Vorlesung Stochastik II

Christof Schütte
auf Grundlage eines Kurzskripts von C. Lasser

Sommersemester 2011

Inhaltsverzeichnis

1 Werkzeug 1: Maß- und Integrationstheorie

Literatur: [E, Dei]

1.1 Maßproblem, Mengensysteme, Maßraum

Definition 1 (Kongruenz) Zwei Mengen $A, B \subseteq \mathbb{R}^n$ heißen kongruent, falls es eine orthogonale Matrix $U \in \mathbb{R}^{n\times n}$ und einen Vektor $v \in \mathbb{R}^n$ gibt mit $B = U(A) + v$.

Satz 1 (Vitali, 1905) Sei $n \in \mathbb{N}$. Es gibt keine Abbildung $\mu : \mathcal{P}(\mathbb{R}^n) \to [0, \infty]$, welche die folgenden Eigenschaften erfüllt:

i) σ-Additivität: Für jede Folge $(A_i)_{i\in\mathbb{N}}$ paarweiser disjunkter Mengen $A_i \subseteq \mathbb{R}^n$ gilt

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

ii) Bewegungsinvarianz: Für kongruente Mengen $A, B \subset \mathbb{R}^n$ gilt $\mu(A) = \mu(B)$.
iii) Normiertheit: $\mu([0, 1]^n) = 1$

Definition 2 (σ-Algebra) Sei Ω eine Menge. Ein Mengensystem $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt σ-Algebra über Ω, falls gilt:

i) $\Omega \in \mathcal{A}$
ii) $\forall A \in \mathcal{A} : A^c \in \mathcal{A}$
iii) Für jede Folge $(A_i)_{i\in\mathbb{N}}$ in $\mathcal{A}$ ist $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Definition 3 (Erzeugte σ-Algebra) Sei $M \subseteq \mathcal{P}(\Omega)$ ein Mengensystem, Σ die Menge aller σ-Algebren über Ω, die M enthalten. Dann ist

$$\sigma(M) = \bigcap_{\mathcal{A}\in\Sigma} \mathcal{A}$$

die kleinste σ-Algebra, die M enthält. Sie heisst die von M erzeugte σ-Algebra.

Definition 4 (Durchschnittsstabil) Ein Mengensystem $M \subseteq \mathcal{P}(\Omega)$ heisst durchschnittsstabil, wenn aus $A, B \in M$ folgt, dass $A \cap B \in M$.

Proposition 1 Eine σ-Algebra ist durchschnittsstabil.

Proposition 2 (Dynkin-Trick) Sei $M \subseteq \mathcal{P}(\Omega)$ ein Mengensystem und $D_0 \subseteq \mathcal{P}(\Omega)$ das von M erzeugte Dynkin-System, d.h. die kleinste Menge, die alle abzählbaren Vereinigungen $\cup_i A_i$ disjunkter $A_i \in D_0$ ebenfalls wieder enthält. Wenn D_0 durchschnittsstabil ist, dann gilt $D_0 = \sigma(M)$.

Definition 5 (Borel'sche σ-Algebra) Die Borel'sche σ-Algebra $\mathcal{B}(\mathbb{R}^n)$ über $\mathbb{R}^n$ ist die vom System der offenen Mengen $\mathcal{O}$ auf $\mathbb{R}^n$ erzeugte σ-Algebra.

Proposition 3 (Borel'sche σ-Algebra) Die Borel'sche σ-Algebra über $\mathbb{R}^n$ wird auch von folgenden Mengensystemen erzeugt:

$$\begin{aligned} \mathcal{C} &:= \{A \subseteq \mathbb{R}^n : A \text{ abgeschlossen}\}, \\ \mathcal{I} &:= \{]a,b] : a,b \in \mathbb{R}^n, \forall i : a_i \leq b_i\}, \\ \mathcal{I}_\infty &:= \{]-\infty, c] : c \in \mathbb{R}^n\}. \end{aligned}$$

Definition 6 Die Borel'sche σ-Algebra $\mathcal{B}[-\infty,\infty] = \mathcal{B}(\mathbb{R} \cup \{\pm\infty\})$ ist definiert als $\mathcal{B}[-\infty,\infty] = \sigma(\mathcal{B}(\mathbb{R}) \cup \{\pm\infty\})$.

Definition 7 (Maß, Maßraum, σ-endlich) Sei $\mathcal{A}$ eine σ-Algebra auf Ω. Eine Funktion $\mu : \mathcal{A} \to [0,\infty]$ heißt Maß auf $\mathcal{A}$, falls folgendes gilt:
i) Normiertheit: $\mu(\emptyset) = 0$,
ii) σ-Additivität: Für jede Folge $(A_i)_{i\in\mathbb{N}}$ paarweiser disjunkter Mengen aus $\mathcal{A}$ ist

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

Das Tripel $(\Omega, \mathcal{A}, \mu)$ heißt Maßraum.
Gibt es $(A_i)_{i\in\mathbb{N}}$ in $\mathcal{A}$ mit $\bigcup_{i=1}^{\infty} A_i = \Omega$, $\forall i \in \mathbb{N} : \mu(A_i) < \infty$, so heißt μ σ-endlich.

Beispiel 1 Das Dirac-Maß $\delta_\omega : \mathcal{A} \to [0,\infty]$ in $\omega \in \Omega$ ist definiert durch

$$\delta_\omega(A) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{sonst} \end{cases}$$

Beispiel 2 Das Zähl-Maß $\mu : \mathcal{A} \to [0,\infty]$ ist definiert durch

$$\mu(A) = \begin{cases} |A| & \text{falls } A \text{ endlich} \\ \infty & \text{sonst} \end{cases}$$

1.2 Lebesgue-Maß und -Integral

Lemma 1 (Stetigkeit) Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $A, A_n \in \mathcal{A}$ für $n \in \mathbb{N}$.
i) Stetigkeit von unten: Falls $A_1 \subseteq A_2 \subseteq \dots$ und $A = \bigcup_{n=1}^{\infty} A_n$, so gilt $\mu(A_n) \uparrow \mu(A)$ für $n \to \infty$.
ii) Stetigkeit von oben: Falls $A_1 \supseteq A_2 \supseteq \dots$, $A = \bigcap_{n=1}^{\infty} A_n$ und $\mu(A_1) < \infty$, so gilt $\mu(A_n) \downarrow \mu(A)$ für $n \to \infty$.

Satz 2 (Lebesgue-Maß)
i) Es gibt genau ein Maß $\lambda : \mathcal{B}(\mathbb{R}^n) \to [0,\infty]$, so daß für alle $a, b \in \mathbb{R}^n$ gilt $\lambda(]a,b]) = \prod_{i=1}^{n}(b_i - a_i)$.
ii) λ ist bewegungsinvariant.
iii) λ ist σ-endlich, aber nicht endlich.
iv) Für abzählbares $A \in \mathcal{B}(\mathbb{R}^n)$ gilt $\lambda(A) = 0$.

Definition 8 (Meßbar) Sind $\mathcal{A}_1, \mathcal{A}_2$ σ-Algebren über Ω_1, Ω_2, so heißt eine Abbildung $f : \Omega_1 \to \Omega_2$ $\mathcal{A}_1$-$\mathcal{A}_2$-meßbar, falls $f^{-1}(\mathcal{A}_2) \subseteq \mathcal{A}_1$.

Lemma 2 (Meßbar auf Erzeuger) Seien $\mathcal{A}_1, \sigma(\mathcal{M})$ σ-Algebren über Ω_1, Ω_2. Eine Abbildung $f : \Omega_1 \to \Omega_2$ ist genau dann $\mathcal{A}_1$-$\sigma(\mathcal{M})$-meßbar, wenn $f^{-1}(\mathcal{M}) \subseteq \mathcal{A}_1$.

Definition 9 (Bildmaß) Ist $(\Omega_1, \mathcal{A}_1, \mu)$ ein Maßraum, $\mathcal{A}_2$ eine σ-Algebra über Ω_2 und $f : \Omega_1 \to \Omega_2$ meßbar, so heißt $\mu_f : \mathcal{A}_2 \to [0, \infty]$, $A \mapsto \mu(f^{-1}(A))$ das Bildmaß von μ unter f.

Definition 10 (Treppenfunktion) Sei $(\Omega, \mathcal{A}, \mu)$ Maßraum. Sind $A_1, \ldots, A_n \in \mathcal{A}$ paarweise disjunkte meßbare Mengen und $y_1, \ldots, y_n \in \mathbb{R}$ für $n \in \mathbb{N}$, so heißt

$$f = y_1 \chi_{A_1} + \ldots + y_n \chi_{A_n}$$

eine Treppenfunktion, wobei χ_A die Indikatorfunktion der Menge $A \subseteq \Omega$ bezeichnet, d.h. $\chi_A(\omega) = 1$ falls $\omega \in A$ und $= 0$ sonst. Ist f eine nichtnegative Treppenfunktion, so ordnet man ihr das Lebesgue-Integral

$$\int f \mathrm{d}\mu = y_1 \mu(A_1) + \ldots + y_n \mu(A_n)$$

zu.

Lemma 3 (Approximation durch Treppenfunktionen) Jede nichtnegative meßbare Funktion $f : \Omega \to [0, \infty]$ ist durch eine monotone Folge von nichtnegativen Treppenfunktionen $(f_n)_{n \in \mathbb{N}}$ punktweise approximierbar: $f_n \uparrow f$ für $n \to \infty$.

1.3 Lebesgue-Integral und Konvergenzsätze

Definition 11 (Integral für nichtnegative meßbare Funktionen) Einer nichtnegativen meßbaren Funktion $f : \Omega \to [0, \infty]$ ordnet man das Integral

$$\int f \mathrm{d}\mu = \lim_{n \to \infty} \int f_n \mathrm{d}\mu$$

zu, wobei $(f_n)_{n \in \mathbb{N}}$ eine Folge nichtnegativer Treppenfunktionen mit $f_n \uparrow f$ ist.

Definition 12 (Integral für meßbare Funktionen) Für ein meßbare Funktion $f : \Omega \to [-\infty, +\infty]$ bezeichnen $f^+ = \max(f, 0)$ und $f^- = \max(-f, 0)$ den Positiv- und Negativteil. Falls $\int f^+ \mathrm{d}\mu < \infty$ und $\int f^- \mathrm{d}\mu < \infty$ gilt, so heißt f integrierbar mit Integral

$$\int f \mathrm{d}\mu = \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu.$$

Gilt entweder $\int f^+ \mathrm{d}\mu = \infty$ oder $\int f^- \mathrm{d}\mu = \infty$, so definieren wir $\int f \mathrm{d}\mu$ genauso und erlauben die Werte $\pm\infty$.
Falls μ das Lebesgue-Maß bezeichnet, $\mu = \lambda$, dann heisst das so definierte Integral auch Lebesgue-Integral.

Satz 3 (Beziehung zwischen Riemann- & Lebesgue-Integral)
Sei $I \subseteq \mathbb{R}$ ein Intervall, $f : I \to \mathbb{R}$ eine meßbare Funktion.
i) Sei $I = [a, b]$ und f Riemann-integrierbar. Dann ist f Lebesgue-integrierbar und es gilt $\int f \mathrm{d}\mu = \int_a^b f(x)\mathrm{d}x$.
ii) Sei f auf jedem kompakten Teilinterval von I Riemann-integrierbar. f ist genau dann Lebesgue-integrierbar, wenn $|f|$ uneigentlich Riemann-integrierbar ist. In diesem Fall gilt $\int f \mathrm{d}\lambda = \int_I f(x)\mathrm{d}x$.

Proposition 4 (Linearität und Monotonie) Sind $f, g : \Omega \to [-\infty, +\infty]$ integrierbar und $\alpha, \beta \in \mathbb{R}$, so gilt
i) $\alpha f + \beta g$ ist integrierbar und $\int(\alpha f + \beta g)\mathrm{d}\mu = \alpha \int f \mathrm{d}\mu + \beta \int g \mathrm{d}\mu$.
ii) Aus $f \leq g$ folgt $\int f \mathrm{d}\mu \leq \int g \mathrm{d}\mu$.

Satz 4 (Monotone Konvergenz, Beppo Levi 1906) Für jede monoton wachsende Folge integrierbarer Funktionen $(f_n)_{n\in\mathbb{N}}$ gilt $\lim_{n\to\infty} f_n \mathrm{d}\mu = \int \lim_{n\to\infty} f_n \mathrm{d}\mu$, wobei beide Seiten die Werte $+\infty$ annehmen können.

Bemerkung 1 Die Eigenschaft, dass die Folge monoton wachsend sein muss, ist essentiell, wie man am Beispiel der Folge $f_n = \chi_{[0,n]}/n$ mit $f_n \to 0$ aber $\int f_n \mathrm{d}\mu = 1$ für alle $n \in \mathbb{N}$ sieht.

Korollar 1 (Lemma von Fatou, 1906) Es sei f integrierbar und $(f_n)_{n\in\mathbb{N}}$ eine Folge meßbarer Funktionen mit $f_n \geq f$ für alle $n \in \mathbb{N}$. Dann gilt

$$\int \liminf_{n\to\infty} f_n \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \mathrm{d}\mu.$$

Satz 5 (Dominierte Konvergenz, Lebesgue 1910) Seien $(f_n)_{n\in\mathbb{N}}, (g_n)_{n\in\mathbb{N}}$ Folgen meßbarer Funktionen mit (1) $f = \lim_{n\to\infty} f_n$ und (2) $g = \lim_{n\to\infty} g_n$ und mit (3) $|f_n| \leq g_n$ für alle $n \in \mathbb{N}$. Sind nun g_n und g integrierbar und gilt $\lim_{n\to\infty} \int g_n \mathrm{d}\mu \to \int g \mathrm{d}\mu$, dann sind auch alle f_n und f integrierbar, und es gilt

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu = \int f \mathrm{d}\mu, \qquad \lim_{n\to\infty} \int |f_n - f| \mathrm{d}\mu = 0.$$

Die Aussage bleibt richtig, wenn die Bedingungen (1), (2) und (3) nur μ fast sicher erfüllt sind, d.h. wenn z.B. nur (3)' $\mu(\{|f_n| > g_n\}) = 0$ statt (3) gilt.

1.4 Satz von Fubini & Radon-Nikodym, L^p-Räume

Proposition 5 (Produktmaß)
Zu σ-endlichen Maßräumen $(\Omega_1, \mathcal{A}_1, \mu)$ und $(\Omega_2, \mathcal{A}_2, \nu)$ gibt es genau ein Maß $\mu \otimes \nu$ auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, so daß

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$$

für alle $A \in \mathcal{A}_1$ und $B \in \mathcal{A}_2$ gilt. Außerdem sind für jedes $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$ die Funktionen $\omega_1 \mapsto \int \chi_C(\omega_1, \omega_2)\mathrm{d}\nu(\omega_2)$ und $\omega_2 \mapsto \int \chi_C(\omega_1, \omega_2)\mathrm{d}\mu(\omega_1)$ meßbar, und es gilt

$$(\mu \otimes \nu)(C) = \int \int \chi_C(\omega_1, \omega_2)\mathrm{d}\mu(\omega_1)\mathrm{d}\nu(\omega_2) = \int \int \chi_C(\omega_1, \omega_2)\mathrm{d}\nu(\omega_2)\mathrm{d}\mu(\omega_1).$$

Satz 6 (Fubini, 1907) Es seien $(\Omega_1, \mathcal{A}_1, \mu)$ und $(\omega_2, \mathcal{A}_2, \nu)$ σ-endliche Maßräume. Ist $f : \Omega_1 \times \Omega_2 \to [-\infty, +\infty]$ eine nichtnegative meßbare oder eine $(\mu \otimes \nu)$-integrierbare Funktion, so gilt

$$\int f \mathrm{d}(\mu \otimes \nu) = \int \int f(\omega_1, \omega_2) \mathrm{d}\mu(\omega_1) \mathrm{d}\nu(\omega_2) = \int \int f(\omega_1, \omega_2) \mathrm{d}\nu(\omega_2) \mathrm{d}\mu(\omega_1).$$

Satz 7 (Hölder- & Minkowski-Ungleichung) Definiere für eine meßbare Funktion $f : \Omega \to [-\infty, \infty]$ und $p \in [1, \infty[$

$$\|f\|_p := \left(\int |f|^p \mathrm{d}\mu \right)^{1/p}, \qquad \|f\|_\infty := \inf\{K \geq 0 : \mu(\{|f| > K\}) = 0\}.$$

Ist $g : \Omega \to [-\infty, \infty]$ ebenfalls meßbar, so gilt:

i) Für $p, q \in [1, \infty]$ mit $1/p + 1/q = 1$ ist $\|fg\|_1 \leq \|f\|_p \|g\|_q$.
ii) Für $p \in [1, \infty]$ ist $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

Definition 13 ($\mathcal{L}^p$- und L^p-Räume) Setze für $p \in [1, \infty]$

$$\begin{aligned} \mathcal{L}^p(\mu) &:= \{f : \Omega \to [-\infty, +\infty],\ f \text{ meßbar},\ \|f\|_p < \infty\} \\ L^p(\mu) &:= \mathcal{L}^p / \{f : \Omega \to [-\infty, +\infty],\ f \text{ meßbar},\ \mu(f \neq 0) = 0\} \end{aligned}$$

Bemerkung 2
i) $(L^p(\mu), \|\cdot\|_p)$ ist ein Banachraum.
ii) $(L^2(\mu), \|\cdot\|_2)$ ist ein Hilbertraum mit Skalarprodukt $\langle f, g \rangle = \int fg \mathrm{d}\mu$.
iii) Falls $\mu(\Omega) < \infty$ und $1 \leq p \leq q$, dann existiert ein $c > 0$ so dass $\|f\|_p \leq c\|f\|_q$ und $L^q(\mu) \subseteq L^p(\mu)$.

Definition 14 (Konvergenz) Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

i) Eine Folge von Funktionen $(f_n)_{n \in \mathbb{N}}$ in L^p heißt L^p-konvergent, wenn es ein $f \in L^p$ gibt mit

$$\lim_{n \to \infty} \|f_n - f\|_p = 0.$$

ii) Eine Folge von meßbaren Funktionen $(f_n)_{n \in \mathbb{N}}$ heißt μ-fast sicher konvergent, wenn es eine meßbare Funktion f gibt mit

$$\mu(\{\lim_{n \to \infty} f_n \neq f\}) = 0.$$

Proposition 6 (Konvergenzkriterium) Sei $p \in [1, \infty]$ und $f, f_n \in L^p(\mu)$, $n \in \mathbb{N}$ mit $f_n \to f$ μ-fast sicher. Dann gilt $f_n \to f$ in $L^p(\mu)$ genau dann, wenn $\|f_n\|_p \to \|f\|_p$.

Der Beweis ist in der $\Rightarrow$ Richtung mittels der Minkowski-Ungleichung trivial. Für die Rückrichtung definiere $g_n = 2^p(|f_n|^p + |f|^p)$ und $g = 2^{p+1}|f|^p$. Da $f_n \to f$ fast sicher, ist auch $g_n \to g$ fast sicher. Ausserdem $\int g_n \mathrm{d}\mu \to \int g \mathrm{d}\mu$ wegen der Voraussetzung $\|f_n\|_p \to \|f\|_p$. Wegen $|x - y|^p \leq 2^p(|x|^p + |y|^p)$ für alle $x, y \in \mathbb{R}$ gilt auch $|g_n| \geq |f_n - f|^p \to 0$ fast sicher. Daher nach dem Satz 5 über dominierte Konvergenz $\lim_{n \to \infty} \int |f_n - f|^p \mathrm{d}\mu = \int \lim_{n \to 0} |f_n - f|^p \mathrm{d}\mu = 0$, was die $L^p(\mu)$-Konvergenz von f_n gegen f aussagt.

1.5 Absolute Stetigkeit

Proposition 7 (Dichte) Für jede nichtnegative meßbare Funktion $f : \Omega \to [0,\infty]$ und ein Maß μ auf Ω definiert $\nu = f\mu : \mathcal{A} \to [0,\infty]$,

$$\nu(A) = \int \chi_A f \mathrm{d}\mu = \int_A f \mathrm{d}\mu$$

ein Maß auf Ω. f heisst dann Dichte von ν bzgl. μ.

Definition 15 (Absolute Stetigkeit) Sind μ, ν Maße auf $(\Omega, \mathcal{A})$, so heißt ν absolut stetig bezüglich μ, wenn für alle $A \in \mathcal{A}$ aus $\mu(A) = 0$ auch $\nu(A) = 0$ folgt. In Zeichen: $\nu \ll \mu$.

Bemerkung 3 Mit $\nu = f\mu$ für nichtnegatives meßbares f ist $\nu \ll \mu$, denn falls $\mu(A) = 0$, dann auch $\nu(A) = f\mu(A) = 0$, denn aus $\mu(\{f\chi_A > 0\}) = 0$ folgt $\nu(A) = \int f\chi_A \mathrm{d}\mu = 0$.

Lemma 4 (Endliche Maße) Seien μ und ν endlich Maße auf $(\Omega, \mathcal{A})$ mit $\nu \le \mu$, d.h., es gelte $\nu(\Omega), \mu(\Omega) < \infty$ und $\nu(A) \le \mu(A)$, für alle $A \in \mathcal{A}$. Dann gibt es ein meßbares $g : \Omega \to [0,1]$ mit $\nu = g\mu$.

Satz 8 (Darstellungssatz von Riesz) Sei H ein Hilbertraum mit Skalarprodukt $\langle \cdot, \cdot \rangle$ und zugehöriger Norm $\|\cdot\|$, und $l : H \to \mathbb{R}$ eine lineare Abbildung mit $|l(f)| \le c\|f\|$ für alle $f \in H$. Dann existiert ein eindeutig bestimmtes $g \in H$ so dass $l(f) = \langle g, f \rangle$.

Der Nachweis dieses Lemmas stellt eine interessante Verknüpfung zur Funktionalanalysis dar: Wegen der Voraussetzungen ist $L^2(\mu) \subset L^2(\nu) \subset L^1(\nu)$. Daher ist die Linearform $l(f) = \int f \mathrm{d}\nu$ wohldefiniert auf dem Hilbertraum $L^2(\mu)$ und erfüllt $|l(f)| \le c\|f\|_{2,\mu}$ und der Rieszsche Darstellungssatz besagt, dass es ein $g \in L^2(\mu)$ gibt, so dass für alle $f \in L^2(\mu)$ gilt: $l(f) = \int f \mathrm{d}\nu = \int gf \mathrm{d}\mu = \langle f, g \rangle$. g ist die gesuchte Dichte und $g : \Omega \to [0,1]$ leicht durch Widerspruchsbeweis nachweisbar: Nimmt man z.B. $\mu(\{g < 0\}) > 0$ an, so folgt $\nu(\{g < 0\}) = \int_{g<0} \mathrm{d}\mu < 0$.

Satz 9 (Radon-Nikodym)
Es seien μ, ν Maße auf $(\Omega, \mathcal{A})$ und μ σ-endlich. Dann sind äquivalent:

i) $\nu \ll \mu$
ii) Es gibt ein meßbares $f \ge 0$ mit $\nu = f\mu$, das heißt

$$\nu(A) = \int_A f \mathrm{d}\mu$$

für alle $A \in \mathcal{A}$.

Der Beweis ist für endliche Maße ν, μ recht einfach: Da ii)$\Rightarrow$i) trivial ist, betrachten wir i)$\Rightarrow$ ii) und definieren das endliche Maß $\phi = \mu + \nu$ mit $\nu, \mu \le \phi$. Dann existieren aufgrund des letzten Lemmas meßbare g, h so daß $\mu = g\phi$ und $\nu = h\phi$. Wegen $\nu \ll \mu$ folgt aus $\mu(\{g = 0\}) = \int_{\{g=0\}} g \mathrm{d}\phi = 0$ auch $\nu(\{g = 0\}) = 0$ und daher $\nu(A) = \nu(A \cap \{g \ne 0\})$. Nun kann man $f : \Omega \to [0,\infty]$ definieren mit $f(x) = h(x)/g(x)$ falls $g(x) \ne 0$ und 0 sonst. Diese Funktion ist nichtnegativ und meßbar mit $\nu(A) = \int_{A\cap\{g\ne0\}} h \mathrm{d}\phi = \int_{A\cap\{g\ne0\}} fg \mathrm{d}\phi = \int_{A\cap\{g\ne0\}} f \mathrm{d}\mu = \int_A f \mathrm{d}\mu = f\mu(A)$.

2 Werkzeug 2: Wahrscheinlichkeitstheorie

Literatur: [Dur, G, K]

2.1 Wahrscheinlichkeitsraum, Verteilungsfunktion

Definition 16 (W-Raum) Ein Maßraum $(\Omega, \mathcal{A}, \mu)$ mit $\mu(\Omega) = 1$ heißt Wahrscheinlichkeitsraum. μ heißt dann Wahrscheinlichkeitsmaß.
Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ heißt diskret, wenn es ein abzählbares $T \in \mathcal{A}$ mit $\mu(T) = 1$ gibt. T heißt Träger.

Beispiel 3 Betrachte $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \delta_x)$ mit dem Dirac-Maß δ_x getragen in $x \in \mathbb{R}$. Dieser Raum ist als diskret mit Träger $T = \{x\}$, obwohl $\Omega = \mathbb{R}$ offensichtlich kontinuierlich ist.

Proposition 8 (Zähldichte) Ist $(\Omega, \mathcal{P}(\Omega), \mu)$ ein diskreter W-Raum mit Träger T, so ist die zugehörige Zähldichte $f : \Omega \to [0, 1]$, $\omega \mapsto \mu(\{\omega\})$ ein Dichte von μ bezüglich des Zählmaßes μ_Z, und es gilt $\mu(A) = \sum_{\omega \in A \cap T} f(\omega)$ für alle $A \in \mathcal{A}$.

Beispiel 4 (Poisson-Verteilung)
Sei $\Omega = \mathbb{N}$ und $k > 0$. Das W-Maß mit $\mu(\{n\}) = e^{-k}\, k^n/k!$ heisst Poisson-Verteilung mit Rate k. Die Rate gibt den Mittelwert an, $\sum_{n \geq 0} n\mu(\{n\}) = k$. Die Zähldichte ist einfach $f(n) = e^{-k}\, k^n/k!$ mit $\mu(A) = \sum_{n \in A} f(n)$ für alle $A \subseteq \Omega$.

Satz 10 (Maßeindeutigkeit und Lebesgue-Stieltjes-Maß) i) Es seien μ, ν Maße auf $(\Omega, \sigma(\mathcal{M}))$ mit $\forall A, B \in \mathcal{M} : A \cap B \in \mathcal{M}$. Gibt es eine Folge $(A_n)_{n \in \mathbb{N}}$ in $\mathcal{M}$ mit $\mu(A_n) < \infty$, $n \in \mathbb{N}$, und $\bigcup_{n \in \mathbb{N}} A_n = \Omega$, und gilt $\mu(A) = \nu(A)$ für alle $A \in \mathcal{M}$, so ist $\mu = \nu$.

ii) Zu jeder monoton wachsenden, rechtsstetigen Funktion $F : \mathbb{R} \to \mathbb{R}$ gibt es genau ein Maß $\lambda_F : \mathcal{B}(\mathbb{R}) \to [0, \infty]$, so daß für alle $a, b \in \mathbb{R}$, $a \leq b$, gilt:

$$\lambda_F(]a, b]) = F(b) - F(a).$$

λ_F heisst Lebesgue-Stieltjes-Maß zu F; für $F = id$ ist $\lambda_F = \lambda$ identisch mit dem Lebesgue-Maß.

Definition 17 (Verteilungsfunktion) i) Ist μ ein W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so heißt $F_\mu : \mathbb{R} \to [0, 1]$, $x \mapsto \mu(] - \infty, x])$ die Verteilungsfunktion von μ.

ii) Eine monoton wachsende, rechtsseitig stetige Funktion $F : \mathbb{R} \to [0, 1]$ mit $\lim_{x \to -\infty} F(x) = 0$ und $\lim_{x \to +\infty} F(x) = 1$ heißt Verteilungsfunktion.

Proposition 9 (Korrespondenz) Für jede Verteilungsfunktion $F : \mathbb{R} \to [0, 1]$ ist $\mu = \lambda_F$ ein W-Maß auf $(\mathbb{R}, \mathcal{B}(R))$ mit $F_\mu = F$. Umgekehrt ist für jedes W-Maß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ die Funktion $G = F_\mu$ eine Verteilungsfunktion mit $\lambda_G = \mu$.

Korollar 2 Es sei μ ein W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und F_μ seine Verteilungsfunktion. Dann sind äquivalent:

i) F_μ ist stetig.
ii) $\mu(\{x\}) = 0$ für alle $x \in \mathbb{R}$.

Das Korollar ergibt sich leicht aus folgender Überlegung: Betrachte $x \in \mathbb{R}$ und eine Folge (x_n) mit $x_n \uparrow x$. Dann ergibt sich aus der Stetigkeit von μ:

$$\mu(\{x\}) \quad = \quad \lim_{n\to\infty} \mu(]x_n, x]) = \lim_{n\to\infty} (\mu(]-\infty, x]) - \mu(]-\infty, x_n]) = F_\mu(x) - \lim_{n\to\infty} F_\mu(x_n),$$

woraus sofort beide Richtungen folgen.

Beispiel 5 (Normalverteilung)
Die Normalverteilung $\mathcal{N}(\bar{x}, \sigma^2)$ mit Mittelwert $\bar{x} \in \mathbb{R}$ und Standardabweichung $\sigma > 0$ hat die Dichte

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-\bar{x})^2}{2\sigma^2}\Big)$$

und die Verteilungsfunktion $F(x) = \int_{-\infty}^{x} f \mathrm{d}\lambda = \Phi((x-\bar{x})/\sigma)$ wobei Φ die Gaußsche Fehlerfunktion ist. Offensichtlich ist für das zugehörige W-Maß $\mu(\{x\}) = 0$.

Beispiel 6 (Exponentialverteilung)
Die Exponentialverteilung $Exp(k)$, $k > 0$ hat die Dichte $f(x) = k\exp(-kx)\chi_{(0,\infty)}(x)$. Der Mittelwert ist $1/k = \int_{-\infty}^{\infty} f(x)dx$.

2.2 Zufallsvariablen und Momente

Definition 18 (Zufallsvariable) Sei $(\Omega_1, \mathcal{A}_1, \mu)$ ein W-Raum und $\mathcal{A}_2$ eine σ-Algebra auf Ω_2. Dann heißt eine $\mathcal{A}_1$-$\mathcal{A}_2$-meßbare Abbildung $X : \Omega_1 \to \Omega_2$ Zufallsvariable. Das Bildmaß $\mu_X : \mathcal{A}_2 \to [0,1]$, $\mu_X(A) = \mu(X^{-1}(A))$ heißt Verteilung von X.
Die von X erzeugte σ-Algebra, bezeichnet mit $\sigma(X) \subseteq \mathcal{P}(\Omega_1)$ ist die kleinste σ-Algebra, in der X meßbar ist.

Lemma 5 (Erzeugte σ-Algebra) Es gilt $\sigma(X) = X^{-1}(\mathcal{A}_2)$ mit den Bezeichungen aus der letzten Definition.

Lemma 6 (Integration bzgl. einer Verteilung) Sei $X : \Omega \to \mathbb{R}^n$ eine Zufallsvariable und $f : \mathbb{R}^n \to \mathbb{R}$ meßbar, so daß $f \circ X : \Omega \to \mathbb{R}$ μ-integrierbar ist. Dann gilt

$$\int_{\mathbb{R}^n} f(x)\mathrm{d}\mu_X(x) = \int_\Omega (f \circ X)(\omega)\mathrm{d}\mu(\omega).$$

Betrachte zuerst $f = \chi_A$ für $A \in \mathcal{B}(\mathbb{R}^n)$. Dann ist $f \circ X = \chi_{X^{-1}(A)}$ und

$$\int f\mathrm{d}\mu_X = \int \chi_A \mathrm{d}\mu_X = \mu_X(A) = \mu(X^{-1}(A)) = \int \chi_{X^{-1}(A)}\mathrm{d}\mu = \int (f \circ X)\mathrm{d}\mu.$$

Alles weitere folgt durch Approximation eines beliebigen f durch Treppenfunktionen.

Definition 19 (Erwartungswert & Varianz) Sei $X : \Omega \to \mathbb{R}$ Zufallsvariable.

i) Für $X \in \mathcal{L}^1(\mu)$ heißt $\mathbb{E}(X) := \int X \mathrm{d}\mu$ Erwartungswert von X.
ii) Für $X \in \mathcal{L}^2(\mu)$ heißt $\mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2)$ Varianz von X.

Definition 20 (Momente) Ist $X \in \mathcal{L}^n(\mu)$, $n \in \mathbb{N}$, so heißen $\mathbb{E}(|X|^n)$, $\mathbb{E}(X^n)$ und $\mathbb{E}((X - \mathbb{E}(X))^n)$ n-tes absolutes Moment, n-tes Moment und n-tes zentiertes Moment.

Beispiel 7 (Verschiedene Verteilungen)
i) X deterministisch: Definiert durch $\mathbb{V}(X) = 0$, also $X = \mathbb{E}(X)$ fast sicher.
ii) Sei $X : \Omega \to \mathbb{N}_0$ Poisson-verteilt mit Rate k. Dann $\mathbb{E}(X) = k$ und $\mathbb{V}(X) = \sum_{n \geq 0} n^2 e^k k^n / n! - k^2 = k$
iii) Sei X gemäß $\mathcal{N}(\bar{x}, \sigma^2)$ normalverteilt. Dann $\mathbb{E}(X) = \bar{x}$ und $\mathbb{V}(X) = \sigma^2$.

Satz 11 (Markov- und Tschebyschev-Ungleichung) Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, $\varepsilon > 0$ und $n \in \mathbb{N}$. Es gilt

$$\mu(|X| \geq \varepsilon) \leq \tfrac{1}{\varepsilon^n} \int_{\{|X| \geq \varepsilon\}} |X|^n \mathrm{d}\mu \leq \tfrac{1}{\varepsilon^n} \mathbb{E}(|X|^n).$$

Insbesondere folgen die Markov- und die Tschebyschev-Ungleichung:

$$\mu(|X| \geq \varepsilon) \leq \tfrac{1}{\varepsilon} \mathbb{E}(|X|^n), \qquad \mu(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \tfrac{1}{\varepsilon^2} \mathbb{V}(|X|).$$

Satz 12 (Jensen'sche Ungleichung) Sei $\phi : I \to \mathbb{R}$ eine konvexe Funktion auf einem Intervall $I \subseteq \mathbb{R}$ und $X : \omega \to I$ eine integrierbare Zufallsvariable. Dann ist $\mathbb{E}(X) \in I$ und es gilt

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)),$$

wobei der Fall $\mathbb{E}(\phi(X)) = +\infty$ möglich ist.

Definition 21 (Randverteilung) Sei $X : \Omega \to \mathbb{R}^n$ eine Zufallsvariable und $g : \mathbb{R}^n \to \mathbb{R}^d$ eine meßbare Funktion. Dann ist $g(X)$ eine d-dimensionale Zufallsvariable mit Verteilung $\mu_{g(X)}(A) = \mu_X(g^{-1}(A))$. Falls g die Projektion auf eine Koordinate ist, d.h. $g(X) = X_j$, ergibt sich $\mu_{X_j}(A) = \mu(X_j \in A)$ und $g(X) = X_j$ heisst Randverteilung.

Lemma 7 (Dichte der Randverteilung) Besitzt X die Dichte f, so hat X_j die Dichte f_j gegeben durch

$$f_j(x) = \int_{\mathbb{R}^{n-1}} f(x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_{j-1} \mathrm{d}x_{j+1} \ldots \mathrm{d}x_n.$$

Lemma 8 (Randverteilung bei Normalverteilung) Sei $X : \Omega \to \mathbb{R}^n$ normalverteilt gemäß $\mathcal{N}(\bar{x}, \Sigma)$ mit $\bar{x} = (\bar{x}_i)$ und $\Sigma = (\Sigma_{ij})$ positiv definit. Dann ist X_j normalverteilt gemäß $\mathcal{N}(\bar{x}_j, \Sigma_{jj})$.

2.3 Bedingte Wahrscheinlichkeit und Unabhängigkeit

Definition 22 (Bedingte Wahrscheinlichkeit) Ist $(\Omega, \mathcal{A}, \mu)$ ein W-Raum und $B \in \mathcal{A}$ mit $\mu(B) > 0$, so heißt

$$\mu(\cdot|B) : \mathcal{A} \to [0,1], \quad \mu(A|B) = \begin{cases} \frac{\mu(A \cap B)}{\mu(B)} & \text{falls } \mu(B) > 0 \\ 0 & \text{sonst} \end{cases}$$

die bedingte Wahrscheinlichkeit unter Bedingung B.

Lemma 9 (Totale Wahrscheinlichkeit & Formel von Bayes) Sei $(\Omega, \mathcal{A}, \mu)$ ein W-Raum, $A, B \in \mathcal{A}$ und $(A_n)_{n\in\mathbb{N}}$ eine disjunkte Folge in $\mathcal{A}$ mit $\mu(\bigcup_{n\in\mathbb{N}} A_n) = 1$. Es gilt:

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)\mu(A|A_n), \qquad \mu(A|B) = \frac{\mu(B|A)\mu(A)}{\mu(B)} = \frac{\mu(B|A)\mu(A)}{\sum_{n=1}^{\infty} \mu(A_n)\mu(B|A_n)}.$$

Proposition 10 (Gedächtnislosigkeit) Eine exponentialverteilte Zufallsvariable X ist gedächtnislos:

$$\forall s, t \geq 0 : \mu(X > t + s | X > t) = \mu(X > s).$$

Umgekehrt ist jede gedächtnislose positive Zufallsvariable exponentialverteilt.

Die erste Aussage sieht man wie folgt: Zuerst ist einmal $\mu(X > s) = 1 - \mu(X \leq s) = 1 - \int_0^s ke^{-kx}dx = e^{-ks}$. Daher gilt $\mu(X > t+s|X > t) = \mu(X > t+s)/\mu(X > t) = e^{-k(t+s)}/e^{-kt} = e^{-ks} = \mu(X > s)$.
Für die zweite Aussage definiere $g(t) = \mu(X > t)$ mittels des gegebenen X. Dann ist g monoton fallend mit $g(t+s) = \mu(X > t+s) = \mu(X > t)\mu(X > t+s|X > t) = \mu(X > t)\mu(X > s) = g(t)g(s)$ und $g(0) = 1$. Definiere dann k via $e^{-k} = g(1)$. Daraus folgt dann $g(t) = e^{-kt}$ und daher ist X exponentialverteilt.

Definition 23 (Unabhängigkeit) Sei $(\Omega, \mathcal{A}, \mu)$ ein W-Raum.
i) Eine Familie $(A_i)_{i\in I}$ in $\mathcal{A}$ heißt unabhängig, wenn für jede endliche Teilmenge $\emptyset \neq J \subseteq I$ gilt

$$\mu(\bigcap_{j\in J} A_j) = \prod_{j\in J} \mu(A_j).$$

ii) Eine Familie $(\mathcal{M}_i)_{i\in I}$ von Mengensystemen $\mathcal{M}_i \subseteq \mathcal{A}$ heißt unabhängig, wenn jede Familie $(A_i)_{i\in I}$ von Ereignissen $A_i \in \mathcal{M}_i$ unabhängig ist.

iii) Eine Familie $(X_i)_{i\in I}$ von Zufallsvariablen $X_i : \Omega \to \Omega_i$ mit zugehörigen σ-Algebren $\mathcal{A}_i$ heißt unabhängig, wenn die Mengensysteme $(\sigma(X_i))_{i\in I}$, $\sigma(X_i) = X_i^{-1}(\mathcal{A}_i)$, unabhängig sind.

Proposition 11 (durchschnittsstabiler Erzeuger) Für jede unabhängige Familie $(\mathcal{M}_i)_{i\in I}$ von durchschnittsstabilen Mengensystemen ist auch $(\sigma(\mathcal{M}_i))_{i\in I}$ unabhängig.

Korollar 3 (endlich viele Zufallsvariablen) $X_1, \ldots, X_n$ reelle Zufallsvariablen sind genau dann unabhängig, wenn für alle $c \in \mathbb{R}^n$ gilt

$$\mu(X_1 \leq c_1, \ldots, X_n \leq c_n) = \prod_{j=1}^{n} \mu(X_j \leq c_j).$$

$X_1, \ldots, X_n$ diskrete Zufallsvariablen mit Träger $T_1, \ldots, T_n$ sind genau dann unabhängig, wenn für alle $x \in T_1 \times \ldots \times T_n$ gilt

$$\mu(X_1 = x_1, \ldots, X_n = x_n) = \prod_{j=1}^{n} \mu(X_j = x_j).$$

Satz 13 (Produktsatz) Für unabhängige Zufallsvariablen $X_1, \ldots X_n \in \mathcal{L}^1(\mu)$ gilt

$$\mathbb{E}(\prod_{j=1}^{n} X_j) = \prod_{j=1}^{n} \mathbb{E}(X_j).$$

Definition 24 (Kovarianz) Für reelle Zufallsvariablen $X, Y \in \mathcal{L}^1(\mu)$ heißt

$$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

die Kovarianz von X und Y. Falls $\mathrm{Cov}(X, Y) = 0$ gilt, heißen X und Y unkorreliert.

Bemerkung 4 (Unabhängigkeit versus Unkorreliertheit) Wenn die reellen Zufallsvariablen X, Y unabhängig sind, dann sind sie auch unkorreliert. Umgekehrt gilt das nicht.

2.4 Stochastische Konvergenz

Definition 25 (Stochastische Konvergenz) Sei $(X_i)_{i \in \mathbb{N}}$ eine Folge von reellen Zufallsvariablen im W-Raum $(\Omega, \mathcal{A}, \mu)$. (X_i) konvergiert stochastisch gegen eine reelle Zufallsvariable X, falls für jedes $\epsilon > 0$ gilt:

$$\lim_{i \to \infty} \mu\Big(|X_i - X| \geq \epsilon\Big) = 0.$$

Wir schreiben $X_i \to_\mu X$.

Satz 14 (Fast sichere und stochastische Konvergenz) Seien $(X_i), X$ reelle Zufallsvariablen mit $X_i \to X$ μ-fast sicher. Dann gilt $X_i \to_\mu X$. Die Umkehrung ist im allgemeinen falsch.

Satz 15 (Kriterium für stochastische Konvergenz) Seien $(X_i), X$ reelle Zufallsvariablen. Dann gilt $X_i \to_\mu X$ genau dann, wenn jede Teilfolge (X_{i_k}) von (X_i) wiederum eine Teilfolge besitzt, die μ-fast sicher konvergiert.

Satz 16 (L^p- und stochastische Konvergenz) Seien $(X_i), X$ reelle Zufallsvariablen aus $L^p(\mu)$ mit $p \in [1, \infty]$ und $X_i \to X$ in $L^p(\mu)$, d.h. $\lim_{i \to \infty} \|X_i - X\|_p = 0$. Dann gilt $X_i \to_\mu X$. Die Umkehrung ist im allgemeinen falsch.

Lemma 10 (Stochastische Konvergenz unter Transformation)
Seien $(X_i), X$ reelle Zufallsvariablen mit $X_i \to_\mu X$ und $f : \mathbb{R} \to \mathbb{R}$ eine stetige Funktion. Dann gilt auch $f(X_i) \to_\mu f(X)$.

2.5 Anwendung: Monte Carlo Integration

Definition 26 (iid) Seien $X_i : \Omega \to \mathbb{R}^n$ unabhängige Zufallsvariablen mit gleicher Verteilung μ_X. Die Folge $(X_i)_{i\in\mathbb{N}}$ wird dann unabhängig identisch verteilt oder iid für *independent and identically distributed* genannt.

Bemerkung 5 Eine iid Folge $(X_i)_{i\in\mathbb{N}}$ beschreibt z.B. die Situation, in der man das immer gleich Zufallsexperiment vielmals hintereinander und unabhängig voneinander wiederholt, zum Beispiel in dem man ein und denselben Würfel immer wieder wirft. In gewisser Weise stellt das Zufallsexperiment *Werfen dieses Würfels* eine **Realisierung** der Zufallsvariable $X : \Omega \to W = \{1, 2, 3, 4, 5, 6\}$ mit Gleichverteilung $\mu_X(\{i\}) = 1/6,\ i \in W$ dar und die X_i können als unabhängige Kopien von X verstanden werden. Das n-malige Werfen des Würfels ist dann eine Realisierung der n-dimensionalen Zufallsvariable $Y_n = (X_1, \ldots, X_n)$; diese spezielle Realisierung wird dann durch $Y_n(\omega)$ für ein spezielles $\omega \in \Omega$ beschrieben. Obwohl wir die X_i als Kopien von X verstehen, bedeutet das aber nicht, das ihre Werte in $Y_n(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ identisch sind, sondern lediglich, dass sie Resultat eines identisch durchgeführten Zufallsexperiments sind.

Algorithmus 1 (Monte-Carlo Integration) Sei $(X_i)_{i\in\mathbb{N}}$ eine iid Folge von Zufallsvariablen $X_i : \Omega \to \mathbb{R}^n$ mit Verteilung μ und $f : \mathbb{R}^d \to \mathbb{R}$ eine μ-integrierbare Funktion. Dann approximieren wir das Integral $\int f \mathrm{d}\mu$ mittels einer Realisierung $(X_1(\omega), \ldots, X_n(\omega))$ für ausreichend großes n, in dem wir

$$S_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i),$$

realisieren, d.h. wir wollen $S_n(\omega)$ als Approximation von $\int f \mathrm{d}\mu$ verwenden.

Experiment 1 (Experiment zur MC-Integration) In Abbildung 1 ist das Ergebnis des folgenden numerischen Experiments gezeigt: Die iid Folge $(X_i)_{i\in\mathbb{N}}$ ist standard-normalverteilt und $f(x) = x^2$. Dann ist

$$\int f \mathrm{d}\mu = \frac{1}{\sqrt{2\pi}} \int x^2 \exp(-x^2/2) dx = 1$$

analytisch berechenbar. Abbildung 1 zeigt die Approximation von $\int f \mathrm{d}\mu$ durch eine Realisierung $S_n(\omega)$ mit wachsendem n. S_n ist dabei wie oben definiert.

Bemerkung 6 Die Idee der Monte-Carlo Integration stellt uns vor die Frage, wann und in welchem Sinne die Konvergenz

$$S_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \int f \mathrm{d}\mu, \qquad n \to \infty,$$

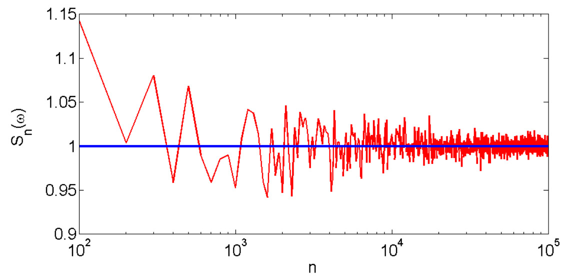

Abbildung 1: Ergebnis einer Monte Carlo Integration wie in Experiment 1 beschrieben. Gezeigt ist $S_n(\omega)$ für wachsendes n im Vergleich zum zu approximierenden Wert $\int f \mathrm{d}\mu = 1$.

sicher gestellt ist und wie schnell sie erfolgt. Diese Frage wird uns durch die nächsten Kapitel begleiten und durch die Aussagen des Gesetzes der großen Zahlen und des zentralen Grenzwertsatzes beantwortet werden.

2.6 Gesetze der großen Zahlen

Definition 27 (Gesetze der großen Zahlen) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen in $\mathcal{L}^1(\mu)$.

i) Es gilt das **schwache Gesetz der großen Zahlen**, wenn $\frac{1}{n}\sum_{j=1}^n (X_j - \mathbb{E}(X_j)) \to 0$ stochastisch.

ii) Es gilt das **starke Gesetz der großen Zahlen**, wenn $\frac{1}{n}\sum_{j=1}^n (X_j - \mathbb{E}(X_j)) \to 0$ fast sicher.

Satz 17 (Schwaches Gesetz für iid Zufallsvariablen) Für unabhängige, identisch verteilte Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{L}^2(\mu)$ mit beschränkter Varianz $\sigma^2 = \mathbb{V}(X_i) < \infty$ gilt das schwache Gesetz der großen Zahlen.

Die Unabhängigkeit liefert uns für $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ die Unkorreliertheit und diese sofort

$$\mathbb{V}(\bar{X}_n) = \mathbb{V}(\frac{1}{n}\sum_{j=1}^n X_j) = \frac{1}{n^2}\mathbb{V}(\sum_{j=1}^n X_j) = \frac{1}{n^2}\sum_{j=1}^n \mathbb{V}(X_j) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Setze $m = \mathbb{E}(X_i)$ und die Tschebyschev-Ungleichung für $\bar{X}_n$ mit $\mathbb{E}(\bar{X}_n) = m$ liefert

$$\mu(|\bar{X}_n - m| \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n\epsilon^2}.$$

Für $n \to \infty$ geht die linke Seite für alle $\epsilon > 0$ gegen 0 und damit folgt das schwache Gesetz.

Satz 18 (Schwaches Gesetz) Für paarweise unkorrelierte Zufallsvariablen $(X_j)_{j\in\mathbb{N}}$ aus $\mathcal{L}^1(\mu)$ mit $\lim_{n\to\infty} \frac{1}{n^2}\sum_{j=1}^n \mathbb{V}(X_j) = 0$ gilt das schwache Gesetz der großen Zahlen.

Wie im letzten Beweis nutzen wir die Unkorreliertheit und erhalten $\mathbb{V}(\bar{X}_n) = \frac{1}{n^2}\sum_{j=1}^n \mathbb{V}(X_j) = \sigma_n^2$ und dann mit der Tschebyschev-Ungleichung für $\bar{X}_n$ sofort $\mu(|\bar{X}_n - \mathbb{E}(\bar{X}_n)| \geq \epsilon) \leq \frac{1}{\epsilon^2}\sigma_n^2$, was wegen $\sigma_n^2 \to 0$ für $n \to \infty$ die Behauptung liefert.

Korollar 4 Für paarweise unkorrelierte Zufallsvariablen mit gleichmäßig beschränkter Varianz gilt das schwache Gesetz der großen Zahlen.

Satz 19 (Maximalungleichung von Kolmogorov)
Seien $X_1, \ldots, X_n \in \mathcal{L}^2(\mu)$ unabhängige Zufallsvariablen mit $\mathbb{E}(X_j) = 0$ und . Dann gilt für jedes $\varepsilon > 0$

$$\mu\left(\max_{1 \leq k \leq n} |X_1 + \ldots + X_k| \geq \varepsilon\right) \leq \varepsilon^{-2}\, \mathbb{V}(X_1 + \ldots X_n).$$

Definition 28 (Fast sichere Konvergenz einer Reihe) Sei (X_i) eine Folge von Zufallsvariablen. Die Reihe $\sum_{i=1}^{\infty} X_i$ heisst μ-fast sicher konvergent, falls eine Zufallsvariable Z mit $\mu(|Z| < \infty) = 1$ existiert, so dass $\sum_{i=1}^{n} X_i \to Z$ μ-fast sicher.

Satz 20 (Fast sichere Konvergenz von unabhängigen Zufallsvariablen)
Sei (X_i) eine Folge von unabhängigen Zufallsvariablen mit $\mathbb{E}(X_i) = 0$ für alle i. Ist $\sum_{i=1}^{\infty} \mathbb{E}(X_i^2) = \sum_{i=1}^{\infty} \mathbb{V}(X_i) < \infty$ so konvergiert (X_i) μ-fast sicher.

Satz 21 (Starkes Gesetz für unabhängige Zufallsvariablen) Für unabhängige, reelle Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ mit $\mathbb{V}(X_i) < \infty$ für alle i, sei ausserdem die Kolmogorov-Bedingung

$$\sum_{i=1}^{\infty} \frac{1}{i^2} \mathbb{V}(X_i) < \infty$$

erfüllt. Dann gilt das starke Gesetz der großen Zahlen für (X_n).

Lemma 11 (Kronecker) Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}$ mit $\sum_{n=1}^{\infty} x_n/b_n < \infty$ für eine monoton wachsende Folge $(b_n)_{n\in\mathbb{N}}$ in $]0, \infty[$ mit $\lim_{n\to\infty} b_n = \infty$. Dann gilt $\lim_{n\to\infty} \frac{1}{b_n} \sum_{j=1}^{n} x_j = 0$.

Beweis zu Satz 21: Setze $Y_j = (X_j - \mathbb{E}(X_j))/j$. Nach Voraussetzung gilt $\sum_{j=1}^{\infty} \mathbb{V}(Y_j) = \sum_{j=1}^{\infty} \mathbb{V}(X_j)/j^2 = \infty$, so dass nach Satz 20 die Reihe $\sum_j Y_j$ fast sicher konvergiert. Dann ist nach Kroneckers Lemma $\frac{1}{n} \sum_{j=1}^{n} (X_j - \mathbb{E}(X_j)) = 0$ fast sicher.

Satz 22 (Lemma von Borel-Cantelli) Für jede Folge $(A_n)_{n\in\mathbb{N}}$ von Ereignissen in $\mathcal{A}$ mit $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ ist $\mu(\limsup_{n\to\infty} A_n) = 0$.
Ist $(A_n)_{n\in\mathbb{N}}$ unabhängig mit $\sum_{n=1}^{\infty} \mu(A_n) = \infty$, so folgt $\mu(\limsup_{n\to\infty} A_n) = 1$.

Lemma 12 (Cesaros Lemma) Sei (a_n) eine Folge reeller Zahlen mit $a_n \to a \in \mathbb{R}$ und (b_n) eine andere Folge mit $b_n \to \infty$ monoton von unten. Dann gilt mit $a_0 = b_0 = 0$,

$$\frac{1}{b_n} \sum_{i=1}^{n} (b_i - b_{i-1}) a_{i-1} \to a.$$

Definition 29 Betrachte die Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$. Dann heißen die $Y_n = X_n \chi_{\{|X_n| \leq n\}}$ gestutzte Zufallsvariablen.

Lemma 13 (Auf gestutzte Zufallsvariablen vererbte Eigenschaften) Für iid Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{L}^1(\mu)$ gilt für die zugehörigen gestutzten Zufallsvariablen $Y_n = X_n \chi_{\{|X_n|\leq n\}}$:

i) $\mathbb{E}(Y_n) \to \mathbb{E}(X_1)$ für $n \to \infty$
ii) $\mu(X_n = Y_n \text{ für fast alle } n) = 1$
ii) (Y_n) erfüllt die Kolmogorov-Bedingung $\sum_{j=1}^{\infty} \mathbb{V}(Y_j)/j^2 < \infty$.

Satz 23 (Starkes Gesetz für iid Zufallsvariablen) Für iid Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{L}^1(\mu)$ gilt das starke Gesetz der großen Zahlen, d.h.

$$\frac{1}{n}\sum_{j=1}^{n} X_j \to \mathbb{E}(X_1), \quad \mu-\text{fast sicher.}$$

3 Der zentrale Grenzwertsatz

Literatur: [Dur, K, MS]

3.1 Schwache Konvergenz

Definition 30 (Schwache Konvergenz) Sei $(\Omega, \mathcal{A}, \mu)$ ein W-Raum.

i) Sei Ω ein topologischer Raum und $\mathcal{A} = \mathcal{B}(\Omega)$. Eine Folge von W-Maßen auf $(\Omega, \mathcal{A})$ konvergiert schwach gegen μ, falls

$$\forall f \in C_b(\Omega) := \{f \in C(\Omega) : f \text{ beschränkt}\} : \int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu.$$

ii) Seien $(X_n)_{n\in\mathbb{N}}$, X reelle Zufallsvariablen auf $(\Omega, \mathcal{A})$. $(X_n)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen X, falls $\mu_{X_n} \to \mu_X$ schwach.
Wir notieren die schwache Konvergenz als $X_n \to_d X$ oder $\mu_n \to_d \mu$.

Lemma 14 Konvergenz in Verteilung folgt sowohl aus fast sicherer Konvergenz als auch aus stochastischer Konvergenz.

Satz 24 (Skorokhod-Darstellung) Seien $(X_n)_{n\in\mathbb{N}}$, X reelle Zufallsvariablen und $(F_n)_{n\in\mathbb{N}}$, F die zugehörigen Verteilungsfunktionen. Gilt für alle $x \in \mathbb{R}$, in denen F stetig ist, $\lim_{n\to\infty} F_n(x) = F(x)$, so gibt es reelle Zufallsvariablen $(Y_n)_{n\in\mathbb{N}}$, Y auf einem W-Raum $(\Omega, \mathcal{A}, \mu)$, so daß gilt:

$$\forall n \in \mathbb{N} : \mu_{X_n} = \mu_{Y_n}, \quad \mu_X = \mu_Y, \quad Y_n \to Y \text{ fast sicher.}$$

Satz 25 (Teil des Portemanteau-Theorems) Für reelle Zufallsvariablen (X_n), X mit zugehörigen Verteilungsfunktionen (F_n), F sind äquivalent:

i) $X_n \to X$ in Verteilung.
ii) $F_n(x) \to F(x)$ für alle $x \in \mathbb{R}$, in denen F stetig ist.
iii) $\mu(X_n \in A) \to \mu(X \in A)$ für alle $A \in \mathcal{B}(\mathbb{R})$ mit $\mu(X \in \partial A) = 0$.

3.2 Levy'scher Stetigkeitssatz und charakteristische Funktionen

Definition 31 (Charakteristische Funktion) Für eine n-dimensionale reelle Zufallsvariable $X : \Omega \to \mathbb{R}^n$ heißt $\phi_X : \mathbb{R}^n \to \mathbb{C}, t \mapsto \mathbb{E}(\mathrm{e}^{it\cdot x}) = \int_{\mathbb{R}^n} \mathrm{e}^{it\cdot x} \mathrm{d}\mu_X(x)$ charakteristische Funktion.

Für ein Wahrscheinlichkeitsmaß μ auf $\mathbb{R}^n$ heißt $\phi_\mu : \mathbb{R}^n \to \mathbb{C}$, $t \mapsto \int_{\mathbb{R}^n} \mathrm{e}^{it\cdot x} \mathrm{d}\mu(x)$ charakteristische Funktion.

Lemma 15 (Eigenschaften) Für charakteristische Funktionen $\phi : \mathbb{R}^n \to \mathbb{C}$ gilt

i) $\forall t \in \mathbb{R}^n : |\phi(t)| \leq \phi(0) = 1$
ii) $\forall t \in \mathbb{R}^n : \phi(-t) = \overline{\phi}(t)$
iii) Sei $C \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$. $\forall t \in \mathbb{R}^n : \phi_{CX+b}(t) = \mathrm{e}^{it\cdot b}\phi(C^T t)$
iv) ϕ ist gleichmäßig stetig.

Lemma 16 (Summen) Für unabhängige Zufallsvariablen $X_1, \dots X_n$ gilt

$$\phi_{X_1+\ldots+X_n} = \phi_{X_1} \cdot \ldots \cdot \phi_{X_n}.$$

Satz 26 (Levy'scher Stetigkeitssatz) Es seien μ, μ_n W-Maße auf $\mathbb{R}^d$ mit charakteristischen Funktionen ϕ, ϕ_n. Es gilt $\lim_{n\to\infty} \mu_n = \mu$ schwach genau dann, wenn $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ für alle $t \in \mathbb{R}^d$.

Korollar 5 (Eindeutigkeit) Charakteristische Funktionen legen Wahrscheinlichkeitsmaße eindeutig fest, d.h. wenn ϕ und ψ die charakteristischen Funktionen zu den Maßen μ und nu auf $\mathbb{R}^d$ sind, so folgt aus $\phi = \psi$ sofort $\mu = \nu$.

Proposition 12 (Differenzierbarkeit) Sei X eine reelle Zufallsvariable in $\mathcal{L}^n(\mu)$. Dann ist die charakteristische Funktion ϕ_X n-mal differenzierbar und es gilt

$$\forall k \leq n, t \in \mathbb{R} : \phi_X^{(k)}(t) = \mathbb{E}((iX)^k \mathrm{e}^{itX}).$$

Insbesondere ist $\phi_X^{(k)}(0) = i^k \mathbb{E}(X^k)$.

Proposition 13 (Normalverteilung) Sind $m, \sigma \in \mathbb{R}$, so ist für eine $N(m, \sigma^2)$-verteilte Zufallsvariable X die charakteristische Funktion $\phi_X(t) = \mathrm{e}^{itm}\mathrm{e}^{-\sigma^2 t^2/2}$.

Ist $m \in \mathbb{R}^n$ und $C \in \mathbb{R}^{n\times n}$ symmetrisch positiv definit, so ist für eine $N(m, C)$-verteilte Zufallsvariable $X = (X_1, \dots, X_n)$ die charakteristische Funktion $\phi_X(t) = \mathrm{e}^{it\cdot m}\mathrm{e}^{-t\cdot Ct/2}$.

3.3 Momentenerzeugende Funktion

Definition 32 (Momentenerzeugende Funktion) Für eine n-dimensionale reelle Zufallsvariable $X : \Omega \to \mathbb{R}^n$ heißt

$$M : D \to \mathbb{R}, \quad M(s) = \mathbb{E}[\exp(sX)] = \int \exp(sx) \mathrm{d}\mu_X$$

auf $D = \{s \in \mathbb{R} : \mathbb{E}[\exp(sX)] < \infty\}$ momentenerzeugende Funktion von X.

Satz 27 (Momente charakterisieren Verteilung) Sei X eine reellwertige Zufallsvariable mit momentenerzeugender Funktion $M : D \to \mathbb{R}$. Ist $(-a, a) \subset D$ für ein positives $a \in \mathbb{R}$, dann sind alle Momente $\mathbb{E}(X^n)$ endlich und es gilt

$$M(s) = \sum_{j=1}^{\infty} \frac{s^n}{n!} \mathbb{E}(X^n)$$

für alle $s \in (-a, a)$. Ausserdem ist M in $(-a, a)$ unendlich oft differenzierbar mit n-ter Ableitung $M^{(n)}(0) = \mathbb{E}(X^n)$. Letztlich ist dann auch μ_X eindeutig durch die Momente bestimmt.

Proposition 14 (Normalverteilung) Sind $m, \sigma \in \mathbb{R}$, so ist für eine $N(m, \sigma^2)$-verteilte Zufallsvariable X die momentenerzeugende Funktion $M(s) = \mathrm{e}^{sm}\mathrm{e}^{\sigma^2 s^2/2}$.

3.4 Zentraler Grenzwertsatz

Satz 28 (Zentraler Grenzwertsatz für Zufallsspaziergang) Sei (X_n) eine Folge von iid Zufallsvariablen auf $\Omega = \{-1, +1\}$ mit $\mu(X_1 = +1) = 1/2$ und $\mu(X_1 = -1) = 1/2$. Dann gilt für $S_n = X_1 + \ldots + X_n$ dass

$$\frac{S_n}{\sqrt{n}} \to_d \chi$$

wobei χ eine $\mathcal{N}(0, 1)$ verteilte Zufallsvariable ist.

Beweis: Alle X_n besitzt eine gemeinsame charakteristische Funktion ϕ, die aufgrund der Definition der X_n die einfache Form

$$\phi(t) = \frac{1}{2}(e^{it} + e^{-it}) = \cos(t).$$

Daraus erhalten wir sofort die charakteristische Funktion für $S_n^* = S_n/\sqrt{n}$:

$$\psi_n(t) = \phi(t/\sqrt{n})^n = \cos^n(t/\sqrt{n})$$

hat. Nach dem Stetigkeitssatz von Levy müssen wir nun nur noch zeigen, dass ψ_n gegen die charakteristische Funktion $\exp(-t^2/2)$ einer skalaren standard-normalverteilten Zufallsvariable konvergiert. D.h. für alle $t \in \mathbb{R}$ ist nachzuweisen, dass $\psi_n(t) \to \exp(-t^2/2)$ für $n \to \infty$. Dazu wählen wir t beliebig und betrachten ausreichend grosse n so dass $\cos(t/\sqrt{n}) > 0$. Zu zeigen ist $\log \psi_n(t) = n \log \cos(t/\sqrt{n}) \to -t^2/2$, was sich aus der Regel von l'Hospital ergibt:

$$\lim_{n\to\infty} n \log \cos(t/\sqrt{n}) = \lim_{x\to 0} \frac{\log \cos(x)}{x^2/t^2} = -\frac{t^2}{2}.$$

Satz 29 (Zentraler Grenzwertsatz für Dreiecksschema) Sei

$$\{X_{nj} : j = 1, \ldots k_n, n \in \mathbb{N}\}$$

eine doppelt indizierte Folge von Zufallsvariablen, so daß die $X_{n1}, \ldots, X_{nk_n}$ für jedes $n \in \mathbb{N}$ unabhängig sind. Gilt
i) Standardisierung: $\mathbb{E}(X_{nj}) = 0$ und $\sum_{j=1}^{k_n} \mathbb{E}(X_{nj}^2) = 1$ für $j = 1, \ldots, k_n$, $n \in \mathbb{N}$,
ii) Lindeberg-Bedingung: $\forall \varepsilon > 0 : \lim_{n\to\infty} \sum_{j=1}^{k_n} \mathbb{E}(X_{nj}^2 \chi_{\{|X_{nj}| \geq \varepsilon\}}) = 0$,
so konvergiert die Summe $S_n^* = \sum_{j=1}^{k_n} X_{nj}$ in Verteilung gegen eine standardnormalverteilte Zufallsvariable X, $\lim_{n\to\infty} S_n^* = X$ in Verteilung.

Korollar 6 (Zentraler Grenzwertsatz für iid Zufallsvariablen)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen in $\mathcal{L}^2(\mu)$, $\sigma^2 := \mathbb{V}(X_1)$. Dann konvergiert die standardisierte Summe

$$S_n^* = \frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}(X_j - \mathbb{E}(X_j))$$

in Verteilung gegen eine standardnormalverteilte Zufallsvariable.

Beweis: Setze $k_n = n$ in Satz 29 und

$$X_{ni} = \frac{1}{\sigma\sqrt{n}}\Big(X_i - \mathbb{E}(X_i)\Big), \quad i = 1, \ldots, k_n.$$

Nun prüfen wir die Lindeberg-Bedingung: Dazu sei $\epsilon > 0$ beliebig und wir finden

$$\begin{aligned}\sum_{j=1}^{k_n}\mathbb{E}(X_{nj}^2\chi_{\{|X_{nj}|\geq\varepsilon\}}) &= \frac{1}{\sigma^2 n}\sum_{j=1}^{k_n}\mathbb{E}\Big[(X_j-\mathbb{E}(X_j))^2\chi_{\{|X_j-\mathbb{E}(X_j)|\geq\varepsilon\sigma/\sqrt{n}\}}\Big]\\ &= \frac{1}{\sigma^2}\mathbb{E}\Big[(X_1-\mathbb{E}(X_1))^2\chi_{\{|X_1-\mathbb{E}(X_1)|\geq\varepsilon\sigma/\sqrt{n}\}}\Big]\end{aligned}$$

Da $\mathbb{E}(X_1 - \mathbb{E}(X_1)) = \mathbb{V}(X_1) < \infty$ und $\{|X_1 - \mathbb{E}(X_1)| \geq \varepsilon\sigma/\sqrt{n}\}$ für $n \to \infty$ gegen die leere Menge konvergiert, konvergiert der letzte Ausdruck in obiger Gleichung gegen 0, was die Lindeberg-Bedingung beweist. Dann folgt aus Satz 29 die Behauptung.

Korollar 7 (Satz von deMoivre-Laplace) Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger $B(1,p)$-verteilter Zufallsvariablen, so konvergiert die standardisierte Summe $S_n^* = (S_n - np)/\sqrt{np(1-p)}$ in Verteilung gegen eine $N(0,1)$-verteilte Zufallsvariable χ.

Proposition 15 (Lyapunov-Bedingung) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsvariablen in $\mathcal{L}^2(\mu)$, $s_n^2 := \sum_{j=1}^n \mathbb{V}(X_j)$. Wenn eine der beiden folgenden Bedingungen erfüllt ist:

i) Lyapunov-Bedingung: $\exists\delta > 0 : \lim_{n\to\infty} s_n^{-2-\delta}\sum_{j=1}^n \mathbb{E}(|X_j - \mathbb{E}(X_j)|^{2+\delta}) = 0$,
oder

ii) gleichmäßige Beschränktheit: $\exists c > 0 \forall n \in \mathbb{N} : |X_n| < c$ und $\lim_{n\to\infty} s_n = \infty$,

so konvergiert die standardisierte Summe

$$S_n^* = \frac{1}{s_n}\sum_{j=1}^{n}(X_j - \mathbb{E}(X_j))$$

in Verteilung gegen eine $N(0,1)$-verteilte Zufallsvariable.

3.5 Drei-Reihen-Satz

Satz 30 (Drei-Reihen-Satz) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger reeller Zufallsvariablen. Die Reihe $\sum_{n=1}^\infty X_n$ konvergiert fast sicher genau dann, wenn es ein $c > 0$ gibt so daß

$$\sum_{n=1}^{\infty}\mu(|X_n| < c) < \infty, \quad \sum_{n=1}^{\infty}\mathbb{V}(X_n\chi_{\{|X_n|<c\}}) < \infty, \quad \sum_{n=1}^{\infty}\mathbb{E}(X_n\chi_{\{|X_n|<c\}}) \text{ konv.}$$

4 Markov-Ketten

Literatur: [C, G, MS]

4.1 Existenz

Definition 33 (Stochastische Matrix) Sei S abzählbar. Eine $|S|\times|S|$-Matrix $P = (p_{kl})_{k,l\in S}$ mit nichtnegativen Einträgen und $\sum_{l\in S} p_{kl} = 1$ für alle $k \in S$ heißt stochastische Matrix.

Definition 34 (Markov-Kette) Sei S abzählbar, α ein Wmaß auf S und P eine stochastische Matrix. Eine Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}_0}$, $X_n : \Omega \to S$, heißt Markov-Kette mit Übergangsmatrix P und Startverteilung α, falls $\mu_{X_0} = \alpha$ und für alle $n \in \mathbb{N}_0$, $x_0, \ldots, x_{n+1} \in S$ mit $\mu(X_0 = x_0, \ldots, X_n = x_n) > 0$ die Markov-Eigenschaft gilt:

$$\mu(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = \mu(X_{n+1} = x_{n+1} \mid X_n = x_n) = p_{x_n,x_{n+1}}.$$

Satz 31 (Existenz) Zu jeder Startverteilung α und stochastischen Matrix P gibt es eine Markov-Kette $(X_n)_{n\in\mathbb{N}_0}$.

Satz 32 (Markov-Eigenschaft) Ist $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette, so gilt für alle $n \in \mathbb{N}_0$, $A \in \mathcal{P}(S)^{\otimes\mathbb{N}_0}$, $B \subseteq S^n$, $x \in S$

$$\mu((X_n, X_{n+1}, \ldots) \in A \mid (X_0, \ldots, X_{n-1}) \in B, X_n = x) = \mu((X_0, X_1, \ldots) \in A \mid X_0 = x).$$

Dabei bezeichnet $\mathcal{P}(S)^{\otimes\mathbb{N}_0}$ die Produkt-σ-Algebra $\mathcal{P}(S) \otimes \mathcal{P}(S) \otimes \ldots$.

Definition 35 (Stationäre Verteilung) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette. Ein Wahrscheinlichkeitsmaß π auf dem Zustandsraum S heißt stationär, falls $\pi P = \pi$, das heißt $\sum_{y\in S} \pi(y)p_{yx} = \pi(x)$ für alle $x \in S$ gilt.

Satz 33 (Chapman-Kolmogorov) Ist $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette mit Übergangsmatrix P und $n, m \in \mathbb{N}_0$ beliebig. Definiere die n-Schritt Übergangswahrscheinlichkeit als

$$p^n(x_0, x_n) = \mu(X_n = x_n \mid X_0 = x_0).$$

Dann gilt

$$p^{n+m}(x, y) = \sum_z p^n(x, z)p^m(z, y) = P^{n+m}_{x,y}.$$

4.2 Starke Markov-Eigenschaft und Rekurrenz

Definition 36 (Stoppzeit) Es sei $(X_n)_{n\in\mathbb{N}_0}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{A}, \mu)$. Eine Funktion $\tau : \Omega \to \mathbb{N}_0 \cup \{+\infty\}$ heißt Stoppzeit bezüglich $(X_n)_{n\in\mathbb{N}_0}$, falls $\{\tau = n\} \in \sigma(X_0, \ldots, X_n) = \mathcal{F}_n$ für jedes $n \in \mathbb{N}_0$.

Beispiel 8 (Eintrittszeit) Sei $A \subset S$ und $\tau_A : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ definiert durch $\tau_A(\omega) = \inf\{n \geq 0 : \ X_n(\omega) \in A\}$ mit $\inf \emptyset = \infty$. τ_A ist eine Stoffzeit, da $\{\tau_A = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\} \in \mathcal{F}_n$ für jedes $n \in \mathbb{N}_0$.

Beispiel 9 (Rückkehrzeit) Sei $x \in S$ und $\tau_x : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ definiert durch $\tau_x(\omega) = \inf\{n \geq 1 : \ X_n(\omega) = x\}$ mit $\inf \emptyset = \infty$. τ_x ist eine Stoffzeit, da $\{\tau_i = n\} = \{X_1 \neq x, \ldots, X_{n-1} \neq x, X_n = x\} \in \mathcal{F}_n$ für jedes $n \in \mathbb{N}_0$. Man beachte, dass im Gegensatz zur Eintrittszeit aus $X_0 = x$ nicht $\tau_x = 0$ folgt.

Satz 34 (Starke Markov-Eigenschaft) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette und τ eine Stoppzeit mit $\mu(\tau < \infty) = 1$ und $(X_{\tau+n})_{n\in\mathbb{N}_0}$ durch $X_{\tau+n}(\omega) := X_{\tau(\omega)+n}(\omega)$ definiert. Dann gilt für alle $A \in \mathcal{P}(S)^{\mathbb{N}_0}$, $x \in S$ und

$$B \in \mathcal{F}_\tau := \{B \in \mathcal{A} : \forall n \in \mathbb{N}_0 \cup \{+\infty\} : B \cap \{\tau \leq n\} \in \mathcal{F}_n\}$$

die starke Markov-Eigenschaft

$$\mu((X_\tau, X_{\tau+1}, \ldots) \in A \mid B, X_\tau = x) = \mu((X_0, X_1, \ldots) \in A \mid X_0 = x).$$

Definition 37 (Rekurrenz, Transienz) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette. Ein Zustand $z \in S$ heißt rekurrent (transient), falls $\mu_z(\tau_z < \infty) = 1$ ($\mu_z(\tau_z < \infty) < 1$), wobei $\tau_z := \inf\{n \geq 1 : X_n = z\}$ die Rückkehrzeit nach z bezeichnet und $\mu_z(A) := \mu(A \mid X_0 = z)$ für $A \in \mathcal{A}$.

Proposition 16 (Kriterium) Es sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette und $N_z := \sum_{n=0}^{\infty} \chi_{\{X_n = z\}}$ die Anzahl der Besuche im Zustand $z \in S$.
i) Ist z rekurrent, so gilt $\mu_z(\limsup_{n\to\infty}\{X_n = z\}) = 1$, $\mathbb{E}_z(N_z) = \infty$.
i) Ist z transient, so gilt $\mu_z(\limsup_{n\to\infty}\{X_n = z\}) = 0$, $\mathbb{E}_z(N_z) = \frac{1}{1-\mu_z(\tau_z<\infty)}$.

Definition 38 (kommunizierend, irreduzibel) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette. $x, y \in S$ heißen kommunizierend, $x \leftrightarrow y$, falls es $n, m \in \mathbb{N}_0$ gibt mit $(P^n)_{xy}, (P^m)_{yx} > 0$. Falls für alle $x, y \in S$ gilt, dass $x \leftrightarrow y$, dann heißt die Kette irreduzibel.

Lemma 17 Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette.
i) $\leftrightarrow$ ist eine Äquivalenzrelation auf S.
ii) Rekurrenz und Transienz sind Klasseneigenschaften.
iii) Rekurrente Klassen sind abgeschlossen: $\sum_{y\in R} p_{xy} = 1$ für alle $x \in R$.

4.3 Existenz einer stationären Verteilung

Satz 35 (Zerlegung des Zustandsraums) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette. Es gibt eine disjunkte Zerlegung des Zustandsraums $S = T \cup \bigcup_{l\in L} R_l$, wobei $L \subseteq \mathbb{N}$, T die Menge der transienten Zustände und R_l abgeschlossene Klassen rekurrenter Zustände sind.

Proposition 17 Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette mit stationärer Startverteilung π. Dann sind alle X_n π-verteilt und es gilt

$$\mu((X_n, X_{n+1}, \ldots) \in A) = \mu((X_0, X_1, \ldots) \in A)$$

für alle $A \in \mathcal{P}(S)^{\otimes \mathbb{N}_0}$ und $n \in \mathbb{N}_0$.

Definition 39 (Positiv und Null-rekurrent) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette. Ein rekurrenter Zustand $z \in S$ mit $\mathbb{E}_z(\tau_z) < \infty$ heißt positiv rekurrent. Ein rekurrenter Zustand $z \in S$ mit $\mathbb{E}_z(\tau_z) = \infty$ heißt Null-rekurrent.

Satz 36 (Charakterisierung) Für irreduzible Markov-Ketten sind äquivalent:

i) Es gibt einen positiv rekurrenten Zustand $z \in S$.
ii) Es gibt eine stationäre Verteilung π.
iii) Alle Zustände sind positiv rekurrent.

Für die stationäre Verteilung gilt $\pi(x) = 1/\mathbb{E}_x(\tau_x)$ für alle $x \in S$.

Satz 37 (Invariantes Maß) Sei (X_n) eine irreduzible and rekurrente Markov-Kette und $x \in S$ ein beliebiger Zustand. Definiere $\phi = (\phi(y))_{y\in S}$ durch

$$\phi(y) = \mathbb{E}_x \left[\sum_{n=1}^{T_x} 1_{\{X_n = y\}} \right], \tag{1}$$

die erwartete Anzahl von Besuchen in y vor Rückkehr nach x. Dann gilt

i) $0 < \phi(y) < \infty$ für alle $y \in S$ und $\phi(x) = 1$.
ii) $\phi = \phi P$.
iii) Falls $\nu = \nu P$ für ein Maß ν, dann $\nu = \alpha\phi$ mit $\alpha \in \mathbb{R}$.
Dabei ist $Z = \sum_{y\in S} \phi(y) < \infty$ (nur) für positiv rekurrente Ketten; in diesem Fall ist dann $\pi(x) = \phi(x)/Z$ eine stationäre Verteilung.

Beweis: Ad i) Wegen der Rekurrenz von x und Definition von ϕ haben wir

$$\begin{aligned}
\phi(x) &= \mathbb{E}_x \left[\sum_{n=1}^{T_x} 1_{\{X_n = x\}} \right] = \sum_{n=1}^{\infty} \mathbb{E}_x [1_{\{X_n = x\}} 1_{\{n \le T_x\}}] \\
&= \sum_{n=1}^{\infty} \mu_x [X_n = x, n \le T_x] = \sum_{n=1}^{\infty} \mu_x [T_x = n] = \mu_x [T_x < \infty] = 1.
\end{aligned}$$

Die Beschränktheit von ϕ verschieben wir auf später. Für ii) wählen wir zuerst ein $n \in \mathbb{N}$ und beobachten, dass $\{T_x \ge n\}$ nur von $X_0, X_1, \ldots, X_{n-1}$ abhängt. Daher

$$\mu_x[X_n = z, X_{n-1} = y, T_x \ge n] = \mu_x[X_{n-1} = y, T_x \ge n]\, P(y, z).$$

Daher für beliebiges $z \in S$

$$\begin{aligned}
\sum_{y\in S} \phi(y) P(y, z) &= \phi(x) P(x, z) + \sum_{y \ne x} \mu(y) P(y, z) \\
&= P(x, z) + \sum_{y \ne x} \sum_{n=1}^{\infty} \mu_x [X_n = y, n \le T_x]\, P(y, z)
\end{aligned}$$

$$
\begin{aligned}
&= P(x,z) + \sum_{n=1}^{\infty} \sum_{y \neq x} \mu_x[X_{n+1} = z, X_n = y, n \le T_x] \\
&= \mu_x[X_1 = z] + \sum_{n=1}^{\infty} \mu_x[X_{n+1} = z, n+1 \le T_x] \\
&= \mu_x[X_1 = z, 1 \le T_x] + \sum_{n=2}^{\infty} \mu_x[X_n = z, n \le T_x] \\
&= \sum_{n=1}^{\infty} \mu_x[X_n = z, n \le T_x] \; = \; \phi(z),
\end{aligned}
$$

wo wir für die vierte Ungleichung benutzten, dass $X_n = y$, $n \le T_x$ and $x \neq y$ sofort $n+1 \le T_x$ implizieren. Unsere Gleichungskette zeigt $\phi P = \phi$. Nun setzen wir i) fort: Da P irreduzible ist, existieren ganze Zahlen $k, j \in \mathbb{N}$ so dass $P^k(x,y) > 0$ and $P^j(y,x) > 0$ für jedes $y \in S$. Daher für jedes $k \in \mathbb{N}$ und unter Ausnutzung von ii):

$$0 < \phi(x) P^k(x,y) \le \sum_{z \in S} \phi(z) P^k(z,y) = \phi(y).$$

Andererseits

$$\phi(y) = \frac{\phi(y) P^j(y,x)}{P^j(y,x)} \le \frac{\sum_{z \in S} \phi(z) P^j(z,x)}{P^j(y,x)} = \frac{\phi(x)}{P^j(y,x)} < \infty,$$

womit i) bewiesen ist. Den Teil iii) sparen wir uns hier.

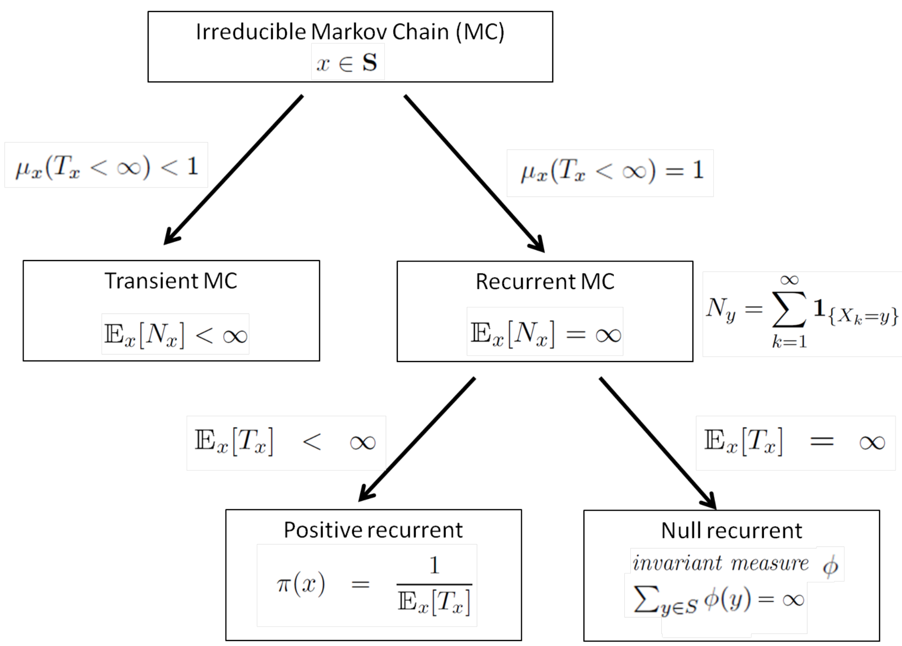

Abbildung 2: Verschiedene irreduzible Markov-Ketten.

4.4 Kopplung von Markov-Ketten

Definition 40 (Kopplungspaar) Seien $(X_n)_{n \in \mathbb{N}_0}$ und $(Y_n)_{n \in \mathbb{N}_0}$ zwei Markov-Ketten mit gleichem Zustandsraum S. (X_n) und (Y_n) bilden ein Kopplungspaar, falls es eine fast sicher endliche Stoppzeit τ gibt, so dass für alle $\omega \in \Omega$ gilt

$$n \ge \tau(\omega) \Rightarrow X_n(\omega) = Y_n(\omega).$$

τ heisst dann Kopplungszeit.

Satz 38 (Gleichverteilung von Kopplungspaaren) Sei τ eine Kopplungszeit von zwei Markov-Ketten (X_n) und (Y_n) mit gleichem Zustandsraum S. Dann gilt

$$\mu(X_n \in A) - \mu(Y_n \in A) \to 0, \quad \forall A \in \mathcal{A}, \quad n \to \infty.$$

Beweis durch Zerlegung in Zeiten vor und nach τ:

$$\begin{aligned} |\mu(X_n \in A) - \mu(Y_n \in A)| &= |\mu(X_n \in A, \tau \leq n) + \mu(X_n \in A, \tau > n)) \\ &\quad -\mu(Y_n \in A, \tau \leq n) - \mu(Y_n \in A, \tau > n)| \\ &= |\mu(X_n \in A, \tau > n)) - \mu(Y_n \in A, \tau > n)| \\ &\leq 2\,\mu(\tau > n)) \to 0, \end{aligned}$$

da die Stoppzeit fast sicher endlich ist.

Definition 41 (Produkt-Markov-Kette) Seien $(X_n)_{n\in\mathbb{N}_0}$ und $(Y_n)_{n\in\mathbb{N}_0}$ zwei Markov-Ketten mit gleichem Zustandsraum S. Die Übergangsmatrix und Startverteilung zu (X_n) seien (P, α), die zu (Y_n) seien (Q, β). Die Produkt-Markov-Kette $(Z_n) = (X_n, Y_n)$ hat dann Zustandsraum $S \times S$, die Übergangsmatrix R gegeben durch $R_{(i,j),(k,l)} = P_{i,k}Q_{j,l}$ und die Startverteilung γ gegeben durch $\gamma_{ij} = \alpha_i\beta_j$.

Satz 39 (Unabhängige Kopplung) Gegeben seien zwei unabhängige Markov-Ketten (X_n) und (Y_n) mit gleichem Zustandsraum S, gleicher Übergangsmatrix P und Startverteilungen α und β. (Z_n) sei die zugehörige Produkt-Markov-Kette. Falls (Z_n) irreduzibel und rekurrent ist, gelten die folgenden Aussagen:
i) $T = \inf\{n \geq 1 : X_n = Y_n = i\}$ für ein $i \in S$ ist fast sicher endlich.
ii) (W_n) definiert als $W_n = X_n$ für $n \leq T$ und $W_n = Y_n$ für $n > T$ ist eine Markov-Kette mit Übergangsmatrix P und Startverteilung α.
iii) $\mu(X_n \in A) - \mu(Y_n \in A) \to 0, \quad \forall A \in \mathcal{A}, \quad n \to \infty.$

Beweis: i) folgt aus der Beobachtung, dass T die Rückkehrzeit $T_{(i,i)}$ der Produkt-Markov-Kette (Z_n) in den Zustand (i,i) ist. Da (Z_n) irreduzibel und rekurrent ist, folgt sofort $\mu(T < \infty) = 1$.
ii) ist eine direkte Konsequenz der starken Markov-Eigenschaft. Das besagt auch, dass (X_n) und (W_n) gleich verteilt sind: $\mu(X_n \in A) = \mu(W_n \in A)$.
iii) folgt, da (W_n) und (Y_n) ein Kopplungspaar bilden (da T fast sicher endlich ist). Damit folgt aus Satz 38 und Teil ii) sofort:

$$\mu(X_n \in A) - \mu(Y_n \in A) = \mu(W_n \in A) - \mu(Y_n \in A) \to 0, \qquad n \to \infty.$$

4.5 Ergodensatz, Teil 1

Definition 42 (Aperiodisch) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette. $x \in S$ heißt aperiodisch, wenn 1 der größte gemeinsame Teiler von $\{n \geq 1 : (P^n)_{xx} > 0\}$ ist.

Lemma 18 Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette.
i) $x \in S$ ist genau dann aperiodisch, wenn $\exists n_0 \forall n \geq n_0 : (P^n)_{xx} > 0$.
ii) Aperiodizität ist eine Klasseneigenschaft.

Satz 40 (Ergodensatz, Teil 1) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine irreduzible, aperiodische Markov-Kette mit stationärer Verteilung π und Übergangsmatrix P. Dann gilt $\mu(X_n \in A) \to \pi(A)$ für alle $A \subseteq S$, also insbesondere $(P^n)_{xy} \to \pi(y)$ für alle $x, y \in S$.

Beweis: Sei (Y_n) eine von (X_n) unabhängige Markov-Kette auf selbem Zustandsraum, gleicher Übergangsmatrix P und Startverteilung π. Dann gilt $\mu(Y_n = i) = \pi(i)$. Die Produkt-Markov-Kette $(Z_n) = (X_n, Y_n)$ ist irreduzibel, da (X_n) und (Y_n) irreduzibel und aperiodisch sind. Das sieht man wie folgt: Aperiodizität impliziert die Existenz eines r mit $P^r_{jj}, P^r_{ll} > 0$. Irreduzibilität liefert ein n mit $P^n_{ij}, P^n_{kl} > 0$. Daraus folgt

$$R^{n+r}_{(i,k),(j,l)} = P^{n+r}_{ij} P^{n+r}_{kl} \geq P^n_{ij} P^r_{ij} P^n_{kl} P^r_{ll} > 0.$$

Die Produkt-Kette besitzt ausserdem die stationäre Verteilung $\nu(i, j) = \pi(i)\pi(j)$ und ist daher (positiv) rekurrent. Dann folgt aus Satz 39 sofort

$$\mu(X_n \in A) - \mu(Y_n \in A) = \mu(X_n \in A) - \pi(A) \to 0, \qquad n \to \infty.$$

4.6 Ergodensatz, Teil 2

Satz 41 (Ergodensatz, Teil 2) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine irreduzible, aperiodische, rekurrente Markov-Kette mit $\mathbb{E}_x(\tau_x) = \infty$ für alle $x \in S$, d.h. alle Zustände sind Null-rekurrent. Dann existiert keine positive stationäre Verteilung und es gilt $(P^n)_{xy} \to 0$ aber $\sum_{n=0}^{\infty}(P^n)_{xy} = \infty$ für alle $x, y \in S$.

Satz 42 (Klassifikation) Für eine irreduzible, aperiodische Markov-Kette X gilt genau einer der folgenden Fälle:

i) X ist transient. Dann gilt $\forall x, y \in S : \sum_{n=0}^{\infty}(P^n)_{xy} < \infty, \lim_{n\to\infty}(P^n)_{xy} = 0$.

ii) X ist rekurrent. Dann gilt $\forall x, y \in S : \sum_{n=0}^{\infty}(P^n)_{xy} = \infty$. Eine stationäre Verteilung π gibt es genau dann, wenn X positiv rekurrent ist.

Ist X positiv rekurrent, so gilt $\forall x, y \in S : \lim\limits_{n\to\infty}(P^n)_{xy} = \pi(y) = 1/\mathbb{E}_y(\tau_y) > 0$.

Ist X nullrekurrent, so gilt $\forall x, y \in S : \lim\limits_{n\to\infty}(P^n)_{xy} = 1/\mathbb{E}_y(\tau_y) = 0$.

4.7 Gesetz der großen Zahlen

Satz 43 (Starkes Gesetz der großen Zahlen für Markov Ketten) Sei (X_n) eine irreduzible Markov-Kette mit stationärer Verteilung π und $f \in L^1(\pi)$ auf diskretem Zustandsraum S. Für jeden Anfangszustand $x \in S$, d.h. für $X_0 = x$ gilt

$$\frac{1}{n+1}\sum_{k=0}^{n} f(X_k) \longrightarrow \mathbb{E}_\pi[f] = \sum_{x\in S} f(x)\pi(x) \tag{2}$$

für $n \to \infty$ und μ_x–fast sicher.

Beweis: Laut der Voraussetzungen ist (X_n) irreduzible und positiv rekurrent, so dass $\nu(y) = \mathbb{E}_x[\sum_{n=0}^{T_x} \chi_{\{X_n=y\}}]$ ein invariantes Maß definiert, welches mit der stationären Verteilung über

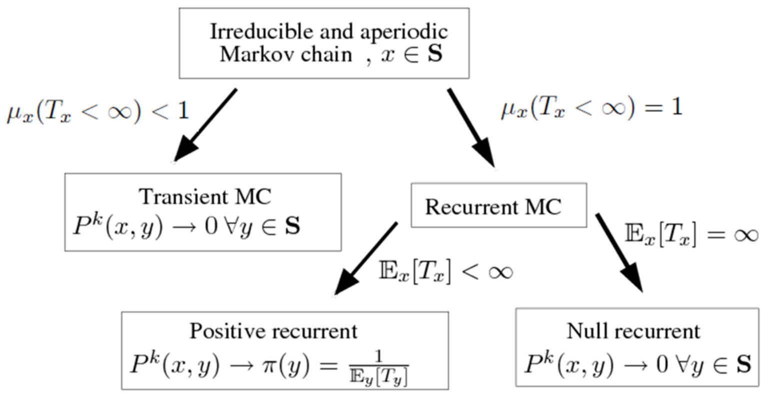

Abbildung 3: Verschiedene irreduzible, aperiodische Markov-Ketten.

$\pi(y) = \frac{\nu(y)}{Z}$ zusammenhängt, wobei $Z = \sum_{y \in S} \nu(y)$, vgl. Satz 37. Für die Zufallsvariable $U_0 = \sum_{k=0}^{T_x} f(X_k)$ finden wir den Erwartungswert

$$\begin{aligned} \mathbb{E}[U_0] &= \mathbb{E}_x\left[\sum_{k=0}^{T_x} f(X_k)\right] = \mathbb{E}_x\left[\sum_{k=0}^{T_x}\sum_{y\in S} f(y)\chi_{\{X_k=y\}}\right] \\ &= \sum_{y\in S} f(y)\mathbb{E}_x\left[\sum_{k=0}^{T_x}\chi_{\{X_k=y\}}\right] = \sum_{y\in S} f(y)\nu(y) \end{aligned} \tag{3}$$

Betrachte $U_p = \sum_{k=\tau_p+1}^{\tau_{p+1}} f(X_k)$ mit $p \geq 1$ und $T_x = \tau_0, \tau_1, \tau_2, \ldots$ die sukzessiven Rückkehrzeiten nach x. Nach der starken Markov-Eigenschaft sind die $U_0, U_1, U_2, \ldots$ iid Zufallsvariablen (man betrachte sie als parallel in $X_0 = x$ gestartete unabhängige Ketten). Da laut (3) auch gilt, dass $\mathbb{E}[|U_0|] < \infty$, können wir das starke Gesetz der grossen Zahlen für unabhängige Zufallsvariablen anwenden und erhalten fast sicher, dass

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n} U_k = \sum_{y\in S} f(y)\nu(y) \Leftrightarrow \lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{\tau_{n+1}} f(X_k) = \sum_{y\in S} f(y)\nu(y).$$

Nun nehmen wir für den Augenblick an, dass $f \geq 0$ und definieren $N_x(n) := \sum_{k=0}^{n}\chi_{\{X_k=x\}}$, die Anzahl der Besuche in x in den ersten n Schritten. Wegen $\tau_{N_x(n)} \leq n < \tau_{N_x(n)+1}$ und $f \geq 0$ folgt dann

$$\frac{1}{N_x(n)}\sum_{k=0}^{\tau_{N_x(n)}} f(X_k) \leq \frac{1}{N_x(n)}\sum_{k=0}^{n} f(X_k) \leq \frac{1}{N_x(n)}\sum_{k=0}^{\tau_{N_x(n)+1}} f(X_k). \tag{4}$$

Da unsere Markov-Kette rekurrent ist haben wir $\lim_{n\to\infty} N_x(n) = \infty$, so dass die obere und untere Schranke in (4) gegen $\sum_{y\in S} f(y)\nu(y)$ konvergiert und daher liefert, dass

$$\lim_{n\to\infty}\frac{1}{N_x(n)}\sum_{k=0}^{n} f(X_k) = \sum_{y\in S} f(y)\nu(y) = Z\sum_{y\in S} f(y)\pi(y).$$

Nun betrachte $g \equiv 1$ (eine positive Funktion mit $g \in L^1(\pi)$). Aus obigem folgt

$$\lim_{n\to\infty}\frac{1}{N_x(n)}\sum_{k=0}^{n} g(X_k) = \lim_{n\to\infty}\frac{n+1}{N_x(n)} = Z \Rightarrow \lim_{n\to\infty}\frac{N_x(n)}{n+1} = \frac{1}{Z},$$

und letztlich

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n} f(X_k) &= \lim_{n\to\infty}\frac{1}{N_x(n)}\frac{N_x(n)}{n+1}\sum_{k=0}^{n} f(X_k)\\ &= \frac{1}{Z}\sum_{y\in S} f(y)\nu(y) = \sum_{y\in S} f(y)\pi(y).\end{aligned}$$

Für beliebiges f betrachte $f^+ = \max(0, f)$ und $f^- = \max(0, -f)$ und dann die Differenz der entsprechenden Grenzwerte.

4.8 Zeitumkehr und Reversibilität

Definition 43 (Zeit-Umkehr) Sei (X_n) eine Markov-Kette mit Übergangsmatrix P und stationärer Verteilung $\pi > 0$. Dann heißt die Markov Kette $\{Y_n\}$ mit Übergangsmatrix Q definiert durch

$$Q(y, x) = \frac{\pi(x)P(x, y)}{\pi(y)} \tag{5}$$

die zeit-umgekehrte Markov-Kette assoziiert mit (X_n).

Definition 44 (reversibel) Sei (X_n) eine Markov-Kette mit Übergangsmatrix P und stationärer Verteilung $\pi > 0$. Die zeit-umgekehrte Markov-Kette habe Übergangsmatrix Q. Dann heißt (X_n) reversibel in Bezug auf π, wenn $P(x, y) = Q(x, y)$ für alle $x, y \in S$.

Satz 44 (detailed balance) Sei (X_n) eine Markov-Kette mit Übergangsmatrix P und $\pi > 0$ eine gegebene Verteilung. Sei die detailed balance Bedingung erfüllt,

$$\pi(x)P(x, y) = \pi(y)P(y, x), \qquad \forall x, y \in S.$$

Dann ist (X_n) reversibel in Bezug auf und π ist eine stationäre Verteilung von (X_n), i.e., $\pi P = \pi$.

4.9 Markov-Ketten Monte-Carlo (MCMC)

Bemerkung 7 (Aufgabenstellung MCMC) Wir konzentieren uns hier auf sogenanntes *Metropolis MCMC*. Die Grundlage eines MCMC-Verfahrens ist eine gegebene Verteilung π in einem Zustandsraum S. In vielen Fällen kann man $\pi(x)$ für gegebenes x nicht berechnen; das ist vor allem dann der Fall, wenn π die Form $\pi(x) = \frac{1}{Z}\mu(x)$ mit einer Normalisierungkonstante $Z = \sum_{x\in S}\mu(x)$ hat, wo $\mu(x)$ zwar leicht explizit berechnet werden kann, Z wegen der Größe des Zustandsraums aber nicht. Für ein solches π sollen Erwartungswerte $\mathbb{E}_\pi[f] = \sum_{x\in S} f(x)\pi(x)$ für eine gegebene Funktion $f \in L^1(\pi)$ berechnet werden.

Bemerkung 8 (Grundidee MCMC) Wir wollen eine Markov-Kette (X_n) konstruieren, die (B1) π als stationäre Verteilung hat und (B2) deren Realisierungen lediglich die Berechnung von μ erfordern und nicht die von π bzw. Z.

Diese Kette soll irreduzibel sein, so dass das starke Gesetz der großen Zahlen die Berechnung der Erwartungswerte mittels

$$S_n(f) = \frac{1}{n+1}\sum_{k=0}^{n} f(X_k) \longrightarrow \mathbb{E}_\pi[f]$$

erlaubt.

Definition 45 (Vorschlagsmatrix) Sei Q die Übergangsmatrix einer irreduziblen Markov-Kette auf S. Q heißt Vorschlagsmatrix, falls $Q_{x,y} \neq 0 \Leftrightarrow Q_{y,x} \neq 0$.

Definition 46 (Metropolis-Funktion) Eine Funktion $\Psi : (0,\infty) \to (0,1]$ mit $\frac{\Psi(x)}{\Psi(1/x)} = x$ für alle $x \in (0,\infty)$ heißt Metropolis-Funktion.

Bemerkung 9 (Beispiel für eine Metropolis-Funktion) $\Psi(x) = \min\{1, x\}$ ist eine Metropolis-Funktion.

Satz 45 (Metropolis-MCMC) Sei $\pi > 0$ eine Wahrscheinlichkeitsverteilung und Q eine Vorschlagsmatrix auf S und Ψ eine Matropolis-Funktion. Die *Akzeptanz-Funktion* $A : S \times S \to (0,1]$ sei gegeben durch $A(x,y) = 0$ falls $Q(x,y) = 0$ und falls $Q(x,y) \neq 0$ durch

$$A(x,y) = \Psi\Big(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\Big) \in (0,1].$$

Die Übergangsmatrix P sei definiert durch

$$P(x,y) = \begin{cases} Q(x,y)A(x,y) & \text{if } x \neq y \\ 1 - \sum\limits_{z\in S, z\neq x} Q(x,z)A(x,z) & \text{if } x = y \end{cases} . \tag{6}$$

Dann ist die von P auf S erzeugte Markov-Kette irreduzibel und hat π als stationäre Verteilung, d.h., es gilt $\pi P = \pi$.

Beweis: Zuerst zeigen wir, dass P eine Übergangsmatrix ist:

$$\sum_{y\in S} P(x,y) = \sum_{z\in S, z\neq x} Q(x,z)A(x,z) + 1 - \sum_{z\in S, z\neq x} Q(x,z)A(x,z) = 1.$$

Da ausserdem $\pi > 0$ and $Q \geq 0$ per definitionem, haben wir $A \geq 0$ and $P \geq 0$. Ausserdem folgt aus $Q(x,y) > 0$ für $x \neq y$ auch $Q(y,x) > 0$ und daraus $A(x,y) > 0$, $A(y,x) > 0$ and $P(x,y) > 0$, $P(y,x) > 0$. P erbt also die Irreduzibilität von Q.

Wählen wir also $x, y \in S$ so dass $Q(x,y) \neq 0$, dann liefert uns die Eigenschaft von Ψ:

$$\frac{A(x,y)}{A(y,x)} = \frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)},$$

und daher

$$\pi(x)P(x,y) = \pi(y)P(y,x). \tag{7}$$

Das gleich gilt auch für $x = y$. Falls $x \neq y$ so dass $Q(x,y) = 0$ dann $P(x,y) = 0$ und Satz (7) gilt ebenfalls. Also gilt die detailed balance Bedingung (7) für alle $x, y \in S$. Daher ist π laut 44 eine stationäre Verteilung von (X_n) und $\pi P = \pi$.

Algorithmus 2 (MCMC Algorithmus) Die Berechnung einer Realisierung $(x_k)_{k=0,1,2,\ldots}$ der Markov Kette (X_n) ist sehr einfach:

1. Starte in x_0 mit Index $k = 0$.
2. x_k sei der derzeitige Zustand.
3. Ziehe y gemäß der Verteilung $Q(x_k, \cdot)$.
4. Berechne die Akzeptanz-Wahrscheinlichkeit
$$a = A(x_k, y) = \Psi\Big(\frac{\pi(y)Q(y, x_k)}{\pi(x_k)Q(x_k, y)}\Big) = \Psi\Big(\frac{\mu(y)Q(y, x_k)}{\mu(x_k)Q(x_k, y)}\Big).$$
5. Ziehe $r \in [0, 1)$ zufällig gemäß der Gleichverteilung auf $[0, 1)$.
6. Setze
$$x_{k+1} = \begin{cases} y & \text{falls } r \le a \\ x_k & \text{falls } r > a \end{cases}.$$
7. Setze $k := k + 1$ und gehe zurück zu Schritt 2.

Bemerkung 10 (Symmetrischer Vorschlag) Falls Q eine symmetrische Matrix ist, also $Q(x, y) = Q(y, x)$ gilt für alle Paare $x, y \in S$, dann bekommt die Akzeptanz-Funktion die besonders einfache Form $A(x_k, y) = \Psi\Big(\frac{\mu(y)}{\mu(x_k)}\Big)$.

Bemerkung 11 Tatsächlich benötigen wir die Normalisierungs-Konstante Z von $\pi = \mu/Z$ nicht; sondern lediglich Verhältnisse der Form $\mu(x)/\mu(y)$.

4.10 Anwendung von MCMC

Bemerkung 12 (Ising-Modell: Gitter, Spins, Zustandsraum) Wir betrachten das sogenannte Ising-Modell, das als einfachstes Modell zur Analyse von Magnetisierungprozessen in Kristallen verwendet wird. Im Ising-Modell ist ein reguläres Gitter G gegeben, im einfachsten Fall in 2d z.B. von der Form $G_N = \{(i, j) \in \mathbb{N}\times\mathbb{N}, i, j \le n\}$ mit $N = n^2$ vielen Gitterplätzen. An jedem Gitter-Platz sitzt ein sogenannter Spin (Elementarmagnet), der lediglich die beiden Werte 1 oder -1 annehmen kann. Wenn wir die Gitterplätze nummerieren, $i = 1, \ldots, N$, dann haben wir den Zustandsraum

$$S = \{s = (s_1, \ldots, s_N) : \ s_i = \pm 1, \ i = 1, \ldots, N\},$$

der (für grosse N) sehr groß sein kann: $|S| = 2^N = 2^{n^2}$.

Bemerkung 13 (Ising-Modell: Nachbar-Spins, Energie) Für jeden Spin i definieren wir eine Nachbarschaft $\mathcal{N}_i \subset \{1, \ldots, N\}$. Typischerweise ist die Anzahl von Spins in der Nachbarschaft eine Konstante des Modells, d.h. $|\mathcal{N}_i| = m$ unabhängig von i, mit möglichen Ausnahmen an den Rändern des Gitters. Nur

Spins aus $\mathcal{N}_i$ interagieren direkt mit Spin i. Die Energie eines Spin-Zustandes $s \in S$ ist gegeben durch

$$E(s) = -\frac{1}{2} \sum_{\substack{i=1,\dots,N \\ j \in \mathcal{N}_i}} J_{ij} s_i s_j \; - \; \mathcal{M} \sum_{i=1}^{N} s_i,$$

wobei der Skalar $\mathcal{M}$ das externe magnetische Feld bezeichnet und J_{ij} die sogenannte Kopplungskonstanten, die normalerweise nicht von ij abhängen, $J_{ij} = J$. Die ferromagnestische Wechselwirkung ($J > 0$) versucht die Energie zu minimieren, d.h. benachbarte Spins parallel auszurichten.

Bemerkung 14 (Ising-Modell: Stationäre Verteilung) Thermischen Fluktuationen der Umgebung ändern den Spinzustand des Systems. Die zugehörige stationäre Verteilung ist gegeben durch

$$\pi(s) = \frac{1}{Z}\mu(s), \quad \mu(s) = \exp(-\beta E(s)), \quad Z = \sum_{s \in S} \mu(s),$$

wobei $\beta = (k_B T)^{-1} > 0$ die sogenannte inverse Temperatur ist (T ist die physikalische Temperatur des Magneten und k_B eine physikalische Konstante). In diesem Fall können wir für große N die Normalisierungkonstante Z nicht direkt berechnen und daher auch nicht die Erwartungswerte $\mathbb{E}_\pi(f)$ für physikalisch interessante Observablen $f : S \to \mathbb{R}$.

Bemerkung 15 (Ising-Modell: Magnetisierung und spezifische Wärme) Als physikalisch interessante Observablen $f : S \to \mathbb{R}$ werden meistens betrachtet: (1) die Magnetisierung $M(s) = \sum_i s_i$, (2) die Energie $E(s)$, (3) die spezifische Wärme $c_v(s) = \beta(E(s) - \mathbb{E}_\pi(E))^2/T$ und (4) die magnetische Suszeptibilität $\chi(s) = \beta(M(s) - \mathbb{E}_\pi(M))^2$. Aus der Abhängigkeit der zugehörigen Erwartungswerte von externen Parametern, z.B. von der Temperatur, läßt sich viel über die physikalischen Eigenschaften des Magneten ablesen. Also interessieren wir uns in aller erster Linie für die Temperatur-Abhängigkeit der ersten zwei Momente von E und M.

Definition 47 (MCMC Vorschlagsmatrix) Um die Vorschlagsmatrix zu definieren, betrachten wir zuerst die Spin-Zustände, die durch Umkehrung eines einzelnen Spins aus dem Zustand $s \in S$ hervorgehen:

$$\begin{aligned} \text{switch}(s) = \{s' \in S; \quad & s'_j = s_j, \text{ mit Ausnahme eines einzelnen } i \in \{1, \dots, N\} \\ & \text{mit } s'_i = -s_i\}, \end{aligned}$$

so dass $|\text{switch}(s)| = N$ für alle $s \in S$. Für den Vorschlag ziehen wir eine Gitterposition i aus einer Gleichverteilung auf $\{1, \dots, N\}$ und gehen dann von s nach $s' \in \text{switch}(s)$ durch Umkehrung des Spins in i:

$$Q(s, s') = \begin{cases} 1/N & \text{falls } s' \in \text{switch}(s) \\ 0 & \text{sonst} \end{cases}.$$

Offensichtlich ist die durch Q definierte Markov-Ketten irreduzibel. Ausserdem finden wir, dass $s' \in \text{switch}(s) \Leftrightarrow s \in \text{switch}(s')$, mit der Konsequenz, dass $Q(s, s') = Q(s', s)$.

Bemerkung 16 (Akzeptanz-Wahrscheinlichkeit) Durch die Definition von Q und mit Metropolis-Funktion $\Psi(x) = \min\{1, x\}$ erhalten wir

$$A(s, s') = \begin{cases} 1 & \text{falls } s' \in \text{switch}(s) \text{ und } \Delta E_{s,s'} \leq 0 \\ \exp\Big(- \beta \Delta E_{s,s'} \Big) & \text{falls } s' \in \text{switch}(s) \text{ and } \Delta E_{s,s'} > 0 \\ 0 & \text{sonst} \end{cases}, \tag{8}$$

mit

$$\Delta E_{s,s'} = E(s') - E(s) = 2 \sum_{j \in \mathcal{N}_i} J_{ij} s_i s_j \; + \; 2\mathcal{M},$$

da $\mu(s')/\mu(s) = \exp\Big(- \beta \Delta E_{s,s'} \Big)$, wobei i die Gitterposition der Spin-Umkehr bezeichnet.

Algorithmus 3 (MCMC-Algorithmus für das Ising-Modell) Der MCMC Algorithmus erhält folgende einfache Form:

1. Starte in Spinzustand $s^{(0)}$ mit Iteration Index $k = 0$.
2. Sei $s^{(k)}$ der gegenwärtige Zustand.
3. Bestimme den Vorschlag $s' \in \text{switch}(s^{(k)})$ durch Spinumkehr in Position i, wobei i zufällig gleichverteilt aus $\{1, \ldots, N\}$ gezogen wird.
4. Berechne $\Delta E_{s^{(k)},s'} = E(s') - E(s^{(k)}) = 2 \sum_{j \in \mathcal{N}_i} J_{ij} s_i^{(k)} s_j^{(k)} \; + \; 2\mathcal{M}$.
5. Berechne die Akzeptanz-Wahrscheinlichkeit $a = A(s^{(k)}, s')$ gemäß (8).
6. Ziehe $r \in [0, 1)$ zufällig gleichverteilt aus $[0, 1)$.
7. Setze
$$s^{(k+1)} = \begin{cases} s' & \text{falls } r \leq a \\ s^{(k)} & \text{falls } r > a \end{cases}.$$
8. Setze $k := k + 1$ und kehre zu Schritt 2. zurück.

Bemerkung 17 (Aufwand) Im Vergleich zur Größe des Zustandsraum mit 2^N Zuständen erfordert jede einzelne Iteration des MCMC-Verfahrens sehr wenig Rechenaufwand, $|\mathcal{N}_i|$ Additionen und Multiplikationen und die Bestimmung von zwei Zufallsvariablen.

Bemerkung 18 (Verallgemeinerung) Wann immer die betrachtete stationäre Verteilung die Form $\pi(x) = \exp(-\beta E(x))/Z$ hat, wird die Akzeptanzfunktion im Wesentlichen die obige Form haben. Solche Fälle sind in der physikalischen, chemischen und biologischen Literatur sehr häufig und machen die obigen Form von MCMC zu einem der am häufigsten verwendeten Algorithmen der heutigen Zeit.

5 Martingale

Literatur: [Dud, Dur, MS]

5.1 Bedingte Erwartung

Bemerkung 19 (Motivation) Betrachte eine Zufallsvariable Z auf dem W-Raum $(\Omega, \mathcal{A}, \mu)$. Für ein $A \in \mathcal{A}$ mit $\mu(A) > 0$ erhalten wir mittels der Definition der bedingten Wahrscheinlichkeit sofort $\mathbb{E}(Z|A) = \mathbb{E}(Z\chi_A)/\mu(A)$. Für eine weitere reellwertige Zufallsvariable X mit diskretem Träger erlaubt uns das, die Funktion $g : \mathbb{R} \to \mathbb{R}$ wie folgt zu definieren: $g(x) = \mathbb{E}(Z|X = x)$ falls $\mu(X = x) > 0$ und $g(x) = 0$ falls $\mu(X = x) = 0$. Damit wird dann die Zufallsvariable $Y = g \circ X$ definiert. Diese ist $\sigma(X)$-meßbar und erfüllt für alle $B \in \sigma(X)$:

$$\mathbb{E}(Z\chi_{\{X\in B\}}) = \sum_{\substack{x\in B\\ \mu(X=x)>0}} \mathbb{E}(Z\chi_{\{X=x\}}) = \sum_{\substack{x\in B\\ \mu(X=x)>0}} g(x)\mu(X = x) = \mathbb{E}(Y\chi_{\{X\in B\}}).$$

Definition 48 (Bedingte Erwartung) Seien $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum, $\mathcal{G} \subseteq \mathcal{A}$ eine σ-Algebra und $Z \in \mathcal{L}^1(\mu)$. Eine relle Zufallsvariable Y heißt bedingte Erwartung von Z unter $\mathcal{G}$, wenn Y $\mathcal{G}$-meßbar ist und $\mathbb{E}(Z\chi_B) = \mathbb{E}(Y\chi_B)$ für alle $B \in \mathcal{G}$ gilt.

Satz 46 (Existenz & Eindeutigkeit) Sei $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum, $\mathcal{G} \subseteq \mathcal{A}$ eine σ-Algebra und $Z \in \mathcal{L}^1(\mu)$. Dann existiert eine bedingte Erwartung Y von Z unter $\mathcal{G}$. Y ist fast sicher eindeutig bestimmt. (Man schreibt $Y = \mathbb{E}(Z|\mathcal{G})$.)

Beweis: Die Existenz ergibt sich aus Radon-Nikodym: O.B.d.A setzen wir $Z \geq 0$ voraus; dann definiert $\nu(B) = \int_B Z d\mu$ für alle $B \in \mathcal{G}$ ein Maß auf $\mathcal{G}$. ν ist absolut stetig bzgl μ und endlich, da $Z \in \mathcal{L}^1(\mu)$. Dann ergibt Radon-Nikodym, das ein $\mathcal{G}$-meßbares f existiert, so dass

$$\mathbb{E}(Z\chi_B) = \nu(B) = \int_B f d\mu = \mathbb{E}(f\chi_B), \qquad B \in \mathcal{G},$$

so dass f ein Version von $\mathbb{E}(Z|\mathcal{G})$ ist. Für die fast-sichere Eindeutigkeit seien Y und X zwei Versionen von $\mathbb{E}(Z|\mathcal{G})$. Dann folgt $\int_B (Y - X)d\mu = 0$ für alle $B \in \mathcal{G}$ und damit insbesondere für $B_> = \{Y > X\} \in \mathcal{G}$ und $B_< = \{Y < X\} \in \mathcal{G}$. Daraus folgt aber $\mu(Y = X) = 1$, also die fast-sichere Identität beider.

Beispiel 10 Wir betrachten $Z = (X, Y)$ mit $X, Y : \Omega \to \{1, 2\}$ und $\mu(Z = (i, j)) = p_{ij}$, wobei $p_{11} = p_{22} = 0.4$ und $p_{21} = p_{12} = 0.1$. Dann finden wir mit $\mu(X = 1) = \mu(X = 2) = 0.5$ sofort $\mathbb{E}(Z|X = 1) = (1, 1.2)$ und $\mathbb{E}(Z|X = 2) = (2, 1.8)$. Daher $\mathbb{E}(Z|\sigma(X)) = (X, 0.6 \cdot (1 + X))$.

Satz 47 (Elementare Eigenschaften) Seien $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum, $X, X_1, X_2 \in \mathcal{L}^1(\mu)$ und $\mathcal{G} \subseteq \mathcal{A}$ eine σ-Algebra.

i) $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$.
ii) Ist X $\mathcal{G}$-meßbar, so ist $\mathbb{E}(X|\mathcal{G}) = X$ fast sicher.
iii) Linearität: Für $a, b \in \mathbb{R}$ gilt $\mathbb{E}(aX_1 + bX_2|\mathcal{G}) = a\mathbb{E}(X_1|\mathcal{G}) + b\mathbb{E}(X_2|\mathcal{G})$ fast

sicher.
iv) Monotonie: Aus $X_1 \leq X_2$ folgt $\mathbb{E}(X_1|\mathcal{G}) \leq \mathbb{E}(X_2|\mathcal{G})$ fast sicher.
v) $|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X||\mathcal{G})$ fast sicher.

Satz 48 (Bedingte Konvergenzsätze) Seien $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum, $X, Y, X_n \in \mathcal{L}^1(\mu)$ und $\mathcal{G} \subseteq \mathcal{A}$ eine σ-Algebra.

i) Monotone Konvergenz: Aus $X_n \uparrow X$ f.s. folgt $\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$ f.s.
ii) Lemma von Fatou: Aus $X_n \geq Y$ f.s., $\liminf \mathbb{E}(X_n) < \infty$ folgt $\liminf X_n \in \mathcal{L}^1(\mu)$, $\mathbb{E}(\liminf X_n|\mathcal{G}) \leq \liminf \mathbb{E}(X_n|\mathcal{G})$ f.s.
iii) Dominierte Konvergenz: Aus $X_n \to X$ f.s. und $|X_n| \leq Y$ folgt $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$ f.s.

Satz 49 Seien $X, Y, XY \in \mathcal{L}^1(\mu)$ und $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{A}$ σ-Algebren.

i) Projektionseigenschaft: $\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G})$ f.s.
ii) Produkteigenschaft: Ist X $\mathcal{G}$-meßbar, so gilt $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$ f.s.
iii) Unabhängigkeit: Sind $\sigma(X)$ und $\mathcal{G}$ unabhängig, so gilt $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ f.s.

Satz 50 (Bedingte Jensen'sche Ungleichung) Sei $\phi : I \to \mathbb{R}$ eine konvexe Funktion auf einem Intervall $I \subseteq \mathbb{R}$ und $X : \Omega \to I$ eine integrierbare Zufallsvariable mit $\phi \circ X \in \mathcal{L}^1(\mu)$. Ist $\mathcal{G} \subseteq \mathcal{A}$ eine σ-Algebra, so gilt fast sicher

$$\mathbb{E}(X|\mathcal{G}) \in I, \qquad \phi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\phi(X)|\mathcal{G}).$$

5.2 Gleichgradige Integrierbarkeit

Definition 49 (Gleichgradig integrierbar) Eine Familie von Zufallsvariablen $(X_i)_{i\in I}$ in $\mathcal{L}^1(\mu)$ heißt gleichgradig integrierbar, wenn $\sup_{i\in I} \mathbb{E}(|X_i|\chi_{\{|X_i|\geq c\}}) \to 0$ für $c \to \infty$.

Proposition 18 Sei $X \in \mathcal{L}^1(\mu)$ und $(\mathcal{A}_i)_{i\in I}$ eine Familie von Sub-σ-Algebren in $\mathcal{A}$. Dann definiert $X_i = \mathbb{E}(X|\mathcal{A}_i)$ eine gleichgradig integrierbare Familie.

Satz 51 (Äquivalenz zur L^1-Beschränkheit) Für eine Familie $X = (X_i)_{i\in I}$ in $L^1(\mu)$ ist äquivalent:

i) X ist gleichgradig integrierbar.
ii) X ist in $L^1(\mu)$ beschränkt und für alle $\varepsilon > 0$ existiert ein $\delta > 0$, so daß für alle $A \in \mathcal{A}$ mit $\mu(A) < \delta$ gilt $\sup_{i\in I} \mathbb{E}(|X_i|\chi_A) < \varepsilon$.

Satz 52 (Äquivalenz zur L^1-Konvergenz) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $L^1(\mu)$ und X eine reelle Zufallsvariable. Dann sind äquivalent:

i) X ist gleichgradig integrierbar und $X_n \to X$ stochastisch.
ii) $X \in L^1(\mu)$ und $X_n \to X$ in $L^1(\mu)$.

Korollar 8 Gilt $X_n \to X$ fast sicher und ist $(X_n^p)_{n\in\mathbb{N}}$, $p > 1$, gleichgradig integrierbar, so folgt $X_n \to X$ in $L^p(\mu)$.

5.3 Martingale: Definition und Beispiele

Definition 50 (Filtration, Martingal) Es seien $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ eine aufsteigende Folge von σ-Algebren, $\mathcal{A}_n \subseteq \mathcal{A}_m \subseteq \mathcal{A}$ für $n \leq m$, eine sogenannte Filtration. Eine Folge $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{L}^1(\mu)$, für die X_n $\mathcal{A}_n$-meßbar ist und $\mathbb{E}(X_n|\mathcal{A}_{n-1}) = X_{n-1}$ $(\geq, \leq)$ für alle $n \in \mathbb{N}$ gilt, heißt Martingal (Submartinagal, Supermartingal).

Satz 53 (Elementare Eigenschaften) Sei (X_n) ein Martingal bzgl. der Filtration $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$. Dann gilt
i) Für alle $m \in \mathbb{N}$: $\mathbb{E}(X_{n+m}|\mathcal{A}_n) = X_n$.
ii) $\mathbb{E}(X_n) = \mathbb{E}(X_0)$ für alle $n \in \mathbb{N}_0$.

Der Beweis von i) erfolgt über die Projektionseigenschaft der bedingten Erwartung:

$$\mathbb{E}(X_{n+m}|\mathcal{A}_n) = \mathbb{E}(\mathbb{E}(X_{n+m}|\mathcal{A}_n)|\mathcal{A}_n) = \mathbb{E}(X_{n+m-1}|\mathcal{A}_n) = \ldots = \mathbb{E}(X_{n+m-m}|\mathcal{A}_n) = X_n.$$

Der Beweis von ii) nimmt zuerst aus i), dass $\mathbb{E}(X_n|\mathcal{A}_0) = X_0$ und daher $\mathbb{E}(\mathbb{E}(X_n|\mathcal{A}_0)) = \mathbb{E}(X_0)$. Wegen der Projektionseigenschaft wiederum gilt

$$\mathbb{E}(\mathbb{E}(X_n|\mathcal{A}_0)) = \mathbb{E}(\mathbb{E}(X_n|\mathcal{A}_0)|\{\emptyset, \Omega\}) = \mathbb{E}(X_n|\{\emptyset, \Omega\}) = \mathbb{E}(X_n),$$

und somit $\mathbb{E}(X_n) = \mathbb{E}(X_0)$.

Beispiel 11 (Summe unabhängiger Zufallsvariablen) Sei $\{X_i\}$ eine Folge unabhängiger Zufallsvariablen, $X_0 = 0$ und $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Dann definiert $M_n = \sum_{i=1}^n (X_i - \mathbb{E}(X_i))$, $n \in \mathbb{N}_0$ ein Martingal, denn wegen iii) aus Satz 49:

$$\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = \mathbb{E}(Y_{n+1} - \mathbb{E}(Y_{n+1})|\mathcal{F}_n) = \mathbb{E}(Y_{n+1} - \mathbb{E}(Y_{n+1}) = 0.$$

Beispiel 12 (Faire und unfaire Spiele) Sei M_n der stochastische Prozess, der die Entwicklung der Gewinnsumme eines Spielers nach n Spielrunden modelliert. Dann ist $M_n - M_{n+1}$ der Gewinn in der $n$ten Runde. Sei zudem $\mathcal{F}_n$ die durch die Spielrunden erzeugte Filtration. Wenn $M = (M_n)$ ein Martingal ist, dann $\mathbb{E}(M_n - M_{n-1}|\mathcal{F}_n) = 0$ und das Spiel ist fair. Wenn M ein Supermartingal ist, so gilt $\mathbb{E}(M_n - M_{n-1}|\mathcal{F}_n) \leq 0$ und das Spiel ist zum Nachteil des Spielenden.

Beispiel 13 (Produkte unabhängiger Zufallsvariablen) Sei $\{X_i\}$ eine Folge unabhängiger Zufallsvariablen, $X_i \geq 0$ und $\mathbb{E}(X_i) = 1$ für alle i. Wir setzen $M_0 = 1$ und $\mathcal{F}_0 = \{\emptyset, \Omega\}$ und $M_n = X_1 \cdot \ldots \cdot X_n$ und $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Dann ist M_n $\mathcal{F}_{n-1}$-meßbar und X_n und $\mathcal{F}_{n-1}$ ist unabhängig. Daher

$$\mathbb{E}(M_n|\mathcal{F}_{n-1}) = \mathbb{E}(M_{n-1}X_n|\mathcal{F}_{n-1}) = M_{n-1}\mathbb{E}(X_n|\mathcal{F}_{n-1}) = M_{n-1}\mathbb{E}(X_n) = M_{n-1},$$

und daher ist M_n ein Martingal.

Satz 54 (Filtration definiert Martingal) Es seien $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ eine Filtration bzgl. $\mathcal{A}$. $X \in L^1(\mu)$ sei eine Zufallsvariable. Dann definiert $M_n = \mathbb{E}(X|\mathcal{A}_n)$ ein Martingal.

Der Beweis ist einfach: Da X integrierbar ist, ist es auch M_n. M_n ist wegen der Definition der bedingten Erwartung auch $\mathcal{A}_n$-meßbar und es gilt wegen der Projektionseigenschaft der bedingten Erwartung:

$$\mathbb{E}(M_{n+1}|\mathcal{A}_n) = \mathbb{E}(\mathbb{E}(X|\mathcal{A}_{n+1})|\mathcal{A}_n) = \mathbb{E}(X|\mathcal{A}_n) = M_n.$$

Bemerkung 20 (Spielsystem allgemein) Sei (M_n) der Gewinnprozess bei einem Spiel, so dass $M_n - M_{n-1}$ den Gewinn *pro Euro Einsatz* in der n-ten Spielrunde beschreibt; wir setzen $M_0 = 0$. Sei $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ die zugehörige Filtration, so dass M_n $\mathcal{A}_n$-meßbar ist. Ausserdem sei H_n der stochastische Prozeß der den Einsatz in der n-ten Runde beschreibt. Dieser ist nur dann *erlaubt*, wenn er nur von der Information $\mathcal{A}_1, \ldots, \mathcal{A}_{n-1}$ abhängen, muss also $\mathcal{A}_{n-1}$-meßbar sein. Dann ist $H_n(M_n - M_{n-1})$ der Gewinn in der n-ten Runde und $(H.M)_n = \sum_{k=1}^n H_k(M_k - M_{k-1})$ der Gesamtgewinn nach der n-ten Runde.

Satz 55 (Martingaltransformierte) Sei $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ eine Filtration, $(Y_n)_{n\in\mathbb{N}}$ eine Folge $\mathcal{A}_{n-1}$-$\mathcal{B}(\mathbb{R})$-meßbarer Zufallsvariablen und $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal, so daß $Y_n(X_n - X_{n-1}) \in \mathcal{L}^1(\mu)$. Dann definiert die Martingaltransformierte

$$(Y.X)_n = X_0 + \sum_{k=1}^n Y_k(X_k - X_{k-1})$$

ein Martingal.

Der Beweis ist einfach: $(Y.X)_n$ ist integrierbar und $\mathcal{A}_n$-meßbar. Da Y_n $\mathcal{A}_{n-1}$-meßbar ist, gilt nach ii) aus Satz 49 und wegen der Martingal-Eigenschaft von X_n , dass

$$\mathbb{E}((Y.X)_n - (Y.X)_{n-1}|\mathcal{A}_{n-1}) \quad = \quad \mathbb{E}(Y_n(X_n - X_{n-1})|\mathcal{A}_{n-1}) = Y_n\mathbb{E}(X_n - X_{n-1})|\mathcal{A}_{n-1}) = 0.$$

Bemerkung 21 (Spielsystem bei fairen Spielen) Wenn wir die Bezeichnungen wie in Bemerkung 20 wählen, dann handelt es sich um ein faires Spiel, wenn M_n ein Martingal ist, siehe Beispiel 12. Dann ist der Gesamtgewinn $(H.M)_n$ laut Satz ebenfalls ein Martingal und somit $\mathbb{E}((H.M)_n) = \mathbb{E}(M_0) = 0$, was zeigt, dass bei einem fairen Spiel keine erlaubte Spielstrategie (H_n) existiert, die einen erwarteten Gewinn garantiert.

5.4 Doob'scher Stoppsatz

Definition 51 (Gestoppter Prozess) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Folge von Zufallsvariablen, $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ eine Filtration, so daß X_n $\mathcal{A}_n$-meßbar ist. Sei $\tau : \Omega \to \mathbb{N}_0 \cup \{+\infty\}$ eine Stoppzeit, das heißt $\{\tau = n\} \in \mathcal{A}_n$ für jedes $n \in \mathbb{N}_0$. Dann ist der gestoppte Prozess X^τ durch $X_n^\tau = X_{\min(\tau,n)}$ definiert.

Satz 56 (Doob'scher Stoppsatz) Seien X ein Martingal und τ eine Stoppzeit bezüglich $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$. Dann ist X^τ ebenfalls ein Martingal und es gilt $\mathbb{E}(X_n^\tau) = \mathbb{E}(X_0)$ für alle $n \in \mathbb{N}_0$. Ist τ fast sicher beschränkt oder X^τ gleichgradig integrierbar und $\mu(\tau < \infty) = 1$, so gilt $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$.

Beweis dafür, dass X^τ ein Martingal ist: Wir definieren $Y_n = \chi_{\tau \geq n} = 1 - \chi_{\tau < n}$. Y_n ist $\mathcal{A}_{n-1}$-meßbar, da τ eine Stoppzeit ist und es gilt $0 \leq Y_n \leq 1$. Daher sind die Bedingungen von Satz 21 erfüllt und $Y.X$ ist ein Martingal und es gilt

$$(Y.X)_n = X_0 + \sum_{k=1}^{n} \chi_{\tau \geq k}(X_k - X_{k-1}) = X_{\min(\tau,n)} = X_n^\tau,$$

also $Y.X = X^\tau$ und daher laut ii) aus Satz 53: $\mathbb{E}(X_n^\tau) = \mathbb{E}(X_0^\tau) = \mathbb{E}(X_0)$ für alle $n \in \mathbb{N}_0$.

Beweis für $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$: Ist τ fast sicher beschränkt oder $\tau < \infty$ fast sicher, dann $\min(\tau, n) \to \tau$ fast sicher für $n \to \infty$ und daher $X_n^\tau = X_{\min(\tau,n)} \to X_\tau$. Ist nun $\tau \leq N$ fast sicher, so ist $\max_{i=1,\ldots,N} |X_i|$ eine integrierbare Majorante zu X_n^τ und aus $X_n^\tau \to X_\tau$ folgt $\mathbb{E}(X_n^\tau) = \mathbb{E}(X_0) = \mathbb{E}(X_\tau)$ aus dem Satz der dominierten Konvergenz. Ist X_n^τ gleichgradig integrierbar, so gilt sogar $X_n^\tau \to X_\tau$ in $L^1(\mu)$ und dann ebenfalls $\mathbb{E}(X_n^\tau) = \mathbb{E}(X_0) = \mathbb{E}(X_\tau)$.

Beispiel 14 (Ruinwahrscheinlichkeit) Betrachtet wird ein faires Spiel mit zwei Spielausgängen. Der Spielausgang in Runde i sei beschrieben durch die Zufallsvariable $Y_i \in \{\pm 1\}$ mit $\mu(Y_i = \pm 1) = 1/2$. Ein Spieler mit Kapital a Euro spielt gegen die Bank mit Kapital b Euro. In jeder Runde setzen beide den Betrag 1 Euro; bei $Y_i = +1$ erhält der Spieler den Einsatz, bei $Y_i = -1$ die Bank. Wir setzen $X_n = \sum_{i=1}^n Y_i$ und $X_0 = 0$. Da die Y_i unabhängig sind mit $\mathbb{E}(Y_i) = 0$ ist X_n laut Beispiel 11 ein Martingal. X_n beschreibt die Bilanz des Spiels, in dem Sinne, dass $X_n = -c$ mit $c > 0$ heisst, dass der Spieler nach n Runden c Euro verloren hat, während $X_n = c$ heisst, dass die Bank c Euro verloren hat. Das Spiel endet mit dem Ruin eines der beiden Spielenden, also an der Stoppzeit $\tau = \inf\{n \geq 0 : X_n \in (-a, b)\}$. Dann ist der gestoppte Prozess X^τ ebenfalls ein Martingal. Man weist leicht nach, dass τ fast sicher endlich ist und X^τ beschränkt, da $|X_n^\tau| \leq \max\{a, b\}$ und daher gleichgradig stetig. Daher gilt $\mathbb{E}(X_\tau) = \mathbb{E}(X_0) = 0$. Da $X_\tau \in \{-a, b\}$, laut Definition der Stoppzeit, haben wir $\mathbb{E}(X_\tau) = -a\mu(X_\tau = -a) + b\mu(X_\tau = b)$. Mit der Ruinwahrscheinlichkeit für den Spieler, $p_s = \mu(X_\tau = -a)$, folgt $(a + b)p_s = b$ und wir erhalten die Ruinwahrscheinlichkeit $p_s = b/(a + b)$.

5.5 Martingalkonvergenz

Lemma 19 Es seien $X = (X_n)_{n\in\mathbb{N}_0}$ ein Submartingal und $g : \mathbb{R} \to \mathbb{R}$ konvex mit $g(X_n) \in \mathcal{L}^1(\mu)$ für alle $n \in \mathbb{N}_0$. Ist g monoton wachsend oder X ein Martingal, so ist $g(X)$ ein Submartingal.

Lemma 20 (Überquerungen) Sei X ein Submartingal, $a < b \in \mathbb{R}$ und $N \in \mathbb{N}_0$. Man definiert $S_0 = T_0 = 0$,

$$S_k = \inf\{n \geq T_{k-1} : X_n \leq a\}, \qquad T_k = \inf\{n \geq S_k : X_n \geq b\}$$

und $U_N = \sup\{k \in \mathbb{N}_0 : T_k \leq N\}$. Dann gilt $\mathbb{E}(U_N) \leq \mathbb{E}((X_N - a)_+)/(b - a)$.

Satz 57 (L^1-Konvergenz) Sei X ein Submartingal (Supermartingal) mit

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}(X_n^+) < \infty \qquad (\sup_{n\in\mathbb{N}_0} \mathbb{E}(X_n^-) < \infty).$$

Dann existiert ein $X_\infty \in L^1(\mu)$ mit $X_\infty = \lim_{n\to\infty} X_n$ fast sicher. Ist X gleichgradig integrierbar, so folgt sogar $X_n \to X_\infty$ in $L^1(\mu)$.

5.6 Optional Sampling Theorem

Korollar 9 (Martingalkonvergenz) Für ein Martingal X sind äquivalent:

i) X ist gleichgradig integrierbar.
ii) Es gibt ein $\mathcal{A}_\infty$-meßbares $X_\infty \in \mathcal{L}^1(\mu)$ mit $X_n = \mathbb{E}(X_\infty|\mathcal{A}_n)$ für alle $n \in \mathbb{N}_0$, wobei $\mathcal{A}_\infty = \sigma(\bigcup_{n\in\mathbb{N}_0} \mathcal{A}_n)$.
iii) X konvergiert in $L^1(\mu)$ gegen ein $\mathcal{A}_\infty$-meßbares X_∞.

Satz 58 (Optional Sampling) Es seien X ein Martingal und σ, τ Stoppzeiten mit $\sigma \leq \tau$. Ist τ fast sicher beschränkt oder X gleichgradig integrierbar, so sind X_σ, X_τ wohldefinierte, integrierbare Zufallsvariablen, und es gilt

$$\mathbb{E}(X_\tau|A_\sigma) = X_\sigma,$$

wobei $A_\sigma = \{A \in \mathcal{A} : A \cap \{\tau \leq n\} \in \mathcal{A}_n \text{ für alle } n \in \mathbb{N}_0\}$.

Literatur

[C] K. Chung: Markov chains with stationary transition probabilities, Springer-Verlag, 1960.

[Dei] O. Deiser: Reelle Zahlen, Springer-Verlag, 2007.

[Dud] R. Dudley: Real analysis and probability, Wadsworth, 1989.

[Dur] R. Durrett: Probability, theory and examples, Wadsworth, 1991.

[E] J. Elstrodt: Maß- und Integrationstheorie, 4. korrigierte Auflage, Springer-Verlag, 2005.

[G] H.-O. Georgii: Stochastik, de Gruyter-Verlag, 2007.

[K] A. Klenke: Wahrscheinlichkeitstheorie, Springer-Verlag, 2008.

[MS] D. Meintrup, S. Schäffler: Stochastik, Springer-Verlag, 2005.

[RC] C. Robert, G. Casella: Monte Carlo Statistical Methods, Springer-Verlag, 2004.