Vorlesungsskript
Stochastik

Sommersemester 2020

Friedrich-Schiller-Universität Jena

Michael H. Neumann

Wichtige Informationen

- Vorlesungszeit: 04. Mai – 17. Juli
  - Bis zu einer eventuellen Aufnahme des Prï¿$\frac{1}{2}$senzunterrichts erhalten Sie wöchentlich einen Teil des Vorlesungsskriptes. Das Material wird unter folgendem Link abgelegt: `https://users.fmi.uni-jena.de/~jschum/lehre/lectures.php?name=Neumann`
  - Die Bearbeitung der darin enthaltenen Übungsaufgaben ist freiwillig, aber empfohlen. Ich bin gern zur Durchsicht zugesandter Lösungen bereit.
- Prüfungszeit: 20. Juli – 14. August,
  Termine für mündliche Prüfungen werden rechtzeitig abgestimmt.
- Kontakt
  - Büro: Raum 3516
  - Telefon: 03641/946260
  - Email: michael.neumann@uni-jena.de

Literatur

[1] Elstrodt, J. (2011). *Maß- und Integrationstheorie*. Springer, Heidelberg.

[2] Klenke, A. (2006). *Wahrscheinlichkeitstheorie*. Springer, Berlin.

Inhalt

1 Charakteristische Funktionen

1.1 Definition und elementare Eigenschaften

Sogenannte charakteristische Funktionen sind ein nützliches Hilfsmittel der Wahrscheinlichkeitstheorie. Sie sind ein analytische Werkzeug für den Umgang mit Wahrscheinlichkeitsmaßen und erlauben beispielsweise eine einfache Definition der mehrdimensionalen Normalverteilung mit einer singulären Kovarianzmatrix. (In diesem Fall existiert bekanntermaßen keine Dichte bezüglich des Lebesgue-Maßes.) Wie im Weiteren gezeigt wird, ist das Konzept der charakteristischen Funktionen verträglich mit stochastischer Unabhängigkeit, Faltung von Verteilungen sowie schwacher Konvergenz von Wahrscheinlichkeitsmaßen. Charakteristische Funktionen erweisen sich als sehr hilfreich beim Beweis von Konvergenzaussagen. Während ein direkter Beweis des Zentralen Grenzwertsatzes doch recht mühsam ist, so ist ein Beweis mit dem Hilfsmittel der charakteristischen Funktionen vergleichsweise einfach. Selbst der Beweis einer mehrdimensionalen Version des Zentralen Grenzwertsatzes (d.h., für Zufallsvektoren) ist dann nur unwesentlich aufwendiger als im univariaten Fall. An dieser Stelle fallen wir gleich mit der Tür ins Haus:

Definition 1.1. *(i) μ sei ein endliches Maß auf $(\mathbb{R}^d, \mathcal{B}^d)$. Dann heißt $\widehat{\mu}$ mit*

$$\widehat{\mu}(t) := \int_{\mathbb{R}^d} e^{it^T x}\, d\mu(x) \quad \forall t \in \mathbb{R}^d$$

die **Fouriertransformierte** *von μ.*

(ii) $(\Omega, \mathcal{A}, P)$ sei ein Wahrscheinlichkeitsraum, $X\colon\ \Omega \to \mathbb{R}^d$ sei $(\mathcal{A} - \mathcal{B}^d)$-messbar. Dann heißt φ_X mit

$$\varphi_X(t) := \widehat{P^X}(t) = \int_{\mathbb{R}^d} e^{it^T x}\, dP^X(dx) = \int_\Omega e^{it^T X(\omega)}\, dP(\omega) = E\left[e^{it^T X}\right]$$

charakteristische Funktion *von X (bzw. P^X).*

Hier tauchen nun Integrale mit komplexwertigen Integranden auf. Die Definition solcher Integrale ist jedoch, basierend auf dem aus der Vorlesung Maßtheorie bekannten Integralbegriff, einfach. Es seien also $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f\colon\ \Omega \to \mathbb{C}$ eine komplexwertige Funktion. Zu definieren sei $\int_\Omega f\, d\mu$. Die Idee ist naheliegend. Man zerlegt f in seinen Real- und Imaginaärteil, $f(\omega) = \Re f(\omega) + i\, \Im f(\omega)$, wobei $\Re f$ und $\Im f$ reellwertige Funktionen sind, und definiert

$$\int_\Omega f\, d\mu := \int_\Omega \Re f\, d\mu + i \int_\Omega \Im f\, d\mu.$$

Dies ist möglich, wenn

- f geeignet messbar ist (d.h., Real- und Imaginärteil sind jeweils messbar),
- beide Integrale auf der rechten Seite endlich sind.

Hinsichtlich der Messbarkeit komplexwertiger Funktionen benutzen wir einfach die bekannte Definition der Messbarkeit reellwertiger bzw. $\mathbb{R}^d$-wertiger Funktionen.

Definition 1.2. $(\Omega, \mathcal{A})$ *sein ein messbarer Raum. Dann heißt* $f\colon\ \Omega \to \mathbb{C}$ $(\mathcal{A} - \mathcal{B}^2)$-**messbar**, *falls*

$$\widehat{f} := \begin{pmatrix} \Re f \\ \Im f \end{pmatrix} :\ \Omega \to \mathbb{R}^2 \qquad (\mathcal{A} - \mathcal{B}^2) - \textit{messbar ist.}$$

Bemerkung 1.3. (siehe Lemma 2.7 aus VL Maßtheorie, WS 2019/20)
f ist $(\mathcal{A} - \mathcal{B}^2)$-messbar $\Longleftrightarrow$ $\Re f$ und $\Im f$ sind $(\mathcal{A} - \mathcal{B})$-messbar.

Beweis. ($\Longrightarrow$)
Es sei $B \in \mathcal{B}$ beliebig. Da $B \times \mathbb{R}, \mathbb{R} \times B \in \mathcal{B}^2$ sind, so folgen

$$\begin{aligned} (\Re f)^{-1}(B) &= \widehat{f}^{-1}(B \times \mathbb{R}) \in \mathcal{A}, \\ (\Im f)^{-1}(B) &= \widehat{f}^{-1}(\mathbb{R} \times B) \in \mathcal{A}. \end{aligned}$$

($\Longleftarrow$)
Es seien $B_1, B_2 \in \mathcal{B}$ beliebig. Dann gelten

$$\begin{aligned} \widehat{f}^{-1}(B_1 \times \mathbb{R}) &= (\Re f)^{-1}(B_1) \in \mathcal{A}, \\ \widehat{f}^{-1}(\mathbb{R} \times B_2) &= (\Im f)^{-1}(B_1) \in \mathcal{A}. \end{aligned}$$

Daraus folgt, dass

$$\widehat{f}^{-1}(B_1 \times B_2) = \widehat{f}^{-1}(B_1 \times \mathbb{R}) \cap \widehat{f}^{-1}(\mathbb{R} \times B_2) \in \mathcal{A}.$$

Da jedoch

$$\sigma\left(\{B_1 \times B_2 :\quad B_1, B_2 \in \mathcal{B}\}\right) = \mathcal{B}^2$$

gilt, so folgt nach Lemma 2.4 aus der Vorlesung Maßtheorie, WS 2019/20, dass f $(\mathcal{A} - \mathcal{B}^2)$-messbar ist. □

Nun können wir auch Integrale mit komplexwertigem Integranden definieren:

Definition 1.4. $(\Omega, \mathcal{A}, \mu)$ *sei ein Maßraum,* $f\colon\ \Omega \to \mathbb{C}$.
f *heißt* μ-**integrierbar über** Ω *(Bezeichnung:* $f \in L^1(\mu)$*.), falls* f $(\mathcal{A} - \mathcal{B}^2)$*-messbar ist und die vier Integrale*

$$\int_\Omega (\Re f)^{\pm}\, d\mu, \qquad \int_\Omega (\Im f)^{\pm}\, d\mu$$

alle endlich sind. Dann heißt

$$\int_\Omega f\, d\mu := \int_\Omega \Re f\, d\mu + i \int_\Omega \Im f\, d\mu$$

das μ-**Integral von** f **über** Ω *(Lebesgue-Integral von* f *über* Ω *bezüglich* μ*).*

Erinnerung: $(\Omega, \mathcal{A}, \mu)$ sei ein Maßraum, $f\colon\ \Omega \to \bar{\mathbb{R}}$ sei $(\mathcal{A} - \bar{\mathcal{B}})$-messbar.

- Falls $\int_\Omega f^+ \, d\mu < \infty$ und $\int_\Omega f^- \, d\mu < \infty$, so heißt f μ-integrierbar über Ω.
- Falls genau eines der Integrale $\int_\Omega f^+ \, d\mu$, $\int_\Omega f^- \, d\mu$ endlich ist, so heißt f quasi-μ-integrierbar über Ω.

Für Integrale mit komplexwertigen Integranden gelten im Wesentlichen dieselben Eigenschaften wie für Integrale mit reellwertigen Integranden. Eine einfache Eigenschaft ist durch folgendes Lemma gegeben.

Lemma 1.5. *Es seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f \in L^1(\mu)$.*
Dann gilt

$$\left| \int_\Omega f \, d\mu \right| \le \int_\Omega |f| \, d\mu.$$

Beweis. Es sei $\int_\Omega |f| \, d\mu < \infty$. Wegen $f^+, f^- \le |f|$ folgt, dass $|\int_\Omega f \, d\mu| < \infty$. Nun gilt

$$\begin{aligned}
\left| \int_\Omega f \, d\mu \right|^2 &= \int_\Omega f \, d\mu \; \overline{\int_\Omega f \, d\mu} \\
&= \Re\left[\int_\Omega f \, d\mu \; \overline{\int_\Omega f \, d\mu} \right] \\
&= \Re\left[\int_\Omega \left(f \; \overline{\int_\Omega f \, d\mu} \right) \, d\mu \right] \\
&= \int_\Omega \underbrace{\Re\left[f \; \overline{\int_\Omega f \, d\mu} \right]}_{\le |f| \, |\int_\Omega f \, d\mu|} \, d\mu \\
&\le \int_\Omega |f| \, \left| \int_\Omega f \, d\mu \right| \, d\mu \\
&= \int_\Omega |f| \, d\mu \; \left| \int_\Omega f \, d\mu \right|.
\end{aligned}$$

□

Durch Zerlegung in Real- und Imaginärteil ist leicht zu sehen, dass die aus der Vorlesung über Maßtheorie bekannten Sätze von Lebesgue über majorisierte Konvergenz und von Fubini auf den Fall von Integralen mit komplexwertigen Integranden übertragen werden können.

Theorem 1.6. $(\Omega, \mathcal{A}, \mu)$ *sei ein Maßraum. Die Funktionen $f, f_n\colon\ \Omega \to \mathbb{C}$ ($n \in \mathbb{N}$) seien $(\mathcal{A} - \mathcal{B}^2)$-messbar und es gelte $\lim_{n\to\infty} f_n = f$ μ-fast überall. Ferner gebe es eine Funktion $g \in L^1(\mu)$, so dass für alle $n \in \mathbb{N}$ $|f_n| \le g$ μ-fast überall gilt. Dann sind f und alle f_n ($n \in \mathbb{N}$) μ-integrierbar und es gelten*

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mu = \int_\Omega f \, d\mu$$

und

$$\lim_{n\to\infty} \int_\Omega |f_n - f| \, d\mu = 0.$$

Theorem 1.7. *$(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ seien σ-endliche Maßräume, $f\colon \Omega_1 \times \Omega_2 \to \mathbb{C}$ sei $(\mu_1 \otimes \mu_2)$-integrierbar (d.h., f ist $(\mathcal{A}_1 \otimes \mathcal{A}_2) - \mathcal{B}^2)$-messbar und es gelte $\int_{\Omega_1 \times \Omega_2} |f| \, d(\mu_1 \otimes \mu_2) < \infty$). Dann gelten:*

(i) *$N_1 := \{\omega_1 \in \Omega_1 \colon \; f(\omega_1, \cdot)$ ist nicht $\mu_2 -$ integrierbar$\} \in \mathcal{A}_1$,*
$N_2 := \{\omega_2 \in \Omega_2 \colon \; f(\cdot, \omega_2)$ ist nicht $\mu_1 -$ integrierbar$\} \in \mathcal{A}_2$
und $\mu_1(N_1) = \mu_2(N_2) = 0$.

(ii) *Die Funktionen $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) \, d\mu_2(\omega_2)$ bzw. $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2) \, d\mu_1(\omega_1)$ sind μ_1-integrierbar über $\Omega_1 \backslash N_1$ bzw. μ_2-integrierbar über $\Omega_2 \backslash N_2$ und es gilt*

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f \, d(\mu_1 \otimes \mu_2) &= \int_{\Omega_1 \backslash N_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) \, d\mu_2(\omega_2) \right] d\mu_1(\omega_1) \\ &= \int_{\Omega_2 \backslash N_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) \, d\mu_1(\omega_1) \right] d\mu_2(\omega_2). \end{aligned}$$

Beispiele für charakteristische Funktionen

1) δ_a sei das Dirac-Maß auf $(\mathbb{R}^d, \mathcal{B}^d)$, $a \in \mathbb{R}^d$, d.h., $\delta_a(B) = 1$, falls $a \in B$ und $\delta_a = 0$, falls $a \notin B$. Dann
$$\widehat{\delta_a}(t) \; = \; e^{it^T a} \quad \forall t \in \mathbb{R}^d.$$

2) Es sei $\mu = \sum_{n=1}^{\infty} p_n \delta_{a_n}$ eine diskrete Wahrscheinlichkeitsverteilung auf $(\mathbb{R}^d, \mathcal{B}^d)$. Dann
$$\widehat{\mu}(t) \; = \; \sum_{n=1}^{\infty} p_n \, e^{it^T a_n} \quad \forall t \in \mathbb{R}^d.$$

3) Es sei $X \sim N(0,1)$. Dann gilt
$$\varphi_X(t) \; = \; \widehat{N(0,1)}(t) \; = \; e^{-t^2/2} \quad \forall t \in \mathbb{R}^d.$$

Beweis. Es sei
$$g(t) \; := \; E[e^{itX}] \; = \; \int_{-\infty}^{\infty} e^{itx} \, \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx \; = \; \int_{-\infty}^{\infty} \cos(tx) \, \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx.$$

Dann gilt
$$\begin{aligned} \frac{d}{dt} g(t) &= \int_{-\infty}^{\infty} \frac{d}{dt} \left\{ \cos(tx) \, \frac{e^{-x^2/2}}{\sqrt{2\pi}} \right\} dx \\ &= \int_{-\infty}^{\infty} \sin(tx) \, (-x) \, \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx \\ &= \underbrace{\left[\sin(tx) \frac{e^{-x^2/2}}{\sqrt{2\pi}} \right]_{-\infty}^{\infty}}_{=0} - \int_{-\infty}^{\infty} t \cos(tx) \, \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx \\ &= -t g(t). \end{aligned}$$

In der ersten Zeile kann man tatsächlich Differentiation und Integration vertauschen. Man betrachte dazu den Differenzenquotienten und erhält wegen

$$\left| \frac{\cos(t_n x) - \cos(tx)}{t_n - t} - (-x)\ \sin(tx) \right| \leq 2\,|x|$$

eine konvergente Majorante, so dass bei $t_n \to t$ nach dem Satz von Lebesgue über majorisierte Konvergenz das gewünschte Resultat folgt.

Die Differentialgleichung $g'(t) = -tg(t)$ hat unter der Nebenbedingung $g(0) = 1$ die eindeutig bestimmte Lösung

$$g(t) = e^{-t^2/2} \quad \forall t \in \mathbb{R}.$$

□

4) Die Dichte einer d-dimensionalen Standardnormalverteilung $N(0_d, I_d)$ ist gegeben durch

$$\varphi_{0_d,I_d}(x) = \frac{1}{(2\pi)^{d/2}}\, e^{-\|x\|^2/2}.$$

Mit $h_0(t,x) = \cos(tx)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ und $h_1(t,x) = \sin(tx)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ erhält man nach Beispiel 3) die charakterische Funktion von $N(0_d, I_d)$:

$$\begin{aligned}
\widehat{N(0_d, I_d)}(t) &= \int_{\mathbb{R}^d} e^{it^T x}\, \frac{1}{(2\pi)^{d/2}}\, e^{-\|x\|^2/2}\, d\lambda^d(x) \\
&= \int_{\mathbb{R}^d} \prod_{j=1}^{d} \left[\big(\cos(t_j x_j) + i\ \sin(t_j x_j)\big)\, \frac{1}{\sqrt{2\pi}}\, e^{-x_j^2/2} \right] d\lambda^d(x) \\
&= \int_{\mathbb{R}^d} \sum_{k_1,\dots,k_d \in \{0,1\}} i^{k_1+\cdots+k_d}\, h_{k_1}(t_1,x_1)\cdots h_{k_d}(t_d,x_d)\, d\lambda^d(x) \\
&= \sum_{k_1,\dots,k_d \in \{0,1\}} i^{k_1+\cdots+k_d} \int_{\mathbb{R}^d} h_{k_1}(t_1,x_1)\cdots h_{k_d}(t_d,x_d)\, d\lambda^d(x) \\
&= \sum_{k_1,\dots,k_d \in \{0,1\}} i^{k_1+\cdots+k_d} \prod_{j=1}^{d} \int_{\mathbb{R}} h_{k_j}(t_j,x_j)\, d\lambda(x_j) \\
&= \prod_{j=1}^{d} \int_{\mathbb{R}} h_0(t_j,x_j) + i\ h_1(t_j,x_j)\, d\lambda(x_j) \\
&= \prod_{j=1}^{d} e^{-t_j^2/2} = e^{-\|t\|^2/2}.
\end{aligned}$$

Eigenschaften von charakteristischen Funktionen

Lemma 1.8. *P sei ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^d, \mathcal{B}^d)$. Dann gelten*

(i) $\widehat{P}$ *ist gleichmäßig stetig auf* $\mathbb{R}^d$,

(ii) $|\widehat{P}(t)| \leq \widehat{P}(0_d) = 1 \quad \forall t \in \mathbb{R}^d$.

Beweis. (i) Es gilt

$$\begin{aligned}
&\left|\widehat{P}(s) - \widehat{P}(t)\right| \\
&\le \int_{\mathbb{R}^d} |e^{is^Tx} - e^{it^Tx}|\, dP(x) \\
&\le \int_{\{x:\ \|x\|\le K\}} \underbrace{|e^{is^Tx} - e^{it^Tx}|}_{\le|s^Tx-t^Tx|\le K\|s-t\|}\, dP(x) \\
&\qquad + \int_{\{x:\ \|x\|>K\}} \underbrace{|e^{is^Tx} - e^{it^Tx}|}_{\le 2}\, dP(x) \\
&\le K\,\|s-t\|\, P(\{x:\ \|x\|\le K\}) + \underbrace{2\, P(\{x:\ \|x\|>K\})}_{\to_{K\to\infty} 0}.
\end{aligned}$$

Da die rechte Seite nur von $\|s-t\|$ abhängt, folgt die behauptete gleichmäßige Stetigkeit.
(ii) is offensichtlich. □

Bevor wir nun weitere Eigenschaften von charakteristischen Funktionen herleiten, überlegen wir uns zunächst, dass durch die charakteristische Funktion die entsprechende Wahrscheinlichkeitsverteilung eindeutig bestimmt ist. Das rechtfertigt schließlich auch die Bezeichnung „charakteristische Funktion“.

Theorem 1.9. *P_1 und P_2 seien Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}^d)$. Falls $\widehat{P_1}(t) = \widehat{P_2}(t)$ $\forall t \in \mathbb{R}^d$ gilt, dann sind P_1 und P_2 gleich.*

Beweis. Es sei $[a, b] := [a_1, b_1] \times \cdots \times [a_d, b_d]$, wobei $a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{R}$ beliebig sind mit $a_i < b_i$ $\forall i$. Wir zeigen, dass

$$P_1([a,b]) = \int_{\mathbb{R}^d} \mathbb{1}_{[a,b]}\, dP_1 = \int_{\mathbb{R}^d} \mathbb{1}_{[a,b]}\, dP_2 = P_2([a,b]). \tag{1}$$

Da $\{[a,b]:\ a, b \in \mathbb{R}^d\}$ ein durchschnittsstabiles Erzeugendensystem von $\mathcal{B}^d$ ist, wobei insbesondere $\cup_{n=1}^{\infty}[-n,n]^d = \mathbb{R}^d$ gilt, so folgt nach dem Eindeutigkeitssatz der Maßtheorie (Satz 1.8 aus der Vorlesung Maßtheorie, WS 2019/20), dass P_1 und P_2 auf $\mathcal{B}^d$ übereinstimmen.

Zum Beweis von (1) approximieren wir $\mathbb{1}_{[a,b]}$ durch eine Folge $(g_n)_{n\in\mathbb{N}}$, wobei

$$g_n(z) = \mathbb{1}_{[a-d_n 1_d, b+d_n 1_d]} \star \varphi_{0_d, c_n^2 I_d}(z),$$

und $(c_n)_{n\in\mathbb{N}}$, $(d_n)_{n\in\mathbb{N}}$ Nullfolgen positiver Zahlen sind mit $c_n/d_n \to_{n\to\infty} 0$. $1_d = (1, \ldots, 1)^T$ und $0_d = (0, \ldots, 0)^T$ bezeichnen Vektoren der Länge d und $\varphi_{0_d, c_n^2 I_d}$ bezeichnet die Dichte einer mehrdimensionalen Normalverteilung mit Erwartungswertvektor 0_d und Kovarianzmatrix $c_n^2 I_d$, d.h., $\varphi_{0_d, c_n^2 I_d}(x) = (2\pi c_n^2)^{-d/2} \exp\{-\|x\|^2/(2c_n^2)\}$. Dann gelten

$$0 \le g_n(z) = \int_{\mathbb{R}^d} \underbrace{\mathbb{1}_{[a-d_n 1_d, b+d_n 1_d]}(x)}_{\le 1}\, \varphi_{0_d, c_n^2 I_d}(z-x)\, d\lambda^d(x) \le 1$$

sowie

$$g_n(z) \underset{n\to\infty}{\longrightarrow} \mathbb{1}_{[a,b]}(z) \quad \forall z \in \mathbb{R}^d. \tag{2}$$

Somit folgt nach dem Satz von Lebesgue über majorisierte Konvergenz, dass

$$\int_{\mathbb{R}^d} g_n \, dP_i \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} \mathbb{1}_{[a,b]} \, dP_i \;=\; P_i([a,b]), \quad i = 1, 2. \tag{3}$$

Außerdem gilt

$$g_n(z) \;=\; \int_{\mathbb{R}^d} e^{iz^T x} \widetilde{g}_n(x) \, d\lambda^d(x), \tag{4}$$

wobei $\widetilde{g}_n \in L^1(\lambda^d)$. Aus (4) folgt, dass

$$\begin{aligned}
\int_{\mathbb{R}^d} g_n \, dP_1 &= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} e^{iz^T x} \widetilde{g}_n(x) \, d\lambda^d(x) \right] dP_1(z) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{iz^T x} \widetilde{g}_n(x) \, d(\lambda^d \otimes P_1)(x, z) \\
&= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} e^{iz^T x} \widetilde{g}_n(x) \, dP_1(z) \right] d\lambda^d(x) \\
&= \int_{\mathbb{R}^d} \underbrace{\left[\int_{\mathbb{R}^d} e^{iz^T x} \, dP_1(z) \right]}_{=\widehat{P_1}(x)} \widetilde{g}_n(x) \, d\lambda^d(x) \\
&= \int_{\mathbb{R}^d} \underbrace{\widehat{P_1}(x)}_{=\widehat{P_2}(x)} \, \widetilde{g}_n(x) \, d\lambda^d(x) \\
&= \int_{\mathbb{R}^d} \widehat{P_2}(x) \, \widetilde{g}_n(x) \, d\lambda^d(x) \;=\; \ldots \;=\; \int_{\mathbb{R}^d} g_n \, dP_2.
\end{aligned} \tag{5}$$

Aus (3) und (5) folgt schließlich (1). □

Ergänzung: Beweis von (2) und (4)
Beweis von (2)
Fall 1: $z \in [a, b]$
Es gilt

$$\begin{aligned}
1 &\geq \mathbb{1}_{[a - d_n 1_d, b + d_n 1_d]} \star \varphi_{0_d, c_n^2 I_d}(z) \\
&= \int_{[a - d_n 1_d, b + d_n 1_d]} \varphi_{0_d, c_n^2 I_d}(z - x) \, d\lambda^d(x) \\
&= 1 - \int_{[a - d_n 1_d, b + d_n 1_d]^c} \varphi_{0_d, c_n^2 I_d}(z - x) \, d\lambda^d(x) \\
&\geq 1 - \int_{\{x:\ \|z-x\| > d_n\}} \frac{1}{(2\pi c_n^2)^{d/2}} \, e^{-\|z-x\|^2/(2c_n^2)} \, d\lambda^d(x) \\
&= 1 - \int_{\{u:\ \|u\| > d_n/c_n\}} \frac{1}{(2\pi)^{d/2}} \, e^{-\|z-x\|^2/2} \, d\lambda^d(x) \\
&\underset{n\to\infty}{\longrightarrow} 1.
\end{aligned}$$

Fall 2: $z \notin [a, b]$.
Dann gilt offensichtlich auch $z \notin [a - 2d_n 1_d, b + 2d_n 1_d]$ für alle hinreichend großen n.

Somit folgt

$$
\begin{aligned}
0 &\le \mathbb{1}_{[a-d_n 1_d, b+d_n 1_d]} \star \varphi_{0_d, c_n^2 I_d}(z) \\
&\le \int_{\{x:\ \|z-x\|>d_n\}} \varphi_{0_d, c_n^2 I_d}(x)\, d\lambda^d(x) \\
&= \int_{\{u:\ \|u\|>d_n/c_n\}} \varphi_{0_d, I_d}(u)\, d\lambda^d(u) \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

Beweis von (4)
Es gilt

$$
\begin{aligned}
g_n(z) &= \int_{\mathbb{R}^d} \underbrace{\mathbb{1}_{[a-d_n 1_d, b+d_n 1_d]}(x)}_{=:I_n(x)} \frac{1}{(2\pi c_n^2)^{d/2}} \underbrace{e^{-\|z-x\|^2/(2c_n^2)}}_{\widehat{N(0_d, I_d)}(\frac{z-x}{c_n})} d\lambda^d(x) \\
&= \int_{\mathbb{R}^d} I_n(x) \frac{1}{(2\pi c_n^2)^{d/2}} \left[\int_{\mathbb{R}^d} e^{i(\frac{z-x}{c_n})^T y} \frac{1}{(2\pi)^{d/2}} e^{-\|y\|^2/2}\, d\lambda^d(y)\right] d\lambda^d(x) \\
&= \int_{\mathbb{R}^d} \frac{1}{(2\pi c_n^2)^d} e^{i(\frac{z}{c_n})^T y - \|y\|^2/2} \underbrace{\left[\int_{\mathbb{R}^d} e^{-ix^T(\frac{y}{c_n})} I_n(x)\, d\lambda^d(x)\right]}_{\widehat{I_n}(-\frac{y}{c_n})} d\lambda^d(y) \\
&= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{iz^T y - \frac{c_n^2\|y\|^2}{2}} \widehat{I_n}(-y)\, d\lambda^d(y) \\
&= \int_{\mathbb{R}^d} e^{iz^T y} \underbrace{\left[\frac{1}{(2\pi)^d} e^{-\frac{c_n^2\|y\|^2}{2}} \widehat{I_n}(-y)\right]}_{=:\widetilde{g}_n(y)} d\lambda^d(y).
\end{aligned}
$$

Da aber offensichtlich $\widetilde{g}_n \in L^1(\lambda^d)$ ist, so folgt (4).

Übungsaufgaben

ÜA 1.1 Es seien $X_n \sim \text{Bin}(n, p/n)\ \forall n \in \mathbb{N}$ und $X \sim \text{Poisson}(p)$, wobei $p > 0$.

(i) Zeigen Sie, dass

$$P(X_n = k) \underset{n\to\infty}{\longrightarrow} P(X = k) \quad \forall k = 0, 1, \ldots$$

gilt!

(ii) Berechnen Sie die charakterischen Funktionen φ_{X_n} und φ_X und zeigen Sie, dass

$$\varphi_{X_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_X(t) \quad \forall t \in \mathbb{R}$$

gilt!
Hinweis: Nutzen Sie, dass $(1 - c/n)^n \underset{n\to\infty}{\longrightarrow} e^c\ \forall c \in \mathbb{C}$ *gilt.*

1.2 Charakteristische Funktionen und stochastische Unabhängigkeit

In diesem Kapitel wollen wir eine nützliche Charakterisierung der stochastischen Unabhängigkeit mit Hilfe des Kalküls der charakteristischen Funktionen herleiten. Darüber hinaus wird gezeigt, dass die Verteilung eines Zufallsvektors durch die Verteilungen der Linearkombinationen seiner Komponenten bestimmt ist. Zunächst zur Charakterisierung der stochastischen Unabhängigkeit:

Theorem 1.10. *Auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *seien Zufallsvariable* $X_j\colon\ \Omega \to \mathbb{R}^{d_j}$ *($j = 1, \ldots, n$) gegeben. Dann sind folgende Aussagen äquivalent:*

(i) $X_1, \ldots, X_n$ *sind stochastisch unabhängig.*

(ii) *Für* $X = (X_1^T, \ldots, X_n^T)^T$ *gilt*

$$\varphi_X(t) = \prod_{j=1}^n \varphi_{X_j}(t_j) \quad \forall t = (t_1^T, \ldots, t_n^T)^T \in \mathbb{R}^{d_1+\cdots+d_n}.$$

Beweis. ((i) $\Longrightarrow$ (ii)) Es gilt

$$\begin{aligned}\varphi_X(t) &= Ee^{it^TX} = Ee^{i(t_1^TX_1+\cdots+t_n^TX_n)} \\ &= E\left[e^{it_1^TX_1}\cdots e^{it_n^TX_n}\right] \\ &= E\left[\big(\cos(t_1^TX_1) + i\,\sin(t_1^TX_1)\big)\cdots\big(\cos(t_n^TX_n) + i\,\sin(t_n^TX_n)\big)\right].\end{aligned}$$

An der Stelle können wir nicht direkt die Eigenschaft nutzen, dass der Erwartungswert eines Produktes unabhängiger Zufallsvariablen gleich dem Produkt der Erwartungswerte ist. Diese Eigenschaft ist aus der Vorlesung EWMS für reellwertige, nicht jedoch für komplexwertige Zufallsvariable bekannt. Eine Zerlegung in Real- und Imaginärteil wie beim Beipiel 4 auf Seite 8 dieses Skriptes liefert jedoch die notwendige Erweiterung. Es seien $f_j^{(0)}(x) = \cos(t_j^Tx)$ und $f_j^{(1)}(x) = \sin(t_j^Tx)$ ($j = 1, \ldots, n$). Dann folgt aus den obigen Rechnungen

$$\begin{aligned}\varphi_X(t) &= E\left[\sum_{k_1,\ldots,k_n\in\{0,1\}} i^{k_1+\cdots+k_n} \prod_{j=1}^n f_j^{(k_j)}(X_j)\right] \\ &= \sum_{k_1,\ldots,k_n\in\{0,1\}} i^{k_1+\cdots+k_n}\, E\left[\prod_{j=1}^n f_j^{(k_j)}(X_j)\right] \\ &= \sum_{k_1,\ldots,k_n\in\{0,1\}} i^{k_1+\cdots+k_n} \prod_{j=1}^n E\left[f_j^{(k_j)}(X_j)\right] \\ &= \prod_{j=1}^n E\left[f_j^{(0)}(X_j) + i\, f_j^{(1)}(X_j)\right] \\ &= \prod_{j=1}^n \varphi_{X_j}(t_j).\end{aligned}$$

((ii) $\Longrightarrow$ (i)) Es gelte $\varphi_X(t) = \prod_{j=1}^n \varphi_{X_j}(t_j)$. Wir beweisen die stochastische Unabhängigkeit mit einem „Kopplungstrick". Wir zeigen, dass X dieselbe Verteilung wie $\widetilde{X} = (\widetilde{X_1}^T, \ldots, \widetilde{X_n}^T)^T$ besitzt, wobei $\widetilde{X_1}, \ldots, \widetilde{X_n}$ nach deren Konstruktion unabhängig sind und jeweils dieselben Verteilungen wie $X_1, \ldots, X_n$ besitzen. Es seien also $\widetilde{X_1}, \ldots, \widetilde{X_n}$ stochastisch unabhängige Zufallsvariable, welche auf einem geeigneten Wahrscheinlichkeitsraum $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ definiert sind und es sei $\widetilde{P}^{\widetilde{X}_j} = P^{X_j}$ für $j = 1, \ldots, n$. Dann gilt

$$\varphi_{\widetilde{X}}(t) = \prod_{i=1}^n \varphi_{\widetilde{X}_j}(t_j) = \prod_{j=1}^n \varphi_{X_j}(t_j) = \varphi_X(t) \quad \forall t = (t_1^T, \ldots, t_n^T)^T.$$

Nach Theorem 1.9 folgt damit, dass $\widetilde{P}^{\widetilde{X}} = P^X$. Es seien nun $B_1 \in \mathcal{B}^{d_1}, \ldots, B_n \in \mathcal{B}^{d_n}$ beliebig. Nun gilt

$$P^X(B_1 \times \cdots \times B_n) = \widetilde{P}^{\widetilde{X}}(B_1 \times \cdots \times B_n) = \prod_{j=1}^n \widetilde{P}^{\widetilde{X}_j}(B_j) = \prod_{j=1}^n P^{X_j}(B_j),$$

d.h., $X_1, \ldots, X_n$ sind stochastisch unabhängig. $\square$

Mit ähnlichen Überlegungen lässt sich die charakteristische Funktion einer Summe von unabhängigen Zufallsvariablen berechnen.

Lemma 1.11. *$X_1, \ldots, X_n$ seien unabhängige Zufallsvariable mit Werten in $\mathbb{R}^d$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann gilt*

$$\varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t) \quad \forall t \in \mathbb{R}^d.$$

Beweis. Für beliebiges $t \in \mathbb{R}^d$ gilt nach Theorem 1.10, dass

$$\varphi_{\sum_{j=1}^n X_j}(t) = Ee^{it^T(X_1+\cdots+X_n)} = Ee^{i(t^TX_1+\cdots+t^TX_n)} = \varphi_{(X_1^T,\ldots,X_n^T)^T}((t^T, \ldots, t^T)^T)) = \prod_{j=1}^n \varphi_{X_j}(t).$$

$\square$

An dieser Stelle sei nachdrücklich darauf hingewiesen, dass die Umkehrung der Aussage von Lemma 1.11 im Allgemeinen nicht gilt. Ein einfaches Gegenbeispiel lässt sich mit Cauchy-verteilen Zufallsvariablen konstruieren. Die Standard-Cauchy-Verteilung hat eine Wahrscheinlichkeitsdichte p mit $p(x) = 1/(\pi\,(1 + x^2))$ $\forall x \in \mathbb{R}$ sowie eine charakterische Funktion φ mit $\varphi(t) = e^{-|t|}$ $\forall t \in \mathbb{R}$. Falls nun X_1 und X_2 stochastisch unabängig und Standard-Cauchy-verteilt sind, so gilt nach Lemma 1.11

$$\varphi_{X_1+X_2}(t) = e^{-|t|}\,e^{-|t|} = e^{-2|t|}.$$

Falls jedoch X_1 Standard-Cauchy-verteilt ist und $X_2 := X_1$, so gilt ebenfalls

$$\varphi_{X_1+X_2}(t) = \varphi_{2X_1}(t) = Ee^{it(2X_1)} = Ee^{i(2t)X_1} = e^{-2|t|}.$$

Im letzteren Fall sind jedoch X_1 und X_2 alles andere als stochastisch unabhängig.

Die folgende Aussage stellt einen Zusammenhang zwischen der Verteilung eines Zufallsvektors und den Verteilungen der Linearkombinationen seiner Komponenten her. Dieser Zusammenhang ist insbesondere dann hilfreich sein, wenn aus Konvergenzsätzen für reellwertige Zufallsvariable entsprechende Konvergenzaussagen für Zufallsvektoren hergeleitet werden.

Theorem 1.12. *X_1 und X_2 seien $\mathbb{R}^d$-wertige Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann sind die folgenden Aussagen äquivalent:*

(i) $P^{X_1} = P^{X_2}$,

(ii) $P^{c^T X_1} = P^{c^T X_2} \quad \forall c \in \mathbb{R}^d$.

Beweis. ((i) $\Longrightarrow$ (ii)) Es sei $P^{X_1} = P^{X_2}$. Dann gilt $\varphi_{X_1}(t) = \varphi_{X_2}(t)$ $\forall t \in \mathbb{R}^d$. Es seien nun $c \in \mathbb{R}^d$ und $u \in \mathbb{R}$ beliebig. Dann gilt

$$\begin{aligned} \varphi_{c^T X_1}(u) &= E e^{i(c^T X_1)u} = E e^{i(uc)^T X_1} = \varphi_{X_1}(u\,c) \\ &= \varphi_{X_2}(u\,c) = \ldots = \varphi_{c^T X_2}(u). \end{aligned}$$

Nach Theorem 1.9 folgt, dass $P^{c^T X_1} = P^{c^T X_2}$.
((ii) $\Longrightarrow$ (i)) Es sei $t \in \mathbb{R}^d$ beliebig. Da $P^{t^T X_1} = P^{t^T X_2}$, so folgt insbesondere

$$\varphi_{t^T X_1}(1) = \varphi_{t^T X_2}(1).$$

Somit gilt

$$\begin{aligned} \varphi_{X_1}(t) &= E e^{it^T X_1} = E e^{i1(t^T X_1)} = \varphi_{t^T X_1}(1) \\ &= \varphi_{t^T X_2}(1) = \ldots = \varphi_{X_2}(t), \end{aligned}$$

woraus wiederum nach Theorem 1.9 die Gleichheit $P^{X_1} = P^{X_2}$ folgt. □

1.3 Charakteristische Funktionen und schwache Konvergenz

Wir beginnen dieses Kapitel mit einer Wiederholung der Definition zweier Konvergenzarten.

Definition 1.13. *$(P_n)_{n \geq 0}$ seien Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}^d)$. Die Folge $(P_n)_{n \in \mathbb{N}}$* **konvergiert schwach** *gegen P_0, falls*

$$\int_{\mathbb{R}^d} f\, dP_n \underset{n \to \infty}{\longrightarrow} \int_{\mathbb{R}^d} f\, dP_0$$

für alle stetigen und beschränkten Funktionen $f\colon \mathbb{R}^d \to \mathbb{R}$ gilt.
(Bezeichung: $P_n \Longrightarrow P_0$)
$(X_n)_{n \geq 0}$ seien $\mathbb{R}^d$-wertige Zufallsvariable auf jeweiligen Wahrscheinlichkeitsräumen $(\Omega_n, \mathcal{A}_n, P_n)$. Die Folge $(X_n)_{n \in \mathbb{N}}$ **konvergiert in Verteilung** *gegen X_0 genau dann, wenn $P_n^{X_n} \Longrightarrow P_0^{X_0}$.*
(Bezeichnung: $X_n \stackrel{d}{\longrightarrow} X_0$)

Das folgende Lemma liefert eine Modifikation der obigen Definition, wobei anstelle der Stetigkeit nun die gleichmäßige Stetigkeit der „Testfunktionen“ f gefordert wird. Dieses modifizierte Kriterium erweist sich als recht nützlich beim Nachweis der Konvergenz in Verteilung.

Lemma 1.14. *$(X_n)_{n\geq 0}$ sei eine Folge von $\mathbb{R}^d$-wertigen Zufallsvariablen auf jeweiligen Wahrscheinlichkeitsräumen $(\Omega_n, \mathcal{A}_n, P_n)$. Dann gilt*

$$X_n \xrightarrow{d} X_0$$

genau dann, wenn

$$\int_{\mathbb{R}^d} f\, dP_n^{X_n} \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} f\, dP_0^{X_0}$$

für alle **gleichmäßig** *stetigen und beschränkten Funktionen $f\colon\ \mathbb{R}^d \to \mathbb{R}$ gilt.*

Beweis. $(\Longrightarrow)$ Dies ist offensichtlich, da jede gleichmäßig stetige Funktion auch stetig ist.
$(\Longleftarrow)$ $f\colon\ \mathbb{R}^d \to \mathbb{R}$ sei eine beliebige stetige und beschränkte Funktion. Es sei

$$g_K(x) := \begin{cases} 1, & \text{falls } \|x\| \leq K, \\ K+1-\|x\|, & \text{falls } K \leq \|x\| \leq K+1, \\ 0, & \text{falls } \|x\| \geq K. \end{cases}$$

Nun sind $f \cdot g_K$ und $1 - g_K$ gleichmäßig stetig. Es gilt

$$\begin{aligned} &|Ef(X_n) - Ef(X_0)| \\ &\quad\leq\ |E[f(X_n)g_K(X_n)] - E[f(X_0)g_K(X_0)]| + E|f(X_n)(1-g_K(X_n))| + E|f(X_0)(1-g_K(X_0))| \\ &\quad=:\ R_{n,1}(K) + R_{n,2}(K) + R_{n,3}(K). \end{aligned}$$

Es sei nun $\varepsilon > 0$ beliebig. Dann gilt

$$R_{n,3}(K) \leq \|f\|_\infty\, E[(1-g_K)(X_0)] \leq \|f\|_\infty\, P_0(\|X_0\| > K) \leq \varepsilon/3$$

für $K = K(\varepsilon)$ hinreichend groß. Außerdem gilt

$$\begin{aligned} R_{n,2}(K(\varepsilon)) &\leq \|f\|_\infty\, E[(1-g_{K(\varepsilon)})(X_n)] \\ &\leq \|f\|_\infty \left\{E[(1-g_{K(\varepsilon)})(X_0)] + \left|E[(1-g_{K(\varepsilon)})(X_n)] - E[(1-g_{K(\varepsilon)})(X_0)]\right|\right\} \\ &\leq \varepsilon/3 + \|f\|_\infty \left|E[(1-g_{K(\varepsilon)})(X_n)] - E[(1-g_{K(\varepsilon)})(X_0)]\right|. \end{aligned}$$

Da aber $f \cdot g_K$ und $1 - g_K$ gleichmäßig stetig sind, so folgt

$$R_{n,1}(K(\varepsilon)) + \|f\|_\infty \left|E[(1-g_{K(\varepsilon)})(X_n)] - E[(1-g_{K(\varepsilon)})(X_0)]\right| \leq \varepsilon/3 \quad \forall n \geq N(\varepsilon),$$

wobei $N(\varepsilon) \in \mathbb{N}$ hinreichend groß ist. Zusammenfassend erhalten wir

$$|Ef(X_n) - Ef(X_0)| \leq \varepsilon \quad \forall n \geq N(\varepsilon),$$

woraus $X_n \xrightarrow{d} X_0$ folgt. □

Als Hauptresultat in diesem Kapitel beweisen nur nun den sogenannten Stetigkeitssatz welcher besagt, dass Konvergenz in Verteilung äquivalent ist zur punktweisen Konvergenz der charakteristischen Funktionen.

Theorem 1.15. *$(X_n)_{n\geq 0}$ sei eine Folge von $\mathbb{R}^d$-wertigen Zufallsvariablen auf jeweiligen Wahrscheinlichkeitsräumen $(\Omega_n, \mathcal{A}_n, P_n)$. Dann gilt*

$$X_n \xrightarrow{d} X_0$$

genau dann, wenn

$$\varphi_{X_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_X(t) \quad \forall t \in \mathbb{R}^d.$$

Bevor wir zum Beweis dieses Satzes kommen, leiten wir noch zwei Hilfsaussagen her. Das erste Lemma besagt, dass Konvergenz der charakteristischen Funktionen zunächst die Konvergenz in Verteilung (beliebig leicht) „verrauschter“ Versionen der Zufallsvariablen impliziert. Das darauffolgende Lemma zeigt, dass aus der Verteilungskonvergenz der verrauschten Variablen schließlich auch die Konvergenz in Verteilung der interessierenden Variablen folgt. Hier nun zunächst die erste Hilfsaussage:

Lemma 1.16. *$(X_n)_{n\geq 0}$ sei eine Folge von $\mathbb{R}^d$-wertigen Zufallsvariablen auf jeweiligen Wahrscheinlichkeitsräumen $(\Omega_n, \mathcal{A}_n, P_n)$ und es gelte $\varphi_{X_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_{X_0}(t)$ $\forall t \in \mathbb{R}^d$. Für alle $n \geq 0$ seien $Y_n \sim N(0_n, I_d)$ unabhängig von X_n und $\sigma > 0$ sei beliebig. Dann gelten:*

(i) Die Zufallsvariablen $X_n + \sigma Y_n$ besitzen jeweils Dichten p_n und es gilt

$$p_n(z) \underset{n\to\infty}{\longrightarrow} p_0(z) \quad \forall z \in \mathbb{R}^d.$$

(ii) Es gilt

$$X_n + \sigma Y_n \xrightarrow{d} X_0 + \sigma Y_0.$$

Beweis. (i) Nach der Faltungsformel gilt für beliebiges $B \in \mathcal{B}^d$ und alle $n \geq 0$, dass

$$\begin{aligned} P_n^{X_n+\sigma Y_n}(B) &= \int_{\mathbb{R}^d} P_n^{\sigma Y_n}(B-x)\, dP_n^{X_n}(x) \\ &= \int_{\mathbb{R}^d} \left[\int_B p_{\sigma Y_n}(z-x)\, d\lambda^d(z) \right] dP_n^{X_n}(x) \\ &= \int_B \underbrace{\left[\int_{\mathbb{R}^d} p_{\sigma Y_n}(z-x)\, dP_n^{X_n}(x) \right]}_{=:p_n(z)} d\lambda^d(z). \end{aligned}$$

Das heißt, $X_n + \sigma Y_n$ besitzt die Dichte p_n mit

$$\begin{aligned} p_n(z) &= \int_{\mathbb{R}^d} \frac{1}{(2\pi\sigma^2)^{d/2}} \underbrace{e^{-\frac{\|z-x\|^2}{2\sigma^2}}}_{=\varphi_{Y_n}((z-x)/\sigma)} dP_n^{X_n}(x) \\ &= \int_{\mathbb{R}^d} \frac{1}{(2\pi\sigma^2)^{d/2}} \int_{\mathbb{R}^d} e^{i(\frac{z-x}{\sigma})^T y} \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|y\|^2}{2}} d\lambda^d(y)\, dP_n^{X_n}(x) \\ &= \int_{\mathbb{R}^d} \frac{1}{(2\pi\sigma)^d} e^{i(\frac{z}{\sigma})^T y - \frac{\|y\|^2}{2}} \underbrace{\int_{\mathbb{R}^d} e^{i(-\frac{y}{\sigma})^T x}\, dP_n^{X_n}(x)}_{=\varphi_{X_n}(-y/\sigma)} d\lambda^d(y). \end{aligned}$$

Wegen $\varphi_{X_n}(-y/\sigma) \underset{n\to\infty}{\longrightarrow} \varphi_{X_0}(-y/\sigma)$ folgt nun mit dem Satz von Lebesgue über majorisierte Konvergenz, dass

$$p_n(z) \underset{n\to\infty}{\longrightarrow} p_0(z) \quad \forall z \in \mathbb{R}^d.$$

(ii) $f\colon \mathbb{R}^d \to \mathbb{R}$ sei eine beliebige stetige und beschränkte Funktion. Dann gilt

$$\begin{aligned}
&|Ef(X_n + \sigma Y_n) \;-\; Ef(X_0 + \sigma Y_0)| \\
&\quad \leq \; \|f\|_\infty \int |p_n(z) \;-\; p_0(z)| \, d\lambda^d(z) \\
&\quad = \; \|f\|_\infty \left\{ \int (p_0(z) \;-\; p_n(z))_+ \, d\lambda^d(z) \;+\; \int (p_0(z) \;-\; p_n(z))_- \, d\lambda^d(z) \right\}.
\end{aligned}$$

Nach den Satz von Lebesgue über majorisierte Konvergenz gilt

$$\int (p_0(z) \;-\; p_n(z))_+ \, d\lambda^d(z) \underset{n\to\infty}{\longrightarrow} 0.$$

Wegen $\int (p_0(z) \;-\; p_n(z))_+ \, d\lambda^d(z) = \int (p_0(z) \;-\; p_n(z))_- \, d\lambda^d(z)$ folgt jedoch, dass

$$Ef(X_n + \sigma Y_n) \underset{n\to\infty}{\longrightarrow} Ef(X_0 + \sigma Y_0).$$

□

Lemma 1.17. *Für $n \geq 0$ seien X_n und Y_n $\mathbb{R}^d$-wertige Zufallsvariable auf jeweiligen Wahrscheinlichkeitsräumen $(\Omega_n, \mathcal{A}_n, P_n)$ und es gelte $P_n^{Y_n} = Q$ $\forall n \geq 0$. Falls*

$$X_n \,+\, \sigma Y_n \overset{d}{\longrightarrow} X_0 \,+\, \sigma Y_0 \quad \forall \sigma > 0,$$

so folgt

$$X_n \overset{d}{\longrightarrow} X_0.$$

Beweis. $f\colon \mathbb{R}^d \to \mathbb{R}$ sei eine beliebige **gleichmäßig** stetige und beschränkte Funktion. Nach Lemma 1.14 genügt es zu zeigen, dass

$$Ef(X_n) \underset{n\to\infty}{\longrightarrow} Ef(X_0).$$

Es sei $\varepsilon > 0$ beliebig. Wegen der gleichmäßigen Stetigkeit existiert ein $\delta > 0$ mit

$$|f(x) \;-\; f(y)| \;\leq\; \varepsilon \quad \forall x, y \text{ mit } \|x - y\| \leq \delta.$$

Daher gilt

$$|Ef(X_n) \;-\; Ef(X_n + \sigma Y_n)| \;\leq\; \varepsilon \;+\; \underbrace{P_n(\sigma\|Y_n\| > \delta)}_{Q(\{y\colon\ \|y\| > \delta/\sigma\})} \;\leq\; 2\varepsilon \quad \forall n \geq 0,$$

falls $\sigma = \sigma_\varepsilon$ hinreichend klein ist. Außerdem gilt

$$|Ef(X_n + \sigma_\varepsilon Y_n) \;-\; Ef(X_0 + \sigma_\varepsilon Y_0)| \;\leq\; \varepsilon \quad \forall n \geq N(\varepsilon),$$

wobei $N(\varepsilon) \in \mathbb{N}$ hinreichend groß ist. Damit erhalten wir

$$|Ef(X_n) \;-\; Ef(X_0)| \;\leq\; 5\,\varepsilon \quad \forall n \geq N(\varepsilon),$$

woraus schließlich die Behauptung folgt. □

Mit Hilfe der Lemmas 1.16 und 1.17 ergibt sich nun unmittelbar der Beweis von Theorem 1.15.

Beweis von Theorem 1.15. ($\Longrightarrow$) Dies folgt direkt aus der Zerlegung $e^{it^T x} = \cos(t^T x) + i\,\sin(t^T x)$ und der Definition der Konvergenz in Verteilung.
($\Longleftarrow$) Folgt aus Lemma 1.16 und 1.17. $\square$

Die Charakterisierung der Konvergenz in Verteilung durch Theorem 1.15 erlaubt nun auch eine Zurückführung von Konvergenzaussagen für Folgen von Zufallsvektoren auf bereits bekannte Aussagen fi¿½r Folgen reellwertiger Zufallsvariablen. Später werden wir sehen, dass für Zufallsvektoren ein Analogon des Zentralen Grenzwertsatzes ohne besonderen Aufwand von einem bekannten Zentralen Grenzwertsatz fi¿½r reellwertige Zufallsvariable gewonnen werden kann.

Theorem 1.18. *(Satz von Cramér-Wold)*
Für $n \geq 0$ seien X_n und Y_n $\mathbb{R}^d$-wertige Zufallsvariable auf jeweiligen Wahrscheinlichkeitsräumen $(\Omega_n, \mathcal{A}_n, P_n)$. Dann sind äquivalent:

(i) $X_n \xrightarrow{d} X_0,$

(ii) $c^T X_n \xrightarrow{d} c^T X_0 \quad \forall c \in \mathbb{R}^d.$

Beweis. ((i) $\Longrightarrow$ (ii))
Nach Theorem 1.15 folgt aus $X_n \xrightarrow{d} X_0$, dass

$$\varphi_{X_n}(t) \;=\; Ee^{it^T X_n} \underset{n\to\infty}{\longrightarrow} Ee^{it^T X_0} \;=\; \varphi_{X_0}(t) \quad \forall t \in \mathbb{R}^d.$$

Es sei nun $c \in \mathbb{R}^d$ beliebig. Dann folgt

$$\varphi_{c^T X_n}(u) \;=\; Ee^{iu(c^T X_n)} \;=\; Ee^{i\,(uc)^T X_n} \underset{n\to\infty}{\longrightarrow} Ee^{i\,(uc)^T X_0} \;=\; \varphi_{c^T X_0}(u) \quad \forall u \in \mathbb{R},$$

woraus wiederum nach Theorem 1.15 $c^T X_n \xrightarrow{d} c^T X_0$ folgt.
((ii) $\Longrightarrow$ (i))
Es sei nun $t \in \mathbb{R}^d$ beliebig. Wegen $t^T X_n \xrightarrow{d} t^T X_0$ folgt aus Theorem 1.15, dass

$$\varphi_{t^T X_n}(u) \underset{n\to\infty}{\longrightarrow} \varphi_{t^T X_0}(u) \quad \forall u \in \mathbb{R}.$$

Daher gilt insbesondere

$$\varphi_{X_n}(t) \;=\; \varphi_{t^T X_n}(1) \underset{n\to\infty}{\longrightarrow} \varphi_{t^T X_0}(1) \;=\; \varphi_{X_0}(t).$$

Eine erneute Anwendung von Theorem 1.15 ergibt nun $X_n \xrightarrow{d} X_0$. $\square$

Übungsaufgaben

ÜA 1.2 Für $n \in \mathbb{N}$ und $j = 1, \dots, n$ seien $X_{n,j} \sim \text{Uniform}([-\sqrt{3/n}, \sqrt{3/n}])$ gegeben.

(i) Berechnen Sie Erwartungswert, Varianz und die charakteristische Funktion von $X_{n,j}$!

(ii) Für $n \in \mathbb{N}$ seien $X_{n,1}, \dots, X_{n,n}$ stochastisch unabhängig.
Konvergiert $\varphi_{\sum_{j=1}^n X_{n,j}}(t)$? Wenn ja, wogegen?
Hinweis: Nutzen Sie dabei die Taylorreihenentwicklung $sin(u) = u - u^3/6 + O(u^4)$ *sowie die Tatsache, dass* $(1 - c_n/n)^n \underset{n\to\infty}{\longrightarrow} e^c$ *gilt, falls* $(c_n)_{n\in\mathbb{N}}$ *eine Folge komplexer Zahlen mit* $c_n \underset{n\to\infty}{\longrightarrow} c$ *ist.*

1.4 Differenzierbarkeit charakteristischer Funktionen und Momente

In diesem Abschnitt stellen wird zunächst den Zusammenhang zwischen Differenzierbarkeitseigenschaften einer charakteristischen Funktion und der Existenz von Momenten der zugehörigen Wahrscheinlichkeitsverteilung her. Je nach Anzahl der endlichen Momente kann somit die charakteristische Funktion in eine Taylorreihe entsprechender Ordnung entwickelt werden. Diese Approximation erlaubt sehr einfache Beweise von Grenzwertsätzen. Zunächst wird gezeigt, dass die Existenz und Endlichkeit von Momenten bis zur Ordnung m die m-malige Differenzierbarkeit der charakteristische Funktion und somit auch eine Taylorreihenentwicklung dieser Ordnung erlauben.

Theorem 1.19. $(\Omega, \mathcal{A}, P)$ *sei ein Wahrscheinlichkeitsraum und* $X\colon (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ *sei eine Zufallsvariable mit* $E|X|^m < \infty$ *für eine* $m \in \mathbb{N}$.

Dann gilt $E|X|^k < \infty \; \forall k < m$. *Die charakterische Funktion* φ_X *ist* m*-mal gleichmäßig stetig differenzierbar mit*

$$\frac{d^k}{dt^k}\varphi_X(t) \;=\; i^k\, E[X^k\, e^{itX}] \quad \forall k = 0, \dots, m.$$

Für die Momente von X *gilt*

$$E[X^k] \;=\; \frac{1}{i^k}\; \frac{d^k}{dt^k}\varphi_X(t)\bigg|_{t=0} \qquad \forall k = 0, \dots, m.$$

und es gilt die folgende Taylorreihenentwicklung

$$\varphi_X(t) \;=\; \sum_{k=0}^{m} \frac{i^k\, E[X^k]}{k!}\, t^k \;+\; o(|t|^m).$$

Beweis. a) Momente

Für $k < m$ gilt $|X|^k \le |X|^m + 1$ und somit

$$E[|X|^k] \;\le\; E[|X|^m] \,+\, 1 \;<\; \infty.$$

b) Existenz der Ableitungen

Wir beweisen diese Aussage durch vollständige Induktion. Für $k = 0$ ist diese Aussage trivial. Wir nehmen nun an, dass für ein $k \in \{0, \dots, m-1\}$

$$\frac{d^k}{dt^k}\varphi_X(t) \;=\; i^k\, E[X^k\, e^{itX}]$$

gilt. Um die nächsthöhere Ableitung zu erhalten, betrachten wir mit $h \neq 0$ den Differenzenquotienten

$$\frac{\frac{d^k}{dt^k}\varphi_X(t+h) \;-\; \frac{d^k}{dt^k}\varphi_X(t)}{h} \;=\; i^k\, E\left[X^k\, e^{itX}\, \frac{e^{ihX} - 1}{h}\right].$$

Wegen $|e^{iy} - 1| \le |y|$ ist durch $|X|^{k+1}$ eine konvergente Majorante für die Zufallsgröße unter dem Erwartungswert gegeben. Somit existiert für alle $h \neq 0$ der obige Erwartungswert und wir erhalten nach dem Satz von Lebesgue über majorisierte Konvergenz (Theorem 1.6), dass

$$\frac{d^{k+1}}{dt^{k+1}}\varphi_X(t+h) \;=\; \lim_{h\to 0} i^k\, E\left[X^k\, e^{itX}\, \frac{e^{ihX}-1}{h}\right]$$
$$\;=\; i^k\, E\left[X^k\, e^{itX} \underbrace{\lim_{h\to 0}\frac{e^{ihX}-e^{i0X}}{h}}_{=iX}\right] \;=\; i^{k+1}\, E[X^{k+1}\, e^{itX}].$$

c) Gleichmäßige Stetigkeit der Ableitungen

Für $l \in \{0,\ldots,m\}$ gilt folgende Abschätzung:

$$\left|\frac{d^k}{dt^k}\varphi_X(t+h) \;-\; \frac{d^k}{dt^k}\varphi_X(t)\right| \;\le\; \left|E\left[X^k\, e^{itX}\,(e^{ihX}\;-\;1)\right]\right| \;\le\; E\left[|X|^k\, |e^{ihX}\;-\;1|\right].$$

Die Zufallsvariable innerhalb des rechten Terms wird durch $2|X|^l$ majorisiert und mit dem Satz von Lebesgue über majorisierte Konvergenz folgt

$$E\left[|X|^k\, |e^{ihX}\;-\;1|\right] \underset{h\to 0}{\longrightarrow} 0,$$

womit die gleichmäßige Stetigkeit der Ableitungen folgt.

d) Taylorreihenentwicklung

Wir werden den bekannten Satz von Taylor für reellwertige Funktionen benutzen und zerlegen dazu die charakterische Funktion in Real- und Imaginärteil. Für den Realteil von φ_X gilt

$$\Re(\varphi_X(t)) \;=\; \sum_{k=0}^{m} \frac{\frac{d^k}{dt^k}\Re(\varphi_X(0))}{k!}\, t^k \;+\; \frac{t^m}{m!}\left\{\frac{d^m}{dt^m}\Re(\varphi_X(\widetilde{t})) \;-\; \frac{d^m}{dt^m}\Re(\varphi_X(0))\right\}$$

für ein $\widetilde{t}$ zwischen t und 0. Wegen Stetigkeit der m-ten Ableitung von φ_X folgt, dass der Term in geschweiften Klammern von der Ordnung $o(|t|^m)$ ist. Analog kann man den Imaginärteil in eine Taylorreihe entwickeln und zum Schluss werden beide Entwicklungen zusammengesetzt. □

An dieser Stelle sei noch die folgende einfache Rechenregel angefügt.

Lemma 1.20. *X sei eine $\mathbb{R}^d$-wertige Zufallsvariable mit charakteristischer Funktion φ_X. A sei eine $(k\times d)$-Matrix und $b \in \mathbb{R}^k$. Dann besitzt die Zufallsvariable $Y = AX + b$ die charakteristische Funktion φ_Y mit*

$$\varphi_Y(t) \;=\; e^{it^Tb}\varphi_X(A^Tt) \quad \forall t \in \mathbb{R}^k.$$

Beweis. Es gilt

$$\varphi_Y(t) \;=\; E^{it^T(AX+b)} \;=\; E[e^{it^Tb}\, e^{i(A^Tt)^TX}] \;=\; e^{it^Tb}\varphi_X(A^Tt) \quad \forall t \in \mathbb{R}^k.$$

□

Beispiel

Falls $X \sim N(0,1)$, so gilt für die charakteristische Funktion φ_Y von $Y = \sigma X + b$, dass

$$\varphi_Y(t) \;=\; e^{itb}\, e^{-\sigma^2t^2/2} \quad \forall t \in \mathbb{R}.$$

2 Grenzwertsätze

In diesem Kapitel werden eine Reihe von wichtigen Grenzwertsätzen hergeleitet. Dabei wird es sich als extrem hilfreich erweisen, dass in Theorem 1.15 gezeigt wurde, dass die schwache Konvergenz einer Folge von Zufallsvariablen äquivalent zur punktweisen Konvergenz der entsprechenden charakteristischen Funktionen ist. Der Zugewinn wird sich sehr bald beim Beweis verschiedener Versionen des Zentralen Grenzwertsatzes offenbaren. In der Vorlesung „Elementare Wahrscheinlichkeitstheorie und Mathematische Statistik“ im Sommersemester 2019 wurde mit dem Grenzwertsatz von de Moivre-Laplace eine Vorstufe zum Zentralen Grenzwertsatz bewiesen, welcher eine Approximation der Binomialverteilung durch die Normalverteilung beinhaltet. Ohne das Hilfsmittel der charakteristischen Funktionen war der rein analytische Beweis dieser Aussage recht mühsam. Dagegen wird sich der Beweis der einfachsten Version des Zentralen Grenzwertsatzes, wovon der Satz von de Moivre-Laplace ein Spezialfall ist, als *“no-brainer”* erweisen und in wenigen Zeilen erledigt sein. Selbst bei komplizierteren Strukturen wie beim Beweis eines Zentralen Grenzwertsatzes für Dreiecksschemata von unabhängigen Zufallsvariablen wird sich der Aufwand in Grenzen halten. Mit Hilfe des Satzes von Cramér-Wold (Theorem 1.18), welcher ebenfalls mit dem Hilfsmittel charakteristischer Funktionen bewiesen wurde, wurde die Äquivalenz zwischen der Verteilungskonvergenz von Zufallsvektoren und der Verteilungskonvergenz der Linearkombinationen der jeweiligen Komponenten dieser Zufallsverktoren hergestellt. Dies wird die Herleitung eines multivariaten Zentralen Grenzwertsatzes (für Zufallsvektoren) aus einem ZGWS für Summen reellwertiger Zufallsgrößen ermöglichen. Schließlich sei noch erwähnt, dass sich das Hilfsmittel der charakteristischen Funktionen selbst im Zusammenhang mit abhängigen Zufallsvariablen nutzen lässt. In der Vorlesung „Zeitreihenanalyse“ wird dieses Hilfsmittel zum Beweis eines Zentralen Grenzwertsatzes für Summen abhängiger Zufallsvariablen eingesetzt. Abschließend sei jedoch darauf hingewiesen, dass trotz der einfach erscheinenden Beweise in den nächsten Abschnitten der „Erhaltungssatz der Schwierigkeit“ nicht verletzt wurde: Die „harte Arbeit“ wurde in den Abschnitten 1.3 und 1.4 geleistet!

Bevor wir nun die angekündigten Grenzwertsätze herleiten, soll noch ein technisches Hilfsmittel bereitgestellt werden. Sicherlich ist bekannt, dass

$$\left(1 + \frac{x}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e^x \quad \forall x \in \mathbb{R}$$

gilt. Im Hinblick auf die folgenden Beweise wird dieses Resultat verallgemeinert.

Lemma 2.1. $(c_n)_{n\in\mathbb{N}}$ *sei eine Folge komplexer Zahlen mit* $c_n \underset{n\to\infty}{\longrightarrow} c \in \mathbb{C}$. *Dann gilt*

$$\left(1 + \frac{c_n}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e^c.$$

Beweis. Zunächst erhalten wir mit dem binomischen Lehrsatz

$$\left(1 + \frac{c_n}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{c_n}{n}\right)^k = \sum_{k=0}^{\infty} \frac{c_n^k}{k!} \underbrace{\frac{n!}{(n-k)!n^k} \mathbb{1}(n \geq k)}_{=:\alpha_{n,k}}.$$

Nun gelten $0 \leq \alpha_{n,k} \leq 1$ und $\alpha_{n,k} \underset{n\to\infty}{\longrightarrow} 1$. Außerdem ist die Folge $(c_n)_{n\in\mathbb{N}}$ beschränkt. Folglich gilt mit dem Satz von Lebesgue über majorisierte Konvergenz, dass

$$\left(1 + \frac{c_n}{n}\right)^n \underset{n\to\infty}{\longrightarrow} \sum_{k=0}^{\infty} \frac{c^k}{k!}.$$

Um den Grenzwert der unendlichen Reihe auf der rechten Seite zu bestimmen, zerlegen wir $c = a + ib$, wobei $a, b \in \mathbb{R}$ und nutzen, dass $\cos(b) = \sum_{s=0}^{\infty}(-1)^s b^{2s}/(2s)!$ sowie $\sin(b) = \sum_{s=0}^{\infty}(-1)^s b^{2s+1}/(2s+1)!$ gelten. Damit erhalten wir

$$\begin{aligned}
\sum_{k=0}^{\infty} \frac{c^k}{k!} &= \sum_{k=0}^{\infty} \frac{(a+ib)^k}{k!} \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{l=0}^{k} \binom{k}{l} a^l \, (ib)^{k-l} \\
&= \underbrace{\sum_{r=0}^{\infty} \frac{1}{r!} a^r}_{=e^a} \quad \times \quad \underbrace{\sum_{s=0}^{\infty} \frac{1}{s!} (ib)^s}_{=\cos(b)+i\sin(b)} \\
&= e^a \, e^{ib} = e^c.
\end{aligned}$$

Hierbei sei angemerkt, dass alle beteiligten Reihen absolut konvergieren, was die vorgenommene Vertauschung der Summationsreihenfolge erlaubt. □

2.1 Der Zentrale Grenzwertsatz

Zunächst beweisen wir hier die einfachste Version des Zentralen Grenzwertsatzes.

Theorem 2.2. *(Zentraler Grenzwertsatz von Lindeberg-Lévy)*
$(X_n)_{n\in\mathbb{N}}$ *sei eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit* $EX_n = 0$ *und* $EX_n^2 =: \sigma^2 \in (0, \infty)$ $\forall n \in \mathbb{N}$.

Dann gilt

$$Y_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{d} Y \sim N(0, \sigma^2).$$

Beweis. Nach Theorem 1.19 kann die charakteristische Funktion von X_k in folgende Taylorreihe entwickelt werden.

$$\varphi_{X_k}(t) \;=\; 1 + \underbrace{i\,E[X_k]\,t}_{=0} - \underbrace{\frac{EX_k^2}{2}\,t^2}_{=\sigma^2 t^2/2} + o(t^2).$$

Nach Lemma 1.11 und Lemma 2.1 folgt

$$\varphi_{Y_n}(t) \;=\; \prod_{k=1}^{n} \varphi_{X_k}(t/\sqrt{n}) \;=\; \left(1 - \frac{\sigma^2 t^2}{2n} + o(n^{-1})\right)^n \;\underset{n\to\infty}{\longrightarrow}\; e^{-\sigma^2 t^2/2} \quad \forall t \in \mathbb{R}.$$

Nach Lemma 1.20 gilt jedoch $\varphi_Y(t) = e^{-\sigma^2 t^2/2}$, woraus mit Theorem 1.15 die Behauptung folgt. □

Eine Verallgemeinerung dieses Satzes, welche insbesodere Summen nicht notwendigerweise identisch verteilter Zufallsvariablen mit einbezieht, ist durch den folgenden Satz gegeben. Ein weiteres hervorzuhebendes Merkmal ist, dass dieser Satz auf Dreiecksschemata von Zufallsvariablen anwendbar ist, d.h., mit sich veränderndem Stichprobenumfang n darf sich auch die Verteilung der beteiligten Zufallsgrößen ändern.

Theorem 2.3. *(Zentraler Grenzwertsatz von Lindeberg-Feller)*

Für jedes $n \in \mathbb{N}$ seien $(X_{n,k})_{k=1,\dots,k_n}$ stochastisch unabängige Zufallsvariable mit $EX_{n,k} = 0$ und $EX_{n,k}^2 =: \sigma_{n,k}^2 < \infty$. Ferner mögen folgende Bedingungen erfüllt sein:

a) $\sigma_n^2 \;=\; \sum_{k=1}^{k_n} \sigma_{n,k}^2 \;\underset{n\to\infty}{\longrightarrow}\; \sigma^2 \in [0,\infty).$

b) *(Lindeberg-Bedingung)*

$$L_n(\epsilon) \;:=\; \sum_{k=1}^{k_n} E[X_{n,k}^2 \mathbb{1}(|X_{n,k} > \epsilon)] \;\underset{n\to\infty}{\longrightarrow}\; 0 \quad \forall \epsilon > 0.$$

Dann gilt

$$S_n \;:=\; X_{n,1} + \cdots + X_{n,k_n} \;\xrightarrow{d}\; Y \sim N(0,\sigma^2).$$

Bemerkung 2.4. (i) Bei Voraussetzung (i) ist der Fall $\sigma^2 = 0$ durchaus zugelassen. In diesem Fall hat Y eine Dirac-Verteilung im Punkt 0. Der Einschluss dieser Möglichkeit ist angenehm, wenn zunächst asymptotische Normalität gezeigt werden soll und die Varianz der Grenzverteilung anschließend oder auch gar nicht explizit bestimmt wird.

(ii) Die Lindeberg-Bedingung impliziert, dass

$$\max_{1\le k\le k_n} \{\sigma_{n,k}^2\} \;\underset{n\to\infty}{\longrightarrow}\; 0,$$

d.h., mit $n \to \infty$ strebt der Beitrag jedes einzelnen Summanden $X_{n,k}$ zur Varianz σ_n^2 der Partialsumme gegen Null. Es gilt nämlich für beliebiges $\epsilon > 0$

$$\sigma_{n,k}^2 \;\le\; \epsilon^2 + E[X_{n,k}^2 \mathbb{1}(|X_{n,k}| > \epsilon)] \;\le\; \epsilon^2 + L_n(\epsilon) \;\le\; 2\epsilon$$

für alle hinreichend großen n.

Beweis von Theorem 2.3. Für jedes $n \in \mathbb{N}$ seien $(Y_{n,k})_{k=1,\dots,k_n}$ stochastisch unabängig mit $Y_{n,k} \sim N(0, \sigma^2_{n,k})$. Der Hauptteil des Beweises besteht nun in einem Vergleich der charakteristischen Funktionen von $\sum_{k=1}^{k_n} X_{n,k}$ und $T_n := \sum_{k=1}^{k_n} Y_{n,k}$. Dabei wird sich zeigen, dass die jeweilige Übereinstimmung der ersten beiden Momente, nicht jedoch das jeweilige Verteilungsgesetz der $X_{n,k}$ ausschlaggebend ist.

Es sei $t \in \mathbb{R}$ beliebig. Nun gilt

$$\begin{aligned} |\varphi_{S_n}(t) - \varphi_Y(t)| &\leq |\varphi_{S_n}(t) - \varphi_{T_n}(t)| + |\varphi_{T_n}(t) - \varphi_Y(t)| \\ &=: R_{n,1} + R_{n,2}. \end{aligned}$$

Wegen $\sigma_n^2 \underset{n\to\infty}{\longrightarrow} \sigma^2$ folgt sofort

$$R_{n,2} = \left| e^{-\sigma_n^2 t^2/2} - e^{-\sigma^2 t^2/2} \right| \underset{n\to\infty}{\longrightarrow} 0. \tag{6}$$

Nun müssen noch $\varphi_{S_n}(t)$ und $\varphi_{T_n}(t)$ verglichen werden. Da $EX^2_{n,k} < \infty$, so ist nach Theorem 1.19 die charakteristische Funktion von $X_{n,k}$ zweimal stetig differenzierbar und es gilt die folgende Taylorreihenentwicklung:

$$\varphi_{X_{n,k}}(t) = 1 - \frac{t^2}{2}\sigma^2_{n,k} + \underbrace{\frac{t^2}{2}\left\{\Re(\varphi^{(2)}_{X_{n,k}}(\theta_1(t)\,t)) + i\,\Im(\varphi^{(2)}_{X_{n,k}}(\theta_2(t)\,t)) - \varphi^{(2)}_{X_{n,k}}(t)\right\}}_{=:W_{n,k}},$$

wobei $0 \leq \theta_1(t), \theta_2(t) \leq 1$. Wegen $\varphi^{(2)}_{X_{n,k}}(u) = -E[X^2_{n,k} e^{iuX_{n,k}}]$ folgt

$$\begin{aligned} |W_{n,k}(t)| &\leq t^2 \sup_{0\leq u\leq t} E[X^2_{n,k} \underbrace{|e^{iuX_{n.k}} - 1|}_{\leq \min\{u|X_{n,k}|,\,2\}}] \\ &\leq t^2 \left\{\epsilon\, t\, EX^2_{n,k} + 2\,E[X^2_{n,k}\mathbb{1}(|X_{n,k}| > \epsilon)]\right\}, \end{aligned}$$

woraus

$$\sum_{k=1}^{k_n} |W_{n,k}(t)| \underset{n\to\infty}{\longrightarrow} 0 \tag{7}$$

folgt. Auf analoge Weise erhalten wir

$$\varphi_{Y_{n,k}}(t) = 1 - \frac{t^2}{2}\sigma^2_{n,k} + \widetilde{W}_{n,k}(t),$$

wobei

$$|\widetilde{W}_{n,k}(t)| \leq t^3\, E|Y^3_{n,k}| = O(\sigma^2_{n,k} \max_{1\leq j\leq k_n}\{\sigma_{n,j}\}).$$

Damit erhalten wir

$$\sum_{k=1}^{k_n} |\widetilde{W}_{n,k}(t)| \leq \sum_{k=1}^{k_n} \sigma^2_{n,k} \max_{1\leq j\leq k_n}\{\sigma_{n,j}\} \underset{n\to\infty}{\longrightarrow} 0. \tag{8}$$

Nun können wir zur Abschätzung von $R_{n,1} = |\varphi_{S_n}(t) - \varphi_{T_n}(t)|$ kommen. Dabei nutzen wir, dass für beliebige komplexe Zahlen $a_1, \dots, a_m, b_1, \dots, b_m$ mit $|a_j| \le 1$, $|b_j| \le 1$ gilt, dass

$$\left| \prod_{j=1}^m a_j - \prod_{j=1}^m b_j \right| \le \sum_{j=1}^m \left| \left(\prod_{k=1}^{j-1} a_k \right) (a_j - b_j) \left(\prod_{k=j+1}^m b_k \right) \right| \le \sum_{j=1}^m |a_j - a_k|.$$

Folglich gilt

$$\begin{aligned} |\varphi_{S_n}(t) - \varphi_{T_n}(t)| &= \left| \prod_{j=1}^{k_n} \varphi_{X_{n,j}}(t) - \prod_{j=1}^{k_n} \varphi_{Y_{n,j}}(t) \right| \\ &\le \sum_{j=1}^{k_n} \left| \varphi_{X_{n,j}}(t) - \varphi_{Y_{n,j}}(t) \right| \\ &\le \sum_{j=1}^{k_n} |W_{n,k}(t)| + |\widetilde{W}_{n,k}(t)| \underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

□

Übungsaufgaben

ÜA 1.3 Die Zufallsvariable X möge eine Cauchy-Verteilung besitzen, d.h., die Dichte p_X bezüglich des Lebesgue-Maßes ist gegeben durch $p_X(x) = 1/((\pi(1+x^2))$.

Für welche $\alpha > 0$ gilt $E[|X|^\alpha] < \infty$?
(*Das zeigt, dass die Bedingung an die Momente in Theorem 1.19 wesentlich ist.*)

ÜA 1.4 Für $n \in \mathbb{N}$ und $1 \le k \le k_n := n$ seien $X_{n,k} \sim \mathrm{Bin}(1, p/n)$ sowie $Y_{n,k} = X_{n,k} - p/n$.

Zeigen Sie, dass die $Y_{n,k}$ die Lindeberg-Bedingung **nicht** erfüllen!
(*Nach ÜA 1.1 gilt, dass $X_{n,1} + \cdots + X_{n,n} \xrightarrow{d} X \sim$ Poisson(p) und somit $Y_{n,1} + \cdots + Y_{n,n} \xrightarrow{d} X - p$. Obwohl $EY_{n,k} = 0$ und $\sum_{k=1}^n \mathrm{var}(Y_{n,k}) = p(1-p/n) \underset{n\to\infty}{\longrightarrow} 1$ gelten, folgt nicht, dass $Y_{n,1} + \cdots + Y_{n,n} \xrightarrow{d} Y \sim N(0,1)$. Auf die Lindeberg-Bedingung in Theorem 2.3 kann also nicht verzichtet werden.*)

ÜA 1.5 $(X_{n,k})_{k=1,\dots,k_n}$, $n \in \mathbb{N}$ sei ein Dreiecksschema von Zufallsvariablen wie in Theorem 2.3.

Zeigen Sie, dass

$$\sum_{k=1}^{k_n} E\left[|X_{n,k}|^{2+\delta} \right] \underset{n\to\infty}{\longrightarrow} 0 \quad \text{für ein } \delta > 0$$

(Ljapunoff-Bedingung) die Lindeberg-Bedingung impliziert!

2.2 Mehrdimensionale Normalverteilung und der multivariate ZGWS

Die univariate Normalverteilung ist aus der Vorlesung „EWMS“ aus dem Sommersemester 2019 bereits bekannt. Eine reellwertige Zufallsvariable X besitzt eine Standardnormalverteilung ($X \sim N(0,1)$), falls X eine Wahrscheinlichkeitsdichte p_X bezüglich des Lebesgue-Maßes λ mit

$$p_X(x) \;=\; \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \forall x \in \mathbb{R}$$

besitzt. Die entsprechende charakteristische Funktion φ_X ist gegeben durch

$$\varphi_X(t) \;=\; e^{-t^2/2} \quad \forall t \in \mathbb{R},$$

siehe Abschnitt 1.1, Seite 7, des vorliegenden Skriptes.

Falls nun $X \sim N(0,1)$ und $Y = \sigma X + \mu$ mit, so besitzt Y eine Normalverteilung mit Parametern μ und σ^2 ($Y \sim N(\mu, \sigma^2)$). (Der Fall $\sigma = 0$ kann hier eingeschlossen werden, $N(\mu, 0)$ ist dann das Dirac-Maß im Punkt μ.) Nach der Dichtetransformationsformel (siehe Satz 7.6 aus der Vorlesung „EWMS“, SS 2019) besitzt Y im Falle von $\sigma \neq 0$ eine Dichte p_Y mit

$$p_Y(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \quad \forall x \in \mathbb{R}.$$

Die charakteristische Funktion φ_Y berechnet sich nach Lemma 1.20 folgendermaßen:

$$\varphi_Y(t) \;=\; e^{it\mu}\varphi_X(\sigma t) \;=\; e^{it\mu - \sigma^2 t^2/2} \quad \forall t \in \mathbb{R}.$$

An dieser Stelle soll nun eine mehrdimensionale Normalverteilung eingeführt werden. Am intuitivsten geschieht dies auf konstruktivem Wege. Es seien $X_1, \ldots, X_d \sim N(0,1)$ stochastisch unabhängig. Dann besitzt der Zufallsvektor $X = (X_1, \ldots, X_d)^T$ bezüglich des d-dimensionalen Lebesgue-Maßes λ^d eine Dichte p_X mit

$$p_X(x) \;=\; \prod_{j=1}^d \frac{1}{\sqrt{2\pi}} e^{-x_j^2/2} \;=\; \frac{1}{(2\pi)^{d/2}} e^{-\|x\|^2/2} \quad \forall x = (x_1, \ldots, x_d)^T \in \mathbb{R}^d.$$

Die charakteristische Funktion von X erhalten wir nach Theorem 1.10 folgendermaßen:

$$\varphi_X(t) \;=\; \prod_{j=1}^d \varphi_{X_j}(t_j) \;=\; \prod_{j=1}^d e^{-t_j^2/2} \;=\; e^{-\|t\|^2/2} \quad \forall t = (t_1, \ldots, t_d)^T \in \mathbb{R}^d.$$

Wir schreiben $X \sim N(0_d, I_d)$ oder, wenn auf die Dimension des Zufallsvektors X hingewiesen werden soll, $X \sim N_d(0_d, I_d)$. Wie auch im eindimensionalen Fall geben die Parameter den Erwartungswertvektor und die Kovarianzmatrix von X an. Die Verteilung von X ist die d-dimensionale Standardnormalverteilung.

Wie im univariaten Fall lässt sich nun auch eine mehrdimensionale Normalverteilung mit beliebigem Erwartungswertparameter $\mu \in \mathbb{R}^d$ und beliebiger Kovarianzmatrix Σ herleiten. Ausgehend von einer Spektraldarstellung von Σ mit

$$\Sigma \;=\; \Big(e_1 \dots e_d\Big)\,\mathrm{Diag}(\lambda_1, \dots, \lambda_d)\begin{pmatrix} e_1^T \\ \vdots \\ e_d^T \end{pmatrix} \;=\; \sum_{i=1}^d \lambda_i e_i e_i^T,$$

wobei $\lambda_1, \dots, \lambda_d$ die Eigenwerte (entsprechend ihrer Vielfachheit) von Σ und $\{e_1, \dots, e_d\}$ ein zugehöriges Orthonormalsystem von Eigenvektoren sind, können wir die Quadratwurzel von Σ definieren durch

$$\Sigma^{1/2} \;:=\; \sum_{i=1}^d \sqrt{\lambda_i} e_i e_i^T.$$

Da Σ als Kovarianzmatrix positiv semidefinit ist, sind die Eigenwerte $\lambda_1, \dots, \lambda_d$ reell und nichtnegativ. $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}$ sind folglich reell und können ebenfalls nichtnegativ gewählt werden. Die Matrix $\Sigma^{1/2}$ is offensichtlich eine symmetrische $(d \times d)$-Matrix und es gilt $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$. Falls nun $X \sim N(0_d, I_d)$, so legen wir fest, dass $Y = \Sigma^{1/2}X + \mu$ ebenfalls eine d-dimensionale Normalverteilung besitzt. Offensichtlich gelten

$$EY \;=\; \mu \qquad \text{und} \qquad \mathrm{Cov}(Y) \;=\; \Sigma^{1/2}\,\mathrm{Cov}(X)\Sigma^{1/2} \;=\; \Sigma.$$

Dementsprechend schreiben wir

$$Y \sim N(\mu, \Sigma).$$

Gemäß Lemma 1.20 ist die charakteristische Funktion φ_Y von Y gegeben durch

$$\varphi_Y(t) \;=\; e^{it^T\mu}\,\varphi_X(\Sigma^{1/2}t) \;=\; e^{it^T\mu - t^T\Sigma t/2} \quad \forall t \in \mathbb{R}^d.$$

Falls die Matrix Σ regulär ist, so besitzt Y eine Dichte bezüglich λ^d, welche sich nach einem Analogon zur Dichtetransformationsformel aus der Vorlesung „EWMS“ folgendermaßen ergibt:

$$p_Y(x) \;=\; \frac{1}{(2\pi)^{d/2}\det(\Sigma)} e^{-(x-\mu)^T\Sigma^{-1}(x-\mu)/2} \quad \forall x \in \mathbb{R}^d.$$

Falls jedoch Σ singulär ist, so ist die Verteilung $N(\mu, \Sigma)$ **nicht** absolut stetig bezüglich des Lebesgue-Maßes λ^d und besitzt demzufolge auch **keine** Dichte. Um dies zu sehen, betrachten wir die Spektralzerlegung der Kovarianzmatrix Σ,

$$\Sigma \;=\; \sum_{i=1}^m \lambda_i e_i e_i^T,$$

wobei $m < d$ gleich dem Rang der Matrix Σ ist. Gemäß obiger Konstruktion von Y erhalten wir

$$Y \;=\; \mu \,+\, \Sigma^{1/2}X \;=\; \mu \,+\, \sum_{i=1}^m \sqrt{\lambda_i}\, e_i\, (e_i^T X),$$

d.h., mit Wahrscheinlichkeit 1 liegt der Zufallsvektor Y im affinen Teilraum $M := \{\mu + \sum_{i=1}^m \alpha_i e_i\colon\ \alpha_1, \dots, \alpha_m \in \mathbb{R}\}$ des $\mathbb{R}^d$. Wegen $\lambda^d(M) = 0$ folgt, dass $N(\mu, \Sigma)$ nicht absolut stetig bezüglich λ^d ist.

Die folgende Aussage fasst eine Reihe von wichtigen Eigenschaften der multivariaten Normalverteilung zusammen.

Proposition 2.5. *Es sei* $X = (X_1, \dots, X_d)^T \sim N_d(\mu, \Sigma)$*, wobei* $\mu \in \mathbb{R}^d$ *beliebig und* Σ *eine positiv semidefinite* $(d \times d)$*-Matrix ist. Dann gelten:*

(i) $EX = \mu, \quad \mathrm{Cov}(X) = \Sigma.$

(ii) *Falls* A *eine* $(k \times d)$*-Matrix (*$k \in \mathbb{N}$ *beliebig),* $b \in \mathbb{R}^k$*, so*

$$AX + b \sim N_k(A\mu + b, A\Sigma A^T).$$

(iii) *Falls* $c \in \mathbb{R}^d$*, so*

$$c^T X \sim N(c^T\mu, c^T\Sigma c).$$

(iv) *Falls* $\{j_1, \dots, j_k\} \subseteq \{1, \dots, d\}$ *so*

$$\begin{pmatrix} X_{j_1} \\ \vdots \\ X_{j_k} \end{pmatrix} \sim N_k\left(\begin{pmatrix} \mu_{j_1} \\ \vdots \\ \mu_{j_k} \end{pmatrix}, \begin{pmatrix} \Sigma_{j_1,j_1} & \dots & \Sigma_{j_1,j_k} \\ \vdots & \ddots & \vdots \\ \Sigma_{j_k,j_1} & \dots & \Sigma_{j_k,j_k} \end{pmatrix}\right).$$

(v) *Es sind äquivalent:*

a) $X_1, \dots, X_d$ *sind stochastisch unabhängig,*

b) Σ *ist eine Diagonalmatrix,*

c) $X_1, \dots, X_d$ *sind paarweise unkorreliert.*

Ferner gilt: Falls $X_1 \sim N_d(\mu_1, \Sigma_1)$ *und* $X_2 \sim N_d(\mu_2, \Sigma_2)$ *stochastisch unabhängig sind, so*

$$X_1 + X_2 \sim N_d(\mu_1 + \mu_2, \Sigma_+\Sigma_2).$$

Beweis. Die Zufallsvariable X besitzt eine charakteristische Funktion φ_X mit

$$\varphi_X(t) = e^{it^T\mu - t^T\Sigma t/2} \quad \forall t \in \mathbb{R}^d.$$

Wir beweisen nun nacheinander die einzelnen Aussagen.
(i) Es seien $Y_1, \dots, Y_d \sim N(0,1)$ stochastisch unabhängig. Dann gilt

$$Y = (Y_1, \dots, Y_d)^T \sim N_d(0_d, I_d)$$

und $\widetilde{X} := \Sigma^{1/2}Y + \mu$ besitzt nach Definition eine $N_d(\mu, \Sigma)$-Verteilung. Offensichtlich gelten jedoch

$$EY = 0_d \qquad \text{und} \qquad \mathrm{Cov}(Y) = I_d,$$

woraus

$$E\widetilde{X} = \mu \qquad \text{und} \qquad \mathrm{Cov}(\widetilde{X}) = \Sigma$$

folgen. Da aber Xund $\widetilde{X}$ dieselbe Verteilung besitzen, so folgt, dass

$$EX = \mu \qquad \text{und} \qquad \mathrm{Cov}(X) = \Sigma.$$

(ii) Dazu berechnen wir einfach die charakteristische Funktion von $AX+b$. Nach Lemma 1.20 gilt

$$\begin{aligned}\varphi_{AX+b}(u) &= e^{iu^Tb}\varphi_X(A^Tu)\\ &= e^{iu^Tb}\,e^{iu^TA\mu-u^TA\Sigma A^Tu/2}\\ &= e^{iu^T(A\mu+b)-u^T(A\Sigma A^T)u/2} \qquad \forall u\in\mathbb{R}^k.\end{aligned}$$

Somit gilt $AX+b\sim N_k(A\mu+b, A\Sigma A^T)$.
(iii), (iv) folgen direkt aus (ii).
(v) [a) $\Longrightarrow$ $b)$] Falls $X_1,\dots,X_d$ stochastisch unabhängig sind, so gilt nach Theorem 1.10

$$\begin{aligned}\varphi_X(t) &= \prod_{j=1}^d \varphi_{X_j}(t_j) = \prod_{j=1}^d e^{it_j\mu_j-t_j^2\Sigma_{jj}/2}\\ &= e^{it^T\mu-\sum_{j=1}^d\Sigma_{jj}t_j^2/2} \qquad \forall t=(t_1,\dots,t_d)^T\in\mathbb{R}^d.\end{aligned}$$

Anderseits gilt nach Voraussetzung, dass

$$\varphi_X(t) = e^{it^T\mu-t^T\Sigma t/2} \quad \forall t\in\mathbb{R}^d.$$

Da Σ symmetrisch ist, muss Σ eine Diagonalmatrix sein.
[b) $\Longrightarrow$ $c)$] Für $j\neq k$ gilt $\mathrm{cov}(X_j,X_k)=\Sigma_{jk}=0$. Also sind $X_1,\dots,X_d$ paarweise unkorreliert.
[c) $\Longrightarrow$ $a)$] Falls $X_1,\dots,X_d$ paarweise unkorreliert sind, so folgt wegen $\Sigma=\mathrm{Cov}(X)$, dass Σ eine Diagonalmatrix ist.

Falls $X_1\sim N_d(\mu_1,\Sigma_1)$ und $X_2\sim N_d(\mu_2,\Sigma_2)$ stochastisch unabhängig sind, so folgt nach Lemma 1.11

$$\begin{aligned}\varphi_{X_1+X_2}(t) &= \varphi_{X_1}(t)\,\varphi_{X_2}(t)\\ &= e^{it^T\mu_1-t^T\Sigma_1t/2}\,e^{it^T\mu_2-t^T\Sigma_2t/2}\\ &= e^{it^T(\mu_1+\mu_2)-t^T(\Sigma_1+\Sigma_2)t/2} \qquad \forall t\in\mathbb{R}^d.\end{aligned}$$

Daraus folgt, dass $X_1+X_2\sim N_d(\mu_1+\mu_2,\Sigma_1+\Sigma_2)$. □

Übungsaufgabe

ÜA 1.6 Es sei $X\sim N_n(\mu,\sigma^2I_n)$, wobei $\mu\in\mathbb{R}^n$ und $\sigma^2>0$. A sei eine $(k\times n)$-Matrix und B eine $(l\times n)$-Matrix, wobei $k,l\in\mathbb{N}$.

(i) Bestimmen Sie die charakteristischen Funktionen von AX und BX!

(ii) Nehmen Sie jetzt an, dass $AB^T=0_{k\times l}$ gilt. ($0_{k\times l}$ bezeichnet die aus Nullen bestehende Matrix mit k Zeilen und l Spalten.)

Beweisen Sie, dass dann AX und BX stochastisch unabhängige Zufallsvektoren sind!

Hier kommt nun das mehrdimensionale Analogon zum Zentralen Grenzwertsatz von Lindeberg-Lévy.

Theorem 2.6. $(X_n)_{n\in\mathbb{N}}$ *sei eine Folge von unabhängigen und identisch verteilten Zufallsvektoren mit Werten in* $\mathbb{R}^d$*. Es seien* $EX_n = 0_d$ *und* $\mathrm{Cov}(X_n) = \Sigma\ \forall n \in \mathbb{N}$*. Dann gilt*

$$Y_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{d} Y \sim N_d(0_d, \Sigma).$$

Beweis. Der Beweis basiert im Wesentlichen auf dem bereits bewiesenen Zentralen Grenzwertsatz von Lindeberg-Lévy (Theorem 2.2). Nach dem Satz von Cramér-Wold (Theorem 1.18) genügt es zu zeigen, dass

$$c^T Y_n \xrightarrow{d} c^T Y \quad \forall c \in \mathbb{R}^d. \tag{9}$$

Es sei $c \in \mathbb{R}^d$ beliebig. Nach Übungsaufgabe 1.6 gilt $c^T Y \sim N(0, c^T \Sigma c)$. Da nun $(c^T X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit $E[c^T X_n] = 0$ und $\mathrm{var}(c^T X_n) = c^T \Sigma c$ ist, so folgt nach Theorem 2.3, dass

$$c^T Y_n = \frac{c^T X_1 + \cdots + c^T X_n}{\sqrt{n}} \xrightarrow{d} c^T Y.$$

Damit folgt die Behauptung. □

2.3 Gesetze der Großen Zahlen

In diesem Kapitel werden wir die aus der Vorlesung „EWMS“ bereits bekannten Gesetze der Großen Zahlen verschärfen. Anstelle der damals vorausgesetzten Quadratintegrierbarkeit der beteiligten Zufallsvariablen werden wir diese Gesetze unter der schwächeren Voraussetzung der Integrierbarkeit beweisen. Für das Schwache Gesetz der Großen Zahlen ist diese Verallgemeinerung mit dem Hilfsmittel der charakteristischen Funktionen eine einfache „Fingerübung“:

Proposition 2.7. $(X_n)_{n\in\mathbb{N}}$ *sei eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit* $EX_n = \mu \in (-\infty, \infty)$*. Dann gilt*

$$\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{P} \mu.$$

Beweis. Nach Theorem 1.19 kann wegen $E|X_j| < \infty$ die charakteristische Funktion von X_j in folgende Taylorreihe entwickelt werden:

$$\varphi_{X_j}(u) = 1 + iu\mu + o(|u|).$$

Daraus folgt mit Lemma 2.1, dass für beliebiges $t \in \mathbb{R}$

$$\begin{aligned}
\varphi_{n^{-1}\sum_{j=1}^n X_j}(t) &= \varphi_{\sum_{j=1}^n X_j}(t/n) = \prod_{j=1}^n \varphi_{X_j}(t/n) \\
&= \Big(1 + i(\frac{t}{n})\mu + o(\frac{|t|}{n})\Big)^n \\
&= \Big(1 + \frac{it\mu + o(1)}{n}\Big)^n \underset{n\to\infty}{\longrightarrow} e^{it\mu} = \widehat{\delta_\mu}(t)
\end{aligned}$$

folgt, wobei δ_μ die Dirac-Verteilung im Punkt μ bezeichnet. Nach dem Stetigkeitssatz (Theorem 1.15) erhalten wir

$$\frac{1}{n}\sum_{j=1}^n X_j \xrightarrow{d} \mu,$$

woraus gemäss der folgenden Übungsaufgabe 1.7

$$\frac{1}{n}\sum_{j=1}^n X_j \xrightarrow{P} \mu$$

folgt. □

Übungsaufgabe

ÜA 1.7 Auf einem Wahrscheinlichkeitraum $(\Omega, \mathcal{A}, P)$ sei eine Folge $(Y_n)_{n\in\mathbb{N}}$ von Zufallsvariablen mit Werten in $\mathbb{R}^d$ gegeben. Zeigen Sie, dass aus $Y_n \xrightarrow{d} \mu$ für ein $\mu \in \mathbb{R}^d$ die Eigenschaft

$$Y_n \xrightarrow{P} \mu$$

folgt!

In der Vorlesung EWMS im Sommersemester 2019 wurde das Starke Gesetz der Großen Zahlen für Folgen von Zufallsvariablen mit beschränkten zweiten Momenten bewiesen. Hier folgt eine auf Etamadi zurückgehende Verschärfung dieser Aussage, wobei lediglich die Integrierbarkeit der Zufallsvariablen gefordert wird.

Theorem 2.8. *(Starkes Gesetz der Großen Zahlen von Etemadi)*
$(X_n)_{n\in\mathbb{N}}$ sei eine Folge reeller, identisch verteilter und paarweise unabhängiger Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und EX_1 sei endlich.
Dann genügt $(X_n)_{n\in\mathbb{N}}$ dem Starken Gesetz der Großen Zahlen, d.h.,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \stackrel{P-f.s.}{\longrightarrow} EX_1.$$

Beweis. Es genügt zu zeigen, dass

$$\frac{1}{n}\sum_{i=1}^{n} X_i^+ \stackrel{P-f.s.}{\longrightarrow} EX_1^+ \tag{10a}$$

und

$$\frac{1}{n}\sum_{i=1}^{n} X_i^- \stackrel{P-f.s.}{\longrightarrow} EX_1^-. \tag{10b}$$

Diese Fokussierung auf nichtnegative Zufallsvariable liefert Monotonie der entsprechenden Partialsummenprozesse, was sich im letzten Beweisschritt als hilfreich erweisen wird. Wir beweisen o.B.d.A. (10a). (10b) folgt mit analogen Argumenten. Zunächst sei noch bemerkt, dass $(X_n^+)_{n\in\mathbb{N}}$ ebenfalls eine Folge reeller, identisch verteilter und paarweise unabhängiger Zufallsvariablen mit endlichem Erwartungswert ist. Der Hauptteil des Beweises wird der Übersichtlichkeit halber in zwei große Schritte aufgeteilt. Das zentrale Hilfsmittel beim Beweis der fast sicheren Konvergenz wird das aus der Vorlesung EWMS bekannte Lemma von Borel-Cantelli sein. Damit dieses Argument durchschlägt, beweisen wir das Konvergenzresultat zunächst für eine geeignete Teilfolge.

(i) (Konvergenz einer Teilfolge)

Es sei $\varepsilon > 0$ beliebig und $n_k := \lfloor (1+\varepsilon)^k \rfloor$, wobei $\lfloor a \rfloor$ die größte ganze Zahl, welche kleiner oder gleich a ist, bezeichnet. Wir zeigen zunächst, dass mit $k \to \infty$

$$\frac{1}{n_k}\sum_{i=1}^{n_k} X_i^+ \stackrel{P-f.s.}{\longrightarrow} EX_1^+ \tag{11}$$

gilt, d.h., die **Teilfolge** $(X_{n_k})_{k\in\mathbb{N}}$ genügt dem Starken Gesetz der Großen Zahlen. Da für die Zufallsvariablen X_i nur die Integrierbarkeit, nicht jedoch die Endlichkeit des zweiten Moments vorausgesetzt wurde, wenden wir einen typischen Abschneidetrick an und definieren gestutzte Zufallsvariable

$$Y_i := X_i^+ \mathbb{1}(X_i^+ \le i).$$

Im Weiteren betrachten wir entsprechende Partialsummen,

$$S_m = \sum_{i=1}^{m} (Y_i - EY_i).$$

Bevor wir nun (11) beweisen, leiten wir noch zwei dazu benötigte Abschätzungen her. Wegen $n_k = \lfloor (1+\varepsilon)^k \rfloor \geq (1+\varepsilon)^k/2$ folgt

$$\sum_{k:\ n_k \geq m} \frac{1}{n_k^2} \leq \sum_{k:\ (1+\varepsilon)^k \geq m} 4\,(1+\varepsilon)^{-2k} \leq \frac{1}{m^2}\,\frac{4}{1-(1+\varepsilon)^{-2}}. \tag{12}$$

Außerdem gilt wegen $\sum_{m=i}^{\infty} \frac{1}{m^2} \leq \frac{1}{i^2} + \int_i^{\infty} \frac{1}{x^2}\,dx = \frac{1}{i^2} + \frac{1}{i} \leq \frac{2}{i}$, dass

$$\begin{aligned}
\sum_{m=1}^{\infty} \frac{EY_m^2}{m^2} &= \sum_{m=1}^{\infty} \frac{1}{m^2} \sum_{i=1}^{m} E\left[X_1^{+2}\,\mathbb{1}\big(X_1^+ \in (i-1,i]\big)\right] \\
&\leq \sum_{m=1}^{\infty} \sum_{i=1}^{m} \frac{i}{m^2}\, E\left[X_1^+\,\mathbb{1}\big(X_1^+ \in (i-1,i]\big)\right] \\
&= \sum_{i=1}^{\infty} E\left[X_1^+\,\mathbb{1}\big(X_1^+ \in (i-1,i]\big)\right] \underbrace{\left(\sum_{m:\ m\geq i} \frac{i}{m^2}\right)}_{\leq 2} \leq 2\, EX_1^+. \qquad (13)
\end{aligned}$$

Es sei nun $\delta > 0$ beliebig. Mit der Tschebyscheff-Ungleichung erhalten wir

$$P\left(\left|\frac{S_{n_k}}{n_k}\right| > \delta\right) \leq \frac{1}{\delta^2}\ \mathrm{var}\left(\frac{S_{n_k}}{n_k}\right) \leq \frac{1}{\delta^2\, n_k^2} \sum_{m=1}^{n_k} EY_m^2.$$

Mit den Abschätzungen (13) und (14) erhalten wir nun

$$\begin{aligned}
\sum_{k=1}^{\infty} P\left(\left|\frac{S_{n_k}}{n_k}\right| > \delta\right) &\leq \frac{1}{\delta^2} \sum_{k=1}^{\infty} \frac{1}{n_k^2} \sum_{m=1}^{n_k} EY_m^2 \\
&= \frac{1}{\delta^2} \sum_{m=1}^{\infty} \underbrace{\left(\sum_{k:\ n_k \geq m} \frac{1}{n_k^2}\right)}_{\leq \frac{1}{m^2}\,\frac{4}{1-(1-\varepsilon)^{-2}}} EY_m^2 \\
&\leq \frac{8}{\delta^2\,(1-(1-\varepsilon)^{-2})}\, EX_1^2. \qquad (14)
\end{aligned}$$

Es sei $\Omega_\delta = \{\omega\colon\ |S_{n_k}(\omega)/n_k| > \delta$ für unendlich viele $k\}$. Aus (14) folgt nach dem Lemma von Borel-Cantelli, dass

$$P(\Omega_\delta) = 0 \quad \forall \delta > 0.$$

Daraus folgt jedoch

$$P\left(\frac{S_{n_k}}{n_k} \not\longrightarrow 0\right) = P\left(\bigcup_{m=1}^{\infty} \Omega_{1/m}\right) \leq \sum_{m=1}^{\infty} P(\Omega_{1/m}) = 0. \tag{15}$$

Wir müssen jetzt noch zeigen, dass die Stutzung der Zufallsgrößen X_i^+ asymptotisch vernachlässigbar ist. Nach dem Satz über monotone Konvergenz gilt

$$EY_i = E[X_1^+ \mathbb{1}(X_1^+ \le i)] \underset{i\to\infty}{\longrightarrow} EX_1^+,$$

woraus auch

$$\frac{1}{n}\big(EY_1 + \cdots + EY_n\big) \underset{n\to\infty}{\longrightarrow} EX_1^+$$

folgt. Somit erhalten wir aus (15), dass

$$\frac{1}{n_k}\sum_{i=1}^{n_k} Y_i \overset{P-f.s.}{\longrightarrow} EX_1^+. \tag{16}$$

Weiterhin gilt

$$\begin{aligned}
\sum_{n=1}^{\infty} P(X_n^+ \ne Y_n) &= \sum_{n=1}^{\infty} P(X_n^+ > n) \\
&= \sum_{n=1}^{\infty}\sum_{i=1}^{\infty} P\big(X_n^+ \in (i,i+1]\big) \\
&= \sum_{i=1}^{\infty} P\big(X_1^+ \in (i,i+1]\big)\ \#\{n\in\mathbb{N}\colon\ n \le i\} \\
&= \sum_{i=1}^{\infty} i\ P\big(X_1^+ \in (i,i+1]\big) \le EX_1^+ < \infty,
\end{aligned}$$

woraus mit dem Lemma von Borel-Cantelli folgt, dass

$$P\big(X_n^+ \ne Y_n \text{ für unendlich viele } n\big) = 0.$$

Daher folgt aus (16), dass (11) gilt.

(ii) (Konvergenz der vollen Folge)

Wir beginnen wieder mit einer Nebenrechnung. Da offensichtlich $(1+\varepsilon)^k\varepsilon \ge 1+2\varepsilon$ $\forall k \ge k_0(\varepsilon)$ und $k_0(\varepsilon)$ hinreichend groß gilt, so folgt

$$n_{k+1} \le (1+\varepsilon)^{k+1} \le (1+\varepsilon)^{k+1}+(1+\varepsilon)^k\varepsilon-(1+2\varepsilon) = [(1+\varepsilon)^k-1](1+2\varepsilon) \le n_k(1+2\varepsilon) \quad \forall k \ge k_0. \tag{17}$$

Für $n > n_1$ existiert stets ein $k(n)\in\mathbb{N}$ mit $n_{k(n)} < n \le n_{k(n)+1}$. Falls nun n so groß, dass $k(n) \ge k_0$, so folgen

$$\frac{n_{k(n)+1}}{n} < \frac{n_{k(n)+1}}{n_{k(n)}} \le 1+2\varepsilon \qquad \text{und} \qquad \frac{n_{k(n)}}{n} \ge \frac{n_{k(n)}}{n_{k(n)+1}} \ge \frac{1}{1+2\varepsilon}.$$

Wegen Monotonie des Partialsummenprozesses $\big(\sum_{i=1}^n X_i^+\big)_{n\in\mathbb{N}}$ erhalten wir

$$\frac{1}{1+2\varepsilon}\,\frac{1}{n_{k(n)}}\sum_{i=1}^{n_{k(n)}} X_i^+ \le \frac{1}{n}\sum_{i=1}^{n} X_i^+ \le (1+2\varepsilon)\,\frac{1}{n_{k(n)+1}}\sum_{i=1}^{n_{k(n)+1}} X_i^+ \quad \forall n > n_{k_0}. \tag{18}$$

Nach (11) konvergiert die linke Seite von (18) gegen $\frac{1}{1+2\varepsilon}EX_1^+$, die rechte gegen $(1+2\varepsilon)EX_1^+$. Da jedoch $\varepsilon > 0$ beliebig war, so folgt schließlich

$$\frac{1}{n}\sum_{i=1}^{n} X_i^+ \overset{P-f.s.}{\longrightarrow} EX_1^+,$$

d.h., (10a) ist bewiesen. □

Die Aussage von Theorem 2.8 kann dahingehend weiter verschärft werden, dass die bloße Existenz des Erwartungswertes von X_1 ausreicht. Im Falle von $EX_1 = \pm\infty$ konvergiert die Folge der arithmetischen Mittel mit Wahrscheinlichkeit 1 entsprechend gegen $\pm\infty$.

Proposition 2.9. $(X_n)_{n\in\mathbb{N}}$ *sei eine Folge reeller, identisch verteilter und paarweise unabhängiger Zufallsvariablen auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *und es gelte* $EX_1 \in \{-\infty, \infty\}$.
Dann gilt

$$\frac{1}{n}\sum_{i=1}^{n} X_i \overset{P-f.s.}{\longrightarrow} EX_1.$$

Beweis. Es sei o.B.d.A. $EX_1 = +\infty$. Die Existenz des Erwartungswertes bedeutet, dass dann $EX_1^- < \infty$ gilt. Da auch $(X_n^-)_{n\in\mathbb{N}}$ eine Folge reeller, identisch verteilter und paarweise unabhängiger Zufallsvariablen ist, so folgt aus Theorem 2.8

$$\frac{1}{n}\sum_{i=1}^{n} X_i^- \overset{P-f.s.}{\longrightarrow} EX_1^-.$$

Es verbleibt zu zeigen, dass

$$\frac{1}{n}\sum_{i=1}^{n} X_i^+ \overset{P-f.s.}{\longrightarrow} \infty. \tag{19}$$

Wir betrachten die gestutzten Zufallsvariablen $X_i^+ \wedge M$ für $M < \infty$. $(X_m^+ \wedge M)_{n\in\mathbb{N}}$ ist wiederum eine Folge von identisch verteilten und paarweise unabhängigen Zufallsvariablen mit endlichem ERwartungswert. Somit gilt nach Theorem 2.6, dass

$$\frac{1}{n}\sum_{i=1}^{n}(X_i^+ \wedge M) \overset{P-f.s.}{\longrightarrow} E[X_1^+ \wedge M] \quad \forall M < \infty.$$

Wegen $\frac{1}{n}\sum_{i=1}^{n} X_i^+ \geq \frac{1}{n}\sum_{i=1}^{n}(X_i^+ \wedge M)$ gilt jedoch für beliebiges $M < \infty$

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^+ \geq E[X_1^+ \wedge M] \qquad P-\text{fast sicher}.$$

Anderseits gilt nach dem Satz von Lebesgue über majorisierte Konvergenz, dass

$$E[X_1^+ \wedge M] \underset{M\to\infty}{\longrightarrow} EX_1^+ = \infty,$$

woraus (19) unmittelbar folgt. □

Wir möchten diesen Abschnitt mit einer typischen und wichtigen Anwendung des Starken Gesetzes der Großen Zahlen abschließen. Wir setzen voraus, dass Zufallsvariable $X_1, \ldots, X_n$ jeweils die Verteilungsfunktion F besitzen. Diese kann approximiert („geschätzt") werden durch die sogenannte **empirische Verteilungsfunktion** F_n, wobei

$$F_n(x) \;=\; \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq x) \quad \forall x \in \mathbb{R}.$$

Theorem 2.10. *(Satz von Glivenko-Cantelli)*
$(X_n)_{n\in\mathbb{N}}$ *sei eine Folge reeller, identisch verteilter und paarweise unabhängiger Zufallsvariablen auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$*, wobei* X_1 *die Wahrscheinlichkeitsverteilungsfunktion* F *besitzen möge. Dann gilt*

$$\sup_{x\in\mathbb{R}} |F_n(x) \;-\; F(x)| \stackrel{P-f.s.}{\longrightarrow} 0.$$

Beweis. (i) (punktweise Konvergenz)
Theorem 2.8 liefert direkt die punktweise Konvergenz. Da für beliebiges $x \in \mathbb{R}$ $(\mathbb{1}(X_n \leq x))_{n\in\mathbb{N}}$ eine Folge von identisch verteilten und paarweise unabhängigen Zufallsvariablen ist, folgt

$$F_n(x) \stackrel{P-f.s.}{\longrightarrow} F(x). \tag{20a}$$

Ein analoges Konvergenzresultat gilt für linksseitige Grenzwerte. Wegen Stetigkeit von unten gilt

$$F(x-) \;=\; \sup_{y:\; y<x} F(y) \;=\; P\left(\bigcup_{y:\; y<x} \{\omega:\; X_1(\omega) \leq y\}\right) \;=\; P(X_1 < x).$$

Offensichtlich ist, dass $F_n(x-) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_1 < x)$. Nun ist $(\mathbb{1}(X_n \leq x))_{n\in\mathbb{N}}$ ebenfalls eine Folge von identisch verteilten und paarweise unabhängigen Zufallsvariablen und es folgt nach Theorem 2.8, dass

$$F_n(x-) \stackrel{P-f.s.}{\longrightarrow} F(x-). \tag{20b}$$

(ii) (gleichmäßige Konvergenz)

Im zweiten Schritt wird nun aus den punktweisen Konvergenzresultaten (20a) und (20b) die fast sichere gleichmäßige Konvergenz hergeleitet.
Für $N \in \mathbb{N}$ und $k \in \{1, \ldots, N-1\}$ sei

$$x_k \;:=\; F^{-1}(k/N),$$

wobei $F^{-1}(t) := \inf\{x \in \mathbb{R}\colon\; F(x) \geq t\}\ \forall t \in (0,1)$ die sogenannte **verallgemeinerte Inverse** von F ist. Zusätzlich setzen wir $x_0 = -\infty$ und $x_N = \infty$. Es sei

$$R_n \;=\; \max_{k=1,\ldots,N} \left\{|F_n(x_{k-1}) - F(x_{k-1})| \;+\; |F_n(x_k-) - F(x_k-)|\right\}.$$

Nach (20a) und (20b) folgt, dass

$$R_n \overset{P-f.s.}{\longrightarrow} 0,$$

was

$$R_n \leq \frac{1}{N} \qquad \forall n \geq n_0 \tag{21}$$

für ein $n_0 = n_0(N, \omega) \in \mathbb{N}$ impliziert.
Es sei nun $x \in \mathbb{R}$ beliebig. Dann existiert ein $k \in \{1, \ldots, N\}$ mit $x \in [x_{k-1}, x_k)$. Wegen Monotonie von F und F_n gelten

$$F_n(x) \leq F_n(x_k-) \leq F(x_k-) + R_n \leq F(x) + \frac{1}{N} + R_n$$

sowie

$$F_n(x) \geq F_n(x_{k-1}) \geq F(x_{k-1}) - R_n \geq F(x) - \frac{1}{N} - R_n.$$

Damit erhalten wir unter Berücksichtigung von (21), dass

$$\max_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{2}{N} \qquad \forall n \geq n_0(N, \omega).$$

Da dies für alle $N \in \mathbb{N}$ gilt, folgt die Behauptung. $\square$

Übungsaufgaben

ÜA 1.8 $(X_n)_{n \in \mathbb{N}}$ sei eine Folge von reellwertigen und stochastisch unabhängigen Zufallsvariablen mit einer jeweiligen Verteilungsfunktion F, F_n mit $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$ sei die empirische Verteilungsfunktion. $x_1, \ldots, x_k \in \mathbb{R}$ seien beliebig.
Wogegen konvergiert $\sqrt{n}\big(F_n(x_1) - F(x_1), \ldots, F_n(x_k) - F(x_k)\big)^T$, falls $n \to \infty$?

ÜA 1.9 $f\colon [0,1] \to \mathbb{R}$ sei eine stetige Funktion.
Zeigen Sie unter Benutzung des Starken Gesetzes der Großen Zahlen, dass die Folge der Bernstein-Polynome $(f_n)_{n \in \mathbb{N}}$, wobei $f_n(p) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k}$ ist, punktweise gegen f konvergiert!

Hinweis: Betrachten Sie eine Folge von unabhängigen Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ *mit* $X_n \sim$ *Bin*$(1, p)$ *auf* $(\Omega, \mathcal{A}, P)$ *und zeigen Sie, dass* $E\left[f\left(\frac{X_1 + \cdots + X_n}{n}\right)\right] \underset{n \to \infty}{\longrightarrow} f(p)$ *gilt!*

3 Stochastische Prozesse

In der Vorlesung „EWMS“ im Sommersemester 2019 und auch in den ersten Kapiteln dieses Skriptes haben wir uns mit der Modellierung von aufeinanderfolgenden, sich nicht gegenseitig beeinflussenden Zufallsexperimenten befasst. Dies wird beschrieben durch Folgen von unabhängigen und identisch verteilten Zufallsvariablen. Wichtige Gesetzmäßigkeiten waren die Gesetze der Großen Zahlen oder auch der Zentrale Grenzwertsatz.

Anderseits begegnet man in der Praxis häufig Phänomenen, welche sich über einen gewissen Zeitraum zufällig entwickeln. Neben gekünstelten Beispielen wie der Anzahl der Gewinne nach $1, 2, \ldots$ Glücksspielen sei hier beispielsweise die Entwicklung von Börsenkursen erwähnt. Ausgehend von der Entwicklung eines solchen Kurses bis hin zu einem Zeitpunkt t_0 ist jedoch dessen weitere Entwicklung nicht exakt bestimmbar, da diese von zahlreichen in der Zukunft liegenden Faktoren abhängt, welche zum Zeitpunkt t_0 noch nicht bekannt sind. Daher wird die Entwicklung von Börsenkursen üblicherweise durch Zufallsmodelle beschrieben. Klar ist, dass der (in solch einem Modell zufällige) Wert einer Aktie zum Zeitpunkt t_0 auch einen gewissen Einfluss auf den Wert zu nachfolgenden Zeitpunkten hat. Die Entwicklung von Börsenkursen über ein Zeitintervall $[T_0, T_1]$ wird daher nicht durch stochastisch unabhängige Zufallsvariable $(X_t)_{t\in[T_0,T_1]}$, sondern besser durch Zufallsgrößen, welche eine geeignete Abhängigkeitstruktur aufweisen, beschrieben. Dies führt zum Begriff des **stochastischen Prozesses**, welchen wir in diesem Kapitel behandeln werden. Stochastische Prozesse sind also mathematische Modelle zur Beschreibung und Untersuchung der zeitlichen und ggf. auch räumlichen Entwicklung zufallsbeeinflusster Systeme. Dementsprechend vielfältig sind die Anwendungen dieser Modelle in den Natur-, Wirtschafts-, Sozial- und Ingenieurwissenschaften.

3.1 Definitionen und einfache Beispiele

Definition 3.1. $(\Omega, \mathcal{A}, P)$ *sei ein Wahrscheinlichkeitsraum. Ein* **stochastischer Prozess** *ist eine Familie von Zufallsvariablen* $(X_t)_{t\in T}$ *mit Werten in einem gemeinsamen messbaren Raum* $(\Omega_X, \mathcal{A}_X)$*, wobei* T *eine beliebige nichtleere Indexmenge ist.*

- *Der Raum* $(\Omega_X, \mathcal{A}_X)$ *(bzw.* Ω_X*) heißt* **Zustandsraum** *des Prozesses.*
- *Für jedes* $\omega \in \Omega$ *heißt die Abbildung* $t \mapsto X_t(\omega)$ **Pfad** (**Trajektorie**, **Realisierung**) *des Prozesses* $(X_t)_{t\in T}$.
- *Für* $t_1, \ldots, t_k \in T$, $k \in \mathbb{N}$ *ist* $P^{X_{t_1},\ldots,X_{t_k}}$ *eine* **endlichdimensionale Verteilung** *des Prozesses* $(X_t)_{t\in T}$.

In diesem Kurs werden wir uns auf reellwertige stochastische Prozesse beschränken, d.h., der Zustandandsraum Ω_X ist gleich der Menge der reellen Zahlen $\mathbb{R}$. Als σ-Algebra $\mathcal{A}_X$ wählen wir das System der Borelmengen $\mathcal{B}$. Andere Möglichkeiten sind durch $\Omega_X = \mathbb{C}$ oder $\Omega_X = \mathbb{R}^d$ gegeben. In diesen Fällen sprechen wir von einem komplexwertigen bzw. vektorwertigen Prozess.
Hinsichtlich der Menge T werden wir uns in diesem Kurs auf Teilmengen von $\mathbb{R}$ beschränken, z.B. $\mathbb{R}$ selbst oder auch $T = [0, \infty)$, $T = \mathbb{N}$. In diesen Fällen kann der Parameter $t \in T$ als Zeit interpretiert werden, was tatsächlich typischen Anwendungen entspricht. Darüber hinaus kann T auch mehrdimensional sein, z.B. $T = \mathbb{R}^d$. Ein entsprechender Prozess heißt dann zufälliges Feld. Auch hier kann eine Komponente von t die Zeit darstellen, während die anderen Komponenten anders interpretiert werden, beispielsweise als räumliche Koordinaten.

Ein einfaches Beispiel für einen stochastischen Prozess ist durch eine Folge $(X_n)_{n\in\mathbb{N}}$ von unabhängigen und identisch verteilten Zufallsvariablen gegeben, welche auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ definiert sind. Damit hatten wir beispielsweise bereits bei der Formulierung der Gesetze der Großen Zahlen oder auch beim Zentralen Grenzwertsatz zu tun. Häufiger beabsichtigt man jedoch, mit einem stochastischen Prozess zufällige Phänomene zu modellieren, welche nicht stochastisch unabhängig sind. Ein wichtiges Beispiel für einen stochastischen Prozess ist durch die folgende Definition gegeben.

Definition 3.2. *Ein stochastischer Prozess* $W = (W_t)_{t\in[0,\infty)}$ *heißt* **Wiener Prozess** (**Standard-Brownsche Bewegung**) *auf* $(\Omega, \mathcal{A}, P)$, *falls*

(i) $W_0 = 0$ *P-fast sicher,*

(ii) *W besitzt unabhängige Zuwächse, d.h., für beliebige* $0 < t_1 < \ldots < t_k$, $k \in \mathbb{N}$ *sind* $W_{t_1} - W_0$, $W_{t_2} - W_{t_1}$, $\ldots$, $W_{t_k} - W_{t_{k-1}}$ *stochastisch unabhängig,*

(iii) $W_t - W_s \sim N(0, t-s)$ *für alle* $0 \leq s < t$,

(iv) *W besitzt P-fast sicher stetige Pfade, d.h.,* $t \mapsto W_t(\omega)$ *ist stetig für P-fast alle ω.*

Übungsaufgabe

ÜA 3.1 $(W_t)_{t\in[0,\infty)}$ sei ein Wiener Prozess auf $(\Omega, \mathcal{A}, P)$.
Bestimmen Sie die endlichdimensionalen Verteilungen $P^{X_{t_1},\ldots,X_{t_k}}$ für $t_1, \ldots, t_k \in [0, \infty)$, $k \in \mathbb{N}$!

Hinweis: Setzen zunächst voraus, dass $0 < t_1 \leq t_2 \leq \ldots \leq t_k$ *und bestimmen Sie zuerst die Verteilung von* $(W_{t_1}, \ldots, W_{t_k} - W_{t_{k-1}})^T$.

3.2 Der Satz von Kolmogoroff

Bisher wurde sowohl hier als auch in der Vorlesung „EWMS“ ein wesentlicher Punkt stillschweigend übergangen. Die **„Existenz“** eines Prozesses mit den obigen Eigenschaften oder selbst die Existenz einer Folge $(X_n)_{n\in\mathbb{N}}$ von unabhängigen und identisch verteilten Zufallsvariablen mit einer jeweiligen Verteilung $P^{X_n} = Q$ ist nicht ohne Weiteres klar. Dabei ist, wie in der Mathematik üblich, nicht die „physische Existenz“ eines solchen Prozesses gemeint, sondern nur die Tatsache, dass ein Prozess mit den gewünschten Eigenschaften **widerspruchsfrei konstruierbar** ist. Bevor wir einen Satz beweisen, der uns von all diesen Zweifeln befreien wird, sehen wir uns noch kurz zwei einfache Eigenschaften der Familie der endlichdimensionalen Verteilungen eines Prozesses an. Falls $(X_t)_{t\in T}$ ein (reellwertiger) stochastischer Prozess auf $(\Omega, \mathcal{A}, P)$ ist, so gelten

$$\begin{aligned} &P^{X_{t_1},\ldots,X_{t_k}}\big(B_1 \times \cdots \times B_{i-1} \times \mathbb{R} \times B_{i+1} \times \cdots \times B_k\big) \\ &\quad= P^{X_{t_1},\ldots,X_{t_{i-1}},X_{t_{i+1}},\ldots,X_{t_k}}\big(B_1 \times \cdots \times B_{i-1} \times B_{i+1} \times \cdots \times B_k\big) \\ &\qquad\qquad \forall t_1,\ldots,t_k \in T,\ \forall B_1,\ldots,B_{i-1},B_{i+1},\ldots,B_k \in \mathcal{B},\ \forall k \geq 2. \end{aligned} \tag{22}$$

und

$$\begin{aligned} &P^{X_{t_1},\ldots,X_{t_k}}\big(B_1 \times \cdots \times B_k\big) = P^{X_{t_{\pi(1)}},\ldots,X_{t_{\pi(k)}}}\big(B_{\pi(1)} \times \cdots \times B_{\pi(k)}\big) \\ &\qquad\qquad \forall t_1,\ldots,t_k \in T,\ \forall B_1,\ldots,B_k \in \mathcal{B},\ \forall \pi \in \mathcal{P}_k,\ \forall k \geq 2, \end{aligned} \tag{23}$$

wobei $\mathcal{P}_k = \{\pi\colon\ \{1,\ldots,k\} \to \{1,\ldots,k\}$ ist eine Permutation von k Elementen$\}$ ist. Während die Gültigkeit dieser Aussage auf den ersten Blick einleuchtend ist, ist deren Umkehrung weniger offensichtlich, jedoch in ihrer Tragweite kaum zu überschätzen. Wir definieren die folgende Verträglichkeitsbedingung.

Definition 3.3. *T sei eine nichtleere Menge. Dann ist*

$$\{\pi_{t_1,\ldots,t_k} \text{ ist ein Wahrscheinlichkeitsmaß auf } (\mathbb{R}^k, \mathcal{B}^k)\colon\ t_1,\ldots,t_k \in T,\ t_i \neq t_j \text{ für } i \neq j,\ k \in \mathbb{N}\}$$

eine **konsistente Familie von Verteilungen***, falls*

$$\begin{aligned} &\pi_{t_1,\ldots,t_k}\big(B_1 \times \cdots \times B_{i-1} \times \mathbb{R} \times B_{i+1} \times \cdots \times B_k\big) \\ &\quad= \pi_{t_1,\ldots,t_{i-1},t_{i+1},\ldots,t_k}\big(B_1 \times \cdots \times B_{i-1} \times B_{i+1} \times \cdots \times B_k\big) \\ &\qquad \forall t_1,\ldots,t_k \in T,\ t_i \neq t_j \text{ für } i \neq j,\ \forall B_1,\ldots,B_{i-1},B_{i+1},\ldots,B_k \in \mathcal{B},\ \forall k \geq 2. \end{aligned} \tag{24}$$

und

$$\begin{aligned} &\pi_{t_1,\ldots,t_k}\big(B_1 \times \cdots \times B_k\big) = \pi_{t_{\pi(1)},\ldots,t_{\pi(k)}}\big(B_{\pi(1)} \times \cdots \times B_{\pi(k)}\big) \\ &\qquad \forall t_1,\ldots,t_k \in T,\ t_i \neq t_j \text{ für } i \neq j,\ \forall B_1,\ldots,B_k \in \mathcal{B},\ \forall \pi \in \mathcal{P}_k,\ \forall k \geq 2, \end{aligned} \tag{25}$$

gelten.

Theorem 3.4. *T sei eine nichtleere Indexmenge.*

$$\{\pi_{t_1,\dots,t_k} \text{ ist ein Wahrscheinlichkeitsmaß auf } (\mathbb{R}^k, \mathcal{B}^k): \ t_1,\dots,t_k \in T,\ t_i \neq t_j \text{ für } i \neq j,\ k \in \mathbb{N}\}$$

sei eine konsistente Familie von Wahrscheinlichkeitsmaßen, d.h., die Verträglichkeitsbedingungen (24) und (25) seien erfüllt.
Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, auf welchem ein stochastischer Prozess $(X_t)_{t\in T}$ definiert werden kann mit

$$P^{X_{t_1},\dots,X_{t_k}} = \pi_{t_1,\dots,t_k} \qquad \forall t_1,\dots,t_k \in T,\ t_i \neq t_j \text{ für } i \neq j,\ \forall k \in \mathbb{N}.$$

Beweis. Bevor wir zu den entscheidenden Beweisschritten kommen, seien ein paar vorbereitende Definitionen und Erklärungen vorangestellt. Wir wählen den Grundraum Ω gleich dem Raum aller möglichen Pfade des Prozesses. Da die zu konstruierenden Zufallsvariablen X_t reellwertig sind, so definieren wir

$$\Omega := \Big\{\omega = (\omega_t)_{t\in T}: \quad \omega_t \in \mathbb{R} \quad \forall t \in T\Big\}. \tag{26}$$

Falls beispielweise $T = \mathbb{N}$ ist, so entspricht Ω der Menge aller Folgen reeller Zahlen, jedes $\omega \in \Omega$ ist somit eine Folge reeller Zahlen. Die Festlegung der σ-Algebra wird zunächst aufgeschoben und erst dann beschrieben, wenn etwas klarer geworden ist, wie weit die Vorgabe der gewünschten endlichdimensionalen Verteilungen das zu konstruierende Wahrscheinlichkeitsmaß P bestimmt. Der gesuchte Prozess $(X_t)_{t\in T}$ wird durch Abgreifen der Komponenten von ω definiert:

$$X_t(\omega) := \omega_t \qquad \forall t \in T, \forall \omega \in \Omega. \tag{27}$$

$\mathcal{A}$ und P müssen nun so gewählt werden, dass alle Zufallsvariablen X_t $(\mathcal{A}-\mathcal{B})$-messbar sind und dass die endlichdimensionalen Verteilungen $P^{X_{t_1},\dots,X_{t_k}}$ nicht im Widerspruch zu den „Vorgaben“ $\pi_{t_1,\dots,t_k}$ stehen. Aus der Vorlesung über Maßtheorie ist bekannt, dass die Messbarkeit aller Zufallsvariablen zur Messbarkeit der daraus bildbaren Zufallsvektoren äquivalent ist. Folglich muss

$$(X_{t_1},\dots,X_{t_k})^{-1}(B) = \big\{\omega:\ (\omega_{t_1},\dots,\omega_{t_k})^T \in B\big\} \in \mathcal{A} \tag{28}$$

für alle $B \in \mathcal{B}^k$, $t_1,\dots,t_k \in T$ und $k \in \mathbb{N}$ gelten. Wir bezeichnen im Weiteren mit

$$\mathcal{Z}_k := \Big\{\big\{\omega:\ (\omega_{t_1},\dots,\omega_{t_k})^T \in B\big\}\Big|\, t_1,\dots,t_k \in T,\ t_i \neq t_j \text{ für } i \neq j,\ B \in \mathcal{B}^k\Big\}$$

das System der **k-dimensionalen Zylindermengen** und mit

$$\mathcal{Z} := \bigcup_{k=1}^{\infty} \mathcal{Z}_k$$

das System aller **Zylindermengen**. Mit (28) und der Vorgabe für die endlichdimensionalen Verteilungen wird nun sofort klar, wie das Wahrscheinlichkeitsmaß P auf den Zylindermengen zu wählen ist:

$$P\left(\big\{\omega:\ (\omega_{t_1},\dots,\omega_{t_k})^T \in B\big\}\right) = P^{X_{t_1},\dots,X_{t_k}}(B) = \pi_{t_1,\dots,t_k}(B)$$
$$\forall t_1,\dots,t_k \in T,\ t_i \neq t_j \text{ für } i \neq j,\ \forall B \in \mathcal{B}^k, \forall k \in \mathbb{N}. \tag{29}$$

An dieser Stelle sei angemerkt, dass wegen der vorausgesetzten Verträglichkeitsbedingungen (24) und (25) die Festlegung (29) auf keinen Widerspruch stößt. Damit sind wir aber noch nicht fertig. Es ist leicht zu sehen, dass das System der Zylindermengen $\mathcal{Z}$ zwar eine Mengenalgebra, jedoch keine σ-Algebra in Ω ist. Nach dieser Beweisskizze wird noch ausführlich gezeigt, dass P_0 mit

$$P_0\big(\{\omega\colon\ (\omega_{t_1},\ldots,\omega_{t_k})^T \in B\}\big) \;=\; \pi_{t_1,\ldots,t_k}(B)$$
$$\forall t_1,\ldots,t_k \in T,\ t_i \neq t_j \text{ für } i \neq j,\ \forall B \in \mathcal{B}^k,\ \forall k \in \mathbb{N}. \tag{30}$$

ein Prämaß auf der Mengenalgebra $\mathcal{Z}$ ist. Eine explizite Festlegung von P über $\mathcal{Z}$ hinausgehend ist im Moment erst einmal nicht kanonisch, da die Familie $\{\pi_{t_1,\ldots,t_k}$ ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^k,\mathcal{B}^k)\colon\quad t_1,\ldots,t_k \in T,\ t_i \neq t_j$ für $i \neq j,\ k \in \mathbb{N}\}$ hierzu keinen direkten Hinweis gibt. Wir geben uns also bei der Wahl von $\mathcal{A}$ so bescheiden wie möglich und definieren

$$\mathcal{A} := \sigma\big(\mathcal{Z}\big). \tag{31}$$

Was nun die Festlegung von P auf den Mengen in $\mathcal{A} \setminus \mathcal{Z}$ betrifft, so können wir uns zurücklehnen und dem Satz von Carathéodory den Rest der Arbeit überlassen: Nach diesem Satz existiert eine Erweiterung des Prämaßes P_0 zu einem Maß P auf $\mathcal{A}$. Da darüber hinaus das System $\mathcal{Z}$ durchschnittsstabil ist, so ist nach dem Eindeutigkeitssatz der Maßtheorie diese Erweiterung eindeutig bestimmt. $\square$

Übungsaufgabe

ÜA 3.2 Zeigen Sie, dass P_0 ein Inhalt auf $\mathcal{Z}$ ist!

Hier wird jetzt noch der letzte fehlende Schritt, nämlich der Beweis, dass P_0 sogar ein Prämaß auf $\mathcal{Z}$ ist, nachgeliefert.

Lemma 3.5. *P_0 ist ein Prämaß auf $\mathcal{Z}$.*

Beweis. ÜA 3.2 enthält bereits die Aussage, dass P_0 ein Inhalt auf der Mengenalgebra $\mathcal{Z}$ ist. Damit sind die folgenden Eigenschaften bereits klar:

- $P_0(A) \geq 0 \quad \forall A \in \mathcal{Z}$,
- $P_0(\emptyset) \;=\; 0$,
- Falls $A, B \in \mathcal{Z}$ beliebige disjunkte Mengen sind, so folgt
  $P_0(A \cup B) \;=\; P_0(A) \;+\; P_0(B)$.

Es ist lediglich noch die Eigenschaft der σ-Additivität der Mengenfunktion P_0 nachzuweisen. Es seien also $A_1, A_2, \ldots \in \mathcal{Z}$ beliebige disjunkte Mengen mit $A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{Z}$. Es ist zu zeigen, dass dann

$$P_0(A) = \sum_{i=1}^{\infty} P_0(A_i) \tag{32}$$

gilt. Wegen endlicher Additivität von P_0 gilt für die rechte Seite von (32)

$$\sum_{i=1}^{\infty} P_0(A_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P_0(A_i) = \lim_{n\to\infty} P_0\big(A_1 \cup \cdots \cup A_n\big).$$

Da $P_0(A) \le P_0(\Omega) = 1$ gilt, ist (32) wegen Subtraktivität von P_0 äquivalent zu

$$P_0\big(\underbrace{A \setminus (A_1 \cup \cdots \cup A_n)}_{:=B_n}\big) \underset{n\to\infty}{\longrightarrow} 0, \tag{33}$$

was im Weiteren gezeigt wird. Die Mengen B_n bilden eine nichtwachsende Folge mit

$$\bigcap_{n=1}^{\infty} B_n = \emptyset. \tag{34}$$

Wir werden (33) indirekt beweisen. Da aus der Additivität von P_0 auf $\mathcal{Z}$ auch die Eigenschaft der Isotonie folgt, so nehmen wir an, dass ein $\varepsilon > 0$ existiert mit

$$P_0(B_n) \ge \varepsilon \quad \forall n \in \mathbb{N}. \tag{35}$$

Wir müssen nun noch einen Widerspruch zu $\bigcap_{n=1}^{\infty} B_n = \emptyset$ herleiten. [1]
Die Mengen B_n haben folgende Gestalt:

$$B_n = \big\{\omega: \quad (\omega_{t_1^{(n)}}, \ldots, \omega_{t_{k(n)}^{(n)}})^T \in C_n\big\},$$

wobei $t_1^{(n)}, \ldots, t_{k(n)}^{(n)} \in T$, $C_n \in \mathcal{B}^{k(n)}$ und $k(n) \in \mathbb{N}$ sind. Indem wir ggf. weitere Komponenten hinzunehmen können wir erreichen, dass $\{t_1^{(n)}, \ldots, t_{k(n)}^{(n)}\} \subseteq \{t_1^{(n+1)}, \ldots, t_{k(n+1)}^{(n+1)}\}$ für alle $n \in \mathbb{N}$ gilt. Zur Vereinfachung der Notation sei $(t_k)_{k\in\mathbb{N}}$ eine Folge von Elementen aus T mit $\{t_1^{(n)}, \ldots, t_{k(n)}^{(n)}\} = \{t_1, \ldots, t_{d_n}\}$ $\forall n \in \mathbb{N}$. Indem wir wiederum ggf. weitere Komponenten hinzunehmen, erreichen wir $d_{n+1} > d_n$ für alle $n \in \mathbb{N}$ gilt. Die Mengen B_n können somit auch folgendermaßen dargestellt werden:

$$B_n = \big\{\omega: \quad (\omega_{t_1}, \ldots, \omega_{t_{d_n}})^T \in D_n\big\}, \tag{36}$$

wobei $D_n \in \mathcal{B}^{d_n}$ entsprechend gewählt ist. Da $\pi_{t_1,\ldots,t_{d_n}}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^{d_n}, \mathcal{B}^{d_n})$ ist, so existieren **kompakte** Mengen $E_n \in \mathbb{R}^{d_n}$ mit $E_n \subseteq D_n$ und

$$\pi_{t_1,\ldots,t_{d_n}}\big(D_n \setminus E_n\big) \le \varepsilon\, 2^{-n};$$

siehe Satz 1.12 aus der Vorlesung „Maßtheorie" im WS 2019/20 für ein ähnliches Resultat.

[1] *In ähnlich gelagerten Fällen wäre dies vergleichsweise einfach. Wenn die Mengen B_n* **kompakte** *Teilmengen des $\mathbb{R}^d$ wären, so würde aus (35) und der Nulltreue jedes Wahrscheinlichkeitsmaßes folgen, dass alle B_n nichtleer sind. Da wir es hier mit einer nichtwachsenden Folge kompakter Mengen zu tun hätten, so würde jedoch $\bigcap_{n=1}^{\infty} B_n \neq \emptyset$ und somit der gesuchte Widerspruch folgen. In der vorliegenden Situation haben wir es mit Mengen zu tun, welche weder endlichdimensional noch kompakt sind. Nichtsdestotrotz wird jedoch die eben geschilderte Idee in angepasster Weise umgesetzt.*

Daraus folgt, dass

$$P_0\Big(\big\{\omega\colon \quad (\omega_{t_1},\ldots,\omega_{t_{d_n}})^T \in D_n \setminus E_n\big\}\Big) \;=\; \pi_{t_1,\ldots,t_{d_n}}\big(D_n \setminus E_n\big) \;\le\; \varepsilon\, 2^{-n}. \qquad (37)$$

Es seien nun $F_k = \{\omega\colon\; (\omega_{t_1},\ldots,\omega_{t_{d_k}})^T \in E_k\}$ und

$$G_n \;:=\; \bigcap_{k=1}^{n} F_k.$$

Dann gelten $G_n \in \mathcal{Z}$, $G_n \subseteq B_n$ und aus (37) folgt $P_0(G_n) > 0$. (Dies folgt aus $P_0(B_n) \ge \varepsilon$ und $P_0(B_n \setminus G_n) \le \sum_{k=1}^{n} P_0(B_n \setminus F_k) \le \varepsilon(2^{-1} + \cdots + 2^{-n})$.) Somit gilt aber $G_n \ne \emptyset$ für alle $n \in \mathbb{N}$. Daraus folgt jedoch, dass

$$\bigcap_{n=1}^{\infty} G_n \ne \emptyset. \qquad (38)$$

Wegen $G_n \subseteq B_n$ folgt nun

$$\bigcap_{n=1}^{\infty} B_n \ne \emptyset,$$

was einen Widerspruch zu (34) darstellt. Somit war die Annahme (35) falsch und die σ-Additivität von P_0 ist bewiesen. $\square$

Nachtrag: Beweis von (38)
Wir zeigen, dass ein $\omega \in \mathcal{Z}$ existiert mit $\omega \in \bigcap_{n=1}^{\infty} G_n$, d.h., es muss ein ω gefunden werden mit $\omega \in G_n$ $\forall n \in \mathbb{N}$. Die Menge G_n hat folgende Gestalt:

$$G_n \;=\; \big\{\omega\colon \quad (\omega_{t_1},\ldots,\omega_{t_{d_n}})^T \in H_n\big\},$$

wobei $H_n = \cap_{k=1}^{n-1}\{(\omega_{t_1},\ldots,\omega_{t_{d_n}})\colon\; (\omega_{t_1},\ldots,\omega_{t_{d_k}})^T \in E_k\} \cap E_n$ als Durchschnitt abgeschlossener Mengen und der kompakten Menge D_n eine kompakte Teilmenge des $\mathbb{R}^{d_n}$ ist. Wir wählen $\omega^{(n)} \in G_n$ beliebig. Wegen $d_{n+1} > d_n$ $\forall n \in \mathbb{N}$ folgt, dass $d_k \ge k$ und somit liegen $\omega_{t_k}^{(k)}, \omega_{t_k}^{(k+1)}, \ldots$ in einer kompakten Menge. Nun folgt ein typisches Teilfolgenargument:
Im Hinblick auf die ersten Komponenten $\omega_{t_1}^{(n)}$ findet man eine Teilfolge $(n_k^{(1)})_{k\in\mathbb{N}}$ von $\mathbb{N}$ mit

$$\omega_{t_1}^{(n_k^{(1)})} \underset{k\to\infty}{\longrightarrow} \omega_{t_1}^{\infty}.$$

Im nächsten Schritt wählt man eine Teilfolge $(n_k^{(2)})_{k\in\mathbb{N}}$ von $(n_k^{(1)})_{k\in\mathbb{N}}$ mit

$$\omega_{t_2}^{(n_k^{(2)})} \underset{k\to\infty}{\longrightarrow} \omega_{t_2}^{\infty}.$$

Dies wird sukzessive für alle $k \in \mathbb{N}$ fortgesetzt. Wir betrachten die „Diagonalfolge“ $(n_k)_{k\in\mathbb{N}}$ mit $n_k := n_k^{(k)}$. Dann gilt

$$\omega_{t_i}^{(n_k)} \underset{k\to\infty}{\longrightarrow} \omega_{t_i}^{\infty} \quad \forall i \in \mathbb{N}.$$

Es sei $\omega \in \mathcal{Z}$ so, dass $\omega_{t_i} = \omega_{t_i}^{\infty}$ für alle $i \in \mathbb{N}$ und ω_t beliebig für $t \notin \{t_1, t_2, \ldots\}$ ist. Nun gilt wegen Kompaktheit der Mengen H_n, dass $(\omega_{t_1},\ldots,\omega_{t_{d_n}})^T \in H_n$. Das bedeutet aber, dass $\omega \in G_n$. Da dies für alle $n \in \mathbb{N}$ gilt, so folgt, dass

$$\omega \;\in\; \bigcap_{n=1}^{\infty} G_n \;\subseteq\; \bigcap_{n=1}^{\infty} B_n.$$

Somit ist (38) bewiesen.

3.3 Der Wiener Prozess

In diesem Abschnitt werden wir den Wiener Prozess näher untersuchen. Zur Erinnerung noch einmal die Definition, welche bereits am Ende des Abschnittes 3.1 gegeben wurde.

Definition 3.6. *Ein stochastischer Prozess* $W = (W_t)_{t\in[0,\infty)}$ *heißt* **Wiener Prozess** (**Standard-Brownsche Bewegung**) *auf* $(\Omega, \mathcal{A}, P)$*, falls*

(i) $W_0 = 0$ *P-fast sicher,*

(ii) *W besitzt unabhängige Zuwächse, d.h., für beliebige* $0 < t_1 < \ldots < t_k$*,* $k \in \mathbb{N}$ *sind* $W_{t_1} - W_0$*,* $W_{t_2} - W_{t_1}$*, …,* $W_{t_k} - W_{t_{k-1}}$ *stochastisch unabhängig,*

(iii) $W_t - W_s \sim N(0, t-s)$ *für alle* $0 \le s < t$*,*

(iv) *W besitzt P-fast sicher stetige Pfade, d.h.,* $t \mapsto W_t(\omega)$ *ist stetig für P-fast alle* ω*.*

Wir untersuchen zunächst die Frage der Existenz (d.h., der widerspruchsfreien Konstruierbarkeit) eines Prozesses, der Punkte (i) bis (iv) der obigen „Wunschliste“ erfüllt. Wir nehmen an, dass ein solcher Prozess tatsächlich existiert. (i), (ii) und (iii) erlauben die Identifizierung der endlichdimensionalen Verteilungen des noch zu konstruierenden Prozesses (siehe auch ÜA 3.1). Für beliebige $t_1, \ldots, t_k \in [0, \infty)$, $k \in \mathbb{N}$, gilt

$$P^{W_{t_1},\ldots,W_{t_k}} = N_k\left(0_k, \begin{pmatrix} t_1 \wedge t_1 & \ldots & t_1 \wedge t_k \\ \vdots & \ddots & \vdots \\ t_k \wedge t_1 & \ldots & t_k \wedge t_k \end{pmatrix}\right),$$

wobei 0_k den k-dimensionalen Nullvektor bezeichnet. Es sei also die Verteilungsfamilie

$$\left\{\pi_{t_1,\ldots,t_k} : \ t_1, \ldots, t_k \in [0, \infty), \ t_i \neq t_j \text{ für } i \neq j, \ k \in \mathbb{N}\right\}$$

gegeben mit

$$\pi_{t_1,\ldots,t_k} = N_k\left(0_k, \begin{pmatrix} t_1 \wedge t_1 & \ldots & t_1 \wedge t_k \\ \vdots & \ddots & \vdots \\ t_k \wedge t_1 & \ldots & t_k \wedge t_k \end{pmatrix}\right).$$

Nach Proposition 2.5 erfüllt diese Verteilungsfamilie die Konsistenzbedingungen (24) und (25) und der Existenzsatz von Kolmogoroff (Theorem 3.4) besagt, dass auf einem geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ ein Prozess $(W_t)_{t\in[0,\infty)}$ mit

$$P^{W_{t_1},\ldots,W_{t_k}} = \pi_{t_1,\ldots,t_k} \qquad t_1, \ldots, t_k \in [0, \infty), \ t_i \neq t_j \text{ für } i \neq j, \ k \in \mathbb{N}$$

konstruiert werden kann. Dieser Prozess erfüllt dann auch die eingangs geforderten Bedingungen (i) bis (iii). Im Buch “*Theory and Statistical Applications of Stochastic Prozesses*” von Y. Mishura und G. Shevchenko (Wiley, 2017) wird der Wiener Prozess tatsächlich so definiert, dass er nur die obigen Bedingungen (i) bis (iii) erfüllt. Es ist jedoch weit mehr verbreitet, Bedingung (iv) (d.h., die fast sichere Stetigkeit der Pfade) in die Definition eines Wiener Prozesse mit einzubeziehen, siehe z.B. die Bücher “*Probability: Theory and Examples*” von R. Durrett (Duxbury Press, 1995) oder “*Stochastic Processes*” von R.F. Bass (Cambridge University Press, 2011).

Obwohl der Satz von Kolmogoroff völlig ausreicht für die Konstruktion von stochastischen Prozessen mit einem höchstens abzählbaren Indexbereich T, so ist die erwünschte pfadweise Stetigkeitseigenschaft eines durch diesen Satz gewonnenen Prozesses nicht klar. Um das Problem zu sehen, stelle man sich einen Wiener Prozess $(W_t)_{t\in[0,\infty)}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ vor, welcher P-fast sicher stetige Pfade besitzt. Wir nehmen an, dass auf $(\Omega, \mathcal{A}, P)$ eine auf dem Intervall $[0,1]$ gleichverteilte Zufallsvariable definiert werden kann und wir nehmen folgende Modifikation von $(W_t)_{t\in[0,\infty)}$ vor:

$$\widetilde{W_t}(\omega) \;=\; \begin{cases} W_t(\omega) + 1, & \text{falls } U(\omega) = t, \\ W_t(\omega), & \text{falls } U(\omega) \neq t, \end{cases}$$

Der Prozess $(\widetilde{W_t})_{t\in[0,\infty)}$ besitzt Pfade mit einer einzigen Unstetigkeit im zufälligen Punkt $U(\omega)$. Da jedoch $U(\omega)$ eine stetige Verteilung besitzt, sind die endlichdimensionalen Verteilungen des Prozesses $(\widetilde{W_t})_{t\in[0,\infty)}$ dieselben wie die von $(W_t)_{t\in[0,\infty)}$. Der Existenzsatz von Kolmogoroff vermittelt nun lediglich eine Aussage über die Existenz eines Prozesses mit vorgegebenen endlichdimensionalen Verteilungen. Daher ist im Falle des Wiener Prozesses noch nicht klar, ob tatsächlich eine Version mit fast sicher stetigen Pfaden existiert. Ein weiteres Hinweis auf diese Problematik ist durch die folgende Aussage gegeben.

Für **Folgen** reeller Zahlen definieren wir die Menge k-dimensionale Zylindermengen durch

$$\mathcal{F}_k \;:=\; \Big\{ \big\{ x = (x_n)_{n\in\mathbb{N}} \colon \quad (x_1, \ldots, x_k)^T \in B \big\} \Big| B \in \mathcal{B}^k \Big\}$$

und die dadurch erzeugte σ-Algebra durch

$$\mathcal{F} \;:=\; \sigma\Big(\bigcup_{k=1}^{\infty} \mathcal{F}_k \Big).$$

Für **Funktionen** $f \colon [0,\infty) \to \mathbb{R}$ definieren wir wie bisher

$$\mathcal{Z}_k \;:=\; \Big\{ \big\{ \omega = (\omega_t)_{t\in[0,\infty)} \colon \quad (\omega_{t_1}, \ldots, \omega_{t_k})^T \in B \big\} \Big| t_1, \ldots, t_k \in [0,\infty),\; t_i \neq t_j \text{ für } i \neq j,\; B \in \mathcal{B}^k \Big\}$$

und

$$\mathcal{Z} \;:=\; \sigma\Big(\bigcup_{k=1}^{\infty} \mathcal{Z}_k \Big).$$

Lemma 3.7. *Falls $A \in \mathcal{Z}$, so existiert eine Folge $(t_k)_{k\in\mathbb{N}}$ und eine Menge $B \in \mathcal{F}$, sodass*

$$A \;=\; \big\{ \omega = (\omega_t)_{t\in[0,\infty)} \colon \quad (\omega_{t_1}, \omega_{t_2}, \ldots) \in B \big\}.$$

Lemma 3.7 besagt, dass alle Ereignisse in $\mathcal{Z}$ nur von den Zuständen an abzählbar vielen zeitpunkten abhängen. Insbesondere gilt

$$\{\omega \colon \quad t \mapsto \omega_t \text{ ist eine stetige Funktion auf } [0,\infty)\} \;\notin\; \mathcal{Z}.$$

Die folgende Aussage liefert nun zunächst das Zwischenergebnis, dass ein beliebiger Prozess, welche die Bedingungen (i) bis (iii) aus der Definition eines Wiener Prozesses erfüllt, die gewünschte Stetigkeitseigenschaft auf einer geeigneten **abzählbaren** Teilmenge von $[0, \infty)$ besitzt.

Lemma 3.8. *Auf einem beliebigen Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *sei* $(W_t)_{t\in[0,\infty)}$ *ein stochastischer Prozess mit*

$$P^{W_{t_1},\dots,W_{t_k}} = N_k\left(0_k, \begin{pmatrix} t_1 \wedge t_1 & \dots & t_1 \wedge t_k \\ \vdots & \ddots & \vdots \\ t_k \wedge t_1 & \dots & t_k \wedge t_k \end{pmatrix}\right) \qquad \forall t_1, \dots, t_k \geq 0,\ \forall k \in \mathbb{N}.$$

Für $M \in \mathbb{N}$ *sei*

$$D_M = \Big\{t = x_0 + \sum_{j=1}^{k} x_j 2^{-j}: \quad x_0 \in \{0, 1, \dots, M-1\},\ x_j \in \{0,1\},\ k \in \mathbb{N}\Big\}$$

die Menge der dyadischen Zahlen aus $[0, M)$.
Dann ist mit Wahrscheinlichkeit 1 die Funktion $t \mapsto W_t$ *gleichmäßig stetig auf* D_M.

Beweis. Wir wählen ein $\delta \in (0, 1/2)$ und definieren die folgenden „guten“ Mengen:

$$A_m^{(M)} := \bigcap_{j=1}^{M2^m} \Big\{\omega: \quad \big|W_{j2^{-m}}(\omega) - W_{(j-1)2^{-m}}(\omega)\big| \leq 2^{-m(1/2-\delta)}\Big\}.$$

Für $Y_0 \sim N(0,1)$ erhalten wir wegen $e^{-(x+t)^2/2} \leq e^{-x^2/2} e^{-t^2/2}\ \forall x, t \geq 0$ die folgende Abschätzung für $t \geq 0$:

$$\begin{aligned} P\big(|Y_0| > t\big) &= 2 \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 2 \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(x+t)^2/2}\, dx \\ &\leq 2\, e^{-t^2/2} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = e^{-t^2/2}. \end{aligned}$$

Da für beliebiges $K > 0$ ein $L_K < \infty$ existiert mit $e^{-t^2/2} \leq L_K t^{-K}\ \forall t \geq 0$, so folgt mit $Y_m \sim N(0, 2^{-m})$

$$\begin{aligned} P\Big(\big(A_m^{(M)}\big)^c\Big) &\leq M 2^m\, P\big(|Y_m| > 2^{-m(1/2-\delta)}\big) \\ &= M2^m\, P\big(|Y_0| > 2^{m\delta}\big) \\ &\leq M2^m e^{-(2^{m\delta})^2/2} \leq M2^m\, L_{2/\delta} 2^{m\delta(-2/\delta)} = L_{2/\delta}\, 2^{-m}. \end{aligned}$$

Daher gilt $\sum_{m=1}^\infty P\big((A_m^{(M)})^c\big) < \infty$, woraus mit dem Lemma von Borel-Cantelli folgt, dass

$$P\Big((A_m^{(M)})^c \text{ für unendlich viele } m\Big) = P\Big(\bigcap_{k=1}^\infty \bigcup_{m=k}^\infty \big(A_m^{(M)}\big)^c\Big) = 0. \tag{39}$$

Es sei nun

$$\omega \in \Big(\bigcap_{k=1}^{\infty} \bigcup_{m=k}^{\infty} \big(A_m^{(M)}\big)^c\Big) = \bigcup_{k=1}^{\infty} \bigcap_{m=k}^{\infty} A_m^{(M)}$$

Dann existiert ein $k_0 = k_0(\omega)$ mit $\omega \in \bigcap_{m=k_0}^{\infty} A_m^{(M)}$. Es seien nun $s, t \in D_M$ mit $0 < |s-t| \leq 2^{-m}$, wobei $m \geq k_0(\omega)$. Dann gelten

$$\begin{aligned} s &= \tau 2^{-m} \quad + \sum_{l=m+1}^{L} \sigma_l 2^{-l}, \\ t &= (\tau + \tau_m) 2^{-m} + \sum_{l=m+1}^{L} \tau_l 2^{-l} \end{aligned}$$

für gewisse $\tau \in \{0, \ldots, M2^m - 1\}$, $\sigma_l, \tau_l \in \{0, 1\}$ und $L < \infty$. Nun gelten

$$\begin{aligned} &\big|W_s(\omega) - W_{\tau 2^{-m}}(\omega)\big| \\ &= \big|W_{\tau 2^{-m} + \sigma_{m+1} 2^{-m+1}}(\omega) - W_{\tau 2^{-m}}(\omega)\big| + \cdots + \big|W_s(\omega) - W_{\tau 2^{-m} + \sum_{l=m+1}^{L-1} \sigma_l 2^{-l}}(\omega)\big| \\ &\leq \sum_{l=m+1}^{L} \sigma_l\, 2^{-l(1/2-\delta)} \\ &\leq 2^{-(m+1)(1/2-\delta)} \underbrace{\sum_{l=0}^{\infty} 2^{-l(1/2-\delta)}}_{=: C_\delta < \infty} \end{aligned}$$

sowie

$$\begin{aligned} \big|W_t(\omega) - W_{\tau 2^{-m}}(\omega)\big| &\leq \sum_{l=m}^{L} \tau_l\, 2^{-l(1/2-\delta)} \\ &\leq 2^{-m(1/2-\delta)}\, C_\delta. \end{aligned}$$

Somit ist die Funktion $t \mapsto W_t(\omega)$ gleichmäßig stetig auf D_M. □

Der folgende Satz liefert nun die Existenz eines Wiener Prozesses mit stetigen Pfaden. Dabei wird sogar über Bedingung (iv) der Definition hinaus gegangen, welche nur die P-fast sichere Stetigkeit verlangte. Die Konstruktion des gesuchten Prozesses beruht wesentlich auf dem vorangegangen Lemma 3.8, welches die fast sichere Stetigkeit der Pfade auf den **abzählbaren** Mengen D_M liefert.

Theorem 3.9. $(\Omega, \mathcal{A}, P)$ *sei ein beliebiger Wahrscheinlichkeitsraum, auf dem ein stochastischer Prozess* $(W_t)_{t\in[0,\infty)}$ *mit endlichdimensionalen Verteilungen*

$$P^{W_{t_1},\dots,W_{t_k}} = N_k\left(0_k, \begin{pmatrix} t_1\wedge t_1 & \dots & t_1\wedge t_k \\ \vdots & \ddots & \vdots \\ t_k\wedge t_1 & \dots & t_k\wedge t_k \end{pmatrix}\right) \qquad \forall t_1,\dots,t_k \geq 0,\ \forall k\in\mathbb{N}.$$

definiert ist.
Dann lässt sich auf $(\Omega, \mathcal{A}, P)$ *ein Prozess* $(\widetilde{W}_t)_{t\in[0,\infty)}$ *konstruieren, welcher für alle* $\omega\in\Omega$ *stetige Pfade besitzt.*

Beweis. $(W_t)_{t\in[0,\infty)}$ sei ein Prozess auf $(\Omega, \mathcal{A}, P)$ mit den endlichdimensionalen Verteilungen

$$P^{W_{t_1},\dots,W_{t_k}} = N_k\left(0_k, \begin{pmatrix} t_1\wedge t_1 & \dots & t_1\wedge t_k \\ \vdots & \ddots & \vdots \\ t_k\wedge t_1 & \dots & t_k\wedge t_k \end{pmatrix}\right) \qquad \forall t_1,\dots,t_k \geq 0,\ \forall k\in\mathbb{N}. \tag{40}$$

Für $M\in\mathbb{N}$ sei $A^{(M)} := \bigcup_{k=1}^{\infty}\bigcap_{m=k}^{\infty} A_m^{(M)}$, wobei $A_m^{(M)}$ die im Beweis von Lemma 3.8 verwendeten „guten Mengen“ sind. Da $P(A^{(M)}) = 1\ \forall M\in\mathbb{N}$, so folgt für $A := \bigcap_{M=1}^{\infty} A^{(M)}$, dass

$$P\big(A\big) = 1.$$

Weiter sei $D = \bigcup_{M=1}^{\infty} D_M$ die Menge aller dyadischen Zahlen in $[0,\infty)$. Wir definieren

$$\widetilde{W}_t(\omega) := \begin{cases} \lim_{t_n\to t,\,(t_n)_n\subseteq D} W_{t_n}(\omega), & \text{falls } \omega\in A, \\ 0, & \text{falls } \omega\notin A. \end{cases}$$

Wegen **gleichmäßiger** Stetigkeit von $t\mapsto W_t(\omega)$ auf den Mengen D_M für $\omega\in A$ ist nun auch $t\mapsto \widetilde{W}_t(\omega)$ stetig.

Es ist noch zu zeigen, dass die endlichdimensionalen Verteilungen von $(\widetilde{W}_t)_{t\in[0,\infty)}$ denen eines Wiener Prozesses entsprechen. Wir zeigen, dass für beliebiges **festes** $t\in[0,\infty)$ gilt, dass

$$P\Big(\big\{\omega:\quad \widetilde{W}_t(\omega)\neq W_t(\omega)\big\}\Big) = 0. \tag{41}$$

Aus (40) und (41) folgt dann, dass

$$P^{\widetilde{W}_{t_1},\dots,\widetilde{W}_{t_k}} = P^{W_{t_1},\dots,W_{t_k}} \qquad \forall t_1,\dots,t_k\geq 0,\ \forall k\in\mathbb{N},$$

womit der Satz bewiesen wäre.

Es fehlt noch die Begründung für (41). Da für $\omega \in A$ der Pfad $t \mapsto W_t(\omega)$ auf D stetig ist, folgt nach der Definition von $\widetilde{W}_t$, dass

$$P\big(\widetilde{W}_t \;=\; W_t \quad \forall t \in D\big) \;=\; 1.$$

Für $t \notin D$ können wir folgendermaßen argumentieren. Es sei $(t_n)_{n\in\mathbb{N}} \subseteq D$ mit $t_n \underset{n\to\infty}{\longrightarrow} t$. Wegen

$$P\big(|W_{t_n} \;-\; W_t| \;\geq\; \epsilon\big) \;\leq\; \epsilon^{-2}\, E\big[(W_{t_n} \;-\; W_t)^2\big] \;=\; \epsilon^{-2}\, |t_n - t|$$

folgt

$$W_{t_n} \;\stackrel{P}{\longrightarrow}\; W_t.$$

Andererseits gilt aber nach der Definition von $\widetilde{W}_t$, dass

$$W_{t_n} \;\stackrel{P-f.s.}{\longrightarrow}\; \widetilde{W}_t,$$

was bekanntermaßen

$$W_{t_n} \;\stackrel{P}{\longrightarrow}\; \widetilde{W}_t$$

impliziert. Wegen der fast sicheren Eindeutigkeit des Grenzwertes bei P-stochastischer Konvergenz folgt (41). $\square$

Übungsaufgabe

ÜA 3.3 Beweisen sie Lemma 3.7!

Hinweis: Zeigen Sie, dass

$$\mathcal{Z} \;\subseteq\; \Big\{\big\{\omega = (\omega_t)_{t\in[0,\infty)}: \quad (\omega_{t_1}, \omega_{t_2}, \ldots) \in B\big\}\Big|\, t_1, t_2, \ldots \in [0,\infty),\; t_i \neq t_j \text{ für } i \neq j,\; B \in \mathcal{F}\Big\}$$

gilt!

Bemerkung 3.10. Der Wiener Prozess besitzt eine Reihe von Invarianzeigenschaften. Falls $(W_t)_{t\in[0,\infty)}$ ein Wiener Prozess ist, so ist leicht zu sehen, dass auch $(W_t^{(i)})_{t\in[0,\infty)}$ mit

- $W_t^{(1)} := -W_t$,
- $W_t^{(2)} := W_{t+t_0} - W_{t_0} \quad$ für $t_0 \geq 0$,
- $W_t^{(3)} := \sqrt{c}\, W_{t/c} \quad$ für ein $c > 0$

jeweils Wiener Prozesse sind. (Die fast sichere Stetigkeit der Pfade dieser Prozesse ist offensichtlich. Ein direktes Überprüfen der endlichdimensionalen Verteilungen ergibt dann die Behauptung.)

Zum Ende dieses Kapitels werfen wir noch einen kurzen Blick auf einen vom Wiener Prozess abgeleiteten Prozess, auf die sogenannte **Brownsche Brücke**. Zur Erinnerung sei nochmals erwähnt, dass eine Prozess $(W_t)_{t\in[0,\infty)}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ die Eigenschaften eines Wiener Prozesses (einer Standard-Brownschen Bewegung) besitzt, falls

(i) für beliebige $t_1, \ldots, t_k \geq 0$, $k \in \mathbb{N}$

$$P^{W_{t_1},\ldots,W_{t_k}} = N_k\left(0_k, \begin{pmatrix} t_1 \wedge t_1 & \ldots & t_1 \wedge t_k \\ \vdots & \ddots & \vdots \\ t_k \wedge t_1 & \ldots & t_k \wedge t_k \end{pmatrix}\right),$$

gilt und

(ii) die Pfade $t \mapsto W_t(\omega)$ für P-fast alle ω stetig sind.

Ausgehend von einem Wiener Prozess $(W_t)_{t\in[0,\infty)}$ auf $(\Omega, \mathcal{A}, P)$ können wir nun einen weiteren Prozess $(B_t)_{t\in[0,1]}$ herleiten durch

$$B_t = W_t - t\,W_1 \qquad \forall t \in [0,1].$$

Da es sich hierbei um lineare Transformationen eines **Gaußprozesses** (d.h., eines Prozesses, dessen endlichdimensionale Verteilungen Normalverteilungen sind) handelt, so folgt aus Proposition 2.5(ii), dass $(B_t)_{t\in[0,1]}$ ebenfalls ein Gaußprozess ist. Es gelten $EB_t = 0$ $\forall t \in [0,1]$ und

$$\begin{aligned} \operatorname{cov}(B_s, B_t) &= E\big[(W_s - sW_1)(W_t - tW_1)\big] \\ &= \underbrace{E[W_sW_t]}_{=s\wedge t} - \underbrace{s\,E[W_1W_t]}_{=st} - \underbrace{t\,E[W_sW_1]}_{=ts} + \underbrace{st\,E[W_1^2]}_{=st} \\ &= s\wedge t - st \qquad \forall s,t \in [0,1]. \end{aligned}$$

Darüber hinaus erbt $(B_t)_{t\in[0,1]}$ die Stetigkeitseigenschaft der Pfade von $(W_t)_{t\in[0,\infty)}$. Hier ist eine formale Definition einer Brownschen Brücke:

Definition 3.11. *Ein stochastischer Prozess* $(B_t)_{t\in[0,1]}$ *auf auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *heißt* **Brownsche Brücke**, *falls*

(i) *für beliebige* $t_1, \ldots, t_k \in [0,1]$, $k \in \mathbb{N}$

$$P^{B_{t_1},\ldots,B_{t_k}} = N_k\left(0_k, \begin{pmatrix} t_1 \wedge t_1 - t_1t_1 & \ldots & t_1 \wedge t_k - t_1t_k \\ \vdots & \ddots & \vdots \\ t_k \wedge t_1 - t_kt_1 & \ldots & t_k \wedge t_k - t_kt_k \end{pmatrix}\right),$$

gilt und

(ii) *die Pfade* $t \mapsto B_t(\omega)$ *für* P*-fast alle* ω *stetig sind.*

Bei Übungsaufgabe 1.8 war bereits zu sehen, dass eine Struktur wie bei der Brownschen Brücke beim asymptotischen Verhalten der empirischen Verteilungsfunktion auftritt. Um dies nochmals aufzugreifen setzen wir voraus, dass $(X_n)_{n\in\mathbb{N}}$ eine Folge von reellwertigen und stochastisch unabhängigen Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ mit einer jeweiligen Verteilungsfunktion F sei. $F_n\colon \mathbb{R} \to [0,1]$ mit $F_n(x) = n^{-1}\sum_{i=1}^n \mathbb{1}(X_i \le x)$ ist die zur Stichprobe $X_1, \ldots, X_n$ gehörende **empirische Verteilungsfunktion** und $(G_n(x))_{x\in\mathbb{R}}$ mit $G_n(x) = \sqrt{n}\big(F_n(x) - F(x)\big)$ der entprechende **empirische Prozess**. Aus dem mehrdimensionalen Zentralen Grenzwertsatz (Theorem 2.6; siehe auch ÜA 1.8) folgt, dass die endlichdimensionalen Verteilungen des empririschen Prozesses gegen mehrdimensionale Normalverteilungen konvergieren, d.h., für $x_1, \ldots, x_k \in \mathbb{R}$, $k \in \mathbb{N}$ gilt

$$\big(G_n(x_1), \ldots, G_n(x_k)\big)^T \xrightarrow{d} Z \sim N(0_k, \Sigma),$$

wobei

$$\begin{aligned} \Sigma_{k,l} &= \operatorname{cov}\big(\mathbb{1}(X_1 \le x_k), \mathbb{1}(X_1 \le x_l)\big) \\ &= E\big[\mathbb{1}(X_1 \le x_k)\mathbb{1}(X_1 \le x_l)\big] - E[\mathbb{1}(X_1 \le x_k)]E[\mathbb{1}(X_1 \le x_l)] \\ &= F(x_k) \wedge F(x_l) - F(x_k)F(x_l). \end{aligned}$$

Mit anderen Worten: Die endlichdimensionalen Verteilungen des empirischen Prozesses konvergieren gegen jene des Prozesses $(B_{F(x)})_{x\in\mathbb{R}}$, falls $(B_t)_{t\in[0,1]}$ eine Brownsche Brücke ist. Im Spezialfall $X_i \sim \text{Uniform}([0,1])$ gelten $F(x) = F_n(x) = 0$ für $x < 0$ sowie $F(x) = F_n(x) = 1$ für $x > 1$. Daher genügt es, den empirischen Prozess mit Indexmenge $[0,1]$ zu betrachten und es folgt, dass die endlichdimensionalen Verteilungen von $(G_n(x))_{x\in[0,1]}$ gegen jene der Brownschen Brücke konvergieren. Aus diesem Grund spielt die Brownsche Brücke auch in der Mathematischen Statistik eine wichtige Rolle, nämlich wenn der empirische Prozess als Grundlage für einen Anpassungstest benutzt und ein kritischer Wert entsprechend der Grenzverteilung der Teststatistik gewählt wird.

Übungsaufgabe

ÜA 3.4 $(B_t)_{t\in[0,1]}$ sei eine Brownsche Brücke.
Zeigen Sie, dass $(\widetilde{B}_t)_{t\in[0,1]}$ mit $\widetilde{B}_t = B_{1-t}$ $\forall t \in [0,1]$ ebenfalls eine Brownsche Brücke ist!

3.4 Der Poissonprozess

In diesem Abschnitt werden wir den sogenannten Poissonprozess einführen und die Hintergründe für die spezielle Wahl der entsprechenden Verteilungseigenschaften diskutieren. Wir fallen gleich mit der Tür ins Haus und beginnen direkt mit der Definition.

Definition 3.12. *Ein stochastischer Prozess* $(N_t)_{t\in[0,\infty)}$ *auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *heißt* **Poissonprozess mit Intensität** λ*, falls*

(i) $N_0 = 0$ *P-fast sicher,* $(N_t)_{t\in[0,\infty)}$ *besitzt unabhängige Zuwächse, d.h., für* $0 = t_0 < t_1 < \ldots < t_k$, $k \in \mathbb{N}$*, sind* $N_{t_1} - N_{t_0}, \ldots, N_{t_k} - N_{t_{k-1}}$ *stochastisch unabhängig und es gilt*

$$N_t - N_s \sim \textit{Poisson}(\lambda(t-s)) \qquad \textit{für } 0 \le s < t,$$

(ii) $(N_t)_{t\in[0,\infty)}$ *besitzt rechtsstetige Pfade.*

Wie schon beim Wiener Prozess ist nach der bloßen Definition nicht klar, ob ein solcher Prozess existiert, d.h., widerspruchsfrei konstruierbar ist. Mit Hilfe des Existenzsatzes von Kolmogoroff (Theorem 3.4) können wir einen ersten Versuch starten. Dazu müssen wir uns zunächst ein passende konsistente Familie von Verteilungen zurechtlegen. Es sei zunächst $0 = t_0 \le t_1 < \ldots < t_k < \infty$, $k \in \mathbb{N}$. Die Verteilung $\pi_{t_1,\ldots,t_k}$ auf $(\mathbb{N}_0^k, 2^{\mathbb{N}_0^k})$ soll so gewählt werden, dass sie der endlichdimensionalen Verteilung $P^{N_{t_1},\ldots,N_{t_k}}$ des zu konstruierenden Prozesses entspricht, also

$$\pi_{t_1,\ldots,t_k}\big(\{(r_1,\ldots,r_k)^T\}\big) \;=\; p_{\lambda(t_1-t_0)}(r_1-r_0)\,\cdots\,p_{\lambda(t_k-t_{k-1})}(r_k-r_{k-1}) \quad (0 = r_0 \le r_1 \le r_2 \le \ldots \le r_k),$$

wobei $p_a(r) = e^{-a}a^r/r!$ $(r \in \mathbb{N}_0)$ die Poisson-Wahrscheinlichkeiten sind. Nun gilt für $i \in \{1,\ldots,k\}$

$$\begin{aligned}
&\pi_{t_1,\ldots,t_k}\big(\{(r_1,\ldots,r_{i-1})^T\} \times \mathbb{N}_0 \times \{(r_{i+1},\ldots,r_k)^T\}\big)\\
&= p_{\lambda(t_1-t_0)}(r_1-r_0)\,\cdots\,p_{\lambda(t_{i-1}-t_{i-2})}(r_{i-1}-r_{i-2})\underbrace{\sum_r p_{\lambda(t_i-t_{i-1})}(r-r_{i-1})p_{\lambda(t_{i+1}-t_i)}(r_{i+1}-r)}_{=p_{\lambda(t_{i+1}-t_{i-1})}(r_{i+1}-r_{i-1})}\\
&\qquad \times p_{\lambda(t_{i+2}-t_{i+1})}(r_{i+2}-r_{i+1})\cdots p_{\lambda(t_k-t_{k-1})}(r_k-r_{k-1})\\
&= p_{\lambda(t_1-t_0)}(r_1-r_0)\,\cdots\,p_{\lambda(t_{i-1}-t_{i-2})}(r_{i-1}-r_{i-2})p_{\lambda(t_{i+1}-t_{i-1})}(r_{i+1}-r_{i-1})\,\cdots\,p_{\lambda(t_k-t_{k-1})}(r_k-r_{k-1})\\
&= \pi_{t_1,\ldots,t_{i-1},t_{i+1},\ldots,t_k}\big(\{(r_1,\ldots,r_{i-1},r_{i+1},\ldots,r_k)^T\}\big). \qquad (42)
\end{aligned}$$

Für möglicherweise ungeordnete $t_1,\ldots,t_k$ $(t_i \ne t_j$ für $i \ne j)$ existiert eine Permutation $\rho\colon\ \{1,\ldots,k\} \to \{1,\ldots,k\}$ mit $t_{\rho(1)} < \ldots < t_{\rho(k)}$. Die Verteilung $\pi_{t_1,\ldots,t_k}$ auf $(\mathbb{N}_0^k, 2^{\mathbb{N}_0^k})$ definiert man nun (logischerweise) so, dass

$$\pi_{t_1,\ldots,t_k}\big(\{(r_1,\ldots,r_k)^T\}\big) \;=\; \pi_{t_{\rho(1)},\ldots,t_{\rho(k)}}\big(\{(r_{\rho(1)},\ldots,r_{\rho(k)})^T\}\big). \qquad (43)$$

Wenn wir nun noch die Eigenschaft der σ-Additivität nutzen, so erhalten wir aus (42) und (43) für beliebige $t_1,\ldots,t_k \in [0,\infty)$ $(t_i \ne t_j$ für $i \ne j)$, $k \ge 2$

$$\pi_{t_1,\ldots,t_k}\big(B \times \mathbb{N}_0\big) \;=\; \pi_{t_1,\ldots,t_{k-1}}\big(B\big) \qquad B \subseteq \mathbb{N}_0^{k-1}.$$

Folglich ist $\{\pi_{t_1,\ldots,t_k}\colon\ t_1,\ldots,t_k \in [0,\infty),\ t_i \ne t_j \text{ für } i \ne j,\ k \in \mathbb{N}\}$ eine konsistente Verteilungsfamilie. Nach dem Existenzsatz von Kolmogoroff (Theorem 3.4) kann nun auf einem geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ ein stochastischer Prozess $(N_t)_{t\in[0,\infty)}$ definiert werden mit

$$P^{N_{t_1},\ldots,N_{t_k}} \;=\; \pi_{t_1,\ldots,t_k} \qquad \forall t_1,\ldots,t_k \in [0,\infty), k \in \mathbb{N}.$$

Mit der obigen Wahl der Verteilungen $\pi_{t_1,\dots,t_k}$ erfüllt solch ein Prozess Forderung (i) aus der Definition eines Poissonprozesses. Wie schon im Falle des Wiener Prozesses ist die zusätzlich geforderte Stetigkeitseigenschaft (ii) durch die bloße Festlegung der endlichdimensionalen Verteilungen nicht garantiert. Selbst wenn der Prozess $(N_t)_{t\in[0,\infty)}$ rechtsseitig stetige Pfade besitzt, so kann man diesen bei Beibehaltung seiner endlichdimensionalen Verteilungen leicht abändern, sodass der neuen Prozess die Stetigkeitseigenschaft (ii) nicht mehr aufweist. Wir nehmen an, dass auf $(\Omega, \mathcal{A}, P)$ eine auf dem Intervall $[0,1]$ gleichverteilte Zufallsvariable definiert werden kann und wir nehmen folgende Modifikation von $(N_t)_{t\in[0,\infty)}$ vor:

$$\widetilde{N}_t(\omega) \;=\; \begin{cases} N_t(\omega) + 1/2, & \text{falls } U(\omega) = t, \\ N_t(\omega), & \text{falls } U(\omega) \neq t, \end{cases}$$

Der Prozess $(\widetilde{N}_t)_{t\in[0,\infty)}$ besitzt Pfade mit einer zweiseitigen Unstetigkeit im zufälligen Punkt $U(\omega)$. Da jedoch U eine stetige Verteilung besitzt, sind die endlichdimensionalen Verteilungen des Prozesses $(\widetilde{N}_t)_{t\in[0,\infty)}$ dieselben wie die von $(N_t)_{t\in[0,\infty)}$.

Die im Folgenden dargestellte explizite Konstruktion liefert nun eine Version des Poissonprozesses, der alle Forderungen aus der obigen Definition erfüllt.

Theorem 3.13. *$(X_n)_{n\in\mathbb{N}}$ sei eine Folge von unabhängigen Exp(λ)-verteilten Zufallsvariablen ($\lambda > 0$) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Weiter seien $Y_n = X_1 + \cdots + X_n$ und*

$$N_t \;:=\; \#\big\{n \in \mathbb{N}: \quad Y_n \leq t\big\}.$$

Dann ist $(N_t)_{t\in[0,\infty)}$ ein Poissonprozess mit Intensität λ.

Beweis. Zunächst ist festzuhalten, dass nach dem Starken Gesetz der Großen Zahlen gilt

$$Y_n/n \;\stackrel{P-f.s.}{\longrightarrow}\; EX_1 \;=\; 1/\lambda.$$

Folglich gilt mit Wahrscheinlichkeit eins $N_t < \infty$ für alle $t \in [0,\infty)$. In diesem Fall sind die Pfade des Prozesses (nach Konstruktion) rechtsstetig.

Wir berechnen nun zunächst die Dichte des Zufallsvektors $Y^{(n)} := \big(Y_1, \dots, Y_n\big)^T$. Der Zufallsvektor $X^{(n)} := \big(X_1, \dots, X_n\big)^T$ besitzt bezüglich λ^n eine Dichte $p_X^{(n)}$ mit

$$p_X^{(n)}\big(x_1, \dots, x_n\big) \;=\; \prod_{k=1}^{n} \lambda e^{-\lambda x_k} \mathbb{1}(x_k \geq 0) \;=\; \lambda^n \, e^{-\lambda(x_1+\cdots+x_n)} \, \mathbb{1}(x_1, \dots, x_n \geq 0).$$

Wegen

$$Y^{(n)} \;=\; \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \dots & \dots & 1 \end{pmatrix}}_{=:A} X^{(n)}$$

folgt nach der Dichtetransformationsformel, dass $Y^{(n)}$ eine Dichte $p_Y^{(n)}$ besitzt mit

$$\begin{aligned} p_Y^{(n)}\big(y_1,\dots,y_n\big) &= p_X^{(n)}\big(y_1,y_2-y_1,\dots,y_n-y_{n-1}\big)\,\big|\underbrace{\det(A^{-1})}_{=1}\big|\,\mathbb{1}\big(0\le y_1\le\dots\le y_n\big) \\ &= \lambda^n\,e^{-\lambda y_n}\,\mathbb{1}\big(0\le y_1\le\dots\le y_n\big). \end{aligned}$$

Es seien nun $0<t_1<\dots<t_k$, $r_1\le r_2\le\dots\le r_k$, $k\in\mathbb{N}$ beliebig. Wir werden zeigen, dass (mit $t_0=r_0=0$)

$$\begin{aligned} &P\big(N_{t_1}=r_1,\dots,N_{t_k}=r_k\big) \\ &= e^{-\lambda(t_1-t_0)}\frac{(\lambda(t_1-t_0))^{r_1-r_0}}{(r_1-r_0)!}\cdots e^{-\lambda(t_k-t_{k-1})}\frac{(\lambda(t_k-t_{k-1}))^{r_k-r_{k-1}}}{(r_k-r_{k-1})!} \end{aligned} \tag{44}$$

gilt. Nun ist

$$P^* := P\big(N_{t_1}=r_1,\dots,N_{t_k}=r_k\big) = P\big((Y_1,\dots,Y_{r_k+1})^T\in R\big),$$

wobei

$$R=\big\{(y_1,\dots,y_{r_k+1})\colon\ 0\le y_1\le\dots\le y_{r_1}\le t_1<y_{r_1+1}\le\dots\le y_{r_2}\le t_2<\dots\le y_{r_k}\le t_k<y_{r_k+1}\big\}.$$

Daraus folgt

$$\begin{aligned} P^* &= \int_R \lambda^{r_k+1}\,e^{-\lambda y_{r_k+1}}\,d\lambda^{r_k+1}(y_1,\dots,y_k) \\ &= \bigg(\int_{R_1} d\lambda^{r_1-r_0}(y_{r_0+1},\dots,y_{r_1})\bigg)\cdots\bigg(\int_{R_k} d\lambda^{r_k-r_{k-1}}(y_{r_{k-1}+1},\dots,y_{r_k})\bigg) \\ &\qquad\times\underbrace{\int_{t_k}^{\infty}\lambda^{r_k+1}e^{-\lambda y_{r_k+1}}\,dy_{r_k+1}}_{=\lambda^{r_k}e^{-\lambda t_k}}, \end{aligned}$$

wobei

$$R_j=\big\{\big(y_{r_{k-1}+1},\dots,y_k\big)\colon\quad t_{k-1}<y_{r_{k-1}+1}\le\dots\le y_k\le t_k\big\}.$$

Mit vollständiger Induktion kann gezeigt werden, dass

$$\int_{R_j} d\lambda^{r_j-r_{j-1}}(y_{r_{j-1}+1},\dots,y_{r_j}) = \frac{(t_j-t_{j-1})^{r_j-r_{j-1}}}{(r_j-r_{j-1})!}.$$

Daraus folgt schließlich

$$\begin{aligned} P^* &= \frac{(t_1-t_0)^{r_1-r_0}}{(r_1-r_0)!}\cdots\frac{(t_k-t_{k-1})^{r_k-r_{k-1}}}{(r_k-r_{k-1})!}\,\lambda^{r_k}e^{-\lambda t_k} \\ &= e^{-\lambda(t_1-t_0)}\frac{(\lambda(t_1-t_0))^{r_1-r_0}}{(r_1-r_0)!}\quad\cdots\quad e^{-\lambda(t_k-t_{k-1})}\frac{(\lambda(t_k-t_{k-1}))^{r_k-r_{k-1}}}{(r_k-r_{k-1})!} \end{aligned}$$

gilt. Somit ist (44) bewiesen und der Beweis ist vollständig. □

Sicherlich stellt sich an dieser Stelle noch die Frage, warum ausgerechnet ein auf der Poissonverteilung basierender Prozess von besonderem Interesse ist. Einen möglichen Hinweis liefert die in Theorem 3.13 geschilderte Konstruktion eines Poissonprozesses. Nehmen wir an, dass der Ausfallprozess technischer Bauteile (Glühlampen,...) modelliert werden soll, wobei ein defektes Bauteil unverzüglich durch ein neues ersetzt wird. Die Lebensdauern der einzelnen Bauteile können nun durch eine Folge von unabhängigen und identisch verteilten Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ beschrieben werden. Nimmt man nun noch an, dass ein Ausfall eher spontan passiert und weniger mit der bisherigen Betriebsdauer des Bauteils zusammenhängt, so bietet sich eine Exponentialverteilung als Verteilungsgesetz für die Zufallsvariablen X_n an. Aus dem Kurs „Einführung in die Wahrscheinlichkeitstheorie und Mathematische Statistik“ im Sommersemester 2019 ist bekannt (siehe Übungsaufgabe 20, Übungsblatt 5), dass die Zufallsvariable $X \sim \operatorname{Exp}(\lambda)$ die sogenannte **Nichtalterungseigenschaft** besitzt. d.h.,

$$P\Big(X > t+s \big| X > t\Big) \;=\; P\big(X > s\big) \qquad \forall s,t > 0.$$

Weitere Bespiele für die Adäquatheit eines Poissonprozesse finden sich in der Literatur, z.B. für die Beschreibung der ankommenden Anrufe bei einem Callcenter in gegebenen Zeiträumen.

Übungsaufgabe

ÜA 3.5 $(N_t)_{t\in[0,\infty)}$ sei ein Poissonprozess mit Intensität $\lambda > 0$ auf $(\Omega, \mathcal{A}, P)$. Ein Prozess $(\widetilde{N}_t)_{t\in\mathbb{N}_0}$ sei gegeben durch $\widetilde{N}_t = N_t \quad \forall t \in \mathbb{N}_0 = \{0,1,\ldots\}$.
Zeigen Sie, dass für $0 = r_0 \le r_1 \le \ldots \le r_t$

$$P\Big(\widetilde{N}_t = r_t \big| \widetilde{N}_0 = r_0, \ldots, \widetilde{N}_{t-1} = r_{t-1}\Big) \;=\; e^{-\lambda}\,\frac{\lambda^{r_t - r_{t-1}}}{(r_t - r_{t-1})!} \;=\; P\Big(\widetilde{N}_t = r_t \big| \widetilde{N}_{t-1} = r_{t-1}\Big)$$

gilt!
($(\widetilde{N}_t)_{t\in\mathbb{N}_0}$ besitzt die sogenannte **Markoff-Eigenschaft**.)

3.5 Markoff-Ketten

In diesem Abschnitt werden wir eine weitere Klasse von stochastischen Prozessen untersuchen, sogenannte Markoffprozesse. Um den technischen Aufwand möglichst gering zu halten, konzentrieren wir uns einen den Spezialfall, wobei wir eine diskrete Menge ($T = \mathbb{N}_0$) für den Zeitbereich sowie einen höchstens abzählbaren Zustandsraum E für die Werte der Zufallsvariablen annehmen. Verallgemeinerungen auf den Fall überabzählbarer Indexmengen (z.B. $T = [0, \infty)$) oder auch reellwertiger Zufallsvariablen können mit ähnlichen Methoden behandelt werden, erfordern aber einen abstrakteren technischen Apparat. Hier kommen wir mit dem Begriff der elementaren bedingten Wahrscheinlichkeit aus, während gewisse Verallgemeinerungen den in der Vorlesung „Maßtheorie" behandelten Begriff bedingter Wahrscheinlichkeiten bzw. noch abstaktere Konzepte erfordern. Bevor wir uns einige Beispiele ansehen, stellen wir eine formale Definition voran:

Definition 3.14. *Eine Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}_0}$ auf $(\Omega, \mathcal{A}, P)$ mit Werten in einer endlichen oder abzählbar unendlichen Menge E heißt* **Markoff-Kette***, falls für alle $n \in \mathbb{N}_0$ und alle $x_0, \ldots, x_{n+1} \in E$ mit $P\big(X_n = x_n, \ldots, X_0 = x_0\big) > 0$ gilt, dass*

$$P\big(X_{n+1} = x_{n+1}\big|X_n = x_n, \ldots, X_0 = x_0\big) \;=\; P\big(X_{n+1} = x_{n+1}\big|X_n = x_n\big). \tag{45}$$

Die Menge E heißt **Zustandsraum** *der Markoff-Kette.*
Die Verteilung P^{X_0} von X_0 unter P heißt **Anfangsverteilung***,*
$p_n(x, y) := P\big(X_{n+1} = y\big|X_n = x\big)$ ist eine sogenannte **Übergangswahrscheinlichkeit***.*
Falls für alle $x, y \in E$ und $m, n \in \mathbb{N}_0$

$$P\big(X_{m+1} = y\big|X_m = x\big) \;=\; P\big(X_{n+1} = y\big|X_n = x\big),$$

so ist $(X_n)_{n\in\mathbb{N}_0}$ eine **(zeitlich) homogene Markoff-Kette***.*

Die wesentliche Eigenschaft einer Markoff-Kette ist die durch (45) beschriebene Eigenschaft der „Gedächtnislosigkeit", d.h., die bedingte Verteilung von X_{n+1} (und wie wir später sehen werden von $X_{n+1}, X_{n+2}, \ldots$) hängt nur vom „aktuellen Zustand" $X_n = x_n$, jedoch nicht von den vorangegangenen Zuständen ab.

Ein erstes Beispiel einer Markoff-Kette war bereits bei Übungsaufgabe 3.5 zu sehen. Falls $(N_t)_{t\in[0,\infty)}$ ein Poissonprozess mit Intensität $\lambda > 0$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ ist, so ist der „ausgedünnte" Prozess $(\widetilde{N}_t)_{t\in\mathbb{N}_0}$ mit $\widetilde{N}_t = N_t \quad \forall t \in \mathbb{N}_0$ ein Markoff-Prozess. Um dies zu sehen, wählen wir $n \in \mathbb{N}_0$ und $x_0, \ldots, x_{n+1} \in \mathbb{N}_0$. Es genügt, den Fall $0 = x_0 \leq x_1 \leq \ldots \leq x_n$ zu betrachten; andernfalls ist $P\big(X_n = x_n, \ldots, X_0 = x_0\big) = 0$. Nun folgt aus der Unabhängigkeit der Zuwächse sowie wegen $P(X_0 = x_0) = 0$, dass

$$\begin{aligned}
&P\big(X_{n+1} = x_{n+1}, X_n = x_n, \ldots, X_0 = x_0\big)\\
&\quad = \; P\big(X_{n+1} - X_n = x_{n+1} - x_n, X_n - X_{n-1} = x_n - x_{n-1}, \ldots, X_1 - X_0 = x_1 - x_0, X_0 = x_0\big)\\
&\quad = \; P\big(X_{n+1} - X_n = x_{n+1} - x_n, X_n - X_{n-1} = x_n - x_{n-1}, \ldots, X_1 - X_0 = x_1 - x_0\big)\\
&\quad = \; P\big(X_{n+1} - X_n = x_{n+1} - x_n\big) P\big(X_n - X_{n-1} = x_n - x_{n-1}, \ldots, X_1 - X_0 = x_1 - x_0\big)\\
&\quad = \; P\big(X_{n+1} - X_n = x_{n+1} - x_n\big) P\big(X_n - X_{n-1} = x_n - x_{n-1}, \ldots, X_1 - X_0 = x_1 - x_0, X_0 = x_0\big)\\
&\quad = \; P\big(X_{n+1} - X_n = x_{n+1} - x_n\big) P\big(X_n = x_n, \ldots, X_0 = x_0\big)
\end{aligned}$$

sowie

$$\begin{aligned}
&P\big(X_{n+1} = x_{n+1}, X_n = x_n\big)\\
&\quad = \; P\big(X_{n+1} - X_n = x_{n+1} - x_n, X_n - X_0 = x_n - x_0\big)\\
&\quad = \; P\big(X_{n+1} - X_n = x_{n+1} - x_n\big) P\big(X_n - X_0 = x_n - x_0\big)\\
&\quad = \; P\big(X_{n+1} - X_n = x_{n+1} - x_n\big) P\big(X_n = x_n\big).
\end{aligned}$$

Daraus folgt wegen $P\big(X_n = x_n\big) \geq P\big(X_n = x_n, \ldots, X_0 = x_0\big) > 0$, dass

$$\begin{aligned}
P\big(X_{n+1} = x_{n+1} \big| X_n = x_n, \ldots, X_0 = 0\big) \; &= \; P\big(X_{n+1} - X_n = x_{n+1} - x_n\big)\\
&= \; P\big(X_{n+1} = x_{n+1} \big| X_n = x_n\big),
\end{aligned}$$

d.h., die Markoff-Eigenschaft ist erfüllt.

Hier sind einige weitere Beispiele für Markoff-Ketten:

1) **Zufällige Irrfahrt auf $\mathbb{R}^d$**

Es seien $X_0, Y_1, Y_2, \ldots$ unabhängige Zufallsvariable mit Werten in $\mathbb{R}^d$. Dann ist der stochastische Prozess $(X_n)_{n \in \mathbb{N}_0}$ mit

$$X_n \; = \; X_0 \; + \; Y_1 \; + \; \cdots \; + \; Y_n$$

eine zufällige Irrfahrt auf $\mathbb{R}^d$. Üblicherweise ist $X_0 = x_0$ deterministisch und die Zuwächse $Y_1, Y_2, \ldots$ werden als identisch verteilt vorausgesetzt. Analog zu den obigen Betrachtungen folgt, dass dann $(X_n)_{n \in \mathbb{N}_0}$ eine Markoff-Kette ist.

2) **Der Galton-Watson-Prozess**

Ausgangspunkt für dieses Modell war die Beobachtung, dass immer wieder Adelsgeschlechter aus Mangel an männlichen Nachkommen ausstarben und somit immer mehr traditionsreiche Namen aus der adligen Gesellschaft verschwanden. Der Naturforscher Francis Galton und der Mathematiker Henry William Watson veöffentlichten im Jahr 1874 eine Arbeit, in der ein entsprechendes Modell vorgestellt wurde. Später wurde dieses Modell auch in anderen Gebieten eingesetzt, zur Modellierung der Ausbreitung von Populationen oder in der Warteschlangentheorie. Der Galton-Watson-Prozess zeichnet sich durch folgende Modellannahmen aus:

- Die erste Generation besteht aus einem Individuum. (Diese Einschränkung ist unproblematisch; andernfalls betrachtet man mehrere sich parallel entwickelnde Prozesse.)
- Jedes Individuum lebt exakt ein Zeitintervall lang.
- Das i-te Individuum im n-ten Zeitintervall produziert unahängig von allen anderen Individuen eine zufällige Anzahl $N_i^{(n)}$ von Nachkommen.
- Die Zufallsgrößen $N_i^{(n)}$ sind identisch verteilt mit Werten in $\mathbb{N}_0$.

Wenn wir nun mit X_n die zufällige Anzahl der Individuen der n-ten Generation bezeichnen, so haben wir nach Voraussetzung $X_0 = 1$ und für $n \geq 1$ erhalten wir

$$X_{n+1} = \sum_{i=1}^{X_n} N_i^{(n)}.$$

Da dies eine Summe von unabhängigen und identisch verteilten Zufallsvariablen ist, so hängt die bedingte Verteilung von X_{n+1} nur von X_n, jedoch nicht von den vorangegangenen Werten $X_{n-1}, X_{n-2}, \ldots$ ab. Für die Übergangswahrscheinlichkeiten erhalten wir

$$P\big(X_{n+1} = y \big| X_n = x\big) = P\Big(\sum_{i=1}^{x} N_i^{(n)} = y\Big).$$

Diese hängen offenbar nicht von n ab. Somit ist die Markoff-Kette homogen.

3) **Das Ehrenfest-Modell**

Dieses Modell wurde am Anfang des 20. Jahrhunderts von Paul Ehrenfest und seiner Ehefrau Tatjana Ehrenfest-Afanassjeva zur Beschreibung von Diffusionsvorgängen von Molekülen durch eine Membran eingeführt. Dabei wird vorausgesetzt, dass sich insgesamt N Teilchen in zwei miteinander durchlässig verbundenen, aber nach außen abgeschlossenen Behältern befinden. Zum Zeitpunkt 0 mögen sich x_0 Teilchen in Behälter A sowie $N - x_0$ Teilchen in Behälter B befinden. Zu den Zeiten $n \in \mathbb{N}$ bewegt sich jeweils ein rein zufällig ausgewähltes Teilchen in den anderen Behälter. Falls nun X_n die Anzahl der Teilchen in Behälter A zum Zeitpunkt n beschreibt, so ist $(X_n)_{n \in \mathbb{N}}$ eine Markoff-Kette und es gilt

$$P\big(X_{n+1} = y \big| X_n = x\big) = \begin{cases} \frac{x}{N}, & \text{falls } y = x - 1, \\ \frac{N-x}{N}, & \text{falls } y = x + 1, \\ 0, & \text{falls } |x - y| \neq 1. \end{cases}$$

Ein typische Fragestellung ist hierbei, ob sich die Markoff-Kette „einpendelt“, d.h., ob sich die Verteilung von X_n einem Äquilibrium nähert, falls $n \to \infty$.

4) **Ruinproblem**

Zwei Spieler spielen wiederholt gegeneinander. Spieler 1 beginnt mit einem Guthaben von n Euro, Spieler 2 mit $N-n$ Euro. Bei jedem Spiel gewinnt Spieler 1 mit Wahrscheinlichkeit p und erhält daraufhin von seinem Gegenspieler einen Euro. Spieler 2 gewinnt mit Wahrscheinlichkeit $q = 1-p$ und erhält in diesem Fall ebenfalls einen Euro von seinem Gegner. Das Spiel endet, wenn einer der beiden Spieler sein gesamtes Geld verspielt hat. Typische Fragestellungen sind hierbei, ob das Spiel jemals endet und mit welchen Wahrscheinlichkeiten Spieler 1 bzw. Spieler 2 verlieren.
Dieses Spiel lässt sich nun als Markoff-Kette mit Zustandsraum $E = \{0, 1, \ldots, N\}$ modellieren, wobei $X_n = k$ bedeutet, dass Spieler 1 zum Zeitpunkt n ein Guthaben von k Euro hat. Es gilt

$$P\big(X_{n+1} = y \big| X_n = x\big) = \begin{cases} p, & \text{falls } y = x+1, \\ q, & \text{falls } y = x-1, \\ 0, & \text{falls } |x-y| \neq 1. \end{cases}$$

Die Definition der Markoff-Eigenschaft besagt, dass das Verhalten einer Markoff-Kette zum Zeitpunkt $n+1$ nur vom Zustand zur Zeit n, nicht jedoch von den Zuständen zu den vorangegangenen Zeitpunkten abhängt. Die folgenden Aussagen beschreiben, dass auch die gesamte „Zukunft“ nach n diese Eigenschaft besitzt.

Lemma 3.15. *$(X_n)_{n\in\mathbb{N}_0}$ sei eine auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ definierte Markoff-Kette mit einem höchstens abzählbaren Zustandsraum E. Dann gilt für alle $n, k \in \mathbb{N}$, $x_0, \ldots, x_{n+k} \in E$ mit $P\big(X_n = x_n, \ldots, X_0 = x_0\big) > 0$:*

$$\begin{aligned} &P\big(X_{n+k} = x_{n+k}, \ldots, X_{n+1} = x_{n+1} \big| X_n = x_n, \ldots, X_0 = x_0\big) \\ &= P\big(X_{n+k} = x_{n+k}, \ldots, X_{n+1} = x_{n+1} \big| X_n = x_n\big). \end{aligned}$$

Beweis. Aus dem Multiplikationssatz und der Markoff-Eigenschaft folgen

$$\begin{aligned} &P\big(X_{n+k} = x_{n+k}, \ldots, X_0 = x_0\big) \\ &= P\big(X_{n+k} = x_{n+k} \big| X_{n+k-1} = x_{n+k-1}, \ldots, X_0 = x_0\big) \cdots P\big(X_{n+1} = x_{n+1} \big| X_n = x_n, \ldots, X_0 = x_0\big) \\ &\qquad P\big(X_n = x_n, \ldots, X_0 = x_0\big) \\ &= P\big(X_{n+k} = x_{n+k} \big| X_{n+k-1} = x_{n+k-1}\big) \cdots P\big(X_{n+1} = x_{n+1} \big| X_n = x_n\big) \\ &\qquad P\big(X_n = x_n, \ldots, X_0 = x_0\big) \end{aligned}$$

sowie

$$\begin{aligned} &P\big(X_{n+k} = x_{n+k}, \ldots, X_n = x_n\big) \\ &= P\big(X_{n+k} = x_{n+k} \big| X_{n+k-1} = x_{n+k-1}, \ldots, X_n = x_n\big) \cdots P\big(X_{n+1} = x_{n+1} \big| X_n = x_n\big) \\ &\qquad P\big(X_n = x_n\big) \\ &= P\big(X_{n+k} = x_{n+k} \big| X_{n+k-1} = x_{n+k-1}\big) \cdots P\big(X_{n+1} = x_{n+1} \big| X_n = x_n\big) \\ &\qquad P\big(X_n = x_n\big). \end{aligned}$$

Unter Berücksichtigung von $P\big(X_n = x_n\big) \geq P\big(X_n = x_n, \ldots, X_0 = x_0\big) > 0$ folgt daraus

$$\begin{aligned} &P\big(X_{n+k} = x_{n+k}, \ldots, X_{n+1} = x_{n+1} \big| X_n = x_n, \ldots, X_0 = x_0\big) \\ &= P\big(X_{n+k} = x_{n+k}, \ldots, X_{n+1} = x_{n+1} \big| X_n = x_n\big). \end{aligned}$$

□

Die Aussage von Lemma 3.15 lässt sich nun leicht auf das zukünftige Verhalten der Markoff-Kette über unendliche Zeiträume erweitern. Dazu definieren wir für $k \in \mathbb{N}$

$$\mathcal{Z}_k := \left\{ B \times E \times E \times \cdots \middle| B \in E^k \right\}$$

das System der k-dimensionalen Zylindermengen in $E^\infty := E \times E \times \cdots$, mit

$$\mathcal{Z} := \bigcup_{k=1}^{\infty} \mathcal{Z}_k$$

das System aller Zylindermengen sowie mit

$$\mathcal{A}^\infty = \sigma\big(\mathcal{Z}\big)$$

die davon erzeugte σ-Algebra in $E^\infty = E \times E \times \cdots$.

Theorem 3.16. *$(X_n)_{n\in\mathbb{N}_0}$ sei eine auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ definierte Markoff-Kette mit einem höchstens abzählbaren Zustandsraum E. Dann gilt für alle $n \in \mathbb{N}$, $x_0, \dots, x_n \in E$ mit $P\big(X_n = x_n, \dots, X_0 = x_0\big) > 0$:*

$$\begin{aligned} &P\big((X_{n+1}, X_{n+2}, \dots) \in A \big| X_n = x_n, \dots, X_0 = x_0\big) \\ &= P\big((X_{n+1}, X_{n+2}, \dots) \in A \big| X_n = x_n\big) \qquad \forall A \in \mathcal{A}^\infty. \end{aligned}$$

Beweis. Es sei zunächst $A \in \mathcal{Z}_k$, d.h., $A = B \times E \times E \times \cdots$ für ein $B \in E^k$. Nun folgt aus Lemma 3.15 und der σ-Additivität

$$\begin{aligned} &P\big((X_{n+1}, X_{n+2}, \dots) \in A \big| X_n = x_n, \dots, X_0 = x_0\big) \\ &= P\big((X_{n+1}, \dots, X_{n+k}) \in B \big| X_n = x_n, \dots, X_0 = x_0\big) \\ &= \sum_{(x_{n+1},\dots,x_{n+k}) \in B} P\big(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \in B \big| X_n = x_n, \dots, X_0 = x_0\big) \\ &= \sum_{(x_{n+1},\dots,x_{n+k}) \in B} P\big(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \in B \big| X_n = x_n\big) \\ &= P\big((X_{n+1}, \dots, X_{n+k}) \in B \big| X_n = x_n\big) \\ &= P\big((X_{n+1}, X_{n+2}, \dots) \in A \big| X_n = x_n\big). \end{aligned}$$

Nun sind jedoch $P\big((X_{n+1}, X_{n+2}, \dots) \in \cdot \big| X_n = x_n, \dots, X_0 = x_0\big)$ und $P\big((X_{n+1}, X_{n+2}, \dots) \in \cdot \big| X_n = x_n\big)$ Wahrscheinlichkeitsmaße, welche auf dem durchschnittsstabilen Mengensystem $\mathcal{Z}$ übereinstimmen. Nach dem Eindeutigkeitssatz der Maßtheorie folgt die Gleichheit dieser Wahrscheinlichkeitsmaße auf $\mathcal{A}^\infty$. □

Übungsaufgabe

ÜA 3.6 $(Y_n)_{n\in\mathbb{N}_0}$ sei eine Folge von unabhängigen Zufallsvariablen mit $P(Y_n = 1) = p$ und $P(Y_n = 0) = 1 - p$. Es sei $X_n := 2Y_n + Y_{n+1}$.
Begründen Sie, dass $(X_n)_{n\in\mathbb{N}}$ eine Markoff-Kette ist und bestimmen Sie die Übergangswahrscheinlichkeiten!

Wir werden als Nächstes das Verhalten von Markoff-Ketten untersuchen, welche schon lange gelaufen sind. Unter geeigneten Voraussetzungen kommt es zu dem Effekt, dass sich das zufällige Verhalten der Markoff-Kette „einpendelt", d.h., unabhängig von Start zum Zeitpunkt $n = 0$ werden die Verteilungen der X_n mit $n \to \infty$ gegen eine Grenzverteilung streben. Maximale Stabilität liegt vor, wenn die Verteilung der X_n gleich bleibt. Ein solches Verhalten der Markoff-Kette ist nur denkbar, wenn auch die Übergangswahrscheinlichkeiten sich nicht über die Zeit ändern, d.h., wenn sie zeitlich **homogen** sind. Es sei also im Weiteren $(X_n)_{n\geq 0}$ eine Markoff-Kette auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Die Markoff-Kette möge einen höchstens abzählbaren Zustandsraum E sowie homogene Übergangswahrscheinlichkeiten $p_{xy} = P\big(X_{n+1} = y \big| X_n = x\big)$ besitzen. Die Übergangswahrscheinlichkeiten können nun in einer sogenannten Übergangsmatrix P angeodnet werden, wobei $P := (p_{xy})_{x,y\in E}$. Falls E eine endliche Menge mit N Elementen ist, z.B. $E = \{1, 2, \dots, N\}$, so enthält die i-te Zeile von P die Übergangswahrscheinlichkeiten $P\big(X_{n+1} = 1 \big| X_n = i\big), P\big(X_{n+1} = 2 \big| X_n = i\big), \dots, P\big(X_{n+1} = N \big| X_n = i\big)$. Wir bezeichnen mit $p_x^{(0)} = P\big(X_0 = x\big)$ die Wahrscheinlichkeiten der Anfangsverteilung und fassen diese in einem Zeilenvektor $p^{(0)} := (p_x^{(0)})_{x\in E}$ zusammen. Dementsprechend ist $p^{(0)}P = (P(X_1 = x))_{x\in E}$ der Vektor der Wahrscheinlichkeiten zum Zeitpunkt $n = 1$, pP^2 der Vektor der Wahrscheinlichkeiten zum Zeitpunkt $n = 2$ u.s.w. Hier kommt nun die Definition einer stationären Verteilung:

Definition 3.17. $(X_n)_{n\in\mathbb{N}_0}$ *sei eine Markoff-Kette auf* $(\Omega, \mathcal{A}, P)$ *mit Zustandsraum* E *und Übergangsmatrix* P*. Eine Wahrscheinlichkeitsverteilung* π *auf* E*, gegeben durch einen Vektor* $\pi = (\pi_x)_{x\in E}$*, heißt* **stationäre Verteilung***, falls*

$$\pi P = \pi,$$

d.h.,

$$\pi_y = \sum_{x\in E} \pi_x p_{xy} \qquad \forall y \in E.$$

Für einige der bereits erwähnten Beispiele lassen sich die stationären Verteilungen leicht bestimmen.

- „Ruinproblem"

Damit die Markoff-Kette zu allen Zeiten $n \in \mathbb{N}_0$ wohldefiniert ist legen wir fest, dass die Kette nach dem „Ruin" eines Spielers im zuletzt erreichten Zustand liegen bleibt. Damit ergeben sich folgende Übergangswahrscheinlichkeiten:

$$p_{xy} = P\big(X_{n+1} = y \big| X_n = x\big) = \begin{cases} p, & \text{falls} \quad 0 < x < N,\ y = x+1, \\ q, & \text{falls} \quad 0 < x < N,\ y = x-1, \\ 1, & \text{falls} \quad x = y = 0 \quad \text{oder} \quad x = y = N, \\ 0, & \text{sonst.} \end{cases}$$

Die Übergangsmatrix hat dann folgende Gestalt:

$$P = \big(p_{xy}\big)_{x,y=0,\dots,N} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ q & 0 & p & \ddots & & \vdots \\ 0 & q & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & p & 0 \\ \vdots & & \ddots & q & 0 & p \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Mögliche stationäre Verteilungen können nun relativ leicht erraten werden. Wegen des dann eintretenden Ruins eines der Spieler sind die Zustände 0 und N **"absorbierend"**, d.h., erreicht die Markoff-Kette einen dieser Zustände, so verbleibt sie in diesem für immer. Folglich sind $\pi^{(0)} = (1, 0, \dots, 0)$ sowie $\pi^{(N)} = (0, \dots, 0, 1)$ offensichtlich stationäre Verteilungen. Man kann dies auch formal nachprüfen; es gilt jeweils $\pi^{(i)} = \pi^{(i)}P$. Aus dem gleichen Grund sind auch $\pi := (\alpha, 0, \dots, 0, 1-\alpha)$ für $\alpha \in (0,1)$ stationäre Verteilungen. Ein etwas genaueres Hinsehen ist notwendig um festzustellen, dass damit aber die Menge der stationären Verteilungen erschöpft ist. Angenommen, $\pi = \big(\pi_0, \dots, \pi_N\big)$ mit $\pi_x > 0$ für ein $x \in \{1, \dots, N-1\}$ sei eine stationäre Verteilung. Wegen $p_{x,x+1} = p > 0$ und $p_{x,x-1} = q > 0$ folgt dann auch, dass $\pi_{x+1} > 0$ und $\pi_{x-1} > 0$ sind. Auf diese Art folgt, dass insbesondere auch $\pi_1 > 0$ und $\pi_{N-1} > 0$ gelten. Nun folgen aber sowohl

$$\sum_{x\in E} \pi_x p_{x0} = \pi_0 + \pi_1 \underbrace{p_{10}}_{=q>0} > \pi_0$$

als auch

$$\sum_{x\in E} \pi_x p_{xN} = \pi_N + \pi_{N-1} \underbrace{p_{N-1,N}}_{=p>0} > \pi_N.$$

Damit ist klar, dass π keine stationäre Verteilung ist. Anhand dieses Beispiels wird außerdem klar, dass eine stationäre Verteilung nicht eindeutig bestimmt sein muss. Kriteriem für die Existenz und auch für die Eindeutigkeit der stationären Verteilung werden im Folgenden hergeleitet. Wir beschränken uns hier auf den Fall eines endlichen Zustandsraumes E.

Theorem 3.18. *$(X_n)_{n\in\mathbb{N}_0}$ sei eine Markoff-Kette auf $(\Omega, \mathcal{A}, P)$ mit einem endlichen Zustandsraum E und homogenen Übergangswahrscheinlichkeiten $p_{xy} = P\big(X_{n+1} = y \big| X_n = x\big)$ für $n \in \mathbb{N}_0$ und $x, y \in E$.*
Dann besitzt $(X_n)_{n\in\mathbb{N}_0}$ mindestens eine stationäre Verteilung $\pi = \big(\pi_x\big)_{x\in E}$.

Beweis. Wir nehmen an, dass die Markoff-Kette gemäß einer beliebigen Anfangsverteilung $p^{(0)} := (p_x^{(0)})_{x\in E}$ gestartet wird. Es sei $p^{(n)}$ die Verteilung der Markoff-Kette nach n Schritten, d.h.,

$$p^{(n)} \;=\; \big(P(X_n = x)\big)_{x\in E} \;=\; p^{(0)}P^n,$$

wobei P^n das n-fache Produkt der Übergangsmatrix P ist. Wir betrachten die folgenden Cesàro-Mittel:

$$\pi^{(n)} \;:=\; \frac{1}{n}\sum_{k=1}^{n} p^{(k)}.$$

Die $\pi^{(n)}$ liegen in der kompakten (da abgeschlossen und beschränkt) Teilmenge $M := \big\{(a_x)_{x\in E} \colon a_x \in [0,1]\ \forall x \in E,\ \ \sum_{x\in E} a_x = 1\big\}$ des $\mathbb{R}^K$, wobei K die Anzahl der Elemente von E ist. Daher existiert eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ von $\mathbb{N}$ mit

$$\pi^{(n_k)} \underset{k\to\infty}{\longrightarrow} \pi^{(\infty)}$$

für ein gewisses $\pi^{(\infty)} \in M$. Nun gilt:

$$\begin{aligned}
\sum_{x\in E} \pi_x^{(n_k)} p_{xy} \;&=\; \sum_{x\in E}\Big(\frac{1}{n_k}\sum_{l=1}^{n_k} p_x^{(l)}\Big)p_{xy} \\
&=\; \frac{1}{n_k}\sum_{l=1}^{n_k}\underbrace{\sum_{x\in E} p_x^{(l)}p_{xy}}_{p_y^{(l+1)}} \\
&=\; \underbrace{\frac{1}{n_k}\sum_{l=1}^{n_k} p_y^{(l)}}_{=\pi_y^{(n_k)}} + \underbrace{\frac{1}{n_k}\Big\{p_y^{(n_k+1)} \;-\; p_y^{(1)}\Big\}}_{\underset{k\to\infty}{\longrightarrow}\,0}.
\end{aligned}$$

An dieser Stelle sehen wir, warum wir Cesàro-Mittel $\pi_y^{(n)}$ betrachtet haben. Bis auf einen gegen 0 konvergierenden Restterm stimmt $\sum_{x\in E}\pi_x^{(n_k)}p_{xy}$ mit $\pi_y^{(n_k)}$ überein. Mit $k\to\infty$ erhalten wir nun

$$\sum_{x\in E}\pi_x^{(\infty)}p_{xy} \;=\; \pi_y^{(\infty)} \qquad \forall y\in E,$$

d.h., der Vektor $\pi^{(\infty)} = \big(\pi_x^{(\infty)}\big)_{x\in E}$ beschreibt eine stationäre Verteilung. □

Das eben bewiesene Theorem 3.18 liefert zwar die Existenz einer stationären Verteilung, nicht aber deren Eindeutigkeit. Um die Nichteindeutigkeit auszuschließen, müssen wir eine Situation wie beim Ruinproblem vermeiden, wo die beiden absorbierenden Zustände 0 und N dafür sorgen, dass mehr als eine stationäre Verteilung existiert. Die folgende Aussage liefert ein entsprechendes Resultat.

Theorem 3.19. *$(X_n)_{n\in\mathbb{N}_0}$ sei eine Markoff-Kette auf $(\Omega,\mathcal{A},P)$ mit einem endlichen Zustandsraum E und homogenen Übergangswahrscheinlichkeiten $p_{xy}=P\big(X_{n+1}=y\big|X_n=x\big)$ für $n\in\mathbb{N}_0$ und $x,y\in E$. $p_{xy}^{(n)}$ $(x,y\in E,\ n\in\mathbb{N})$ seien die n-Schritt-Übergangswahrscheinlichkeiten. Zusätzlich setzen wir voraus, dass ein $n_0\in\mathbb{N}$ und ein $y_0\in E$ existieren mit*

$$p_{x,y_0}^{(n_0)}>0 \qquad \forall x\in E.$$

(i) Es existiert eine Wahrscheinlichkeitsverteilung $\pi=(\pi_x)_{x\in E}$, sodass bei einer beliebigen Startverteilung $p^{(0)}=P^{X_0}$ gilt

$$P\big(X_n=x\big)\underset{n\to\infty}{\longrightarrow}\pi_x \qquad \forall x\in E.$$

(ii) π ist die eindeutig bestimmte stationäre Verteilung von $(X_n)_{n\in\mathbb{N}_0}$.

Beweis. (i) Nach Theorem 3.18 besitzt die Markoff-Kette mindestens eine stationäre Verteilung. π sei eine beliebige dieser Verteilungen. Wir zeigen, dass dann

$$P\big(X_n=x\big)\underset{n\to\infty}{\longrightarrow}\pi_x \qquad \forall x\in E \tag{46}$$

gilt. Dazu benutzen wir eine typische **Kopplungsmethode**: Auf einem geeigneten Wahrscheinlichkeitsraum $(\widetilde{\Omega},\widetilde{\mathcal{A}},\widetilde{P})$ werden zwei Markoff-Ketten $(\widetilde{X}_n)_{n\in\mathbb{N}_0}$ und $(\widetilde{X}'_n)_{n\in\mathbb{N}_0}$ mit jeweiligen Übergangswahrscheinlichkeiten p_{xy} so konstruiert, dass $\widetilde{X}_0\sim p^{(0)}$ und $\widetilde{X}'_0\sim\pi$ erfüllt sind. D.h., die erste Kette wird mit der beliebig gewählten Verteilung $p^{(0)}$ gestartet, während die zweite mit der beliebig gewählten stationären Verteilung π gestartet wird. Nun gelten

$$\widetilde{P}^{\widetilde{X}_n}=P^{X_n}$$

und, wegen Stationarität der zweiten Kette,

$$\widetilde{P}^{\widetilde{X}'_n}=\pi.$$

Die Kette $(\widetilde{X}_n)_{n\in\mathbb{N}_0}$ imitiert somit das zufällige Verhalten der Originalkette $(X_n)_{n\in\mathbb{N}_0}$, während die zweite Kette $(\widetilde{X}'_n)_{n\in\mathbb{N}_0}$ den Vergleich zum stationären Verhalten liefert. Beide Ketten sollen nun unter Berücksichtigung der vorgegebenen Übergangswahrscheinlichkeiten derart miteinander gekoppelt werden, dass

$$\widetilde{P}\big(\widetilde{X}_n\neq\widetilde{X}'_n\big)\underset{n\to\infty}{\longrightarrow}0$$

gilt. Daraus folgt dann (46).
Die durch Kopplung erzeugte Markoff-Kette hat Werte in $E\times E$, eine Anfangsverteilung mit $\big(\widetilde{X}_0,\widetilde{X}'_0\big)^T\sim p^{(0)}\otimes\pi$ und besitzt die folgenden Übergangswahrscheinlichkeiten:

$$\widetilde{P}\left(\begin{pmatrix}\widetilde{X}_{n+1}\\ \widetilde{X}'_{n+1}\end{pmatrix}=\begin{pmatrix}y\\ y'\end{pmatrix}\middle|\begin{pmatrix}\widetilde{X}_n\\ \widetilde{X}'_n\end{pmatrix}=\begin{pmatrix}x\\ x'\end{pmatrix}\right)=\begin{cases}p_{xy}\,p_{x'y'}, & \text{falls } x\neq x',\\ p_{xy}, & \text{falls } x=x',\ y=y',\\ 0, & \text{falls } x=x',\ y\neq y'.\end{cases}$$

Diese Vorschrift bedeutet, dass sich die beiden Markoff-Ketten $(\widetilde{X}_n)_{n\in\mathbb{N}_0}$ und $(\widetilde{X}'_n)_{n\in\mathbb{N}_0}$ voneinander **unabhängig** entwickeln bis sie zu einem zufälligen Zeitpunkt $T := \inf\big\{n \geq 0 : \ \widetilde{X}_n = \widetilde{X}'_n\big\}$ erstmals zusammentreffen. Ab diesem Zeitpunkt fallen die beiden Ketten zusammen, d.h., $\widetilde{P}\big(\widetilde{X}_n = \widetilde{X}'_n \quad \forall n \geq T\big) = 1$. Die Konstruktion sichert außerdem, dass sich beide Markoff-Ketten gemäß der vorgegebenen Übergangswahrscheinlichkeiten p_{xy} entwickeln. Daher gilt die folgende entscheidende Abschätzung:

$$
\begin{aligned}
&\sum_{x\in E}\Big|\underbrace{P(X_n = x)}_{=\widetilde{P}(\widetilde{X}_n = x)} - \underbrace{\pi_x}_{=\widetilde{P}(\widetilde{X}'_n = x)}\Big| \\
&\quad = \sum_{x\in E}\big|\widetilde{P}(\widetilde{X}_n = x) - \widetilde{P}(\widetilde{X}'_n = x)\big| \\
&\quad \leq \sum_{x\in E}\Big|\underbrace{\widetilde{P}\big(\widetilde{X}_n = x, \widetilde{X}_n = \widetilde{X}'_n\big) - \widetilde{P}\big(\widetilde{X}'_n = x, \widetilde{X}_n = \widetilde{X}'_n\big)}_{=0} \\
&\qquad\qquad + \widetilde{P}\big(\widetilde{X}_n = x, \widetilde{X}_n \neq \widetilde{X}'_n\big) - \widetilde{P}\big(\widetilde{X}'_n = x, \widetilde{X}_n \neq \widetilde{X}'_n\big)\Big| \\
&\quad = \sum_{x\in E}\Big|\widetilde{P}\big(\widetilde{X}_n = x, \widetilde{X}_n \neq \widetilde{X}'_n\big) - \widetilde{P}\big(\widetilde{X}'_n = x, \widetilde{X}_n \neq \widetilde{X}'_n\big)\Big| \\
&\quad \leq \sum_{x\in E}\widetilde{P}\big(\widetilde{X}_n = x, \widetilde{X}_n \neq \widetilde{X}'_n\big) + \sum_{x\in E}\widetilde{P}\big(\widetilde{X}'_n = x, \widetilde{X}_n \neq \widetilde{X}'_n\big) \\
&\quad = 2\,\widetilde{P}\big(\widetilde{X}_n \neq \widetilde{X}'_n\big).
\end{aligned}
\tag{47}
$$

Es ist nun noch zu zeigen, dass $\widetilde{P}\big(\widetilde{X}_n \neq \widetilde{X}'_n\big) \underset{n\to\infty}{\longrightarrow} 0$. Es sei

$$\alpha := \min\big\{p^{(n_0)}_{x,y_0} : \quad x \in E\big\}.$$

Nun gilt

$$
\begin{aligned}
&\widetilde{P}\big(\widetilde{X}_{k+n_0} \neq \widetilde{X}'_{k+n_0}\big|\widetilde{X}_k \neq \widetilde{X}'_k\big) \\
&\quad = \frac{\sum_{x,x'\in E,\, x\neq x'}\widetilde{P}\big(\widetilde{X}_{k+n_0} \neq \widetilde{X}'_{k+n_0},\ \widetilde{X}_k = x,\ \widetilde{X}'_k = x'\big)}{\sum_{x,x'\in E,\, x\neq x'}\widetilde{P}\big(\widetilde{X}_k = x,\ \widetilde{X}'_k = x'\big)} \\
&\quad = \frac{\sum_{x,x'\in E,\, x\neq x'}\widetilde{P}\big(\widetilde{X}_{k+n_0} \neq \widetilde{X}'_{k+n_0}\big|\widetilde{X}_k = x,\ \widetilde{X}'_k = x'\big)\widetilde{P}\big(\widetilde{X}_k = x,\ \widetilde{X}'_k = x'\big)}{\sum_{x,x'\in E,\, x\neq x'}\widetilde{P}\big(\widetilde{X}_k = x,\ \widetilde{X}'_k = x'\big)} \\
&\quad \leq 1 - \min\big\{\widetilde{P}\big(\widetilde{X}_{k+n_0} = \widetilde{X}'_{k+n_0} = y_0\big|\widetilde{X}_k = x, \widetilde{X}'_k = x'\big) : \quad x, x' \in E,\ \widetilde{P}\big(\widetilde{X}_k = x,\ \widetilde{X}'_k = x'\big) > 0\big\} \\
&\quad \leq 1 - \alpha^2.
\end{aligned}
$$

Daher erhalten wir mit $k_n := \big[n/n_0\big]$ (k_n ist die größte ganze Zahl, welche kleiner oder gleich n/n_0 ist.)

$$
\begin{aligned}
&\widetilde{P}\big(\widetilde{X}_n \neq \widetilde{X}'_n\big) \\
&\quad \leq \widetilde{P}\big(\widetilde{X}_{k_n n_0} \neq \widetilde{X}'_{k_n n_0}\big) \\
&\quad = \underbrace{\widetilde{P}\big(\widetilde{X}_{n_0} \neq \widetilde{X}'_{n_0}\big)}_{\leq 1-\alpha^2}\underbrace{\widetilde{P}\big(\widetilde{X}_{2n_0} \neq \widetilde{X}'_{2n_0}\big|\widetilde{X}_{n_0} \neq \widetilde{X}'_{n_0}\big)}_{\leq 1-\alpha^2}\cdots\underbrace{\widetilde{P}\big(\widetilde{X}_{k_n n_0} \neq \widetilde{X}'_{k_n n_0}\big|\widetilde{X}_{(k_n-1)n_0} \neq \widetilde{X}'_{(k_n-1)n_0}\big)}_{\leq 1-\alpha^2} \\
&\quad \leq \big(1 - \alpha^2\big)^{k_n} \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

Dies ergibt zusammen mit (47) Behauptung (i).

(ii) Die Verteilung π wurde am Anfang des Beweises von (i) als eine stationäre Verteilung gewählt. Es ist noch deren Eindeutigkeit zu begründen. Falls nun $\pi' = (\pi'_x)_{x\in E}$ eine beliebige stationäre Verteilung ist, so folgt mit denselben Schritten wie im Beweis von (i), dass

$$P\big(X_n = x\big) \underset{n\to\infty}{\longrightarrow} \pi'_x \qquad \forall x \in E.$$

Damit folgt aber wegen (46), dass $\pi'_x = \pi_x$ für alle $x \in E$ gilt. Somit ist die stationäre Verteilung eindeutig bestimmt. □

- Ehrenfest-Modell

Es sei daran erinnert, dass bei diesem Modell der Zustandsraum $E = \{0, 1, \dots, N\}$ ist und die Übergangswahrscheinlichkeiten durch

$$p_{xy} = \begin{cases} \frac{x}{N}, & \text{falls } y = x - 1, \\ \frac{N-x}{N}, & \text{falls } y = x + 1, \\ 0, & \text{falls } |x - y| \neq 1 \end{cases}$$

gegeben sind. Wir erhalten folgende Matrix der Übergangswahrscheinlichkeiten:

$$P = (p_{xy})_{x,y=0,\dots,N} = \begin{pmatrix} 0 & \frac{N}{N} & 0 & \dots & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & \ddots & \vdots \\ 0 & \frac{2}{N} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & \frac{1}{N} \\ 0 & \dots & 0 & \frac{N}{N} & 0 \end{pmatrix}.$$

Nach Theorem 3.18 lässt sich ein Kandidat für eine stationäre Verteilung erahnen. Angenommen, die Markoff-Kette läuft schon sehr lange (n groß), so wird wohl jedes der N Teilchen mehrmals den Behälter gewechselt haben. Folglich könnte ein stationäres Regime so aussehen, dass sich ein jedes Teilchen jeweils mit Wahrscheinlichkeit 1/2 in Behälter A oder Behälter B befindet. Dies würde bedeuten, dass die Anzahl der Teilchen in Behälter A eine Binomialverteilung mit Parametern N und 1/2 besitzt. Das lässt sich folgendermaßen verifizieren. Es sei

$$\pi = \big(\pi_0, \pi_1, \dots, \pi_N\big) = \left(\binom{N}{0}2^{-N}, \binom{N}{1}2^{-N}, \dots, \binom{N}{N}2^{-N}\right).$$

Nun gelten

$$\begin{aligned} \sum_{x\in E} \pi_x p_{x0} &= \pi_1 p_{10} = \binom{N}{1}2^{-N}\,\frac{1}{N} = 2^{-N} = \pi_0, \\ \sum_{x\in E} \pi_x p_{xN} &= \pi_{N-1} p_{N-1,N} = \binom{N}{N-1}2^{-N}\,\frac{1}{N} = 2^{-N} = \pi_N \end{aligned}$$

sowie, für $y \in \{1, \dots, N-1\}$,

$$\begin{aligned} \sum_{x\in E} \pi_x p_{xy} &= \pi_{y-1} p_{y-1,y} + \pi_{y+1} p_{y+1,y} \\ &= \binom{N}{y-1}2^{-N}\,\frac{N-(y-1)}{N} + \binom{N}{y+1}2^{-N}\,\frac{y+1}{N} \\ &= \left\{\binom{N-1}{y-1} + \binom{N-1}{y}\right\}2^{-N} = \binom{N}{y}2^{-N} = \pi_y. \end{aligned}$$

Somit ist π eine stationäre Verteilung. Die Eindeutigkeit der stationären Verteilung ist hier weniger offensichtlich, folgt aber nach Theorem 3.19.

Übungsaufgabe

ÜA 3.7 Ein stochastischer Prozess $(X_n)_{n\in\mathbb{N}_0}$ auf $(\Omega, \mathcal{A}, P)$ mit einem höchstens abzählbaren Zustandsraum E ist eine **Markoff-Kette** m**-ter Ordnung**, falls für alle $n \geq m-1$ und $x_0, \ldots, x_{n+1} \in E$ mit $P\big(X_n = x_n, \ldots, X_0 = x_0\big) > 0$ gilt, dass

$$P\big(X_{n+1} = x_{n+1} \big| X_n = x_n, \ldots, X_0 = x_0\big) \;=\; P\big(X_{n+1} = x_{n+1} \big| X_n = x_n, \ldots, X_{n-m+1} = x_{n-m+1}\big).$$

Es seien nun $E' := E^m$ das m-fache Produkt des Zustandsraumes E und $X'_n := (X_{n+m-1}, \ldots, X_n)$.
Zeigen sie, dass $(X_n)_{n\in\mathbb{N}_0}$ eine Markoff-Kette der Ordnung m genau dann ist, wenn $(X'_n)_{n\in\mathbb{N}_0}$ eine Markoff-Kette der Ordnung 1 ist!

ÜA 3.8 $(Y_n)_{n\in\mathbb{N}_0}$ sei eine Folge von unabhängigen Zufallsvariablen mit $P(Y_n = 1) = p = 1 - P(Y_n = -1)$ für alle $n \in \mathbb{N}_0$.
Zeigen sie, dass $(X_n)_{n\in\mathbb{N}_0}$ mit $X_n = Y_0 \cdots Y_n$ eine Markoff-Kette ist, bestimmen Sie die Übergangsmatrix sowie die stationäre Verteilung!

Absorptionswahrscheinlichkeiten

Bei einer Markoff-Kette kann es gewisse Zustände geben, welche nach einem erstmaligen Erreichen nicht mehr verlassen werden. Die Wahrscheinlichkeit, dass ein solcher Zustand erreicht wird, kann dabei durchaus von Interesse sein. Beim Galton-Watson-Prozess (Beispiel 2) bedeutet das Erreichen des Zustandes 0, dass die durch diesen Prozess modellierte Population ausstirbt. Beim Ruinproblem (Beispiel 4) gewinnt Spieler 1, falls der Zustand N erreicht wird und Spieler 2, falls 0 erreicht wird. Ausgehend vom Anfangskapital x des Spielers 1 kann man sich hier nach der Wahrscheinlichkeit eines Gewinns durch diesen Spieler fragen. Im Weiteren untersuchen wir die Wahrscheinlichkeit, dass eine im Punkt x gestartete Markoff-Kette diesen Zustand erreicht. Zur klaren Kennzeichnung des Falles einer in x deterministisch gestarteten Markoff-Kette bezeichnen wir das entsprechende Wahrscheinlichkeitsmaß mit P^x.

Definition 3.20. $(X_n)_{n\in\mathbb{N}_0}$ *sei eine Markoff-Kette mit einem höchstens abzählbaren Zustandsraum* E *und homogenen Übergangswahrscheinlichkeiten* p_{xy}. *Ein Zustand* $z \in E$ *heißt* **absorbierend**, *falls* $p_{zz} = 1$. *In diesem Fall heißt*

$$\begin{aligned} h_z(x) &:= P^x\big(X_k = z \text{ für alle } k \geq n \text{ und ein } n \in \mathbb{N}\big) \\ &= P^x\big(\bigcup_{n\geq 0}\bigcap_{k\geq n}\{X_k = z\}\big|X_0 = x\big) \end{aligned}$$

die **Absorptionswahrscheinlichkeit** *in* z *beim Start in* $x \in E$.

Theorem 3.21. $(X_n)_{n\in\mathbb{N}_0}$ *sei eine Markoff-Kette mit einem höchstens abzählbaren Zustandsraum* E *und homogenen Übergangswahrscheinlichkeiten* p_{xy}. $z \in E$ *sei ein absorbierender Zustand und* $x \in E$ *sei beliebig. Dann gelten:*

(i) $h_z(x) = \lim_{n\to\infty} P^x\big(X_n = z\big)$,

(ii) h_z *ist die kleinste nichtnegative Funktion mit* $h_z(z) = 1$ *und*

$$h_z(x) = \sum_{y\in E} p_{xy}h_z(y) \qquad \forall x \in E. \tag{48}$$

Beweis. (i) Aus der Stetigkeit von unten folgt

$$h_z(x) = P^x\big(\bigcup_{n\geq 0}\bigcap_{k\geq n}\{X_k = z\}\big) = \lim_{n\to\infty} P^x\big(\underbrace{\bigcap_{k\geq n}\{X_k = z\}}_{=\{X_n=z\}}\big) = \lim_{n\to\infty} P^x\big(\{X_n = z\}\big).$$

(ii) $h_z(x) \geq 0$ und $h_z(z) = 1$ gelten offensichtlich. Weiterhin gilt

$$\begin{aligned} \sum_{y\in E} p_{xy}h_z(y) &= \sum_{y\in E} P^x(X_1 = y)P^x\big(\bigcup_{n\geq 1}\bigcap_{k\geq n}\{X_k = z\}\big|X_1 = y\big) \\ &= P^x\big(\bigcup_{n\geq 1}\bigcap_{k\geq n}\{X_k = z\}\big) \\ &= P^x\big(\bigcup_{n\geq 0}\bigcap_{k\geq n}\{X_k = z\}\big) = h_z(x). \end{aligned}$$

Wir bezeichnen mit $p_{xy}^{(n)}$ die n-Schritt-Übergangswahrscheinlichkeiten. Es sei nun h eine nichtnegative Funktion mit $h(z) = 1$, welche (48) erfüllt. Dann gilt

$$\begin{aligned} h(x) &= \sum_{y\in E} p_{xy} h(y) = \sum_{y_1 \in E} p_{x,y_1} \sum_{y_2\in E} p_{y_1,y_2} h(y_2) \\ &= \sum_{y_2\in E} p_{x,y_2}^{(2)} h(y_2) = \ldots = \sum_{y\in E} p_{xy}^{(n)} h(y) \\ &\geq p_{xz}^{(n)} h(z) = p_{xz}^{(n)} = P^x\big(X_n = z\big). \end{aligned}$$

Mit $n \to \infty$ erhalten wir schließlich

$$h(x) \geq \lim_{n\to\infty} P^x\big(X_n = z\big) = h_z(x),$$

d.h., h_z ist die kleinste nichtnegative Funktion mit $h_z(z) = 1$, welche (48) erfüllt. □

Die zweite Aussage von Theorem 3.21 kann benutzt werden um die Absorptionswahrscheinlichkeiten bei den oben angeführten Beispielen zu bestimmen.

Ruinproblem (Beispiel 4)
Es wird angenommen, dass zwei Spieler wiederholt gegeneinander spielen. Spieler 1 habe ein Anfangskapital von x Euro, Spieler 2 von $N - x$ Euro. Der Gewinner eines Spiels erhält jeweils einen Euro von seinem Gegner. Wir nehmen vereinfachend an, dass die Spielausgänge unabhängig sind und dass jeder Spieler ein jedes einzelne Spiel mit Wahrscheinlichkeit 1/2 gewinnt. Wir bezeichnen mit X_n das Kapital von Spieler 1 nach n Spielen. Das Spiel endet, wenn einer der Spieler sein gesamtes Kapital verspielt hat. In diesem Fall verbleibt der Prozess in dem entsprechenden absorbierenden Zustand. Wir erhalten eine Markoff-Kette $(X_n)_{n\in\mathbb{N}_0}$ mit Zustandsraum $E = \{0, \ldots, N\}$, $X_0 = x$ und homogenen Übergangswahrscheinlichkeiten

$$p_{xy} = P\big(X_{n+1} = y \big| X_n = x\big) = \begin{cases} 1/2, & \text{falls } \ 0 < x < N,\ |x-y| = 1, \\ 1, & \text{falls } \ x = y = 0 \quad \text{oder} \quad x = y = N, \\ 0, & \text{sonst.} \end{cases}$$

(Um weitere Komplikationen zu vermeiden, definieren wir diese Übergangswahrscheinlichkeiten so, auch wenn $P(X_n = x) > 0$ nicht klar ist.)
Gesucht sind nun die Wahrscheinlichkleiten $h_0(x)$ und $h_N(x)$, dass Spieler 1 bzw. 2 verliert. Nach Theorem 3.21 ist $h_0\colon E \to [0,1]$ die kleinste Funktion, welche die Bedingungen $h_0(0) = 1$ sowie

$$h_0(x) = \sum_{y\in E} p_{xy} h_0(y) \qquad \forall x \in E$$

erfüllt. Die triviale Lösung $h_0(x) = 1$ für alle $x \in E$ scheidet von vornherein aus, da der Zustand N ebenfalls absorbierend ist und daher $h_0(N) = 0$ gilt.

Für $0 < x < N$ erhalten wir

$$h_0(x) \;=\; \frac{1}{2}h_0(x-1) \;+\; \frac{1}{2}h_0(x+1),$$

woraus

$$h_0(x) \;-\; h_0(x-1) \;=\; h_0(x+1) \;-\; h_0(x) \qquad \forall x \in \{1,\ldots,N-1\}$$

folgt. Mit den obigen Nebenbedingungen $h_0(0) = 1$ und $h_0(N) = 0$ erhalten wir als einzige Lösung

$$h_0(x) \;=\; \frac{N-x}{N},$$

d.h., Spieler 1 verliert mit einer Wahrscheinlichkeit von $1-x/N$. Mit analogen Rechnungen findet man, dass $h_N(x) = x/N$ ist, d.h., das Spiel endet mit Wahrscheinlichkeit 1 und Spieler 2 verliert mit einer Wahrscheinlichkeit von x/N.

Galton-Watson-Prozess (Beispiel 2)

Es sei $(N_i^{(n)})_{i\in\mathbb{N},\, n\in\mathbb{N}_0}$ ein Dreiecksschema von unabhängigen und identisch verteilten Zufallsvariablen mit Werten in $\mathbb{N}_0$. Falls die n-te Generation aus mindestens i Individuen besteht, so bezeichnet $N_i^{(n)}$ die Anzahl der Nachkommen des i-ten Individuums der n-ten Generation. X_n sei die Anzahl der Individuen der n-ten Generation. Die Entwicklung des Prozesses $(X_n)_{n\in\mathbb{N}_0}$ wird durch die Rekursionsgleichungen

$$X_{n+1} \;=\; \sum_{i=1}^{X_n} N_i^{(n)}$$

beschrieben. Somit erhalten wir folgende Übergangswahrscheinlichkeiten:

$$p_{xy} \;=\; P\big(X_{n+1} = y \big| X_n = x\big) \;=\; P\Big(\sum_{i=1}^{x} N_i^{(n)} = y\Big).$$

Die Verteilung von $N_i^{(n)}$ werde durch den Vektor $\rho \;=\; \big(\rho(k)\big)_{k\in\mathbb{N}_0}$ beschrieben, wobei $\rho(k) \;=\; P\big(N_i^{(n)} \;=\; k\big)$. Die Verteilung von $\sum_{i=1}^{x} N_i^{(n)}$ ist somit gleich der x-fachen Faltung von ρ. (Wir identifizieren der Einfachheit halber die Verteilung mit dem Vektor der Wahrscheinlichkeiten.) Der Zustand 0 ist offensichtlich absorbierend und beschreibt das Aussterben der Population. Gesucht ist jetzt die Wahrscheinlichkeit $h_0(k)$, dass die gesamte Population, welche zum Zeitpunkt 0 aus k Individuen besteht ($X_0 = k$), ausstirbt. Wir werden diese Absorptionswahrscheinlichkeiten mit Hilfe von Theorem 3.21 bestimmen. Dabei werden wir uns auf den nichttrivialen Fall $\rho(0) \in (0,1)$ beschränken, da die Population bei $\rho(0) = 0$ niemals und bei $\rho(0) = 1$ sofort ausstirbt. Zunächst halten wir fest, dass

$$h_0(k) \;=\; \big(h_0(1)\big)^k \qquad \forall k \in \mathbb{N} \tag{49}$$

gilt.

Das ist inhaltlich sofort klar, wenn man anstelle von einer mit k Individuen startenden Population die Entwicklung von k unabhängigen Populationen $((X_n^{(i)})_{n\in\mathbb{N}_0},\ i=1,\dots,k)$ mit $X_0^{(i)}=1$ betrachtet. Dann gilt natürlich wegen der vorausgesetzten Unabhängigkeit, dass

$$\begin{aligned} h_0(k) &= P\big(\text{alle } k \text{ Populationen sterben aus}\big) \\ &= P\big(\text{erste Population stirbt aus}\big)\cdots P\big(k\text{-te Population stirbt aus}\big) \;=\; \big(h_0(1)\big)^k. \end{aligned}$$

Eine formale Begründung von (49) kann folgendermaßen gegeben werden: Es gilt

$$p_{kl} \;=\; P\big(N_1^{(1)}+\cdots+N_k^{(1)}=l\big) \;=\; \sum_{l_1+\cdots+l_k=l} P\big(N_1^{(1)}=l_1,\dots,N_k^{(1)}=l_k\big) \;=\; \sum_{l_1+\cdots+l_k=l} p_{1,l_1}\cdots p_{1,l_k},$$

woraus insbesondere

$$P^k\big(X_1=0\big) \;=\; \Big(P^1\big(X_1=0\big)\Big)^k$$

folgt. Mit vollständiger Induktion bekommen wir

$$\begin{aligned} P^k\big(X_n=0\big) &= \sum_{l=0}^{\infty} P^k\big(X_1=l\big)P\big(X_n=0\big|X_1=l\big) \\ &= \sum_{l=0}^{\infty} p_{kl}\, \underbrace{P^l\big(X_{n-1}=0\big)}_{=\left(P^1(X_{n-1}=0)\right)^l} \\ &= \sum_{l=0}^{\infty} \sum_{l_1+\cdots+l_k=l} p_{1,l_1}\cdots p_{1,l_k}\Big(P^1\big(X_{n-1}=0\big)\Big)^{l_1+\cdots+l_k} \\ &= \Big(\sum_{l=0}^{\infty} p_{1l}\, P^l\big(X_{n-1}=0\big)\Big)^k \\ &= \Big(P^1\big(X_n=0\big)\Big)^k. \end{aligned}$$

Nach (i) von Theorem 3.21 gilt aber nun $P^k\big(X_n=0\big) \underset{n\to\infty}{\longrightarrow} h_0(k)$, woraus (49) folgt.

Zur Bestimmung der Wahrscheinlichkeit des Aussterbens einer Population mit einem Individuum zum Zeitpunkt 0 müssen wir nach (ii) von Theorem 3.21 nach der kleinsten nichtnegativen Funktion h suchen mit $h(0)=1$ und

$$h(k) \;=\; \sum_{l=0}^{\infty} p_{kl}\, h(l) \qquad \forall k\in\mathbb{N}_0.$$

Wegen (49) genügt es, die kleinste nichtnegative Zahl q zu finden mit

$$q^k \;=\; \sum_{l=0}^{\infty} p_{kl}\, q^l \qquad \forall k\in\mathbb{N}_0.$$

Dafür wiederum ist hinreichend, dass

$$q = \sum_{l=0}^{\infty} p_{1l}\, q^l \tag{50}$$

gilt. Wegen $p_{kl} = \sum_{l_1+\cdots+l_k=l} p_{1,l_1} \cdots p_{1,l_k}$ folgt dann

$$\sum_{l=0}^{\infty} p_{kl}\, q^l = \sum_{l=0}^{\infty} \sum_{l_1+\cdots+l_k=l} p_{1,l_1} \cdots p_{1,l_k}\, q^{l_1+\cdots+l_k} = \Big(\sum_{l=0}^{\infty} p_{1l}\, q^l\Big)^k = q^k.$$

Wir definieren eine Funktion f mit

$$f(s) = \sum_{k=0}^{\infty} \rho(k) s^k - s \qquad \forall s \in [0,1]$$

und suchen die kleinste nichtnegative Nullstelle dieser Funktion.

Fall 1: $EN_i^{(n)} \leq 1$

Falls $\rho(k) = 0$ für alle $k \geq 2$, so folgt wegen $\rho(1) = 1 - \rho(0) < 1$

$$\begin{aligned} f(s) &= \rho(0) + \rho(1)s - s \\ &= \rho(0) + \underbrace{\big(\rho(1) - 1\big)}_{<0} s \\ &> \rho(0) + \big(\rho(1) - 1\big) = f(1) = 0 \qquad \forall s \in [0,1). \end{aligned}$$

Falls jedoch $\rho(k_0) > 0$ für ein $k_0 \geq 2$, so folgt wegen der Abschätzung $1 - s^k = k \int_s^1 u^{k-1}\, du < k(1-s)$ für $s \in [0,1)$ und $k \geq 2$, dass

$$\begin{aligned} f(s) &= f(s) - f(1) \\ &= \sum_{k=0}^{\infty} \rho(k)\big(s^k - 1\big) + (1 - s) \\ &= \rho(1)(s - 1) + \underbrace{\sum_{k=2}^{\infty} \rho(k)\big(s^k - 1\big)}_{> \sum_{k=2}^{\infty} k\rho(k)(s-1)} + (1 - s) \\ &> \Big(1 - \sum_{k=1}^{\infty} k\rho(k)\Big)\big(1 - s\big) \geq 0 \qquad \forall s \in [0,1). \end{aligned}$$

In beiden Fällen sehen wir, dass die Funktion f die kleinste nichtnegative Nullstelle im Punkt 1 hat. Demzufolge gilt $h_0(1) = 1$ und somit auch

$$h_0(k) = 1,$$

d.h., unabhängig von der Anzahl der Individuen zum Zeitpunkt 0 stirbt die Population mit Wahrscheinlichkeit 1 aus.

Fall 2: $EN_i^{(n)} > 1$

Wegen $EN_i^{(n)} = \sum_{k=0}^{\infty} k\rho(k) > 1$ existiert ein $K \in \mathbb{N}$ mit

$$\sum_{k=0}^{K} k\rho(k) > 1.$$

Nun gilt für $s \in [0,1)$

$$\begin{aligned} f(s) - f(1) &= \sum_{k=0}^{\infty} \rho(k) \underbrace{\left(s^k - 1\right)}_{<0} + (1-s) \\ &\leq \sum_{k=0}^{K} \rho(k)\left(s^k - 1\right) + (1-s) =: g(s). \end{aligned}$$

Nun gilt $g(1) = 0$ und

$$\frac{d}{ds} g(s)\Big|_{s=1} = \sum_{k=0}^{K} k\rho(k) - 1 > 0.$$

Damit existiert ein $s_0 \in [0,1)$ mit $g(s_0) < 0$, woraus $f(s_0) < f(1)$ folgt. Wegen $f(0) = \rho(0) \geq 0$ folgt jedoch, dass f mindestens eine Nullstelle in $[0,1)$ besitzt. Die kleinste Nullstelle ist dann gleich der gesuchten Aussterbewahrscheinlichkeit $h_0(1)$. Das können wir noch präzisieren: Falls sogar $\rho(0) = 0$, so ist diese wegen $f(0) = \rho(0) = 0$. Falls $\rho(0) > 0$, so gilt $f(0) > 0$ und es folgt, dass $h_0(1) \in (0,1)$.

Übungsaufgabe

ÜA 3.9 Zwei Spieler A und B seien am Beginn im Besitz von a bzw. b Euro. Sie spielen wiederholt ein Spiel, wobei angenommen wird, dass die jeweiligen Spielausgänge stochastisch unabhängig sind und dass Spieler A ein einzelnes Spiel mit einer Wahrscheinlichkeit $p > 0$ gewinnt sowie mit Wahrscheinlichkeit $q > 0$ verliert. Mit einer Wahrscheinlichkeit von $r = 1 - p - q \geq 0$ endet das Spiel unentschieden (z.B. beim Schach). Gewinnt A, so erhält er einen Euro von B, falls er verliert, zahlt er einen Euro an B, bei einem Unentschieden wird nichts gezahlt. Das Spiel ist beendet, wenn einer der Spieler sein Kapital verloren hat. Es sei X_n das aktuelle Kapital (nicht der Gewinn!) von Spieler A zum Zeitpunkt n ($X_0 = a$).

(i) Bestimmen Sie den Zustandsraum E und die Übergangsmatrix $P = (p_{xy})_{x,y \in E}$ der Markoff-Kette $(X_n)_{n \in \mathbb{N}_0}$!

(ii) Welche Zustände sind absorbierend und wie sind die Absorptionswahrscheinlichkeiten beim Start im Zustand a?