# Stochastische Modelle

Peter Eichelsbacher

Sommersemester 2004

# Inhaltsverzeichnis

KAPITEL 1

# Bedingte Erwartungswerte

Bisher haben wir ausführlich Eigenschaften unabhängiger Zufallsgrößen untersucht. Wir wollen nun beginnen, stochastische Abhängigkeiten in einem stochastischen Modell mathematisch präzise zu beschreiben. Der Begriff des bedingten Erwartungswerts ist zentral hierzu.

Gegeben sei fortan ein W-Raum $(\Omega, \mathcal{A}, P)$. Zu $A, B \in \mathcal{A}$ mit $P(B) > 0$ heißt

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

(elementare) *bedingte Wahrscheinlichkeit* von $A$ unter (der Hypothese) $B$. Die Mengenfunktion $P(\cdot|B) : \mathcal{A} \to [0,1]$ bildet ein W-Maß auf $(\Omega, \mathcal{A})$. Seien $X, Y$ zwei Zufallsgrößen auf $(\mathbb{R}, \mathcal{B}, P)$, $Y$ sei diskret. Dann können wir

$$P(X \in B|Y = y) = P(\{X \in B\}|\{Y = y\}) = \frac{P(X \in B, Y = y)}{P(Y = y)}$$

für alle $B \in \mathcal{B}$ und $y$ mit $P(Y = y) > 0$ definieren. Für jedes $y$ ist $P(X \in \cdot|Y = y)$ ein W-Maß auf $(\mathbb{R}, \mathcal{B})$, welches häufig *bedingte Verteilung* von $X$, gegeben $Y = y$, genannt wird, in Zeichen $P^{X|Y=y}$. Der zugehörige Erwartungswert

$$\mathbb{E}(X|Y = y) := \int x \, P^{X|Y=y}(dx) = \frac{1}{P(Y = y)} \int_{\{Y=y\}} X \, dP \, , \tag{1.1}$$

sofern dieser existiert, heißt *bedingter Erwartungswert* von $X$ gegeben $Y = y$. Die zweite Gleichheit folgt mittels des Funktions-Erweiterungsarguments[1]. Bei diskretem $Y$ sind bedingte Wahrscheinlichkeit und bedingter Erwartungswert gegeben $Y = y$ für $P^Y$-fast alle $y$ definiert. Was ist aber zu tun, wenn $Y$ keine diskrete Verteilung besitzt? Nun kann sogar $P(Y = y) = 0$ für alle $y \in \mathbb{R}$ gelten. Wir geben dazu, also auf der Suche nach einer allgemeinen Definition, eine *andere* Interpretation bedingter Wahrscheinlichkeit bzw. bedingter Erwartungswerte.

Sei zunächst wieder $Y$ diskret und $|X| \leq c$. Dann folgt $|\mathbb{E}(X|Y = y)| \leq c$. Also bildet

$$f(y) := \begin{cases} \mathbb{E}(X|Y = y), & \text{falls } P(Y = y) > 0, \\ 0, & \text{sonst} \end{cases}$$

eine durch $c$ beschränkte, messbare Funktion. Folglich ist $f(Y) =: \mathbb{E}(X|Y)$ eine beschränkte $\sigma(Y)$-messbare Zufallsgröße. Sei $X$ eine Variable, deren Wert

[1]siehe Wahrscheinlichkeitstheorie, Kapitel 1

wir nicht beobachten können. Stattdessen erhalten wir den Wert von $Y$ und suchen nach einer Approximation $g(Y)$, eine $\sigma(Y)$-messbare Variable, die möglichst gut $X$ approximiert (wobei wir mittels einer Metrik klären müssen, was hier möglichst gut bedeuten soll). Bei diskretem $Y$ soll das Ergebnis $f(Y) = \mathbb{E}(X|Y)$ sein. Gibt es eine Metrik, bezüglich der $f(Y)$ die beste Approximation liefert?

Wir beobachten zunächst: Für jedes $A = \{Y \in B\}$ gilt

$$\begin{aligned}\int_A \mathbb{E}(X|Y)\,dP &= \int_B \mathbb{E}(X|Y=y)\,P^Y(dy) = \sum_{y\in B} \mathbb{E}(X|Y=y)P(Y=y)\\ &= \sum_{y\in B}\Big(\frac{1}{P(Y=y)}\int_{\{Y=y\}} X\,dP\Big)P(Y=y) = \int_{\{Y\in B\}} X\,dP.\end{aligned}$$

Hierbei haben wir die Transformationsformel[2], sowie (1.1) verwendet. Es gilt also:

(a) $\mathbb{E}(X|Y)$ ist $\sigma(Y)$-messbar

(b) $\int_A \mathbb{E}(X|Y)\,dP = \int_A X\,dP$ für alle $A \in \sigma(Y)$.

Durch (a) und (b) ist $\mathbb{E}(X|Y)$ fast sicher eindeutig bestimmt, denn sei $Z$ eine $\sigma(Y)$-messbare Zufallsgröße mit $\int_A Z\,dP = \int_A X\,dP$ für alle $A \in \sigma(Y)$, so gilt

$$\int_A (\mathbb{E}(X|Y) - Z)\,dP = 0 \;\text{ für alle } A \in \sigma(Y)\;.$$

Da $\mathbb{E}(X|Y) - Z$ $\sigma(Y)$-messbar, folgt $\mathbb{E}(X|Y) - Z = 0$ fast sicher, siehe 32.3(i), Analysis III.

(b) können wir auch so schreiben:

$$\text{(b')}\; \mathbb{E}(1_A\mathbb{E}(X|Y)) = \mathbb{E}(1_A X) \text{ für alle } A \in \sigma(Y)\;.$$

Nun sind -im Moment nach Voraussetzung- $\mathbb{E}(X|Y)$ und $X$ beschränkt, also existiert $\mathbb{E}\big(Z\,\mathbb{E}(X|Y)\big)$ und $\mathbb{E}(ZX)$ für beliebige integrierbare Zufallsgrößen $Z$. Mit Hilfe des Funktions-Erweiterungsarguments[3] liefert (b') dann

$$\text{(c)}\; \mathbb{E}\big(Z\,\mathbb{E}(X|Y)\big) = \mathbb{E}(ZX), \text{ also } \mathbb{E}\Big(Z\big(X - \mathbb{E}(X|Y)\big)\Big) = 0$$

für alle $\sigma(Y)$-messbaren und integrierbaren Zufallsgrößen $Z$.

Ist $Z$ quadrat-integrierbar und $\sigma(Y)$-messbar, folgt

$$\begin{aligned}\mathbb{E}(X-Z)^2 &= \mathbb{E}\big(X - \mathbb{E}(X|Y)\big)^2 + \mathbb{E}\big(\mathbb{E}(X|Y) - Z\big)^2\\ &\quad + 2\mathbb{E}\Big(\big(\mathbb{E}(X|Y) - Z\big)\big(X - \mathbb{E}(X|Y)\big)\Big)\\ &= \mathbb{E}\big(X - \mathbb{E}(X|Y)\big)^2 + \mathbb{E}\big(\mathbb{E}(X|Y) - Z\big)^2 \geq \mathbb{E}\big(X - \mathbb{E}(X|Y)\big)^2\;,\end{aligned}$$

[2]Wahrscheinlichkeitstheorie, Satz 1.22

[3]Wahrscheinlichkeitstheorie, Kapitel 1

denn $\mathbb{E}(X|Y) - Z$ ist $\sigma(Y)$-messbar. Also minimiert $\mathbb{E}(X|Y)$ den $\mathcal{L}^2$-Pseudo-Abstand zu $X$ unter allen $\sigma(Y)$-messbaren $Z$!

Wegen der Vollständigkeit von $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ bildet dieser Raum mit $\langle X, Y\rangle := \mathbb{E}(XY)$ einen Hilbertraum (siehe Satz 33.12, Analysis III), womit orthogonale Projektionen existieren, die bei der Suche nach besten Approximationen bzgl. der Norm die natürlichen Kandidaten darstellen.

Es sei nun $\mathcal{F} \subset \mathcal{A}$ eine $\sigma$-Algebra und $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Obiges (a) und (b) übertragen wir nun zu

$$\begin{array}{ll} \text{(I)} & \text{Die Zufallsgröße } \mathbb{E}(X|\mathcal{F}) \text{ ist } \mathcal{F}\text{-messbar} \\ \text{(II)} & \displaystyle\int_A \mathbb{E}(X|\mathcal{F})\, dP = \int_A X\, dP \text{ für alle } A \in \mathcal{F}\ . \end{array}$$

Es gilt $\mathcal{L}^2(\Omega, \mathcal{F}, P) \subset \mathcal{L}^2(\Omega, \mathcal{A}, P)$.

Die *orthogonale Projektion* $\mathbb{P}_{\mathcal{F}}X$ von $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ ist durch

$$||X - \mathbb{P}_{\mathcal{F}}X||_2 = \min_{Z \in \mathcal{L}^2(\Omega,\mathcal{F},P)} ||X - Z||_2$$

fast sicher eindeutig bestimmt und auch durch

$$\langle X - \mathbb{P}_{\mathcal{F}}X, Z\rangle = \mathbb{E}\big((X - \mathbb{P}_{\mathcal{F}}X)Z\big) = 0 \tag{1.2}$$

für alle $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ charakterisiert. Man nennt $\mathbb{P}_{\mathcal{F}}X$ *beste Approximation* von $X$ in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$.

**1.1 Satz und Definition** Seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $\mathcal{F}$ eine Unter-$\sigma$-Algebra von $\mathcal{A}$ und $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Bezeichnet dann $\mathbb{E}(X|\mathcal{F})$ die $P$-f.s. eindeutige beste Approximation von $X$ in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, so erfüllt diese die Bedingung (I) und (II). $\mathbb{E}(X|\mathcal{F})$ heißt *bedingter Erwartungswert* oder *bedingte Erwartung* von $X$ gegeben $\mathcal{F}$.

*Beweis:* Mit (1.2) gilt für alle $A \in \mathcal{F}$ (womit $1_A \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$)

$$\int_A \big(X - \mathbb{E}(X|\mathcal{F})\big)\, dP = \langle X - \mathbb{E}(X|\mathcal{F}), 1_A\rangle = 0. \qquad \square$$

Der bedingte Erwartungswert ist außerhalb von $P$-Nullmengen $\in \mathcal{F}$ eindeutig bestimmt. Man bezeichnet jede zulässige Wahl auch als *Version* von $\mathbb{E}(X|\mathcal{F})$. Wir setzen $\mathbb{E}(X|Y) := \mathbb{E}\big(X|\sigma(Y)\big)$.

**1.2 Lemma** *Seien $(\Omega, \mathcal{A}, P)$ ein W-Raum und $\mathcal{F}$ eine Unter-$\sigma$-Algebra von $\mathcal{A}$ und $X, Y, X_1, X_2, \ldots \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Dann gilt*

(a) $\mathbb{E}(\alpha X + \beta Y|\mathcal{F}) = \alpha\mathbb{E}(X|\mathcal{F}) + \beta\mathbb{E}(Y|\mathcal{F})$ *P-f.s. für alle* $\alpha, \beta \in \mathbb{R}$.

(b) $X \leq Y$ *P-f.s.* $\Rightarrow \mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F})$ *P-f.s.*

(c) $X_n \nearrow X$ *P-f.s.* $\Rightarrow \mathbb{E}(X_n|\mathcal{F}) \nearrow \mathbb{E}(X|\mathcal{F})$ *P-f.s.*

*Beweis:*

(a) Es gilt

$$\begin{aligned} 0 &= \alpha \langle X - \mathbb{E}(X|\mathcal{F}), Z\rangle + \beta \langle Y - \mathbb{E}(Y|\mathcal{F}), Z\rangle \\ &= \big\langle (\alpha X + \beta Y) - \big(\alpha\mathbb{E}(X|\mathcal{F}) + \beta\mathbb{E}(Y|\mathcal{F})\big), Z\big\rangle \end{aligned}$$

für alle $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Die $P$-f.s. Eindeutigkeit des bedingten Erwartungswertes liefert die Behauptung.

(b) $A := \{\mathbb{E}(X|\mathcal{F}) > \mathbb{E}(Y|\mathcal{F})\} \in \mathcal{F}$. Zu zeigen: $P(A) = 0$. Mit (II) gilt nach Definition von $A$

$$0 \leq \int_A \big(\mathbb{E}(X|\mathcal{F}) - \mathbb{E}(Y|\mathcal{F})\big)\, dP = \int_A (X - Y)\, dP \leq 0\ ,$$

da $X \leq Y$ $P$-f.s. Damit folgt wegen der strikten Positivität von $\mathbb{E}(X|\mathcal{F}) - \mathbb{E}(Y|\mathcal{F})$ auf $A$, dass $P(A) = 0$.

(c) Nach (b) existiert ein $\mathcal{F}$-messbares $W$ mit $\mathbb{E}(X_n|\mathcal{F}) \nearrow W$ $P$-f.s. Da $\mathbb{E}(X_1|\mathcal{F}) \leq W \leq \mathbb{E}(X|\mathcal{F})$ $P$-f.s., folgt $W \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Es gilt

$$\begin{aligned} \langle X - W, Z\rangle &= \langle X - X_n, Z\rangle + \langle X_n - \mathbb{E}(X_n|\mathcal{F}), Z\rangle + \langle \mathbb{E}(X_n|\mathcal{F}) - W, Z\rangle \\ &= \langle X - X_n, Z\rangle + \langle \mathbb{E}(X_n|\mathcal{F}) - W, Z\rangle \\ &\leq \big(||X - X_n||_2 + ||W - \mathbb{E}(X_n|\mathcal{F})||_2\big)||Z||_2 \end{aligned}$$

nach Cauchy-Schwarz. Da $0 \leq X - X_n \leq X - X_1$ und $0 \leq W - \mathbb{E}(X_n|\mathcal{F}) \leq W$ $P$-f.s., liefert der Satz von der dominierten Konvergenz, dass der letzte Ausdruck für $n \to \infty$ gegen 0 strebt. Wegen der $P$-f.s. Eindeutigkeit des bedingten Erwartungswertes muss $W = \mathbb{E}(X|\mathcal{F})$ $P$-f.s. gelten. □

Wir wollen den Begriff des bedingten Erwartungswertes auf quasi-integrierbare Zufallsvariablen ausdehnen (siehe 31.16, Analysis III). Sei zunächst $X \geq 0$ $P$-f.s. und $X_n := X1_{\{X\leq n\}}$, $n \geq 1$. Dann ist $X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ und mit Lemma 1.2 (b) existiert zu jeder Unter-$\sigma$-Algebra $\mathcal{F} \subset \mathcal{A}$ eine $\mathcal{F}$-messbare, nicht-negative Zufallsgröße $W$ mit $\mathbb{E}(X_n|\mathcal{F}) \nearrow W$ $P$-f.s. Der Satz von der monotonen Konvergenz und (II) liefern

$$\int_A X\, dP = \lim_{n\to\infty} \int_A X_n\, dP = \lim_{n\to\infty} \int_A \mathbb{E}(X_n|\mathcal{F})\, dP = \int_A W\, dP$$

für alle $A \in \mathcal{F}$. $W$ erfüllt somit (I) und (II) und wir setzen $\mathbb{E}(X|\mathcal{F}) := W$. Diese Definition stimmt für quadratisch integrierbare $X$ mit der vorher gegebenen Definition gemäß (c) in Lemma 1.2 überein. $\mathbb{E}(X|\mathcal{F})$ ist durch (I) und (II) $P$-f.s. eindeutig bestimmt (einfach)! Mit $A = \Omega$ in (II) ist

$$X \text{ integrierbar } \Leftrightarrow \mathbb{E}(X|\mathcal{F}) \text{ integrierbar.}$$

Damit folgt für quasi-integrierbares $X$ vermöge $X = X^+ - X^-$:

$$\mathbb{E}(X|\mathcal{F}) := \mathbb{E}(X^+|\mathcal{F}) - \mathbb{E}(X^-|\mathcal{F}) \ .$$

**1.3 Satz** *Es seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $\mathcal{F}$ eine Unter-$\sigma$-Algebra von $\mathcal{A}$ und $X, Y, X_1, X_2, \ldots$ nicht-negative bzw. integrierbare Zufallsgrößen. Dann gilt:*

(a) $\mathbb{E}(\alpha X + \beta Y|\mathcal{F}) = \alpha\mathbb{E}(X|\mathcal{F}) + \beta\mathbb{E}(Y|\mathcal{F})$ *$P$-f.s. für alle $\alpha, \beta \geq 0$ bzw. $\in \mathbb{R}$.*

(b) $X \leq Y$ *$P$-f.s.* $\Rightarrow \mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F})$ *$P$-f.s.*

(c) $\mathbb{E}\big(\mathbb{E}(X|\mathcal{F})\big) = \mathbb{E}X$ *und* $|\mathbb{E}(X|\mathcal{F})| \leq \mathbb{E}(|X|\,|\mathcal{F})$ *$P$-f.s.*

(d) $X_n \nearrow X$ *$P$-f.s.* $\Rightarrow \mathbb{E}(X_n|\mathcal{F}) \nearrow \mathbb{E}(X|\mathcal{F})$ *$P$-f.s.*

(e) $X_n \to X$ *und* $\sup_{n\geq 1} |X_n| \leq Y$ *$P$-f.s.*
$\Rightarrow \quad \mathbb{E}(X_n|\mathcal{F}) \to \mathbb{E}(X|\mathcal{F})$ *$P$-f.s.*

*Beweis:* Lemma 1.2 und Integrationstheorie □

Wir werden die folgenden Bezeichnungen verwenden:

Falls $X = 1_A$, $A \in \mathcal{A}$, so schreiben wir

$$P(A|\mathcal{F}) := \mathbb{E}(1_A|\mathcal{F}) \ ,$$

und nennen $P(A|\mathcal{F})$ die *bedingte Wahrscheinlichkeit* von $A$ gegeben $\mathcal{F}$.

Ist $\mathcal{F} = \sigma(A) = \{\Omega, A, A^c, \emptyset\}$, so schreibt man auch $\mathbb{E}(X|A)$ für $\mathbb{E}(X|\mathcal{F})$ und analog für $\mathcal{F} = \sigma((A_i)_{i\in I})$ auch $\mathbb{E}(X|(A_i)_{i\in I})$ und $\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$, sowie $\mathbb{E}(X|(Y_i)_{i\in I})$ im Fall $Y = (Y_i)_{i\in I}$.

**1.4 Beispiele**

(a) Es sei $\mathcal{F} = \{\emptyset, \Omega\}$. Aus (I) folgt, dass $\mathbb{E}(X|\mathcal{F})$ konstant ist und mit (II) folgt $\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}X$. Ist $\mathcal{F} \supset \sigma(X)$, so folgt $\mathbb{E}(X|\mathcal{F}) = X$ $P$-f.s., denn $X$ ist $\mathcal{F}$-messbar und es gilt (II).

(b) Ist $\mathcal{F} = \sigma(A) = \{\Omega, A, A^c, \emptyset\}$ mit $P(A) \in (0, 1)$, so muss $\mathbb{E}(X|\mathcal{F})$ wegen der $\mathcal{F}$-Messbarkeit auf $A$ und $A^c$ jeweils konstant sein. Mit (II) folgt

$$\mathbb{E}(X|A) = \Big(P(A)^{-1} \int_A X \, dP\Big) 1_A + \Big(P(A^c)^{-1} \int_{A^c} X \, dP\Big) 1_{A^c} \ .$$

Hier ist $P$-f.s. überflüssig, da $\emptyset$ die einzige $P$-Nullmenge in $\mathcal{F}$ ist. Analog: Ist $\Omega = \dot{\bigcup}_{j=1}^{n} A_j$, $A_j \in \mathcal{A}$, $j = 1, \ldots, n$, so gilt

$$\mathbb{E}(X|A_1, \ldots, A_n) = \sum_{j=1}^{n} \Big( P(A_j)^{-1} \int_{A_j} X\, dP \Big) 1_{A_j} \quad P\text{-}f.s.$$

Hierbei setzen wir $\infty \cdot 0 := 0$, sofern $P(A_j) = 0$ für ein $j$. Sind alle $P(A_j)$ positiv, kann auf den Zusatz $P$-f.s. verzichtet werden.

**1.5 Satz** *Eine $\mathcal{F}$-messbare und integrierbare (nicht-negative) Zufallsgröße $Y$ ist genau dann Version des bedingten Erwartungswertes $\mathbb{E}(X|\mathcal{F})$ einer integrierbaren (nicht-negativen) Zufallsgröße $X$, wenn*

$$\mathbb{E}(XZ) = \mathbb{E}(YZ)$$

*für alle beschränkten (nicht-negativen) und $\mathcal{F}$-messbaren Zufallsgrößen $Z$ gilt. Im Fall $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, $p \in [1, \infty]$, folgt letztere Behauptung sogar für alle $Z \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$ mit $\frac{1}{p} + \frac{1}{q} = 1$.*

*Beweis:* Aus $\mathbb{E}(XZ) = \mathbb{E}(YZ)$ für alle $Z$ wie angegeben folgt sofort (II), also auch $Y = \mathbb{E}(X|\mathcal{F})$ $P$-f.s., denn $Y$ ist nach Voraussetzung $\mathcal{F}$-messbar. Ist umgekehrt $Y = \mathbb{E}(X|\mathcal{F})$ $P$-f.s., so folgt (II) und deshalb für nicht-negatives $X$ auch $\mathbb{E}(XZ) = \mathbb{E}(YZ)$ für alle $\mathcal{F}$-messbaren, nicht-negativen $Z$, indem man $Z$ durch eine monotone Folge aus $\mathcal{E}(\Omega, \mathcal{F})$ approximiert. Für integrierbares $X$ erhält man die Aussage durch Zerlegung in Positiv- und Negativteil. Details bleiben als Übung. $\square$

**1.6 Satz** *Seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2$ Unter-$\sigma$-Algebren von $\mathcal{A}$ und $X, Y$ Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$. Dann gilt:*

(a) *Sind $X$, $Y$ nicht-negativ oder $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, $Y \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$ für $1 \leq p, q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, so gilt:*

$$Y \ \mathcal{F}\text{-messbar} \quad \Rightarrow \quad \mathbb{E}(XY|\mathcal{F}) = Y\mathbb{E}(X|\mathcal{F}) \ P\text{-f.s.}$$

(b) *Ist $X$ quasi-integrierbar, so gilt:*

$$\mathcal{F}_1 \subset \mathcal{F}_2 \quad \Rightarrow \quad \mathbb{E}\Big(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1\Big) = \mathbb{E}\Big(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2\Big) = \mathbb{E}(X|\mathcal{F}_1) \ P\text{-f.s.}$$

(c) *Ist $\mathcal{F} = \sigma(\mathcal{F}_1 \cup \mathcal{F}_2)$, $X$ quasi-integrierbar und die von $X$ und $\mathcal{F}_1$ erzeugte $\sigma$-Algebra $\sigma(X, \mathcal{F}_1)$ unabhängig von $\mathcal{F}_2$, so gilt:*

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X|\mathcal{F}_1) \quad P\text{-f.s.}$$

(d) *Ist $X$ quasi-integrierbar, sowie unabhängig von $\mathcal{F}$, so gilt:*

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}X \quad P\text{-f.s.}$$

*Beweis:*

(a) Seien $X$, $Y$ nicht-negativ und $Y$ $\mathcal{F}$-messbar. Dann gilt mit Satz 1.5 für alle nicht-negativen $\mathcal{F}$-messbaren $Z$

$$\int \mathbb{E}(XY|\mathcal{F})Z\,dP = \int XYZ\,dP = \int X(YZ)\,dP = \int \mathbb{E}(X|\mathcal{F})YZ\,dP\ ,$$

denn mit $Z$ ist auch $YZ$ nicht-negativ und $\mathcal{F}$-messbar. Da $\mathbb{E}(X|\mathcal{F})Y$ $\mathcal{F}$-messbar, folgt die Behauptung aus Satz 1.5. Für $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ und $Y \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$ ist $XY$ integrierbar und $YZ \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$ für alle $\mathcal{F}$-messbaren und beschränkten $Z$. Der Beweis geht nun analog.

(b) Es gilt

$$\int_A \mathbb{E}\Big(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1\Big)\,dP = \int_A \mathbb{E}(X|\mathcal{F}_2)\,dP = \int_A X\,dP$$

für alle $A \in \mathcal{F}_1$ (und $P$-f.s. Eindeutigkeit des bedingten Erwartungswertes). Die zweite Gleichung folgt aus Beispiel 1.4 (a).

(c) Sei $Y$ eine Version von $\mathbb{E}(X|\mathcal{F}_1)$. Diese ist $\mathcal{F}_1$- und somit auch $\mathcal{F}$-messbar. Wir zeigen, dass $Y$ auch eine Version von $\mathbb{E}(X|\mathcal{F})$ ist. Dafür ist zu zeigen:

$$\int_A X\,dP = \int_A Y\,dP \text{ für alle } A \in \mathcal{F}\ . \tag{1.3}$$

Das System aller Mengen, welche diese Gleichheit erfüllen, ist ein Dynkin-System (Übung). Es genügt daher, (1.3) für alle $A$ aus einem durchschnittstabilen Erzeuger von $\mathcal{F}$ zu zeigen. Wähle $\mathcal{E} := \{A_1 \cap A_2 \ :\ A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$. Wir bekommen dann

$$\begin{aligned}\int_{A_1\cap A_2} Y\,dP = \mathbb{E}(1_{A_1}\,1_{A_2}\,Y) &= P(A_2)\,\mathbb{E}(1_{A_1}\,Y) \\ &= P(A_2)\mathbb{E}(1_{A_1}\,X) = \int_{A_1\cap A_2} X\,dP,\end{aligned}$$

denn sowohl $1_{A_2}$ und $1_{A_1}Y$ als auch $1_{A_2}$ und $1_{A_1}X$ sind unabhängig. Wir verwenden Satz 5.21 aus der Wahrscheinlichkeitstheorie.

(d) Wir wählen in (c) $\mathcal{F}_1 = \{\Omega, \emptyset\}$. □

**1.7 (Jensensche Ungleichung)** Seien $X$ eine integrierbare Zufallsgröße mit Werten in einem offenen Intervall $I \subset \mathbb{R}$ und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion.

Dann gilt $\mathbb{E}(X|\mathcal{F}) \in I$ $P$-f.s. für jede Unter-$\sigma$-Algebra $\mathcal{F} \subset \mathcal{A}$ und im Fall der Integrierbarkeit von $\varphi \circ X$ gilt

$$\varphi\big(\mathbb{E}(X|\mathcal{F})\big) \leq \mathbb{E}\big(\varphi(X)|\mathcal{F}\big) \quad P\text{-f.s.}$$

*Beweis:* Mit $P(X \in I) = 1$ folgt $P\big(\mathbb{E}(X|\mathcal{F}) \in I\big) = 1$, wobei wir die Monotonie, Satz 1.3 (b), verwenden: Aus $X(\omega) < \beta$ für alle $\omega \in \Omega$ folgt $\mathbb{E}(X|\mathcal{F}) \leq \beta$ $P$-f.s., also $\mathbb{E}(\beta - X|\mathcal{F}) \geq 0$ $P$-f.s. Nun ist $C := \{\mathbb{E}(\beta - X|\mathcal{F}) = 0\} \in \mathcal{F}$ und $\int_C (\beta - X)\, dP = 0$, somit $P(C) = 0$, also $\mathbb{E}(X|\mathcal{F}) < \beta$ $P$-f.s. Aus $\alpha < X(\omega)$ für alle $\omega \in \Omega$ folgt analog $\alpha < \mathbb{E}(X|\mathcal{F})$ $P$-f.s. In der Analysis lernt man, dass die konvexe Funktion $\varphi$ in $I$ rechts- und linksseitig differenzierbar ist, also auch stetig. Die rechtsseitige Ableitung $\varphi'_+(x)$ ist auf $I$ isoton. Ferner gilt

$$\varphi(y) \geq \varphi(x) + \varphi'_+(x)(y - x)\ ,\ x, y \in I \tag{1.4}$$

Im Fall $x < y$ und $t \in (x, y)$ folgt dies aus

$$\frac{\varphi(x) - \varphi(t)}{x - t} \leq \frac{\varphi(x) - \varphi(y)}{x - y}$$

(was aus der Konvexität gefolgert werden kann). Für $x = y$ folgt Gleichheit in (1.4), also

$$\varphi(y) = \sup_{x \in I} \Big(\varphi(x) + \varphi'_+(x)(y - x)\Big), \quad y \in I\ .$$

Es gilt

$$\varphi(y) = \sup_{x \in I \cap \mathbb{Q}} \Big(\varphi(x) + \varphi'_+(x)(y - x)\Big), \quad y \in I\ , \tag{1.5}$$

denn $\varphi'_+$ ist isoton, also lokal beschränkt, und $\varphi$ ist stetig. Also $\varphi \circ X \geq \varphi(x) + \varphi'_+(x)(X - x)$, und somit

$$\mathbb{E}(\varphi(X)|\mathcal{F}) \geq \varphi(x) + \varphi'_+(x)(\mathbb{E}(X|\mathcal{F}) - x)$$

$P$-f.s. für jedes $x \in I \cap \mathbb{Q}$. Wegen der Abzählbarkeit von $I \cap \mathbb{Q}$ gibt es ein $C \in \mathcal{F}$, $P(C) = 0$, mit

$$\mathbb{E}(\varphi(X)|\mathcal{F}) \geq \varphi(x) + \varphi'_+(x)(\mathbb{E}(X|\mathcal{F}) - x) \quad \text{auf } \Omega \setminus C$$

für jedes $x \in I \cap \mathbb{Q}$. Also folgt die Aussage auf $\Omega \setminus (C \cup D)$ mittels Anwendung von (1.5), wenn $D$ ein Ereignis mit Wahrscheinlichkeit Null mit $\mathbb{E}(X|\mathcal{F})(\omega) \in I$ für alle $\omega \in \Omega \setminus D$ ist. $\square$

**1.8 Korollar** *Seien $p \geq 1$, $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ und $\mathcal{F} \subset \mathcal{A}$. Dann $|\mathbb{E}(X|\mathcal{F})|^p \leq \mathbb{E}(|X|^p|\mathcal{F})$ $P$-f.s., sowie $\mathbb{E}|\mathbb{E}(X|\mathcal{F})|^p \leq \mathbb{E}|X|^p$ (siehe Satz 1.5 (c)). Insbesondere ist $\mathbb{E}(X|\mathcal{F}) \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$.*

*Beweis:* $p = \infty$: mit $X$ ist auch $\mathbb{E}(X|\mathcal{F})$ fast sicher beschränkt, siehe Satz 1.3 (c), sowie

$$X = \text{const.} = \alpha \quad \Rightarrow \quad \mathbb{E}(X|\mathcal{F}) = \alpha \ P\text{-f.s.}$$

□

**1.9 Satz** *$X, Y$ seien zwei Zufallsgrößen auf einem W-Raum $(\Omega, \mathcal{A}, P)$ und $\mathcal{F} \subset \mathcal{A}$ eine Unter-$\sigma$-Algebra. Dann gilt:*

(a) (Hölder-Ungleichung) *Aus $\mathbb{E}|X|^p < \infty$ und $\mathbb{E}|Y|^q < \infty$ für $1 \le p, q \le \infty$ mit $\frac{1}{p} + \frac{1}{q} = 1$ folgt*

$$\big|\mathbb{E}(XY|\mathcal{F})\big| \le \mathbb{E}\big(|XY|\big|\mathcal{F}\big) \le \Big(\mathbb{E}\big(|X|^p\big|\mathcal{F}\big)\Big)^{\frac{1}{p}} \Big(\mathbb{E}\big(|Y|^q\big|\mathcal{F}\big)\Big)^{\frac{1}{q}} < \infty \quad P\text{-f.s.}$$

(b) (Cauchy-Schwarz-Ungleichung) *Die Aussage aus (i) für $p = q = 2$.*

(c) (Minkowski-Ungleichung) *Aus $\mathbb{E}|X|^p < \infty$ und $\mathbb{E}|Y|^p < \infty$ für $p \ge 1$ folgt*

$$\Big(\mathbb{E}\big(|X+Y|^p\big|\mathcal{F}\big)\Big)^{\frac{1}{p}} \le \Big(\mathbb{E}\big(|X|^p\big|\mathcal{F}\big)\Big)^{\frac{1}{p}} + \Big(\mathbb{E}\big(|Y|^p\big|\mathcal{F}\big)\Big)^{\frac{1}{p}} < \infty \quad P\text{-f.s.}$$

*Beweis:* Übung.

**1.10 Bemerkungen**

(a) Eine Beweisvariante zur Existenz bedingter Erwartungswerte verwendet den Satz von Radon-Nikodym[4]. $X$ sei eine nicht-negative Zufallsgröße auf $(\Omega, \mathcal{A}, P)$, $\mathcal{F} \subset \mathcal{A}$ eine Unter-$\sigma$-Algebra. Es sei

$$Q_{X,\mathcal{F}}(A) := \int_A X \, dP \ , \quad A \in \mathcal{F},$$

ein Maß auf $(\Omega, \mathcal{F})$. Ist $P|_{\mathcal{F}}$ die Einschränkung von $P$ auf $\mathcal{F}$, so ist jede $P|_{\mathcal{F}}$-Nullmenge auch eine $Q_{X,\mathcal{F}}$-Nullmenge. $P|_{\mathcal{F}}$ ist als W-Maß $\sigma$-endlich, also existiert nach Satz 1.19, Wahrscheinlichkeitstheorie, eine $\mathcal{F}$-messbare Zufallsgröße $Y$ ($P|_{\mathcal{F}}$-f.s. eindeutig) mit

$$\int_A X \, dP = Q_{X,\mathcal{F}}(A) = \int_A Y \, dP|_{\mathcal{F}} = \int_A Y \, dP \ , \quad A \in \mathcal{F} \ .$$

$Y$ besitzt also die Eigenschaften (I) und (II) einer Version des bedingten Erwartungswertes $\mathbb{E}(X|\mathcal{F})$.

[4]siehe Wahrscheinlichkeitstheorie, Satz 1.19

(b) $P(A|\mathcal{F})$, wie nach Satz 1.5 eingeführt, ist eine $\mathcal{F}$-messbare, nicht-negative Funktion (f.s.) mit

$$\int_C P(A|\mathcal{F})\, dP = \int_C 1_A\, dP = P(A \cap C)\ , \quad C \in \mathcal{F}\ .$$

Wenn wir in der Situation von Beispiel 1.4 (b) sind, so folgt

$$P(A|A_1, \dots, A_n) = \sum_{j=1}^{n} \frac{P(A \cap A_j)}{P(A_j)} 1_{A_j} \quad P\text{-f.s.}$$

Es gilt $0 \leq P(A|\mathcal{F}) \leq 1$ f.s., $P(\emptyset|\mathcal{F}) = 0$ f.s., $P(\Omega|\mathcal{F}) = 1$ f.s. Aus $A_1 \subset A_2$, $A_1, A_2 \in \mathcal{A}$, folgt $P(A_1|\mathcal{F}) \leq P(A_2|\mathcal{F})$ $P$-f.s. und für eine Folge $(A_n)_n$ paarweise fremder Ereignisse aus $\mathcal{A}$ gilt:

$$P\Big(\bigcup_{n\geq 1} A_n|\mathcal{F}\Big) = \sum_{n\geq 1} P(A_n|\mathcal{F}) \quad P\text{-f.s.} \tag{1.6}$$

(für jede Folge von Versionen der bedingten Wahrscheinlichkeiten $P(A_n|\mathcal{F})$ liefert die rechte Seite eine Version der bedingten Wahrscheinlichkeit auf der linken Seite!)

*Es gilt nicht:* Für fast alle $\omega \in \Omega$ ist

$$A \mapsto P(A|\mathcal{F})(\omega)$$

ein W-Maß auf $\mathcal{A}$! Denn in jeder der genannten Eigenschaften tritt eine von der betreffenden Gegebenheit, zum Beispiel in (1.6) von der Folge $(A_n)_n$ abhängige $P$-Nullmenge, als Ausnahmemenge auf. Die Vereinigung dieser im Allgemeinen überabzählbar vielen Mengen ist daher meist keine Nullmenge mehr!

Wir greifen das in (b) gestellte Problem auf und teilen inoffiziell mit:

**1.11 Definition** Es seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ zwei Messräume. Eine Funktion $K : \Omega \times \mathcal{A}' \to \mathbb{R}$ heißt Markov*scher Kern* von $(\Omega, \mathcal{A})$ nach $(\Omega', \mathcal{A}')$, wenn

(a) $K(\cdot, A') : \Omega \to \mathbb{R}$ $\mathcal{A}$-messbar ist für jedes $A' \in \mathcal{A}'$

(b) $K(\omega, \cdot) : \mathcal{A}' \to \mathbb{R}$ ein W-Maß auf $(\Omega', \mathcal{A}')$ für alle $\omega \in \Omega$ ist.

Es sei weiter $(\Omega, \mathcal{A}, P)$ ein W-Raum und $\mathcal{F}$ eine Unter-$\sigma$-Algebra von $\mathcal{A}$. Ein Markovscher Kern von $(\Omega, \mathcal{F})$ nach $(\Omega, \mathcal{A})$ mit

$$K(\cdot, A) = P(A|\mathcal{F})\ P\text{-f.s.}, \quad A \in \mathcal{A}\ ,$$

heißt *reguläre bedingte Wahrscheinlichkeit* bezüglich $\mathcal{F}$.

**1.12 Satz** *Ist $\Omega$ ein polnischer Raum, $\mathcal{A}$ die $\sigma$-Algebra der* BOREL*schen Mengen und $P$ ein W-Maß auf $(\Omega, \mathcal{A})$. Dann existiert für jede $\sigma$-Algebra $\mathcal{F} \subset \mathcal{A}$ eine reguläre bedingte Wahrscheinlichkeit bezüglich $\mathcal{F}$. (ohne Beweis)*

Wir wenden uns noch $\mathbb{E}(X|Y) := \mathbb{E}\big(X\big|\sigma(Y)\big)$ zu. Sei $X$ eine integrierbare reelle Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Weiter seien $Y : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ eine Zufallsvariable und $\mathbb{E}(X|Y)$ eine reelle integrierbare Version der bedingten Erwartung. Da $\mathbb{E}(X|Y)$ $\sigma(Y)$-messbar ist, existiert eine von der gewählten Version abhängige messbare reelle Funktion $g$ auf $(\Omega', \mathcal{A}')$ mit

$$\mathbb{E}(X|Y) = g \circ Y \ . \tag{1.7}$$

Dies folgt aus dem

**1.13 Lemma (Faktorisierungslemma)** Sei $Y : \Omega \to \Omega'$ eine Abbildung der Menge $\Omega$ in den Messraum $(\Omega', \mathcal{A}')$ und $Z : \Omega \to \bar{\mathbb{R}}$ eine numerische Funktion auf $\Omega$. Genau dann ist $Z$ $\sigma(Y)$-$\mathcal{B}$-messbar, wenn es eine messbare numerische Funktion $g$ auf $(\Omega', \mathcal{A}')$ gibt mit $Z = g \circ Y$.

*Beweis:* Siehe Übung 3 der Wahrscheinlichkeitstheorie. $Z = g \circ Y$ ist als Komposition einer $\sigma(Y)$-$\mathcal{A}'$-messbaren und einer $\mathcal{A}'$-$\bar{\mathcal{B}}$-messbaren Abbildung $\sigma(Y)$-$\bar{\mathcal{B}}$-messbar! Sei umgekehrt $Z = \sum_{j=1}^n \alpha_j 1_{A_j} \in \mathcal{E}\big(\Omega, \sigma(Y)\big)$, also $A_j \in \sigma(Y)$ und $\alpha_j \geq 0$, $j = 1, \ldots, n$. Dann existieren $A'_j \in \mathcal{A}'$ mit $A_j = Y^{-1}(A'_j)$, und die Behauptung folgt mit $g = \sum_{j=1}^n \alpha_j 1_{A'_j}$. Jetzt folgt noch das Funktions-Erweiterungsargument (Übung). $\square$

**1.14 Satz** *Jede $\mathcal{A}'$-messbare reelle Funktion $g$, welche sich mit (1.7) aus einer reellen Version $\mathbb{E}(X|Y)$ ableitet, ist $P^Y$-integrierbar und es gilt*

$$\int_{A'} g \, dP^Y = \int_{\{Y \in A'\}} X \, dP \ , \quad A' \in \mathcal{A}' \tag{1.8}$$

*Sie ist $P^Y$-fast sicher eindeutig bestimmt. Ist $g$ eine $\mathcal{A}'$-messbare, reelle $P^Y$-integrierbare Funktion auf $\Omega'$, welche (1.8) erfüllt, so ist $g \circ Y$ eine Version von $\mathbb{E}(X|Y)$.*

*Beweis:* Mit $X$ ist $\mathbb{E}(X|Y)$ $P$-integrierbar und der Transformationssatz liefert für jedes $A' \in \mathcal{A}'$

$$\int_{\{Y \in A'\}} X \, dP = \int_{\{Y \in A'\}} \mathbb{E}(X|Y) \, dP = \int_{\{Y \in A'\}} g \circ Y \, dP = \int_{A'} g \, dP^Y \ ,$$

also ist $g$ $P^Y$-integrierbar. Damit gilt (1.8). Ist $h$ eine weitere $\mathcal{A}'$-messbare reelle Funktion mit $\mathbb{E}(X|Y) = h \circ Y$, so folgt $\int_{A'} g \, dP^Y = \int_{A'} h \, dP^Y$, $A' \in \mathcal{A}'$.

$g$ und $h$ sind $P^Y$-integrierbar und

$$\int_{A'} (g^+ + h^-)\, dP^Y = \int_{A'} (g^- + h^+)\, dP^Y \ , \quad A' \in \mathcal{A}' \ ,$$

also $g^+ + h^- = g^- + h^+$, also $g = h$ $P^Y$-f.s.

Gilt umgekehrt (1.8) für $g$, reell und $P^Y$-integrierbar, so ist $g \circ Y$ $\sigma(Y)$-messbar. Es liegt eine Version von $\mathbb{E}(X|Y)$ vor, denn

$$\int_{\{Y \in A'\}} g \circ Y \, dP = \int_{A'} g \, dP^Y = \int_{\{Y \in A'\}} X \, dP \ , \quad A' \in \mathcal{A}' \ ,$$

also $\int_C g \circ Y \, dP = \int_C X \, dP$ für alle $C \in \sigma(Y)$. $\square$

Sei nun $\{y\} \in \mathcal{A}'$, $y \in \Omega'$, so liefert (1.8):

$$g(y) P(Y = y) = \int_{\{Y=y\}} X \, dP.$$

Wenn $P(Y = y) > 0$, folgt

$$g(y) = \frac{1}{P(Y = y)} \int_{\{Y=y\}} X \, dP =: \mathbb{E}_{\{Y=y\}}(X) =: \mathbb{E}(X|Y = y)$$

(siehe unsere anfängliche Diskussion).

$$\Rightarrow \mathbb{E}(X|Y)(\omega) = \mathbb{E}(X|Y = y) \quad \text{für alle } \omega \in \{Y = y\}.$$

Auf $\{Y = y\}$ mit $P(Y = y) > 0$ ist somit jede Version der bedingten Erwartung $\mathbb{E}(X|Y)$ konstant und gleich der bedingten Erwartung von $X$ unter der Bedingung $\{Y = y\}$.

Im allgemeinen ist jedoch $P(Y = y)$ für viele $y \in \Omega'$ gleich Null (Lebesgue-stetig verteilte Zufallsgrößen sind eine Beispielklasse). Aber der Wert von $g$ an einer Stelle $y \in \Omega'$ steht zur Verfügung.

**1.15 Definition** Für integrierbares, reelles $X$ sei $g$ eine der Bedingung (1.8) genügende $\mathcal{A}'$-messbare, $P^Y$-integrierbare, reelle Funktion. Dann heißt $g(y)$ für jedes $y \in \Omega'$ *bedingte Erwartung* von $X$ unter der Bedingung, dass $Y$ gleich $y$ ist, in Zeichen

$$\mathbb{E}(X|Y = y) := g(y) \ .$$

$\mathbb{E}(X|Y = y)$ ist eine Zahl! Nach Satz 1.14 gewinnt man aus der Vergabe von $g(y) = \mathbb{E}(X|Y = y)$ eine Version der bedingten Erwartung $\mathbb{E}(X|Y)$ in der Form $\omega \mapsto \mathbb{E}\big(X \big| Y = Y(\omega)\big)$. Also $\mathbb{E}(X|Y)(\omega) = \mathbb{E}\big(X \big| Y = Y(\omega)\big)$ für fast alle $\omega \in \Omega$. Manchmal kann man $y \mapsto \mathbb{E}(X|Y = y)$ berechnen:

**1.16 Satz** *$X, Y$ seien Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$, deren gemeinsame Verteilung $P^{(X,Y)}$ eine W-Dichte $f$ besitzt: $P^{(X,Y)} = f\lambda^2$, $f : \mathbb{R}^2 \to \mathbb{R}_+$ eine $\mathcal{B}^2$-messbare Funktion. Ferner sei $X$ integrierbar und*

$$f_0(y) := \int f(x,y)\,dx > 0 \quad \text{für alle } y \in \mathbb{R}\ .$$

*Dann ist*

$$\mathbb{E}(X|Y = y) = \frac{1}{f_0(y)} \int x f(x,y)\,dx \quad \text{für } P^Y\text{-fast alle } y \in \mathbb{R}\ . \tag{1.9}$$

*Es gilt weiter*

$$\mathbb{E}(X|Y) = \frac{1}{f_0(Y)} \int x f(x,Y)\,dx \quad P\text{-f.s.}$$

*Beweis:* Es gilt

$$\begin{aligned}
\int_{\{Y \in \mathcal{A}'\}} X\,dP &= \int X 1_{A'}(Y)\,dP = \int X 1_{\mathbb{R}\times A'}(X,Y)\,dP \\
&= \int x 1_{\mathbb{R}\times A'}(x,y)\,P^{(X,Y)}\big(d(x,y)\big) = \int_{\mathbb{R}\times A'} x f(x,y)\,dx\,dy\ .
\end{aligned}$$

Insbesondere ist ($A' = \mathbb{R}$): $\mathbb{E}(X) = \int_{\mathbb{R}\times\mathbb{R}} x f(x,y)\,dx\,dy$, also $(x,y) \mapsto x f(x,y)$ $\lambda^2$-integrierbar, also $y \mapsto \int x f(x,y)\,dx$ $\lambda^1$-fast überall definiert und $\lambda^1$-integrierbar (Fubini).

Da $\int f\,d\lambda^2 = 1$, folgt analog $f_0(y) < \infty$ $\lambda^1$-fast überall. Also existiert eine messbare Funktion $g : \mathbb{R} \to \mathbb{R}$:

$$g(y) := \begin{cases} \frac{1}{f_0(y)} \int x\,f(x,y)\,dx, & \text{falls } y \in \mathbb{R} \setminus N\ , \\ 0, & y \in N\ , \end{cases}$$

mit $\lambda^1(N) = 0$ und $N$ so, dass $\int |x| f(x,y)\,dx < \infty$ und $f_0(y) < \infty$ für alle $y \in N^c$. $N$ ist auch eine $P^Y$-Nullmenge, denn

$$P^Y(N) = P\big((X,Y) \in \mathbb{R} \times N\big) = \int_{\mathbb{R}\times N} f(x,y)\,dx\,dy = 0\ .$$

Sei $g$ $P^Y$-integrierbar, so gilt für alle $A' \in \mathcal{B}^1$:

$$\begin{aligned}
\int_{A'} g\,dP^Y &= \int_{\{Y \in A'\}} g \circ Y\,dP = \int 1_{\mathbb{R}\times A'}(X,Y) g \circ Y\,dP \\
&= \int_{\mathbb{R}\times A'} g(y) f(x,y)\,dx\,dy = \int_{A'} g(y) \Big( \int_{\mathbb{R}} f(x,y)\,dx \Big)\,dy \\
&= \int_{A'} g(y) f_0(y)\,dy = \int_{A'} \int_{\mathbb{R}} x f(x,y)\,dx\,dy = \int_{\mathbb{R}\times A'} x f(x,y)\,dx\,dy.
\end{aligned}$$

Mit der anfänglichen Rechnung folgt

$$\int_{A'} g\, dP^Y = \int_{\{Y \in A'\}} X\, dP \; .$$

Für $A' = \mathbb{R}$ und $|g|$ anstelle von $g$ folgt analog

$$\begin{aligned} \int |g|\, dP^Y &= \int |g| f_0\, d\lambda^1 = \int \Big| \int x f(x,y)\, dy \Big|\, dx \\ &\leq \int \int |x f(x,y)|\, dx\, dy = \mathbb{E}(|X|) < \infty \; , \end{aligned}$$

also die $P^Y$-Integrierbarkeit von $g$. Nun ist $g$ also eine Funktion $y \mapsto \mathbb{E}(X|Y = y)$ wie in Definition 1.15 und Satz 1.14 liefert die Behauptungen. $\square$

**1.17 Beispiel** $(X,Y)$ besitze eine $N\Big(0, \frac{1}{1-\rho^2}\binom{\sigma^2\ \rho\sigma^2}{\rho\sigma^2\ \sigma^2}\Big)$-Verteilung, mit Dichte

$$f(x,y) = \frac{(1-\rho^2)^{1/2}}{2\pi\sigma^2} \exp\Big( - \frac{x^2 - 2\rho xy + y^2}{2\sigma^2} \Big)$$

für $(\sigma, \rho) \in (0,\infty) \times (-1,1)$. Dann ist

$$\begin{aligned} f_0(y) &= \int_{\mathbb{R}} f(x,y)\, dx \\ &= \frac{(1-\rho^2)^{1/2}}{2\pi\sigma^2} \exp\Big( - \frac{y^2(1-\rho^2)}{2\sigma^2} \Big) \int_{\mathbb{R}} \exp\Big( - \frac{(x-\rho y)^2}{2\sigma^2} \Big)\, dx \\ &= \frac{(1-\rho^2)^{1/2}}{2\pi\sigma^2} \exp\Big( - \frac{y^2(1-\rho^2)}{2\sigma^2} \Big) (2\pi\sigma^2)^{1/2} \\ &= \Big( \frac{(1-\rho^2)}{2\pi\sigma^2} \Big)^{1/2} \exp\Big( - \frac{y^2(1-\rho^2)}{2\sigma^2} \Big) \; . \end{aligned}$$

Also ist $Y$ $N\Big(0, \frac{\sigma^2}{1-\rho^2}\Big)$-verteilt. Gemäß (1.9) folgt nun

$$\mathbb{E}(X|Y=y) = \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{\mathbb{R}} x \exp\Big( - \frac{(x-\rho y)^2}{2\sigma^2} \Big)\, dx = \rho\, y \; ,$$

also $\mathbb{E}(X|Y) = \rho Y$ $P$-f.s. $\square$

KAPITEL 2

# Die Theorie der Martingale

Die von J. L. DOOB begründete Theorie der Martingale ist zentral für die Untersuchung der fast sicheren Konvergenz von Folgen von Zufallsvariablen. Der Zugang des Begriffs erfolgt über den Begriff der bedingten Erwartung.

Es sei $(X_n)_n$ eine unabhängige Folge integrierbarer reeller Zufallsvariablen, $S_n := X_1 + \cdots + X_n$. Offenbar gilt $\sigma(S_1, \ldots, S_n) = \sigma(X_1, \ldots, X_n)$. Da $X_{n+1}$ unabhängig von $\sigma(X_1, \ldots, X_n)$ ist, also auch von $\sigma(S_1, \ldots, S_n)$, folgt

$$\mathbb{E}\big(X_{n+1}|S_1, \ldots, S_n\big) = \mathbb{E}(X_{n+1}) \quad \text{P-f.s.}$$

Weiter gilt $\mathbb{E}(X_i|S_1, \ldots, S_n) = X_i$ P-f.s. für alle $i = 1, \ldots, n$, also bei Summation

$$\mathbb{E}(S_{n+1}|S_1, \ldots, S_n) = S_n + \mathbb{E}(X_{n+1}) \quad \text{P-f.s.}$$

Sind alle $X_n$ zentriert, so folgt

$$\mathbb{E}(S_{n+1}|S_1, \ldots, S_n) = S_n \quad \text{P-f.s.}$$

Gilt $\mathbb{E}(X_n) \leq 0$ (bzw. $\geq 0$) für alle $n$, so folgt entsprechend

$$\mathbb{E}(S_{n+1}|S_1, \ldots, S_n) \leq S_n \quad (\geq S_n) \quad \text{P-f.s.}$$

jeweils für alle $n \in \mathbb{N}$.

Sei speziell $X_n$ $\pm 1$-wertig mit $P(X_n = 1) = p$ und somit $\mathbb{E}(X_n) = 2p - 1$ (Spiel zwischen Spieler und Bank eines Spielcasinos). Der Spieler verfolge die folgende Strategie:

Vor Eintritt in das Spiel wählt er eine Folge $(e_n)_n$, $e_n : \{-1, +1\}^n \to \mathbb{R}^+$. Für die $(n+1)$-te Spielrunde wird der vom Ausgang der vorangegangenen $n$ Runden abhängige Einsatz $e_n(X_1, \ldots, X_n)$ geleistet. Die erste Runde wird ohne vorherigen Einsatz gespielt. Der Gegenspieler leiste keine Einsätze. $S_n$ bezeichne den Gesamtgewinn des Spielers nach der $n$-ten Runde:

$$\begin{aligned} S_1 &:= X_1 \\ S_{n+1} &:= S_n + e_n(X_1, \ldots, X_n) \cdot X_{n+1}\,. \end{aligned}$$

Es gilt:

$$\begin{aligned} \mathbb{E}(S_{n+1}|X_1, \ldots, X_n) &= S_n + e_n(X_1, \ldots, X_n)\mathbb{E}(X_{n+1}|X_1, \ldots, X_n) \\ &= S_n + e_n(X_1, \ldots, X_n)\mathbb{E}(X_{n+1}) \\ &= S_n + (2p-1)e_n(X_1, \ldots, X_n). \end{aligned}$$

Für $p = 1/2$ folgt

$$\mathbb{E}(S_{n+1}|X_1, \ldots, X_n) = S_n \quad \text{P-f.s.}$$

Analog für $p < 1/2$ ($p > 1/2$):

$$\mathbb{E}(S_{n+1}|X_1, \ldots, X_n) \leq S_n \quad (\geq S_n) \quad \text{P-f.s.}$$

Wir nehmen diese Beobachtungen zum Anlass für die folgenden Definitionen. Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum.

**2.1 Definition**

(a) Eine *Filtrierung* von $\mathcal{A}$ ist eine Folge $(\mathcal{A}_n)_n$ von Teil-$\sigma$-Algebren von $\mathcal{A}$ mit $\mathcal{A}_m \subset \mathcal{A}_n$ für $m \leq n$. Eine Folge von Zufallsvariablen $(X_n)_n$ auf $(\Omega, \mathcal{A}, P)$ heißt $(\mathcal{A}_n)_n$-*angepasst* (oder *adaptiert*), wenn $X_n$ für jedes $n \in \mathbb{N}$ $\mathcal{A}_n$-messbar ist.

(b) Sei $(X_n)_n$ eine $(\mathcal{A}_n)_n$-adaptierte Folge von integrierbaren Zufallsgrößen. $(X_n)_n$ heißt

- $(\mathcal{A}_n)_n$-*Martingal*, falls $\mathbb{E}(X_n|\mathcal{A}_m) = X_m$ P-f.s. für alle $m, n \in \mathbb{N}$ mit $m \leq n$ gilt.
- $(\mathcal{A}_n)_n$-*Submartingal*, falls $\mathbb{E}(X_n|\mathcal{A}_m) \geq X_m$ P-f.s. für alle $m, n \in \mathbb{N}$ mit $m \leq n$ gilt.
- $(\mathcal{A}_n)_n$-*Supermartingal*, falls $\mathbb{E}(X_n|\mathcal{A}_m) \leq X_m$ P-f.s. für alle $m, n \in \mathbb{N}$ mit $m \leq n$ gilt.

**2.2 Bemerkungen**

(a) $(X_n)_n$ ist adaptiert an $(\mathcal{F}_n^X)_n$ mit $\mathcal{F}_n^X := \sigma(X_j, j \leq n)$. Man spricht bei dieser Filtrierung auch von der *kanonischen Filtrierung*.

(b) Um nachzuprüfen, dass $(X_n)_n$ ein Martingal ist, reicht es natürlich, dass

$$\mathbb{E}(X_{n+1}|\mathcal{A}_n) = X_n \quad \text{P-f.s.}$$

für alle $n \in \mathbb{N}$ gilt. (Entsprechendes für Sub- bzw. Supermartingale.) Dies folgt aus der P-f.sicheren Gleichheit

$$\mathbb{E}(X_{n+2}|\mathcal{A}_n) = E\big(E(X_{n+2}|\mathcal{A}_{n+1})|\mathcal{A}_n\big) = E(X_{n+1}|\mathcal{A}_n) = X_n.$$

(c) Das Wort „Martingal“ hat mehrere Bedeutungen. Es steht für „Hilfszügel beim Zaumzeug“ eines Reitpferdes, welches zu starke Kopfbewegungen des Pferdes verhindert, aber auch für ein die Takelage bei Segelschiffen absicherndes Seil. Vor allem bedeutet es aber wohl eine Spielstrategie beim Roulettespiel, im Provenzalischen genannt „jouga a la martegalo“. Diese Strategie besteht in der jeweiligen Verdoppelung des beim vorausgegangenen Spiel verlorenen Einsatzes, ein Spezialfall des in den Vorbemerkungen betrachteten Beispiels, welches wir in Beispiel 2.3 (g) genauer betrachten.

Bei der Wahl der kanonischen Filtrierung $(\mathcal{F}_n^X)_n$ nennen wir die Filtrierung nich extra.

Wir sammeln Beispiele:

**2.3 Beispiele**

(a) Die eindimensionale symmetrische Irrfahrt auf $\mathbb{Z}$ mit Zeitachse $\mathbb{N}_0$ ist ein Martingal.

(b) Sei $(X_n)_n$ die symmetrische Irrfahrt auf $\mathbb{N}$ mit Start in 1 und Absorbtion in 0. Also $p_{i,i+1} = p_{i,i-1} = 1/2$ für $i \in \mathbb{N}$ und $p_{0,0} = 1$. Also gilt für $n \in \mathbb{N}_0$ fast sicher

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n^X) = \mathbb{E}(X_{n+1}|X_1, \ldots, X_n) = \sum_{j \in \mathbb{N}_0} j p_{X_n,j} = X_n\,,$$

also ist diese Irrfahrt ein Martingal.

(c) In den Vorbemerkungen haben wir schon die Verallgemeinerung von (a) gesehen: Sind $(\xi_n)_n$, $n \in \mathbb{N}$, unabhängige und identisch verteilte, integrierbare Zufallsgrößen mit $\mathbb{E}(\xi_i) = 0$, so sei $X_0 := 0$; $X_n := \sum_{j=1}^n \xi_j$, $n \in \mathbb{N}$, und $\mathcal{A}_0 := \{\emptyset, \Omega\}$ und $\mathcal{A}_n := \sigma(\xi_j,\, j \leq n)$. Dann ist $(X_n)_n$ ein $(\mathcal{A}_n)_n$-Martingal.

(d) (Levy's *Martingal*) Es sei $(\mathcal{A}_n)_n$ eine beliebige Filtrierung und $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Dann ist die Folge der bedingten Erwartungswerte

$$X_n := \mathbb{E}(X|\mathcal{A}_n)\,, \quad n \in \mathbb{N}\,,$$

offensichtlich ein Martingal, denn es gilt P-f.s.

$$\mathbb{E}(X_{n+1}|\mathcal{A}_n) = \mathbb{E}\big(\mathbb{E}(X|\mathcal{A}_{n+1})|\mathcal{A}_n\big) = \mathbb{E}(X|\mathcal{A}_n)\,.$$

Nicht jedes Martingal kann in dieser Weise dargestellt werden, wie wir noch sehen werden.

(e) (*exponentielles Martingal*) Die $\xi_n$, $n \in \mathbb{N}$, und die Filtrierung seien wie in (c) gewählt. Es sei $\mathbb{E}\xi_i \neq 0$ zugelassen. Es existiere ein $\lambda_0 > 0$ mit $M(\lambda) := \mathbb{E}(e^{\lambda \xi_i}) < \infty$ für $|\lambda| \leq \lambda_0$.

Sei $X_0 \equiv 1$ und

$$X_n := \exp\Big(\lambda \sum_{j=1}^n \xi_j\Big) M(\lambda)^{-n}\,, \quad n \in \mathbb{N}\,.$$

Dann ist für die entsprechenden Werte von $\lambda$ die Folge $(X_n)_n$ ein $(\mathcal{A}_n)_n$-Martingal, denn es gilt P-f.s.

$$\begin{aligned}\mathbb{E}(X_{n+1}|\mathcal{A}_n) &= M(\lambda)^{-n-1}\exp\big(\lambda\sum_{j=1}^n \xi_j\big)\mathbb{E}(e^{\lambda\xi_{n+1}}|\mathcal{A}_n)\\ &= M(\lambda)^{-n-1}\exp\big(\lambda\sum_{j=1}^n \xi_j\big)\mathbb{E}(e^{\lambda\xi_{n+1}})\\ &= M(\lambda)^{-n}\exp\big(\lambda\sum_{j=1}^n \xi_j\big) = X_n\,.\end{aligned}$$

(f) Durch Differentiation nach $\lambda$ in $\lambda = 0$ lassen sich aus Beispiel (e) neue Martingale gewinnen. Einmalige Differentiation liefert (c), zweimaliges Differenzieren:

$$\frac{d^2}{d\lambda^2}X_n\Big|_{\lambda=0} = S_n^2 - 2nS_nM'(0) + n(n+1)M'(0)^2 - nM''(0)$$

mit $S_n := \sum_{j=1}^n \xi_j$, $M'(0) = \mathbb{E}\xi_i$, $M''(0) = \mathbb{E}(\xi_i^2)$. Die rechte Seite ist ein $(\mathcal{A}_n)_n$-Martingal (Übung), sofern $\xi_i \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$.

(g) (*Martingal-Wettstrategie*) Sei $(X_n)_n$ die Folge $\pm 1$-wertiger unabhängiger Zufallsgrößen mit $P(X_i = 1) = 1/2$. Der Spieler[1] setzt 1 Einheit Einsatz zu Beginn und verdoppelt seinen Einsatz so lange, bis zum ersten Mal 1 erscheint, um dann aufzuhören! $M_n$ bezeichne den Gewinn nach $n$ Würfen mit $M_0 = 0$. Im Fall eines Gewinns hört der Spieler auf: $P(M_{n+1} = 1|M_n = 1) = 1$. Erscheint $n$ Mal stets $-1$, beträgt der Verlust $1+2+4+\dots+2^{n-1} = 2^n - 1$ Einheiten. Dann wird der Einsatz auf $2^n$ verdoppelt. Es gilt:

$$P\big(M_{n+1} = 1|M_n = -(2^n-1)\big) = P(X_{n+1} = 1) = \frac{1}{2}$$

und

$$P\big(M_{n+1} = -(2^{n+1}-1)|M_n = -(2^n-1)\big) = P(X_{n+1} = -1) = \frac{1}{2}\,,$$

also

$$\mathbb{E}\big(M_{n+1}|M_n = -(2^n-1)\big) = 1\cdot\frac{1}{2} + (-2^{n+1}+1)\cdot\frac{1}{2} = -2^n + 1\,,$$

also ist $(M_n)_n$ ein Martingal bezüglich $(\mathcal{A}_n)_n$ mit $\mathcal{A}_n := \sigma(M_0, X_1, \dots, X_n)$.

[1] siehe einleitendes Beispiel

**2.4 Lemma** *Sei $(X_n)$ ein $(\mathcal{A}_n)_n$-Martingal bzw. Submartingal.*

(a) *Ist $(\mathcal{B}_n)_n$ eine weitere Filtrierung von $\mathcal{A}$ mit $\mathcal{B}_n \subset \mathcal{A}_n$ für alle $n \in \mathbb{N}$ und ist $(X_n)_n$ an $(\mathcal{B}_n)_n$ angepasst, so ist $(X_n)_n$ auch ein $(\mathcal{B}_n)_n$-Martingal (bzw. Submartingal)*

(b) *Es gilt $\mathbb{E}(X_n) = \mathbb{E}(X_m)$ (bzw. $\mathbb{E}(X_n) \geq \mathbb{E}(X_m)$) für alle $m, n \in \mathbb{N}$ mit $m \leq n$.*

(c) *Ist $\varphi$ eine konvexe Funktion mit $\mathbb{E}\,|\varphi(X_n)| < \infty$, $n \in \mathbb{N}$, so ist $\big(\varphi(X_n)\big)_n$ ein Submartingal bzgl. $(\mathcal{A}_n)_n$.*

*Beweis:* Definition 2.1, Satz 1.7. □

Martingale beschreiben im Sinne der einführenden Beispiele faire Spiele. Beispiel 2.3 (g) beschreibt eine Strategie, mit der in einem fairen Spiel Gewinn gemacht werden kann! Es gilt für $\tau := \inf\{n \geq 0; M_n = 1\}$ stets $M_\tau = 1$, während $M_0 = 0$ ist. $\tau$ ist dabei eine zufällige Zeit. Bei gewissen Einschränkungen an diese zufällige Zeit $\tau$ werden wir aber sehen, dass man „in einem fairen Spiel" keinen Gewinn machen kann. Dies wird der Inhalt des sogenannten „Optional-Sampling-Theorems" werden.

**2.5 Definition** Sei $(\mathcal{A}_n)_n$ eine Filtrierung von $\mathcal{A}$. Eine Abbildung $\tau : \Omega \to \mathbb{N} \cup \{+\infty\} =: \bar{\mathbb{N}}$ heißt *Stoppzeit*, falls $\{\tau \leq n\} \in \mathcal{A}_n$ für alle $n \in \mathbb{N}$ gilt.

**2.6 Bemerkungen**

(a) $\tau : \Omega \to \bar{\mathbb{N}}$ ist genau dann eine Stoppzeit, wenn $\{\tau = n\} \in \mathcal{A}_n$ für alle $n \in \mathbb{N}$ gilt. Denn

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\} \in \mathcal{A}_n\,,$$

falls $\tau$ eine Stoppzeit ist. Außerdem ist $\{\tau \leq n\} = \bigcup_{m:m\leq n}\{\tau = m\} \in \mathcal{A}_n$, falls alle $\{\tau = n\} \in \mathcal{A}_n$, $n \in \mathbb{N}$.

(b) Ist $\tau$ eine Stoppzeit, so gilt

$$\{\tau = \infty\} = \Big(\bigcup_{n\in\mathbb{N}}\{\tau \leq n\}\Big)^c \in \mathcal{A}_\infty := \sigma\Big(\bigcup_{n\in\mathbb{N}} \mathcal{A}_n\Big)\,.$$

**2.7 Beispiele**

(a) Konstante Abbildungen $\tau : \Omega \to \bar{\mathbb{N}}$ sind Stoppzeiten.

(b) $(X_n)_n$ sei eine Folge von Zufallsvariablen mit Werten in einem Messraum $(E, \mathcal{E})$, $\mathcal{A}_n := \sigma(X_j, j \leq n)$, $n \in \mathbb{N}$, und $A \in \mathcal{E}$. $\tau_A : \Omega \to \bar{\mathbb{N}}$ sei definiert durch

$$\tau_A(\omega) := \inf\big\{n \in \mathbb{N} : X_n(\omega) \in A\big\}\,.$$

Setze $\inf \emptyset = \infty$. Dann ist $\tau_A$ eine Stoppzeit, denn

$$\{\tau_A \leq n\} = \bigcup_{m=1}^{n} \{X_m \in A\} \in \mathcal{A}_n\,, \quad n \in \mathbb{N}\,.$$

Sei weiter $\tau$ eine Stoppzeit, so setzen wir

$$\mathcal{A}_\tau := \Big\{A \in \mathcal{A}_\infty = \sigma\big(\bigcup_{n\in\mathbb{N}} \mathcal{A}_n\big) | A \cap \{\tau \leq n\} \in \mathcal{A}_n \text{ für alle } n \in \mathbb{N}\Big\}\,.$$

Ein Ereignis liegt in $\mathcal{A}_\tau$, wenn es durch den Ablauf des „Prozesses" bis zum Zeitpunkt $\tau$ festgelegt ist.

**2.8 Lemma** *$\mathcal{A}_\tau$ ist eine $\sigma$-Algebra.*

*Beweis:* $\Omega \in \mathcal{A}_\tau$, da $\tau$ eine Stoppzeit ist. Ist $A \in \mathcal{A}_\tau$, so gilt für jedes $n \in \mathbb{N}$

$$A^c \cap \{\tau \leq n\} = \{\tau \leq n\} \backslash \big(A \cap \{\tau \leq n\}\big) \in \mathcal{A}_n\,.$$

Sind $A_k \in \mathcal{A}_\tau$ für jedes $k \in \mathbb{N}$, so folgt für jedes $n \in \mathbb{N}$

$$\Big(\bigcup_{k\in\mathbb{N}} A_k\Big) \cap \{\tau \leq n\} = \bigcup_{k\in\mathbb{N}} \big(A_k \cap \{\tau \leq n\}\big) \in \mathcal{A}_n\,. \qquad \square$$

**2.9 Lemma** *Ist $\tau \equiv n$ für ein $n \in \mathbb{N}$, so ist $\mathcal{A}_\tau = \mathcal{A}_n$.*

*Beweis:* Übung. $\square$

**2.10 Lemma** *Es seien $\tau$ und $\sigma$ zwei Stoppzeiten.*

(a) *Das Minimum $\tau \wedge \sigma := \min\{\tau, \sigma\}$ und das Maximum $\tau \vee \sigma := \max\{\tau, \sigma\}$ sind Stoppzeiten.*

(b) *Gilt $\sigma(\omega) \leq \tau(\omega)$ für alle $\omega \in \Omega$, so gilt $\mathcal{A}_\sigma \subset \mathcal{A}_\tau$.*

(c) *Die Stoppzeit $\tau$ ist $\mathcal{A}_\tau$-messbar.*

(d) *$\{\sigma \leq \tau\} \in \mathcal{A}_\sigma \cap \mathcal{A}_\tau$ und allgemeiner ist für alle $A \in \mathcal{A}_\sigma$ das Ereignis $A \cap \{\sigma \leq \tau\}$ in $\mathcal{A}_\sigma \cap \mathcal{A}_\tau$.*

(e) *Ist $\tau' : \Omega \to \bar{\mathbb{N}}$ eine $\mathcal{A}_\tau$-messbare Abbildung und gilt $\tau' \geq \tau$, so ist $\tau'$ eine Stoppzeit.*

(f) *$\tau + \sigma$ ist eine Stoppzeit.*

*Beweis:*

(a) $n \in \mathbb{N}$: $\{\tau \vee \sigma \leq n\} = \{\tau \leq n\} \cap \{\sigma \leq n\}$ und $\{\tau \wedge \sigma \leq n\} = \{\tau \leq n\} \cup \{\sigma \leq n\}$.

(b) Sei $A \in \mathcal{A}_\sigma$. Mit $\{\tau \leq n\} \subset \{\sigma \leq n\}$ und $A \cap \{\sigma \leq n\} \in \mathcal{A}_n$ folgt $A \cap \{\tau \leq n\} = \big(A \cap \{\sigma \leq n\}\big) \cap \{\tau \leq n\} \in \mathcal{A}_n$ für jedes $n \in \mathbb{N}$.

(c) Für $n, m \in \mathbb{N}$ gilt

$$\{\tau \leq n\} \cap \{\tau \leq m\} = \{\tau \leq n \wedge m\} \in \mathcal{A}_{n \wedge m} \subset \mathcal{A}_m \,.$$

Dies gilt für alle $m$, also $\{\tau \leq n\} \in \mathcal{A}_\tau$ für alle $n$. Damit ist $\tau$ $\mathcal{A}_\tau$-messbar.

(d) Wir zeigen: $A \cap \{\sigma \leq \tau\} \in \mathcal{A}_{\sigma \wedge \tau}$ für alle $A \in \mathcal{A}_\sigma$. Mit (a) und (b) ist $\mathcal{A}_{\sigma \wedge \tau} \subset \mathcal{A}_\sigma \cap \mathcal{A}_\tau$. Für $n \in \mathbb{N}$ ist

$$\begin{aligned} A \cap \{\sigma \leq \tau\} \cap \{\sigma \wedge \tau \leq n\} &= A \cap \{\sigma \leq \tau\} \cap \{\sigma \leq n\} \\ &= A \cap \{\sigma \leq n\} \cap \bigcup_{m=1}^{n} \{\sigma \leq m\} \cap \{\tau \geq m\} \in \mathcal{A}_n \,, \end{aligned}$$

denn $A \cap \{\sigma \leq n\} \in \mathcal{A}_n$ und $\{\sigma \leq m\} \cap \{\tau \geq m\} \in \mathcal{A}_m \subset \mathcal{A}_n$. Also $A \cap \{\sigma \leq \tau\} \in \mathcal{A}_{\sigma \wedge \tau}$.

(e) Für $n \in \mathbb{N}$ gilt $\{\tau' \leq n\} = \{\tau' \leq n\} \cap \{\tau \leq n\} \in \mathcal{A}_n$, da $\{\tau' \leq n\}$ nach Voraussetzung in $\mathcal{A}_\tau$ ist.

(f)

$$\{\tau + \sigma \leq n\} = \bigcup_{\substack{k,l \in \mathbb{N}: \\ k+l \leq n}} \{\tau \leq k\} \cap \{\sigma \leq l\} \in \mathcal{A}_n.$$

□

Sei $(\mathcal{A}_n)_{n \in \mathbb{N}_0}$ eine Filtrierung von $\mathcal{A}$.

**2.11 Definition**

(a) Eine Folge $\mathbb{V} := (V_n)_{n \in \mathbb{N}}$ von Zufallsgrößen heißt *vorhersehbar* bzgl. $(\mathcal{A}_n)_{n \in \mathbb{N}_0}$, wenn $V_n$ für jedes $n \in \mathbb{N}$ bezüglich $\mathcal{A}_{n-1}$ messbar ist.

(b) Sind $\mathbb{V}$ und $\mathbb{X} := (X_n)_{n \in \mathbb{N}_0}$ zwei Folgen von Zufallsgrößen, so definieren wir eine neue Folge $\mathbb{V} \bullet \mathbb{X} = (Y_n)_{n \in \mathbb{N}_0}$ durch $Y_0 = 0$ und $Y_n := \sum_{k=1}^n V_k(X_k - X_{k-1})$ für alle $n \in \mathbb{N}$. $\mathbb{V} \bullet \mathbb{X}$ heißt *Martingaltransformation* (von $\mathbb{X}$ unter $\mathbb{V}$), wenn $\mathbb{X}$ ein Martingal ist.

**2.12 Bemerkung** Ist $\mathbb{X}$ ein Martingal, so werden die Zuwächse (Martingaldifferenzen) $X_k - X_{k-1}$ durch $V_k$ gewichtet. In einer Spielsituation gibt $X_k - X_{k-1}$ den Ausgang der $k$-ten Spielrunde an und $V_k$ den Einsatz des Spielers, den er vorher festzulegen hat („vorhersehbar“).

**2.13 Lemma**

(a) *Sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal und $\mathbb{V} = (V_n)_{n\in\mathbb{N}}$ ein vorhersehbarer „Prozess" mit $V_n \geq 0$ und $\|V_n\|_\infty < \infty$ für alle $n \in \mathbb{N}$. Dann ist $\mathbb{V} \bullet \mathbb{X}$ ein Supermartingal. Ist $\mathbb{X}$ ein Martingal, so auch $\mathbb{V} \bullet \mathbb{X}$.*

(b) *In (a) kann die Voraussetzung $\|V_n\|_\infty < \infty$ ersetzt werden durch $\|V_n\|_2 < \infty$, falls $\|X_n\|_2 < \infty$ für alle $n \in \mathbb{N}_0$ gilt.*

*Beweis:* Sei $\mathbb{V} \bullet \mathbb{X} = (Y_n)_{n\in\mathbb{N}_0}$. Für $n \in \mathbb{N}$ folgt unter den Voraussetzungen in (a) oder (b) die Integrierbarkeit von $Y_n$. Weiter gilt P-f.s.

$$\mathbb{E}(Y_n|\mathcal{A}_{n-1}) = Y_{n-1} + V_n\mathbb{E}(X_n - X_{n-1}|\mathcal{A}_{n-1}) \leq Y_{n-1}\,.$$

Die letzte Ungleichung ist eine Gleichung, falls $\mathbb{X}$ ein Martingal ist. □

Wir untersuchen den wichtigen Spezialfall von Martingalstransformationen mittels einer Stoppzeit $\tau$:

Sei $V_n^\tau := 1_{\{n\leq\tau\}}$, $n \in \mathbb{N}$. Da $\{n \leq \tau\} = \{\tau \leq n-1\}^c \in \mathcal{A}_{n-1}$, ist $\mathbb{V}^\tau = (V_n^\tau)_{n\in\mathbb{N}}$ vorhersehbar. Für jede Folge $(X_n)_{n\in\mathbb{N}_0}$ folgt dann

$$\begin{aligned}(\mathbb{V}^\tau \bullet \mathbb{X})_n &= \sum_{k=1}^{n} 1_{\{\tau>k-1\}}(X_k - X_{k-1})\\ &= \sum_{k=1}^{\tau\wedge n}(X_k - X_{k-1}) = X_{\tau\wedge n} - X_0\,.\end{aligned}$$

Wir definieren den zur Zeit $\tau$ *gestoppten Prozess* $\mathbb{X}^\tau$ durch $\mathbb{X}^\tau = (X_{\tau\wedge n})_{n\in\mathbb{N}}$.

**2.14 Satz (Optional-Sampling-Theorem)** *Es seien $\mathbb{X}$ ein Supermartingal und $\tau$ eine Stoppzeit. Dann ist $\mathbb{X}^\tau$ ein Supermartingal. Ist $\mathbb{X}$ ein Martingal, so auch $\mathbb{X}^\tau$.*

Es folgt insbesondere $\mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X_0)$ für alle $n \in \mathbb{N}_0$, wenn $\mathbb{X}$ ein Martingal und $\tau$ eine Stoppzeit ist. Aber daraus folgt im Allgemeinen nicht $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$, siehe etwa Beispiel 2.3 (g) oder:

**2.15 Beispiel** $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ sei die symmetrische Irrfahrt auf $\mathbb{N}_0$ mit Start in 1 und Absorbtion in 0. Nach Beispiel 2.3 (b) ist $\mathbb{X}$ ein $(\mathcal{A}_n^{\mathbb{X}})_{n\in\mathbb{N}_0}$-Martingal. Sei $\tau := \inf\{n \in \mathbb{N}_0 | X_n = 0\}$. Wegen der Rekurrenz der eindimensionalen Irrfahrt gilt $P(\tau < \infty) = 1$. Für jedes $n \in \mathbb{N}$ gilt

$$\{\tau \leq n\} = \bigcup_{k=0}^{n}\{X_k = 0\} \in \mathcal{A}_n^{\mathbb{X}}\,,$$

also ist $\tau$ eine $(\mathcal{A}_n^{\mathbb{X}})_{n\in\mathbb{N}_0}$-Stoppzeit. Es gilt $\mathbb{E}(X_0) = 1$, also $\mathbb{E}(X_{\tau\wedge n}) = 1$ für alle $n \in \mathbb{N}_0$ mittels Satz 2.14, aber offensichtlich ist $\mathbb{E}(X_\tau) = 0$.

Es gibt hinreichende Kriterien für $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$ („du kannst in einem fairen Spiel keinen Gewinn machen“):

**2.16 Satz** *Es seien $\tau$ eine Stoppzeit und $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal (bzw. ein Martingal), die eine der folgenden Bedingungen erfüllen:*

(a) *$\tau$ ist beschränkt,*

(b) *$\tau$ ist P-f.s. endlich und $(X_n)_{n\in\mathbb{N}_0}$ ist beschränkt:*

$$P(\tau < \infty) = 1\,, \quad \sup_{n\in\mathbb{N}_0} \|X_n\|_\infty < \infty\,,$$

(c) $\mathbb{E}(\tau) < \infty$ *und* $\sup_{n\in\mathbb{N}} \|X_n - X_{n-1}\|_\infty < \infty$.

*Dann ist $X_\tau$ integrierbar und es gilt*

$$\mathbb{E}(X_\tau) \leq \mathbb{E}(X_0) \quad (\text{bzw. } \mathbb{E}(X_\tau) = \mathbb{E}(X_0))\,.$$

*Beweis:* Nach Satz 2.14 gilt $\mathbb{E}(X_{\tau\wedge n}) \leq \mathbb{E}(X_0)$ (bzw. $\mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X_0)$) für jedes $n \in \mathbb{N}$. Ist (a) erfüllt, so ist $\tau \wedge n = \tau$ für ein genügend großes $n \in \mathbb{N}$. Unter (b) gilt $\lim_{n\to\infty} \mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X_\tau)$ nach dem Satz von der majorisierten Konvergenz. Ist (c) erfüllt, so existiert ein $K > 0$ mit $P(|X_n - X_{n-1}| \leq K) = 1$ für alle $n \in \mathbb{N}$. Also $|X_{\tau\wedge n} - X_0| \leq K\tau$ fast sicher. Wegen $\mathbb{E}(\tau) < \infty$ folgt $\lim_{n\to\infty} \mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X_\tau)$ ebenfalls nach dem Satz von der majorisierten Konvergenz. □

**2.17 Korollar** *Seien $(X_n)_{n\in\mathbb{N}_0}$ ein positives Supermartingal und $\tau$ eine endliche Stoppzeit. Dann gilt $\mathbb{E}(X_\tau) \leq \mathbb{E}(X_0)$.*

*Beweis:* Nach Satz 2.14 gilt $\mathbb{E}(X_{\tau\wedge n}) \leq \mathbb{E}(X_0)$ und mittels des Lemmas von Fatou folgt

$$\mathbb{E}(X_\tau) = \mathbb{E}(\lim_{n\to\infty} X_{\tau\wedge n}) \leq \lim_{n\to\infty} \mathbb{E}(X_{\tau\wedge n}) \leq \mathbb{E}(X_0)\,.$$ □

Die betrachteten Stop- und Transformationstechniken werden wir auch zum Beweis des sogenannten „upcrossing“-Lemmas von Doob verwenden.

Seien $\alpha = (\alpha_n)_{n\in\mathbb{N}_0}$ eine Folge reeller Zahlen und $a, b \in \mathbb{R}$ mit $a < b$. Für $n \in \mathbb{N}_0$ ist die Anzahl der *aufsteigenden Überschreitungen* des Intervalls $[a, b]$ durch die Folge $\alpha$ bis zum Zeitpunkt $n$ definiert durch

$$U_n[a,b](\alpha) := \sup\big\{k \in \mathbb{N} \,|\, \text{es ex. } 0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq n \\ \text{mit } \alpha_{s_i} < a,\, \alpha_{t_i} > b \text{ für alle } i \in \{1,\ldots,k\}\big\}\,,$$

wobei $\sup\emptyset := 0$ gesetzt wird.

Natürlich ist $U_n[a,b](\alpha) \leq n$. Wir setzen

$$U_\infty[a,b](\alpha) := \sup_{n\in\mathbb{N}_0} U_n[a,b](\alpha)\,.$$

**2.18 Lemma** *Eine Folge $\alpha = (\alpha_n)_{n\in\mathbb{N}_0}$ konvergiert genau dann in $\bar{\mathbb{R}}$, wenn $U_\infty[a,b](\alpha) < \infty$ für alle $a, b \in \mathbb{Q}$ mit $a < b$ gilt.*

*Beweis:* Wir argumentieren indirekt:

Es gilt:

$$\begin{aligned} & \liminf_{n\to\infty} \alpha_n < \limsup_{n\to\infty} \alpha_n \\ \Leftrightarrow\quad & \exists a, b \in \mathbb{Q} \text{ mit } \liminf_{n\to\infty} \alpha_n < a < b < \limsup_{n\to\infty} \alpha_n \\ \Leftrightarrow\quad & \exists a, b \in \mathbb{Q} \text{ mit } a < b \text{ und } U_\infty[a,b](\alpha) = \infty\,. \end{aligned} \qquad \square$$

**2.19 Satz (Ungleichung von Doob)** *Es sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal und $a < b$. Dann gilt für jedes $n \in \mathbb{N}_0$:*

$$\mathbb{E}\big(U_n[a,b](\mathbb{X})\big) \leq \frac{1}{b-a}\mathbb{E}\big((X_n - a)^-\big)\,,$$

*wobei $x^- := \max\{0, -x\}$ für alle $x \in \mathbb{R}$.*

Bevor wir den Beweis geben, ziehen wir Konsequenzen:

**2.20 Satz (Doobscher Konvergenzsatz)** *Sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal mit $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$. Dann existiert $X_\infty := \lim_{n\to\infty} X_n$ $P$-f.s. und ist integrierbar.*

*Beweis von Satz 2.20:* Seien $a < b$ zwei reelle Zahlen. Nach dem Satz von der monotonen Konvergenz gilt

$$(b-a)\mathbb{E}\big(U_\infty[a,b](\mathbb{X})\big) = (b-a)\lim_{n\to\infty} \mathbb{E}\big(U_n[a,b](\mathbb{X})\big)\,.$$

Die rechte Seite kann mit 2.19 nach oben durch

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}\big((X_n - a)^-\big) \leq |a| + \sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$$

abgeschätzt werden. Also ist $P\big(U_\infty[a,b](\mathbb{X}) < \infty\big) = 1$, also

$$P\Big(\bigcap_{\substack{a,b\in\mathbb{Q}\\ a<b}} \big\{U_\infty[a,b](\mathbb{X}) < \infty\big\}\Big) = 1\,.$$

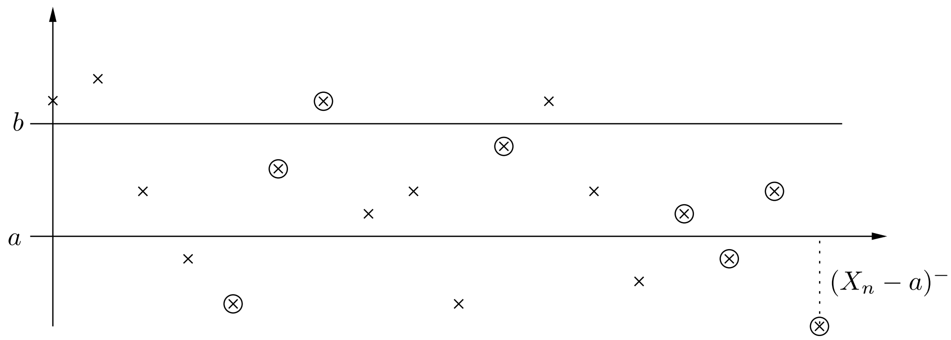



Abb. 2.1: Pfad $\mathbb{N}_0 \ni n \mapsto X_n(\omega)$ eines Supermartingals.

Somit existiert nach Lemma 2.18 $X_\infty = \lim_{n\to\infty} X_n$ in $\bar{\mathbb{R}}$ f.s. Nach dem Lemma von Fatou ist

$$\mathbb{E}|X_\infty| \leq \liminf_{n\to\infty} \mathbb{E}|X_n| \leq \sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty\,. \qquad \square$$

**2.21 Korollar** *Jedes nicht-negative Supermartingal $(X_n)_{n\in\mathbb{N}_0}$ konvergiert f.s. gegen eine integrierbare Zufallsgröße.*

*Beweis:* Aus $X_n \geq 0$ folgt $\mathbb{E}|X_n| = \mathbb{E}(X_n) \leq \mathbb{E}(X_0)$, also $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$. $\square$

**2.22 Bemerkung** Die Bedingung $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$ reicht im Allgemeinen nicht für $\mathcal{L}^1$-Konvergenz aus. Siehe dazu Beispiel 2.15: Dort gilt $\mathbb{E}X_n = 1$ für alle $n \in \mathbb{N}$, $X_n \to 0$ f.s. (siehe Korollar), aber die Erwartungswerte konvergieren nicht, also gilt nicht $\mathcal{L}^1$-Konvergenz.

*Beweis von 2.19:* Wir konstruieren eine Martingaltransformation $\mathbb{Y} = (Y_n)_{n\in\mathbb{N}_0}$ von $\mathbb{X}$. Mit Hilfe der nachfolgenden Regeln nutzt dabei $\mathbb{Y}$ alle aufsteigenden Überschreitungen des Supermartingals $\mathbb{X}$, um möglichst weit nach oben zu gelangen:

(a) Starte mit $Y_0(\omega) = 0$. Ist $X_0 \geq a$, benutze Regel (b), andernfalls Regel (c).

(b) Warte so lange, das heißt setze $Y_n(\omega) = Y_{n-1}(\omega)$, bis $X_n(\omega) < a$ ist. Benutze dann für den nächsten Schritt Regel (c).

(c) Nutze die Zuwächse, das heißt setze $Y_n(\omega) = Y_{n-1}(\omega) + X_n(\omega) - X_{n-1}(\omega)$, bis $X_n > b$. Benutze dann für den nächsten Schritt Regel (b).

Da der Prozess $\mathbb{Y}$ jedes Mal mindestens die Höhe $(b-a)$ gewinnt, wenn $\mathbb{X}$ das Intervall $[a,b]$ aufsteigend überschreitet, und $\mathbb{Y}$ seit der letzten Überschreitung höchstens die Höhe $(X_n - a)^-$ verloren haben kann, gilt für jedes

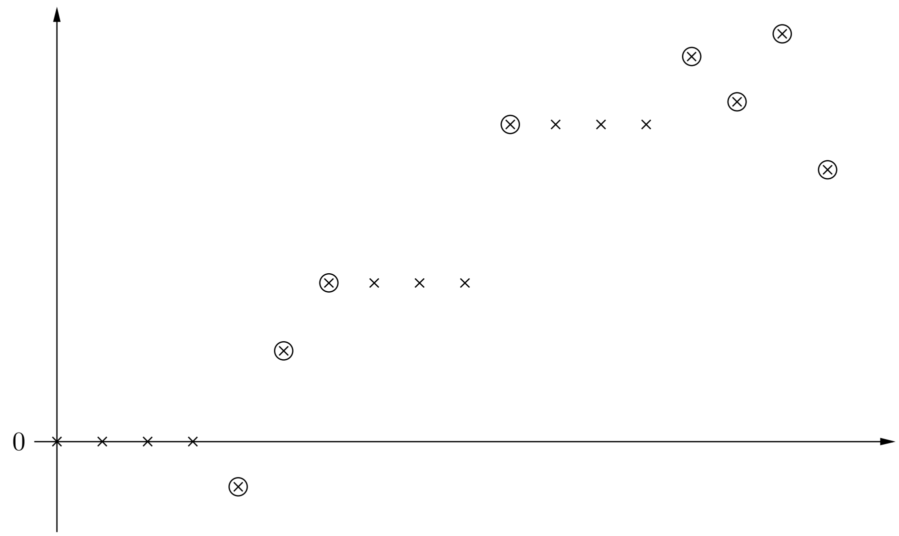



Abb. 2.2: Der zu Abb. 2.1 gehörige Pfad $\mathbb{N}_0 \ni n \mapsto Y_n(\omega)$ der Martingaltransformation.

$n \in \mathbb{N}_0$

$$Y_n \geq (b-a)U_n[a,b](\mathbb{X}) - (X_n - a)^- \,. \tag{2.1}$$

$\mathbb{Y}$ ist eine Martingaltransformation von $\mathbb{X}$:

$$V_n := \begin{cases} 1_{\{X_0 < a\}} & \text{für } n = 1\,, \\ 1_{\{V_{n-1}=1, X_{n-1} \leq b\}} + 1_{\{V_{n-1}=0, X_{n-1} < a\}} & \text{für } n \geq 2\,. \end{cases}$$

Damit ist nach Definition $\mathbb{V} = (V_n)_{n \in \mathbb{N}}$ vorhersehbar und

$$Y_n = \sum_{k=1}^{n} V_k(X_k - X_{k-1})$$

für alle $n \in \mathbb{N}_0$, also ist $\mathbb{Y}$ ein Supermartingal nach Lemma 2.13 (a), also gilt:

$$\mathbb{E}(Y_n) \leq \mathbb{E}(Y_0) = 0\,,$$

und dies liefert mit (2.1) die behauptete Ungleichung. $\square$

**2.23 Beispiel (Polyas Urnenschema)** In einer Urne liegen $R_n$ rote und $S_n$ schwarze Kugeln zum Zeitpunkt $n \in \mathbb{N}_0$. Im Zeitintervall $(n, n+1)$ wird die Urne gut gemischt, eine Kugel zufällig gezogen und zusammen mit einer zusätzlichen Kugel der gleichen Farbe zurückgelegt. Zum Zeitpunkt 0 sei $R_0 = S_0 = 1$. Dann ist $\big\{(R_n, S_n)\big\}_{n \in \mathbb{N}_0}$ eine Markovkette mit Zustandsraum $\mathbb{N}^2$ und Übergangswahrscheinlichkeiten $p\big((r,s),(r+1,s)\big) = r/(r+s)$ und

$p\big((r,s),(r,s+1)\big) = s/(r+s)$, $(r,s) \in \mathbb{N}^2$. $R_n + S_n = n+2$, $n \in \mathbb{N}_0$. Sei $\mathcal{A}_n := \sigma\big((R_i,S_i), i \le n\big)$ und

$$X_n := \frac{R_n}{n+2}\,, \quad n \in \mathbb{N}_0\,.$$

Dann ist $(X_n)_{n\in\mathbb{N}_0}$ ein $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$-Martingal, denn für jedes $n \in \mathbb{N}_0$ gilt:

$$\begin{aligned}\mathbb{E}(X_{n+1}\,|\,\mathcal{A}_n) &= \mathbb{E}\Big(\frac{R_{n+1}}{n+3}\,|(R_n,S_n)\Big)\\ &= \frac{R_n}{n+2}\frac{R_n+1}{n+3} + \frac{(n+2-R_n)}{n+2}\frac{R_n}{n+3}\\ &= \frac{R_n}{n+2} = X_n\,.\end{aligned}$$

Es gilt $X_n \ge 0$, also existiert $X_\infty := \lim_{n\to\infty} X_n$ f.s. nach Korollar 2.21.

Etwas Kombinatorik führt zu $P(R_n = j) = 1/(n+1)$ für jedes $j \in \{1,\dots,n+1\}$. Ist $f : [0,1] \to \mathbb{R}$ stetig, so gilt

$$\mathbb{E}f(X_n) = \frac{1}{n+1}\sum_{j=1}^{n+1} f\Big(\frac{j}{n+2}\Big)\,,$$

was für $n \to \infty$ gegen $\int_0^1 f(x)\,dx$ konvergiert.

Andererseits folgt aus $X_n \to X_\infty$ und dem Satz von Lebesgue von der majorisierten Konvergenz

$$\mathbb{E}f(X_\infty) = \lim_{n\to\infty} \mathbb{E}f(X_n)\,.$$

Also ist $\mathbb{E}f(X_\infty) = \int_0^1 f(x)\,dx$. Dies gilt für jede stetige Funktion, und somit ist die Verteilung von $X_\infty$ das Lebesgue-Maß auf $[0,1]$.

Wir diskutieren nun die $\mathcal{L}^2$-Konvergenz von Martingalen.

**2.24 Definition** Ein Martingal $(X_n)_{n\in\mathbb{N}_0}$ heißt $\mathcal{L}^2$-*Martingal*, wenn für jedes $n \in \mathbb{N}_0$ die Zufallsgröße $X_n$ quadratisch integrierbar ist.

**2.25 Satz** *Sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein $\mathcal{L}^2$-Martingal. Folgende Aussagen sind äquivalent:*

(a) $\sup_{n\in\mathbb{N}_0} \mathbb{E}(X_n^2) < \infty$,

(b) $\sum_{k=1}^\infty \mathbb{E}\big((X_k - X_{k-1})^2\big) < \infty$,

(c) $(X_n)_{n\in\mathbb{N}_0}$ *konvergiert $P$-f.s. und in $\mathcal{L}^2$.*

*Beweis:* Quadratisch integrierbare Martingale haben unkorrelierte Zuwächse: Für alle $m, n \in \mathbb{N}_0$ mit $m < n$ gilt

$$\mathbb{E}\big((X_n - X_m)^2\big) = \sum_{k=m+1}^{n} \mathbb{E}\big((X_k - X_{k-1})^2\big)\,. \tag{2.2}$$

Wir zeigen (2.2) via Induktion nach $n$: Für $n = m + 1$ ist (2.2) klar und für $n \geq m + 2$ gilt

$$\begin{aligned}\mathbb{E}\big((X_n - X_m)^2\big) &= \mathbb{E}\big((X_n - X_{n-1})^2\big) + \mathbb{E}\big((X_{n-1} - X_m)^2\big)\\ &\quad + 2\mathbb{E}\big((X_n - X_{n-1})(X_{n-1} - X_m)\big)\,.\end{aligned}$$

Für den dritten Summanden liefert das Einschieben eines bedingten Erwartungswertes:

$$\begin{aligned}&\mathbb{E}\big((X_n - X_{n-1})(X_{n-1} - X_m)\big)\\ &\qquad = \mathbb{E}\Big(\mathbb{E}\big((X_n - X_{n-1})(X_{n-1} - X_m)|\mathcal{A}_{n-1}\big)\Big)\\ &\qquad = \mathbb{E}\big((X_{n-1} - X_m)\mathbb{E}(X_n - X_{n-1}|\mathcal{A}_{n-1})\big) = 0.\end{aligned}$$

Mit der Induktionsvoraussetzung für $n - 1$ folgt (2.2) für $n$.

Für jedes $n \in \mathbb{N}_0$ folgt aus $\mathbb{E}(X_n\, X_0) = \mathbb{E}\big(\mathbb{E}(X_n|\mathcal{A}_0)X_0\big) = \mathbb{E}(X_0^2)$ und (2.2):

$$\mathbb{E}(X_n^2) - \mathbb{E}(X_0^2) = \mathbb{E}\big((X_n - X_0)^2\big) = \sum_{k=1}^{n} \mathbb{E}\big((X_k - X_{k-1})^2\big)\,,$$

also die Äquivalenz von (a) und (b). (a) folgt aus der $\mathcal{L}^2$-Konvergenz in (c).

Wir zeigen noch, dass (c) aus (a) und (b) folgt: Aus (a) folgt $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$ und somit mit Satz 2.20, dass $X_\infty := \lim_{n\to\infty} X_n$ P-f.s. existiert. Aus dem Lemma von Fatou und (2.2) folgt

$$\begin{aligned}\mathbb{E}\big((X_\infty - X_m)^2\big) &\leq \liminf_{n\to\infty} \mathbb{E}\big((X_n - X_m)^2\big)\\ &= \sum_{k=m+1}^{\infty} \mathbb{E}\big((X_k - X_{k-1})^2\big)\,,\end{aligned}$$

was nach (b) für $m \to \infty$ gegen Null konvergiert. □

**2.26 Bemerkung** Aus der $\mathcal{L}^2$-Beschränktheit eines Martingals folgt Konvergenz in $\mathcal{L}^2$, aus der $\mathcal{L}^1$-Beschränktheit folgt P-fast sichere Konvergenz, aber nicht $\mathcal{L}^1$-Konvergenz, siehe Beispiel 2.15.

Wir diskutieren nun, unter welchen Zusatzbedingungen neben $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$ ein Martingal in $\mathcal{L}^1$ konvergiert.

**2.27 Definition** Eine Teilmenge $\Gamma$ von $\mathcal{L}^1 := \mathcal{L}^1(\Omega, \mathcal{A}, P)$ heißt *gleichgradig integrierbar*, falls

$$\lim_{n\to\infty} \sup_{X\in\Gamma} \int_{\{|X|\geq n\}} |X|\, dP = 0$$

gilt.

*Notation:* Wir schreiben $\mathbb{E}(X; A)$ für $\int_A X\, dP$ also etwa

$$\mathbb{E}\big(|X|; |X| \geq n\big) \quad \text{für} \quad \int_{\{|X|\geq n\}} |X|\, dP\,.$$

**2.28 Lemma** *Es sei $X \in \mathcal{L}^1$. Dann gilt*

$$\lim_{\varepsilon\searrow 0} \sup\big\{\mathbb{E}(|X|; A) | A \in \mathcal{A}\,,\ P(A) \leq \varepsilon\big\} = 0\,.$$

*Beweis:* Angenommen, es existiert eine Folge $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ mit $P(A_n) \leq 2^{-n}$ für alle $n \in \mathbb{N}$ und $\limsup_{n\to\infty} \mathbb{E}(|X|; A_n) > 0$. Für jedes $n \in \mathbb{N}$ sei $B_n := \bigcup_{k=n}^{\infty} A_k$. Dann gilt $1_{B_n} \downarrow \emptyset$ P-f.s. für $n \to \infty$ und mit dem Satz von der majorisierten Konvergenz folgt

$$\limsup_{n\to\infty} \mathbb{E}(|X|; A_n) \leq \lim_{n\to\infty} \mathbb{E}(|X|; B_n) = 0\,.$$

Dies ist ein Widerspruch. $\square$

**2.29 Lemma** *Jede der folgenden Bedingungen ist hinreichend für die gleichgradige Integrierbarkeit einer Familie $\Gamma \subset \mathcal{L}^1$:*

(a) *Es existiert $p \in (1, \infty)$ mit $\sup_{X\in\Gamma} \|X\|_p < \infty$ .*

(b) *Es existiert eine Zufallsgröße $Y \in \mathcal{L}^1$ mit $|X| \leq Y$ f.s. für alle $X \in \Gamma$.*

(c) *Es existiert $Y \in \mathcal{L}^1$ und eine Familie $\Phi$ von Teil-$\sigma$-Algebren von $\mathcal{A}$ mit*

$$\Gamma = \{\mathbb{E}(Y|\mathcal{G}) : \mathcal{G} \in \Phi\}\,.$$

*Beweis:*

(a) Für alle $n \in \mathbb{N}$ und $X \in \Gamma$ gilt

$$\begin{aligned}\mathbb{E}(|X|; |X| \geq n) &\leq \mathbb{E}\Big(|X|\frac{|X|^{p-1}}{n^{p-1}}; |X| \geq n\Big)\\ &\leq \frac{\|X\|_p^p}{n^{p-1}} \leq \frac{1}{n^{p-1}} \sup_{X\in\Gamma} \|X\|_p^p\end{aligned}$$

und dies konvergiert gegen Null für $n \to \infty$.

(b) Für alle $X \in \Gamma$ und $n \in \mathbb{N}$ gilt

$$\mathbb{E}(|X|; |X| \geq n) \leq \mathbb{E}(Y; Y \geq n),$$

und dies konvergiert nach dem Satz von der majorisierten Konvergenz gegen Null für $n \to \infty$.

(c) Für $n \in \mathbb{N}$ und $\mathcal{G} \in \Phi$ sei

$$A_n(\mathcal{G}) := \big\{ |\mathbb{E}(Y|\mathcal{G})| \geq n \big\}.$$

Es gilt $|\mathbb{E}(Y|\mathcal{G})| \leq \mathbb{E}(|Y|\,|\,\mathcal{G})$. Somit liefert die MARKOVsche Ungleichung

$$P\big(A_n(\mathcal{G})\big) \leq \frac{1}{n}\mathbb{E}\big(|\mathbb{E}(Y|\mathcal{G})|\big) \leq \frac{1}{n}\mathbb{E}\big(\mathbb{E}(|Y|\,|\,\mathcal{G})\big) = \frac{1}{n}\mathbb{E}(|Y|).$$

Also gilt $\lim_{n\to\infty} \sup_{\mathcal{G}\in\Phi} P\big(A_n(\mathcal{G})\big) = 0$. Nun ist $A_n(\mathcal{G}) \in \mathcal{G}$, also

$$\begin{aligned} \mathbb{E}\big(|\mathbb{E}(Y|\mathcal{G})|; A_n(\mathcal{G})\big) &\leq \mathbb{E}\big(\mathbb{E}(|Y|\,|\,\mathcal{G}); A_n(\mathcal{G})\big) \\ &= \mathbb{E}\big(|Y|; A_n(\mathcal{G})\big). \end{aligned}$$

Somit liefert Lemma 2.28 die gleichgradige Integrierbarkeit von $\Gamma$. □

**2.30 Satz** *Es seien $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}^1$ und $X \in \mathcal{L}^1$. Die Folge $(X_n)_{n\in\mathbb{N}}$ konvergiert genau dann in $\mathcal{L}^1$ gegen $X$, wenn $(X_n)_{n\in\mathbb{N}}$ in W.keit gegen $X$ konvergiert und $(X_n)_n$ gleichgradig integrierbar ist.*

*Beweis:*

(a) $(X_n)_n$ konvergiere in $\mathcal{L}^1$ gegen $X$. Dann konvergiert $(X_n)_n$ in W.keit gegen $X$ (Satz 4.10 in W-Theorie). Für $k, n \in \mathbb{N}$ ist

$$P(|X_n| \geq k) \leq \frac{\|X_n\|_1}{k}.$$

Konvergiert $(X_n)_n$ im ersten Mittel, so ist $\sup_{n\in\mathbb{N}} \|X_n\|_1 < \infty$. Damit folgt

$$\lim_{k\to\infty} \sup_{n\in\mathbb{N}} P(|X_n| \geq k) = 0. \tag{2.3}$$

Ist $N \in \mathbb{N}$, so gilt

$$\begin{aligned} \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|; |X_n| \geq k) &\leq \sup_{n\leq N} \mathbb{E}(|X_n|; |X_n| \geq k) \\ &\quad + \sup_{n\geq N} \|X_n - X\|_1 + \sup_{n\geq N} \mathbb{E}(|X|; |X_n| \geq k). \end{aligned}$$

Mit (2.3) und Lemma 2.28 konvergiert der dritte Summand für $k \to \infty$ gegen Null. Jede endliche Familie von integrierbaren Zufallsgrößen ist natürlich gleichgradig integrierbar, also konvergiert der erste Summand für $k \to \infty$ gegen Null. Da $N$ beliebig war, folgt die gleichgradige Integrierbarkeit von $(X_n)_n$.

(b) Sei nun $(X_n)_n$ gleichgradig integrierbar und konvergiere in W.keit gegen $X$. Für $k \in \mathbb{N}$ und $x \in \mathbb{R}$ sei

$$\varphi_k(x) := (-k) \vee (x \wedge k)\,.$$

Für $\varepsilon > 0$ und $k, n \in \mathbb{N}$ gilt dann

$$\|\varphi_k(X_n) - \varphi_k(X)\|_1 \le \varepsilon + 2kP(|X_n - X| > \varepsilon)\,.$$

Die Konvergenz in W.keit liefert

$$\lim_{n\to\infty} \|\varphi_k(X_n) - \varphi_k(X)\|_1 = 0\,,$$

da $\varepsilon > 0$ beliebig war, und dies für jedes $k \in \mathbb{N}$. Nun ist

$$\begin{aligned}\|X_n - X\|_1 &\le \|X_n - \varphi_k(X_n)\|_1 + \|\varphi_k(X_n) - \varphi_k(X)\|_1 + \|\varphi_k(X) - X\|_1 \\ &\le \mathbb{E}(|X_n|; |X_n| \ge k) + \|\varphi_k(X_n) - \varphi_k(X)\|_1 + \mathbb{E}(|X|; |X| \ge k)\,,\end{aligned}$$

also

$$\limsup_{n\to\infty} \|X_n - X\|_1 \le \sup_{n\in\mathbb{N}} E(|X_n|; |X_n| \ge k) + \mathbb{E}(|X|; |X| \ge k)\,.$$

Da $k$ beliebig ist, folgt aus der gleichgradigen Integrierbarkeit, dass $\|X_n - X\|_1$ für $n \to \infty$ gegen Null konvergiert. $\square$

Nun zur $\mathcal{L}^1$-Konvergenz von Martingalen:

**2.31 Satz** *Es sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}}$ ein $(\mathcal{A}_n)_n$-Martingal. Dann sind die folgenden Aussagen äquivalent:*

(a) *Das Martingal $(X_n)_n$ ist gleichgradig integrierbar.*

(b) *Das Martingal $(X_n)_n$ konvergiert P-f.s. und in $\mathcal{L}^1$.*

(c) *Es existiert ein $Y \in \mathcal{L}^1$ mit $X_n = \mathbb{E}(Y|\mathcal{A}_n)$ P-f.s. für alle $n \in \mathbb{N}$.*

*Erfüllt $(X_n)_n$ eine dieser Bedingungen, so kann für $Y$ in (c) insbesondere der P-fast sichere und $\mathcal{L}^1$-Limes des Martingals $(X_n)_n$ gewählt werden.*

*Beweis:* (c) $\Rightarrow$ (a) folgt aus Lemma 2.29 (c).

(a) $\Rightarrow$ (b) folgt aus Satz 2.30, wenn wir zeigen, dass $(X_n)_n$ P-f.s. und somit auch in W.keit konvergiert: Es existiert ein $k \in \mathbb{N}$ mit $\mathbb{E}(|X_n|; |X_n| > k) \le 1$ für alle $n \in \mathbb{N}$. Also

$$\begin{aligned}\sup_{n\in\mathbb{N}} \|X_n\|_1 &\le \sup_{n\in\mathbb{N}} \big(\mathbb{E}(|X_n|; |X_n| \le k) + \mathbb{E}(|X_n|; |X_n| > k)\big) \\ &\le k + 1 < \infty\,.\end{aligned}$$

Nun liefert Satz 2.20 die P-fast sichere Konvergenz von $(X_n)_n$.

(b) $\Rightarrow$ (c) Seien $X_\infty$ der $\mathcal{L}^1$-Limes der Folge $(X_n)_n$ und $n \in \mathbb{N}$. Dann gilt für jedes $m \geq n$

$$\begin{aligned}\mathbb{E}\big(|X_n - \mathbb{E}(X_\infty|\mathcal{A}_n)|\big) &= \mathbb{E}\big(|\mathbb{E}(X_m - X_\infty)|\mathcal{A}_n)|\big) \\ &\leq \mathbb{E}\big(\mathbb{E}(|X_m - X_\infty| \,|\, \mathcal{A}_n)\big) \\ &= \|X_m - X_\infty\|_1 \,.\end{aligned}$$

Für $m \to \infty$ folgt $X_n = \mathbb{E}(X_\infty|\mathcal{A}_n)$ P-f.s. Hiermit ist auch die weitere Aussage bewiesen. $\square$

**2.32 Korollar** *Für $Y \in \mathcal{L}^1$ ist $\big(\mathbb{E}(Y|\mathcal{A}_n)\big)_n$ ein $(\mathcal{A}_n)_n$-Martingal, das fast sicher und in $\mathcal{L}^1$ gegen $\mathbb{E}(Y|\mathcal{A}_\infty)$ konvergiert.*

*Beweis:* Für $n \in \mathbb{N}$ sei $X_n := \mathbb{E}(Y|\mathcal{A}_n)$. Dann ist $(X_n)_n$ ein Martingal. Nach Satz 2.31 konvergiert dieses Martingal P-f.s. und in $\mathcal{L}^1$ gegen eine Zufallsgröße $X_\infty$, die $\mathcal{A}_\infty$-messbar ist. Zu zeigen: $X_\infty = \mathbb{E}(Y|\mathcal{A}_\infty)$. Zeige $\mathbb{E}(X_\infty; A) = \mathbb{E}(Y; A)$ für alle $A \in \mathcal{A}_\infty$. Für $A \in \mathcal{A}_n$ folgt dies aus $X_n = \mathbb{E}(X_\infty|\mathcal{A}_n)$ P-f.s., was aus Satz 2.31 folgt. Da $\bigcup_{n\in\mathbb{N}} \mathcal{A}_n$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{A}_\infty$ ist, folgt $\mathbb{E}(X_\infty; A) = \mathbb{E}(Y; A)$ für alle $A \in \mathcal{A}_\infty$. $\square$

Im folgenden Kapitel erleben wir Martingale in Aktion!

KAPITEL 3

# Martingale in Aktion

Wir wollen in diesem Kapitel Anwendungen der Martingaltheorie behandeln. Wir betrachten *Produktmartingale*, den *Satz von* KAKUTANI und Anwendungen in der *asymptotischen Statistik*; wir beweisen den *Satz von* RADON-NIKODYM mittels Martingaltheorie; wir diskutieren KOLMOGOROVS *Kriterium* zu starken Gesetzen der großen Zahlen, untersuchen *verrauschte Beobachtungen* (ein einfaches Modell der sogenannten *Filtertheorie*), betrachten einfache *Verzweigungsprozesse* und ein Modell der *Finanzmathematik*, welches eine diskrete Version der *Black-Scholes-Formel* vorstellt.

## 3.1 Der Satz von Kakutani

Es sei $(\xi_n)_n$ eine Folge von unabhängigen nicht-negativen Zufallsgrößen mit $\mathbb{E}(\xi_n) = 1$. Sei ferner $\mathcal{A}_n := \sigma(\xi_i, i \le n)$, $n \in \mathbb{N}$, und $M_n := \prod_{i=1}^n \xi_i$. Die Folge $(M_n)_n$ ist ein Martingal, denn P-f.s. gilt

$$\begin{aligned}\mathbb{E}(M_{n+1}|\mathcal{A}_n) &= \mathbb{E}(M_n\xi_{n+1}|\mathcal{A}_n)\\ &= M_n\mathbb{E}(\xi_{n+1}|\mathcal{A}_n) = M_n\mathbb{E}(\xi_{n+1}) = M_n\ .\end{aligned}$$

Korollar 2.21 besagt, dass $(M_n)_n$ $P$-f.s. gegen eine nicht-negative Zufallsgröße $M_\infty$ konvergiert. Das Lemma von FATOU liefert $\mathbb{E}(M_\infty) \le \liminf_{n\to\infty} \mathbb{E}(M_n) = 1$. Nach Satz 2.31 konvergiert $(M_n)_n$ genau dann in $\mathcal{L}^1$ gegen $M_\infty$, wenn $(M_n)_n$ gleichgradig integrierbar ist. In diesem Fall ist dann $\mathbb{E}(M_\infty) = 1$.

**3.1 Lemma**

(a) *Das Martingal $(M_n)_n$ ist genau dann gleichgradig integrierbar, wenn*

$$\prod_{n\in\mathbb{N}} \mathbb{E}\big(\sqrt{\xi_n}\big) > 0\ .$$

(b) *Ist $(M_n)_n$ nicht gleichgradig integrierbar, so gilt $M_\infty = 0$ $P$-f.s.*

(c) *Ist $(M_n)_n$ gleichgradig integrierbar und $\xi_n > 0$ $P$-f.s. für alle $n \in \mathbb{N}$, so gilt auch $M_\infty > 0$ $P$-f.s.*

*Beweis:* Für $n \in \mathbb{N}$ seien $a_n := \mathbb{E}\big(\sqrt{\xi_n}\big)$ und $b_n := \prod_{i=1}^n a_i$. Mittels CAUCHY-SCHWARZ folgt $a_n \le 1$, wegen $P(\xi_n = 0) \ne 1$ ist $a_n > 0$.

Für $n \in \mathbb{N}$ sei $N_n := \prod_{i=1}^n \frac{\sqrt{\xi_i}}{a_i}$. Dann ist $(N_n)_n$ ein $(\mathcal{A}_n)_n$-Martingal (s.o.), das nach Korollar 2.21 $P$-f.s. gegen eine nicht-negative Zufallsgröße $N_\infty$ konvergiert. Ist $b_\infty := \prod_{i=1}^\infty a_i > 0$, so folgt

$$K := \sup_{n\in\mathbb{N}} \mathbb{E}(N_n^2) = \sup_{n\in\mathbb{N}} \prod_{i=1}^n \frac{\mathbb{E}(\xi_i)}{a_i^2} = \prod_{i=1}^\infty \frac{1}{a_i^2} = \frac{1}{b_\infty^2} < \infty \ ,$$

nach Satz 2.25 konvergiert $(N_n)_n$ gegen $N_\infty$ in $\mathcal{L}^2$. Es gilt $M_n = b_n^2 N_n^2$, $n \in \mathbb{N}$, also $M_\infty = b_\infty^2 N_\infty^2$ $P$-f.s. Weiter liefert Cauchy-Schwarz

$$\begin{aligned} ||M_n - M_\infty||_1 &= ||(b_n N_n + b_\infty N_\infty)(b_n N_n - b_\infty N_\infty)||_1 \\ &\le ||b_n N_n + b_\infty N_\infty||_2 \, ||b_n N_n - b_\infty N_\infty||_2 \\ &\le 2\sqrt{K}\Big(b_n ||N_n - N_\infty||_2 + |b_n - b_\infty| \, ||N_\infty||_2\Big) \ . \end{aligned}$$

In der letzten Ungleichung haben wir $\|b_n N_n + b_\infty N_\infty\|_2 \le \|b_n N_n\|_2 + \|b_\infty N_\infty\|_2$ sowie

$$\|b_n N_n - b_\infty N_\infty\|_2 \le \|b_n N_n - b_n N_\infty\|_2 + \|b_n N_\infty - b_\infty N_\infty\|_2$$

verwendet. Also konvergiert $(M_n)_n$ gegen $M_\infty$ im ersten Mittel. Nach Satz 2.31 ist $(M_n)_n$ gleichgradig integrierbar. Ist $\prod_{i=1}^\infty a_i = 0$, folgt aus der $P$-f.s. Konvergenz von $(N_n)_n$ gegen $N_\infty$ sofort, dass $(M_n)_n$ $P$-f.s. gegen Null konvergiert. Also gilt $M_\infty = 0$ $P$-f.s. und $(M_n)_n$ konvergiert nicht im ersten Mittel, kann also nach Satz 2.31 nicht gleichgradig integrierbar sein. Somit sind (a) und (b) bewiesen.

(c): $(M_n)_n$ konvergiert $P$-f.s. und in $\mathcal{L}^1$ gegen $M_\infty$. Also ist $\mathbb{E}(M_\infty) = 1$ und $P(M_\infty = 0) \ne 1$. Sei $B_n := \{\prod_{i=n}^\infty \xi_i = 0\}$. Wegen $\prod_{i=1}^{n-1} \xi_i > 0$ $P$-f.s. gilt $P\Big(\{M_\infty = 0\} \Delta B_n\Big) = 0$. Da $(B_n)_n$ eine absteigende Folge ist, gilt

$$\bigcap_{n\in\mathbb{N}} B_n = \limsup_{n\to\infty} B_n \in \bigcap_{k\in\mathbb{N}} \sigma(\xi_j, j \ge k) \ .$$

Nach Kolmogorovs 0-1-Gesetz ist $P(\bigcap_{n\in\mathbb{N}} B_n) \in \{0,1\}$ und

$$P\Big(\{M_\infty = 0\} \Delta \big(\bigcap_{n\in\mathbb{N}} B_n\big)\Big) \le \sum_{n\in\mathbb{N}} P\Big(\{M_\infty = 0\} \Delta B_n\Big) = 0 \ .$$

Aus $P(M_\infty = 0) \ne 1$ folgt somit $P(M_\infty = 0) = 0$. □

Wir betrachten nun die folgende Situation: Es seien $(E, \mathcal{E})$ ein Messraum und $(\mu_n)_n$ und $(\nu_n)_n$ zwei Folgen von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{E})$. Für jedes $n \in \mathbb{N}$ besitze $\nu_n$ eine Dichte $f_n$ bezüglich $\mu_n$ und umgekehrt $\mu_n$ eine Dichte bezüglich $\nu_n$. Wir können ohne Einschränkung annehmen, dass

$f_n(x) > 0$ für alle $x \in E$ gilt. Auf $(\Omega, \mathcal{A}) := (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ seien $P := \otimes_{n\in\mathbb{N}}\mu_n$ und $Q := \otimes_{n\in\mathbb{N}}\nu_n$ die zugehörigen Produktmaße.

Hat $P$ eine Dichte bezüglich $Q$ und umgekehrt $Q$ eine Dichte bezüglich $P$? Kakutanis Satz gibt ein notwendiges und hinreichendes Kriterium.

Für $n \in \mathbb{N}$ seien $\pi_n : \Omega \to E$ die Projektion auf den $n$-ten Faktor und $\mathcal{A}_n = \sigma(\pi_i, i \le n)$. Dann ist $(\mathcal{A}_n)_n$ eine Filtrierung von $\mathcal{A}$.

Sei $\xi_n := f_n \circ \pi_n$. Unter $P$ ist $(\xi_n)_n$ eine Folge unabhängiger Zufallsgrößen mit $\mathbb{E}_P(\xi_n) = 1$ für alle $n \in \mathbb{N}$. Sei $a_n := \mathbb{E}_P\big(\sqrt{\xi_n}\big) = \int_E \sqrt{f_n}\, d\mu_n$.

**3.2 Satz (von Kakutani)**

(a) *Ist $\prod_{n=1}^{\infty} a_n > 0$, so sind $P$ und $Q$ äquivalent, d.h. es existiert eine Dichte von $P$ bzgl. $Q$ und eine Dichte von $Q$ bzgl. $P$. Es gilt*

$$\frac{dQ}{dP} = M_\infty := \lim_{n\to\infty} \prod_{i=1}^{n} \xi_i \quad P\text{-f.s.} \quad \text{und}$$

$$\frac{dP}{dQ} = M_\infty^{-1} \quad P\text{-f.s.}$$

(b) *Ist $\prod_{n=1}^{\infty} a_n = 0$, so sind $P$ und $Q$ zueinander singulär, dass heißt, es existiert ein $A \in \mathcal{A}$ mit $P(A) = 1$ und $Q(A) = 0$.*

*Beweis:*

(a) Für jedes $n \in \mathbb{N}$ seien $M_n := \prod_{i=1}^{n} \xi_i$; $P_n := P|_{\mathcal{A}_n}$, $Q_n := Q|_{\mathcal{A}_n}$. Dann sind $P_n$ und $Q_n$ zueinander äquivalent mit $\frac{dQ_n}{dP_n} = M_n$ $P$-f.s. (zum Beispiel nach Satz 3.7, Wahrscheinlichkeitstheorie). Nach Lemma 3.1 und Satz 2.31 konvergiert $(M_n)_n$ $P$-f.s. in $\mathcal{L}^1(P)$ gegen $M_\infty$. Weiter gilt $M_n = \mathbb{E}_P(M_\infty|\mathcal{A}_n)$ $P$-f.s. für jedes $n \in \mathbb{N}$. Also gilt für $A \in \mathcal{A}_n$

$$Q(A) = Q_n(A) = \int_A M_n\, dP_n = \int_A M_n\, dP = \int_A M_\infty\, dP\ .$$

Dies gilt für alle $A \in \bigcup_{n\in\mathbb{N}} \mathcal{A}_n$, und $\bigcup_{n\in\mathbb{N}} \mathcal{A}_n$ ist ein durchschnittstabiles Erzeugendensystem von $\mathcal{A}$, also gilt die Gleichheit für alle $A \in \mathcal{A}$. Damit ist $\frac{dQ}{dP} = M_\infty$ $P$-f.s. Wegen $M_\infty > 0$ $P$-f.s., siehe Lemma 3.1(c), ist $\frac{dP}{dQ} = M_\infty^{-1}$.

(b) Aus Lemma 3.1 wissen wir, dass $(M_n)_n$ $P$-f.s. gegen Null konvergiert. Wir zeigen nun, dass $(M_n)_n$ $Q$-f.s. gegen unendlich konvergiert. Da $(N_n)_n$ mit $N_n := \frac{\sqrt{M_n}}{\prod_{i=1}^{n} a_i}$ ein $P$-Martingal ist und $a_i \le 1$ für alle $i \in \mathbb{N}$ gilt, ist $(\sqrt{M_n})_n$ ein $P$-Supermartingal: es gilt P-f.s.

$$\mathbb{E}_P\big(\sqrt{M_n}|\mathcal{A}_{n-1}\big) = \Big(\prod_{i=1}^{n} a_i\Big)\, \mathbb{E}_P(N_n|\mathcal{A}_{n-1}) = a_n\sqrt{M_{n-1}} \le \sqrt{M_{n-1}}.$$

Für $B \in \mathcal{A}_n$ gilt

$$\begin{aligned}\int_B \frac{1}{\sqrt{M_n}}\,dQ &= \int_B \frac{1}{\sqrt{M_n}} M_n\,dP = \int_B \sqrt{M_n}\,dP \\ &\geq \int_B \sqrt{M_{n+1}}\,dP = \int_B \frac{1}{\sqrt{M_{n+1}}}\,dQ\end{aligned}$$

($M_n$ und $M_{n+1}$ sind überall strikt positiv gewählt). Also ist $(\frac{1}{\sqrt{M_n}})_n$ ein positives $Q$-Supermartingal und konvergiert gemäß Korollar 2.21 $Q$-f.s. Wegen $\int_\Omega \frac{1}{\sqrt{M_n}}\,dQ = \int_\Omega \sqrt{M_n}\,dP = \prod_{i=1}^n a_i \searrow 0$ für $n \to \infty$ folgt

$$Q\Big(\lim_{n\to\infty} \frac{1}{\sqrt{M_n}} = 0\Big) = 1 \text{ , also } \; Q\Big(\lim_{n\to\infty} M_n = \infty\Big) = 1 \; .$$

Somit ist $A := \{\omega \in \Omega : \lim_{n\to\infty} M_n(\omega) = 0\}$ eine $Q$-Nullmenge mit $P(A) = 1$. □

**3.3 Beispiel** Sei $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$, $\mu_n$ sei $N(0,1)$- und $\nu_n$ $N(\alpha_n, 1)$-verteilt, $\alpha_n \in \mathbb{R}$, $n \in \mathbb{N}$. Dann gilt für jedes $n \in \mathbb{N}$

$$\frac{d\nu_n}{d\mu_n}(x) = \frac{\exp\big(-(x-\alpha_n)^2/2\big)}{\exp\big(-x^2/2\big)} = \exp\big(\alpha_n x - \alpha_n^2/2\big) \; , \quad x \in \mathbb{R} \; .$$

Weiter ist

$$\begin{aligned} a_n &:= \int_{\mathbb{R}} \sqrt{\frac{d\nu_n}{d\mu_n}}\,d\mu_n \\ &= \int_{\mathbb{R}} \exp\Big(\frac{\alpha_n x}{2} - \frac{\alpha_n^2}{4}\Big) \frac{1}{\sqrt{2\pi}} \exp\big(-x^2/2\big)\,dx \\ &= \exp\big(-\alpha_n^2/8\big) \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}\big(x - \frac{\alpha_n}{2}\big)^2\Big)\,dx \\ &= \exp\big(-\alpha_n^2/8\big) \; . \end{aligned}$$

Somit ist $\prod_{n=1}^\infty a_n > 0 \Leftrightarrow \sum_{n=1}^\infty \alpha_n^2 < \infty$. Nach Satz 3.2 sind $P$ und $Q$ genau dann äquivalent, wenn $(\alpha_n)_n \in \ell_2$. Dann ist

$$\frac{dQ}{dP}(x) = \exp\Big(\sum_{n=1}^\infty \alpha_n x_n - \frac{1}{2}\sum_{n=1}^\infty \alpha_n^2\Big)$$

für $P$-fast alle $x = (x_n)_n \in \mathbb{R}^{\mathbb{N}}$. Die Reihe $\sum_{n=1}^\infty \alpha_n x_n$ konvergiert nur $P$-f.s. Außerhalb dieser Menge von $P$- und $Q$-Maß 1 können wir die Dichtefunktion nach Belieben setzen, zum Beispiel 1.

Als Anwendung machen wir einen kleinen Abstecher in die asymptotische Statistik. Wir untersuchen das Verhalten von Folgen von Dichtequotienten. Sind $P$ und $Q$ die obigen Produktmaße auf $(\Omega, \mathcal{A})$, so sind die $(\pi_n)_n$ unabhängige Zufallsgrößen. In der Statistik möchte man auf der Basis von Beobachtungen entscheiden, ob $P$ oder $Q$ vorliegt! Es sei nun $(\rho_n)_n$ eine weitere Folge von W-Maßen auf $(E, \mathcal{E})$ mit: $\mu_n$ besitzt eine strikt positive Dichte $f_n$ bzgl. $\rho_n$ und $\nu_n$ besitzt eine strikt positive Dichte $g_n$ bzgl. $\rho_n$ für jedes $n \in \mathbb{N}$. Dann ist (siehe Notation wie im Beweis von Satz 3.2)

$$\frac{dQ_n}{dP_n} = \prod_{i=1}^{n} \frac{g_i}{f_i} \circ \pi_i =: \prod_{i=1}^{n} \xi_i \quad \text{f.s.}$$

und $\mathbb{E}_P(\xi_i) = \int_E \frac{g_i(x)}{f_i(x)} f_i(x)\, d\rho_i(x) = \int_E g_i(x)\, d\rho_i(x) = \nu_i(E) = 1$, wie bereits gesehen. Weiter ist

$$\mathbb{E}_P(\sqrt{\xi_i}) = \int \frac{\sqrt{g_i(x)}}{\sqrt{f_i(x)}} f_i(x)\, d\rho_i(x) = \int \sqrt{g_i(x) f_i(x)}\, d\rho_i(x) \ ,$$

die Bedingung $\prod_{i=1}^{\infty} a_n > 0$ im Satz von KAKUTANI liest sich hier also zu

$$\prod_{i=1}^{\infty} \int_E \sqrt{g_i(x) f_i(x)}\, d\rho_i(x) > 0 \ , \tag{3.1}$$

was übrigens zu $\sum_{i \geq 1} \int_E \left(f_i(x)^{1/2} - g_i(x)^{1/2}\right)^2 d\rho_i(x) < \infty$ äquivalent ist.

**3.4 Korollar** *$P$ und $Q$ sind, wenn nicht äquivalent, bereits singulär und*

$$Q(M_n \to \infty) = 1 \quad \textit{und} \quad P(M_n \to 0) = 1 \ .$$

*Beweis:* Sind $P$ und $Q$ nicht äquivalent, so existiert keine Dichte von $Q$ bzgl. $P$, also ist das Martingal $(M_n)_n$ nicht gleichgradig integrierbar bzgl. $P$, also $M_n \to 0$ $P$-fast sicher. Analog zu Satz 3.2 ergibt sich $M_n^{-1} \to 0$ $Q$-fast sicher. □

Wir betrachten den Fall $\mu_1 = \mu_2 = \cdots$ und $\nu_1 = \nu_2 = \cdots$; somit sind alle Dichten $f_n$, sowie $g_n$, gleich, in Notation $f$ und $g$. Also ist

$$M_n = \prod_{k=1}^{n} \frac{g(\pi_k)}{f(\pi_k)} \quad , \ n \geq 0 \ .$$

Somit ist $P$ singulär zu $Q$, falls $\mu_1 \neq \nu_1$, siehe (3.1).

Aus der Statistik ist bekannt: Zu $\alpha \in (0,1)$ definiert

$$\varphi_n(x_1, \dots, x_n) := 1_{(k_n, \infty)}\Big(\prod_{k=1}^{n} \frac{g(x_k)}{f(x_k)}\Big)$$

einen „besten“Test zum Niveau $\alpha$ für $\mu_1$ gegen $\nu_1$ bei Beobachtung von $X_1, \ldots, X_n$, wobei $k_n \in [0, \infty)$ so bestimmt wird, dass

$$\mu_1^{\otimes n}\Big(\Big\{(x_1, \ldots, x_n) \in \mathbb{R}^{\mathbb{N}} : \prod_{k=1}^{n} \frac{g(x_k)}{f(x_k)} > k_n\Big\}\Big) = P(M_n > k_n) = \alpha$$

(Fehler 1. Art).

Da mit Korollar 3.4 $P(M_n \to 0) = 1$, folgt $k_n \to 0$. Damit konvergiert der Fehler 2. Art $Q(M_n \leq k_n)$ für $n \to \infty$ gegen 0, denn

$$Q(M_n \leq k_n) = \int_{\{M_n \leq k_n\}} M_n \, dP \leq k_n \to 0 \ .$$

Dies nennt man *Konsistenz* der Testfolge $(\varphi_n)_n$. Analog folgt durch Vertauschen der Rollen von $\mu_1$ und $\nu_1$ die Konsistenz des analogen Testes zum Niveau $\alpha$ für $\nu_1$ gegen $\mu_1$.

Die Folge der Dichtequotienten heißt auch *Likelihood-Prozess*.

## 3.2 Der Satz von Radon-Nikodym

Wir zeigen hier eine einfache Version von Satz 1.20 in der Wahrscheinlichkeitstheorie:

**3.5 Satz** *Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $\mathcal{A}$ sei separabel, das heißt $\mathcal{A} = \sigma(\{A_n, n \geq 1\})$ mit $A_n \subset \Omega$, $n \in \mathbb{N}$. Weiter sei $Q$ ein endliches Maß auf $(\Omega, \mathcal{A})$ und jede $P$-Nullmenge sei eine $Q$-Nullmenge. Dann existiert eine Zufallsvariable $X$ auf $(\Omega, \mathcal{A})$ mit*

$$Q(A) = \int_A X(\omega) \, dP(\omega) \quad \textit{für alle } A \in \mathcal{A} \ .$$

**3.6 Bemerkung** Viele $\sigma$-Algebren sind separabel, zum Beispiel die Borel-$\sigma$-Algebra eines jeden separablen metrischen Raums. Im Buch von Williams kann man lesen, dass diese Bedingung weggelassen werden kann. Der Schritt zu Satz 1.20 in Wahrscheinlichkeitstheorie ist dann eine einfache Übung.

Sei vorbereitend $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{A}_3 \subset \cdots$ eine Folge von Zerlegungen von $\Omega$ in disjunkte messbare Mengen derart, dass $\mathcal{A}_n = \{A_{n1}, \ldots, A_{nk_n}\}$ eine Zerlegung von $\Omega$ ist, bei der jedes Element als Vereinigung von Elementen der Zerlegung $\mathcal{A}_{n+1}$ geschrieben werden kann:

$$A_{ni} = \bigcup_{j \in I(n,i)} A_{n+1,j} \ .$$

Zu $P$, $Q$ und $\mathcal{A}_n$ definieren wir nun

$$M_n := \sum_{i=1}^{k_n} \frac{Q(A_{ni})}{P(A_{ni})} 1_{A_{ni}} \ .$$

Dann ist $(M_n)_n$ unter $P$ ein Martingal bzgl. $(\sigma(\mathcal{A}_n))_n$. $M_n$ ist $\sigma(\mathcal{A}_n)$-messbar. Jedes $G \in \sigma(\mathcal{A}_n)$ ist disjunkte Vereinigung von $A_{ni}$'s. Ohne Einschränkung ist $G = A_{ni}$ für ein $i$. Dann gilt

$$\int_G M_{n+1} \, dP = \sum_{j \in I(n,i)} \frac{Q(A_{n+1,j})}{P(A_{n+1,j})} P(A_{n+1,j}) = Q(A_{ni}) = \int_G M_n \, dP \ .$$

**3.7 Lemma** *Es seien $P$ und $Q$ zwei endliche Maße auf $(\Omega, \mathcal{A})$. Dann ist jede $P$-Nullmenge eine $Q$-Nullmenge genau dann, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit $P(A) \leq \delta$ impliziert $Q(A) \leq \varepsilon$ für jedes $A \in \mathcal{A}$.*

*Beweis:* „$\Leftarrow$“Aus der Bedingung folgt $Q(A) \leq \varepsilon$ für jede $P$-Nullmenge $A \in \mathcal{A}$ und jedes $\varepsilon > 0$. Daher ist $Q(A) = 0$. „$\Rightarrow$“Angenommen, die Bedingung gilt nicht. Dann existiert ein $\varepsilon > 0$ und eine Folge $(A_n)_n$ in $\mathcal{A}$ mit $P(A_n) \leq 2^{-n}$ und $Q(A_n) > \varepsilon$, $n \in \mathbb{N}$. Wähle

$$A := \limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \ ,$$

so ist $A \in \mathcal{A}$ und

$$P(A) \leq P(\bigcup_{m=n}^{\infty} A_m) \leq \sum_{m=n}^{\infty} P(A_m) \leq \sum_{m=n}^{\infty} 2^{-m} = 2^{-n+1} \ ,$$

$n \in \mathbb{N}$, also $P(A) = 0$ und andererseits wegen der Endlichkeit von $Q$

$$Q(A) \geq \limsup_{n \to \infty} Q(A_n) \geq \varepsilon > 0 \ ,$$

im Widerspruch zur Annahme, dass jede $P$-Nullmenge eine $Q$-Nullmenge ist. $\square$

*Beweis von 3.5:* Wir wählen $\mathcal{A}_n := \sigma(A_1, \ldots, A_n)$, $n \in \mathbb{N}$. Diese $\sigma$-Algebren werden erzeugt von den sogenannten Atomen $A_{n1}, \ldots, A_{nk_n}$ (erzeugt bedeutet: alle $2^{k_n}$ möglichen Vereinigungen von den Atomen), wobei jedes Atom der Form $H_1 \cap \cdots \cap H_n$ mit $H_i = A_i$ oder $H_i = A_i^c$, $i = 1, \ldots, n$, ist. Nun wählen wir

$$M_n := \sum_{i=1}^{k_n} \frac{Q(A_{ni})}{P(A_{ni})} 1_{A_{ni}} \ .$$

Für $A \in \mathcal{A}_n$ wähle die Darstellung

$$A = \bigcup_{i\in I} A_{ni} \text{ , } I \text{ endlich , } A_{ni} \text{ Atom in } \mathcal{A}_n.$$

Dann ist

$$Q(A) = \sum_{i\in I} Q(A_{ni}) = \sum_{i\in I} \frac{Q(A_{ni})}{P(A_{ni})} P(A_{ni}) = \sum_{i\in I} \int_{A_{ni}} M_n \, dP = \int_A M_n \, dP \ ,$$

wobei wir beachten, dass im Fall $P(A_{ni}) = 0$ auch $Q(A_{ni}) = 0$ folgt, $\frac{Q(A_{ni})}{P(A_{ni})}$ also beliebig gesetzt werden kann.

$$\text{Also ist } M_n = \frac{dQ|_{\mathcal{A}_n}}{dP|_{\mathcal{A}_n}} \ .$$

$(M_n)_n$ ist ein nicht-negatives Martingal, konvergiert also f.s. gegen ein $M_\infty$. Zu $\varepsilon > 0$ wähle $\delta > 0$ wie in Lemma 3.7 und sei $K \in (0, \infty)$ so, dass $K > \frac{Q(\Omega)}{\delta}$. Dann ist $P(M_n > K) \leq \frac{1}{K}\mathbb{E}(M_n) = \frac{1}{K}Q(\Omega) < \delta$ und somit

$$\int_{\{M_n > K\}} M_n \, dP = Q(M_n > K) \leq \varepsilon \ .$$

Also ist $(M_n)_n$ gleichgradig integrierbar und somit konvergiert $(M_n)_n$ in $\mathcal{L}^1$ gegen $M_\infty$. Also gilt für alle $A \in \mathcal{A}_n$

$$\int_A M_\infty \, dP = \lim_{n\to\infty} \int_A M_n \, dP = Q(A) \ ,$$

und da $\bigcup_n \mathcal{A}_n$ ein durchschnittstabiler Erzeuger von $\mathcal{A}$ ist, folgt die Aussage für alle $A \in \mathcal{A}$, also ist $M_\infty$ die gesuchte Dichte. □

## 3.3 Kolmogorovs Kriterium

Wir erinnern an Satz 6.10 aus der Wahrscheinlichkeitstheorie: Es sei $(X_n)_n$ eine unabhängige Folge von Zufallsgrößen und $(a_n)_n$ eine Zahlenfolge mit $0 < a_n \nearrow \infty$. Dann folgt aus $\sum_{n\geq 1} a_n^{-2} \operatorname{Var}(X_n) < \infty$:

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} (X_k - \mathbb{E}X_k) = 0 \quad P\text{-f.s.}$$

Wir diskutieren dieses Kriterium nun mittels der Martingaltheorie: Ist $(X_n)_n$ fortan eine Folge integrierbarer reeller Zufallsvariablen mit

$$\mathbb{E}(X_{n+1}|X_1, \ldots, X_n) \leq 0 \ , \ n \in \mathbb{N} \ , \ P\text{-f.s.,} \tag{3.2}$$

so ist $S_n := X_1 + \cdots + X_n$ ein Supermartingal, denn zu (3.2) addiere $\mathbb{E}(X_i|X_1,\ldots,X_n) = X_i$, $i = 1,\ldots,n$, und beachte $\sigma(S_1,\ldots,S_n) = \sigma(X_1,\ldots,X_n)$. Sind die $X_i$ auch zentriert, so ist $(S_n)_n$ ein Martingal. Wir hatten dies im Spezialfall einer Folge unabhängiger Zufallsgrößen $(X_n)_n$ zu Beginn von Kapitel 2 gesehen.

Die Bedingung $\sup_{n\in\mathbb{N}} \mathbb{E}|S_n| < \infty$ ist nach Satz 2.20 hinreichend für fast sichere Konvergenz. Im Fall der Unabhängigkeit ist $\sum_{n=1}^\infty \mathrm{Var}(X_n) < \infty$ hinreichend, denn

$$\big(\mathbb{E}|S_n|\big)^2 \leq \mathrm{Var}(S_n) = \sum_{i=1}^n \mathrm{Var}(X_i) \leq \sum_{i=1}^\infty \mathrm{Var}(X_i)\ .$$

Ist $(X_n)_n$ eine Folge integrierbarer, zentrierter, reeller Zufallsvariablen mit

$$\mathbb{E}(X_{n+1}|X_1,\ldots,X_n) = 0 \tag{3.3}$$

und ist $(a_n)_n$ wie in KOLMOGOROVs Kriterium (also $a_n > 0$ und $a_n \nearrow \infty$), so ist

$$\mathbb{E}\bigg(\frac{1}{a_{n+1}}X_{n+1}\Big|\frac{1}{a_1}X_1,\ldots,\frac{1}{a_n}X_n\bigg) = 0 \quad \text{f.s. für alle } n \in \mathbb{N}\ .$$

Folglich ist $\left(\sum_{i=1}^n \frac{1}{a_i}X_i\right)_n$ ein Martingal. Konvergiert dieses f.s. (und eine Bedingung ist $\sup_{n\in\mathbb{N}} \mathbb{E}\Big(\Big|\sum_{i=1}^n \frac{1}{a_i}X_i\Big|\Big) < \infty$ ), so liefert das KRONECKERsche Lemma

$$\lim_{n\to\infty} \frac{1}{a_n}\sum_{i=1}^n X_i = 0 \quad \text{f.s.}$$

(3.3) ist schwächer als die geforderte Unabhängigkeit im Satz 6.10.

### 3.4 Verrauschte Beobachtungen (Filtertheorie)

Es seien $X$ und $(\eta_n)_n$ unabhängige normalverteilte Zufallsgrößen mit $\mathcal{L}(X) = N(0,\sigma^2)$ und $\mathcal{L}(\eta_n) = N(0,1)$, $n \in \mathbb{N}$. Es sei $X$ die Zufallsgröße, an der wir interessiert sind. Angenommen, sie kann nicht direkt beobachtet werden, aber in *verrauschter* Form. Wir nehmen dazu an, dass man die Folge $(Y_n)_n$ mit

$$Y_n := X + c_n\eta_n \quad ,\ n \in \mathbb{N}\ ,$$

beobachten kann, wobei $(c_n)_n$ reelle Zahlen in $(0,\infty)$ sind. Ein *natürlicher Schätzer*, der auf den Beobachtungen $Y_1,\ldots,Y_n$ basieren soll, ist vermutlich $M_n := \mathbb{E}(X|\mathcal{A}_n)$ mit $\mathcal{A}_n := \sigma(Y_i, i \leq n)$. $(M_n)_n$ ist nun ein quadratisch integrierbares Martingal mit

$$\mathbb{E}(M_n^2) \leq \mathbb{E}\big(\mathbb{E}(X^2|\mathcal{A}_n)\big) = \mathbb{E}(X^2) = \sigma^2$$

für alle $n \in \mathbb{N}$. Mit Satz 2.25 und Korollar 2.32 folgt $\lim_{n\to\infty} M_n = M_\infty := \mathbb{E}(X|\mathcal{A}_\infty)$ in $\mathcal{L}^2$ und $P$-f.s., $\mathcal{A}_\infty := \sigma(Y_n, n \in \mathbb{N})$.

Wann ist $M_\infty = X$ $P$-fast sicher?

Dazu werden wir $M_n$ und die Varianzen von $X - M_n$ berechnen. Wir bestimmen dazu $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ so, dass

$$X - \sum_{i=1}^{n} \lambda_i Y_i =: Z$$

und $\mathcal{A}_n$ unabhängig sind. Da alle Zufallsgrößen gemeinsam normalverteilt sind, ist die gewünschte Unabhängigkeit gleichbedeutend mit

$$\operatorname{Cov}(Z, Y_i) = 0 \quad \text{für jedes } i \in \{1, \ldots, n\} \ .$$

Nun ist

$$\operatorname{Cov}(Z, Y_i) = \operatorname{Cov}\Big(X - \sum_{j=1}^{n} \lambda_j (X + c_j \eta_j), X + c_i \eta_i\Big) \ .$$

Eine kleine Nebenrechnung entlang der Definition der Kovarianz ergibt

$$\operatorname{Cov}(Z, Y_i) = \sigma^2 \Big(1 - \sum_{j=1}^{n} \lambda_j\Big) - \lambda_i c_i^2 \ .$$

Wir wollen also das Gleichungssystem

$$\sigma^2 \Big(1 - \sum_{j=1}^{n} \lambda_j\Big) - \lambda_i c_i^2 = 0 \ , \quad i = 1, \ldots, n$$

lösen: Es ergibt sich $\lambda_i = \frac{\sigma^2 c_i^{-2}}{1 + \sigma^2 \sum_{j=1}^{n} c_j^{-2}}$, $i \in \{1, \ldots, n\}$. Nun liefert die Unabhängigkeit von $Z$ und $\mathcal{A}_n$: $\mathbb{E}(Z|\mathcal{A}_n) = \mathbb{E}(Z) = 0$, also

$$M_n = \mathbb{E}(X|\mathcal{A}_n) = \sum_{i=1}^{n} \lambda_i Y_i \quad \text{sowie}$$

$$\mathbb{E}\big((X - M_n)^2\big) = \mathbb{E}Z^2 = \sigma^2 \Big(1 - \sum_{j=1}^{n} \lambda_j\Big)^2 + \sum_{i=1}^{n} \lambda_i^2 c_i^2 = \sigma^2 \Big(1 + \sigma^2 \sum_{i=1}^{n} c_i^{-2}\Big)^{-1} \ .$$

Dies konvergiert genau dann gegen Null für $n \to \infty$, wenn $\sum_{i=1}^{n} c_i^{-2} = \infty$ gilt. Somit erhalten wir

**3.8 Satz** *$(M_n)_n$ konvergiert gegen $X$ $P$-f.s. und in $\mathcal{L}^2$ genau dann, wenn*

$$\sum_{i=1}^{\infty} \frac{1}{c_i^2} = \infty \ .$$

Allgemeine Filterprobleme lesen sich etwa so:

Es seien $\alpha, \beta, \gamma, \rho, \nu$ reelle Konstanten und $X_0$, $(\varepsilon_n)_n$, $(\eta_n)_n$ stochastisch unabhängige Zufallsgrößen mit $X_0 \sim N(\mu_0, \sigma_0^2)$; $\varepsilon_n \sim \eta_n \sim N(0,1)$, $n \in \mathbb{N}$. Der Zustand $X_n$ eines linearen stochastischen Systems zum Zeitpunkt $n \geq 1$ sei durch

$$X_n - X_{n-1} = \alpha X_{n-1} + \beta \varepsilon_n + \gamma$$

beschrieben. Statt $X_n$ selbst kann aber nur eine gestörte (verrauschte) Version $Y_n$, beschrieben durch $Y_n - Y_{n-1} = \rho X_n + \nu \eta_n$, $Y_0 = 0$, beobachtet werden. Der sogenannte Kalman-Bucy-Filter ist ein Verfahren, mit dessen Hilfe der tatsächliche Systemzustand erstaunlich gut geschätzt werden kann. Dies findet Anwendungen bei der Analyse von Zeitreihen oder der Signalverarbeitung (Ortung beweglicher Objekte).

### 3.5 Verzweigungsprozesse

Seien $(X_{n,i})_{n \in \mathbb{N}_0, i \in \mathbb{N}}$ unabhängige und identisch verteilte Zufallsvariablen mit Werten in $\mathbb{N}_0$. Setze $p_k := P(X_{1,1} = k)$, $k \in \mathbb{N}_0$. Seien $m := \mathbb{E}(X_{1,1}) < \infty$ und $\sigma^2 := \mathrm{Var}(X_{1,1}) \in (0, \infty)$. Wir definieren den stochastischen Prozess $Z = (Z_n)_{n \in \mathbb{N}_0}$ durch

$$Z_0 = 1 \ , \quad Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i} \ .$$

Eine mögliche Interpretation ist: $Z_n$ ist die Größe einer Population zur Zeit $n$. Das $i$'te Individuum hat $X_{n,i}$ Nachkommen in der $n+1$'ten Generation. $Z$ heißt *Galton-Watson-Prozess* oder *Verzweigungsprozess* mit Nachkommenverteilung $p = (p_k)_{k \in \mathbb{N}_0}$. Manchmal lernt man im dritten Semester: Ist $q := \lim_{n \to \infty} P(Z_n = 0)$, die *Aussterbewahrscheinlichkeit* , so gilt mit $p_1 \neq 1$:

$$q < 1 \quad \Leftrightarrow \quad \sum_{k=1}^{\infty} k \, p_k = \mathbb{E}(X_{1,1}) > 1$$

(Methode: erzeugende Funktion).

Wir betrachten dieses Modell jetzt im Rahmen der Martingaltheorie: Sei $\mathcal{A}_n := \sigma(X_{k,i}, k < n, i \in \mathbb{N})$. Dann ist $Z = (Z_n)_{n \in \mathbb{N}_0}$ an $(\mathcal{A}_n)_{n \in \mathbb{N}_0}$ adaptiert. Setze

$$W_n := \frac{Z_n}{m^n} \ .$$

**3.9 Lemma** *$W = (W_n)_{n \in \mathbb{N}_0}$ ist ein Martingal.*

*Beweis:*

$$\begin{aligned}
\mathbb{E}(W_{n+1}|\mathcal{A}_n) &= m^{-(n+1)}\mathbb{E}(Z_{n+1}|\mathcal{A}_n) = m^{-(n+1)}\mathbb{E}\Big(\sum_{i=1}^{Z_n} X_{n,i}|\mathcal{A}_n\Big) \\
&= m^{-(n+1)}\sum_{k=1}^{\infty}\mathbb{E}\big(1_{\{Z_n=k\}}\,k\,X_{n,1}|\mathcal{A}_n\big) \\
&= m^{-n}\sum_{k=1}^{\infty}\mathbb{E}\big(k\cdot 1_{\{Z_n=k\}}|\mathcal{A}_n\big) \\
&= m^{-n}Z_n = W_n \quad \text{f.s.}
\end{aligned}$$

□

**3.10 Satz** *Es existiert der fast sichere Limes $W_\infty = \lim_{n\to\infty} W_n$. Es sind äquivalent:*

(a) $m > 1$ ,

(b) $\mathbb{E}(W_\infty) = 1$ ,

(c) $\mathbb{E}(W_\infty) > 0$ .

*Beweis:* Die Limes-Zufallsgröße $W_\infty$ existiert, da $(W_n)_{n\in\mathbb{N}_0}$ ein nicht-negatives Martingal ist, siehe Korollar 2.21. Ist $m \leq 1$, so folgt, dass $(Z_n)_n$ f.s. gegen ein $Z_\infty$ konvergiert. Wegen $\sigma^2 > 0$ kommt nur $Z_\infty = 0$ in Frage.

Es sei nun $m > 1$. Dann wollen wir die Varianz von $W_n$ bestimmen, also die von $Z_n$. Da gibt es eine Formel, die man auch den Satz von Blackwell-Girshick nennt, den wir im Anschluss formulieren und beweisen:

$$\begin{aligned}
\mathrm{Var}(W_n) &= m^{-2n}\big(\sigma^2\mathbb{E}(Z_{n-1}) + m^2\,\mathrm{Var}(Z_{n-1})\big) \\
&= \sigma^2 m^{-(n+1)} + \mathrm{Var}(W_{n-1}) \,.
\end{aligned}$$

Induktiv folgt

$$\mathrm{Var}(W_n) = \sigma^2\sum_{k=2}^{n+1} m^{-k} \leq \frac{\sigma^2 m}{m-1} < \infty \,.$$

Also ist $(W_n)_n$ in $\mathcal{L}^2$ beschränkt, also folgt mit Satz 2.25 $W_n \to W_\infty$ in $\mathcal{L}^2$, somit auch in $\mathcal{L}^1$ und speziell $\mathbb{E}(W_\infty) = \mathbb{E}(W_0) = 1$. □

**3.11 Satz (Waldsche Gleichheit, Satz von Blackwell-Girshick)** *Es sei $(X_n)_n$ eine an eine Filtrierung $(\mathcal{A}_n)_n$ in $\mathcal{A}$ adaptierte Folge identisch verteilter, integrierbarer, reeller Zufallsvariablen über $(\Omega, \mathcal{A}, P)$, wobei $X_{n+1}$ von $\mathcal{A}_n$ unabhängig sei, $n \in \mathbb{N}$. $(S_n)_n$ sei die Folge der Partialsummen $S_n := X_1 + \cdots + X_n$. $T$ sei eine integrierbare Stoppzeit bzgl. $(\mathcal{A}_n)_n$ mit Werten in $\mathbb{N}_0$. Dann gilt:*

(a) *$S_T$ ist integrierbar und*

$$\mathbb{E}(S_T) = \mathbb{E}(T)\,\mathbb{E}(X_1)\ . \quad \textit{(}\text{WALD}\textit{sche Identität)}$$

(b) *Sind die $X_n$ sogar quadratisch integrierbar und zentriert, so trifft dies auch auf $S_T$ zu und es gilt*

$$\mathbb{E}(S_T^2) = \mathbb{E}(T)\,\mathbb{E}(X_1^2)\ . \quad \textit{(}\text{BLACKWELL-GIRSHICK}\textit{)}$$

Wäre die Stoppzeit $T$ beschränkt, so würde man das Martingal $S_n^* := S_n - n\mathbb{E}(X_1)$ betrachten und erhielt mit Satz 2.16 $\mathbb{E}(S_T^*) = \mathbb{E}(S_1^*) = 0$, also (a). Die Bedingung $\mathbb{E}(T) < \infty$ erzwingt ein anderes Vorgehen:

*Beweis:*

(a) Mit der Notation, die wir nach Definition 2.1 eingeführt hatten, gilt

$$\begin{aligned} \mathbb{E}(|S_T|) &= \sum_{n=1}^{\infty} \mathbb{E}(|S_n|; T = n) \le \sum_{n=1}^{\infty}\sum_{i=1}^{n} \mathbb{E}(|X_i|; T = n) \\ &= \sum_{i=1}^{\infty}\sum_{n=i}^{\infty} \mathbb{E}(|X_i|; T = n) = \sum_{i=1}^{\infty} \mathbb{E}(|X_i|; T \ge i)\ . \end{aligned} \tag{3.4}$$

$X_i$ ist von $\mathcal{A}_{i-1}$ unabhängig, also von $1_{\{T\ge i\}}$ $\big($beachte $\{T < i\} = \{T \le i-1\} \in \mathcal{A}_{i-1}\big)$, $i \ge 2$. Es gilt $\{T \ge 1\} = \Omega$. Somit

$$\mathbb{E}(|X_i|; T \ge i) = P(T \ge i)\,\mathbb{E}(|X_i|) = P(T \ge i)\,\mathbb{E}(|X_1|)\ ,$$

also

$$\mathbb{E}(|S_T|) \le \sum_{i=1}^{\infty} P(T \ge i)\,\mathbb{E}(|X_1|) = \mathbb{E}(T)\,\mathbb{E}(|X_1|) < \infty\ ,$$

da $P(T < \infty) = 1$.

Also ist $S_T$ integrierbar. Es folgt die WALDsche Identität, wenn man $S_T$ anstelle von $|S_T|$ betrachtet. Wegen der bewiesenen Konvergenz der Reihe $\sum_{n=1}^{\infty}\sum_{i=1}^{n} \big|\mathbb{E}(X_i; T = n)\big|$ ist die Vertauschung der Summen in (3.4) gerechtfertigt.

(b) Sei $Y_n := X_n 1_{\{n\le T\}}$, also $Y_n \in \mathcal{L}^2$. Nun gilt $\mathbb{E}(Y_m Y_n) = 0$ für $m \ne n$: Für $m < n$ ist $Y_m Y_n = X_n X_m 1_{\{n\le T\}}$ wegen $\{n \le T\} \subset \{m \le T\}$. $X_n$ und $1_{\{n\le T\}} X_m$ sind wieder unabhängig, also

$$\mathbb{E}(Y_m Y_n) = \mathbb{E}(X_n)\,\mathbb{E}(X_m 1_{\{n\le T\}}) = 0\ .$$

Auch $X_n^2$ und $1_{\{n\le T\}}$ sind unabhängig, also

$$\mathbb{E}(Y_n^2) = \mathbb{E}(X_n^2)\,P(n \le T) = \mathbb{E}(X_1^2)\,P(n \le T)\ , \quad n \in \mathbb{N}\ .$$

Somit

$$\sum_{n\geq 1} \mathbb{E}(Y_n^2) = \mathbb{E}(X_1^2)\,\mathbb{E}(T) < \infty\ . \tag{3.5}$$

Weiter ist für $m < n$

$$\mathbb{E}\big((Y_m + \cdots + Y_n)^2\big) = \mathbb{E}\Big(\sum_{i=m}^{n} Y_i^2 + \sum_{i\neq j} Y_i Y_j\Big) = \sum_{i=m}^{n} \mathbb{E}(Y_i^2)\ ,$$

also

$$||Y_m + \cdots + Y_n||_2 = \Big(\sum_{i=m}^{n} \mathbb{E}(Y_i^2)\Big)^{1/2},$$

und somit folgt mit (3.5) die $\mathcal{L}^2$-Konvergenz der Reihe $\sum_{i=1}^{\infty} Y_i$. Es gilt $S_T = \sum_{i=1}^{\infty} X_i 1_{\{i\leq T\}} = \sum_{i=1}^{\infty} Y_i$ fast sicher, denn dies gilt für $\{T < \infty\}$. Also ist $S_T \in \mathcal{L}^2$ und

$$\lim_{n\to\infty} \sum_{i=1}^{n} Y_i = S_T \quad \text{in } \mathcal{L}^2\ .$$

Schließlich gilt nach der majorisierten Konvergenz

$$\mathbb{E}(S_T^2) = \lim_{n\to\infty} \mathbb{E}\Big(\big(\sum_{i=1}^{n} Y_i\big)^2\Big) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}(Y_i^2) = \sum_{i=1}^{\infty} \mathbb{E}(Y_i^2)\ ,$$

also folgt die Identität in (b) aus (3.5) und die WALDsche Identität liefert die Zentriertheit von $S_T$. □

Ist $X_1$ in Teil (b) von Satz 3.11 nicht zentriert und bezeichnet $\mu$ den endlichen Erwartungswert und $\sigma^2$ die endliche Varianz, so liefert die Formel von BLACKWELL-GIRSHICK

$$\mathbb{E}(S_T - \mu\, T)^2 = \sigma^2\, \mathbb{E}(T).$$

Dies kann man nun nach $\mathbb{E}(S_T^2)$ auflösen und erhält

$$\mathbb{E}(S_T^2) = \sigma^2\, \mathbb{E}(T) + 2\mu\, \mathbb{E}(S_T)\, \mathbb{E}(T) - \mu^2\, \mathbb{E}(T^2).$$

Daraus folgt dann

$$\mathrm{Var}(S_T) = \mathbb{E}(S_T^2) - (\mathbb{E}(S_T))^2 = \sigma^2\, \mathbb{E}(T) + \mu^2\, \mathrm{Var}(T).$$

Tatsächlich haben wir diese Formel zur Bestimmung von $\mathrm{Var}(W_n)$ verwendet.

### 3.6 Black-Scholes Formel, Finanzmathematik

Wir betrachten das sogenannte Cox-Ross-Rubinstein *Modell*, eine diskrete Version eines Aktienkursmodells in stetiger Zeit, in welchem ursprünglich die Black-Scholes *Formel* entwickelt wurde. Für Arbeiten zur Finanzmathematik Anfang der siebziger Jahre erhielten M.Scholes und R.C.Merton 1997 den Nobelpreis für Wirtschaftswissenschaften. F.Black verstarb 1996.

Ein Finanzmarkt biete zwei Anlageobjekte (securities): Wertpapiere (bonds) mit fester Verzinsung $r$ und Aktien (stock), deren Kurs zufällig schwankt. Die Kurse der Wertpapiere und Aktien variieren in $n = 0, 1, \ldots, N$ (Zeitpunkte). $V_n = (1+r)^n V_0$ sei der Wert des Wertpapiers im Zeitintervall $[n, n+1)$, $S_n$ der zufällige Wert der Aktie in $[n, n+1)$; $S_0 > 0$. Haben wir in $n = 0$ $W_0$ Wertpapiere und $A_0$ Aktien im Gesamtwert von $x$, so gilt

$$W_0 V_0 + A_0 S_0 = x \ .$$

In $[0, 1)$ können Wertpapiere gegen Aktien (und umgekehrt) getauscht werden. Wir setzen voraus, dass keine *Transaktionskosten* entstehen, der Wert des Portfolios also unverändert bleibt. Unmittelbar vor $n = 1$ gilt

$$W_1 V_0 + A_1 S_0 = x \quad (W_1/A_1 \text{ Bestände an Wertpapieren/Aktien}).$$

In $n = 1$ hat nun das Portfolio den neuen Wert

$$X_1 = W_1 V_1 + A_1 S_1 \ .$$

$(W_{n-1}, A_{n-1})$ sei die Zusammensetzung des Portfolios zum Zeitpunkt $n-1$, $n \geq 1$, und

$$X_{n-1} = W_{n-1} V_{n-1} + A_{n-1} S_{n-1} \ .$$

der Kurswert. Umschichten führt in $(n-1, n)$ zu $(W_n, A_n)$ mit $X_{n-1} = W_n V_{n-1} + A_n S_{n-1}$ und zum Zeitpunkt $n$ ist der neue Wert

$$X_n = W_n V_n + A_n S_n \ .$$

Sei für $n \geq 1$ $R_n$ via $S_n = (1 + R_n) S_{n-1}$ definiert ($R_n$ ist eine Zufallsvariable!), also $S_n - S_{n-1} = R_n S_{n-1}$, so gilt mit $V_n - V_{n-1} = r V_{n-1}$

$$\begin{aligned} X_n - X_{n-1} &= W_n(V_n - V_{n-1}) + A_n(S_n - S_{n-1}) = r W_n V_{n-1} + A_n R_n S_{n-1} \\ &= r X_{n-1} + A_n S_{n-1}(R_n - r) \ . \end{aligned}$$

Sei weiter $Y_n := \frac{1}{(1+r)^n} X_n$ (diskontierter Wert des Portfolios in $n$), so ist

$$Y_n - Y_{n-1} = \frac{1}{(1+r)^n} A_n S_{n-1}(R_n - r) \ . \tag{3.6}$$

Die Zufallsgrößen seien auf $(\Omega, \mathcal{A}, P_\theta)$, $\theta \in (0,1)$, definiert. Sei $\mathcal{A}_n = \sigma(S_1, \ldots, S_n)$, $n \geq 1$, $\mathcal{A}_0$ die triviale $\sigma$-Algebra ($S_0$ konstant). Dann sind $W_n$ und $A_n$ $\mathcal{A}_{n-1}$-messbar, also ist $(W_n, A_n)_{1 \leq n \leq N}$ eine *vorhersagbare Folge* bzgl. $(\mathcal{A}_n)_{0 \leq n \leq N}$, genannt *Handelsstrategie.*

Nun setzen wir voraus, dass $(R_n)_{1 \leq n \leq N}$ eine unabhängige Folge identisch verteilter Zufallsgrößen mit

$$\theta := P_\theta(R_1 = b) = 1 - P_\theta(R_1 = a)$$

mit $-1 < a < r < b < \infty$ und $\theta \in (0,1)$ ist. Natürlich gilt mit Definition der $R_n$: $\mathcal{A}_n = \sigma(R_1, \ldots, R_n)$. Weiter ist $\mathbb{E}_{\theta^*}(R_1 - r) = 0$ für $\theta^* := \frac{r-a}{b-a}$, und somit ist $(Y_n)_{0 \leq n \leq N}$ unter $P_{\theta^*}$ ein Martingal bzgl. $(\mathcal{A}_n)_{0 \leq n \leq N}$, denn

$$\mathbb{E}_{\theta^*}\big(\frac{1}{(1+r)^n} A_n S_{n-1}(R_n - r)\big|\mathcal{A}_{n-1}\big) = \frac{1}{(1+r)^n} A_n S_{n-1} \mathbb{E}_{\theta^*}\big((R_n - r)|\mathcal{A}_{n-1}\big),$$

und es gilt $\mathbb{E}_{\theta^*}\big((R_n - r)|\mathcal{A}_{n-1}\big) = \mathbb{E}_{\theta^*}(R_n - r) = 0$. Die Bedingung $a < r < b$ kommentieren wir später.

Eine *europäische Kauf-Option* (European call option) mit Laufzeit $N$ ist ein Vertrag, abgeschlossen in $n = 0$, der dem Käufer zum Zeitpunkt $n = N$ den Kauf einer Aktie zum Preis $K$ erlaubt. $K$ heißt *Ausübungspreis* (striking price). Man wird die Option ausüben, wenn $S_N > K$, andernfalls wird man die Option fallen lassen. Der Wert der Option in $N$ ist also $(S_N - K)_+$. Wieviel sollte man für die Option zum Zeitpunkt 0 bezahlen? Black und Scholes führten zur Beantwortung eine *selbstfinanzierende Handelsstrategie* (hedging strategy) ein: Die Bewertungsfolge $(X_0, \ldots, X_N)$ genüge:

(a) $X_0 = x$

(b) $X_n \geq 0$ für $0 \leq n \leq N$

(c) $X_N = (S_N - K)_+$ .

Man fordert zwar $X_n \geq 0$ für alle $0 \leq n \leq N$, nicht aber, dass $W_n$ und $A_n$ stets nichtnegativ sein müssen. Dabei bedeutet $W_n < 0$ eine Kreditaufnahme zum Zinssatz $r$. Die sogenannten Leerkäufe $A_n < 0$ müssen wir nicht berücksichtigen, da sie gar nicht auftreten. Dies besagt unter anderem das folgende Resultat:

**3.12 Satz (von Black und Scholes)** *Eine selbstfinanzierende Handelsstrategie $(W_n, A_n)_{1 \leq n \leq N}$ mit Anfangswert $x$ gibt es genau dann, wenn*

$$x = x_0 := \frac{1}{(1+r)^N} \mathbb{E}_{\theta^*}(S_N - K)_+ \ .$$

*In diesem Fall ist $A_n \geq 0$ für alle $1 \leq n \leq N$.*

Alle Zufallsgrößen sind bzgl. $\mathcal{A}_N = \sigma(R_1, \dots, R_N)$ messbar. Wir wählen $\Omega = \{a, b\}^N$ mit Potenzmenge und W-Maß $P_\theta$, definiert durch

$$P_\theta(\{\omega_1, \dots, \omega_N\}) := \theta^{n(\omega)}(1-\theta)^{N-n(\omega)} ,$$

$n(\omega) = |\{i : \omega_i = b\}|$ und $R_n$ als Projektion von $\Omega$ auf die $n$-te Komponente. Die Wahl liefert, dass unter jedem $P_\theta$, $\theta \in (0,1)$, alle Elemente in $\Omega$ positive Wahrscheinlichkeiten haben und somit nur die leere Menge eine Nullmenge bildet.

**3.13 Satz** *Jede bzgl. $(\mathcal{A}_n)_{0\le n\le N}$ adaptierte Folge $(M_n)_{0\le n\le N}$, die unter einem $P_{\theta_0}$ ein Martingal ist, hat die Darstellung*

$$M_n = M_0 + \sum_{k=1}^{n} H_k(R_k - r_{\theta_0}) , \quad 1 \le n \le N ,$$

*wobei $r_{\theta_0} = \mathbb{E}_{\theta_0}(R_1)$ und $H = (H_1, \dots, H_N)$ eine eindeutig bestimmte vorhersagbare Folge bildet.*

*Beweis:* $M_n$ ist $\mathcal{A}_n$-messbar, also existiert $f_n : \{a,b\}^n \to \mathbb{R}$ mit $M_n(\omega) = f_n\big(R_1(\omega), \dots, R_n(\omega)\big) = f_n(\omega_1, \dots, \omega_n)$, siehe das Faktorisierungslemma 1.13. Mit der Martingaleigenschaft folgt *für alle* $\omega \in \Omega$

$$\begin{aligned} 0 &= \mathbb{E}_{\theta_0}(M_n - M_{n-1} | \mathcal{A}_{n-1})(\omega) \\ &= \theta_0 f_n(\omega_1, \dots, \omega_{n-1}, b) + (1-\theta_0) f_n(\omega_1, \dots, \omega_{n-1}, a) - f_{n-1}(\omega_1, \dots, \omega_{n-1}) , \end{aligned}$$

und somit

$$\frac{f_n(\omega_1, \dots, \omega_{n-1}, b) - f_{n-1}(\omega_1, \dots, \omega_{n-1})}{(1-\theta_0)(b-a)} = \frac{f_{n-1}(\omega_1, \dots, \omega_{n-1}) - f_n(\omega_1, \dots, \omega_{n-1}, a)}{\theta_0(b-a)} .$$

Diesen Wert setzen wir $H_n(\omega)$, also ist $H_n$ $\mathcal{A}_{n-1}$-messbar und

$$H_n(R_n - r_{\theta_0}) = f_n(R_1, \dots, R_n) - f_{n-1}(R_1, \dots, R_{n-1}) = M_n - M_{n-1} ,$$

also ist die Darstellung bewiesen. Details hierzu sowie der Beweis der Eindeutigkeit von $H$ sind eine Übung. □

*Beweis von Satz 3.12:* Angenommen, es existiert eine selbstfinanzierende Handelsstrategie $(W_n, A_n)_{1\le n\le N}$ mit Anfangswert $x$ und $X_n$ und $Y_n$ seien wie oben gewählt. Dann ist $(Y_n)_{0\le n\le N}$ unter $P_{\theta^*}$ ein Martingal bzgl. $(\mathcal{A}_n)_{0\le n\le N}$, also

$$\mathbb{E}_{\theta^*} Y_N = \mathbb{E}_{\theta^*} Y_0 = x \quad \text{und} \quad Y_N = (1+r)^{-N} X_N = (1+r)^{-N}(S_N - K)_+ ,$$

also folgt $x = x_0$.

Für die umgekehrte Richtung definieren wir die Folge $(Y_n)_{0\le n\le N}$ *neu*:

$$Y_n := (1+r)^{-N}\,\mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathcal{A}_n\big)\ .$$

Dies ist ein $P_{\theta^*}$-Martingal. Wir suchen $A_n \ge 0$ und $W_n$, $n \in \{1,\dots,N\}$, so dass $Y_n$ (3.6) erfüllt und $(A_n)_n$ und $(W_n)_n$ vorhersagbar sind. Mit Satz 3.13 existiert ein eindeutig bestimmter vorhersehbarer Prozess $(A_0,\dots,A_N)$, so dass $Y_n$ (3.6) erfüllt. Dazu wählen wir $A_0 > 0$ konstant und

$$A_n = \frac{H_n(1+r)^n}{S_{n-1}}, \quad 1 \le n \le N.$$

Wir setzen nun:

$$\begin{aligned} X_n &:= (1+r)^n Y_n \\ W_n &:= \frac{X_n - A_n S_n}{V_n}\ . \end{aligned}$$

Es bleibt die Nicht-Negativität der $A_n$ zu zeigen: Wir erinnern an die Definition von $H_n$ im Beweis des obigen Satzes bei der Wahl $Y_n =: M_n$, dann folgt mit (3.6)

$$A_n = \frac{(1+r)^{N-n}\Big(\mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_{n-1}, R_n = b\big) - \mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_{n-1}\big)\Big)}{S_{n-1}(1-\theta^*)(b-a)}$$

mit $\mathbb{R}_n := (R_1,\dots,R_n)$. Wir betrachten

$$\begin{aligned} &\mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_{n-1}\big) \\ &= \theta^*\mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_{n-1}, R_n=b\big) + (1-\theta^*)\mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_{n-1}, R_n=a\big)\ , \end{aligned}$$

womit $A_n \ge 0$ genau dann gilt, wenn

$$\mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_{n-1}, R_n=b\big) \ge \mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_{n-1}, R_n=a\big)\ .$$

Dies folgt mit $b > a$ und $S_N = S_0\prod_{i=1}^N(1+R_i)$, formal:

$$\begin{aligned} &\mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_n = (r_1,\dots,r_{n-1},b)\big) \\ &\quad - \mathbb{E}_{\theta^*}\big((S_N-K)_+|\mathbb{R}_n = (r_1,\dots,r_{n-1},a)\big) \\ &= \mathbb{E}_{\theta^*}\big(S_0S_{n,N}(1+b)^{k+1}(1+a)^{n-1-k} - K\big)_+ \\ &\quad - \mathbb{E}_{\theta^*}\big(S_0S_{n,N}(1+b)^{k+1}(1+a)^{n-k} - K\big)_+ \end{aligned}$$

mit $k = k(r_1,\dots,r_{n-1}) := \big|\{1 \le i \le n-1 : r_i = b\}\big|$ und $S_{n,N} := \frac{S_N}{S_n}$. $\square$

Die Voraussetzung $a < r < b$ sichert, dass der Markt *Arbitrage-frei* ist. Arbitrage bedeutet: Man kann ohne einen Cent zu besitzen Geld verdienen. Gilt $r \leq a < b$, so ist $\left( \prod_{i=1}^{n} \frac{1+R_i}{1+r} \right)_{0 \leq n \leq N}$ eine monoton wachsende Folge. Ein Anleger, der sich zum Zeitpunkt 0 den Betrag $x_0 > 0$ zum Zinssatz $r$ leiht und den Betrag sofort in Aktien investiert, macht daher mit negativem Anfangskapital den Gewinn $x_0 \Big( \big( \prod_{i=1}^{n} (1+R_i) \big) - (r+1)^N \Big) \geq 0$ zum Zeitpunkt $N$, der mit positiver Wahrscheinlichkeit strikt positiv ist. Gilt aber $a < b \leq r$, so ist obige Folge monoton fallend. Hier erzielt ein Anleger durch Leerkauf[1] von Aktien im Wert von $S_0 = x_0$ und sofortige Anlage des erzielten Erlöses $x_0$ zum Zinssatz $r$ den Gewinn

$$x_0 \Big( (1+r)^N - \big( \prod_{i=1}^{N} (1+R_i) \big) \Big) \geq 0 \ ,$$

wiederum mit positiver Wahrscheinlichkeit strikt positiv. Arbitrage kann immer nur kurzfristig existieren, weil sich der Markt schnell anpasst, um diese Arbitrage zu zerstören. Arbitragefreiheit scheint daher eine realistische Annahme zu sein.

[1] Ein Leerkauf: der Anleger „leiht“ über einen Broker von einem anderen Aktienbesitzer Aktien und verkauft sie auf dem Markt zum gegenwärtigen Kurs. Zu einem späteren Zeitpunkt erwirbt er dieselbe Anzahl von Aktien zum dann gültigen Kurs zwecks Rückgabe an denjenigen, von dem die Aktien geliehen wurden. Er macht einen Gewinn, wenn der Kurs gesunken ist, andernfalls einen Verlust!

KAPITEL 4

# Stationäre Prozesse

Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

**4.1 Definition** Sei $(E, \mathcal{E})$ ein messbarer Raum. Eine Folge $\mathbb{X} = (X_n)_n$ von $(E, \mathcal{E})$-wertigen Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ heißt $(E, \mathcal{E})$-*wertiger stochastischer Prozess*. Im Fall $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ sagen wir einfach *stochastischer Prozess*.

Die Verteilung eines stochastischen Prozesses $\mathbb{X}$ ist einfach seine Verteilung als $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$-wertige Zufallsvariable, d. h. das W-Maß $P\mathbb{X}^{-1}$ auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$. Ist $\mathbb{X}$ ein stochastischer Prozess, so ist die Folge $(\pi_n)_n$ der Projektionen $\pi_n : E^{\mathbb{N}} \to E$ ein auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}, P\mathbb{X}^{-1})$ definierter stochastischer Prozess, der dieselbe Verteilung wie $\mathbb{X}$ hat.

**4.2 Definition** Ist $Q$ ein W-Maß auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ und ist $\pi^{(n)} : E^{\mathbb{N}} \to E^n$, $n \in \mathbb{N}$, durch

$$\pi^{(n)} = (\pi_1, \dots, \pi_n)$$

definiert, so ist

$$Q^{(n)} := Q\big(\pi^{(n)}\big)^{-1}$$

ein W-Maß auf $(E^n, \mathcal{E}^n)$. Die Maße $Q^{(n)}$ heißen *endlichdimensionale Verteilungen* von $Q$.

**4.3 Satz** *Die Folge der endlichdimensionalen Verteilungen $(Q^{(n)})_n$ bestimmt das W-Maß $Q$ eindeutig.*

*Beweis:* Wir geben ein durchschnittstabiles Erzeugendensystem $\mathcal{D}$ von $\mathcal{E}^{\mathbb{N}}$ an:

$$\mathcal{D} := \{A_J,\ J \subset \mathbb{N} \text{ endlich},\ A_j \in \mathcal{E},\ j \in J\}$$

mit

$$A_J := \Big\{(x_i)_{i \in \mathbb{N}} \in E^{\mathbb{N}} : x_j \in A_j \text{ für } j \in J\Big\}.$$

Dann gilt $\bigotimes_{i \in \mathbb{N}} \mathcal{E} = \mathcal{E}^{\mathbb{N}} = \sigma(\mathcal{D})$. $A_J$ nennt man auch „endlichdimensionale Zylinder“. Nach Satz 1.9, Kapitel 1, Wahrscheinlichkeitstheorie, stimmen zwei W-Maße auf $\mathcal{E}^{\mathbb{N}}$ überein, wenn sie auf $\mathcal{D}$ übereinstimmen. $Q$ ist also auf $\mathcal{D}$ festgelegt. Daraus folgt die Behauptung. $\square$.

Im Fall $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ folgt

**4.4 Proposition** Sind $\mathbb{X} = (X_n)_n$ und $\mathbb{X}' = (X'_n)_n$ zwei $(\mathbb{R}, \mathcal{B})$-wertige stochastische Prozesse auf $(\Omega, \mathcal{A}, P)$ bzw. $(\Omega', \mathcal{A}', P')$, so ist $\mathcal{L}(\mathbb{X}) = \mathcal{L}(\mathbb{X}')$ genau dann, wenn

$$P(X_1 \leq t_1, \dots, X_n \leq t_n) = P'(X'_1 \leq t_1, \dots, X'_n \leq t_n)$$

für alle $n \in \mathbb{N}$ und alle $t_1, \dots, t_n \in \mathbb{R}$ gilt.

*Beweis:* $\{x \in \mathbb{R}^{\mathbb{N}} : x_1 \leq t_1, \dots, x_n \leq t_n\}$ bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}^{\mathbb{N}}$. Nach Voraussetzung stimmen $P\mathbb{X}^{-1}$ und $P'\mathbb{X}'^{-1}$ auf diesen Mengen überein. Also folgt $P\mathbb{X}^{-1} = P'\mathbb{X}'^{-1}$. $\square$

**4.5 Definition** Eine $\mathcal{A}$-$\mathcal{A}$-messbare Abbildung $T : \Omega \to \Omega$ heißt *maßerhaltende Transformation*, wenn

$$PT^{-1} = P$$

gilt. Man sagt auch, dass $P$ *$T$-invariant* (oder *invariant unter $T$*) ist. Ein stochastischer Prozess $\mathbb{X}$ heißt *stationär*, wenn $\mathbb{X}$ dieselbe Verteilung wie der Prozess $(X_2, X_3, \dots)$ hat.

**4.6 Lemma** *Es seien $Y$ eine Zufallsgröße und $T$ eine maßerhaltende Transformation auf $(\Omega, \mathcal{A}, P)$. $X_1 := Y$; $X_{n+1} := X_n \circ T$, $n \in \mathbb{N}$. Dann ist $(X_n)_n$ ein stationärer Prozess.*

*Beweis:* Es seien $n \in \mathbb{N}$ und $t_1, \dots, t_n \in \mathbb{R}$. Dann gilt

$$\begin{aligned} P(X_1 \leq t_1, \dots, X_n \leq t_n) &= PT^{-1}(X_1 \leq t_1, \dots, X_n \leq t_n) \\ &= P(X_1 \circ T \leq t_1, \dots, X_n \circ T \leq t_n) \\ &= P(X_2 \leq t_1, \dots, X_{n+1} \leq t_n)\,. \end{aligned}$$

Also folgt die Behauptung mit Proposition 4.4. $\square$

**4.7 Bemerkung** Zu jedem auf $(\Omega, \mathcal{A}, P)$ definierten, stationären Prozess $\mathbb{X}$ existiert eine maßerhaltende Transformation, die einen Prozess liefert, der in Verteilung mit $\mathbb{X}$ übereinstimmt. $\mathbb{X}$ ist eine $\mathcal{A}$-$\mathcal{B}^{\mathbb{N}}$-messbare Abbildung, sei $P' := P\mathbb{X}^{-1}$. Sei $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ definiert durch $T\big((x_1, x_2, \dots)\big) = (x_2, x_3, \dots)$. Dann ist $T$ auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$ maßerhaltend, wenn $\mathbb{X}$ stationär ist:

Für $A \in \mathcal{B}^{\mathbb{N}}$ gilt

$$P'(A) = P(\mathbb{X} \in A) = P(T \circ \mathbb{X} \in A) = P\mathbb{X}^{-1}T^{-1}(A) = P'T^{-1}(A)\,.$$

$P'$ ist auch die Verteilung der Folge der Projektionen $\pi_1, \pi_2, \dots : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, denn $\pi = (\pi_n)_n$ ist die identische Abbildung auf $\mathbb{R}^{\mathbb{N}}$. Weiter gilt $\pi_n = \pi_{n-1} \circ T$, $n \geq 2$. Der Prozess $(\pi_n)_n$ auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ ist somit von der in Lemma 4.6 definierten Form und besitzt dieselbe Verteilung wie $\mathbb{X}$.

**4.8 Beispiele**

(a) Eine Folge von unabhängigen, identisch verteilten Zufallsgrößen ist ein stationärer Prozess. Ist $\mu$ ein W-Maß auf $(\mathbb{R}, \mathcal{B})$, so ist $\mu^{\otimes \mathbb{N}}$ invariant unter der Verschiebungsabbildung $T$, wie in Bemerkung 4.7 definiert.

(b) $\Omega = D := \{z \in \mathbb{C} : |z| = 1\}$. Wir übertragen vom Intervall $[0, 2\pi)$ die Borel-$\sigma$-Algebra und die gleichförmige Verteilung (= Lebesgue-Maß/$2\pi$) auf $D$ und nennen dies $(D, \mathcal{B}_D, \lambda_D)$. Für $c \in D$ sei $T_c : D \to D$ durch $T_c(\omega) := c\,\omega$ definiert. Ist $A \subset D$ ein Intervall, so gilt

$$\lambda_D\big(T_c^{-1}(A)\big) = \lambda_D(A)\,.$$

Intervalle bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_D$, woraus folgt, dass $T_c$ maßerhaltend ist.

(c) Auf $([0,1), \mathcal{B}_{[0,1)}, \lambda_{[0,1)})$ betrachten wir $T : [0,1) \to [0,1)$, definiert durch

$$T(\omega) := \begin{cases} 2\omega & \text{für } \omega \in [0, \frac{1}{2})\,, \\ 2\omega - 1 & \text{für } \omega \in [\frac{1}{2}, 1)\,. \end{cases}$$

Für $0 \le a < b \le 1$ ist

$$T^{-1}\big([a,b)\big) = \Big[\frac{a}{2}, \frac{b}{2}\Big) \cup \Big[\frac{a+1}{2}, \frac{b+1}{2}\Big)\,,$$

also $PT^{-1}\big([a,b)\big) = b - a = P\big([a,b)\big)$. Die Mengen $[a,b)$ mit $0 \le a < b \le 1$ zusammen mit $\emptyset$ bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_{[0,1)}$, also folgt $PT^{-1} = P$.

**4.9 Definition** Für eine Abbildung $T : \Omega \to \Omega$ heißt eine Teilmenge $A$ von $\Omega$ $T$-*invariant*, wenn $T^{-1}(A) = A$ gilt. Eine maßerhaltende Abbildung $T$ auf $(\Omega, \mathcal{A}, P)$ heißt *ergodisch*, wenn für jede $T$-invariante Menge $A \in \mathcal{A}$ gilt:

$$P(A) \in \{0, 1\}\,.$$

**4.10 Bemerkung** Die Familie der $T$-invarianten Mengen aus $\mathcal{A}$ bilden eine Teil-$\sigma$-Algebra von $\mathcal{A}$, die hier mit $\mathcal{J}$ bezeichnet wird. Ist $T$ ergodisch, so sagt man, dass $\mathcal{J}$ $P$-*trivial* ist. Jede $\mathcal{J}$-messbare Zufallsgröße ist fast sicher konstant (im Fall $T$ ergodisch). Dies folgt aus Korollar 5.20, Wahrscheinlichkeitstheorie.

**4.11 Satz** *Es sei $\mathbb{X} = (X_n)_n$ eine Folge von unabhängigen, identisch verteilten Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ mit $P' := P\mathbb{X}^{-1}$. Dann ist die Verschiebungsabbildung $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, $\{x_n\}_{n \in \mathbb{N}} \mapsto \{x_{n+1}\}_{n \in \mathbb{N}}$ ergodisch auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$.*

*Beweis:* Sei $A \in \mathcal{B}^{\mathbb{N}}$ eine $T$-invariante Menge, so gilt $A = T^{-n}(A) := (T^n)^{-1}(A) = \{(x_k)_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} : (x_{n+k})_{k\in\mathbb{N}} \in A\}$. Also gilt

$$\begin{aligned}\mathbb{X}^{-1}(A) &= \{\omega : \big(X_{n+1}(\omega), X_{n+2}(\omega), \dots\big) \in A\} \in \tilde{\mathcal{A}}_{n+1} \\ &:= \sigma\Big(\bigcup_{k=n+1}^{\infty} \big(X_k^{-1}(A)\big)\Big)\,.\end{aligned}$$

Dies gilt für alle $n \in \mathbb{N}$, also

$$\mathbb{X}^{-1}(A) \in \mathcal{A}_\infty := \bigcap_{n=1}^{\infty} \tilde{\mathcal{A}}_n\,.$$

Nach Kolmogorovs 0-1-Gesetz folgt

$$P'(A) = P\big(\mathbb{X}^{-1}(A)\big) \in \{0,1\}\,. \qquad \square$$

**4.12 Bemerkung** Die $\sigma$-Algebra $\mathcal{J}$ der verschiebungsinvarianten messbaren Mengen in $\mathbb{R}^{\mathbb{N}}$ ist eine Teil-$\sigma$-Algebra der terminalen $\sigma$-Algebra

$$\mathcal{A}_\infty = \bigcap_{n=1}^{\infty} \sigma(\pi_k : k \geq n)\,.$$

Es gilt jedoch keinesfalls $\mathcal{J} = \mathcal{A}_\infty$ (ohne Beweis).

**4.13 Satz** *Es seien $D = \{z \in \mathbb{C} : |z| = 1\}$, $\mathcal{B}_D$ die Borel-$\sigma$-Algebra auf $D$ und $\lambda_D$ das normierte Lebesgue-Maß auf $(D, \mathcal{B}_D)$, $T_c : D \to D$ definiert durch $T_c(\omega) := c\,\omega$ für $c, \omega \in D$. Dann ist $T_c$ genau dann ergodisch auf $(D, \mathcal{B}_D, \lambda_D)$, wenn $c$ keine Einheitswurzel ist.*

*Beweis:* Setze $T := T_c$. Ist $c$ eine Einheitswurzel, existiert also ein $n \in \mathbb{N}$ mit $c^n = 1$, so ist $T^n$ die Identität. Dann ist für jedes $A \in \mathcal{B}_D$ die Menge $A \cup T^{-1}A \cup \dots \cup T^{-n+1}A$ invariant unter $T$.

Wähle ein $A \in \mathcal{B}_D$ mit $0 < \lambda_D(A) < 1/n$, so gilt

$$0 < \lambda_D(A \cup \dots \cup T^{-n+1}A) \leq \sum_{k=0}^{n-1} \lambda_D(T^{-k}A) = n\lambda_D(A) < 1\,.$$

Somit ist $T$ nicht ergodisch. Für die Umkehrung treffen wir zunächst Vorbereitungen:

**4.14 Lemma** *Sei $c \in D$ keine Einheitswurzel. Dann liegt $\{c^n, n \in \mathbb{N}_0\}$ dicht in $D$.*

*Beweis:* Da die $c^n$, $n \in \mathbb{N}$, alle verschieden sind, besitzt diese Folge mindestens einen Häufungspunkt $\omega_0 \in D$. Seien $\varepsilon > 0$ und $m > n$ in $\mathbb{N}$ mit $|c^n - \omega_0| < \varepsilon$ und $|c^m - \omega_0| < \varepsilon$. Dann gilt $|c^{m-n} - 1| \in (0, 2\varepsilon)$. Daraus folgt, dass für jedes $\omega \in D$ ein $k \in \mathbb{N}$ existiert mit $|\omega - c^{k(m-n)}| < 2\varepsilon$. Da $\varepsilon > 0$ beliebig war, folgt die Behauptung. $\square$

**4.15 Lemma** *Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $\mathcal{F}$ eine Algebra mit $\mathcal{A} = \sigma(\mathcal{F})$. Zu $\varepsilon > 0$ und $A \in \mathcal{A}$ existiert dann ein $B \in \mathcal{F}$ mit $P(A \triangle B) < \varepsilon$.*

*Beweis:* maßtheoretische Übung. $\square$

Wir führen den Beweis von Satz 4.13 fort: Sei $c$ keine Einheitswurzel. $\mathcal{F}$ sei die Familie der endlichen Vereinigungen paarweise disjunkter Intervalle in $D$; dann ist $\mathcal{F}$ eine Algebra. Sei nun $A$ invariant unter $T_c$ mit $\lambda_D(A) > 0$ und $\varepsilon \in (0, 1/2)$, dann existiert nach 4.15 eine endliche Vereinigung $B = \bigcup_{i=1}^n I_i$ disjunkter Intervalle $I_1, \ldots, I_n \subset D$ mit $\lambda_D(A \triangle B) \leq \varepsilon \lambda_D(A)$.

Ohne Einschränkung können wir $\lambda_D(I_i) \leq \varepsilon$ für $i = 1, \ldots, n$ annehmen. Es folgt nun

$$\begin{aligned}\lambda_D(A \triangle B) &\leq \varepsilon \lambda_D(A) \leq 2\varepsilon(1-\varepsilon)\lambda_D(A) \\ &\leq 2\varepsilon\big(\lambda_D(A) - \lambda_D(A \triangle B)\big) \leq 2\varepsilon \lambda_D(B)\,,\end{aligned}$$

wobei wir $|\lambda_D(A) - \lambda_D(B)| \leq \lambda_D(A \triangle B)$ verwendet haben. Dies gilt für jedes Maß und folgt aus $A \subset B \cup (A \setminus B)$, womit $\lambda_D(A) - \lambda_D(B) \leq \lambda(A \setminus B) \leq \lambda_D(A \triangle B)$, und dies ist symmetrisch in $A$ und $B$. Es folgt nun

$$\begin{aligned}\sum_{i=1}^n \lambda_D(A \cap I_i) &= \lambda_D(A \cap B) \geq \lambda_D(B) - \lambda_D(A \triangle B) \\ &\geq (1 - 2\varepsilon)\lambda_D(B) \\ &= (1 - 2\varepsilon) \sum_{i=1}^n \lambda_D(I_i).\end{aligned}$$

Mindestens eines der $I_i$ erfüllt also die Ungleichung

$$\lambda_D(A \cap I_i) \geq (1 - 2\varepsilon)\lambda_D(I_i)\,.$$

Mit Lemma 4.14 kann gefolgert werden, dass ein $k \in \mathbb{N}$ und $n_1, \ldots, n_k \in \mathbb{N}$ existieren, so dass die Intervalle

$$T^{-n_1} I_i\,,\, T^{-n_2} I_i\,,\, \ldots\,,\, T^{-n_k} I_i$$

paarweise disjunkt sind und $D$ bis auf eine Menge von kleinerem Maß als $2\varepsilon$ ausfüllen. Wegen der $T$-Invarianz von $A$ und $\lambda_D$ gilt für $j = 1, \ldots, k$:

$$\begin{aligned}\lambda_D(A \cap T^{-n_j} I_i) &= \lambda_D(T^{-n_j} A \cap T^{-n_j} I_i) = \lambda_D T^{-n_j}(A \cap I_i) \\ &= \lambda_D(A \cap I_i) \geq (1 - 2\varepsilon)\lambda_D(I_i) \\ &= (1 - 2\varepsilon)\lambda_D(T^{-n_j} I_i)\,,\end{aligned}$$

also

$$\begin{aligned}\lambda_D(A) &\geq \sum_{j=1}^{k} \lambda_D(A \cap T^{-n_j} I_i) \geq (1 - 2\varepsilon) \sum_{j=1}^{k} \lambda_D(T^{-n_j} I_i) \\ &\geq (1 - 2\varepsilon)^2\,.\end{aligned}$$

Daraus folgt $\lambda_D(A) = 1$. □

**4.16 Definition** Eine maßerhaltende Transformation $T$ auf $(\Omega, \mathcal{A}, P)$ heißt *mischend*, wenn für alle $A, B \in \mathcal{A}$ gilt

$$\lim_{n\to\infty} P(A \cap T^{-n} B) = P(A)P(B)\,. \tag{4.1}$$

**4.17 Lemma** *Jede mischende Transformation ist ergodisch.*

*Beweis:* Ist $A \in \mathcal{A}$ eine $T$-invariante Menge, so gilt

$$P(A) = P(A \cap A) = P(A \cap T^{-n} A) \to P(A)P(A)$$

für $n \to \infty$, also $P(A) = P(A)^2$, also $P(A) \in \{0, 1\}$. □

**4.18 Lemma** *Es sei $\mathcal{F}$ eine Algebra, die $\mathcal{A}$ erzeugt. Falls die Gleichung* (4.1) *für alle $A, B \in \mathcal{F}$ gilt, so ist $T$ mischend.*

*Beweis:* $A, B \in \mathcal{A}$, $\varepsilon > 0$. Nach Lemma 4.15 existieren $A_0, B_0 \in \mathcal{F}$ mit $P(A \triangle A_0) < \varepsilon$ und $P(B \triangle B_0) < \varepsilon$. Daraus folgt

$$\begin{aligned}\left|P(A \cap T^{-n} B) - P(A_0 \cap T^{-n} B_0)\right| &\leq P(A \triangle A_0) + P\big(T^{-n} B \triangle T^{-n} B_0\big) \\ &\leq P(A \triangle A_0) + P(B \triangle B_0) < 2\varepsilon\,.\end{aligned}$$

Wieder haben wir $|P(A) - P(B)| \leq P(A \triangle B)$ verwendet. Die Folge $\big(P(A_0 \cap T^{-n} B_0)\big)_n$ konvergiert gegen $P(A_0)P(B_0)$ und es gilt $|P(A_0) - P(A)| \leq P(A \triangle A_0) < \varepsilon$ und analog $|P(B_0) - P(B)| < \varepsilon$. Also folgt

$$\lim_{n\to\infty} P(A \cap T^{-n} B) = P(A)P(B)\,.$$ □

**4.19 Satz** *Die in Beispiel 4.8 (c) definierte Transformation $T$ auf $([0,1), \mathcal{B}_{[0,1)}, \lambda_{[0,1)})$ ist mischend, also ergodisch.*

*Beweis:* Für jede Menge $A \subset [0,1]$ sind $T^{-1}A \cap [0,1/2)$ und $T^{-1}A \cap [1/2,1)$ nur um $1/2$ gegeneinander verschobene Mengen. Also folgt mit $P := \lambda_{[0,1)}$

$$\begin{aligned} P\Big(T^{-1}A \cap \big[\frac{1}{2}, 1\big)\Big) &= P\Big(T^{-1}A \cap \big[0, \frac{1}{2}\big)\Big) = \frac{1}{2} P(T^{-1}A) \\ &= P(A) P\Big(\big[0, \frac{1}{2}\big)\Big) \,. \end{aligned}$$

Völlig analog folgt für $n \in \mathbb{N}$ und $k = 0, 1, \ldots, 2^n - 1$:

$$P\big(T^{-n}A \cap [k2^{-n}, (k+1)2^{-n})\big) = P(A) P\big([k2^{-n}, (k+1)2^{-n})\big) \,.$$

$\mathcal{A}_0$ sei die Familie der Intervalle der Form $\big[k2^{-n}, (k+1)2^{-n}\big)$ mit $n \in \mathbb{N}$ und $0 \le k \le 2^n - 1$. Dann gilt also

$$\lim_{m \to \infty} P(T^{-m}A \cap I) = P(A)P(I)$$

für alle $I \in \mathcal{A}_0$. Diese Gleichung bleibt richtig, wenn $I$ eine endliche Vereinigung paarweise disjunkter Intervalle dieser Form ist. Diese Figuren bilden eine Algebra, die die Borel-$\sigma$-Algebra auf $[0,1)$ erzeugt. Somit folgt der Satz mit Lemma 4.18. $\square$

**4.20 Satz (Ergodensatz von Birkhoff, 1931)** *Es sei $T$ eine maßerhaltende Transformation auf $(\Omega, \mathcal{A}, P)$ und $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Dann konvergiert*

$$S_n := \frac{1}{n} \sum_{j=0}^{n-1} X \circ T^j$$

*fast sicher für $n \to \infty$ gegen eine $\mathcal{J}$-messbare Zufallsgröße $Y$, für die*

$$\int X \, dP = \int Y \, dP$$

*gilt.*

**4.21 Bemerkungen**

(a) Ist $T$ ergodisch, so ist jede $\mathcal{J}$-messbare Zufallsgröße fast sicher konstant. Wegen $\int X \, dP = \int Y \, dP$ muss $Y = \int X \, dP = \mathbb{E}X$ gelten. Insbesondere folgt aus dem Ergodensatz zusammen mit Satz 4.11 das starke Gesetz der großen Zahlen, siehe Korollar 6.5, Wahrscheinlichkeitstheorie.

(b) Mit $A \in \mathcal{J}$ ist $1_A X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Mit Satz 4.20 existiert eine $\mathcal{J}$-messbare Zufallsgröße $Y_A$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} (1_A X) \circ T^j = Y_A$$

und

$$\int Y_A \, dP = \int 1_A X \, dP = \int_A X \, dP .$$

Ist $A$ invariant, so folgt

$$\frac{1}{n} \sum_{j=0}^{n-1} (1_A X) \circ T^j = \frac{1}{n} 1_A \sum_{j=1}^{n-1} X \circ T^j ,$$

also $Y_A = 1_A Y$ P-f. s. Somit haben wir im nicht-ergodischen Fall gezeigt, dass

$$\int_A X \, dP = \int_A Y \, dP \quad \forall A \in \mathcal{J} ,$$

also $Y = \mathbb{E}(X \mid \mathcal{J})$.

(c) Die Konvergenz in Satz 4.20 gilt auch im 1. Mittel. Siehe etwa Billingsley, „Ergodic Theory and Information".

*Beweis von Satz 4.20* (nach Katznelson, Weiss, 1982)*:* Mit der üblichen Zerlegung $X = X^+ - X^-$ reicht es, den Satz für $X \geq 0$ zu zeigen. Es sei

$$\overline{X}(\omega) := \limsup_{n\to\infty} S_n(\omega) , \quad \underline{X}(\omega) := \liminf_{n\to\infty} S_n(\omega) .$$

Es gilt $\underline{X} \circ T = \underline{X}$ und $\overline{X} \circ T = \overline{X}$, also sind $\underline{X}$ und $\overline{X}$ $\mathcal{J}$-messbar. Man zeigt nun

$$\int \overline{X} \, dP \leq \int X \, dP \leq \int \underline{X} \, dP , \tag{4.2}$$

denn mit $\underline{X} \leq \overline{X}$ folgt dann $\underline{X} = \overline{X}$ fast sicher, also konvergiert $(S_n)_n$ f. s. gegen eine $\mathcal{J}$-messbare Zufallsgröße mit den gewünschten Eigenschaften.

Für den Beweis der 1. Ungleichung in (4.2) verfolge man die Idee: Man schaue auf die „Zeitpunkte" $0 = n_0 < n_1 < n_2 \dots$, für die der durchschnittliche Zuwachs

$$\frac{1}{n_{j+1} - n_j} \left[ X(T^{n_j} \omega) + \cdots + X(T^{n_{j+1}-1} \omega) \right]$$

dem lim sup „nahe" kommt. Es ergeben sich zwei technische Probleme. Erstens kann man nicht ausschließen, dass $\overline{X} = \infty$ ist. Man schneide ab: Für $M \in (0, \infty)$ sei $\overline{X}_M = \min(\overline{X}, M)$. $\varepsilon > 0$ sei beliebig. Für $\omega \in \Omega$ sei $n(\omega) \in \mathbb{N}$ die kleinste Zahl $k \in \mathbb{N}$ mit $S_k(\omega) \geq \overline{X}_M(\omega) - \varepsilon$. Die Abbildung $n : \Omega \to \mathbb{N}$ ist $\mathcal{A}$-$\mathcal{P}(\mathbb{N})$-messbar, denn für $k \in \mathbb{N}$ gilt

$$\{\omega : n(\omega) = k\} = \bigcap_{j=1}^{k-1} \{S_j < \overline{X}_M - \varepsilon\} \cap \{S_k \geq \overline{X}_M - \varepsilon\}\,.$$

$n$ als Funktion von $\omega$ braucht nicht beschränkt zu sein – das zweite technische Problem. Allerdings gilt

$$\lim_{N\to\infty} P\big(\{\omega : n(\omega) > N\}\big) = 0\,.$$

Also existiert ein $N_\varepsilon \in \mathbb{N}$ mit $P\big(\{\omega : n(\omega) > N_\varepsilon\}\big) \leq \varepsilon$. Wir schneiden auch $n(\omega)$ und $X$ ab: Mit $A := \{\omega : n(\omega) \leq N_\varepsilon\}$ seien

$$\tilde{X}(\omega) := \begin{cases} X(\omega)\,, & \omega \in A\,, \\ M & \omega \notin A \end{cases} \qquad \text{und} \qquad \tilde{n}(\omega) := \begin{cases} n(\omega)\,, & \omega \in A\,, \\ 1\,, & \omega \notin A\,. \end{cases}$$

Gilt $X(\omega) > M$, so ist $n(\omega) = 1$, wegen $S_n(\omega) = X(\omega)$. Somit ist dann $\omega \in A$. Dann ist

$$X(\omega) \leq \tilde{X}(\omega) \quad \text{für alle } \omega\,, \tag{4.3}$$

denn im Fall $X(\omega) \leq M$ folgt die Ungleichung für $\omega \in A$ ohnehin und für $\omega \in A^c$ ist $M = \tilde{X}(\omega)$. Für $\omega \notin A$ gilt wegen $\tilde{n}(\omega) = 1$ und $\tilde{X}(\omega) = M$

$$\frac{1}{\tilde{n}(\omega)} \sum_{j=0}^{\tilde{n}(\omega)-1} \tilde{X}\big(T^j(\omega)\big) \geq \overline{X}_M(\omega) - \varepsilon\,. \tag{4.4}$$

Diese Ungleichung gilt auch für $\omega \in A$, denn dort ist $\tilde{n}(\omega) = n(\omega)$ und mit (4.3) ist die linke Seite von (4.4) nicht kleiner als $S_{n(\omega)}(\omega)$, also auch nicht kleiner als $\overline{X}_M(\omega) - \varepsilon$ nach der Definition von $n(\omega)$. Es folgt

$$\int \tilde{X}\, dP = \int_A \tilde{X}\, dP + \int_{A^c} \tilde{X}\, dP \leq \int X\, dP + M\varepsilon\,. \tag{4.5}$$

Nun definiert man rekursiv:

$$\begin{aligned} &n_0(\omega) := 0\,, \quad n_1(\omega) := \tilde{n}(\omega)\,, \\ &n_k(\omega) := n_{k-1}(\omega) + \tilde{n}\big(T^{n_{k-1}(\omega)}(\omega)\big) \quad \text{für } k \in \mathbb{N}\backslash\{1\}\,. \end{aligned}$$

Für $m \in \mathbb{N}$ sei $K(\omega)$ die größte Zahl $k \in \mathbb{N}$ mit $n_k(\omega) \leq m$. $K(\omega)$ hängt somit auch von $m$ ab. Mit $\tilde{n}(\omega) \le N_\varepsilon$ folgt $m - n_{K(\omega)}(\omega) \le N_\varepsilon$. Dann folgt

$$\begin{aligned}
\sum_{j=0}^{m-1} \tilde{X}(T^j\omega) &\ge \sum_{j=0}^{n_{K(\omega)}(\omega)-1} \tilde{X}(T^j\omega) \\
&= \sum_{j=0}^{n_1(\omega)-1} \tilde{X}(T^j\omega) + \sum_{j=n_1(\omega)}^{n_2(\omega)-1} \tilde{X}(T^j\omega) \\
&\qquad + \cdots + \sum_{j=n_{K(\omega)-1}(\omega)}^{n_{K(\omega)}(\omega)-1} \tilde{X}(T^j\omega)\,.
\end{aligned}$$

Nun verwenden wir (4.4), angewandt auf

$$\omega, T^{n_1(\omega)}(\omega), T^{n_2(\omega)}(\omega), \ldots, T^{n_{K(\omega)-1}(\omega)}(\omega)\,,$$

und erhalten

$$\begin{aligned}
&\ge n_1(\omega)\big(\overline{X}_M(\omega) - \varepsilon\big) \\
&\qquad + \big(n_2(\omega) - n_1(\omega)\big)\Big(\overline{X}_M\big(T^{n_1(\omega)}(\omega)\big) - \varepsilon\Big) \\
&\qquad + \cdots + \big(n_{K(\omega)}(\omega) - n_{K(\omega)-1}(\omega)\big) \\
&\qquad\qquad \cdot \Big(\overline{X}_M\big(T^{n_{K(\omega)-1}(\omega)}(\omega)\big) - \varepsilon\Big)\,.
\end{aligned}$$

Für alle $\omega \in \Omega$ und $j \in \mathbb{N}$ ist $\overline{X}_M(T^j\omega) = \overline{X}_M(\omega)$. Also folgt

$$\begin{aligned}
&= n_{K(\omega)}(\omega)\overline{X}_M(\omega) - n_{K(\omega)}(\omega)\varepsilon \\
&\ge m\overline{X}_M(\omega) + \big(n_{K(\omega)}(\omega) - m\big)\overline{X}_M(\omega) - m\varepsilon \\
&\ge m\overline{X}_M(\omega) - N_\varepsilon M - m\varepsilon\,.
\end{aligned}$$

Wir dividieren nun durch $m$ und beachten

$$\int \tilde{X}(T^j\omega)\,P(d\omega) = \int \tilde{X}\,dP \quad (T \text{ maßerhaltend})\,,$$

also

$$\int \tilde{X}\,dP \ge \int \overline{X}_M\,dP - \frac{N_\varepsilon M}{m} - \varepsilon\,.$$

Mit (4.5) folgt

$$\int X\,dP \ge \int \overline{X}_M\,dP - \frac{N_\varepsilon M}{m} - \varepsilon - M\varepsilon\,.$$

Da dies für alle $m \in \mathbb{N}$ und $\varepsilon > 0$ gilt, folgt

$$\int X\,dP \ge \int \overline{X}_M\,dP$$

für alle $M > 0$.

Der Satz von der monotonen Konvergenz liefert

$$\int X \, dP \geq \int \overline{X} \, dP \, .$$

Die zweite Ungleichung folgt analog und ist etwas einfacher, da $\underline{X}$ nicht nach unten abgeschnitten werden muss. □

Der Begriff *Ergodensatz* hat seine Ursprünge in der Physik. Man denke sich $\Omega$ als den Phasenraum eines dynamischen Systems, zur Zeit $n = 0$ sei $\omega$ der Zustand des Systems, zur Zeit $n$ sei er gegeben durch $T^n\omega$, wobei $T : \Omega \to \Omega$ eine Bewegung beschreibe. Es sei nun $X(\omega)$ eine Beobachtung von $\omega$. Tatsächlich mißt man in der Regel nicht $X(\omega)$ (die Beobachtungszeit ist häufig zu lang im Vergleich zur Zeitskala, in der molekulare Wechselwirkungen stattfinden), sondern durchschnittliche Werte von $X(\omega)$, die $\omega$ im Laufe seiner dynamischen Entwicklung annimmt, also

$$\frac{1}{n} \sum_{j=0}^{n-1} X \circ T^j(\omega).$$

J.W. Gibbs stellte in den letzten Jahren des 19. Jahrhunderts die Vermutung auf, dass $T^n(\omega)$ den ganzen Phasenraum durchlaufen sollte und die Dichte der Punkte $T^n(\omega)$ in einer kleinen Umgebung gegen eine Limesdicht konvergieren sollte. Diese Limesverteilung sollte dann invariant unter $T$ sein. Wenn die Limesverteilung $P$ ist, sollte man das obige diskrete *Zeitmittel* duch das *Phasenmittel*

$$\int X(\omega) \, P(d\omega)$$

ersetzen können. Birkhoff hat 1931 gezeigt, dass diese Vermutungen von Gibbs in einer geeigneten Formulierung stimmen. Dabei ist ein wichtiger Punkt, dass $T^n\omega$ niemals über den ganzen Phasenraum verteilt sein wird, wenn es Teilbereiche $A \subset \Omega$ gibt, die unter $T$ invariant sind und $P(A) > 0$ und $P(A^c) > 0$ gilt. Daher ist natürlich die einzige Hoffnung auf Bestätigung der Gibbsschen Vermutung die Annahme, dass invariante Mengen Wahrscheinlichkeit 0 oder 1 haben. Dies führte zu der gegebenen Definition ergodischer maßerhaltender Transformationen.

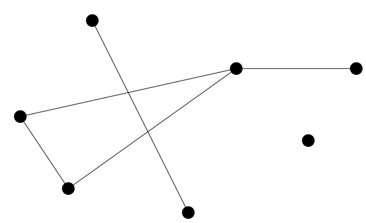
KAPITEL 5

# Zufallsgraphen

In diesem Kapitel wollen wir eine Einführung in die Theorie der zufälligen Graphen geben.

Wir betrachten ein paar Grundbegriffe der Graphentheorie, beschreiben das Modell eines Zufallsgraphen und führen sogenannte Schwellenfunktionen ein. Dieser Begriff wird an Hand einiger Grapheneigenschaften untersucht. Historisch sind die zu beschreibenden Objekte 1959 von Paul Erdős eingeführt worden, und zwar zur Beantwortung einiger Fragen der Graphentheorie. Wir schließen das Kapitel mit dem Beweis eines fundamentalen Satzes von Erdős.

Ein *Graph* ist eine Ansammlung von Punkten (Knoten), wobei manche Punkte durch eine Kante verbunden sind. Etwa



**5.1 Definition** Ein *Graph* ist ein Paar $G = (V, E)$ disjunkter Mengen mit $E \subseteq [V]^2$ (bezeichne die Menge aller 2-elementigen Teilmengen von $V$). Elemente von $V$ nennt man *Ecken* oder *Knoten* (engl. *vertices*) des Graphen, die Elemente von $E$ heißen *Kanten* (engl. *edges*). Der Graph $G$ heißt endlich bzw. unendlich je nachdem, ob $V$ endlich oder unendlich ist. Für $|V|$ schreiben wir auch $|G|$ und nennen es die *Ordnung* von $G$. Zwei Knoten von $G$ sind *benachbart* in $G$, wenn $xy \in E$ ist. $xy$ bezeichnet kurz $\{x, y\}$; $x, y \in V$.

**Zur Historie:** Es ist manchmal schwer – warum auch immer –, einen Graphen mit einer speziellen Eigenschaft zu konstruieren: Gibt es Graphen, die gleichzeitig beliebig große *Taillenweite* und beliebig hohe *chromatische Zahl* haben? Erdős definierte für jedes $n \in \mathbb{N}$ auf der Menge aller Graphen der Ordnung $n$ einen Wahrscheinlichkeitsraum und zeigte, dass bei geeigneter Wahl der Wahrscheinlichkeitsmaße die Wahrscheinlichkeit, dass der gesuchte Graph existiert, für hinreichend große $n$ positiv wird! Man spricht von der *probabilistischen*

*Methode*. Wir werden diese historisch bedeutende Frage später genau untersuchen. Um die Frage aber zumindest formal verstehen zu können, hier die Definition der Begriffe:

**5.2 Definition**

(a) Ein *Weg* ist ein nicht leerer Graph $P = (V, E)$ mit

$$V = \{x_0, x_1, \ldots, x_k\} \quad \text{und} \quad E = \{x_0x_1, x_1x_2, \ldots, x_{k-1}x_k\},$$

wobei die $x_i$ paarweise verschieden sind. Die Anzahl der Kanten eines Weges ist seine *Länge*. Ist $P = x_0 \cdots x_{k-1}$ (Kurzschreibweise) ein Weg und $k \geq 3$, so ist $C := P + x_{k-1}x_0$ ein *Kreis* (wobei $+$ bedeutet: wir vereinen die Kantenmenge bei gleichen $V = \{x_0, \ldots, x_{k-1}\}$). $C = x_0 \cdots x_{k-1}x_0$ (Kurzschreibweise). Ein Kreis der Länge $k$ wird mit $C^k$ bezeichnet.

(b) Die Länge eines kürzesten Kreises in einem Graphen $G$ (bezüglich $\subseteq$: ein Graph $G' = (V', E')$ ist ein Teilgraph von $G$, wenn $V' \subseteq V$ und $E' \subseteq E$ zu $G = (V, E)$) ist die *Taillenweite* $g(G)$ von $G$.

(c) Eine *Eckenfärbung* eines Graphen $G = (V, E)$ ist eine Abbildung $c : V \to S$ mit $c(v) \neq c(w)$ für je zwei benachbarte Ecken $v, w$. $S$ ist die Farbenmenge. Das kleinste $k \in \mathbb{N}$, so dass $G$ eine Eckenfärbung hat ($c : V \to \{1, \ldots, k\}$), nennt man die (Ecken-) *chromatische Zahl* von $G$, $\chi(G)$ in Bezeichnung.

Die oben genannte Frage, dessen Antwort Erdős fand, kann nun wie folgt gestellt werden: Existiert zu jedem $k \in \mathbb{N}$ ein Graph $G$ mit $g(G) > k$ *und* $\chi(G) > k$? Man benötigt Kenntnisse aus der Graphentheorie, um genauer zu verstehen, warum beide Forderungen gleichzeitig einer typischen Konstruktion entgegenstehen. Aus dem Bauch heraus verstehen wir jetzt aber, warum sich die beiden Forderungen an einen zu konstruierenden Graphen in die Quere kommen: Kreise (wie groß sie auch immer sind) benötigen zur Eckenfärbung nur 2 oder 3 Farben. Sie stören also, wenn man eine große Mindestzahl von Farben haben möchte.

Wir wenden uns den Zufallsgraphen zu: Zu $V = \{0, 1, \ldots, n-1\}$ wollen wir die Menge $\mathcal{G}$ der Graphen auf $V$ in einen Wahrscheinlichkeitsraum verwandeln. Dann machen Fragen der Art „mit welcher Wahrscheinlichkeit hat ein Graph $G \in \mathcal{G}$ eine spezielle Eigenschaft“Sinn. Wir betrachten das sogenannte $\mathcal{G}(n, p)$-Modell:

**$\mathcal{G}(n, p)$-Modell:**

Für alle Eckenpaare $e \in [V]^2$ entscheidet man unabhängig voneinander, ob $e$ eine Kante von $G$ sein soll oder nicht. Die Wahrscheinlichkeit, dass $e$

eine Kante sei, soll $0 < p < 1$ sein (zeitliche Interpretation: starte mit einem leeren Graphen auf $n$ Knoten. Jede der $\binom{n}{2}$ Kanten wird mit Wahrscheinlichkeit $p$ eingefügt). Ist $G_0$ ein konkreter Graph auf $V$ mit $m$ Kanten, so hat das Elementarereignis $\{G_0\}$ die Wahrscheinlichkeit

$$p^m \, q^{\binom{n}{2}-m}, \; q := 1-p.$$

Der Wahrscheinlichkeitsraum $\mathcal{G} = \mathcal{G}(n,p)$ ist nun ganz einfach zu konstruieren: zu $e \in [V]^2$ sei $\Omega_e := \{0,1\}$ und $P_e(0) = 1-p$ und $P_e(1) = p$. Dann bezeichne $\mathcal{G}(n,p)$ den Produktraum $(\Omega, P)$ mit

$$\Omega := \prod_{e \in [V]^2} \Omega_e, \quad P \text{ Produktmaß.}$$

Ein $\omega \in \Omega$ identifiziert man mit einem Graphen $G$ auf $V$ und Kantenmenge $E = \{e : \omega(e) = 1\}$. Man sagt nun, $G$ ist ein *Zufallsgraph* auf $V$ mit Kantenwahrscheinlichkeit $p$. Es sei weiter

$$A_e := \{\omega : \omega(e) = 1\}$$

( $e$ ist eine Kante in $G$).

**5.3 Korollar** *Die Ereignisse $(A_e)_e$ sind unabhängige Ereignisse mit Wahrscheinlichkeit $p$.*

Dies ist nach Konstruktion des Produktraumes klar.

Es sei $X : \mathcal{G}(n,p) \to \mathbb{N}$ die Zufallsgröße, die jedem Zufallsgraphen $G$ die Anzahl seiner zu $C^k$ isomorphen Teilgraphen zuordnet. Wir sagen genauer, was dies bedeutet:

**5.4 Definition** $G = (V,E)$ und $G' = (V',E')$ seien zwei Graphen. Gilt $V' \subseteq V$ und $E' \subseteq E$, so ist $G'$ ein *Teilgraph* von $G$. $G$ heißt *isomorph* zu $G'$, $G \simeq G'$ in Zeichen, wenn es eine Bijektion $\varphi : V \to V'$ gibt mit

$$xy \in E \Leftrightarrow \varphi(x)\varphi(y) \in E' \quad \text{für alle } x,y \in V.$$

$\varphi$ heißt *Isomorphismus*.

Um mit den neuen Begriffen ein wenig Übung zu bekommen, zeigen wir

**5.5 Lemma** *Die mittlere Anzahl von Kreisen der Länge $k (\geq 3)$ in $G \in \mathcal{G}(n,p)$ beträgt*

$$\mathbb{E}(X) = \frac{(n)_k}{2k} p^k,$$

*wobei wir $(n)_k := n(n-1)(n-2)\cdots(n-k+1)$ setzen.*

*Beweis:* Für $C := C^k$ mit Ecken aus $V = \{0, \ldots, n-1\}$ sei

$$X_C : G \mapsto \begin{cases} 1, & \text{falls } C \subseteq G \text{ ist,} \\ 0, & \text{sonst.} \end{cases}$$

Dann ist

$$\mathbb{E}(X_C) = P(X_C = 1) = P(\text{ Menge aller Graphen aus } \mathcal{G}(n,p), \text{ die } C \text{ enthalten}).$$

Dies ist nichts anderes als die Wahrscheinlichkeit, dass $C$ ein Kreis in $G$ ist, also $P(C \subseteq G) = p^k$ (da $C$ ein Kreis der Länge $k$ ist). Wie viele dieser Kreise $C = v_0 \cdots v_{k-1} v_0$ gibt es? Es gibt $(n)_k$ Folgen $v_0 \cdots v_{k-1}$ mit (unterschiedlichen) Ecken aus $V$. Jeder Kreis wird durch $2\,k$ dieser Folgen beschrieben: also gibt es $\frac{(n)_k}{2k}$ solche Kreise. Nun ist $X$ die Summe aller $X_C$, wobei $C$ die $\frac{(n)_k}{2k}$ Kreise der Länge $k$ mit Ecken aus $V$ durchläuft, also

$$\mathbb{E}(X) = \sum_C \mathbb{E}(X_C) = \frac{(n)_k}{2k} p^k.$$

Damit ist das Lemma bewiesen. □

**Schwellenfunktionen:**

Man sagt, dass ein Graph $G$ eine *Eigenschaft* $\mathcal{A} \subseteq \mathcal{G}(n,p)$ hat, wenn $G \in \mathcal{A} \subseteq \mathcal{G}(n,p)$ gilt. Erdős und Rényi entdeckten 1960, dass viele Grapheneigenschaften ab einem gewissen $p(n)$ eintreten. Wir lassen nun also $p$ in $n$ variieren. Geht $P(G \in \mathcal{A}) \to 1$ für $n \to \infty$, so sagt man, dass *fast alle* $G$ in $\mathcal{G}(n,p)$ die Eigenschaft $\mathcal{A}$ haben. Strebt die Wahrscheinlichkeit gegen 0, so hat *fast kein* $G$ die Eigenschaft $\mathcal{A}$. (Vorsicht im Vergleich zur üblichen Definition von "fast alle" (fast sicher).)

Ein kleiner Eindruck der *Evolution* eines Zufallsgraphen: ist $p(n)$ wesentlich unterhalb von $1/n^2$, hat fast jeder Graph nur isolierte Ecken. Ab $p(n) = 1/(n\sqrt{n})$ hat fast jeder Graph die ersten Komponenten mit mehr als zwei Ecken (dabei ist eine Komponente ein maximal zusammenhängender nicht leerer Teilgraph, ohne Details). Ab $p(n) = 1/n$ tauchen die ersten Kreise auf. Eine spezielle Komponente beginnt stärker zu wachsen als die anderen und bei etwa $p(n) = \log n/n$ "verschlingt" sie andere und die Graphen werden zusammenhängend.

Die Entwicklung vollzieht sich in Schüben: die genannten $p(n)$-Werte sind Schwellenwerte, unterhalb derer fast kein Graph und oberhalb derer fast jeder Graph die betreffende Eigenschaft hat.

**5.6 Definition** Eine reelle Funktion $t = t(n)$ mit $t(n) \neq 0$ für alle $n \in \mathbb{N}$ heißt *Schwellenfunktion* für eine Grapheneigenschaft $\mathcal{A}$, wenn für jedes $p = p(n)$ und

$G \in \mathcal{G}(n,p)$ gilt:

$$\lim_{n\to\infty} P(G \in \mathcal{A}) = \begin{cases} 0, & \text{falls } p/t \underset{n\to\infty}{\longrightarrow} 0, \\ 1, & \text{falls } p/t \underset{n\to\infty}{\longrightarrow} \infty. \end{cases}$$

Ist $t$ Schwellenfunktion, so natürlich auch $c\,t$ mit $c \in \mathbb{R}_+$. Schwellenfunktionen sind also nur bis auf einen konstanten Faktor eindeutig bestimmt (wenn es sie gibt).

Wir wollen nun eine sehr zentrale Schwellenfunktion bestimmen: Zu einem gegebenen Graphen $H$ bezeichnen wir mit $\mathcal{A}_H$ die Grapheneigenschaft, eine Kopie von $H$ als Teilgraphen zu enthalten:

$$\mathcal{A}_H := \{G : H \subseteq G\}.$$

**5.7 Definition**

(a) Die *Dichte* eines Graphen $G = (V, E)$ ist definiert durch

$$\varrho(G) = \frac{|E|}{|V|}.$$

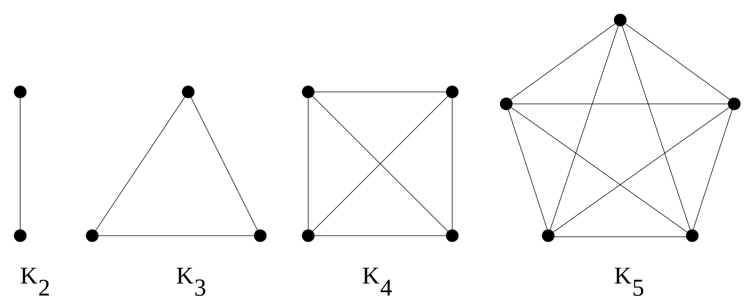
(b) Man nennt $G$ *ausgewogen* oder *balanciert*, wenn für alle Teilgraphen $G'$ von $G$ gilt:

$$\varrho(G') \leq \varrho(G)$$

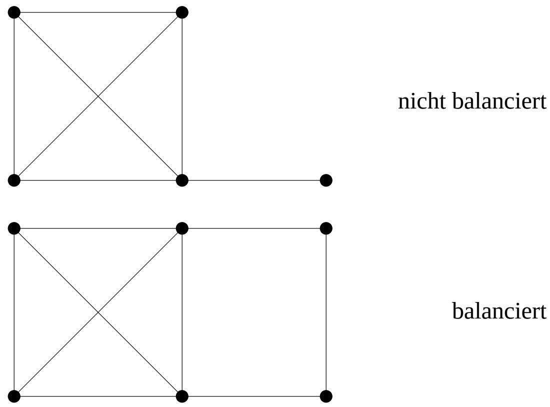
($\mathit{strikt\ ausgewogen}$, *strikt balanciert*, wenn $<$ gilt).

**5.8 Beispiele**

(a) Sind je zwei Ecken von $G$ benachbart, so heißt $G$ *vollständig*. Hat er $k$ Ecken, bezeichnen wir diesen Graphen mit $K_k$. $K_k$ ist strikt balanciert.



(b)




Der erste Graph ist nicht balanciert, denn die Dichte ist $7/5$ und $K_4$ hat Dichte $\frac{3}{2}$. Der zweite ist balanciert, aber nicht strikt (liegt an $K_4$).

**5.9 Satz (von Erdős und Rényi)** *Es sei $H$ ein balancierter Graph mit $k$ Ecken und $l \geq 1$ Kanten. Dann ist $t(n) := n^{-k/l}$ eine Schwellenfunktion für $\mathcal{A}_H$.*

Der Beweis ist recht umfangreich und wird eine sehr wichtige, vielfach angewendete Methode präsentieren. Zuvor sammeln wir ein paar Folgerungen aus diesem Satz:

**5.10 Korollar** *Ist $k \geq 3$ und $\mathcal{A}$ die Eigenschaft, einen Kreis der Länge $k$ als Teilgraph zu enthalten, so ist $t(n) = 1/n$ eine Schwellenfunktion für $\mathcal{A}$.*

*Beweis:* Hier ist $l = k$ und $C^k$ ist balanciert. □

**5.11 Definition** Ein Graph heißt *zusammenhängend*, wenn er für je zwei seiner Ecken $x, y$ einen Weg von $x$ nach $y$ enthält. Ein Graph, der keinen Kreis enthält und zusammenhängend ist, heißt *Baum*. Eine alternative Definition (ohne Beweis): Ein zusammenhängender Graph mit $k$ Ecken ist genau dann ein Baum, wenn er $k-1$ Kanten hat.

**5.12 Korollar** *Ist $T$ ein Baum der Ordnung $k \geq 2$ und $\mathcal{A}$ die Eigenschaft, eine Kopie von $T$ als Teilgraph zu enthalten, so ist $t(n) = n^{-k/k-1}$ eine Schwellenfunktion für $\mathcal{A}$.*

*Beweis:* $k/k-1$ folgt aus der Definition eines Baums. Ein Baum ist balanciert. □

Bemerke, dass die Schwelle beim Kreis unabhängig ist von der Länge des Kreises. Beim Baum ist sie unabhängig von der Gestalt des Baumes.

**5.13 Korollar** *Ist $k \geq 2$ und $\mathcal{A}$ die Eigenschaft, einen $K_k$ als Teilgraph zu enthalten, so ist $t(n) = n^{-2/k-1}$ eine Schwellenfunktion.*

*Beweis:* Es gilt immer

$$\varrho(K_i) = \frac{1}{2}(i-1) < \frac{1}{2}(k-1) = \varrho(K_k)$$

für alle $i < k$. Nun ist $l = \binom{k}{2}$ für $K^k$. Es ist $l = \frac{1}{2}k(k-1)$ und daher ist die Schwelle $n^{-k/l} = n^{-2/k-1}$. □

Die *Cliquenzahl* eines Graphen $G$ ist definiert durch

$$w(G) := \max\{k : \text{ es existiert ein } K_k \text{ in } G\}.$$

Damit ist zum Beispiel die Schwellenfunktion der Eigenschaft

$$\mathcal{A} = \{\text{Existenz einer Clique der Größe mindestens } 4\} = \{w(G) \geq 4\}$$

gleich $n^{-2/3}$. Dies ist eine einfache Monotonie-Überlegung.

Es folgt nun der Beweis von Satz 5.9:

*Vorbereitung:* Es sei $X(G)$ die Anzahl der zu $H$ isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$. Weiter sei $\mathcal{H}$ die Menge aller zu $H$ isomorphen Graphen auf Teilmengen von $\{0,\ldots,n-1\}$ (Eckenmenge der Graphen aus $\mathcal{G}(n,p)$). Formalisiert liefert dies:

$$\mathcal{H} = \{H' : H' \simeq H, V(H') \subseteq \{0,\ldots,n-1\}\}$$

Es sei nun $H' \in \mathcal{H}$ und $G \in \mathcal{G}(n,p)$. Ist $H' \subseteq G$, so bezeichnet dies die Tatsache, dass $H'$ ein Teilgraph von $G$ ist.

Sei $h$ die Anzahl der zu $H$ isomorphen Graphen auf einer festen Menge von $k$ Ecken. Dann ist $h \leq k!$ und somit

$$|\mathcal{H}| = \binom{n}{k} h \leq \binom{n}{k} k! \leq n^k.$$

Zu $p = p(n)$ sei $\gamma = \frac{p}{t}$, also

$$p = \gamma\, n^{-k/l}.$$

Wir müssen nun zeigen:

(a) fast kein $G \in \mathcal{G}(n,p)$ liegt in $\mathcal{A}_H$, falls $\gamma \to 0$ für $n \to \infty$ und

(b) fast alle $G \in \mathcal{G}(n,p)$ liegen in $\mathcal{A}_H$, falls $\gamma \to \infty$ für $n \to \infty$.

*Beweis von Punkt (a):* Hier führen wir den Beweis mit Hilfe der *Methode der ersten Momente*: Dazu eine Erinnerung:

**5.14 Lemma** *Ist $X$ eine nicht-negative, ganzzahlige Zufallsgröße, so ist*

$$P(X > 0) \leq \mathbb{E}(X).$$

*Beweis:* Es gilt

$$\mathbb{E}(X) = \sum_{i \geq 0} i\, P(X = i) \geq \sum_{i \geq 1} P(X = i) = P(X > 0).$$

□

Gilt für eine Folge von Zufallsgrößen $(X_n)_n$, dass $\lim_{n\to\infty} \mathbb{E}(X_n) = 0$, so folgt

$$\lim_{n\to\infty} P(X_n = 0) = 1,$$

und man sagt hierzu, $X_n$ ist fast sicher gleich Null. Wir wollen also $\mathbb{E}(X)$ bestimmen. Es gilt

$$\mathbb{E}(X) = \sum_{H' \in \mathcal{H}} P(H' \subseteq G) = |\mathcal{H}|\, p^l \leq n^k (\gamma n^{-k/l})^l = \gamma^l$$

und mit $\gamma$ geht also auch $\mathbb{E}(X)$ für wachsende $n$ gegen Null. Also liegt fast kein $G$ aus $\mathcal{G}(n,p)$ in $\mathcal{A}_H$.

*Beweis von Punkt (b):* Hier wird uns die *Methode der zweiten Momente* helfen:

**5.15 Lemma** *Ist $X$ eine Zufallsgröße mit Werten in $\mathbb{N}_0$, so gilt*

$$P(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}.$$

*Beweis:*

$$P(X = 0) \leq P\big(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)\big) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}$$

nach der Tschebyschev-Ungleichung. □

Ist nun $\mathrm{Var}(X) = o((\mathbb{E}X)^2)$, so ist $X > 0$ fast sicher. Wir untersuchen also $\mathrm{Var}(X)$: Zunächst ist

$$\mathbb{E}(X^2) = \sum_{(H',H'') \in \mathcal{H}^2} P(H' \cup H'' \subseteq G) = \sum_{(H',H'') \in \mathcal{H}^2} p^{2l - |E(H' \cap H'')|}.$$

Nun ist $\varrho(H' \cap H'') \leq \varrho(H) = l/k$ (balanciert!). Ist $|V(H' \cap H'')| = i$, so folgt $|E(H' \cap H'')| \leq \frac{i\,l}{k}$. Also ist wegen $p < 1$

$$P(H' \cup H'' \subseteq G) \leq p^{2\,l - \frac{i\,l}{k}}.$$

Nun betrachten wir

$$\mathcal{H}_i^2 := \{(H', H'') \in \mathcal{H}^2 : |V(H' \cap H'')| = i\}$$

für $i = 0, \ldots, k$. Wir bestimmen nun

$$\sum_i P(H' \cup H'' \subseteq G) =: A_i,$$

wobei $\sum_i$ die Summe über alle $(H', H'') \in \mathcal{H}_i^2$ bezeichne. Im Fall $i = 0$ sind $H'$ und $H''$ disjunkt und $\{H' \subseteq G\}$ ist somit unabhängig von $\{H'' \subseteq G\}$. Also

$$A_0 \leq \sum_{(H', H'') \in \mathcal{H}^2} P(H' \subseteq G)\, P(H'' \subseteq G) = (\mathbb{E}X)^2.$$

Für $i \geq 1$ ist

$$\sum_i = \sum_{H', H'' \in \mathcal{H} : |V(H' \cap H'')| = i}.$$

Ist $H'$ fest gewählt, so hat die Summe über die $H''$

$$\binom{k}{i}\binom{n-k}{k-i} h$$

Summanden. Damit ist

$$\begin{aligned}
A_i &\leq \sum_{H' \in \mathcal{H}} \binom{k}{i}\binom{n-k}{k-i} h\, p^{2l} p^{-il/k} \\
&= |\mathcal{H}| \binom{k}{i}\binom{n-k}{k-i} h\, p^{2l} \left(\gamma\, n^{-k/l}\right)^{-il/k} \\
&\leq |\mathcal{H}|\, p^l\, c_1\, n^{k-i}\, h\, p^l\, \gamma^{-il/k}\, n^i \\
&= \mathbb{E}(X)\, c_1\, n^k\, h\, p^l\, \gamma^{-il/k} \\
&\leq \mathbb{E}(X)\, c_2 \binom{n}{k} h\, p^l\, \gamma^{-il/k} \\
&= \mathbb{E}(X)^2\, c_2\, \gamma^{-il/k} \\
&\leq \mathbb{E}(X)^2\, c_2\, \gamma^{-l/k}.
\end{aligned}$$

Also folgt mit $c_3 = k\, c_2$:

$$\frac{\mathbb{E}(X^2)}{(\mathbb{E}X)^2} = \left(\frac{A_0}{(\mathbb{E}X)^2} + \sum_{i=1}^{k} A_i \frac{1}{(\mathbb{E}X)^2}\right) \leq 1 + c_3\, \gamma^{-l/k},$$

also

$$\frac{\mathbb{E}(X^2) - (\mathbb{E}X)^2}{(\mathbb{E}X)^2} \leq c_3\,\gamma^{-l/k},$$

und die rechte Seite geht für $\gamma \to \infty$ gegen Null. Somit ist $X(G) > 0$ für fast alle $G \in \mathcal{G}(n,p)$, das heißt fast alle $G \in \mathcal{G}(n,p)$ enthalten eine Kopie von $H$ als Teilgraphen und liegen somit in $\mathcal{A}_H$. Damit ist der Satz bewiesen. $\square$

Für die Anzahl $X(G)$ der zu $H$ isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$ hatten wir $\mathbb{E}(X) = |\mathcal{H}|\,p^l \sim n^k\,p^l$ gesehen. Somit folgt bei der Wahl $\lim_{n\to\infty} p/n^{-k/l} = 0$ für $p = p(n)$, dass $\lim_{n\to\infty} \mathbb{E}(X) = 0$, und für die Wahl $\lim_{n\to\infty} p/n^{-k/l} = \infty$, dass $\lim_{n\to\infty} \mathbb{E}(X) = \infty$ gilt. Dies gilt für alle $H$ mit $k$ Ecken und $l$ Kanten. Trotzdem gilt der folgende Satz:

**5.16 Satz** *Es sei $H$ nicht balanciert mit $k$ Ecken und $l \geq 1$ Kanten. Dann ist $t(n) = n^{-k/l}$ keine Schwellenfunktion für $\mathcal{A}_H$.*

*Beweis:* Es sei $H' \subseteq H$ so gewählt, dass $\varrho(H') > \frac{l}{k}$ ($H$ nicht balanciert). Dann ist

$$\begin{aligned} \mathbb{E}(X') &\sim n^{|V(H')|}\,p^{|E(H')|} \\ &= (n^{|V(H')|/|E(H')|}\,p)^{|E(H')|}, \end{aligned}$$

wobei $X'(G)$ die Anzahl der zu $H'$ isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$ bezeichnet und $|V(H')|$ ($|E(H')|$) die Knotenanzahl (Kantenanzahl) von $H'$. Wir wählen nun $p = p(n)$ so, dass

$$p(n)/n^{-k/l} \to \infty \quad \text{und} \quad p(n)/n^{-|V(H')|/|E(H')|} \to 0$$

(dies ist möglich, da $\frac{|E(H')|}{|V(H')|} > \frac{l}{k}$). Damit folgt aber $\lim\limits_{n\to\infty} \mathbb{E}(X') = 0$, also $\lim\limits_{n\to\infty} P(X' = 0) = 1$. Da weiter $\{X' = 0\} \subseteq \{X = 0\}$, kann $t(n)$ keine Schwellenfunktion für $\mathcal{A}_H$ sein. $\square$

Wir stellen nun die zu Beginn des Kapitels erwähnte *probabilistische Methode* exemplarisch mittels des Beweises des Satzes von Erdős über die Existenz von Graphen mit hoher Taillenweite und chromatischer Zahl dar.

**5.17 Satz (von Erdős, 1959)** *Zu jedem $k \in \mathbb{N}$ gibt es einen Graphen $H$ mit Taillenweite $g(H) > k$ und chromatischer Zahl $\chi(H) > k$.*

Wir bereiten den Beweis vor. Zu einem Graph $G = (V,E)$ heißt eine Teilmenge von $V$ (Eckenmenge) *unabhängig*, wenn ihre Elemente paarweise nicht benachbart sind. Die größte Mächtigkeit einer unabhängigen Eckenmenge in $G$ ist die *Unabhängigkeitszahl* $\alpha(G)$ von $G$.

**5.18 Proposition** Für jedes $k \geq 2$ gilt für die Wahrscheinlichkeit, dass $G \in \mathcal{G}(n,p)$ eine unabhängige Eckenmenge der Mächtigkeit $k$ enthält,

$$P\big(\alpha(G) \geq k\big) \leq \binom{n}{k}(1-p)^{\binom{k}{2}}.$$

*Beweis:* Die Wahrscheinlichkeit, dass eine fest gewählte Eckenmenge $U \in [V]^k$ in $G$ unabhängig ist, beträgt $(1-p)^{\binom{k}{2}}$. Es gibt nur $\binom{n}{k}$ solche Mengen $U$. □

**5.19 Bemerkung** (analog zur Proposition zu beweisen)

$$P(\omega(G) \geq k) \leq \binom{n}{k}p^{\binom{k}{2}}.$$

(Wahrscheinlichkeit, dass $G$ in $\mathcal{G}(n,p)$ einen $K_k$ enthält; Erinnerung: $\omega(G)$ ist die Cliquenzahl).

Wir nennen Kreise der Länge $\leq k$ *kurz* und Mengen von $\frac{|G|}{k}$ oder mehr Knoten von $G$ *groß*. Wir suchen (Erdős suchte) einen Graphen $G$ *ohne kurze Kreise* und *ohne große unabhängige Eckenmengen*: die letzt genannte Eigenschaft impliziert, dass wir mehr als $k$ Farben brauchen, um $G$ zu färben (denn: $\alpha(G) < \frac{|G|}{k}$, also $k < \frac{|G|}{\alpha(G)} \leq \chi(G)$). Die letzte Ungleichung ist eine bekannte Schranke für die chromatische Zahl, die man wie folgt einsehen kann: Die Knoten einer Knotenmenge $\tilde{V}$ von $G$ können genau dann mit derselben Farbe belegt werden, wenn $\tilde{V}$ eine unabhängige Knotenmenge ist. Es sei eine Färbung von $G$ mit $k = \chi(G)$ Farben gegeben. Wir ordnen jeder Farbe $i$ die Menge $V_i \subset V$ der mit $i$ gefärbten Knoten zu. Dann sind alle Mengen $V_1, \ldots, V_k$ unabhängig. Es gilt jeweils $|V_i| \leq \alpha(G)$, also

$$|G| = \sum_{i=1}^{k} |V_i| \leq \alpha(G)\,\chi(G),$$

woraus die behauptet Ungleichung folgt. Wählt man $p$, die Kantenwahrscheinlichkeit, klein genug, so werden die Zufallsgraphen aus $\mathcal{G}(n,p)$ mit hoher Wahrscheinlichkeit keine (kurzen) Kreise haben (wir kennen bereits die Schwelle $p(n) = 1/n$). Intuitiv erwartet man bei $p$ groß genug, dass kaum große unabhängige Eckenmengen auftreten. Es ergibt sich somit die folgende Frage: Kann man $p$ so wählen, dass es für große $n$ gleichzeitig klein genug für $P(g \leq k) < \frac{1}{2}$ und groß genug für $P(\alpha \geq n/k) < \frac{1}{2}$ ist? Wenn ja, so enthält $\mathcal{G}(n,p)$ mindestens einen Graphen ohne kurze Kreise und ohne große unabhängige Eckenmengen. Tatsächlich kann man $p$ so nicht wählen, da $1/n$ auch eine Schwellenfunktion für große unabhängige Mengen ist. Der Beweis des Satzes verfolgt eine andere Idee.

*Beweis von Satz 5.17:* Es sei $\varepsilon > 0$ und $\varepsilon < 1/k$ fest gewählt. Sei $p := p(n) = \frac{n^\varepsilon}{n} = \frac{1}{n^{1-\varepsilon}}$. Wir liegen oberhalb der Schwelle $1/n$ und kümmern uns um die Anzahl $X(G)$ der kurzen Kreise in einem Zufallsgraphen $G \in \mathcal{G}(n,p)$ (Anzahl der Kreise der Länge $\leq k$). Es gilt mit Lemma 5.5:

$$\mathbb{E}(X) = \sum_{i=3}^{k} \frac{(n)_i}{2i} p^i \leq \frac{1}{2} \sum_{i=3}^{k} p^i\, n^i \leq \frac{1}{2}(k-2) n^k\, p^k,$$

denn $(n\,p)^i \leq (n\,p)^k$ wegen $n\,p = n^\varepsilon \geq 1$. Daraus folgt

$$P\left(X \geq \frac{n}{2}\right) \leq \frac{\mathbb{E}(X)}{n/2} \leq (k-2)\, n^{k-1}\, p^k = (k-2)\, n^{k-1}\, n^{(\varepsilon-1)k} = (k-2)\, n^{k\,\varepsilon-1}.$$

Nun ist $k\,\varepsilon - 1 < 0$ nach obiger Wahl, also

$$\lim_{n\to\infty} P\left(X \geq \frac{n}{2}\right) = 0.$$

Wir wählen $n$ so groß, dass $P\big(X \geq n/2\big) < 1/2$. Wir finden also einen Graphen $G \in \mathcal{G}(n,p)$ mit weniger als $n/2$ kurzen Kreisen.

*Behauptung:* $n$ kann so groß gewählt werden, dass $P\big(\alpha \geq \frac{n}{2k}\big) < 1/2$ bei obiger Wahl von $p(n) = \frac{1}{n^{1-\varepsilon}}$ gilt. Wir finden dann einen Graphen $G \in \mathcal{G}(n,p)$ mit $\alpha(G) < \frac{n}{2k}$ und mit weniger als $n/2$ kurzen Kreisen. Dann sind wir aber bereits am Ziel:

Aus *jedem* kurzen Kreis von $G$ entfernen wir eine Ecke. Der entstehende Graph $H$ hat dann noch mindestens $n/2$ Ecken. Er enthält keine kurzen Kreise mehr, also ist $g(H) > k$ und weiter gilt:

$$\chi(H) \geq \frac{|H|}{\alpha(H)} \geq \frac{n/2}{\alpha(G)} > k$$

(die Unabhängigkeitszahl ist nach Konstruktion nicht größer als die von $G$). Damit ist der Satz bewiesen, wenn obige Behauptung nachgeliefert wird. $\square$

**5.20 Lemma** *Es seien $k > 0$ und $p = p(n)$ gegeben. Ist $p \geq \frac{6k\,\log n}{n}$ für große $n$, so gilt*

$$\lim_{n\to\infty} P\left(\alpha \geq \frac{n}{2k}\right) = 0.$$

Mit Hilfe dieses Lemmas folgt dann für $p = n^{\varepsilon-1}$ für beliebig kleines $\varepsilon > 0$, dass ein $G \in \mathcal{G}(n,p)$ existiert mit $\alpha(G) < n/(2k)$.

*Beweis von Lemma 5.20:* Es seien $n, r \in \mathbb{N}$ mit $n \geq r \geq 2$. Weiter sei $G \in \mathcal{G}(n,p)$. Mit $q := 1-p$ gilt nach Proposition 5.18

$$\begin{aligned} P(\alpha \geq r) &\leq \binom{n}{r} q^{\binom{r}{2}} \leq n^r\, q^{\binom{r}{2}} \\ &= \left(n\, q^{\frac{r-1}{2}}\right)^r \leq \left(n\, e^{-\frac{p(r-1)}{2}}\right)^r, \end{aligned}$$

wobei wir $(1-p) \leq e^{-p}$, $0 < p < 1$, verwendet haben. Es sei nun $p \geq \frac{6k \log n}{n}$ und $r \geq \frac{n}{2k}$, so folgt

$$\begin{aligned} n\, e^{-p(r-1)/2} &= n\, e^{-pr/2+p/2} \leq n\, e^{-\frac{3}{2}\log n + p/2} \\ &\leq n\, n^{-3/2} e^{1/2} = \frac{\sqrt{e}}{\sqrt{n}} \underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Mit $r := [\frac{n}{2k}]$ ist

$$\lim_{n\to\infty} P\left(\alpha \geq \frac{n}{2k}\right) = \lim_{n\to\infty} P(\alpha \geq r) = 0,$$

womit das Lemma bewiesen ist. $\square$

KAPITEL 6

# Perkolation

Wir werfen einen porösen Stein in Wasser und fragen nach der Wahrscheinlichkeit, dass das Zentrum des Steins nass wird. Broadbent und Hammersley haben 1957 ein stochastisches Modell zu dieser Frage entwickelt. Es ist die Geburtsstunde des Perkolationsmodells.

In $\mathbb{Z}^d$, $d \geq 1$, sei

$$\delta(x,y) := \sum_{i=1}^{d} |x_i - y_i| \,, \quad x, y \in \mathbb{Z}^d \,.$$

Wir verbinden zwei „Knoten" $x, y \in \mathbb{Z}^d$ durch eine Kante, wenn $\delta(x,y) = 1$. Es entsteht der Graph $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$, wenn $\mathbb{E}^d$ die Menge aller Kanten bezeichnet. Es sei $p$ und $q$ durch $0 \leq p \leq 1$ und $p + q = 1$ festgelegt. Wir nennen eine Kante in $\mathbb{L}^d$ *offen* mit Wahrscheinlichkeit $p$ und *geschlossen* sonst, unabhängig von allen anderen Kanten. Formal meinen wir

$$\Omega = \prod_{e \in \mathbb{E}^d} \{0,1\} \,;$$

Punkte in $\Omega$ heißen Konfigurationen $\omega = \big(\omega(e) : e \in \mathbb{E}^d\big)$, $\omega(e) = 0$ korrespondiere zu einer geschlossenen Kante $e$, $\omega(e) = 1$ zu einer offenen Kante $e$. $\mathcal{A}$ sei die $\sigma$-Algebra erzeugt von den endlich-dimensionalen Projektionen, und das W-Maß sei gegeben durch

$$P_p = \prod_{e \in \mathbb{E}^d} \mu_e \,, \quad \mu_e\big(\omega(e) = 0\big) = q \,, \quad \mu_e\big(\omega(e) = 1\big) = p \,.$$

Es liegt eine eins-zu-eins Korrespondenz zwischen $\Omega$ und der Menge der Teilmengen von $\mathbb{E}^d$ vor.

Für $\omega \in \Omega$ sei

$$K(\omega) := \{e \in \mathbb{E}^d : \omega(e) = 1\}$$

die Menge der offenen Kanten zu $\omega$. Es gilt $\omega_1 \leq \omega_2$ genau dann, wenn $K(\omega_1) \subseteq K(\omega_2)$. Es liegt uns also ein spezieller Zufallsgraph vor (siehe Abb. 6.1).

Angenommen, $\big(X(e), e \in \mathbb{E}^d\big)$ sei eine Familie unabhängiger Zufallsgrößen und $X(e)$ ist gleichmäßig verteilt auf $[0,1]$. Wir „koppeln" nun *alle* Kanten-Perkolationsprozesse auf $\mathbb{L}^d$, wenn $p$ das Intervall $[0,1]$ durchläuft. Sei $0 \leq p \leq 1$ und

$$\eta_p(e) = \begin{cases} 1 \,, & \text{wenn } X(e) < p \,, \\ 0 \,, & \text{wenn } X(e) \geq p \,. \end{cases}$$

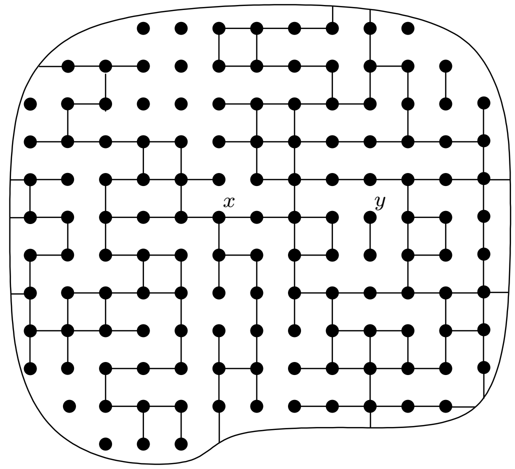



Abb. 6.1: Skizze der Struktur eines zweidimensionalen porösen Steins. Die Linien deuten die offenen Kanten an; geschlossene Kanten wurden weggelassen. Beim Eintauchen des Steins in Wasser wird der Knoten $x$ durch das Eindringen des Wassers nass, während der Knoten $y$ trocken bleibt.

$e$ nennen wir $p$-*offen*, wenn $\eta_p(e) = 1$. Es gilt $P\big(\eta_p(e) = 0\big) = 1 - p$, $P\big(\eta_p(e) = 1\big) = p$, und die Komponenten von $\eta_p$ sind unabhängig. Wenn $p_1 \leq p_2$, so ist $\eta_{p_1} \leq \eta_{p_2}$. Mit Kopplung ist hierbei gemeint, dass wir den Grundraum $\Omega^{[0,1]}$ mit der Produkt-$\sigma$-Algebra und dem Produktmaß $\otimes_{p\in[0,1]} P_p$ versehen und diesen gemeinsamen W-Raum verwenden.

Wir nennen einen Weg oder Kreis *offen*, wenn alle seine Kanten offen sind, und *geschlossen*, wenn alle Kanten geschlossen sind. Teilgraphen von $\mathbb{L}^d$ heißen *Kanten-disjunkt*, wenn sie keine gemeinsamen Kanten haben, disjunkt, wenn sie keine gemeinsamen Kanten und keine gemeinsamen Knoten haben.

Nun betrachten wir den Zufallsgraph in $\mathbb{L}^d$, der aus den Knoten $\mathbb{Z}^d$ und den offenen Kanten besteht. Zusammenhangskomponenten dieses Graphen heißen *offene Cluster*.

$C(x)$ nennen wir den offenen Cluster, der $x$ (Knoten) enthält. $C := C(0)$. $C(x)$ steht auch für die Knotenmenge des Clusters, $|C(x)|$ ist die Ordnung des Clusters. Für zwei Teilmengen $A$ und $B$ von Knoten in $\mathbb{L}^d$ schreiben wir $A \leftrightarrow B$, wenn ein offener Pfad/Weg existiert, der einen Knoten in $A$ mit einem in $B$ verbindet.

$$C(x) = \{y \in \mathbb{Z}^d : x \leftrightarrow y\} .$$

Wir interessieren uns für die *Perkolations-Wahrscheinlichkeit* $\Theta(p)$. Dies ist die Wahrscheinlichkeit, dass ein gegebener Knoten zu einem unendlichen Cluster gehört. Das Gitter und $P_p$ sind translationsinvariant. Daher setzen wir

$$\Theta(p) := P_p(|C| = \infty) = 1 - \sum_{n=1}^{\infty} P_p(|C| = n) .$$

**6.1 Lemma** $\Theta$ *wächst monoton und es gilt*

$$\Theta(0) = 0 \quad \textit{und} \quad \Theta(1) = 1\,.$$

*Beweis:* Die obige Kopplung durch Konstruktion der $\eta_p$ liefert: Für $p_1 \leq p_2$ folgt $\eta_{p_1} \leq \eta_{p_2}$. Wir nennen eine Zufallsgröße $N$ auf $(\Omega, \mathcal{A})$ wachsend, wenn $N(\omega) \leq N(\omega')$ für $\omega \leq \omega'$ folgt. Für $p_1 \leq p_2$ folgt dann $N(\eta_{p_1}) \leq N(\eta_{p_2})$. Nun wählen wir den Produktraum

$$(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, P_{p_1} \otimes P_{p_2}).$$

Dann folgt

$$\mathbb{E}_{P_{p_1}}(N) \leq \mathbb{E}_{P_{p_2}}(N).$$

Nun ist $\{|C| = \infty\}$ ein wachsendes Ereignis in dem Sinne, dass $\omega \in \{|C| = \infty\}$ impliziert dass $\omega' \in \{|C| = \infty\}$ für jedes $\omega \leq \omega'$. Somit ist $1_{\{|C|=\infty\}}$ wachsend, also folgt

$$\Theta(p_1) = P_{p_1}(|C_{p_1}| = \infty) \leq P_{p_2}(|C_{p_2}| = \infty) = \Theta(p_2)\,. \qquad \square$$

Sei nun

$$p_c(d) := \sup\{p : \Theta(p) = 0\}$$

die *kritische Wahrscheinlichkeit*. Es gilt $\Theta(p) = 0$ für $p < p_c$ und $\Theta(p) > 0$ für $p > p_c$. Betrachten wir den Fall $d = 1$. Für $p < 1$ existieren unendlich viele geschlossene Kanten in $\mathbb{L}^1$ links und rechts der Null fast sicher, also $\Theta(p) = 0$ für $p < 1$, also $p_c(1) = 1$. Sei nun fortan $d \geq 2$.

$\mathbb{L}^d$ können wir in $\mathbb{L}^{d+1}$ einbetten (Projektion auf die ersten $d$ Koordinaten). Dann gehört die Null in $\mathbb{L}^{d+1}$ einem unendlichen offenen Cluster für ein $p$ an, wenn die Null dies in $\mathbb{L}^d$ erfüllt. Also ist $\Theta(p) = \Theta_d(p)$ wachsend in $d$, also

$$p_c(d+1) \leq p_c(d)$$

für $d \geq 1$. Es gilt sogar strikte $<$-Relation (ohne Beweis).

Wir wollen die folgenden Sätze betrachten:

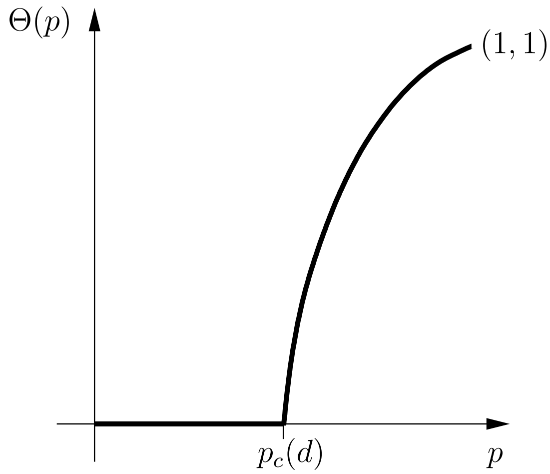
**6.2 Satz** *Für $d \geq 2$ ist $0 < p_c(d) < 1$.*

**6.3 Satz** *Die Wahrscheinlichkeit $\psi(p)$ für die Existenz eines unendlichen offenen Clusters genügt*

$$\psi(p) = \begin{cases} 0\,, & \textit{wenn } \Theta(p) = 0\,, \\ 1\,, & \textit{wenn } \Theta(p) > 0\,. \end{cases}$$

### 6.4 Bemerkungen

(a) Man glaubt den folgenden Verlauf der Perkolations-Wahrscheinlichkeit $\Theta(p)$:




$\Theta$ ist stetig in $p$, außer in $p_c(d)$. Für $3 \leq d < 19$ ist es offen, ob $\Theta$ in $p_c(d)$ springt!

(b) In der *subkritischen Phase* $p < p_c(d)$ ist jeder Knoten (fast sicher) in einem endlichen Cluster. In der *superkritischen Phase* $p > p_c(d)$ hat jeder Knoten strikt positive Wahrscheinlichkeit, in einem unendlichen offenen Cluster zu sein. Es existiert also fast sicher mindestens ein unendlicher offener Cluster. In $p = p_c(d)$ existiert ein unendlicher offener Cluster, wenn $\Theta\big(p_c(d)\big) > 0$. Es ist bekannt, dass für $d = 2$ und $d \geq 19$ *kein* unendlicher offener Cluster existiert, die verbleibenden Dimensionen sind ungeklärt! Man erwartet, dass kein unendlicher offener Cluster existiert.

(c) Wir werden Satz 6.3 mittels eines 0-1-Gesetzes beweisen. Der Satz sagt nichts über die Anzahl von unendlichen offenen Clustern im Fall $\Theta(p) > 0$ aus. Tatsächlich ist aber bekannt, dass der unendliche offene Cluster fast sicher eindeutig ist, wenn er existiert.

(d) $p_c(1) = 1$. Man kennt auch $p_c(2) = 1/2$ (H. Kesten, 1980); $p_c(d)$, $d \geq 3$, sind nicht bekannt! Wir werden im Laufe des Beweises von Satz 6.2 untere und obere Schranken für $p_c(d)$, $d \geq 2$, kennenlernen.

Wir bereiten den Beweis von Satz 6.2 vor. Dazu führen wir den Begriff *selbst-meidender Pfade (self-avoiding)* ein. Ein $n$-Schritt selbst-meidender Pfad $\omega$ in $\mathbb{Z}^d$ mit Start in 0 ist durch $\big(\omega(0) = 0, \omega(1), \ldots, \omega(n)\big)$ mit $\delta\big(\omega(j+1), \omega(j)\big) = 1$ und $\omega(i) \neq \omega(j)$ für alle $i \neq j$ gegeben. $\sigma(n)$ bezeichne die Anzahl der $n$-Schritt selbst-meidenden Pfade mit Start in 0, so ist

$$\sigma(1) = 2d\,, \quad \sigma(2) = 2d(2d-1)\,, \quad \sigma(3) = 2d(2d-1)^2$$

und

$$\sigma(4) = 2d(2d-1)^3 - 2d(2d-2) \quad (!)$$

$\sigma(n)$ ist allgemein schwierig zu bestimmen, aber es gilt

$$\sigma(n) \leq 2d(2d-1)^{n-1}\,, \tag{6.1}$$

denn in jedem neuen Schritt selbst-meidendereines selbst-meidenden Pfades hat man maximal $2d-1$ Schrittmöglichkeiten, da man die derzeitige Position meiden muss. Weiter gilt

$$d^n \leq \sigma(n)\,. \tag{6.2}$$

Dazu betrachten wir einfach die Zahl der Pfade, bei denen jeder Schritt in eine der $d$ positiven Koordinatenrichtungen erfolgt (diese Pfade sind selbst-meidend). Sollte der Limes $\lim_{n\to\infty}\big(\sigma(n)^{1/n}\big)$ existieren, so besagen (6.1) und (6.2)

$$d \leq \lim_{n\to\infty} \sigma(n)^{1/n} =: \lambda(d) \leq 2d-1\,. \tag{6.3}$$

**6.5 Lemma**

(a) *Für* $n, m \in \mathbb{N}$ *gilt*

$$\sigma(n+m) \leq \sigma(n)\cdot\sigma(m)\,,$$

*also*

$$\log\sigma(n+m) \leq \log\sigma(n) + \log\sigma(m)\,.$$

(b) $\lambda(d) = \lim_{n\to\infty}\sigma(n)^{1/n}$ *existiert und*

$$\log\lambda(d) = \inf_{n\geq 1}\frac{1}{n}\log\sigma(n)\,,$$

*also*

$$\lambda(d)^n \leq \sigma(n)\,, \quad n \geq 1\,.$$

*Beweis:* Das Produkt $\sigma(n)\cdot\sigma(m)$ ist die Anzahl der Menge der Pfade der Länge $n+m$, die selbst-meidend bei den ersten $n$ Schritten und bei den letzten $m$ Schritten sind, die aber nicht notwendig insgesamt selbst-meidend sind, also

$$\sigma(n+m) \leq \sigma(n)\sigma(m)\,.$$

Der Teil (b) folgt aus einem einfachen Argument der Analysis und geht auf Hammersley und Morton (1954) zurück:

**6.6 Lemma** *Sei* $(a_n)_n$ *eine Folge reeller Zahlen, die* subadditiv *ist, d. h.* $a_{n+m} \leq a_n + a_m$, $n, m \in \mathbb{N}$. *Dann existiert* $\lim_{n\to\infty} a_n/n$ *in* $[-\infty, +\infty)$ *und*

$$\lim_{n\to\infty}\frac{a_n}{n} = \inf_{n\geq 1}\frac{a_n}{n}\,.$$

Somit ist für $a_n = \log \sigma(n)$ Lemma 6.5 bewiesen. $\square$

*Beweis von Lemma 6.6:* Wir zeigen

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \frac{a_k}{k} \quad \forall k\,, \tag{6.4}$$

denn dann folgt mit $\liminf_{k\to\infty}$ in (6.4) die Existenz des Grenzwertes und die Wahl von $\inf_{k\ge 1}$ in (6.4) liefert die zweite Behauptung.

Sei $k$ fest und $A_k := \max_{1\le r\le k} a_r$. Für $n \in \mathbb{N}\backslash\{0\}$ sei $j$ die größte Zahl strikt kleiner $n/k$. Dann ist $n = jk+r$ mit $1 \le r \le k$, $r \in \mathbb{N}$. Die Subadditivität liefert

$$\frac{a_n}{n} \le \frac{ja_k + a_r}{jk+r}.$$

Dann folgt im $\limsup_{n\to\infty}$

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \limsup_{j\to\infty} \Big(\frac{k}{a_k} + \frac{r}{ja_k}\Big)^{-1} = \frac{a_k}{k}.$$

Die Übereinstimmung mit dem Infimum zeigt $\lim_{n\to\infty} n^{-1}a_n < \infty$. Die Wahl $a_n = -n^2$ zeigt, dass $-\infty$ nicht ausgeschlossen werden kann. $\square$

*Beweis von Satz 6.2:* Wir sahen bereits $p_c(d+1) \le p_c(d)$, $d \ge 1$. Daher zeigen wir $p_c(d) > 0$, $d \ge 2$, und $p_c(2) < 1$.

*Teil (a)* (Beweis $p_c(d) > 0$): Wir zeigen $\Theta(p) = 0$ für $p$ hinreichend nahe bei 0:

$N(n)$ sei die Anzahl der offenen Pfade der Länge $n$ mit Start in 0, dann ist

$$\mathbb{E}_p\big(N(n)\big) = p^n \sigma(n)\,.$$

Wenn 0 einem unendlichen Cluster angehört, so existieren offene Pfade jeder Länge in 0:

$$\Theta(p) \le P_p\big(N(n) \ge 1\big) \le \mathbb{E}_p\big(N(n)\big) = p^n\sigma(n)$$

für alle $n \in \mathbb{N}$. Nun ist nach Lemma 6.5

$$\sigma(n) = \big(\lambda(d) + o(1)\big)^n \quad \text{für } n \to \infty\,,$$

also

$$\Theta(p) \le \big(p\lambda(d) + o(1)\big)^n \overset{n\to\infty}{\longrightarrow} 0\,,$$

falls $p\,\lambda(d) < 1$. Also ist $p_c(d) \ge 1/\lambda(d)$ mit $\lambda(d) \le 2d-1 < \infty$ (siehe Broadbent, Hammersley).

*Teil (b)* (Beweis $p_c(2) < 1$): Wir wollen $\Theta(p) > 0$ für $p$ hinreichend nahe bei 1 zeigen.

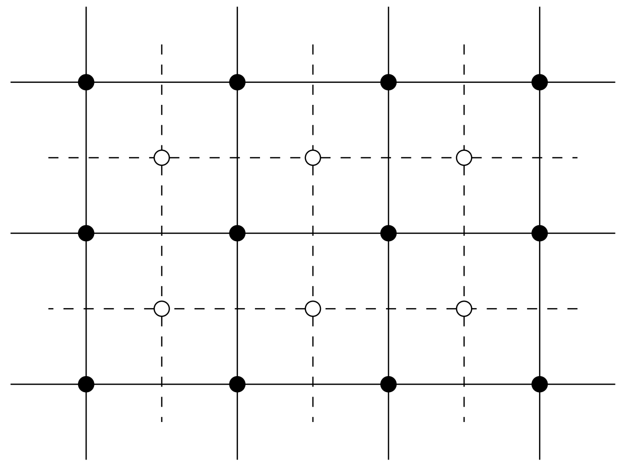


Abb. 6.2: Teil des quadratischen Gitters $\mathbb{L}^2$ zusammen mit seinem Dualen.

Nun betrachten wir das sogenannte *duale Gitter*. Die Knoten sind

$$\Big\{x + \Big(\frac{1}{2}, \frac{1}{2}\Big) : x \in \mathbb{Z}^2\Big\}\,,$$

die Kanten sind geradlinige Verbindungen in $\mathbb{R}^2$, die zwei benachbarte Knoten verbinden (siehe Abb. 6.2).

Dies ergibt eine eins-zu-eins Korrespondenz zwischen den Kanten in $\mathbb{L}^2$ und denen des dualen Graphen. Eine Kante im dualen Gitter heißt *offen*, bzw. *geschlossen*, wenn sie eine offene bzw. geschlossene Kante in $\mathbb{L}^2$ schneidet. Dies liefert ein Kanten-Perkolationsmodell auf dem dualen Graphen mit Kanten-Wahrscheinlichkeit $p$.

Angenommen, der offene Cluster im Ursprung von $\mathbb{L}^2$ sei endlich. Dann ist der Ursprung von einer geschlossenen Kette geschlossener Kanten im dualen Graphen umgeben. Man versuche als Übung, dies rigoros zu notieren (siehe Abb. 6.3).

Es gilt sogar: $|C| < \infty$ genau dann, wenn der Ursprung in $\mathbb{L}^2$ im Inneren eines geschlossenen Kreises im Dualgraph liegt.

Sei nun $\varrho(n)$ die Anzahl der Kreise im Dualgraph mit Länge $n$, die im Inneren den Ursprung von $\mathbb{L}^2$ enthalten. Ein solcher Kreis passiert einen Knoten der Form $(k + 1/2, 1/2)$ für ein $k$ mit $0 \leq k < n$, denn er umrundet den Ursprung, also existiert so ein $(k + 1/2, 1/2)$ für ein $k \geq 0$, und er passiert nicht $(k+1/2, 1/2)$ für ein $k \geq n$, denn sonst wäre seine Länge mindestens $2n$. Also enthält ein solcher Kreis einen selbst-meidenden Pfad der Länge $n - 1$, der in einem Knoten der Form $(k + 1/2, 1/2)$, $0 \leq k < n$, startet. Die Anzahl dieser Pfade ist maximal $n\,\sigma(n - 1)$, also

$$\varrho(n) \leq n\,\sigma(n - 1)\,.$$

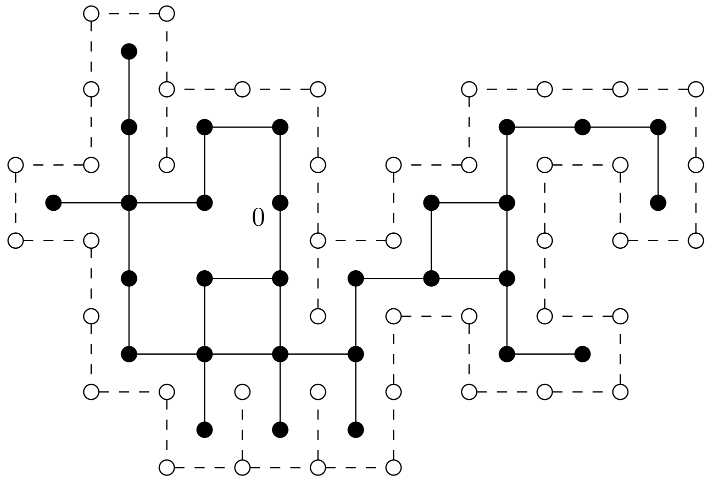



Abb. 6.3: Ein endlicher offener Cluster im Ursprung, umgeben von einer geschlossenen Kette im dualen Gitter.

Nun folgt

$$\begin{aligned}\sum_{\gamma} P_p(\gamma \,\text{ist geschlossen}) &\leq \sum_{n=1}^{\infty} q^n \, n \, \sigma(n-1) \\ &= \sum_{n=1}^{\infty} q \, n \, \{q\lambda(2) + o(1)\}^{n-1} < \infty\end{aligned}$$

für $q\,\lambda(2) < 1$, wobei $q = 1-p$ ist und wir über alle Kreise $\gamma$ des Dualgraphen, die den Ursprung von $\mathbb{L}^2$ im Inneren haben, summieren.

Weiter ist

$$\sum_{\gamma} P_p(\gamma \,\text{ist geschlossen}) \to 0$$

für $q = 1-p \to 0$, also existiert ein $0 < r < 1$ mit

$$\sum_{\gamma} P_p(\gamma \,\text{ist geschlossen}) \leq \frac{1}{2} \quad \text{für}\, p > r\,,$$

Es bezeichne weiter $M(n)$ die Anzahl der geschlossenen Kreise im Dualgraph mit Länge $n$, die im Inneren den Ursprung von $\mathbb{L}^2$ enthalten. Es folgt

$$\begin{aligned}P_p(|C| = \infty) &= P_p\big(M(n) = 0 \,\forall n \in \mathbb{N}\big) \\ &= 1 - P_p\big(M(n) \geq 1 \,\text{für ein}\, n \in \mathbb{N}\big) \\ &\geq 1 - \sum_{\gamma} P_p(\gamma \,\text{ist geschlossen}) \geq \frac{1}{2}\end{aligned}$$

für $p > r$, also $p_c(2) \leq r$.

Tatsächlich kann man sogar

$$\frac{1}{\lambda(2)} \leq p_c(2) \leq 1 - \frac{1}{\lambda(2)}$$

zeigen. Der Beweis der rechten Ungleichung wird in den Übungen diskutiert. $\square$

*Beweis von Satz 6.3:* Mit dem 0-1-Gesetz von Kolmogorov folgt, dass $\psi$ nur die Werte 0 und 1 annimmt. Wenn $\Theta(p) = 0$ ist, so folgt

$$\psi(p) = \sum_{x \in \mathbb{Z}^d} P_p(|C(x)| = \infty) = 0\,.$$

Wenn $\Theta(p) > 0$, so folgt

$$\psi(p) \geq P_p(|C| = \infty) > 0\,,$$

also $\psi(p) = 1$. $\square$

**6.7 Bemerkungen**

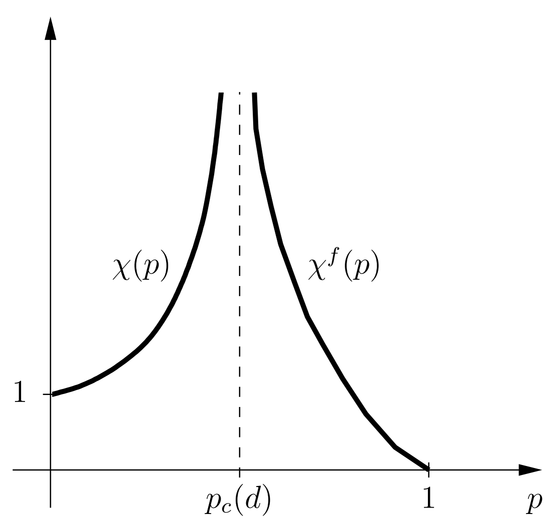
(a) Eine weitere interessante Größe ist

$$\chi(p) := \mathbb{E}_p(|C|)\,,$$

die mittlere Größe eines offenen Clusters. Es ist

$$\begin{aligned}\chi(p) &= \infty P_p(|C| = \infty) + \sum_{n \geq 1} n P_p(|C| = n) \\ &= \infty \Theta(p) + \sum_{n \geq 1} n P_p(|C| = n)\,,\end{aligned}$$

also $\chi(p) = \infty$ für $p > p_c$. Dass $\psi(p) < \infty$ ist für $p < p_c$, ist keinesfalls einfach zu zeigen. Qualitativ sieht $\chi(p)$ so aus:



Wobei $\chi^f(p) := \mathbb{E}_p(|C|; |C| < \infty)$.

(b) Für $p < p_c$ sind die offenen Cluster fast sicher endlich. Weiter gilt dort: Es exisitert ein $\alpha(p)$ mit

$$P_p(|C| = n) \approx e^{-n\,\alpha(p)}$$

für $n \to \infty$, $\alpha(p) > 0$. Hierbei bezeichnet $\approx$ die Eigenschaft, dass $\log P_p(|C| = n)/(-n\,\alpha(p)) \to 1$ für $n \to \infty$. Somit hat $|C|$ endliche Momente jeder Ordnung.

(c) Für $p > p_c$ existiert ein unendlicher Cluster fast sicher: Man kann zeigen, dass dieser fast sicher eindeutig ist (keinesfalls einfach!). Wenn nun $|C| < \infty$, so gibt es $\beta_1(p)$ und $\beta_2(p)$ mit $0 < \beta_2(p) \le \beta_1(p) < \infty$, so dass

$$\exp\bigl(-\beta_1(p) n^{(d-1)/d}\bigr) \le P_p(|C| = n) \le \exp\bigl(-\beta_2(p) n^{(d-1)/d}\bigr)$$

für alle $n \in \mathbb{N}$ gilt. Man vermutet, dass

$$\delta(p) := \lim_{n\to\infty} \{-n^{-(d-1)/d} \log P_p(|C| = n)\}$$

existiert und strikt positiv ist für $p > p_c$. $n^{(d-1)/d}$ ist die Ordnung der Oberfläche der Sphäre in $\mathbb{R}^d$ mit Volumen $n$. $\delta(p)$ existiert für $d = 2$ (Im Jahre 1990 bewiesen von Alexander, Chayes, Chayes) und für $d = 3$ (bewiesen von Cerf).

(d) Existiert ein unendlicher Cluster in $p = p_c$? Für $d = 2$, $d \ge 19$ ist die Antwort nein, für alle anderen Dimensionen vermutet man die gleiche Antwort. Weiter vermutet man

$$P_{p_c}(|C| = n) \approx n^{-1/\delta}$$

für $n \to \infty$ und ein $\delta = \delta(d) > 0$.

(e) Im Fall $d = 1$ sei für ein $k \in \mathbb{Z}$

$$r(k) = \max\{n : k \leftrightarrow k + n\} \quad \text{und}$$
$$R_n = \max\{r(k); 1 \le k \le n\}\,.$$

Dann gilt für den maximalen run

$$P\Bigl(R_n > \frac{(1+\varepsilon)\log n}{\log(1/p)}\Bigr) \to 0$$

für $n \to \infty$ und

$$P\Bigl(R_n < \frac{(1-\varepsilon)\log n}{\log(1/p)}\Bigr) \to 0$$

für $n \to \infty$, $\varepsilon > 0$. Dies folgt unmittelbar aus Satz 7.5, Wahrscheinlichkeitstheorie, dem starken Gesetz für längste Erfolgs-runs.

KAPITEL 7

# Poissonapproximation und die Steinsche Methode

Einer der ältesten Grenzwertsätze in der Wahrscheinlichkeitstheorie ist POISSONs Gesetz der kleinen Zahlen, wonach $b(n,p)(k) \overset{n\to\infty}{\longrightarrow} \pi_\lambda(k)$, wenn $\lim_{n\to\infty} np(n) = \lambda > 0$ gilt. Dies bewies POISSON im Jahre 1837. Wir beschäftigen uns mit einer genaueren Abschätzung des Abstandes zwischen den obigen Wahrscheinlichkeiten.

**7.1 Definition** Für zwei W-Maße $P$, $Q$ auf einem messbaren Raum $(\Omega, \mathcal{A})$ heißt

$$d_{TV}(P,Q) := \sup\{|P(A) - Q(A)| : A \in \mathcal{A}\} \tag{7.1}$$

*Totalvariationsabstand* von $P$ und $Q$.

**7.2 Bemerkungen**

(a) Im Fall $(\Omega, \mathcal{A}) = \big(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0)\big)$ gilt für zwei W-Maße auf $\big(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0)\big)$:

$$d_{TV}(P,Q) = \frac{1}{2} \sum_{k\geq 0} |P(\{k\}) - Q(\{k\})| \ .$$

Dazu betrachte $A := \big\{k : P(\{k\}) \geq Q(\{k\})\big\}$. Das Supremum in (7.1) wird für $A$ oder $A^c$ angenommen, in der Tat sogar stets für beide Mengen, denn $P(\mathbb{N}_0) - Q(\mathbb{N}_0) = 0$ impliziert $P(A) - Q(A) = Q(A^c) - P(A^c)$. Es folgt

$$2d_{TV}(P,Q) = \big(P(A) - Q(A)\big) + \big(Q(A^c) - P(A^c)\big) = \sum_{k\geq 0} |P(\{k\}) - Q(\{k\})| \ .$$

□

(b) Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsgrößen mit Werten in $\mathbb{N}_0$. Diese konvergiert in Verteilung gegen eine W-Maß $P$ genau dann, wenn

$$\lim_{n\to\infty} d_{TV}(\mathcal{L}(X_n), P) = 0 \ .$$

$\mathcal{L}(X_n)$ bezeichnet die Verteilung von $X_n$.

Wir wollen nun beweisen:

**7.3 Satz (von Le Cam, 1960)** *Es seien $I_1, \ldots, I_n$ unabhängige Zufallsgrößen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum, mit*

*$P(I_i = 1) = p_i$ und $P(I_i = 0) = 1 - p_i$ mit $0 < p_i < 1$ für alle $i = 1, \ldots, n$. Sei $W := I_1 + \cdots + I_n$ und $\lambda := \mathbb{E}W = p_1 + \cdots + p_n$. Dann gilt:*

$$d_{TV}(\mathcal{L}(W), \pi_\lambda) \leq \sum_{i=1}^{n} p_i^2,$$

*wobei $\pi_\lambda$ die* Poisson*-Verteilung mit Parameter $\lambda$ bezeichne.*

Es folgt:

**7.4 Korollar**

$$d_{TV}(b(n,p), \pi_{np}) \leq np^2.$$

**7.5 Bemerkung** Für jedes $k \in \mathbb{N}_0$ gilt

$$\left| b(n,p)(k) - \pi_{np}(k) \right| \leq 2\, d_{TV}\big(b(n,p), \pi_{np}\big) \ .$$

Somit folgt aus 7.4 der klassische Poissonsche Grenzwertsatz, da mit $\lim_{n\to\infty} n\,p(n) = \lambda > 0$ auch $\lim_{n\to\infty} n\,p(n)^2 = 0$ gilt. Es darf auch $n\,p(n) \to \infty$ gelten, wenn nur $n\,p(n)^2 \to 0$, siehe zum Beispiel $p(n) = n^{-2/3}$.

Wir verwenden eine Technik, die man *Kopplung* nennt. Nehmen wir an, $f$ und $g$ seien zwei Wahrscheinlichkeitsverteilungen auf $\mathbb{N}_0$: $f, g : \mathbb{N}_0 \to [0,1]$, $\sum_k f(k) = \sum_k g(k) = 1$. Wir wollen zeigen, dass $\sum_{k=0}^{\infty} |f(k) - g(k)|$ klein ist. Wir werden das tun, indem wir Zufallsgrößen $X$, $Y$ auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ konstruieren, die die Verteilung $f$ beziehungsweise $g$ haben, und die „möglichst weitgehend" übereinstimmen. Es soll also die folgende Situation vorliegen:

$$f(k) = P(X = k), \quad g(k) = P(Y = k).$$

$A$ sei ein Ereignis mit $X(\omega) = Y(\omega)$ für $\omega \in A$, das heißt $A \subset \{\omega : X(\omega) = Y(\omega)\}$. Man sagt, $X$ und $Y$ seien auf $A$ „gekoppelt".

**7.6 Lemma** *Unter den obigen Bedingungen gilt*

$$\frac{1}{2} \sum_{k=0}^{\infty} |f(k) - g(k)| \leq P(A^c).$$

*Beweis:* Sei $M := \{k \in \mathbb{N}_0 : f(k) > g(k)\}$. Dann ist

$$\begin{aligned}
\sum_{k=0}^{\infty} |f(k) - g(k)| &= \sum_{k\in M} \big(f(k) - g(k)\big) - \sum_{k\notin M} \big(f(k) - g(k)\big) \\
&= 2 \sum_{k\in M} \big(f(k) - g(k)\big) - \sum_{k=0}^{\infty} \big(f(k) - g(k)\big) \\
&= 2\big(P(X \in M) - P(Y \in M)\big) - (1-1) \\
&= 2\big(P(X \in M, A) + P(X \in M, A^c) - P(Y \in M)\big) \\
&\le 2\big(P(Y \in M, A) + P(A^c) - P(Y \in M)\big) \\
&\le 2P(A^c)\ .
\end{aligned}$$

□

Wir wenden nun dieses Kopplungsargument an, um Satz 7.3 zu beweisen. Der Hauptteil des Beweises besteht in einer geeigneten Wahl des zugrundeliegenden Wahrscheinlichkeitsraumes. Da wir nur die Verteilung von $W$ berechnen müssen, ist es quasi egal, auf welchem Wahrscheinlichkeitsraum die Zufallsgrößen $I_i$ definiert werden. Es ist für uns nur wichtig, dass die Zufallsgrößen unabhängig sind und $P(I_i = 1) = p_i$ sowie $P(I_i = 0) = 1 - p_i$ gilt. Diese Freiheit nutzen wir für eine Wahl derart, dass eine POISSON-verteilte Zufallsgröße zum Parameter $\lambda$ möglichst weitgehend mit $W$ in Verteilung übereinstimmt. Dazu sei $\Omega_i = \{-1, 0, 1, 2, \dots\}$, $P_i(0) = 1 - p_i$ und $P_i(k) = \frac{e^{-p_i}}{k!} p_i^k$ für $k \ge 1$ sowie $P_i(-1) = 1 - P_i(0) - \sum_{k\ge 1} P_i(k) = e^{-p_i} - (1 - p_i)$. Zu den W-Räumen $(\Omega_i, \mathcal{P}(\Omega_i), P_i)$ betrachte dann den Produktraum $(\Omega, \mathcal{A}, P)$ der $(\Omega_i, \mathcal{P}(\Omega_i), P_i)$, $i = 1, \dots, n$. Wir setzen für $\omega \in \Omega$

$$I_i(\omega) := \begin{cases} 0, & \text{falls } \omega_i = 0, \\ 1, & \text{sonst}\ , \end{cases}$$

und

$$Y_i(\omega) := \begin{cases} k, & \text{falls } \omega_i = k, k \ge 1, \\ 0, & \text{sonst}\ . \end{cases}$$

Dann haben nach Definition die Zufallsgrößen $I_i$ die geforderte Verteilung: $P(I_i = 1) = p_i$ und $P(I_i = 0) = 1 - p_i$. Sie sind weiter nach Definition des Produktraumes unabhängig. Die $Y_i$ sind nach Definition POISSON-verteilt zum Parameter $p_i$ und ebenfalls unabhängig. Nun ist $Y = Y_1 + \cdots + Y_n$ POISSON-verteilt zum Parameter $\lambda$ (einfache direkte Rechnung oder Wahrscheinlichkeitstheorie, Beispiel 9.3(a) und Satz 9.5). Nun stimmen die Zufallsgrößen in den Werten 0 und 1 überein, und es ist $P(I_i = Y_i) = P_i(0) + P_i(1) = (1 - p_i) + e^{-p_i} p_i$, und somit

$$P(I_i \ne Y_i) = p_i(1 - e^{-p_i}) \le p_i^2,$$

denn für $x > 0$ gilt $1 - e^{-x} \leq x$. Nach Lemma 7.6 folgt dann

$$\frac{1}{2}\sum_{k=0}^{\infty} |P(W = k) - P(Y = k)| \leq P(W \neq Y) \leq \sum_{i=1}^{n} P(I_i \neq Y_i) \leq \sum_{i=1}^{n} p_i^2.$$

Damit ist Satz 7.3 bewiesen. $\square$

Charles Stein hat 1972 eine Methode entwickelt, mit der er den Fehler in der Approximation einer geeignet normierten Summe von abhängigen Zufallsvariablen durch die *Normalverteilung* abschätzen konnte. Die Methode ist enorm effektiv und ist heute nach ihm benannt. In den Folgejahren zeigte sich, dass die Methode nicht auf den Fall der Normal-Approximation beschränkt bleibt. 1975 hat Louis Chen (ein Schüler von Stein) die Steinsche Methode für die Poisson-Approximation eingeführt. Im Falle der Poisson-Verteilung spricht man in der Literatur daher auch von der Stein-Chen-Methode oder Chen-Stein-Methode. Wir wollen die Methode hier für den Fall der Poisson-Approximation vorstellen. Wir nennen sie fortan einfach die Stein*sche Methode*.

Wir stellen die Grundlagen der Steinschen Methode für Poisson-Approximation vor. Unser erstes Beispiel wird eine Verbesserung des Satzes von Le Cam sein.

**7.7 Satz** *Eine Zufallsgröße $Z$ mit Werten in $\mathbb{N}_0$ ist genau dann $\pi_\lambda$-verteilt, $\lambda \geq 0$, wenn*

$$\lambda\mathbb{E}\big(g(Z+1)\big) - \mathbb{E}\big(Zg(Z)\big) = 0$$

*für jede beschränkte Funktion $g : \mathbb{N}_0 \to \mathbb{R}$ gilt.*

*Beweis:* „$\Rightarrow$ “ Es gilt

$$\mathbb{E}\big(\lambda g(Z+1)\big) = \sum_{j\geq 0} \lambda g(j+1)\frac{\lambda^j}{j!}e^{-\lambda},$$
$$\mathbb{E}\big(Zg(Z)\big) = \sum_{l\geq 1} lg(l)\frac{\lambda^l}{l!}e^{-\lambda} = \sum_{j\geq 0} \lambda g(j+1)\frac{\lambda^j}{j!}e^{-\lambda}.$$

„$\Leftarrow$ “ Es sei $f$ die erzeugende Funktion von $Z$, definiert durch $f(s) := \sum_{k=0}^{\infty} P(Z = k)s^k$. Wir wenden die Voraussetzung auf die beschränkte Funktion $g_s(n) := s^n$, $s \in [-1, 1]$ an. Dann gilt

$$0 = \lambda\mathbb{E}s^{Z+1} - \mathbb{E}Zs^Z = \lambda sf(s) - sf'(s) \text{ für alle } s \in [-1, 1] \, .$$

also löst $f$ die Differentialgleichung $f'(s) = \lambda f(s)$ mit Randwert $f(1) = 1$ (in $s = 0$ sind $f$ und $f'$ stetig), $f(s) = e^{\lambda(s-1)}$ ist die Lösung dieser Randwertaufgabe. Die erzeugende Funktion ist durch ihre Werte in $[-1, 1]$ bereits festgelegt und für eine POISSON-verteilte Zufallsgröße $Z$ ist

$$e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^n s^n}{n!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)} \ ,$$

also liefert der Identitätssatz für Potenzreihen das Resultat. □

Das Vorgehen der STEINschen Methode basiert auf der Idee, dass für jede approximativ $\pi_\lambda$-verteilte Zufallsgröße $X$ und für jedes beschränkte $g : \mathbb{N}_0 \to \mathbb{R}$ die Differenz $\lambda \mathbb{E}\big(g(X+1)\big) - \mathbb{E}\big(Xg(X)\big)$ klein sein sollte!

Sei $f : \mathbb{N}_0 \to \mathbb{R}$ eine beschränkte Abbildung, so kann $f$ in der Form

$$\lambda g_{f,\lambda}(j+1) - j g_{f,\lambda}(j) \ , \ j \geq 0$$

dargestellt werden. $g_{f,\lambda}$ ist definiert durch $g_{f,\lambda}(j+1) = \frac{j!}{\lambda^{j+1}} \sum_{k=0}^{j} \pi_\lambda(k) e^\lambda f(k)$. Dazu sei ohne Einschränkung $g_{f,\lambda}(0) := 0$ , dann ist für $j = 0$

$$g_{f,\lambda}(1) = \frac{f(0)}{\lambda} = \frac{f(0)}{\lambda} e^\lambda \pi_\lambda(0) \ \text{ und mittels Einsetzen}$$

$$g_{f,\lambda}(j+1) = \big(f(j) + j g_{f,\lambda}(j)\big) \frac{1}{\lambda} = \frac{j!}{\lambda^{j+1}} \sum_{k=0}^{j} \pi_\lambda(k) e^\lambda f(k) \ . \tag{7.2}$$

Für $A \subseteq \mathbb{N}_0$ wählen wir

$$f_A(j) := 1_{\{j \in A\}} - \pi_\lambda(A) \ , \ j \geq 0 \ .$$

Dann existiert eine Lösung der Gleichung

$$\lambda g_{A,\lambda}(j+1) - j g_{A,\lambda}(j) = 1_{\{j \in A\}} - \pi_\lambda(A) \ , \ j \geq 0 \quad \text{(STEIN-Gleichungen)} \tag{7.3}$$

für jedes $A \subseteq \mathbb{N}_0$ und mit $U_m := \{0, 1, \ldots, m\}$ folgt

$$\begin{aligned} g_{A,\lambda}(j+1) &= \frac{j!}{\lambda^{j+1}} e^\lambda \big(\pi_\lambda(A \cap U_j) - \pi_\lambda(A) \pi_\lambda(U_j)\big), \\ g_{A,\lambda}(0) &:= 0 \qquad\qquad \text{(STEIN-Lösungen).} \end{aligned} \tag{7.4}$$

Den Nutzen dieser STEIN-Gleichungen und ihrer Lösungen kann man nun schnell erkennen: Setze für $j$ eine Zufallsgröße $W$ mit Werten in $\mathbb{N}_0$ ein, so folgt aus (7.3), wenn man auf beiden Seiten der Gleichung den Erwartungswert bildet:

$$\mathbb{E}\big(\lambda\, g_{A,\lambda}(W+1) - W g_{A,\lambda}(W)\big) = P(W \in A) - \pi_\lambda(A).$$

Wir sind an POISSON-Approximationen interessiert. Somit wollen wir also die rechte Seite der Gleichung gleichmäßig in $A$ abschätzen. Die STEIN-Gleichungen sagen uns nun, dass wir die linke Seite gleichmäßig in $A$ analysieren sollten. Wir hoffen also auf eine gute Berechenbarkeit der linken Seite unter Ausnutzung der STEIN-Lösungen $g_{A,\lambda}(\cdot)$, $A \subseteq \mathbb{N}_0$, und möglicher Schranken dieser Lösungen, gleichmäßig in $A$. Tatsächlich kann man gute Schranken für die Lösungen und für die Zuwächse der Lösungen finden, siehe Satz 7.8 und Bemerkung 7.10. Diese gelten universell in dem Sinne, als dass sie mit der Zufallsgröße $W$, deren Verteilung auf POISSON-Appoximierbarkeit hin untersucht wird, nichts zu tun haben. Die STEINsche Methode besteht also aus zwei Schritten:

(a) Zunächst stellt man die charakterisierende STEINsche Gleichung für die zu untersuchende Verteilung auf (hier die POISSON-Verteilung) und versucht die Lösungen der korrespondierenden Gleichungen gut abzuschätzen. Dieser Schritt betrachtet also ausschließlich die *Target-Verteilung*, losgelöst von der Frage, ob eine Zufallsgröße $W$ mittels der Target-Verteilung approximiert werden kann. Natürlich ist dieser Schritt auch das theoretische Herzstück der Methode, denn eine möglichst optimale Abschätzung der Lösungen und der Zuwächse der Lösungen kann universell verwendet werden.

(b) Dann wird man die Eigenschaften der Zufallsgröße $W$ des stochastischen Modells ins Spiel bringen müssen, hier um den Ausdruck $\mathbb{E}\big(\lambda\, g_{A,\lambda}(W+1)-W g_{A,\lambda}(W)\big)$ abschätzen zu können. In der Anwendung der STEINschen Methode ist dieser Schritt der eigentlich aktive Schritt.

Wir betrachten (als Beispiel) die Zufallsgröße $W$ aus Satz 7.3 (LE CAM):

Es seien $I_1, \ldots, I_n$ unabhängige Zufallsgrößen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum, mit $P(I_j = 1) = p_j = 1 - P(I_j = 0)$, $0 < p_j < 1$, $j = 1, \ldots, n$, und

$$W = \sum_{j=1}^{n} I_j; \quad \lambda := \mathbb{E}(W) = \sum_{j=1}^{n} p_j.$$

Sei weiter

$$W_i := \sum_{j=1, j\neq i}^{n} I_j.$$

Wir kürzen ab: $g_A := g_{A,\lambda}$. Es folgt nun

$$\mathbb{E}\big(I_i g_A(W)\big) = \mathbb{E}\big(I_i\, g_A(W_i+1)\big) = p_i\, \mathbb{E}\big(g_A(W_i+1)\big), \tag{7.5}$$

denn nach Definition ist $I_i$ stochastisch unabhängig von $W_i$. Also ist

$$\mathbb{E}\big(\lambda g_A(W+1) - W g_A(W)\big) = \sum_{i=1}^{n} p_i \big(\mathbb{E}(g_A(W+1)) - \mathbb{E}(g_A(W_i+1))\big).$$

Nun stimmen die Zufallsgrößen $W$ und $W_i$ überein, es sei denn $I_i = 1$. Dieses Ereignis hat Wahrscheinlichkeit $p_i$. Es folgt somit:

$$\begin{aligned} &|P(W \in A) - \pi_\lambda(A)| \le 2 \sup_{j \ge 0} |g_A(j)| \sum_{i=1}^n p_i^2 \\ \text{bzw.} \quad &|P(W \in A) - \pi_\lambda(A)| \le \sup_{j \ge 0} |g_A(j+1) - g_A(j)| \sum_{i=1}^n p_i^2. \end{aligned}$$

Wenn wir nun noch die Größen

$$\begin{aligned} ||g|| \quad &:= \quad ||g_{A,\lambda}|| := \sup_{j \ge 0} |g_{A,\lambda}(j)| \qquad \text{und} \\ \Delta g \quad &:= \quad \Delta g_{A,\lambda} := \sup_{j \ge 0} |g_{A,\lambda}(j+1) - g_{A,\lambda}(j)| \end{aligned}$$

gleichmäßig in $A \subseteq \mathbb{N}_0$ beschränken können, haben wir für den unabhängigen Fall die Funktionsweise der STEINschen Methode bereits vollständig erfaßt.

**7.8 Satz (von Barbour und Eagleson, 1983)** *Für die Zuwächse der STEIN-Lösungen finden wir gleichmäßig in $A \subset \mathbb{N}_0$:*

$$\Delta g \le \frac{1}{\lambda}\big(1 - e^{-\lambda}\big) \le \min\left(1, \frac{1}{\lambda}\right).$$

Damit haben wir gezeigt:

**7.9 Satz (von Barbour und Hall, 1984)** *In der Situation von Satz 7.3 gilt:*

$$d_{TV}\big(\mathcal{L}(W), \pi_{\mathbb{E}W}\big) \le \frac{1}{\lambda}\left(1 - e^{-\lambda}\right) \sum_{i=1}^n p_i^2.$$

*Beweis von Satz 7.8:* Wir wollen $\Delta g$, also die Zuwächse von $g_{A,\lambda}$, abschätzen. Sei $Z \sim \pi_\lambda$. Dann ist

$$f_A(j) = \sum_{k \in A} \big(1_{\{j=k\}} - \pi_\lambda(k)\big),$$

denn Wahrscheinlichkeiten sind abzählbar-additiv und die Indikatorvariable zu einer disjunkten Vereinigung von Ereignissen ist gleich der Summe der einzelnen Indikatoren. Sei nun $g_k$ die zugehörige Lösung der STEIN-Gleichung zu

$$f_k(j) := 1_{\{j=k\}} - \pi_\lambda(k), \quad j \ge 0,$$

so vermutet man, dass die folgende Identität gilt:

$$g_A(j) = \sum_{k \in A} g_k(j), \tag{7.6}$$

denn bei endlich vielen Summanden folgt dies sofort aus der („Linearität"der) Stein-Gleichung. Um (7.6) allgemein einzusehen, betrachten wir eine beliebige Menge $A \subset \mathbb{N}_0$ und führen diesen Fall via Abschneidetechnik auf den endlichen Fall zurück. Es ist für ein $M > 0$

$$f_A = \sum_{\substack{k \in A \\ k \leq M}} f_k + f_{A \cap \{k : k > M\}}.$$

Im Folgenden kürzen wir ab: $\{k : k > M\} =: \{> M\}$. Da $f_A$ so als eine endliche Summe dargestellt wird, gilt für die zu $f_A$ gehörige Stein-Lösung:

$$g_A = \sum_{\substack{k \in A \\ k \leq M}} g_k + g_{A \cap \{> M\}}.$$

Nun ist nach Definition $f_{A \cap \{> M\}}(j) = -\pi_\lambda\big(A \cap \{> M\}\big)$ für alle $j \leq M$. Es sei $j < M$, dann gilt für die Stein-Lösung nach (7.4):

$$g_{A \cap \{> M\}}(j+1) = \frac{-\pi_\lambda\big(A \cap \{> M\}\big)}{\lambda} \left(1 + \frac{j}{\lambda} + \frac{j(j-1)}{\lambda^2} + \cdots + \frac{j!}{\lambda^j}\right),$$

also

$$\begin{aligned} \Big| g_A(j+1) - \sum_{\substack{k \in A \\ k \leq M}} g_k(j+1) \Big| &\leq \operatorname{const}(\lambda, j)\, \pi_\lambda\big(A \cap \{> M\}\big) \\ &\leq \operatorname{const}(\lambda, j) \pi_\lambda\big(> M\big). \end{aligned}$$

Nun konvergiert die rechte Seite für $M \to \infty$ gegen Null, womit (7.6) gezeigt ist. Wir verwenden diese Vorbereitung, indem wir uns nun auf die Zuwächse der $g_k$ konzentrieren.

Wir erinnern uns an (7.2):

$$g_k(j+1) = \frac{1}{\lambda} \left( f_k(j) + \frac{j}{\lambda} f_k(j-1) + \frac{j(j-1)}{\lambda^2} f_k(j-2) + \cdots + \frac{j!}{\lambda^j} f_k(0) \right).$$

Nach Definition ist

$$f_k(j) = \begin{cases} -\pi_\lambda(k), & j < k, \\ 1 - \pi_\lambda(k), & j = k, \\ -\pi_\lambda(k), & j > k. \end{cases}$$

Also folgt für $j < k$:

$$g_k(j+1) = -\frac{\pi_\lambda(k)}{\lambda} \left(1 + \frac{j}{\lambda} + \cdots + \frac{j!}{\lambda^j}\right) \leq 0,$$

und daher ist hier $g_k(j+1) - g_k(j) \leq 0$, also $g_k$ monoton fallend. Es sei nun $j > k$. Mit (7.4) ist

$$
\begin{aligned}
g_k(j+1) &= e^{\lambda}\, j!\, \lambda^{-j-1} \big(\pi_\lambda(k) 1_{\{k \leq j\}} - \pi_\lambda(k)\, \pi_\lambda(U_j)\big) \\
&= e^{\lambda}\, j!\, \lambda^{-j-1} \pi_\lambda(k) \big(1 - \pi_\lambda(U_j)\big) \\
&\geq 0\,.
\end{aligned}
$$

Wir verwenden diese Darstellung und $k \leq j-1$ (da $k < j$):

$$
\begin{aligned}
g_k(j+1) - g_k(j) &= \lambda^{-j}\, (j-1)!\, \pi_\lambda(k) \left( \frac{j}{\lambda} \sum_{l=j+1}^{\infty} \frac{\lambda^l}{l!} - \sum_{l=j}^{\infty} \frac{\lambda^l}{l!} \right) \\
&\leq \lambda^{-j}\, (j-1)!\, \pi_\lambda(k) \left( \sum_{l=j}^{\infty} \frac{\lambda^l}{l!} - \sum_{l=j}^{\infty} \frac{\lambda^l}{l!} \right) \\
&= 0\,.
\end{aligned}
$$

Somit ist $g_k$ in diesem Bereich ebenfalls monoton fallend, womit nur der Zuwachs $g_k(k+1) - g_k(k)$ positiv ist, und es gilt:

$$
\begin{aligned}
g_k(k+1) - g_k(k) &= \frac{k!}{\lambda^{k+1}} \pi_\lambda(k) \left( \sum_{j=k+1}^{\infty} \frac{\lambda^j}{j!} \right) + \frac{(k-1)!}{\lambda^k} \pi_\lambda(k) \left( \sum_{j=0}^{k-1} \frac{\lambda^j}{j!} \right) \\
&= e^{-\lambda} \left( \frac{1}{\lambda} \sum_{j=k+1}^{\infty} \frac{\lambda^j}{j!} + \frac{1}{k} \sum_{j=0}^{k-1} \frac{\lambda^j}{j!} \right) \\
&\leq e^{-\lambda} \left( \frac{1}{\lambda} \sum_{j=k+1}^{\infty} \frac{\lambda^j}{j!} + \frac{1}{\lambda} \sum_{j=1}^{k} \frac{\lambda^j}{j!} \right) \\
&= \frac{e^{-\lambda}}{\lambda} \big(e^{\lambda} - 1\big) = \frac{1}{\lambda} \big(1 - e^{-\lambda}\big) \\
&\leq \min\left(1, \frac{1}{\lambda}\right)\,.
\end{aligned}
$$

Bemerke, dass die Ungleichung in der zweiten Zeile im Fall $k = 1$ eine Gleichung ist. Somit haben wir die Zuwächse von $g_k$ analysiert. Es folgt nun mit (7.6):

$$
g_A(j+1) - g_A(j) = \sum_{k \in A} \big(g_k(j+1) - g_k(j)\big);
$$

hier sind alle Summanden negativ außer im Fall $j = k$. Also ist

$$
g_A(j+1) - g_A(j) \leq g_j(j+1) - g_j(j),
$$

falls $j \in A$, sonst ist $g_A(j+1) - g_A(j) \leq 0$. Es gilt also

$$
g_A(j+1) - g_A(j) \leq \frac{1}{\lambda} \big(1 - e^{-\lambda}\big)
$$

für alle $A \subseteq \mathbb{N}_0$ und für alle $j \geq 0$. Wenn $g_A(j+1) - g_A(j) \geq 0$ für alle $j \geq 0$, so folgt

$$\Delta g \leq \frac{1}{\lambda}\big(1 - e^{-\lambda}\big).$$

Im Fall $g_A(j+1) - g_A(j) < 0$ ist

$$0 < -g_A(j+1) + g_A(j) = g_{A^c}(j+1) - g_{A^c}(j) \leq \frac{1}{\lambda}\big(1 - e^{-\lambda}\big),$$

denn $f_A + f_{A^c} = f_{\mathbb{Z}_+} \equiv 0$, und somit folgt für die korrespondierenden Stein-Lösungen: $g_A = -g_{A^c}$. □

**7.10 Bemerkung** Wir geben eine Schranke für $||g||$ ohne Beweis an:

$$||g|| \leq \min\left(1, \sqrt{\frac{2}{e\,\lambda}}\right).$$

Der eigentliche Erfolg der Steinschen Methode liegt nun darin, dass die Annahme der Unabhängigkeit sehr einfach abgeschwächt werden kann. Im Beispiel (Modell aus Satz 7.3 und 7.9) haben wir die Unabhängigkeit nur für die Identität

$$\mathbb{E}\big(I_i\, g_A(W)\big) = p_i\, \mathbb{E}\big(g_A(W_i + 1)\big)$$

benötigt. Wir versuchen diese Gleichheit zu ersetzen.

Sei fortan $\Gamma$ eine endliche Indexmenge und $(I_\alpha)_{\alpha\in\Gamma}$ eine endliche Menge von Indikatorvariablen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum. Betrachte $W := \sum_{\alpha\in\Gamma} I_\alpha$. Zu jedem $\alpha \in \Gamma$ sei $\Gamma_\alpha$ die Menge der $\beta \in \Gamma \setminus \{\alpha\}$, für die gilt: $I_\beta$ ist abhängig von $I_\alpha$. Es soll also gelten: für jedes $\beta \in (\Gamma_\alpha \cup \{\alpha\})^c$ ist $I_\beta$ unabhängig von $I_\alpha$. Es sei weiter

$$Z_\alpha := \sum_{\beta\in\Gamma_\alpha} I_\beta \quad \text{und} \quad W_\alpha := \sum_{\beta\in(\Gamma_\alpha\cup\{\alpha\})^c} I_\beta.$$

Dann gilt für jedes $\alpha \in \Gamma$:

$$W = I_\alpha + Z_\alpha + W_\alpha.$$

Weiter bezeichnen wir mit $\mu_\alpha := \mathbb{E} I_\alpha$ und $\lambda := \mathbb{E} W = \sum_{\alpha\in\Gamma} \mu_\alpha$. Dann gilt der folgende Satz:

**7.11 Satz (von Arratia, Goldstein, Gordon, 1989)** *Mit den obigen Notationen gilt ganz allgemein:*

$$d_{TV}\big(\mathcal{L}(W), \pi_\lambda\big) \leq \min\left(1, \frac{1}{\lambda}\right)\left\{\sum_{\alpha\in\Gamma} \mu_\alpha^2 + \sum_{\alpha\in\Gamma} \mu_\alpha\, \mathbb{E} Z_\alpha + \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha} \mathbb{E}(I_\alpha\, I_\beta)\right\}.$$

**7.12 Bemerkung** Zunächst besagt dieses abstrakte Resultat nur so viel: sind die $\mu_\alpha$ klein und die $\mathbb{E}(I_\alpha\, I_\beta)$ ebenfalls (keine „starke“Abhängigkeit), so ist die gefundene Schranke gut. In diesen Fällen kann man guten Gewissens die Verteilung von $W$ durch die POISSON-Verteilung zum Parameter $\lambda$ approximieren. Die Beispiele füllen diese vage Interpretation mit Leben.

*Beweis von Satz 7.11:* Wir erinnern uns an den Ansatz von STEIN, nach dem für jede Teilmenge $A \subset \mathbb{N}_0$ die Identität

$$P(W \in A) - \pi_\lambda(A) = \mathbb{E}\big(\lambda\, g_A(W+1) - W g_A(W)\big)$$

gilt, wobei $g_A$ (genauer $g_{A,\lambda}$) die STEIN-Lösung - siehe (7.4) - bezeichnet. Wir bestimmen für jedes $A \subset \mathbb{N}_0$ die rechte Seite dieser Identität. Zunächst berechnen wir $\mathbb{E}\big(I_\alpha\, g_A(W)\big)$:

Es gilt die folgende Identität (zwischen Zufallsgrößen):

$$I_\alpha\, g_A(W) = I_\alpha\, g_A(Z_\alpha + W_\alpha + 1),$$

da $I_\alpha \in \{0,1\}$. Wir schreiben die rechte Seite als

$$I_\alpha\, g_A(W_\alpha+1) + I_\alpha\big\{g_A(Z_\alpha + W_\alpha + 1) - g_A(W_\alpha+1)\big\}.$$

Somit folgt mit der Definition $\Delta g$ der Zuwächse von $g_A$:

$$\big|I_\alpha\, g_A(W) - I_\alpha\, g_A(W_\alpha+1)\big| \le I_\alpha\, Z_\alpha\, \Delta g,$$

also bei Bildung des Erwartungswertes:

$$\big|\mathbb{E}\big(I_\alpha\, g_A(W)\big) - \mathbb{E}\big(I_\alpha\, g_A(W_\alpha+1)\big)\big| \le \mathbb{E}\big(I_\alpha\, Z_\alpha\big)\Delta g.$$

Nun ist nach Definition von $W_\alpha$ diese Zufallsgröße unabhängig von $I_\alpha$, also ist $\mathbb{E}\big(I_\alpha\, g_A(W_\alpha+1)\big) = \mu_\alpha \mathbb{E}\big(g_A(W_\alpha+1)\big)$. Bilden wir nun die Summe über alle $\alpha \in \Gamma$, haben wir insgesamt gezeigt:

$$\left|\mathbb{E}\big(W\, g_A(W)\big) - \sum_{\alpha\in\Gamma} \mu_\alpha \mathbb{E}\big(g_A(W_\alpha+1)\big)\right| \le \Delta g \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha} \mathbb{E}\big(I_\alpha\, I_\beta\big).$$

Weiter gilt

$$\mathbb{E}\big(\lambda\, g_A(W+1)\big) = \sum_{\alpha\in\Gamma} \mu_\alpha \mathbb{E}\big(g_A(W+1)\big)$$

und mit $W - W_\alpha = I_\alpha + Z_\alpha$ folgt

$$|g_A(W+1) - g_A(W_\alpha+1)| \le \Delta g\,(I_\alpha + Z_\alpha).$$

Also

$$\left|\mathbb{E}\big(\lambda g_A(W+1)\big) - \sum_{\alpha\in\Gamma}\mu_\alpha \mathbb{E}\big(g_A(W_\alpha+1)\big)\right| \;\le\; \Delta g \sum_{\alpha\in\Gamma}\mu_\alpha\,\mathbb{E}(I_\alpha+Z_\alpha) \qquad (7.7)$$
$$= \Delta g \sum_{\alpha\in\Gamma}\big(\mu_\alpha^2+\mu_\alpha\,\mathbb{E}Z_\alpha\big).$$

Zusammen folgt somit unsere Behauptung durch Anwendung der Dreiecksungleichung:

$$\begin{aligned}\big|P(W\in A)-\pi_\lambda\{A\}\big| &= \big|\mathbb{E}\big(\lambda g_A(W+1)-W\,g_A(W)\big)\big| \\ &\le \Delta g\Big\{\sum_{\alpha\in\Gamma}\mu_\alpha^2+\sum_{\alpha\in\Gamma}\mu_\alpha\,\mathbb{E}Z_\alpha+\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha}\mathbb{E}(I_\alpha\,I_\beta)\Big\}.\end{aligned}$$

Die Abschätzung für $\Delta g$ nach Satz 7.8 vervollständigt den Beweis. □

Der Satz sollte im Fall von *unabhängigen* Zufallsgrößen $(I_\alpha)_{\alpha\in\Gamma}$ das Resultat von Satz 7.9 reproduzieren. Da zu festem $\alpha\in\Gamma$ hier nun alle anderen Indikatorvariablen $I_\beta$ unabhängig von $I_\alpha$ sind, setzen wir $\Gamma_\alpha=\emptyset$ für jedes $\alpha\in\Gamma$. Dann ist insbesondere $Z_\alpha=0$ für jedes $\alpha\in\Gamma$. Somit folgt unmittelbar

$$d_{TV}\big(\mathcal{L}(W),\pi_\lambda\big)\le\min\left(1,\frac{1}{\lambda}\right)\sum_{\alpha\in\Gamma}\mu_\alpha^2.$$

**7.13 Beispiel** ($k$-runs) Es seien $X_1,\dots,X_n$ unabhängig identisch verteilte Zufallsgrößen mit $P(X_i=1)=p=1-P(X_i=0)$ für $i=1,\dots,n$, $n\ge k$ und $k\in\mathbb{N}$ sei fest gewählt. Wir vereinbaren, dass für alle $1\le i\le n$ und $l\in\mathbb{N}_0$ der Index $i+n\,l$ mit $i$ identifiziert wird. Es sei weiter $\Gamma=\{1,\dots,n\}$ und $I_\alpha:=\prod_{i=\alpha}^{\alpha+k-1}X_i$ und $W=\sum_{\alpha\in\Gamma}I_\alpha$. Dann ist $\mathbb{E}(I_\alpha)=p^k=:\mu$ für alle $\alpha\in\Gamma$ und somit $\mathbb{E}W=n\,p^k=n\mu$. Somit beschreibt die Zufallsgröße $W$ die Anzahl der *runs* von Länge mindestens $k$ (ein $k$-run ist also eine Folge von $k$ aufeinanderfolgenden Einsen). Wir wählen in diesem Beispiel

$$\Gamma_\alpha=\{\alpha-(k-1),\dots,\alpha-1,\alpha+1,\dots,\alpha+k-1\},$$

denn außerhalb dieser Menge sind die zugehörigen Indikatorvariablen $I_\beta$ unabhängig von $I_\alpha$, da die $X_i$ unabhängig gewählt sind. Also sind die Summanden in $W_\alpha$ unabhängig von $I_\alpha$ und wir können Satz 7.11 anwenden. Es gilt

$$\sum_{\alpha\in\Gamma}\big(\mu_\alpha^2+\mu_\alpha\,\mathbb{E}Z_\alpha\big)=\sum_{\alpha\in\Gamma}\big(\mu^2+\mu^2(2k-2)\big)=n\,\mu^2\,(2k-1).$$

Weiter ist einfach zu sehen, dass

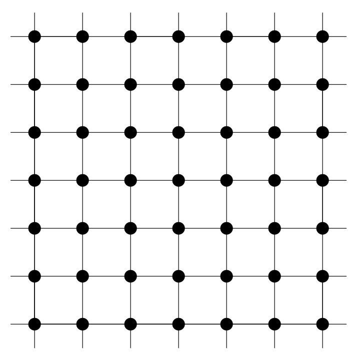
$$\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha}\mathbb{E}(I_\alpha\,I_\beta)=2n\,\mu\sum_{i=1}^{k-1}p^i$$

gilt. Ist zum Beispiel $\beta = \alpha + 1$, so müssen wir $\mathbb{E}(I_\alpha I_\beta) = P(I_\alpha = I_\beta = 1)$ bestimmen: Es soll also an der Position $\alpha$ und an der folgenden $\alpha + 1$ ein $k$-run starten. Dies geschieht mit Wahrscheinlichkeit $p^k\,p = \mu\,p$. Analog gilt für $\beta = \alpha + 2$: $P(I_\alpha = I_\beta = 1) = \mu\,p^2$, und sofort. Mit $\lambda := \mathbb{E}W = n\,\mu$ und Satz 7.11 folgt:

$$\begin{aligned} d_{TV}\big(\mathcal{L}(W), \pi_{n\,\mu}\big) &\leq \frac{1}{n\,\mu}\,(1 - e^{-\lambda})\left((2k-1)n\,\mu^2 + 2n\,\mu \sum_{i=1}^{k-1} p^i\right) \\ &= \big(1 - e^{-\lambda}\big)\left((2k-1)\,p^k + 2\sum_{i=1}^{k-1} p^i\right) \\ &= \big(1 - e^{-\lambda}\big)\left((2k-1)\,p^k + \frac{2p(1-p^{k-1})}{1-p}\right) \\ &= \mathcal{O}(p). \end{aligned}$$

Tatsächlich kann dieses Resultat noch verbessert werden: Mittels einer sogenannten Compound-Poisson Approximation erhält man eine Ordnung $\mathcal{O}(p^k)$.

**7.14 Beispiel** Gegeben sei ein Gitter auf dem Torus mit $n$ Knoten und $N = 2n$ Kanten. Wir vereinbaren die folgende Regel: *Eine Kante kann mit Wahrscheinlichkeit* $1 - p = q$ *entfernt werden, unabhängig von allen anderen Kanten.*



Dies ist ein Modell für einen Multiprozessor, bei dem die Knoten für Prozessoren und die Kanten für Verbindungen (Beziehungen zwischen Prozessoren) stehen. Aus Erfahrung wisse man, dass das System zusammenbricht, wenn $m$ Knoten isoliert sind. Allgemein sagt man nun, dass die Zuverlässigkeit des Systems die Wahrscheinlichkeit ist, dass das System (der Multiprozessor) arbeitet. Dies sei die Wahrscheinlichkeit, dass weniger als $m$ isolierte Knoten auftreten. Man interessiert sich nun für den Wert dieser Wahrscheinlichkeit, den man die *Zuverlässigkeit* oder die *Zuverlässigkeits-Wahrscheinlichkeit* nennt (reliability-probability). Der genaue Wert ist bei komplexen Systemen schwer zu bestimmen. Daher möchte man eine gute Approximation finden. Wir untersuchen hier die POISSON-Approximation.

Es sei $\Gamma = \{1, \ldots, n\}$ und

$$P(\text{eine Kante zwischen } \alpha \text{ und } \beta \text{ wird entfernt}) = q$$

für alle $\alpha$ und $\beta \in \Gamma$, die Nachbar-Gitterpunkte sind. Setze nun

$$I_\alpha := 1_{\{\text{Knoten } \alpha \text{ ist isoliert}\}}$$

für $\alpha \in \Gamma$ und $W = \sum_{\alpha \in \Gamma} I_\alpha$. Hierbei bezeichnet $I(A)$ die Indikatorvariable zum Ereignis $A$, also $1_A$. Dann ist

$$P(\text{Knoten } \alpha \text{ ist isoliert}) = q^4 = \mu_\alpha$$

und $\mathbb{E}W = n\,q^4 = \lambda$. Wie groß ist $P(W \leq m-1)$? Setze

$$\Gamma_\alpha = \{\beta \neq \alpha : \beta \text{ und } \alpha \text{ sind Nachbar-Gitterpunkte}\}.$$

Dann ist $|\Gamma_\alpha| = 4$ und

$$\sum_{\alpha \in \Gamma} \mu_\alpha^2 = n\,q^8$$

und

$$\sum_{\alpha \in \Gamma} \mu_\alpha\, \mathbb{E}Z_\alpha = \sum_{\alpha \in \Gamma} \sum_{\beta \in \Gamma_\alpha} \mu_\alpha\, \mu_\beta = 4\,n\,q^8$$

sowie

$$\sum_{\alpha \in \Gamma} \sum_{\beta \in \Gamma_\alpha} \mathbb{E}(I_\alpha\, I_\beta) = 4\,n\,q^7.$$

Mit Hilfe von Satz 7.11 finden wir

$$d_{TV}\big(\mathcal{L}(W), \pi_{n\,q^4}\big) \leq \big(5\,n\,q^8 + 4\,n\,q^7\big)\left(\frac{1-e^{-\lambda}}{n\,q^4}\right) \leq 5\,q^4 + 4\,q^3 = \mathcal{O}(q^3).$$

Wenn also zum Beispiel $\lim_{n\to\infty} n\,q(n)^4 = \mu_\infty > 1$ gilt, so folgt

$$d_{TV}\big(\mathcal{L}(W), \pi_{n\,q(n)^4}\big) \leq \mathcal{O}(1/n^{3/4}).$$

(Hinweis: wieder kann man mittels einer Compound-POISSON Approximation dieses Resultat verbessern.) Die Zuverlässigkeit des Systems kann mit Hilfe einer POISSON-verteilten Zufallsgröße $Z$ zum Parameter $\lambda = n\,q^4$ mittels

$$P(W \leq m-1) \approx \sum_{i=0}^{m-1} P(Z = i) = \sum_{i=0}^{m-1} e^{-n\,q^4} \frac{(n\,q^4)^i}{i!}$$

approximiert werden.

Zum Abschluss folgt noch ein Ausflug zu den Zufallsgraphen.

Analog zur Notation im Beweis von Satz 5.9 sei fortan $X(G)$ die Anzahl der zu $H$ isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$. Wir beweisen, dass oberhalb der Schwelle – im Fall, dass $H$ balanciert ist – sich $X$ stochastisch wie der Erwartungswert $\mathbb{E}(X)$ verhält. Unterhalb der Schwelle erhält man eine in Null konzentrierte Verteilung. Im Schwellenbereich untersuchen wir anschließend die Poisson-Approximation. Die Steinsche Methode liefert erneut Informationen über die Konvergenzgeschwindigkeit.

Nach der Tschebyschev-Ungleichung gilt immer

$$P\big(|X_n - \mathbb{E}(X_n)| > \varepsilon\, \mathbb{E}(X_n)\big) \le \frac{\mathrm{Var}(X_n)}{\varepsilon^2 (\mathbb{E}(X_n))^2},$$

sobald die Größen existieren. Wenn also $\mathrm{Var}(X_n) = o((\mathbb{E}(X_n))^2)$ gilt, so konvergiert $X_n/\mathbb{E}(X_n)$ stochastisch gegen 1. Wir können zeigen:

**7.15 Satz** *Es sei $H$ ein balancierter Graph mit $k$ Ecken und $l \ge 1$ Kanten. Wenn $\lim_{n\to\infty} p(n)\, n^{k/l} = \infty$ gilt, so folgt*

$$\lim_{n\to\infty} P\left(\left|\frac{X}{\mathbb{E}(X)} - 1\right| > \varepsilon\right) = 0 \quad \textit{für alle } \varepsilon > 0.$$

*Beweis:* Der Kern des Beweises von Satz 5.9 zeigte $\mathrm{Var}(X) = o((\mathbb{E}(X))^2)$, womit wir Satz 7.15 quasi geschenkt bekommen. □

Wir wollen nun die Verteilung von $X$ analysieren. Ist $p(n) \ll n^{-k/l}$, so erwarten wir im Limes eine Verteilung von $X$, die in der Null konzentriert ist.

**7.16 Bemerkung** Es gilt, dass $X$ in Verteilung gegen 0 für $n \to \infty$ konvergiert, wenn $p(n) \ll n^{-k/l}$. Wir formalisieren dies nicht weiter.

Was passiert, wenn

$$\lim_{n\to\infty} p(n)\, n^{k/l} = c > 0$$

gilt ($p(n)$ im Schwellenbereich)? Wir untersuchen die Poisson-Approximation mittels der Steinschen Methode. Wir gleichen die Notation an:

Sei $H$ ein fester Graph mit $k$ Ecken und $l \ge 1$ Kanten. $\Gamma$ sei die Menge aller Kopien von $H$ in $K_n$ (alle Teilgraphen von $K_n$ isomorph zu $H$). Zu $\alpha \in \Gamma$ ist $I_\alpha := 1_{\{\alpha \subset \mathcal{G}(n,p)\}}$ und

$$X = W = \sum_{\alpha\in\Gamma} I_\alpha.$$

Dann ist $\mathbb{E}(I_\alpha) = \mu_\alpha = p^l$. Weiter ist $|\Gamma| = \binom{n}{k}\, h$ und $\lambda = \mathbb{E}W = \binom{n}{k}\, h\, p^l \sim c^l$ (laut Voraussetzung). Zu $\alpha \in \Gamma$ ist $\Gamma_\alpha$ die Menge der $\beta \in \Gamma \setminus \{\alpha\}$, für die $I_\beta$ abhängig ist von $I_\alpha$. $\beta$ muß also mindestens eine gemeinsame Kante, deshalb mindestens 2 gemeinsame Knoten mit $\alpha$ haben. Mit der Definition

$$Z_\alpha := \sum_{\beta \in \Gamma_\alpha} I_\beta = \sum_{s=2}^{k} \sum_{|V(\alpha \cap \beta)|=s} I_\beta$$

folgt die Abschätzung

$$\begin{aligned} \mathbb{E}Z_\alpha &\le \sum_{s=2}^{k} \binom{k}{s} \binom{n-k}{k-s} p^l\, k! \\ &= \mathcal{O}\Big(\sum_{s=2}^{k} p^l\, n^{k-s}\Big) = \mathcal{O}\big(n^{k-2}\, p^l\big) = \mathcal{O}\big(n^{-2}\big). \end{aligned}$$

Weiter folgt analog zu den Rechnungen auf Seite 75 für ein balanciertes $H$: $\mathbb{E}(I_\alpha I_\beta) \le p^{2\,l - s\,\frac{l}{k}} = p^{\frac{l}{k}(2k-s)}$, wenn $\alpha$ und $\beta$ genau $s$ gemeinsame Knoten haben. Also folgt

$$\begin{aligned} \sum_{\alpha \in \Gamma} \sum_{\beta \in \Gamma_\alpha} \mathbb{E}(I_\alpha I_\beta) &\le |\Gamma| \sum_{s=2}^{k} \binom{k}{s} \binom{n-k}{k-s} k!\, p^{\frac{l}{k}(2k-s)} \\ &= \mathcal{O}\Big(\sum_{s=2}^{k} n^k\, n^{k-s} p^{\frac{l}{k}\,(2k-s)}\Big) = \mathcal{O}\Big(\sum_{s=2}^{k} (n\, p^{l/k})^{2\,k-s}\Big) \\ &= \mathcal{O}\,(\text{const.})\ . \end{aligned}$$

Mit $\sum_{\alpha \in \Gamma} \mu_\alpha^2 = |\Gamma| p^{2l}$ und $\lambda \sim c^l$ liefert Satz 7.11 insgesamt

$$d_{TV}\big(\mathcal{L}(W), \pi_\lambda\big) \le p^l + \mathcal{O}(n^{-2}) + \mathcal{O}\,(\text{const.}).$$

Diese Rechnung zeigt, dass die Steinsche Methode (Anwendung von Satz 7.11) im allgemeinen im Schwellenbereich für $p(n)$ keine verwendbare Poisson-Approximation liefert. Es gilt aber:

**7.17 Satz (von Barbour, 1982)** *Es sei $H$ ein strikt balancierter Graph mit $k$ Ecken und $l \ge 1$ Kanten. Ist $p(n)$ so gewählt, dass $\lim_{n\to\infty} p(n)\, n^{k/l} = c > 0$ gilt für ein $c > 0$, so folgt*

$$\lim_{n\to\infty} d_{TV}\big(\mathcal{L}(W), \pi_{\mathbb{E}W}\big) = 0.$$

Im Schwellenbereich liegt für strikt balancierte Teilgraphen stets Poisson-Approximation für die Anzahl $W$ vor. Wir beweisen Satz 7.17 hier nicht.

KAPITEL 8

# Normalapproximation und die Steinsche Methode

Charles Stein entwickelte 1972 die nach ihm benannte Methode ursprünglich zur Untersuchung von Normal-Approximation bei abhängigen Zufallsgrößen. Er untersuchte speziell stationäre, mischende Zufallsgrößen. In Satz 8.3, Wahrscheinlichkeitstheorie, haben wir den zentralen Grenzwertsatz für Folgen von unabhängigen, quadratisch integrierbaren Zufallsgrößen mit positiver Varianz, die der Lindeberg-Bedingung genügen, hergeleitet. Berry und Esséen konnten zeigen:

**8.1 Satz (von Berry und Esséen, 1945)** *Gegeben sei eine Folge $(X_n)_n$ unabhängiger, quadratisch integrierbarer, reeller Zufallsvariablen mit positiver Varianz $\sigma_n^2 := \mathrm{Var}(X_n)$. Es seien $s_n := (\sigma_1^2 + \cdots + \sigma_n^2)^{1/2}$ und $\eta_n := \mathbb{E}(X_n)$. Dann gilt*

$$\sup_{t\in\mathbb{R}} \Big| P\Big( \frac{1}{s_n} \sum_{j=1}^n (X_j - \eta_j) \leq t \Big) - \Phi(t) \Big| \leq 6 s_n^{-3} \sum_{j=1}^n \mathbb{E}(|X_j - \eta_j|^3)\,,$$

*sobald die rechte Seite existiert.*

**8.2 Korollar** *Sind die $(X_n)_n$ in Satz 8.1 zusätzlich identisch verteilt mit $\sigma^2 :=$ $\mathrm{Var}(X_n)$, $\mathbb{E}(X_n) = 0$, so gilt mit $S_n := \sum_{j=1}^n X_j$:*

$$\sup_{t\in\mathbb{R}} \Big| P\Big( \frac{S_n}{\sqrt{n}\sigma} \leq t \Big) - \Phi(t) \Big| \leq \frac{K\xi}{\sigma^3 \sqrt{n}}\,,$$

*falls zusätzlich $\xi := \mathbb{E}|X_i|^3 < \infty$ gilt. $K$ ist eine universelle Konstante.*

Wir beweisen Satz 8.1 mittels der Steinschen Methode. Wieder hat die Methode den Vorteil, auch schwach abhängige Zufallsgrößen studieren zu können. Es gilt die von Stein hergeleitete Charakterisierung der Standardnormalverteilung:

**8.3 Lemma (von Stein, 1972)** *Es sei $Z$ eine Zufallsgröße. $Z$ ist genau dann standard-normalverteilt, wenn für jede stetige und stückweise stetig differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $\int_{\mathbb{R}} |f'(x)| e^{-x^2/2}\, dx < \infty$ gilt:*

$$\mathbb{E}\big( f'(Z) - Z f(Z) \big) = 0\,. \tag{8.1}$$

*Beweis:* „$\Rightarrow$“ Sei $Z$ $N(0,1)$-verteilt und $f$ wie gefordert. Es gilt

$$\int_x^\infty y g_{0,1}(y)\,dy = -\int_{-\infty}^x y g_{0,1}(y)\,dy = g_{0,1}(x)\,.$$

Mit Hilfe von Fubini folgt

$$\begin{aligned}
\mathbb{E}f'(Z) &= \int_{-\infty}^\infty f'(x) g_{0,1}(x)\,dx \\
&= \int_0^\infty f'(x)\int_x^\infty y g_{0,1}(y)\,dy\,dx - \int_{-\infty}^0 f'(x)\int_{-\infty}^x y g_{0,1}(y)\,dy\,dx \\
&= \int_0^\infty y g_{0,1}(y)\int_0^y f'(x)\,dx\,dy - \int_{-\infty}^0 y g_{0,1}(y)\int_y^0 f'(x)\,dx\,dy \\
&= \int_{-\infty}^\infty \big(f(y)-f(0)\big) y g_{0,1}(y)\,dy \\
&= \mathbb{E}Zf(Z) - f(0)\mathbb{E}(Z) \\
&= \mathbb{E}Zf(Z)\,,
\end{aligned}$$

wie behauptet.

„$\Leftarrow$“ Es sei $U$ der lineare Operator auf dem Raum der beschränkten, stückweise stetigen Funktionen von $\mathbb{R}$ nach $\mathbb{R}$, der jedem Element $h$ die Funktion

$$(Uh)(x) := \frac{1}{g_{0,1}(x)}\int_{-\infty}^x \big(h(y)-N(h)\big) g_{0,1}(y) dy$$

mit

$$N(h) := \int_{\mathbb{R}} h(y) g_{0,1}(y)\,dy$$

zuordnet.

Da $\int_{\mathbb{R}}\big(h(y)-N(h)\big)g_{0,1}(y)\,dy = 0$ gilt, folgt

$$\begin{aligned}
U(h)(x) &= e^{x^2/2}\int_{-\infty}^x \big(h(y)-N(h)\big)e^{-y^2/2}\,dy \\
&= -e^{x^2/2}\int_x^\infty \big(h(y)-N(h)\big)e^{-y^2/2}\,dy\,.
\end{aligned}$$

Es gilt $g_{0,1}'(x) = -x g_{0,1}(x)$ für $x \in \mathbb{R}$, und somit

$$(Uh)'(x) = x(Uh)(x) + h(x) - N(h)$$

für alle $x \in \mathbb{R}$. Damit ist $Uh$ stetig und stückweise stetig differenzierbar. Mit $C := \|h\|_\infty + |N(h)|$ und $\Phi(x) := \int_{-\infty}^x g_{0,1}(y)\,dy$ folgt

$$|Uh(x)| \le \begin{cases} C\Phi(x)/g_{0,1}(x) & \text{für } x < 0\,, \\ C\big(1-\Phi(x)\big)/g_{0,1}(x) & \text{für } x > 0\,. \end{cases}$$

Nun folgt mit $\Phi(x) = 1 - \Phi(-x)$ im Fall $x < 0$ und unter Verwendung von Satz 2.20, Wahrscheinlichkeitstheorie,

$$|Uh(x)| \leq \frac{C}{|x|}\,.$$

Insbesondere kann nun

$$\int_{\mathbb{R}} \big|(Uh)'(x)\big| g_{0,1}(x)\, dx < \infty$$

gefolgert werden.

Nun wenden wir also (8.1) auf $Uh$ an und erhalten

$$0 = \mathbb{E}(Uh)'(Z) - \mathbb{E}\big(Z(Uh)(Z)\big) = \mathbb{E}h(Z) - N(h)\,.$$

Dies führt bei Wahl $h = 1_{(-\infty,t]}$, $t \in \mathbb{R}$ beliebig, zu

$$0 = P(Z \leq t) - \Phi(t)\,,$$

also ist $Z$ standard-normalverteilt. □

Das Lemma lässt vermuten: Wenn $\mathbb{E}\big(f'(W) - Wf(W)\big)$ klein ist für eine große Klasse von Funktionen $f$, so ist die Verteilung von $W$ dicht bei der Standardnormalverteilung. Weiter entnehmen wir dem Beweis des Lemmas, dass es zu jedem $h$ (hier beschränkt und stückweise stetig) ein $Uh$ gibt, welches die *Steinsche Gleichung*

$$f'(x) - xf(x) = h(x) - N(h) \qquad \text{S.G.}$$

löst. Es gilt dann

$$\mathbb{E}\big((Uh)'(W) - W(Uh)(W)\big) = \mathbb{E}h(W) - N(h)\,. \tag{8.2}$$

Wir suchen bei der Untersuchung von Normalapproximation somit Schranken für die linke Seite in (8.2) und wollen dabei die spezielle Struktur von $W$ sowie Schranken für die Lösung $Uh(\cdot)$ verwenden.

Wir bezeichnen fortan mit $L(n,\gamma)$ die Menge der $n$-Tupel $X = (X_1,\ldots,X_n)$ von unabhängigen, reellwertigen Zufallsvariablen mit $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$ und $\mathbb{E}|X_i|^3 \leq \gamma$. Die Jensensche Ungleichung liefert

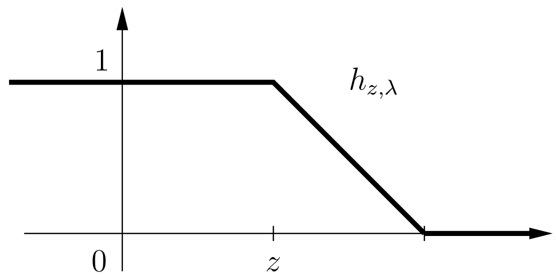
$$\mathbb{E}|X_i|^3 = \mathbb{E}\big((X_i^2)^{3/2}\big) \geq (\mathbb{E}X_i^2)^{3/2}\,.$$

Also muss man $\gamma \geq 1$ annehmen, damit die Menge $L(n,\gamma)$ nicht leer ist. Die Zufallsgrößen müssen *nicht identisch verteilt sein*. In diesem Kapitel sei

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i\,.$$

Für $z, x \in \mathbb{R}$ sei

$$h_{z,\lambda}(x) := \begin{cases} \big((1 + (z-x)/\lambda) \wedge 1\big) \vee 0 & \text{für } \lambda > 0\,, \\ 1_{(-\infty,z]}(x) & \text{für } \lambda = 0\,. \end{cases}$$




Weiter sei

$$\delta(\lambda,\gamma,n) := \sup\big\{|\mathbb{E}h_{z,\lambda}(S_n) - N(h_{z,\lambda})| : z \in \mathbb{R},\ X \in L(n,\gamma)\big\}$$

und $\delta(\gamma, n) := \delta(0, \gamma, n)$.

Wir suchen für $\gamma, n \geq 1$ eine Schranke für $\delta(\gamma, n)$. Wir stellen einen Beweis von Bolthausen vor.

Zu $h := h_{z,\lambda}$ betrachten wir die zugehörige Lösung der Steinschen Gleichung

$$f := f_{z,\lambda} := U h_{z,\lambda}$$

Und setzen

$$S_{n,i} := \frac{1}{\sqrt{n}} \sum_{j:j\neq i} X_j\,.$$

Dann gilt

$$\begin{aligned}
\mathbb{E}h(S_n) - N(h) &= \mathbb{E}\big(f'(S_n) - S_n f(S_n)\big) \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big(f'(S_n) - \sqrt{n}X_i f(S_n)\big) \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big(f'(S_n) - f'(S_{n,i})\big) \\
&\qquad + \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big(-\sqrt{n}X_i f(S_n) + \sqrt{n}X_i f(S_{n,i}) + X_i^2 f'(S_{n,i})\Big)\,,
\end{aligned}$$

denn $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$ und $X_i$ ist unabhängig von $S_{n,i}$.

Nun stellen wir die Differenz $f(S_n) - f(S_{n,i})$ als Integral über $f'$ in den Grenzen $S_{n,i}$ und $S_n$ dar:

$$f(S_n) - f(S_{n,i}) = \int_{S_{n,i}}^{S_n} f'(z)\,dz = \frac{X_i}{\sqrt{n}} \int_0^1 f'\Big(S_{n,i} + t\frac{X_i}{\sqrt{n}}\Big)\,dt,$$

also

$$\mathbb{E}h(S_n) - N(h) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big(f'(S_n) - f'(S_{n,i})\big)$$
$$- \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i^2\int_0^1\Big[f'\Big(S_{n,i} + t\frac{X}{\sqrt{n}}\Big) - f'(S_{n,i})\Big]\,dt\right). \tag{8.3}$$

Nun stellen wir ein paar Eigenschaften der Lösung $f = f_{z,\lambda}$ bereit:

**8.4 Lemma**

(a) $|f(x)| \le 3$, $|xf(x)| \le 1$, $|f'(x)| \le 2$.

(b) $|f'(x+\varepsilon) - f'(x)| \le |\varepsilon|\big(3 + 2|x| + (1/\lambda)\int_0^1 1_{(z,z+\lambda]}(x+s\varepsilon)\,ds\big)$.

(c) *Für jedes Intervall* $(a,b]$ *und jede Konstante* $c$ *gilt:*

$$\big|\mathbb{E}\big(1_{(a,b]}(S_{n,i}+c)\big)\big| \le \sqrt{\frac{n}{2\pi(n-1)}}(b-a) + 2\delta(\gamma, n-1)\,.$$

*Beweis:* (a) Wir hatten im Beweis von Lemma 8.4 bereits $|xf(x)| \le C$ gesehen, dieses $C$ war eine Schranke von $|h - N(h)|$. Hier gilt nun $|h - N(h)| \le 1$. Also ist $|f(x)| \le 1$ für $|x| \ge 1$. Für $-1 \le x \le 0$ folgt

$$|f(x)| \le e^{1/2}\int_{-\infty}^{0} e^{-z^2/2}\,dz \le 3$$

und analog $|f(x)| \le 3$ für $0 \le x \le 1$. Mit $f'(x) - xf(x) = h(x) - N(h)$ folgt $|f'(x)| \le 2$.

(b) Mit (S.G.) folgt

$$|f'(x+\varepsilon) - f'(x)| = \big|\varepsilon f(x+\varepsilon) + x\big(f(x+\varepsilon) - f(x)\big) + h(x+\varepsilon) - h(x)\big|\,.$$

Dies ist beschränkt durch

$$|\varepsilon|\left(\|f\|_\infty + |x|\|f'\|_\infty + \int_0^1 h'(x+s\varepsilon)\,ds\right).$$

Nun ist $h'(x) = -(1/\lambda)1_{(z,z+\lambda]}(x)$ für alle bis auf zwei $x$. Mit (a) folgt somit (b).

(c) Nach Definition von $\delta$ folgt für alle $z$

$$\left|\mathbb{E}h_{z,0}\Big(\sqrt{\frac{n}{n-1}}S_{n,i}\Big) - N(h_{z,0})\right| \le \delta(\gamma, n-1)\,. \tag{8.4}$$

Nun ist

$$
\begin{aligned}
&\big|\mathbb{E}\big(1_{(a,b]}(S_{n,i}+c)\big)\big| = \\
&\qquad \Big|\mathbb{E}h_{\sqrt{n/(n-1)}(b-c),0}\Big(\sqrt{\frac{n}{n-1}}S_{n,i}\Big) - \mathbb{E}h_{\sqrt{n/(n-1)}(a-c),0}\Big(\sqrt{\frac{n}{n-1}}S_{n,i}\Big)\Big|,
\end{aligned}
$$

und dies ist kleiner oder gleich

$$
\big|N(h_{\sqrt{n/(n-1)}(b-c),0}) - N(h_{\sqrt{n/(n-1)}(a-c),0})\big| + 2\delta(\gamma, n-1)
$$

mit (8.4). □

Nun wollen wir (8.3) abschätzen. Es gilt für $n \geq 2$

$$
\begin{aligned}
&\mathbb{E}|f'(S_n) - f'(S_{n,i})| \\
&\quad \leq \mathbb{E}\left[\frac{|X_i|}{\sqrt{n}}\left(3 + 2|S_{n,i}| + \frac{1}{\lambda}\int_0^1 1_{(z,z+\lambda]}\Big(S_{n,i} + s\frac{X_i}{\sqrt{n}}\Big)\,ds\right)\right] \\
&\quad = \mathbb{E}\left[\frac{|X_i|}{\sqrt{n}}\left(3 + 2\mathbb{E}\big[|S_{n,i}|\big|X_i\big] + \frac{1}{\lambda}\int_0^1 \mathbb{E}(1_{z,z+\lambda}|X_i)\big(S_{n,i} + s\frac{X_i}{\sqrt{n}}\big)\,ds\right)\right] \\
&\quad \leq \frac{\mathbb{E}|X_i|}{\sqrt{n}}\Big(3 + 2 + \frac{1}{\lambda}\Big[\sqrt{\frac{n}{2\pi(n-1)}}\lambda + 2\delta(\gamma, n-1)\Big]\Big) \\
&\quad \leq \frac{1}{\sqrt{n}}\Big(6 + \frac{2}{\lambda}\delta(\gamma, n-1)\Big).
\end{aligned}
$$

Dabei haben wir (b) und (c) in Lemma 8.4 verwendet, sowie

$$
\begin{aligned}
\mathbb{E}|S_{n,i}| &\leq \sqrt{\mathbb{E}S_{n,i}^2} = \sqrt{\frac{n-1}{n}} < 1\,, \\
\mathbb{E}|X_i| &\leq \sqrt{\mathbb{E}X_i^2} = 1
\end{aligned}
$$

und die Tatsache, dass $X_i$ unabhängig von $S_{n,i}$ ist.

Den zweiten Term in (8.3) schätzen wir nun analog ab:

$$
\begin{aligned}
\mathbb{E}\Bigg| & X_i^2 \int_0^1 \Bigg[ f'\Big(S_{n,i} + t\frac{X_i}{\sqrt{n}}\Big) - f'(S_{n,i})\Bigg] dt \Bigg| \\
&\le \mathbb{E}X_i^2 \int_0^1 t\frac{|X_i|}{\sqrt{n}} \Bigg[ 3 + 2|S_{n,i}| + \frac{1}{\lambda}\int_0^1 1_{(z,z+\lambda]}\Big(S_{n,i} + st\frac{X_n}{\sqrt{n}}\Big) ds \Bigg] dt \\
&\le \frac{\mathbb{E}|X_i|^3}{\sqrt{n}} \Bigg( \frac{3}{2} + \frac{2}{2} + \frac{1}{\lambda}\int_0^1\int_0^1 t\Big[\sqrt{\frac{n}{2\pi(n-1)}}\lambda + 2\delta(\gamma, n-1)\Big] ds\, dt \Bigg) \\
&\le \frac{\gamma}{\sqrt{n}} \Bigg( \frac{5}{2} + \frac{1}{\lambda}\Big[\frac{1}{2\sqrt{\pi}}\lambda + \delta(\gamma, n-1)\Big] \Bigg) \\
&\le \frac{\gamma}{\sqrt{n}} \Big( 3 + \frac{1}{\lambda}\delta(\gamma, n-1) \Big) .
\end{aligned}
$$

Es folgt somit

$$
\begin{aligned}
\delta(\lambda, \gamma, n) &= \sup |\mathbb{E}h_{z,\lambda} - N(h_{z,\lambda})| \\
&\le \frac{1}{\sqrt{n}}\Big[6 + \frac{2}{\lambda}\delta(\gamma, n-1)\Big] + \frac{\gamma}{\sqrt{n}}\Big[3 + \frac{1}{\lambda}\delta(\gamma, n-1)\Big] \\
&\le \frac{\gamma}{\sqrt{n}}\Big[9 + \frac{3}{\lambda}\delta(\gamma, n-1)\Big] .
\end{aligned}
$$

Die Abschätzung (c) in Lemma 8.4 liefert eine obere Schranke in $\delta(\gamma, n-1)$, woraus wir nun ein Induktions-Argument basteln.

**8.5 Lemma** *Es gilt*

$$
\delta(\gamma, n) \le \delta(\lambda, \gamma, n) + \frac{\lambda}{\sqrt{2\pi}} .
$$

*Beweis:* Es gilt $h_{z-\lambda,\lambda} \le h_{z,0} \le h_{z,\lambda}$, also

$$
\mathbb{E}h_{z-\lambda,\lambda} - N(h_{z,\lambda}) \le \mathbb{E}h_{z,0} - N(h_{z,0}) \le \mathbb{E}h_{z,\lambda} - N(h_{z-\lambda}, \lambda) .
$$

Es folgt

$$
\begin{aligned}
|\mathbb{E}h_{z,0} - N(h_{z,0})| \le \max\big\{ & |\mathbb{E}h_{z,\lambda} - N(h_{z,\lambda})|, |\mathbb{E}h_{z-\lambda,\lambda} - N(h_{z-\lambda,\lambda})| \big\} \\
& + |N(h_{z-\lambda,\lambda}) - N(h_{z,\lambda})| .
\end{aligned}
$$

Nun gilt

$$
|N(h_{z-\lambda,\lambda}) - N(h_{z,\lambda})| \le \frac{1}{\sqrt{2\pi}} \int_{z-\lambda}^{z+\lambda} \Big(1 - \frac{|w-z|}{\lambda}\Big) dw. \qquad \square
$$

Mit Lemma 8.5 erhalten wir also

$$\delta(\gamma,n) \le \frac{\gamma}{\sqrt{n}}\Big[9 + \frac{3}{\lambda}\delta(\gamma,n-1)\Big] + \frac{\lambda}{\sqrt{2\pi}}\,.$$

$\lambda \ge 0$ war beliebig. Nun wählen wir $\lambda$ optimal,

$$\lambda = 6\sqrt{2}\gamma n^{-1/2}\,,$$

und erhalten

$$\delta(\gamma,n) \le 12.5\frac{\gamma}{\sqrt{n}} + \frac{1}{2\sqrt{2}}\delta(\gamma,n-1)\,, \quad n \ge 2\,.$$

Für $n = 1$ ist $\delta(\gamma,n) \le 1$.

Angenommen, wir haben per Induktionsannahme eine Schranke für $\delta(\gamma,n-1)$ der Form

$$\delta(\gamma,n-1) \le \frac{25\gamma}{\sqrt{n-1}}\,,$$

dann folgt

$$\delta(\gamma,n) \le 12.5\frac{\gamma}{\sqrt{n}} + \frac{25\gamma}{2\sqrt{2}\sqrt{n-1}} \le 12.5\frac{\gamma}{\sqrt{n}} + 12.5\frac{\gamma}{\sqrt{n}} = \frac{25\gamma}{\sqrt{n}}\,.$$

Dies ist unser Resultat:

**8.6 Satz (von Berry und Esséen; Bolthausen, 1982)** *Für* $\gamma$, $n \ge 1$ *gilt*

$$\delta(\gamma,n) \le 25\frac{\gamma}{\sqrt{n}}\,.$$

**8.7 Bemerkungen**

(a) Der Induktions-Beweis von Bolthausen liefert sehr elegant die optimale Rate $\gamma/\sqrt{n}$. Die Konstante kann mit anderen Techniken verbessert werden. Grundsätzlich kann mit Hilfe der Steinschen Methode sehr elegant, auf ca. 1.5 Seiten, der zentrale Grenzwertsatz im Falle unabhängig und identisch verteilter Zufallsgrößen hergeleitet werden. Dazu verwendet man im Wesentlichen die Tatsache, dass $Uh$ beschränkt und stetig differenzierbar ist (ohne Lemma 8.4) und schätzt dann die Ausdrücke in (8.3) ab. Mit dieser Philosophie kann auch Satz 8.3, Wahrscheinlichkeitstheorie, bewiesen werden.

(b) Für ein $h : \mathbb{R} \to \mathbb{R}$ beschränkt und absolut stetig (also f. s. differenzierbar) gilt allgemein für $Uh$

$$\begin{aligned} \|Uh\| &\le \sqrt{\frac{\pi}{2}}\|h - N(h)\| \\ \|(Uh)'\| &\le 2\|h - N(h)\| \quad \text{und} \\ \|(Uh)''\| &\le 2\|h'\| \end{aligned}$$

jeweils in der Supremumsnorm. Dies kann man bei STEIN (1986) oder im Skript *Die Steinsche Methode* (EICHELSBACHER, 2003) nachlesen.

Der Aufwand, die Steinsche Methode vorzustellen, lohnt – analog zur Poissonapproximation –, da auch stochastische Modelle mit Abhängigkeit untersucht werden können.

Für das Modell eines Zufallsgraphen $\mathcal{G}(n,p)$ ergab sich in den Kapiteln 5 und 7 das folgende Bild:

Es sei $H$ ein (strikt) *balancierter* Graph mit $k$ Ecken und $l \geq 1$ Kanten und $X(G)$ die Anzahl der zu $H$ isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$. Dann hatten wir bewiesen:

(a) Die Grapheneigenschaft $\mathcal{A}_H$, eine Kopie von $H$ als Teilgraphen zu enthalten, hat $t(n) = n^{-k/l}$ als Schwellenfunktion (gilt für alle $H$, die balanciert sind).

(b) Ist $p(n) \ll n^{-k/l}$, so ist $\mathcal{L}(X) \to \delta_0$. Dies gilt für alle balancierten Graphen $H$.

(c) Ist $H$ *nicht balanciert*, so ist $t(n) = n^{-k/l}$ *keine Schwellenfunktion* für $\mathcal{A}_H$.

(d) Ist $H$ *strikt balanciert* und $p(n)$ so gewählt, dass $\lim_{n\to\infty} p(n)\, n^{k/l} = c > 0$ gilt, so folgt

$$\lim_{n\to\infty} d_{TV}\big(\mathcal{L}(X), \pi_{\mathbb{E}(X)}\big) = 0.$$

Was passiert nun oberhalb der Schwellenfunktion?

Wir betrachten jetzt

$$W := \frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}(X)}}$$

und wollen Normal-Approximation untersuchen. Es gilt:

**8.8 Satz (von Barbour, Karoński und Ruciński, 1989)** *Für jede beschränkte Funktion $f : \mathbb{R} \to \mathbb{R}$ mit beschränkter erster und zweiter Ableitung gilt:*

$$\big|\mathbb{E}(W\,f(W) - f'(W))\big| \leq C\,k(H) \begin{cases} \psi^{-1/2}, & \text{für } p \leq \frac{1}{2}, \\ n^{-1}(1-p)^{-1/2}, & \text{für } p > \frac{1}{2}, \end{cases}$$

*mit*

$$\psi := \min_{G\subseteq H, |E(G)|>0} \{n^{|V(G)|}\, p^{|E(G)|}\}$$

*und $k(H)$ eine Konstante, die nur von $H$ abhängt; $C := ||f''||$.*

Wir werden zunächst die Aussage dieses Satzes genauer diskutieren. Im Anschluß stellen wir eine spezielle Zerlegung von Zufallsgrößen vor, die dem lokalen Ansatz in Satz 7.11 entspricht. Wir leiten ein allgemeines Resultat für sogenannte *zerlegbare* Zufallsgrößen her und folgern Satz 8.8.

Zunächst betrachten wir Folgerungen aus Satz 8.8. Wir untersuchen den Fall, wenn $H$ balanciert ist. Dann ist $\varrho(H) = \max\{\varrho(G) : G \subseteq H\}$. Also ist

$$n^{|V(G)|/|E(G)|}\, p \geq n^{k/l}\, p$$

für alle $G \subset H$ und $|E(G)| > 0$ und

$$\min_{G \subset H, |E(G)|>0} \{n^{|V(G)|/|E(G)|}\, p\} \;= n^{k/l}\, p,$$

und somit $\psi = n^k\, p^l = (n^{k/l} p)^l$. Ist $p(n)$ nun so gewählt, dass $\lim_{n\to\infty} p(n)\, n^{k/l} = \infty$, so folgt $\psi \to +\infty$. Für $p \leq \frac{1}{2}$ liefert der Satz Konvergenz: Zu geeigneten Testfunktionen $h : \mathbb{R} \to \mathbb{R}$ (z. B. die Klasse der stetig differenzierbaren $h$, die außerhalb eines endlichen Intervalls verschwinden, mit $||h'|| < \infty$) ist die Steinlösung $U_N(h)$ beschränkt und die erste und zweite Ableitung ebenfalls (siehe Bemerkung 8.7). Also folgt für diese Konvergenz-determinierende Klasse:

**8.9 Korollar** *Ist $H$ balanciert und $p(n) \gg n^{-k/l}$ und $1 - p(n) \gg n^{-2}$, so konvergiert $W$ schwach gegen die Standardnormalverteilung.*

Wir wollen nun Satz 8.8 beweisen. Dazu betrachten wir allgemeiner die folgende Situation sogenannter *zerlegbarer* Zufallsgrößen. Wir schießen dabei etwas über das Ziel hinaus; wir wollen nur für unsere Zählstatistik $X$ einen zentralen Grenzwertsatz beweisen. Die Struktur, die hier betrachtet wird, hat vielseitige Anwendungsmöglichkeiten (siehe Skript zur Steinschen Methode und Literaturliste dort).

Sei $I$ eine endliche Menge und

$$W = \sum_{i \in I} X_i, \quad X_i \text{ jeweils quadrat-integrierbar.}$$

Es sei weiter $\mathbb{E}X_i = 0$, $i \in I$, und $\mathbb{E}W^2 = 1$. Weiter sei $W = W_i + Z_i$, wobei $W_i$ von $X_i$ unabhängig ist, und

$$Z_i := \sum_{k \in K_i} Z_{ik}, i \in I, K_i \subset I,$$

$W_i := W_{ik} + V_{ik}, i \in I, k \in K_i$, und $W_{ik}$ sei jeweils unabhängig vom Paar $(X_i, Z_{ik})$. Dann gilt:

**8.10 Lemma** *Ist $W$ zerlegbar wie oben und alle Zufallsgrößen der Zerlegung seien quadrat-integrierbar. Dann gilt für jede beschränkte Funktion $f : \mathbb{R} \to \mathbb{R}$ mit beschränkter erster und zweiter Ableitung:*

$$\left|\mathbb{E}(f'(W) - W\, f(W))\right| \leq C\, \varepsilon,$$

*wobei* $C := \sup_x |f''(x)|$ *und*

$$\varepsilon := \frac{1}{2} \sum_{i \in I} \mathbb{E}(|X_i| Z_i^2) + \sum_{i \in I} \sum_{k \in K_i} \big(\mathbb{E}|X_i Z_{ik} V_{ik}| + \mathbb{E}|X_i Z_{ik}|\, \mathbb{E}|Z_i + V_{ik}|\big).$$

Wir werden Satz 8.8 aus Lemma 8.10 folgern.

*Beweis von Lemma 8.10:* Es gilt

$$\begin{aligned}
\sum_{i \in I} \sum_{k \in K_i} \mathbb{E}(X_i Z_{ik}) &= \sum_{i \in I} \mathbb{E}(X_i\, Z_i) = \sum_{i \in I} \big( \, E(X_i W) - \, E(X_i W_i)\big) \\
&= \sum_{i \in I} \mathbb{E}(X_i W) = \mathbb{E} W^2 = 1.
\end{aligned}$$

So kann man schreiben:

$$\begin{aligned}
\mathbb{E}(W\, f(W) - f'(W)) &= \left\{ \mathbb{E}(W\, f(W)) - \sum_{i \in I} \mathbb{E}(X_i Z_i\, f'(W_i)) \right\} \\
&\quad + \left\{ \sum_{i \in I} \mathbb{E}(X_i Z_i\, f'(W_i)) - \sum_{i \in I} \sum_{k \in K_i} \mathbb{E}(X_i Z_{ik})\, \mathbb{E}(f'(W_{ik})) \right\} \\
&\quad + \sum_{i \in I} \sum_{k \in K_i} \left\{ \mathbb{E}(X_i Z_{ik}) \big[\mathbb{E}(f'(W_{ik}) - \mathbb{E}(f'(W))\big] \right\}.
\end{aligned}$$

Taylor liefert

$$\begin{aligned}
W\, f(W) &= \sum_{i \in I} X_i\, f(W) \\
&= \sum_{i \in I} X_i \left\{ f(W_i) + Z_i\, f'(W_i) + \frac{1}{2} Z_i^2\, f''(W_i + \theta_i\, Z_i) \right\}
\end{aligned}$$

für $\theta_i \in [0, 1]$. Nun ist nach Zerlegungs-Konstruktion

$$\begin{aligned}
&\left| \mathbb{E}(W\, f(W)) - \sum_{i \in I} \mathbb{E}\big(X_i Z_i\, f'(W_i)\big) \right| \\
&\quad \leq \left| \mathbb{E}\left( \sum_{i \in I} X_i\, f(W_i) \right) \right| + \frac{1}{2}\, C \sum_{i \in I} \mathbb{E}\big(|X_i| Z_i^2\big).
\end{aligned}$$

Der erste Summand auf der rechten Seite verschwindet, da $X_i$ und $W_i$ stochastisch unabhängig sind. Erneut Taylor angewandt liefert

$$\begin{aligned} X_i Z_i\, f'(W_i) &= \sum_{k\in K_i} X_i Z_{ik}\, f'(W_i) \\ &= \sum_{k\in K_i} X_i Z_{ik} \big\{ f'(W_{ik}) + V_{ik}\, f''(W_{ik} + \theta_{ik}\, V_{ik}) \big\}. \end{aligned}$$

Da $W_{ik}$ unabhängig vom Paar $(X_i, Z_{ik})$ ist, folgt

$$\begin{aligned} &\left| \sum_{i\in I} \mathbb{E}(X_i Z_i\, f'(W_i)) - \sum_{i\in I} \sum_{k\in K_i} \mathbb{E}(X_i Z_{ik})\, \mathbb{E}(f'(W_{ik})) \right| \\ &\le C \sum_{i\in I} \sum_{k\in K_i} E|X_i Z_{ik} V_{ik}|. \end{aligned}$$

Da $W_{ik} = W_i - V_{ik} = W - Z_i - V_{ik}$, folgt nach Taylor

$$f'(W_{ik}) = f'(W) - (Z_i + V_{ik})\, f''(W - \theta_{ik}(Z_i + V_{ik})),$$

also

$$\big|\mathbb{E} f'(W_{ik}) - \mathbb{E} f'(W)\big| \le C\, \mathbb{E}|Z_i + V_{ik}|,$$

womit das Lemma bewiesen ist. □

Sei speziell $I = \{1, \ldots, n\}^r$ und

$$K_i = \{k \in I : \{k_1, \ldots, k_r\} \cap \{i_1, \ldots, i_r\} \neq \emptyset\}.$$

Wähle für die obige Zerlegung

$$Z_{ik} = X_k \;\text{ und }\; V_{ik} = \sum_{l\in K_k\setminus K_i} X_l, \;\; i \in I, k \in K_i.$$

Dann ist

$$\frac{1}{2} \sum_{i\in I} \mathbb{E}\left( |X_i| \left( \sum_{k\in K_i} X_k \right) \left( \sum_{l\in K_i} X_l \right) \right) \le \frac{1}{2} \sum_{i\in I} \sum_{k,l\in K_i} \mathbb{E}\big(|X_i X_k X_l|\big)$$

und

$$\sum_{i\in I} \sum_{k\in K_i} \mathbb{E}\left( \left| X_i X_k \left( \sum_{l\in K_k\setminus K_i} X_l \right) \right| \right) \le \sum_{i\in I} \sum_{k\in K_i, l\in K_i} \mathbb{E}\big(|X_i X_k X_l|\big)$$

sowie

$$\sum_{i\in I} \sum_{k\in K_i} \mathbb{E}|X_i X_k|\;\; E\left| \sum_{l\in K_i} X_l + \sum_{l\in K_k\setminus K_i} X_l \right| \le \sum_{i\in I} \sum_{k\in K_i, l\in K_i} \Big( \mathbb{E}|X_i X_k| \mathbb{E}|X_l| \Big),$$

also gilt in diesem Fall:

$$\varepsilon \leq 2\Bigg\{\sum_{i\in I}\sum_{k,l\in K_i}\mathbb{E}\big(|X_iX_kX_l|\big)+\sum_{i\in I}\sum_{k,l\in K_i}\mathbb{E}|X_iX_k|\,\mathbb{E}|X_l|\Bigg\}.$$

Diese Schranke wollen wir nun für die Zählstatistik im Modell des Zufallsgraphen anwenden. Hier liegt die folgende Zerlegung vor:

Sei $\Gamma_n$ die Menge der Kopien von $H$ in $K_n$, kantenweise beschrieben:

$$\Gamma_n=\Bigg\{i=(i_1,\dots,i_l):1\leq i_1<\cdots<i_l\leq\binom{n}{2}:\{e_{i_1},\dots,e_{i_l}\}\text{ Kopie von }H\Bigg\},$$

wobei $e_j$ die Kante Nr. $j$ bezeichnet. Es ist $X=\sum_{i\in\Gamma_n}I_i$ mit $I_i=I($die Kanten zu $i$ bilden eine Kopie von $H)$. Bezeichnen $J_1^{(n)},\dots,J_{\binom{n}{2}}^{(n)}$ die *Kantenindikatoren*: $P(J_k^{(n)}=1)=p(n)$ für alle $k=1,\dots,\binom{n}{2}$, so ist $I_i=\prod_{k\in i}J_k^{(n)}$. Mit der Funktion $\phi(x_1,\dots,x_l):=(\prod_{k=1}^l x_k-p^l)\frac{1}{\sigma}$ und $\sigma^2:=V(X)$ ist dann

$$W=\frac{X-\mathbb{E}X}{\sqrt{V(X)}}=\sum_{i\in\Gamma_n}\phi(J_{i_1}^{(n)},\dots,J_{i_l}^{(n)}).$$

Diese Form nennt man *unvollständige* oder *gewichtete* $U$-Statistik.

Wir untersuchen $\sigma^2$, also

$$\sigma^2=\sum_{i\in\Gamma_n}\sum_{j\in K_i}\operatorname{cov}(Y_i,Y_j)$$

mit $Y_i:=\prod_{k=1}^l J_{i_k}^{(n)}$. Es gilt mit $G_i:=\{e_{i_1},\dots,e_{i_l}\}$ und für $i\in\Gamma_n$

$$\begin{aligned}\sigma^2 &= \sum_{G\subseteq H,|E(G)|\geq 1}\ \sum_{i,j\in\Gamma_n,G_i\cap G_j=G}p^{2\,l-|E(G)|}(1-p^{|E(G)|})\\ &\sim \sum_{G\subseteq H,|E(G)|\geq 1}C_G\,n^{2k-|V(G)|}\,p^{2l-|E(G)|}\,(1-p^{|E(G)|}),\end{aligned}$$

wobei $C_G$ eine kombinatorische Konstante ist, die von $H$ und $G$ abhängt. Wir erhalten nun für den letzten Ausdruck eine Abschätzung nach unten:

$$\begin{aligned}\sigma &\geq (1-p)\,n^{2k}\,p^{2l}\sum_{G\subseteq H,|E(G)|\geq 1}C_G\,n^{-|V(G)|}\,p^{-|E(G)|}\\ &\geq (1-p)\,n^{2k}\,p^{2l}C_{G^\star}\,\psi^{-1},\end{aligned}$$

wenn $G^\star$ der Graph ist, an dem das Minimum in der Definition von $\psi$ angenommen wird. Wir wollen wie in Lemma 8.10 vorgegeben abschätzen:

$$\varepsilon\leq C\Bigg(\frac{16}{\sigma^3}\sum_{i\in\Gamma_n}\sum_{k,l\in K_i}\mathbb{E}(Y_iY_kY_l)+2\sum_{i\in\Gamma_n}\sum_{k,l\in K_i}\mathbb{E}\big(|X_iX_k|\big)\,\mathbb{E}|X_l|\Bigg).$$

Nun gilt: $|X_i| = \left|\frac{Y_i - \mathbb{E}Y_i}{\sigma}\right|$, also

$$\sigma|X_i| = |Y_i - \mathbb{E}Y_i| = |\prod_{k=1}^{l} J_{i_k}^{(n)} - p^l| \leq 1$$

und

$$\sigma\mathbb{E}|X_i| = \mathbb{E}|Y_i - \mathbb{E}Y_i| = \mathbb{E}|1 - Y_i - \mathbb{E}(1 - Y_i)| \leq 2\mathbb{E}(1 - Y_i).$$

Also ist $\mathbb{E}|X_i X_k|\, \mathbb{E}|X_l| \leq \frac{2\mathbb{E}(1-Y_i)}{\sigma^3}$ und somit

$$\varepsilon \leq \left\{\frac{32}{\sigma^3} \sum_{i\in\Gamma_n} \sum_{k,l\in K_i} \mathbb{E}(Y_i Y_k Y_l)\right\} \wedge \left\{\frac{8}{\sigma^3} \sum_{i\in\Gamma_n} \sum_{k,l\in K_i} \mathbb{E}(1 - Y_i)\right\}.$$

Der zweite Ausdruck ist für $p \geq 1/2$

$$\begin{aligned} &\leq \text{const.}\, \sigma^{-3}\, n^{3k-4}\, (1 - p^l) \\ &\leq \text{const.}\, n^{3k-4}\, (1 - p^l) \left(n^{2k-2}(1-p)\right)^{-3/2} \asymp n^{-1}(1-p)^{-1/2} \end{aligned}$$

(die Konstanten sind uniform in $p \geq 1/2$). Der erste Ausdruck liefert

$$\begin{aligned} \varepsilon &\leq \frac{32}{\sigma^3} \sum_{\substack{G\subseteq H \\ |E(G)|\geq 1}} \sum_{\substack{i,k\in\Gamma_n \\ G_i\cap G_k = G}} \sum_{\substack{K\subset(G_i\cup G_k) \\ |\, E(K)|\geq 1}} \sum_{\substack{l\in\Gamma_n: \\ G_l\cap(G_i\cup G_K)=K}} p^{3l-|E(G)|-|E(K)|} \\ &\leq \frac{32}{\sigma^3} \sum_{\text{“}} \sum_{\text{“}} \sum_{\substack{K\subset(G_i\cup G_k): K\subseteq G_1 \text{ für ein } l \\ |E(K)|\geq 1}} c_K\, p^{2l-|E(G)|}\, n^{k-|V(K)|}\, p^{l-|E(K)|} \\ &\leq \text{const.}\, \sigma^{-1}\, \psi^{-1}\, n^k\, p^l \end{aligned}$$

mit const. uniform in $p \leq 1/2$. Es folgt somit insgesamt $\varepsilon \leq$ const. $\psi^{-1/2}$ für $p \leq 1/2$ und Satz 8.8 ist somit bewiesen.

*Fast alles ist prima, Janosch*

# Literaturverzeichnis

[1] G. Ahlsmeyer. *Stochastische Prozesse.* Skripten zur Mathematischen Statistik, Universität Münster, 2002.

[2] N. Alon, J.H. Spencer. *The probabilistic method.* Wiley-Interscience Series in Discrete Mathematics and Optimization, With an appendix on the life and work of Paul Erdős, Wiley-Interscience, John Wiley & Sons, New York, 2000. ISBN 0-471-37046-0.

[3] A.D. Barbour, L. Holst and S. Janson. *Poisson approximation.* Oxford Studies in Probability, 2, Oxford Science Publications, The Clarendon Press Oxford University Press, New York, 1992. ISBN 0-19-852235-5.

[4] H. Bauer. *Wahrscheinlichkeitstheorie.* de Gruyter Lehrbuch, Walter de Gruyter & Co., Berlin, fifth edition, 2002. ISBN 3-11-017236-4.

[5] P. Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. ISBN 0-471-00710-2.

[6] P. Billingsley. *Ergodic theory and information.* Reprint of the 1965 original, Robert E. Krieger Publishing Co., Huntington, N.Y., 1978. ISBN 0-88275-666-4.

[7] B. Bollobás. *Random graphs.* Cambridge Studies in Advanced Mathematics, 73, Cambridge University Press, Cambridge, 2001. ISBN 0-521-80920-7; 0-521-79722-5.

[8] B. Bollobás. *Modern graph theory.* Graduate Texts in Mathematics, 184, Springer-Verlag, New York, 1998. ISBN 0-387-98488-7.

[9] L. Breiman. *Probability.* Addison-Wesley Publishing Company, Reading, Mass., 1968.

[10] R. Diestel. *Graphentheorie.* Springer-Lehrbuch, Springer-Verlag, Berlin, 1996. ISBN 3-540-60918-0.

[11] P. Eichelsbacher. *Analysis III.* Vorlesungsskript, WS 2002/2003, 2003.

[12] P. Eichelsbacher. *Wahrscheinlichkeitstheorie*. Vorlesungsskript, WS 2003/2004, 2004.

[13] P. Eichelsbacher. *Steinsche Methode*. Skripten zur Mathematischen Statistik Nr. 38 , Münster, 2003.

[14] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[15] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[16] H.-O. Georgii. *Stochastik*. Walter de Gruyter, Berlin, 2002.

[17] G. Grimmett. *Percolation*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1999. ISBN 3-540-64902-6.

[18] S. Janson, T. Łuczak and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. ISBN 0-471-17541-2.

[19] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

[20] Charles Stein. *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7, Institute of Mathematical Statistics, Hayward, CA, 1986. ISBN 0-940600-08-0.

[21] D. W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, Cambridge, 1993. ISBN 0-521-43123-9.

[22] S. R. S. Varadhan. *Probability theory*, volume 7 of *Courant Lecture Notes in Mathematics*. New York University Courant Institute of Mathematical Sciences, New York, 2001. ISBN 0-8218-2852-2.

[23] D. Williams. *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991. ISBN 0-521-40455-X; 0-521-40605-6.

# Index