariable X gibt, sodass das Martingal Informationen über X akkumuliert.

Satz 14 (Konvergenzsatz für uniform integrierbare Martingale)**.** *Wenn das Martingal (X_n) is uniform integrierbar ist, so gibt es eine reellwertige Zufallsvariable X, sodass*

$$\lim_{n\to\infty} X_n = X \textit{ fast sicher und in } L^1.$$

Ausserdem gilt, für jedes n, dass $X_n = \mathbb{E}[X|\mathcal{F}_n]$ fast sicher.

Beweis: Da die Folge (X_n) uniform integrierbar ist, ist sie insbesondere L^1-beschränkt und nach dem Doobschen Martingalkonvergenzsatz fast sicher konvergent gegen eine reellwertige Zufallsvariable X. Nach dem Konvergenzsatz für uniform integrierbare Folgen gilt diese Konvergenz auch im L^1 Sinne. Um die letzte Aussage zu verifizieren prüfen wir für X_n die Eigenschaften einer bedingten Erwartung. Offensichtlich ist X_n messbar bezüglich $\mathcal{F}_n$. Sei nun $F \in \mathcal{F}_n$. Für alle $m \geq n$ gilt nach der Martingaleigenschaft

$$\int 1_F X_m \, dP = \int 1_F X_n \, dP \, .$$

Wir lassen $m \to \infty$, dann gilt

$$\Big| \int 1_F X_m \, dP - \int 1_F X \, dP \Big| \leq \int |X_m - X| \, dP \to 0 \, ,$$

und erhalten, wie gefordert,

$$\int 1_F X \, dP = \int 1_F X_n d\, P \, .$$

Beispiel: Exponentielles Wachstum einer Population. Sei (X_n) ein Galton-Watson Prozess mit mittlerer Nachkommenzahl μ. Im *kritischen Fall* $\mu = 1$ haben wir gesehen, dass der Prozess fast sicher in endlicher Zeit ausstirbt. Nun formulieren wir Voraussetzungen, dass der Prozess *exponentiell wächst* mit positiver Wahrscheinlichkeit. Angenommen die Nachkommensvertiuelung (gegeben durch die Folge (p_n)) hat

- Mittelwert $\mu = \sum_{n=0}^{\infty} n p_n > 1$ (*superkritischer Fall*),
- positive und endliche Varianz $\sigma^2 = \sum_{n=0}^{\infty} n^2 p_n - \mu^2$.

Wir haben bereits herausgefunden, dass

$$\mathbb{E}\big[X_{n+1}\,\big|\mathcal{F}_n\big] = \mu X_n,$$

und somit definiert $M_n = X_n/\mu^n$ ein Martingal. Da $M_k \geq 0$ existiert eine Zufallsvariable $M \geq 0$ mist

$$\lim_{n\to\infty} M_n = M \text{ fast sicher.}$$

Wir wolen zeigen, dass $P\{M > 0\} > 0$. Dazu zeigen wir, dass (M_n) uniform integrierbar ist. Daraus folgt nämlich, dass

$$\mathbb{E}M = \lim_{n\to\infty} \mathbb{E}M_n = 1\,,$$

und somit kann $M \geq 0$ nicht null sein, woraus folgt, dass $P\{M > 0\} > 0$. Um uniforme Integrierbarkeit nachzuprüfen, genügt es zu seigen, dass (M_n) L^2-beschränkt ist (siehe Proposition 2.6(c)). Es gilt, fast sicher

$$\begin{aligned}
\mathbb{E}[M_n^2|\mathcal{F}_{n-1}] &= \mathbb{E}[M_{n-1}^2|\mathcal{F}_{n-1}] + \mathbb{E}[2M_{n-1}(M_n - M_{n-1})|\mathcal{F}_{n-1}] \\
&\quad + \mathbb{E}[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}] \\
&= M_{n-1}^2 + \mathbb{E}[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}]\,.
\end{aligned}$$

Zur Berechnung des zweiten Terms betrachte

$$\mathbb{E}[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}] = \mu^{-2n}\mathbb{E}[(X_n - \mu X_{n-1})^2|\mathcal{F}_{n-1}]$$

und

$$\begin{aligned}
\mathbb{E}\Big[(X_n - \mu X_{n-1})^2\,|\,\mathcal{F}_{n-1}\Big] &= \mathbb{E}\Big[\Big(\sum_{k=1}^{X_{n-1}} (Y_k^n - \mu)\Big)^2\Big|\mathcal{F}_{n-1}\Big] \\
&= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} 1\{k \leq X_{n-1}\}\,1\{\ell \leq X_{n-1}\}\;\mathbb{E}\big[(Y_k^n - \mu)(Y_\ell^n - \mu)\,|\,\mathcal{F}_{n-1}\big] \\
&= \sum_{k=1}^{X_{n-1}} \mathbb{E}\big[(Y_k^n - \mu)^2\big] = X_{n-1}\sigma^2\,.
\end{aligned}$$

Zusammengefasst

$$\mathbb{E}[M_n^2] = \mathbb{E}[M_{n-1}^2] + (\sigma^2/\mu^{2n})\mathbb{E}[X_{n-1}]\,.$$

Nun gilt $\mathbb{E}[X_{n-1}] = \mu^{n-1}\mathbb{E}[M_{n-1}] = \mu^{n-1}$ und somit

$$\mathbb{E}[M_n^2] = \mathbb{E}[M_{n-1}^2] + \frac{\sigma^2}{\mu^{n+1}}\,,$$

woraus induktiv folgt, dass,

$$\mathbb{E}[M_n^2] \leq 1 + \sigma^2 \sum_{k=1}^{\infty} \frac{1}{\mu^{k+1}} < \infty\,,$$

und wir folgern, dass (M_n) L^2-beschränkt ist Insgesamt haben wir also gezeigt, dass fast sicher gilt, für alle hinreichend großen n,

$$X_n = M_n \mu^n \geq \frac{M}{2}\mu^n\,,$$

sodass wenn $M > 0$, und damit mit positiver Wahrscheinlichkeit, der Galton-Watson Prozess exponentiell wächst.

2.3.3 Die Konvergenzsätze von Lévy

Satz 14 zeigt, dass jedes uniform integrierbare Martingal *abgeschlossen* ist, was heißt dass es eine (versteckte) Zufallsvariable X gibt mit $X_n = \mathbb{E}[X|\mathcal{F}_n]$ fast sicher. Umgekehrt ist jedes abgeschlossene Martingal uniform integrierbar und konvergiert gegen die versteckte Zufallsvariable, wie der folgende Satz zeigt.

Satz 15 (Lévyscher Konvergenzsatz für abgeschlossene Martingale)**.** *Sei X eine integrierbare Zufallsvariable und $(\mathcal{F}_n)$ eine Filtration, so dass X messbar ist bezüglich der σ-Algebra, die von der Vereinigung aller $\mathcal{F}_n$ erzeugt wird. Definiere $X_n = \mathbb{E}[X|\mathcal{F}_n]$. Dann ist (X_n) ein uniform integrierbares Martingal und*

$$\lim_{n\to\infty} X_n = X \textit{ fast sicher und in } L^1.$$

Der wichtigste Beweisschritt ist das folgende Lemma.

Lemma 2.7. *Sei X eine integrierbare Zufallsvariable und $(\mathcal{F}_n)$ eine Folge von σ-Algebren. Dann ist die Folge $(\mathbb{E}[X|\mathcal{F}_n])$ uniform integrierbar.*

Beweis: Sei $\varepsilon > 0$ gegeben. Wähle nach Lemma 2.6 ein $\delta > 0$ so, dass für alle $F \in \mathcal{A}$,

$$P(F) < \delta \Longrightarrow \mathbb{E}\big[1_F |X|\big] < \varepsilon\,.$$

Nun wählen wir K größer als $\mathbb{E}|X|/\delta$. Durch getrennte Betrachtung des Positiv- und Negativteils von X erhalten wir, fast sicher,

$$\big|\mathbb{E}\big[X|\mathcal{F}_n\big]\big| \leq \mathbb{E}\big[|X|\,\big|\,\mathcal{F}_n\big],$$

und somit

$$\begin{aligned} KP\{|\mathbb{E}[X|\mathcal{F}_n]| > K\} &= \mathbb{E}\big[K1_{\{|\mathbb{E}[X|\mathcal{F}_n]|>K\}}\big] \leq \mathbb{E}\big|\mathbb{E}[X|\mathcal{F}_n]\big| \\ &\leq \mathbb{E}\big[\mathbb{E}[|X|\,|\mathcal{F}_n]\big] = \mathbb{E}|X|\,, \end{aligned}$$

und schließlich

$$P\{|\mathbb{E}[X|\mathcal{F}_n]| > K\} \leq \frac{\mathbb{E}|X|}{K} < \delta\,.$$

Dieses Ereignis liegt in $\mathcal{F}_n$. Nach Definition der bedingten Erwartung gilt daher, für alle n,

$$\begin{aligned} \mathbb{E}\Big[1_{\{|\mathbb{E}[X|\mathcal{F}_n]|>K\}}|\mathbb{E}[X|\mathcal{F}_n]|\Big] &\leq \mathbb{E}\big[1_{\{|\mathbb{E}[X|\mathcal{F}_n]|>K\}}\mathbb{E}[|X|\,|\mathcal{F}_n]\big] \\ &= \mathbb{E}\big[1_{\{|\mathbb{E}[X|\mathcal{F}_n]|>K\}}|X|\big] < \varepsilon\,. \end{aligned}$$

Das komplettiert den Beweis.

Beweis des Konvergenzsatzes. Wir wissen bereits, dass (X_n) ein Martingal und uniform integrierbar ist. Also gibt es eine Zufallsvariable Y, sodass

$$\lim_{n\to\infty} X_n = Y \text{ fast sicher und in } L^1.$$

Um zu zeigen, dass $X = Y$ fast sicher, nehmen wir oBdA an, dass $X \geq 0$, und somit auch $X_n \geq 0$ und $Y \geq 0$ fast sicher. Beobachte, dass

$$\mathbb{E}[X] = \mathbb{E}[X_n] \to \mathbb{E}[Y]\,.$$

Definiere Wahrscheinlichkeitsmaße P_1 und P_2 auf $\mathcal{A}$ (definiert auf der σ-Algebra, die von der Vereinigung der $\mathcal{F}_n$ erzeugt wird) durch

$$P_1(A) = \frac{1}{\mathbb{E}X}\int 1_A X\,dP \qquad \text{und} \qquad P_2(A) = \frac{1}{\mathbb{E}X}\int 1_A Y\,dP\,.$$

Nun ist $\mathcal{F} := \bigcup \mathcal{F}_n$ ein schnittstabiles Erzeugendensystem ist, das $\mathcal{A}$ erzeugt. Wenn $A \in \mathcal{F}_n$, so gilt für alle $m \geq n$,

$$P_1(A) = \frac{1}{\mathbb{E}X}\int 1_A X\,dP = \frac{1}{\mathbb{E}X}\int 1_A X_m\,dP\,.$$

Des weiteren,

$$\lim_{m\to\infty} \frac{1}{\mathbb{E}X} \int 1_A X_m \, dP = \frac{1}{\mathbb{E}X} \int 1_A Y \, dP = P_2(A) ,$$

wobei die erste Gleichheit aus der L^1-Konvergenz folgt. Der Eindeutigkeitssatz liefert nun $P_1(A) = P_2(A)$ für alle $A \in \mathcal{A}$. Sei $A = \{X > Y\} \in \mathcal{A}$. Dann würde $P(A) > 0$ nach Definition, $P_1(A) > P_2(A)$ implizieren, ein Widerspruch. Also ist $P(A) = 0$, und somit $X \leq Y$ fast sicher. Der Beweis folgt mit Vertauschung der Rollen von X und Y.

Der obige Satz heißt auch *Lévy's upward theorem*. Sein Partner, *Lévy's downward theorem* betrachtet schrumpfende, statt wachsender, σ-Algebren.

Satz 16 (Lévyscher Rückwärtskonvergenzsatz)**.** *Sei $\{\mathcal{G}_{-n} \ : \ n \geq 0\}$ eine Familie von σ-Algebren, sodass*

$$\mathcal{G}_{-\infty} := \bigcap_{k=0}^{\infty} \mathcal{G}_{-k} \subseteq \cdots \subseteq \mathcal{G}_{-n} \subseteq \cdots \subseteq \mathcal{G}_{-2} \subseteq \mathcal{G}_{-1} .$$

Sei X eine Zufalllsvariable und

$$X_{-n} = \mathbb{E}[X|\mathcal{G}_{-n}] .$$

Dann gilt

$$\lim_{n\to\infty} X_{-n} = \mathbb{E}[X|\mathcal{G}_{-\infty}] \text{ fast sicher und in } L^1.$$

Beweis: Sei N eine positive ganze Zahl. Wir betrachten die Filtration $(\mathcal{F}_n)$ gegeben durch

$$\mathcal{F}_n = \mathcal{G}_{(n-N)\wedge(-1)}$$

und den adaptierten Prozess $Y_n = X_{(n-N)\wedge(-1)}$. Nach der Turmeigenschaft ist

$$\begin{aligned}\mathbb{E}[Y_n \,|\, \mathcal{F}_{n-1}] &= \mathbb{E}\Big[\mathbb{E}[X \,|\, \mathcal{G}_{(n-N)\wedge(-1)}] \,\Big|\, \mathcal{G}_{(n-1-N)\wedge(-1)}\Big] \\ &= \mathbb{E}\Big[X \,\Big|\, \mathcal{G}_{(n-1-N)\wedge(-1)}\Big] = Y_{n-1} ,\end{aligned}$$

und somit ist (Y_n) ein Martingal. Wir erhalten vom Überquerungslemma, dass

$$(b-a)\mathbb{E}[U_N[a,b]] \leq \mathbb{E}[(X_{-1} - a)^-] .$$

Grenzübergang $N \to \infty$ ergibt, dass fast sicher die Anzahl der Überquerungen von $[a,b]$ durch den Prozess (X_{-n}) endlich ist. Da dies simultan für alle rationalen Paare $a < b$ gilt, kann man wie im Martingalkonvergenzsatz argumentieren und erhält , dass $\lim X_{-n} = X_{-\infty}$ fast sicher existiert. Nach Lemma 2.7

ist die Folge sogar uniform integrierbar und somit gilt die Konvergenz auch in L^1. Um zu prüfen, dass $X_{-\infty} = \mathbb{E}[X|\mathcal{G}_{-\infty}]$, beobachtet man, dass für alle m und $G \in \mathcal{G}_{-\infty} \subset \mathcal{G}_{-m}$,

$$\int 1_G X \, dP = \int 1_G X_{-m} \, dP \,,$$

und durch $m \to \infty$ erhält man, dass $X_{-\infty}$ eine bedingte Erwartung von X gegeben $\mathcal{G}_{-\infty}$ ist.

2.3.4 Der Doobsche Zerlegungssatz

Wie weit ist eine allgemeiner, adaptierter Prozess X davon entfernt ein Martingal zu sein? Kann man vielleicht eine *Drift* definieren, die man von X abzieht, um es zu einem Martingal zu machen? Die Antwort ist, unter schwachen Voraussetzungen, positiv, wie das folgende wichtige Resultat zeigt.

Satz 17 (Doobscher Zerlegungsssatz)**.** *Sei $(\mathcal{F}_n : n \geq 0)$ eine Filtration und $(X_n : n \geq 0)$ ein adaptierter Prozess mit $\mathbb{E}|X_n| < \infty$ für alle $n \geq 0$. Dann existiert*

- *ein Martingal $(M_n : n \geq 0)$ with $M_0 = 0$, and*
- *ein prävisibler Process $(A_n : n \geq 1)$ mit $A_0 = 0$,*

sodass

$$X_n - X_0 = M_n + A_n \text{ für alle } n \geq 0. \tag{2.2}$$

Zudem stimmen alle Prozesspaare (M, A), die dies erfüllen, fast sicher überein und X ist ein Submartingal genau dann wenn der Prozess A monoton wachsend ist, das heißt wenn $P\{A_n \leq A_{n+1} \forall n\} = 1$.

Beweis: Sei (M, A) ein Prozesspaar wie oben. Dann gilt, fast sicher, dass

$$\begin{aligned} \mathbb{E}\big[X_n - X_{n-1} \,\big|\, \mathcal{F}_{n-1}\big] &= \mathbb{E}\big[M_n - M_{n-1} \,\big|\, \mathcal{F}_{n-1}\big] + \mathbb{E}\big[A_n - A_{n-1} \,\big|\, \mathcal{F}_{n-1}\big] \\ &= A_n - A_{n-1}, \end{aligned}$$

da (A_n) prävisibel ist und (M_n) ein Martingal. Daraus ergibt sich eine explizite Formel für A_n, nämlich

$$A_n = \sum_{k=1}^{n} \mathbb{E}\big[X_k - X_{k-1} \,\big|\, \mathcal{F}(k-1)\big] \text{ fast sicher.}$$

Daraus folgt die Eindeutigkeit von (A_n) und schließlich auch von (M_n). Umgekehrt kann diese Formel benutzt werden, um (A_n) aus (X_n) zu definieren.

Dann ist (A_n) prävisibel und $M_n = X_n - A_n - X_0$ definiert ein Martingal. Aus der Definition von A_n sieht man auch, dass die Submartingaleigenschaft von (X_n) äquivalent dazu ist, dass (A_n) monoton wachsend ist.

Wir betrachten nun einen besonders interessanten Fall. Sei $(M_n : n \geq 0)$ ein Martingal mit $M_0 = 0$ und $\mathbb{E}[M_n^2] < \infty$ für alle $n \geq 0$. Im allgemeinen definiert $X_n = M_n^2$ *kein* Martingal, wohl aber ein Submartingal. Der Prozess $(M_n^2 : n \geq 0)$ hat daher eine Doobzerlegung

$$M_n^2 = N_n + A_n, \text{ für alle } n \geq 0,$$

mit $N_0 = A_0 = 0$, wobei der Prozess $(A_n \colon n \geq 0)$ wachsend ist und meist als spitze Klammer oder *wachsender Prozess* von M bezeichnet und durch $\langle M \rangle_n := A_n$ notiert wird. Wir definieren die Zufallsvariable

$$\langle M \rangle_\infty = \lim_{n\to\infty} \langle M \rangle_n.$$

Der nächste Satz beinhaltet Konvergenzresutate, die nicht im Martingalkonvergenzsatz enthalten sind.

Satz 18. *Sei $(M_n \colon n \geq 0)$ quadratisch integrierbares Martingal mit $M_0 = 0$.*

(a) $\lim_{n\to\infty} M_n$ *existiert (und ist endlich) fast sicher auf* $\{\langle M \rangle_\infty < \infty\}$.

(b) $\lim_{n\to\infty} M_n / \langle M \rangle_n = 0$ *fast sicher auf* $\{\langle M \rangle_\infty = \infty\}$.

Der Beweis benutzt die beiden folgenden Lemmas.

Lemma 2.8. *Ist $(M_n : n \geq 1)$ ein Martingal mit $\mathbb{E}M_n^2 < \infty$, so gilt*

$$\mathbb{E}\big[(M_n - M_{n-1})^2 \,\big|\, \mathcal{F}_{n-1}\big] = \mathbb{E}\big[M_n^2 - M_{n-1}^2 \,\big|\, \mathcal{F}_{n-1}\big].$$

Beweis: Wir benutzen die binomische Formel, ziehen bekanntes raus und erhalten

$$\begin{aligned}\mathbb{E}\big[(M_n - M_{n-1})^2 \,\big|\, \mathcal{F}_{n-1}\big] &= \mathbb{E}\big[M_n^2 \,\big|\, \mathcal{F}_{n-1}\big] - 2M_{n-1}\mathbb{E}\big[M_n \,\big|\, \mathcal{F}_{n-1}\big] \\ &\quad + \mathbb{E}\big[M_{n-1}^2 \,\big|\, \mathcal{F}_{n-1}\big] \\ &= \mathbb{E}\big[M_n^2 - M_{n-1}^2 \,\big|\, \mathcal{F}_{n-1}\big].\end{aligned}$$

Lemma 2.9 (Kroneckers Lemma)**.** *Ist (b_n) eine Folge nichtnegativer Zahlen mit $b_n \uparrow \infty$ und (x_n) eine reelle Folge. Dann gilt für $s_n = x_1 + x_2 + \cdots + x_n$ dass*

$$\sum_{n=1}^{\infty} \frac{x_n}{b_n} \text{ konvergiert genau dann, wenn } \lim_{n\to\infty} \frac{s_n}{b_n} = 0.$$

Beweis: Übungsaufgabe.

Nun beweisen wir Satz 18. Wir nehmen an, dass M_n quadratisch integrierbar ist, aber nicht dass es L^2-beschränkt ist, sodass Konvergenz nicht aus dem Martingalkonvergenzsatz folgt.

Beweis von (a). Da $A = \langle M \rangle$ prävisibel ist gilt für alle $k \geq 1$, dass

$$S := S(k) := \inf \{ n \geq 0 : A_{n+1} > k \}$$

eine Stoppzeit ist. Wir zeigen nun, dass $\langle M^S \rangle = \langle M \rangle^S$. Beachte, dass

$$\{A_\infty < \infty\} = \bigcup_{k=1}^{\infty} \{S(k) = \infty\}. \tag{2.3}$$

Für jedes feste k, ist der gestoppte Prozess A^S prävisibel, da für jede Borelmenge $B \subseteq \mathbb{R}$,

$$\{A_n^S \in B\} = \bigcup_{r=0}^{n-1} \{S = r, A_r \in B\} \cup \Big(\{A_n \in B\} \cap \{S \leq n-1\}^c \Big).$$

Daher ist $(M^S)^2 - A^S = (M^2 - A)^S = N^S$ ein Martingal und daraus folgt $\langle M^S \rangle = A^S$. Der Prozess A^S ist durch k beschränkt und

$$\mathbb{E}[(M_n^S)^2] = \mathbb{E}[A_n^S] + \mathbb{E}[N_n^S] = \mathbb{E}[A_n^S] \leq k,$$

woraus folgt, dass M^S L^2-beschränkt ist. Wir schließen daraus, dass

$$\lim_{n \to \infty} M_n^{S(k)}$$

fast sicher existiert. Nach (2.3) gilt auf $\{\langle M \rangle_\infty < \infty\}$ dass ein k existiert mit $S(k) = \infty$, was den Beweis von (a) abschließt.

Beweis von (b) Definiere den Prozess $(W_n : n \geq 0)$ durch $W_0 = 0$ und

$$W_n := \sum_{k=1}^{n} \frac{M_k - M_{k-1}}{1 + A_k} = (Y \bullet M)_n,$$

wobei $Y_n = 1/(1 + A_n)$ ein beschränkter, prävisibler Prozess. Nach Proposition 2.4 ist (W_n) ein Martingal. Fast sicher gilt, nach Lemma 2.8,

$$\begin{aligned}
\mathbb{E}\big[(W_n - W_{n-1})^2 | \mathcal{F}_{n-1}\big] &= \frac{1}{(1 + A_n)^2} \mathbb{E}\big[(M_n - M_{n-1})^2 | \mathcal{F}_{n-1}\big] \\
&= \frac{1}{(1 + A_n)^2} \mathbb{E}\big[M_n^2 - M_{n-1}^2 | \mathcal{F}_{n-1}\big] = \frac{1}{(1 + A_n)^2} (A_n - A_{n-1}) \\
&\leq \frac{(1 + A_n) - (1 + A_{n-1})}{(1 + A_n)(1 + A_{n-1})} = \frac{1}{1 + A_{n-1}} - \frac{1}{1 + A_n},
\end{aligned}$$

und somit nach Lemma 2.8,

$$\mathbb{E}\langle W\rangle_n = \mathbb{E}W_n^2 = \sum_{k=1}^{n} \mathbb{E}[W_k^2 - W_{k-1}^2] = \sum_{k=1}^{n} \mathbb{E}[(W_k - W_{k-1})^2] \leq 1.$$

Mit dem Lemma von Fatou folgt, dass $\mathbb{E}\langle W\rangle_\infty = \mathbb{E}[\liminf_{n\to\infty}\langle W\rangle_n] \leq \liminf_{n\to\infty}\mathbb{E}[\langle W\rangle_n] \leq 1$, und somit $\langle W\rangle_\infty < \infty$ fast sicher, sodass nach (a) $\lim_{n\to\infty} W_n$ existiert. Mit dem Lemma von Kronecker schließen wir, dass fast sicher

$$\lim_{n\to\infty} \frac{M_n}{1+A_n} = \lim_{n\to\infty} \frac{\sum_{k=1}^n (M_k - M_{k-1})}{1+A_n} = 0,$$

woraus dann die Aussage folgt.

2.4 Anwendungen der Martingaltheorie

2.4.1 Eine elementare Black-Scholes Formel

Die Black-Scholes Formel zur Bestimmung von Optionspreisen ist ein zentrales Ergebnis der probabilistischen Finanzmathematik. Wir wollen in diesem Abschnitt eine sehr einfache diskrete Variante dieses Resultats mit Hilfe der Martingaltheorie diskutieren.

In unserem ökonomischen Modell gibt es zwei Arten von Wertpapiere: *Anleihen*, die mit einer Rate r fest verzinst werden, und *Aktien*, deren Wert fluktuiert. Wir schreiben S_n für den Wert einer Aktie und $B_n = (1+r)^n B_0$ für den Wert der Anleihe im Zeitintervall $(n, n+1)$. Ein Portfolio (A_n, V_n) besteht aus A_n Aktien und B_n Anleihen, die der Eigner zur Zeit n hält. Damit ergibt sch das Startkapital des Eigners als

$$x = X_0 := A_0 S_0 + V_0 B_0$$

und durch Handel im Interval $(0,1)$ kann der Eigner jedes Portfolio (A_1, V_1) erhalten, das die Bedingung $x = A_1 S_0 + V_1 B_0$ erfüllt. Der Wert seines Portfolios im Interval $(1,2)$ ist dann

$$X_1 := A_1 S_1 + V_1 B_1$$

und so weiter. Der Gewinn im $n$ten Spiel ist damit

$$X_n - X_{n-1} = A_n(S_n - S_{n-1}) + V_n(B_n - B_{n-1}).$$

Wir haben $B_n - B_{n-1} = rB_{n-1}$ und $S_n - S_{n-1} = R_n S_{n-1}$ für eine fluktuierende Rate R_n und damit ergibt sich

$$X_n - X_{n-1} = rX_{n-1} + A_n S_{n-1}(R_n - r).$$

Betrachtet man den diskontierten Portfoliowert

$$Y_n := (1+r)^{-n} X_n$$

so gilt

$$Y_n - Y_{n-1} = (1+r)^{-n+1} A_n S_{n-1}(R_n - r).$$

In unserem Modell kann R_n nur zwei möglich Werte annehmen, nämlich $a \in (-1, r)$ und $b \in (r, \infty)$.

Problem: Eine europäische Option erlaubt es ihrem Besitzer eine Aktie zur Zeit $n = N$ zum Preis K zu erwerben. Was ist ein fairer Preis für eine solche Option zur Zeit $n = 0$?

Eine *Hedging Strategie* mit Anfangswert x für die Option ist ein prävisibler Prozeß $((A_n, V_n) \colon n \in \mathbb{N})$ wie oben beschrieben, so dass $X_n \geq 0$ für alle $n \in \{0, \ldots, N\}$ und

$$X_N = (S_N - K)^+.$$

Beachte, dass die linke Seite der Wert des Portfolios zur Zeit N ist, und die rechte Seite der Wert der Option zur Zeit N. Black and Scholes postulieren, dass x ein fairer Preis der Option ist, wenn es eine Hedging Strategie mit Anfangswert x gibt, sodass der Wert des Portfolios zur Zeit N unabhängig von Wertverlauf der Aktie gleich dem Wert der Option zu dieser Zeit ist. Man beachte, dass diesem Ansatz kein stochastisches Modell zugrundeliegt.

Satz 19 (Black-Scholes Formel)**.** *Eine Hedging-Strategie mit Anfangswert x existiert genau dann wenn*

$$x = \mathbb{E}\big[(1+r)^{-N}(S_N - k)^+\big],$$

wobei sich der Erwartungswert auf das Modell von unabhängig identisch verteilten R_n bezieht, die mit Wahrscheinlichkeit $p = \frac{r-a}{b-a}$ den Wert b und mit Wahrscheinlichkeit $1 - p = \frac{b-r}{b-a}$ den Wert a annehmen.

Beweis: Angenommen, eine Hedging Strategie existiert mit Anfangswert x existiert. Wir betrachten u.i.v. Zufallsvariable $\varepsilon_1, \varepsilon_2, \ldots$ mit $P(\varepsilon_i = 1) = p = 1 - P(\varepsilon_i = -1)$ und setzen

$$Z_n = \sum_{k=1}^{n} (\varepsilon_k - 2p + 1).$$

(Z_n) ist ein Martingal und wir können R_n in unserem Modell ausdrücken als

$$R_n = r + \frac{1}{2}(b-a)(Z_n - Z_{n-1}).$$

Damit können wir Y_n als Martingaltransformierte ausdrücken, nämlich

$$Y_n = Y_0 + (F \bullet Z)_n,$$

für $F_n = (1+r)^{-n+1} A_n S_{n-1}$. Indebsondere ist (F_n) prävisibel und beschränkt und (Z_n) ein Martingal. Damit ist auch (Y_n) ein Martingal nach Proposition 2.4. Für $Y_0 = x$ ist folglich

$$x = \mathbb{E} Y_N = \mathbb{E}\big[(1+r)^{-N} X_N\big] = \mathbb{E}\big[(1+r)^{-N} (S_N - K)^+\big],$$

wie behauptet. Nichthegativität von X_n wurde übrigens nicht verwendet.

Nun müssen wir zeigen, dass für unsere Wahl von x eine Hedging Strategie existiert. Wir definieren nun für unser Modell, dass

$$Y_n := \mathbb{E}\big[(1+r)^{-N} (S_N - K)^+ \big| \mathcal{F}_n\big].$$

Dann ist (Y_n) ein Martingal und wir zeigen gleich, dass ein prävisibler Prozess (A_n) existiert, sodass

$$Y_n - Y_{n-1} = (1+r)^{-n+1} A_n S_{n-1} (R_n - r). \tag{2.4}$$

Wenn wir dies haben setzen wir $X_n = (1+r)^n Y_n$ und $V_n = (X_n - A_n S_n)/B_n$. Dann gelten die Formeln, die (X_n) und (A_n, V_n) in Beziehung setzen und insbesondere ist $((A_n, V_n))$ prävisibel. Da $X_0 = Y_0 = x$ und $X_N = \mathbb{E}\big[(S_N - K)^+ \big| \mathcal{F}_N\big] = (S_N - K)^+$ haben wir die Hedging Strategie gefunden. Es bleibt nur folgendes Lemma zu zeigen, das (2.4) impliziert.

Lemma 2.10. *Sei $(M_n \colon 0 \leq n \leq N)$ ein Martingal bezüglich der durch $\mathcal{F}_n = \sigma(\varepsilon_1, \ldots, \varepsilon_n)$ gegebenen Filtration und $Z_n = \sum_{k=1}^n (\varepsilon_k - 2p + 1)$. Dann existiert ein prävisibler Prozess (H_n), sodass*

$$M_n = M_0 + \sum_{k=1}^n H_k (Z_k - Z_{k-1}).$$

Beweis: Da M_n messbar bezüglich $\mathcal{F}_n$ ist, gibt es eine Funktion

$$f_n \colon \{-1, 1\}^n \to \mathbb{R}$$

mit $M_n = f_n(\varepsilon_1, \ldots, \varepsilon_n)$. Die Martingaleigenschaft liefert

$$\begin{aligned} 0 &= \mathbb{E}\big[M_n - M_{n-1} | \mathcal{F}_{n-1}\big] \\ &= p f_n(\varepsilon_1, \ldots, \varepsilon_{n-1}, 1) + (1-p) f_n(\varepsilon_1, \ldots, \varepsilon_{n-1}, -1) - f_{n-1}(\varepsilon_1, \ldots, \varepsilon_{n-1}). \end{aligned}$$

Setzt man

$$\begin{aligned} H_n &= \frac{f_n(\varepsilon_1, \ldots, \varepsilon_{n-1}, 1) - f_{n-1}(\varepsilon_1, \ldots, \varepsilon_{n-1})}{2(1-p)} \\ &= \frac{f_{n-1}(\varepsilon_1, \ldots, \varepsilon_{n-1}) - f_n(\varepsilon_1, \ldots, \varepsilon_{n-1}, -1)}{2p} \end{aligned}$$

so folgt das Resultat durch einfach algebraische Manipulation.

2.4.2 Klassische Sätze der Wahrscheinlichkeitstheorie

Das Null-Eins Gesetz von Kolmogorov ist ein nützliches Hilfsmittel, um zu erkennen, dass bestimmte asymptotische Ereignisse nicht zufällig sind.

Satz 20 (Kolmogorovsches Null-Eins Gesetz). *Sei (X_n) eine Folge unabhängig identisch verteilter Zufallsvariablen. Definiere die σ-Algebren*

$$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$$

und die terminale *σ-Algebra*

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n.$$

Jedes Ereignis $A \in \mathcal{T}$ hat $P(A) = 0$ oder $P(A) = 1$.

Beispiele terminaler Ereignisse sind

- $\{X_n \in A_n$ für unendlich viele $n\}$,
- $\{\lim_{n\to\infty} X_n$ existiert $\}$, oder $\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right\}$.

Beweis: Sei $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$ und $\mathcal{F}_\infty = \sigma(X_1, X_2, \ldots)$. Beachte, dass $\mathcal{F}_n$ und $\mathcal{T}_n$ unabhängig sind. Für $A \in \mathcal{T}$ definiere $X = 1_A$. Dann ist X beschränkt und messbar bezüglich $\mathcal{F}_\infty$. Der Lévysche Konvergenzsatz liefert

$$X = \lim_{n\to\infty} \mathbb{E}[X \mid \mathcal{F}_n] \text{ fast sicher.}$$

Da X messbar ist bezüglich $\mathcal{T}_n$ ist es unabhänig von $\mathcal{F}_n$, und die rechte Seite ist somit $\mathbb{E}[X] = P(A)$. Das Ergebnis folgt, da die linke Seite nur Werte in $\{0,1\}$ annimmt.

Wir geben nun einen sehr eleganten Beweis des starken Gestzes der großen Zahlen von Kolmogorov mit Hilfe der Martingaltheorie. Dies ist ein Highlight der Vorlesung.

Satz 21 (Starkes Gesetz der großen Zahlen von Kolmogorov). *Sei (X_n) eine Folge unabhängig identisch verteilter integrierbarer Zufallsvariablen und μ ihr gemeinsamer Erwartungswert. Dann gilt*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = \mu \textit{ fast sicher und in } L^1.$$

Beweis: Setze $S_n = \sum_{k=1}^n X_k$. Sei $\mathcal{G}(-n)$ die von den Zufallsvariablen

$$S_n, S_{n+1}, S_{n+2}, \ldots$$

erzeugte σ-Algebra und $\mathcal{G}(-\infty)$ die Schnittmenge all dieser Mengensysteme. Wir zeigen zunächst, dass

$$\mathbb{E}[X_1|\mathcal{G}(-n)] = \mathbb{E}[X_1|\sigma(S_n)] = \frac{S_n}{n} \text{ fast sicher.} \tag{2.5}$$

Da $\mathcal{G}(-n)$ auch von den Zufallsvariablen $S_n, X_{n+1}, X_{n+2}, \ldots$ erzeugt wird und X_1 unabhängig von $X_{n+1}, X_{n+2}, \ldots$ ist, folgt $\mathbb{E}[X_1|\mathcal{G}(-n)] = \mathbb{E}[X_1|\sigma(S_n)]$. Es gilt nach Definition der bedingten Erwartung

$$\int 1_{\{S_n \in B\}} \, \mathbb{E}[X_1|\sigma(S_n)] \, dP = \int 1_{\{S_n \in B\}} X_1 \, dP = \int f_1 P_X \otimes \overset{n \text{ times}}{\cdots} \otimes P_X.$$

wobei

$$f_k(x_1, \ldots, x_n) = 1_{\{x_1 + \cdots + x_n \in B\}} x_k$$

und P_X die Verteilung der X_i ist. Aufgrund der Symmetrie des Produktmasses gilt, für alle $k \in \{1, \ldots, n\}$, dass

$$\int f_1 P_X \otimes \overset{n \text{ times}}{\cdots} \otimes P_X = \int f_k P_X \otimes \overset{n \text{ times}}{\cdots} \otimes P_X,$$

und somit

$$\int 1_{\{S_n \in B\}} \, \mathbb{E}[X_1|\sigma(S_n)] \, dP = \int 1_{\{S_n \in B\}} X_k \, dP.$$

Insbesondere ist

$$\mathbb{E}[X_1|\sigma(S_n)] = \mathbb{E}[X_2|\sigma(S_n)] = \cdots = \mathbb{E}[X_n|\sigma(S_n)],$$

und da die Summe dieser Terme $\mathbb{E}[S_n|\sigma(S_n)] = S_n$ ergibt, folgt (2.5).

Nach dem Lévyschen Rückwärtskonvergenzsatz gilt somit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n X_k = \lim_{n\to\infty} \mathbb{E}[X_1|\mathcal{G}(-n)] = \mathbb{E}[X_1|\mathcal{G}(-\infty)] \text{ fast sicher und in } L^1.$$

Da der Grenzwert messbar ist bezüglich der terminalen σ-Algebra $\mathcal{T}$ und diese nach dem Null-Eins Gesetz von Kolmogorov nur Ereignisse mit Wahrscheinlichkeit 0 oder 1 enthält, ist $\mathbb{E}[X_1|\mathcal{G}(-\infty)]$ fast sicher konstant und somit gleich μ.

Als Anwendung des Doobschen Zerlegungssatzes diskutieren wir eine Verallgemeinerung des Borel-Cantelli Lemmas.

Satz 22 (Borel-Cantelli Lemma)**.** *Sei* $(E_n : n \geq 0)$ *eine Folge von Ereignissen mit* $E_n \in \mathcal{F}_n$*. Definiere* $X_n = \mathbb{E}[1_{E_n} \,|\, \mathcal{F}_{n-1}]$*. Dann gilt, fast sicher,*

(a) Ist $\sum_{n=1}^{\infty} X_n < \infty$*, so treten nur endlich viele der Ereignisse* $E_1, E_2, \ldots$ *ein.*

(b) Ist $\sum_{n=1}^{\infty} X_n = \infty$*, so treten unendlich viele der Ereignisse* $E_1, E_2, \ldots$ *ein.*

Bemerkung: Das klassische Borel-Cantelli Lemma folgt hieraus:

(a) $\mathbb{E}[X_k] = P(E_k)$ und daher impliziert $\sum_{n=1}^{\infty} P(E_n) < \infty$ dass $\mathbb{E}[\sum_{n=1}^{\infty} X_n] < \infty$ und somit $\sum_{n=1}^{\infty} X_n < \infty$ fast sicher.

(b) Sind die Ereignisse $(E_n : n \geq 0)$ unabhängig, so setze $\mathcal{F}_n = \sigma(E_1, \ldots, E_n)$ und beobachte, dass $X_k = P(E_k)$.

Beweis: Sei $Z_n = \sum_{k=1}^{n} 1_{E_k}$ die Anzahl der Ereignisse, die bis zur Zeit n eintreten, und $Y_n = \sum_{k=1}^{n} X_k$. Dann definiert $M_n = Z_n - Y_n$ ein Martingal, und $Z_n = M_n + Y_n$ die Doobzerlegung des Submartingals $(Z_n : n \geq 0)$. Wir zeigen, dass

$$\langle M \rangle_n = \sum_{k=1}^{n} X_k(1 - X_k) \leq Y_n, \text{ fast sicher.}$$

Dazu prüft man nach, dass die Differenz von M_n^2 und der rechten Seite ein Martingal ist, es gilt nämlich

$$\begin{aligned}
&\mathbb{E}\big[M_n^2 - \sum_{k=1}^{n} X_k(1 - X_k) \,\big|\, \mathcal{F}_{n-1}\big] \\
&\qquad = M_{n-1}^2 + \mathbb{E}\big[(1_{E_n} - X_n)^2 \,\big|\, \mathcal{F}_{n-1}\big] - \sum_{k=1}^{n} X_k(1 - X_k) \\
&\qquad = M_{n-1}^2 - \sum_{k=1}^{n-1} X_k(1 - X_k) \text{ fast sicher.}
\end{aligned}$$

Falls $Y_\infty := \sum_{n=1}^{\infty} X_n < \infty$ ist $\langle M \rangle_\infty < \infty$ und Lemma 18(a) impliziert, dass $\lim M_n$ existiert und endlich ist. Dann gilt $\sum_{k=1}^{\infty} 1_{E_k} < \infty$. Falls aber $Y_\infty = \infty$ und $\langle M \rangle_\infty < \infty$, so ist auch $\lim M_n$ existent und endlich und somit $\sum_{k=1}^{\infty} 1_{E_k} = \infty$. Es bliebt der Fall $\langle M \rangle_\infty = \infty$. Dann liefert Lemma 18(b) dass $\lim_{n\to\infty} M_n/\langle M \rangle_n = 0$, somit $\lim_{n\to\infty} M_n/Y_n = 0$ and schließlich $\lim_{n\to\infty} Z_n/Y_n = 1$. Daraus folgt aber $\sum_{k=1}^{\infty} 1_{E_k} = \lim_{n\to\infty} Z_n = \infty$.

2.4.3 Stochastische Optimierungsprobleme

Das Bellmansche Optimalitätsprinzip

In einem für den Spieler vorteilhaften Spiel seien die Gewinne $X_1, X_2, \ldots, X_N$ für den Einsatz einer Geldeinheit unabhängig identische verteilte Zufallsvariable mit

$$P\{X_n = 1\} = p, P\{X_n = -1\} = q := 1 - p \text{ für } \tfrac{1}{2} < p < 1.$$

Bei einem Anfangskapital von Y_0 möchte man die Einsätze C_n so wählen, dass die erwartete Verzinsung

$$\mathbb{E}\Big[\log\Big(\frac{Y_N}{Y_0}\Big)\Big]$$

maximiert wird.

Sei $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Wir nehmen an, dass

- $Y_n = Y_{n-1} + C_n X_n$ das Kapital zur Zeit n ist,
- die Einsätze C_n prävisibel sind und $C_n \le Y_{n-1}$.

Was ist die beste Strategie und welche erwartetete Verzinsung wird dadurch erreicht?

Um diese Frage zu beantworten setzen wir $R_n = C_n / Y_{n-1}$, das ist der Anteil des Kapitals, den wir im $n$ten Spiel einsetzen. Dann repräsentiert $R = (R_1, \ldots, R_N)$ eine Strategie.

Lemma 2.11. *Sei R eine (prävisible) Strategie und definiere die* relative Entropie *von (p, q) als*

$$\alpha = p \log(2p) + q \log(2q)$$

Dann definiert

$$M_n = \log\Big(\frac{Y_n}{Y_0}\Big) - n\alpha,\ M_0 = 0$$

ein Supermartingal.

Beweis: Es gilt

$$\begin{aligned}
\mathbb{E}\big[M_n \,|\, \mathcal{F}_{n-1}\big] &= \mathbb{E}\Big[\log\Big(\frac{Y_n}{Y_0}\Big) - n\alpha \Big| \mathcal{F}_{n-1}\Big] \\
&= \log\Big(\frac{Y_{n-1}}{Y_0}\Big) - n\alpha + \mathbb{E}\Big[\log\Big(\frac{Y_n}{Y_{n-1}}\Big)\Big| \mathcal{F}_{n-1}\Big].
\end{aligned}$$

Der letzte Ausdruck ist

$$\mathbb{E}\Big[\log\Big(\frac{Y_n}{Y_{n-1}}\Big)\,\Big|\,\mathcal{F}_{n-1}\Big] = \mathbb{E}\big[\log\big(1+R_nX_n\big)\,\big|\,\mathcal{F}_{n-1}\big] = p\log\big(1+R_n\big)+q\log\big(1-R_n\big).$$

Wir zeigen, dass dies $\leq \alpha$ ist, was äquivalent ist zu

$$p\log\Big(\frac{1+R_n}{2p}\Big)+q\log\Big(\frac{1-R_n}{2q}\Big)\leq 0.$$

Da log eine konkave Funktion ist gilt

$$p\log\Big(\frac{1+R_n}{2p}\Big)+q\log\Big(\frac{1-R_n}{2q}\Big)\leq \log\Big(p\frac{1+R_n}{2p}+q\frac{1-R_n}{2q}\Big)=0,$$

wie behauptet. Zusammengefasst

$$\mathbb{E}\big[M_n\,|\,\mathcal{F}_{n-1}\big]\leq \log\Big(\frac{Y_{n-1}}{Y_0}\Big)-n\alpha+\alpha = M_{n-1}.$$

Dies ist die Supermartingaleigenschaft.

Der Beweis zeigt auch, für welche Strategie wir ein Martingal erhalten.

Lemma 2.12. *Sei $R^* = (2p-1,\dots,2p-1)$. Dann ist $(M_n : n\geq 0)$ ein Martingal.*

Beweis: Wir haben Gleichheit im Beweis von Lemma 2.11 indem wir R_n so wählen, dass

$$p\log\Big(\frac{1+R_n}{2p}\Big)+q\log\Big(\frac{1-R_n}{2q}\Big)= 0.$$

Das ist dann der Fall, wenn $1+R_n = 2p$ und $1-R_n = 2q$, und beide Gleichungen sind äquivalent zu $R_n = 2p-1$.

Um die Anfangsfrage zu beantworten erhalten wir von Lemma 2.11 dass für jede Strategie R gilt

$$\mathbb{E}\Big[\log\Big(\frac{Y_N}{Y_0}\Big)\Big]\leq N\alpha,$$

und für die Strategie $R^* = (2p-1,\dots,2p-1)$ gilt

$$\mathbb{E}\Big[\log\Big(\frac{Y_N}{Y_0}\Big)\Big]= N\alpha,$$

und das macht sie optimal.

Das Sekretärinnenproblem

Das folgende Optimierungsproblem wird auch Sekretärinnenproblem genannt: N Kandidatinnen stellen sich bei einer Arbeitgeberin vor. Die Eignung der $i$ten Kandidatin für den Job sei X_i und $X_1, \dots, X_N$ sind unabhängige uniform verteilte Zufallsvariable auf $(0,1)$. Die Arbeitgeberin interviewt die Kandidatinnen nacheinander und bestimmt den Wert X_i genau. Sie muss sofort entscheiden, ob sie eine Kandidatin nimmt oder wegschickt, und die nächste Kandidatin zum Interview bittet. Kandidatinnen können nicht zurückgeholt werden. Das Problem ist also eine Stoppzeit T zu finden, die $\mathbb{E}X_T$ maximiert.

Lemma 2.13. *Die Stoppzeit $T^* = \inf\{n > 0 : X_n > \alpha_n\}$, mit $\alpha_N = 0$, und*

$$\alpha_{n-1} = \frac{1}{2} + \frac{\alpha_n^2}{2} \text{ für } 1 \le n \le N,$$

maximiert $\mathbb{E}X_T$.

Beweis: Der Beweis erfolgt in vier Schritten. Im *ersten Schritt* zeigen wir, dass für alle $0 \le \alpha \le 1$ gilt

$$\mathbb{E}[X_n \vee \alpha] = \frac{1}{2} + \frac{\alpha^2}{2}. \tag{2.6}$$

Um (2.6) zu prüfen beobachte, dass

$$\begin{aligned} \mathbb{E}[X_n \vee \alpha] &= \int_0^1 x \vee \alpha \, dx \\ &= \int_0^\alpha \alpha \, dx + \int_\alpha^1 x \, dx \\ &= \alpha^2 + \frac{1}{2} - \frac{\alpha^2}{2} = \frac{1}{2} + \frac{\alpha^2}{2}. \end{aligned}$$

Der *zweite Schritt* ist zu zeigen, dass für *jede* Stoppzeit T der durch

$$Y_0 = \alpha_0, \text{ und } Y_n = (X_{T \wedge n}) \vee \alpha_n \text{ für } n \ge 1$$

definierte Prozess ein Supermartingal ist. Es gilt nämlich auf dem Ereignis $\{T \le n-1\}$, dass

$$\mathbb{E}\big[Y_n \,|\, \mathcal{F}_{n-1}\big] = \mathbb{E}\big[X_{T\wedge n} \vee \alpha_n \,|\, \mathcal{F}_{n-1}\big] = X_T \vee \alpha_n \le X_T \vee \alpha_{n-1} = Y_{n-1},$$

wobei wir benutzen, dass α_n fallend ist. Auf dem Ereignis $\{T > n-1\}$ gilt

$$\mathbb{E}\big[Y_n \,|\, \mathcal{F}_{n-1}\big] = \mathbb{E}\big[X_n \vee \alpha_n\big] = \frac{\alpha_n^2}{2} + \frac{1}{2} = \alpha_{n-1} \le Y_{n-1},$$

und damit ist die Supermartingaleigenschaft nachgewiesen.

Als *dritten Schritt* zeigen wir, dass für $T = T^*$ der Prozess (Y_n) ein Martingal ist. Auf dem Ereignis $\{T^* \leq n-1\}$ haben wir, wie oben

$$\mathbb{E}\big[Y_n \,|\, \mathcal{F}_{n-1}\big] = X_{T^*} \vee \alpha_n = X_{T^*},$$

da $X_{T^*} > \alpha_{T^*} \geq \alpha_{n-1} \geq \alpha_n$. Beobachte, dass $Y_{n-1} = X_{T^*} \vee \alpha_{n-1} = X_{T^*}$. Auf dem Ereignis $\{T^* \geq n\}$ gilt

$$\mathbb{E}\big[Y_n \,|\, \mathcal{F}_{n-1}\big] \quad = \quad \mathbb{E}\big[X_n \vee \alpha_n\big] = \frac{\alpha_n^2}{2} + \frac{1}{2} = \alpha_{n-1} = Y_{n-1},$$

und damit ist die Martingaleigenschaft nachgewiesen.

Im *vierten Schritt* zeigen wir schließlich, dass für jede Stoppzeit T gilt $\mathbb{E}X_T \leq \mathbb{E}X_{T^*}$. Dazu nutzen wir den Doobschen Stoppsatz der angewandt auf die (durch N beschränkten) Stoppzeiten liefert

$$\mathbb{E}\big[X_T\big] \leq \mathbb{E}\big[X_T \vee \alpha_T\big] = \mathbb{E}\big[Y_T\big] \leq \mathbb{E}\big[Y_0\big] = \alpha_0,$$

und für unsere Wahl T^*,

$$\mathbb{E}\big[X_{T^*}\big] = \mathbb{E}\big[X_{T^*} \vee \alpha_{T^*}\big] = \mathbb{E}\big[Y_{T^*}\big] = \mathbb{E}\big[Y_0\big] = \alpha_0.$$

Das komplettiert den Beweis.

Kapitel 3

Markovketten in diskreter Zeit

3.1 Markovketten: Definition und Beispiele

Sei $(X_n\colon n \geq 0)$ ein stochastischer Prozess in diskreter Zeit mit Werten in einer abzählbaren oder endlichen Menge I. Sei außerdem $(\mathcal{F}_n\colon n \geq 0)$ die natürliche Filtration dieses Prozesses. Die Intuition dabei ist, dass X_n den *Zustand* des Prozesses zur Zeit n darstellt und $\mathcal{F}_n$ unser Wissen über alle Zustände bis zur Zeit n repräsentiert.

Ein Prozess $(X_n\colon n \geq 0)$ hat die *Markoveigenschaft*, wenn gilt

$$\mathbb{P}\big(X_{n+1} \in A \big| \mathcal{F}_n\big) = \mathbb{P}\big(X_{n+1} \in A \big| \sigma(X_n)\big) \text{ für alle } A \subseteq I, \text{ fast sicher.}$$

Das heißt, dass der Zustand des Prozesses zur Zeit $n+1$ nur vom Zustand zur Zeit n abhängt und, wenn dieser bekannt ist, nicht von dem Pfad des Prozesses davor beeinflusst wird.

Wenn ein Prozess $(X_n\colon n \geq 0)$ die Markoveigenschaft hat, so gibt es Abbildungen $p_n\colon I^2 \to [0,1]$ mit

$$\mathbb{P}\big(X_{n+1} = s \big| \mathcal{F}_n\big) = p_n(X_n, s) \text{ für alle } s \in I, \text{ fast sicher.}$$

Hängen die p_n nicht von n ab, so heißt der Prozess *homogen*.

Definition: Der Prozess $(X_n\colon n \geq 0)$ ist eine *(homogene) Markovkette*, wenn es eine Abbildung $p\colon I^2 \to [0,1]$ gibt mit

$$\mathbb{P}\big(X_{n+1} = s \big| \mathcal{F}_n\big) = p(X_n, s) \text{ für alle } s \in I, \text{ fast sicher.}$$

Die Verteilung der Markovkette $(X_n \,:\, n \in \mathbb{N})$ ist damit gegeben durch eine *Startverteilung* oder *Anfangsverteilung* $(w_i \,:\, i \in I)$ und die Abbildung $p\colon I^2 \to [0,1]$, die wir als *Übergangsmatrix* $P = (p_{ij} \,:\, i,j \in I)$ auffassen. Wir vereinbaren, dass (X_n) auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P}_x)$ stets eine Markovkette mit Anfangswert $x \in I$ und Übergangsmatrix $P = (p_{ij}\colon i,j \in I)$ ist.

Umgekehrt ist eine Matrix $P = (p_{ij} : i, j \in I)$ genau dann die Übergangsmatrix einer Markovkette, wenn gilt $p_{ij} \geq 0$ für alle $i, j \in I$ und

$$\sum_{j \in I} p_{ij} = 1 \qquad \text{für alle } i \in I.$$

Die Verteilung der Markovkette ist dann gegeben durch

$$\begin{aligned}
\mathbb{P}\big(X_0 = x_0, &X_1 = x_1, \dots, X_n = x_n\big) \\
&= \mathbb{P}(X_0 = 0) \prod_{i=1}^{n} \mathbb{P}\big(X_i = x_i | X_{i-1} = x_{i-1}, \dots, X_0 = 0\big) \\
&= \mathbb{P}(X_0 = 0) \prod_{i=1}^{n} \mathbb{P}\big(X_i = x_i | X_{i-1} = x_{i-1}\big) = w_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n},
\end{aligned}$$

für alle $x_0, \dots, x_n \in I$.

Beispiel 3.1. Ein Virus existiert in zwei Formen α, β und in jeder Generation bleibt es entweder gleich oder, mit Wahrscheinlichkeit p, mutiert in die andere Form. Wenn das Virus ursprünglich in Form α ist, was ist die Wahrscheinlichkeit, dass es nach n Generationen in derselben Form ist?

Sei X_n die Form des Virus in der $n$ten Generation. Dann ist (X_n) eine Markovkette mit Werten im Zustandsraum $I = \{\alpha, \beta\}$. Wichtig ist, dass die Zufallsvariablen X_n and X_{n+1} im allgemeinen *nicht unabhängig* sind. Wir beobachten

$$\begin{aligned}
\mathbb{P}\{X_n = \alpha, X_{n+1} = \alpha\} &= \mathbb{P}\{X_n = \alpha \text{ and no mutation occurs in step } n+1\} \\
&= (1-p)\mathbb{P}\{X_n = \alpha\},
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\{X_n = \beta, X_{n+1} = \alpha\} &= \mathbb{P}\{X_n = \beta \text{ and mutation occurs in step } n+1\} \\
&= p\mathbb{P}\{X_n = \beta\},
\end{aligned}$$

und somit

$$\begin{aligned}
\mathbb{P}\{X_{n+1} &= \alpha \,|\, X_n = \alpha\} = 1-p, \quad \mathbb{P}\{X_{n+1} = \beta \,|\, X_n = \alpha\} = p, \\
\mathbb{P}\{X_{n+1} &= \alpha \,|\, X_n = \beta\} = p, \quad \mathbb{P}\{X_{n+1} = \beta \,|\, X_n = \beta\} = 1-p.
\end{aligned}$$

Die Übergangsmatrix hier ist also

$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$

Wir nutzen die Gleichungen

$$\mathbb{P}\{X_{n+1} = \alpha\} = p\mathbb{P}\{X_n = \beta\} + (1-p)\mathbb{P}\{X_n = \alpha\}$$

und

$$\mathbb{P}\{X_n = \beta\} = 1 - \mathbb{P}\{X_n = \alpha\},$$

und erhalten für die relevante Größe $p_n := \mathbb{P}\{X_n = \alpha\}$ die Rekursionsgleichung

$$p_{n+1} = p(1-p_n) + (1-p)p_n = p_n(1-2p) + p, \qquad p_0 = 1.$$

Diese hat die eindeutige Lösung

$$p_n = \frac{1}{2} + \frac{1}{2}(1-2p)^n.$$

Wenn $p \neq 0, 1$ konvergiert p_n gegen $1/2$, unabhängig von der Mutationswahrscheinlichkeit p. Die systematische Untersuchung des Langzeitverhaltens von Markovketten steht im Mittelpunkt dieses Kapitels.

Beispiel 3.2. Die einfache Irrfahrt $S_n = X_1 + \cdots + X_n$ mit $\mathbb{P}(X_i = 1) = p$ und $\mathbb{P}(X_i = -1) = 1 - p$ ist eine Markovkette mit Übergangsmatrix $P = (p_{ij} \colon i, j \in \mathbb{Z})$ gegeben durch

$$p_{ij} = \mathbb{P}\big\{X_{n+1} = j \,|\, X_n = i\big\} = \begin{cases} 0 & \text{if } |j-i| \neq 1, \\ p & \text{if } j - i = 1, \\ 1-p & \text{if } j - i = -1. \end{cases}$$

3.2 Starke Markoveigenschaft, Rekurrenz und Transienz

3.2.1 Die starke Markoveigenschaft

Eine Zufallsvariable T mit Werten in $\mathbb{N}$ heißt Stoppzeit bezüglich der Filtration $(\mathcal{F}_n)$ wenn $\{T \leq n\} \in \mathcal{F}_n$. Gegeben eine Stoppzeit T definieren wir die σ-Algebra $\mathcal{F}_T$ als Mengensystem der Ereignisse A so dass $A \cap \{T \leq n\} \in \mathcal{F}_n$. Insbesondere sind T und X_T messbar bezüglich $\mathcal{F}_T$.

Satz 23 (Starke Markoveigenschaft)**.** *Sei (X_n) eine Markovkette und T eine (endliche) Stoppzeit. Dann gilt*

$$\mathbb{P}\big((X_{T+n})_{n\geq 0} \in A \big| \mathcal{F}_T\big) = \mathbb{P}_{X_T}\big((X_n)_{n\geq 0} \in A\big),$$

für A in der Produkt-σ-Algebra von $I^{\mathbb{N}_0}$.

Beweis: Sei $F \in \mathcal{F}_T$. Gemäß Definition der bedingten Erwartung müssen wir zeigen, dass

$$\mathbb{E}\big[1_F\, \mathbb{P}_{X_T}\big((X_n)_{n\geq 0} \in A\big)\big] = \mathbb{E}\big[1_F 1_{\{(X_{T+n})_{n\geq 0}\in A\}}\big].$$

Sei $A = \big\{(x_n)\colon x_0 \in A_0, \ldots, x_n \in A_n\big\}$ und beobachte, dass diese Mengen einen schnittstabilen Erzeuger der Produkt-σ-Algebra von $I^{\mathbb{N}_0}$ darstellen. Es genügt also, solche Mengen zu betrachten. Ausserdem genügt es für alle $m \in \mathbb{N}$ zu zeigen, dass

$$\mathbb{E}\big[1_{F\cap\{T=m\}}\mathbb{P}_{X_T}\big((X_n)_{n\geq 0} \in A\big)\big] = \mathbb{E}\big[1_{F\cap\{T=m\}}1_{\{(X_{T+n})_{n\geq 0}\in A\}}\big],$$

denn das obige Resultat folgt durch Summation über m. Nun ist $F \cap \{T = m\} \in \mathcal{F}_m$ und damit disjunkte Vereinigung von Mengen der Gestalt $\{X_0 = x_0, \ldots, X_m = x_m\}$. Es genügt daher zu zeigen, dass

$$\mathbb{E}\big[1_{\{X_0=x_0,\ldots,X_m=x_m\}}\mathbb{P}_{x_m}\big((X_n)_{n\geq 0} \in A\big)\big] = \mathbb{E}\big[1_{\{X_0=x_0,\ldots,X_m=x_m\}}1_{\{(X_{m+n})_{n\geq 0}\in A\}}\big].$$

Dies gilt offensichtlich, da beide Seiten gleich

$$\mathbb{P}\big(X_0 = x_0, \ldots, X_m = x_m, X_m \in A_0, \ldots, X_{m+n} \in A_n\big)$$

sind.

3.2.2 Treffwahrscheinlichkeiten und erwartete Rückkehrzeiten

Eine wichtige Methode, um Treffwahrscheinlichkeiten, erwartete Rückkehrzeiten und andere Größen zu berechnen ist die Variation des Startpunkts, die wir jetzt besprechen. Dazu benutzen wir, dass (X_n) auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P}_x)$ eine Markovkette mit Anfangswert $x \in I$ und Übergangsmatrix $P = (p_{ij}\colon i, j \in I)$ ist, das heißt für alle $x_1, \ldots, x_n \in I$,

$$\mathbb{P}_i\big\{X_1 = x_1, \ldots, X_n = x_n\big\} = p_{ix_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}.$$

Wir benutzen diese Notation auch für Erwartungswerte, das heißt $\mathbb{E}_i$ ist er Erwartungswert auf $(\Omega, \mathcal{A}, \mathbb{P}_i)$. Wir definieren die *Rückkehrzeit* T_j des Zustands $j \in I$ durch

$$T_j := \min\{n > 0 \,:\, X_n = j\}$$

und vereinbaren, dass $T_j := \infty$ wenn die Menge leer ist. Beachte, dass stets $T_j > 0$ und $\{T_j = \infty\} = \{X \text{ trifft } j \text{ nicht }\}$.

Satz 24 (Variation des Anfangswertes)**.** *Für jeden festen Zustand $b \in I$ sind*

$$x_i := \mathbb{P}_i\{T_b < \infty\}, \text{ für } i \in I,$$

die kleinste nichtnegative Lösung des Gleichungssystems

$$x_i = \Big(\sum_{j \neq b} p_{ij} x_j\Big) + p_{ib} \qquad \text{für } i \in I. \tag{3.1}$$

Insbesondere, wenn $y_i \geq 0$ das Gleichungssystem

$$y_i = \Big(\sum_{j \neq b} p_{ij} y_j\Big) + p_{ib}, \tag{3.2}$$

lösen, so gilt $y_i \geq x_i$ für alle $i \in I$.

Beachte: $y_i = 1$ für alle $i \in I$ ist stets eine Lösung von (3.2).

Beweis: Wir zeigen zunächst, dass $x_i = \mathbb{P}_i\{T_b < \infty\}$ eine Lösung von (3.1) ist. Sei $E = \{T_b < \infty\}$ das Ereignis, dass b getroffen wird. Dann gilt

$$\begin{aligned} x_i &= \mathbb{P}_i(E) = \sum_{j \in I} \mathbb{P}_i\big(E \cap \{X_1 = j\}\big) \\ &= \sum_{j \in I} \mathbb{P}_i\{X_1 = j\} \mathbb{P}_i(E \,|\, X_1 = j) \\ &= \sum_{j \in I} p_{ij} \mathbb{P}(E \,|\, X_1 = j). \end{aligned}$$

Wir haben

$$\mathbb{P}(E \,|\, X_1 = j) = \begin{cases} 1 & \text{if } j = b, \\ \mathbb{P}_j\{T_b < \infty\} & \text{if } j \neq b. \end{cases}$$

Daraus folgt

$$x_i = \sum_{j \neq b} p_{ij} x_j + p_{ib}.$$

Sei nun $y_i \geq 0$ eine Lösung von (3.2). Wir müssen zeigen, dass $y_i \geq x_i$. Es gilt

$$\begin{aligned} y_i &= \Big(\sum_{j \neq b} p_{ij} y_j\Big) + p_{ib} \\ &= p_{ib} + \sum_{j \neq b} p_{ij} \Big(\sum_{k \neq b} p_{jk} y_k + p_{jb}\Big) \\ &= p_{ib} + \sum_{j \neq b} p_{ij} p_{jb} + \sum_{j \neq b} \sum_{k \neq b} p_{ij} p_{jk} y_k \\ &= \mathbb{P}_i\{X_1 = b\} + \mathbb{P}_i\{X_1 \neq b, X_2 = b\} + \sum_{j \neq b} \sum_{k \neq b} p_{ij} p_{jk} y_k. \end{aligned}$$

Wiederholtes Einsetzen ergibt

$$\begin{aligned} y_i &= \mathbb{P}_i\{X_1 = b\} + \mathbb{P}_i\{X_1 \neq b, X_2 = b\} + \dots \\ &\quad + \mathbb{P}_i\{X_1 \neq b, \dots, X_{n-1} \neq b, X_n = b\} \\ &\quad + \sum_{j_1 \neq b} \cdots \sum_{j_n \neq b} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} y_{j_n}. \end{aligned}$$

Da $y_j \geq 0$ ist der letzte Term positiv und wir erhalten, für alle $n \in \mathbb{N}$,

$$\begin{aligned} y_i &\geq \mathbb{P}_i\{X_1 = b\} + \mathbb{P}_i\{X_1 \neq b, X_2 = b\} + \dots \\ &\quad + \mathbb{P}_i\{X_1 \neq b, \dots, X_{n-1} \neq b, X_n = b\} \\ &= \mathbb{P}_i\{T_b = 1\} + \dots + \mathbb{P}_i\{T_b = n\} \\ &= \mathbb{P}_i\{T_b \leq n\}. \end{aligned}$$

Wenn $n \to \infty$ folgt

$$y_i \geq \lim_{n \to \infty} \mathbb{P}_i\{T_b \leq n\} = \mathbb{P}_i\Big(\bigcup_n \{T_b \leq n\} \Big) = \mathbb{P}_i\{T_b < \infty\} = x_i.$$

Bemerkung: Auf ähnliche Weise kann man auch erwartete Rückkehrzeiten ausrechnen. Ist nämlich $b \in I$ und $y_i := \mathbb{E}_i T_b$, so ist

$$y_i = 1 + \sum_{j \neq b} p_{ij} y_j$$

und y ist die kleinste nichtnegative Lösung dieses Gleichungssystems. Beachte dabei dass $y_i = \infty$ zulässig ist für einige oder alle i und die Vereinbarung $0 \times \infty = 0$ gilt.

Beispiel 3.3. Wir betrachten die einfache Irrfahrt und berechnen $y_i = \mathbb{E}[T_0]$, die mittlere Wartezeit bis wir 0 erreichen wenn wir in $i > 0$ starten. Beachte, dass wir zusätzlich die Gleichung $y_n = n y_1$ haben für alle $n \geq 1$. Zusammen mit unserer Gleichung für $i = 1$ erhalten wir

$$y_1 = 1 + p y_2 = 1 + 2p y_1.$$

Wenn $p < 1/2$ (Drift zur null hin) ergibt sich

$$\mathbb{E}_n[T_0] = \frac{n}{1 - 2p} \text{ für alle } n \geq 1.$$

Wenn $p \geq 1/2$ gilt $y_1 \geq 1 + y_1$, und somit $\mathbb{E}_n[T_0] = \infty$.

3.2.3 Rekurrente und transiente Zustände

Wir betrachten nun das Langzeitverhalten von Markovketten. Setze wieder $T_i = \inf\{n > 0 : X_n = i\}$. Der Zustand i heißt

- *transient* wenn $\mathbb{P}_i\{T_i < \infty\} < 1$, das heißt es gibt eine positive Wahrscheinlichkeit, dass der Zustand i nicht wieder besucht wird. Insbesondere wird der Zustand i fast sicher nur endlich oft besucht.

- *rekurrent* wenn $\mathbb{P}_i\{T_i < \infty\} = 1$, das heißt der Prozess kehrt fast sicher in den Zustand i zurück und besucht ihn daher fast sicher unendlich oft.

Ein rekurrenter Zustand i heißt

- *positiv rekurrent* wenn $\mathbb{E}_i T_i < \infty$, d.h. die erwartete Rückkehrzeit in den Zustand i ist endlich,

- *nullrekurrent* wenn $\mathbb{E}_i[T_i] = \infty$.

Zum Beispiel habe wir bereits herausgefunden, dass für die symmetrische einfache Irrfahrt der Zustand 0 (und jeder andere Zustand) nullrekurrent ist, da $T_0 < \infty$ fast sicher, aber $\mathbb{E}_0 T_0 = \infty$.

3.2.4 Irreduzibilität

Wir sagen, eine Markovkette ist *irreduzibel* wenn für je zwei Zustände $i, j \in I$ ein n existiert mit $\mathbb{P}_i(X_n = j) > 0$.

Lemma 3.1. *Ist (X_n) irreduzibel und der Zustand $j \in I$ rekurrent, so gilt für alle Startverteilungen, dass $\mathbb{P}(T_j < \infty) = 1$.*

Beweis: Nach der Markoveigenschaft gilt

$$\begin{aligned}\mathbb{P}(T_j < \infty) &= \mathbb{E}\big[\mathbb{P}(T_j < \infty | \mathcal{F}_0)\big] = \mathbb{E}\big[\mathbb{P}_{X_0}(T_j < \infty)\big] \\ &= \sum_{i \in I} \mathbb{P}(X_0 = i)\mathbb{P}_i(T_j < \infty).\end{aligned}$$

Es genügt somit zu zeigen, dass $\mathbb{P}_i(T_j < \infty) = 1$ für alle $i \in I$. Aufgrund der Irreduzibilität existiert ein m mit $\mathbb{P}_j(X_m = i) > 0$. Dann gilt

$$\begin{aligned}1 &= \mathbb{P}_j\big(X_n = j \text{ für unendlich viele } n\big) \\ &\le \mathbb{P}_j\big(X_n = j \text{ für ein } n \ge m+1\big) \\ &= \sum_{k \in I} \mathbb{P}_j\big(X_n = j \text{ für ein } n \ge m+1 \big| X_m = k\big)\mathbb{P}_j(X_m = k) \\ &= \sum_{k \in I} \mathbb{P}_k\big(T_j < \infty\big)\mathbb{P}_j(X_m = k).\end{aligned}$$

Da $\sum_{k \in I} \mathbb{P}_j(X_m = k) = 1$ folgt, dass $\mathbb{P}_k(T_j < \infty) = 1$ für alle k.

3.3 Invariante Verteilung, Reversibilität und Gleichgewicht

3.3.1 Invariante Verteilungen

Sei π eine Wahrscheinlichkeitsverteilung auf der diskreten Menge I, d.h. $\pi\colon I \to [0,1]$ with $\sum_{i\in I} \pi_i = 1$. π heißt *invariante* oder *stationäre Verteilung* wenn $\pi P = \pi$, also

$$\sum_{i\in I} \pi_i p_{ij} = \pi_j \text{ für alle } j.$$

Wenn wir ein solches π als Startverteilung wählen, so gilt

$$\mathbb{P}\{X_1 = j\} = \sum_{i\in I} \mathbb{P}\{X_0 = i, X_1 = j\} = \sum_{i\in I} \pi_i p_{ij} = \pi_j = \mathbb{P}\{X_0 = j\},$$

und des weiteren

$$\mathbb{P}\{X_n = j\} = \pi_j \text{ für alle } j \in I.$$

Die Verteilung von X_n ist also gleich π für alle Zeiten n, man sagt der Prozess ist *stationär.*

Satz 25. *Ist (X_n) eine irreduzible Markovkette. Dann sind die folgenden Aussagen äquivalent:*

(i) jeder Zustand $k \in I$ ist positiv rekurrent,

(ii) ein Zustand $k \in I$ ist positiv rekurrent,

(iii) es existiert eine invariante Verteilung $(\pi_i\colon i \in I)$.

Falls sie existiert ist die invariante Verteilung eindeutig und gegeben durch

$$\pi_i = \frac{1}{\mathbb{E}_i[T_i]}.$$

Gilt eine, und damit alle der Bedingungen $(i), (ii), (iii)$ so nennen wir die Markovkette (X_n) *positiv rekurrent.*

Zum Beweis des Satzes zeigen wir zunächst, dass für einen festen rekurrenten Zustand k die erwartete Zeit, die die Kette zwischen zwei Besuchen von k im Zustand i verbringt

$$v_i^{(k)} = \mathbb{E}_k \sum_{n=1}^{T_k} 1_{\{X_n = i\}}$$

die Gleichung $v^{(k)}P = v^{(k)}$ erfüllt. Es gilt nämlich

$$\begin{aligned}
v_i^{(k)} &= \mathbb{E}_k \sum_{n=1}^{\infty} 1_{\{X_n=i, n\le T_k\}} \\
&= \sum_{n=1}^{\infty} \sum_{j\in I} \mathbb{P}_k\{X_{n-1}=j, X_n=i, n\le T_k\} \\
&= \sum_{j\in I} p_{ji} \sum_{n=1}^{\infty} \mathbb{P}_k\{X_{n-1}=j, n\le T_k\} \\
&= \sum_{j\in I} p_{ji} \mathbb{E}_k \sum_{n=0}^{T_k-1} \mathbb{P}_k\{X_m=j\} = \sum_{j\in I} v_j^{(k)} p_{ji}.
\end{aligned}$$

Da $\sum_i v_i^{(k)} = \mathbb{E}_k T_k < \infty$ definiert

$$\pi_i := \frac{v_i^{(k)}}{\mathbb{E}_k T_k}$$

eine invariante Verteilung und somit folgt $(ii) \Rightarrow (iii)$.

Sei nun $(\pi_i\colon i\in I)$ eine invariante Verteilung und $k \in I$. Da die Kette irreduzibel ist, gibt es ein n mit $\mathbb{P}_i(X_n = k) > 0$. Dann folgt

$$\pi_k = \sum_{i\in I} \pi_i \mathbb{P}_i(X_n = k) > 0.$$

Also definiert $\lambda_i = \pi_i/\pi_k$ eine Lösung von $\lambda P = \lambda$ und wir zeigen jetzt, dass $\lambda_j \ge v_j^{(k)}$ für alle $j \in I$. Es gilt nämlich

$$\begin{aligned}
\lambda_j &= \sum_{i_0\in I} \lambda_{i_0} p_{i_0 j} = \sum_{i_0\ne k} \lambda_{i_0} p_{i_0 j} + p_{kj} \\
&= \sum_{i_0, i_1\ne k} \lambda_{i_1} p_{i_1 i_0} p_{i_0 j} + p_{kj} + \sum_{i_0\ne k} \lambda_k p_{k i_0} p_{i_0 j}.
\end{aligned}$$

Iteriert man das n mal und lässt den nichtnegativen ersten Term weg, so erhält man

$$\begin{aligned}
\lambda_j &\ge p_{kj} + \sum_{i_0\ne k} \lambda_k p_{k i_0} p_{i_0 j} + \dots + \sum_{i_0,\dots,i_{n-1}\ne k} \lambda_k p_{k i_{n-1}} p_{i_{n-1} i_{n-2}} \cdots p_{i_1 i_0} p_{i_0 j} \\
&= \mathbb{P}_k(X_1 = j, T_k \ge 1) + \mathbb{P}_k(X_2 = j, T_k \ge 2) + \dots + \mathbb{P}_k(X_n = j, T_k \ge n) \\
&= \mathbb{E}_k \sum_{m=1}^{T_k \wedge n} 1_{\{X_m = j\}} \to v_j^{(k)}.
\end{aligned}$$

Folglich ist

$$\mathbb{E}_k T_k = \sum_{j\in I} v_j^{(k)} \le \sum_{j\in I} \frac{\pi_j}{\pi_k} = \frac{1}{\pi_k} < \infty$$

und somit ist k positiv rekurrent und $(iii) \Rightarrow (i)$.

Da offensichtlich $(i) \Rightarrow (ii)$ bleibt nur zu zeigen, dass in der letzten Ungleichungskette Gleichheit gilt, das heißt es gilt $v_i^{(k)} = \lambda_i$ für alle $i \in I$. Da wir wissen, dass k rekurrent ist erfüllt der Vektor μ gegeben durch $\mu_j = \lambda_j - v_j^{(k)} \ge 0$ auch die Gleichung $\mu P = \mu$. Aufgrund der Irreduzibilität gibt es für jedes i ein n mit $\mathbb{P}_i(X_n = k) > 0$ und damit gilt

$$0 = \mu_k = \sum_{j\in I} \mu_j \mathbb{P}_j(X_n = k) \ge \mu_i \mathbb{P}_i(X_n = k),$$

woraus folgt, dass $\mu_i = 0$ und somit $v_i^{(k)} = \lambda_i$, wie gefordert

3.3.2 Reversibilität und Gleichgewicht

Sei $m = (m_i : i \in I)$ ein Vektor nichtnegativer Zahlen, die nicht alle gleich null sind. Eine Markovkette heißt *m-symmetrisierbar* wenn die *Gleichgewichtsgleichungen* gelten, das heißt

$$m_i p_{ij} = m_j p_{ji} \text{ für alle } i, j \in I.$$

Eine Kette heißt *reversibel*, wenn es ein m wie oben gibt, das die Kette m-symmetrisierbar macht. Aus den Gleichgewichtsgleichungen folgt

$$\sum_{i\in I} m_i p_{ij} = m_j \sum_{i\in I} p_{ji} = m_j \text{ für alle } j \in I.$$

Wenn dann $M = \sum_{i\in I} m_i < \infty$, so gilt

- Die Verteilung π gegeben durch $\pi_i = \frac{m_i}{M}$ löst ebenfalls die Gleichgewichtsgleichungen und ist daher eine invariante Verteilung. Sie heißt *Äquilibrium* oder *Gleichgewichtsverteilung* der Markovkette.

- Wird (X_n) mit Anfangsverteilung π gestartet, so gilt

$$\mathbb{P}\big\{X_0 = i_0, \ldots, X_n = i_n\big\} = \mathbb{P}\big\{X_0 = i_n, \ldots, X_n = i_0\big\},$$

  das heißt $X_0, \ldots, X_n$ hat dieselbe Verteilung wie $X_n, \ldots, X_0$.

Bemerkungen:

- Ist (X_n) π-symmetrisierbar für eine Wahrscheinlichkeitsverteilung π, so ist π eine invariante Verteilung von X. Aber nicht jede invariante Verteilung löst die Gleichgewichtsgleichungen und symmetrisiert (X_n).

- Ist eine Markovkette irreduzibel und die invariante Verteilung löst die Gleichgewichtsgleichungen nicht, so haben sie keine zulässige Lösung.

Beispiel 3.4. Sei (X_n) eine Markovkette mit Zustandsraum $I = \{0, 1, 2\}$ und Übergangsmatrix

$$P = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/6 & 1/3 \\ 1/2 & 1/6 & 1/3 \end{pmatrix}.$$

Diese Kette ist nicht reversibel. Die Gleichgewichtsgleichungen sind

$$\begin{aligned} m_0 p_{01} &= m_1 p_{10} &\Leftrightarrow\quad m_0 &= 2m_1 \\ m_0 p_{02} &= m_2 p_{20} &\Leftrightarrow\quad m_0 &= 2m_2 \\ m_1 p_{12} &= m_2 p_{21} &\Leftrightarrow\quad m_2 &= 2m_1. \end{aligned}$$

Die einzige Lösung $m_0 = m_1 = m_2 = 0$ ist nicht zulässig. Aber es gibt eine invariante Verteilung, nämlich $\pi = (1/2, 5/24, 7/24)$.

3.4 Der Ergodensatz

Wir beweisen nun den Satz, der das fast sichere Langzeitverhalten einer Markovkette beschreibt. Der Satz heißt Ergodensatz, weil er ein zeitliches Mittel mit einem räumlichen Mittel gleichsetzt.

Satz 26 (Ergodensatz für Markovketten)**.** *Sei (X_n) eine irreduzible Markovkette mit beliebiger Startverteilung und $V_i(n)$ die Zeit, die die Kette bis zur Zeit $n-1$ im Zustand i verbringt. Dann gilt, fast sicher*

$$\lim_{n\to\infty} \frac{V_i(n)}{n} = \frac{1}{m_i},$$

wobei $m_i = \mathbb{E}_i[T_i]$ die erwartete Rückkehrzeit in den Zustand i ist.
Ist die Markovkette zudem positive rekurrent so gilt für jede beschränkte Funktion $f\colon I \to \mathbb{R}$, dass

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \int f \, d\pi,$$

wobei π die eindeutige invariante Verteilung der Markovkette ist.

Beweis: Ist der Zustand i transient, so ist die Gesamtzahl V_i der Besuche in i endlich und somit gilt

$$\frac{V_i(n)}{n} \le \frac{V_i}{n} \to 0 = \frac{1}{m_i}.$$

Nun nehmen wir an, der Zustand i sei rekurrent. Dann gilt $\mathbb{P}(T_i < \infty) = 1$ nach Lemma 3.1 und nach der starken Markoveigenschaft ist $(X_{T_i+n} : n \geq 0)$ eine Markovkette mit Übergangsmatrix P und Anfangswert i. Da T_i endlich ist, ergibt der gesuchte Limes denselben Wert für die Markovketten (X_n) und (X_{T+n}). Wir können daher von nun an annehmen, dass (X_n) im Zustand i gestartet wird.

Seien nun $0 = \tau_0, \tau_1, \tau_2, \ldots$ die Folge der Besuchszeiten von Zustand i. Dies ist eine Folge von Stoppzeiten und nach der starken Markoveigenschaft sind die *Exkursionslängen* $S_n := \tau_n - \tau_{n-1}$ für $n \in \mathbb{N}$ unabhängig identisch verteilte Zufallsvariablen mit Erwartungswert $m_i = \mathbb{E}_i[T_i]$. Es gilt

$$S_1 + \cdots + S_{V_i(n)} \geq n \geq S_1 + \cdots + S_{V_i(n)-1},$$

und folglich

$$\frac{S_1 + \cdots + S_{V_i(n)-1}}{V_i(n)} \leq \frac{n}{V_i(n)} \leq \frac{S_1 + \cdots + S_{V_i(n)}}{V_i(n)}.$$

Da die Markovkette rekurrent ist, gilt $V_i(n) \to \infty$ fast sicher und somit nach dem starken Gesetz der grossen Zahlen

$$\frac{S_1 + \cdots + S_{V_i(n)}}{V_i(n)} \to m_i \text{ fast sicher.}$$

Damit konvergieren die linke und rechte Seite in der letzten Ungleichungskette beide gegen m_i und somit auch der mittlere Term, das heisst

$$\frac{n}{V_i(n)} \to m_i \text{ fast sicher.}$$

Daraus folgt, wie gefordert, dass

$$\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ fast sicher.}$$

Nun sei (X_n) positiv rekurrent. Dann ist nach Satz 25 die eindeutige invariante Verteilung durch $(\pi_i = \frac{1}{m_i} : i \in I)$ gegeben. Sei $f : I \to \mathbb{R}$ durch eine

Konstante $K > 0$ beschränkt. Für jedes endliche $J \subseteq I$ gilt

$$\begin{aligned}
\Big|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \sum_{i\in I} f(i)\pi_i\Big| &= \Big|\sum_{i\in I}\Big(\frac{V_i(n)}{n} - \pi_i\Big)f(i)\Big| \\
&\le K\sum_{j\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + K\sum_{j\notin J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| \\
&\le K\sum_{j\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + K\sum_{j\notin J}\Big(\frac{V_i(n)}{n} + \pi_i\Big) \\
&= K\sum_{j\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + K\sum_{j\in J}\Big(\pi_i - \frac{V_i(n)}{n}\Big) + 2K\sum_{j\notin J}\pi_i \\
&\le 2K\sum_{j\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + 2K\sum_{j\notin J}\pi_i.
\end{aligned}$$

Gegeben $\varepsilon > 0$ wähle zuerst J endlich, so dass $\sum_{j\notin J}\pi_i < \varepsilon/(4K)$ und anschließend N so, dass $\sum_{j\in J}|V_i(n)/n - \pi_i| < \varepsilon/(4K)$ für $n \ge N$. Dann folgt, dass

$$\Big|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \int f\,d\pi\Big| < \varepsilon,$$

wie gefordert.