Schwache Konvergenz und Martingalprobleme

G. Kersting

18. Dezember 2021

Inhaltsverzeichnis

Kapitel I

Schwache Konvergenz von W-Maßen

Literatur :

Billingsley, P.: *Convergence of Probability Measures*, Wiley, 1968;
Pollard, D.: *Convergence of Stochastic Processes*, Springer, 1986;
Ethier, S. und Kurtz, T.: *Markov Processes, Characterization and Convergence*, Wiley, 1986.

Die Konvergenz der Verteilungen von Zufallsvariablen ist ein klassisches Thema der Stochastik. Altbekannt ist, dass unter geeigneten Umständen binomiale Zufallsvariablen eine Poisson'sche oder normale Grenzverteilung besitzen. Über die Konvergenz in Verteilung von Summen reeller, unabhängiger Zufallsvariabler gibt der zentrale Grenzwertsatz Auskunft. Neueren Datums sind Konvergenzresultate über die Verteilungen ganzer stochastischer Prozesse, etwa der Aussage, dass reelle Irrfahrten bei geeigneter Skalierung in Verteilung gegen eine Brown'sche Bewegung konvergieren (Donsker, 1951).

Die Theorie der schwachen Konvergenz von W-Maßen, die ihre Ursprünge in Arbeiten von Prohorov (1956) und Skorohod (1956) hat, liefert dafür einen Rahmen. Die Theorie der schwachen Konvergenz bewegt sich vollständig im Kontext metrischer Konvergenz. Es geht um die Konvergenz von W-Maßen auf einem separablen, metrischen Raum S. Wir fassen kurz einige Grundannahmen und Bezeichnungen zusammen:

S bezeichnet einen separablen metrischen Raum mit Metrik d. Ohne Einschränkung der Allgemeinheit nehmen wir

$$d(x,y) \le 1 \qquad \text{für alle } x, y \in S$$

an. Man kann nämlich von $d(x,y)$ immer zu der Metrik $\min(1, d(x,y))$ über gehen, ohne etwas an Konvergenzfragen zu ändern. Die offene ε-Umgebung von $x \in S$ ist wie üblich als

$$U_\varepsilon(x) := \{y \in S : d(x,y) < \varepsilon\}$$

definiert. Separabilität bedeutet, dass es eine abzählbare, dichte Teilmenge Q in S gibt. Dies hat die für uns wichtige Konsequenz, dass sich jede offene Menge $O \subset S$ als abzählbare Vereinigung von Umgebungen darstellen lässt, z.B.

$$O \;=\; \bigcup \{U_\varepsilon(x) \;:\; U_\varepsilon(x) \subset O \;,\; \varepsilon \in \mathbb{Q}^+ \;,\; x \in Q\} \quad .$$

Die Borel-σ-Algebra in S ist die σ-Algebra $\mathscr{B} = \mathscr{B}_S$, die von den offenen Mengen in S erzeugt wird. Aufgrund der Separabilität wird $\mathscr{B}$ auch von den offenen ε-Umgebungen erzeugt. Schließlich ist $\mathscr{B}$ auch die von $C_b(S)$, der Menge aller stetigen, beschränkten Abbildungen $f : S \to \mathbb{R}$, erzeugte σ-Algebra. Man beachte dazu, dass die Indikatorfunktion 1_O für offenes $O \subset S$ monotoner Limes der durch

$$f_n(x) := \min\big(1, n \cdot d(x, O^c)\big) \;, \quad x \in S$$

gegebenen Folge $f_n \in C_b(S)$ ist mit $d(x,B) := \inf\{d(x,y) : y \in B\}$. Man erkennt so auch, dass ein W-Maß μ auf $(S,\mathscr{B})$ durch die Werte aller $\int f\,d\mu$, $f \in C_b(S)$ festgelegt ist. Für offenes O gilt nämlich

$$\mu(O) = \lim_n \int f_n\,d\mu \quad ,$$

und die Behauptung folgt aus dem bekannten Eindeutigkeitssatz für W-Maße. Die Menge aller W-Maße auf S bezeichnen wir mit $\mathcal{P}(S)$.

I.1 Grundlagen

Definition.

a) Eine Folge (μ_n) in $\mathcal{P}(S)$ heißt ***schwach konvergent*** *gegen $\mu \in \mathcal{P}(S)$, falls für $n \to \infty$*

$$\int f\,d\mu_n \to \int f\,d\mu \qquad \textit{für alle } f \in C_b(S) \quad .$$

Wir schreiben dann

$$\mu_n \xrightarrow{s} \mu \quad .$$

b) Eine Folge (X_n) von S-wertigen Zufallsvariablen heißt ***in Verteilung konvergent*** *mit Grenzwert μ (gegen die S-wertige Zufallsvariable X), falls die Verteilungen der X_n schwach gegen μ (gegen die Verteilung von X) konvergieren, falls also für alle $f \in C_b(S)$ gilt*

$$\boldsymbol{E}f(X_n) \to \int f\,d\mu \quad \textit{bzw.} \quad \boldsymbol{E}f(X_n) \to \boldsymbol{E}f(X) \quad .$$

Wir schreiben dann

$$X_n \xrightarrow{v} \mu \quad \textit{bzw.} \quad X_n \xrightarrow{v} X \quad .$$

Beispiele.

1. Dirac-Maße δ_{x_n} konvergieren genau dann schwach gegen das Dirac-Maß δ_x, falls x_n in S gegen x konvergieren. (Übung)

2. Haben die W-Maße μ_n und μ Dichten bzgl. des Maßes λ,

$$d\mu_n = g_n\,d\lambda \ , \quad d\mu = g\,d\lambda$$

und konvergiert g_n λ-fast sicher gegen g, dann gilt $\mu_n \xrightarrow{s} \mu$.

Beweis. Aus

$$\begin{aligned}\int |g_n - g|\,d\lambda &= \int g\,d\lambda + \int g_n\,d\lambda - 2\int \min(g_n,g)\,d\lambda \\ &= 2\big(1 - \int \min(g,g_n)\,d\lambda\big)\end{aligned}$$

folgt mittels majorisierter Konvergenz $\int |g_n - g|\,d\lambda \to 0$. Für beschränktes f gilt daher

$$\left|\int f\,d\mu_n - \int f\,d\mu\right| \le \sup|f| \cdot \int |g_n - g|\,d\lambda \to 0 \quad .$$

Stetigkeit für f ist hier gar nicht erforderlich. (Es liegt Konvergenz in der „Totalvariation“ vor.)

3. Fast sichere Konvergenz (oder auch nur stochastische Konvergenz) von Zufallsvariablen X_n gegen X impliziert Konvergenz in Verteilung, denn mittels majorisierter Konvergenz folgt dann

$$\mathbf{E}f(X_n) \to \mathbf{E}f(X) \qquad \text{für alle } f \in C_b(S) \quad .$$

4. **Bildmaße.** Sind $S,\ S'$ metrische Räume und $\pi : S \to S'$ eine stetige Abbildung, dann gilt für $\mu_n,\ \mu \in \mathcal{P}(S)$
$$\mu_n \xrightarrow{s} \mu \quad \Longrightarrow \quad \pi(\mu_n) \xrightarrow{s} \pi(\mu) \quad .$$
Denn für $f \in C_b(S')$ ist $f \circ \pi \in C_b(S)$, also
$$\int f\, d\pi(\mu_n) = \int f \circ \pi\, d\mu_n \to \int f \circ \pi\, d\mu = \int f\, d\pi(\mu) \quad .$$

Für die Konvergenz der Integrale langt es vorauszusetzen (und das wird später wichtig), dass die Menge der Unstetigkeitspunkte von π eine μ-Nullmenge ist. Es gilt nämlich folgende Charakterisierung der schwachen Konvergenz.

Satz 1.1. *Für $\mu, \mu_n \in \mathcal{P}(S)$ ist äquivalent:*

i) $\mu_n \xrightarrow{s} \mu$,

ii) $\mu(O) \le \liminf_n \mu_n(O)$ *für alle offenen $O \subset S$, oder äquivalent*
$\mu(F) \ge \limsup_n \mu_n(F)$ *für alle abgeschlossenen $F \subset S$* ,

iii) $\mu(B) = \lim_n \mu_n(B)$ *für alle $B \in \mathscr{B}$ mit $\mu(\partial B) = 0$* ,

iv) $\int f\, d\mu_n \to \int f\, d\mu$ *für alle beschränkten Borel-Funktionen $f : S \to \mathbb{R}$ mit $\mu(N_f) = 0$.*

Dabei sei $N_f \subset S$ die Menge aller Unstetigkeitspunkte von f.

Beweis.

i) $\Rightarrow$ ii) für offenes O wähle wie oben $f_m \in C_b(S)$, so dass $0 \le f_m \uparrow 1_O$ für $m \to \infty$. Es folgt
$$\mu(O) = \lim_m \int f_m d\mu = \lim_m \lim_n \int f_m d\mu_n \le \liminf_n \mu_n(O) \quad .$$
Die andere Aussage ergibt sich durch Übergang zu Komplementen.

ii) $\Rightarrow$ iii) Ist O der offene Kern und F der Abschluss von B, so gilt
$$\mu(O) \le \liminf_n \mu_n(O) \le \liminf_n \mu_n(B) \le \limsup_n \mu_n(B) \le \limsup_n \mu_n(F) \le \mu(F) \quad .$$
Wegen $\mu(\partial B) = 0$ gilt aber $\mu(O) = \mu(F) = \mu(B)$ also die Behauptung.

iii) $\Rightarrow$ iv) O.B.d.A. sei $f \ge 0$. Es gilt $\partial\{f \ge u\} \subset \{f = u\} \cup N_f$. Die Menge aller $u \in \mathbb{R}$ mit $\mu(\{f = u\}) > 0$ ist höchstens abzählbar. Nach Voraussetzung gilt dies dann auch für $\partial\{f \ge u\}$. Nach Annahme folgt $\mu_n(f \ge u) \to \mu(f \ge u)$ für alle anderen u. Nach dem Satz von der dominierten Konvergenz folgt mit $s = \sup f$
$$\int f\, d\mu_n = \int_0^s \mu_n(f \ge u)\, du \to \int_0^s \mu(f \ge u)\, du = \int f\, d\mu \ ,$$

iv) $\Rightarrow$ i) ist offenbar. □

Der reelle Fall

Im Fall $S = \mathbb{R}$ lassen sich W-Maße μ bekanntlich durch ihre Verteilungsfunktion F, gegeben durch
$$F(x) := \mu\big((-\infty, x]\big), \qquad x \in \mathbb{R}$$
charakterisieren.

Korollar 1.2. *W-Maße μ_n auf $\mathbb{R}$ konvergieren genau dann schwach gegen μ, wenn für die zugehörigen Verteilungsfunktionen F_n, F gilt:*

$$F_n(x) \to F(x) \qquad \text{für alle } x \in \mathbb{R} \text{ mit } \mu(\{x\}) = 0 \quad .$$

Beweis. Die Notwendigkeit der Bedingung ergibt sich aus Satz 1.1 iii). Umgekehrt ist jede offene Menge O in $\mathbb{R}$ abzählbare Vereinigung von disjunkten offenen Intervallen: $O = \bigcup_{i=1}^{\infty} (a_i, b_i)$. Sei $m \in \mathbb{N}$ und $\varepsilon > 0$ derart, dass $\mu(\{a_i + \varepsilon\}) = \mu(\{b_i - \varepsilon\}) = 0$ für alle $i = 1, \dots, m$, so folgt

$$\begin{aligned}\liminf_n \mu_n(O) &\geq \liminf_n \mu_n\Big(\bigcup_{i=1}^m (a_i+\varepsilon, b_i-\varepsilon]\Big) = \liminf_n \sum_{i=1}^m \big(F_n(b_i-\varepsilon) - F_n(a_i+\varepsilon)\big) \\ &= \sum_{i=1}^m \big(F(b_i-\varepsilon) - F(a_i+\varepsilon)\big) = \mu\Big(\bigcup_{i=1}^m (a_i+\varepsilon, b_i-\varepsilon]\Big) \quad .\end{aligned}$$

Da es nur abzählbar viele $x \in \mathbb{R}$ mit $\mu(\{x\}) > 0$ gibt, sind in dieser Überlegung höchstens abzählbar viele $\varepsilon > 0$ ausgeschlossen. Wir können also den Grenzübergang $\varepsilon \to 0$ durchführen und erhalten $\liminf_n \mu_n(O) \geq \mu\big(\bigcup_{i=1}^m (a_i, b_i)\big)$. Lassen wir noch $m \to \infty$ gehen, so folgt $\liminf_n \mu_n(O) \geq \mu(O)$, also nach Satz 1.1 ii) die Behauptung. □

Produkträume

Seien nun S' und S'' mit den Metriken d' bzw. d'' versehen. Dann wird $S = S' \times S''$ durch

$$d\big((x', x''), (y', y'')\big) := \max\big(d'(x', y'), d(x'', y'')\big) \ , \qquad x', y' \in S',\ x'', y'' \in S''$$

zu einem metrischen Raum, die zugehörige ε-Umgebung von (x', x'') ist

$$U_\varepsilon(x', x'') = U'_\varepsilon(x') \times U''_\varepsilon(x'') \quad .$$

(U'_ε und U''_ε bezeichnen dabei die ε-Umgebungen in S' und S''.) Mit S' und S'' ist auch S separabel, man kann die dichte Teilmenge von der Gestalt $Q = Q' \times Q''$ wählen. Es gilt $\mathscr{B}_S = \mathscr{B}_{S'} \otimes \mathscr{B}_{S''}$.

Wir zeigen nun, dass die Funktionen der Gestalt $f(x', x'') = f'(x') \cdot f''(x'')$ in $C_b(S)$ eine sogenannte **konvergenzbestimmende Teilmenge** von $C_b(S)$ bilden.

Satz 1.3. *Für W-Maße μ_n, μ auf dem Produktraum $S = S' \times S''$ gilt $\mu_n \xrightarrow{s} \mu$ genau dann, wenn für $n \to \infty$*

$$\int f \, d\mu_n \to \int f \, d\mu$$

für alle $f \in C_b(S)$ mit $f(x', x'') = f'(x') f''(x'')$ und $f' \in C_b(S'), f'' \in C_b(S'')$.

Beweis. Wir betrachten erst ‚Rechtecke' $B' \times B''$ mit $B' \in \mathscr{B}', B'' \in \mathscr{B}''$ und offenem Kern $O' \times O''$. Wähle $f'_m \in C_b(S'), f''_m \in C_b(S'')$ so dass $0 \leq f'_m \uparrow 1_{O'}, 0 \leq f''_m \uparrow 1_{O''}$. Nach Annahme folgt für alle m

$$\liminf_n \mu_n(B' \times B'') \geq \lim_n \int f'_m f''_m \, d\mu_n = \int f'_m f''_m \, d\mu$$

Unter der zusätzlichen Annahme $\mu(\delta(B' \times B'')) = 0$ ergibt der Grenzübergang $m \to \infty$

$$\liminf_n \mu_n(B' \times B'') \geq \mu(O' \times O'') = \mu(B' \times B'') \quad .$$

Es folgt

$$\liminf_n \mu_n\Big(\bigcup_{i=1}^k B'_i \times B''_i\Big) \geq \mu\Big(\bigcup_{i=1}^k B'_i \times B''_i\Big) \quad ,$$

sofern $\mu(\delta B_i' \times S'') = \mu(S' \times \delta B_i'') = 0$ für alle $i = 1, \ldots, k$. Jede Vereinigung $\bigcup_{i \le k} B_i' \times B_i''$ läßt sich nämlich durch den Übergang zu Durchschnitten der messbaren Rechtecke in eine disjunkte Vereinigung von messbaren Rechtecken umformen, deren Ränder in $\bigcup_{i \le k} \delta B_i' \times S'' \cup S' \times \delta B_i''$ enthalten sind.

Sei nun $O \subset S$ offen. Dann ist O abzählbare Vereinigung aller Umgebungen $U_\varepsilon'(x') \times U_\varepsilon''(x'') \subset O$ mit $x' \in Q', x'' \in Q''$ und $\varepsilon \in D$, wobei D eine beliebige abzählbare, dichte Teilmenge von $\mathbb{R}_+$ sei. Dabei können wir für diese Umgebungen immer $\mu(\delta U_\varepsilon'(x') \times S'' \cup S' \times \delta U_\varepsilon''(x'')) = 0$ voraussetzen, denn für jedes Paar (x', x'') gibt es höchstens abzählbar viele Ausnahmen. Zu vorgegebenem $\eta > 0$ gibt innerhalb O endlich viele ε_i-Umgebungen, so dass

$$\mu(O) \le \mu\left(\bigcup_{i=1}^k U_{\varepsilon_i}'(x_i') \times U_{\varepsilon_i}''(x_i'')\right) + \eta \quad .$$

Es folgt

$$\liminf_n \mu_n(O) \ge \liminf_n \mu_n\Big(\bigcup_{i=1}^k U_{\varepsilon_i}'(x_i') \times U_{\varepsilon_i}''(x_i'')\Big) \ge \mu\Big(\bigcup_{i=1}^k U_{\varepsilon_i}'(x_i') \times U_{\varepsilon_i}''(x_i'')\Big) \ge \mu(O) - \eta \quad .$$

Mit $\eta \to 0$ folgt nach Satz 1.1 ii) die Behauptung. □

Korollar 1.4. *Gilt $\mu_n' \xrightarrow{s} \mu'$ in S' und $\mu_n'' \xrightarrow{s} \mu''$ in S'', so folgt $\mu_n' \otimes \mu_n'' \xrightarrow{s} \mu' \otimes \mu''$ in $S' \times S''$.*

Nach dem Satz von Fubini gilt dann nämlich

$$\begin{aligned} &\int f'(x')f''(x'')\, d\mu_n' \otimes \mu_n'' \\ &\quad = \int f'\, d\mu_n' \int f''\, d\mu_n'' \to \int f' d\mu' \int f''\, d\mu'' = \int f'(x')f''(x'')\, d\mu' \otimes \mu'' \quad . \end{aligned}$$

Analog behandelt man schwache Konvergenz auf kartesischen Produkten mit mehr als zwei Faktoren.

I.2 $\mathcal{P}(S)$ als metrischer Raum

Die schwache Konvergenz läßt sich metrisieren.

Definition. *Der **Prohorov-Abstand** von $\mu, \nu \in \mathcal{P}(S)$ ist gegeben durch*

$$d_P(\mu, \nu) := \inf\{\varepsilon > 0 : \ \mu(F) \le \nu(F^\varepsilon) + \varepsilon \text{ für alle abgeschlossenen } F \subset S\} \quad .$$

Dabei sei F^ε die offene ε-Umgebung von F,

$$F^\varepsilon := \{x \in S : \ \text{es gibt ein } y \in F \text{ mit } d(x,y) < \varepsilon\} \quad .$$

Beispiel. Für $x, y \in S$ gilt $d_P(\delta_x, \delta_y) = d(x,y)$ (Übung, dabei ist die Annahme $d(x,y) \le 1$ zu beachten.) Man kann also den Prohorov-Abstand als Fortsetzung der Metrik d auf ganz $\mathcal{P}(S)$ auffassen.

Satz 1.5. *Der Prohorov-Abstand ist eine Metrik, durch die $\mathcal{P}(S)$ zu einem separablen metrischen Raum wird. Es gilt $d_P(\mu_n, \mu) \to 0$ genau dann, falls $\mu_n \xrightarrow{s} \mu$.*

Beweis. d_P ist symmetrisch: Sei $\varepsilon > d_P(\mu,\nu)$ und F abgeschlossen. Da das Komplement von F^ε ebenfalls abgeschlossen ist, folgt

$$\mu\big((F^\varepsilon)^c\big) \leq \nu\Big(\big((F^\varepsilon)^c\big)^\varepsilon\Big) + \varepsilon \quad .$$

Wegen $\big((F^\varepsilon)^c\big)^\varepsilon \subset F^c$, ergibt sich durch Übergang zu Komplementen $\nu(F) \leq \mu(F^\varepsilon) + \varepsilon$, also $d_P(\nu,\mu) \leq \varepsilon$. Daher gilt $d_P(\nu,\mu) \leq d_P(\mu,\nu)$, und analog $d_P(\mu,\nu) \leq d_P(\nu,\mu)$.

Gilt $d_P(\mu,\nu) = 0$, so folgt für abgeschlossenes F

$$\mu(F) \leq \inf_{\varepsilon>0}(\nu(F^\varepsilon) + \varepsilon) = \nu(F)$$

und genauso $\nu(F) \leq \mu(F)$. μ und ν stimmen also auf den abgeschlossenen Mengen überein, und es folgt $\mu = \nu$ nach dem Eindeutigkeitssatz für Maße.

Zur Δ-Ungleichung: Ist $\varepsilon > d_P(\mu,\nu)$, $\delta > d_P(\nu,\rho)$ und F abgeschlossen, so gilt

$$\mu(F) \leq \nu(F^\varepsilon) + \varepsilon \leq \rho\left(\overline{F^\varepsilon}^\delta\right) + \varepsilon + \delta \leq \rho\left(F^{\varepsilon+\delta}\right) + \varepsilon + \delta \quad .$$

Daher folgt $d_P(\mu,\rho) \leq \varepsilon + \delta$, also $d_P(\mu,\rho) \leq d_P(\mu,\nu) + d_P(\nu,\rho)$.

d_P ist also eine Metrik. Wir zeigen nun, dass die abzählbare Menge der W-Maße $\nu = \sum_{j=1}^r \alpha_j \cdot \delta_{x_j}$ mit $x_1,\ldots,x_r \in Q$ und $\alpha_1,\ldots,\alpha_r \in \mathbb{Q}^+$, dicht in $\mathcal{P}(S)$ ist.

Sei $\mu \in \mathcal{P}(S)$ und $\varepsilon > 0$. Da Q dicht in S ist, gilt $S = \bigcup_{x\in Q} U_\varepsilon(x)$. Es existieren also $x_1,\ldots,x_r \in Q$, so dass

$$\mu\Big(\bigcup_{i=1}^r U_\varepsilon(x_i)\Big) > 1 - \varepsilon \quad .$$

Für $B_i := \bigcup_{j=1}^i U_\varepsilon(x_j) - \bigcup_{j=1}^{i-1} U_\varepsilon(x_j)$ folgt $B_i \subset U_\varepsilon(x_i)$, $\bigcup_i B_i = \bigcup_i U_\varepsilon(x_i)$ und $\sum_i \mu(B_i) > 1-\varepsilon$. Dann lassen sich $\alpha_i \in \mathbb{Q}^+$ mit $\alpha_1 + \cdots + \alpha_r = 1$ finden, so dass $\sum_{i=1}^r |\mu(B_i) - \alpha_i| \leq \varepsilon$. Für $\nu := \sum_i \alpha_i \cdot \delta_{x_i}$ folgt

$$\nu(F) = \sum_{x_i\in F} \alpha_i \leq \sum_{x_i\in F} \mu(B_i) + \varepsilon = \mu\Big(\bigcup_{x_i\in F} B_i\Big) + \varepsilon \leq \mu\Big(\bigcup_{x_i\in F} U_\varepsilon(x_i)\Big) + \varepsilon \leq \mu(F^\varepsilon) + \varepsilon \quad .$$

Dies bedeutet wie gewünscht $d_P(\nu,\mu) \leq \varepsilon$.

Zur Konvergenz : Gilt $d_P(\mu_n,\mu) \to 0$, so gibt es eine Nullfolge ε_n so dass $\mu_n(F) \leq \mu(F^{\varepsilon_n}) + \varepsilon_n$ für abgeschlossenes F. Es folgt $\limsup \mu_n(F) \leq \mu(F)$, also $\mu_n \xrightarrow{s} \mu$ nach Satz 1.1 ii). Sei umgekehrt μ_n schwach gegen μ konvergent und $\varepsilon > 0$. Wir wollen zeigen, dass $d_P(\mu_n,\mu) \leq 2\varepsilon$ für ausreichend großes n gilt. Wir wählen $U_\varepsilon(x_1),\ldots,U_\varepsilon(x_r)$ wie eben. Dann gibt es nach Satz 1.1 ii) ein m, so dass für $n \geq m$ und jede Teilmenge I von $\{1,\ldots,r\}$

$$\mu_n\Big(\bigcup_{i\in I} \overline{U_\varepsilon(x_i)}\Big) \leq \mu\Big(\bigcup_{i\in I} \overline{U_\varepsilon(x_i)}\Big) + \varepsilon \quad ,$$

sowie

$$\mu_n\Big(\bigcup_{i=1}^r U_\varepsilon(x_i)\Big) \geq 1 - \varepsilon \quad .$$

Setzen wir $I_F = \{i \leq r : U_\varepsilon(x_i) \cap F \neq \emptyset\}$ für abgeschlossenes $F \subset S$, so folgt für $n \geq m$

$$\mu_n(F) \leq \mu_n\Big(\bigcup_{i\in I_F} \overline{U_\varepsilon(x_i)}\Big) + \mu_n\Big(S - \bigcup_{i=1}^r U_\varepsilon(x_i)\Big) \leq \mu\Big(\bigcup_{i\in I_F} \overline{U_\varepsilon(x_i)}\Big) + 2\varepsilon \leq \mu(F^\eta) + \eta$$

für alle $\eta > 2\varepsilon$, und es folgt wie behauptet $d_P(\mu_n,\mu) \leq 2\varepsilon$. □

$\mathcal{P}(S)$ wird damit, versehen mit der zugehörigen Borel-σ-Algebra $\mathscr{B}_{\mathcal{P}}$, zu einem vollwertigen Wertebereich für Zufallsvariablen. Dazu ein Beispiel.

Beispiel. Das Glivenko-Cantelli-Theorem. Seien $Z_1, Z_2, \ldots$ unabhängige, identisch verteilte Zufallsvariablen mit Werten in S und Verteilung μ. Wir bilden die *empirischen Verteilungen*

$$L_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i} \quad .$$

Es gilt $L_n = \varphi_n(Z_1, \ldots, Z_n)$ mit $\varphi_n(z_1, \ldots, z_n) = \frac{1}{n}(\delta_{z_1} + \cdots + \delta_{z_n})$. Man überzeuge sich, dass $\varphi_n : S^n \to \mathcal{P}(S)$ eine stetige Abbildung ist. L_n ist also eine wohldefinierte $\mathcal{P}(S)$-wertige Zufallsvariable.

Behauptung. *Es gilt $d_P(L_n, \mu) \to 0$ fast sicher für $n \to \infty$.*

Beweis. Wir betrachten erneut die Umgebungen $U_\varepsilon(x_i)$ aus dem letzten Beweis. Für Borel-Mengen $B \subset S$ gilt

$$L_n(B) = \frac{1}{n} \sum_{i=1}^{n} I_{\{Z_i \in B\}} \quad ,$$

nach dem starken Gesetz der großen Zahlen folgt also $L_n(B) \to \mu(B)$ fast sicher. Daher gibt es fast sicher ein (zufälliges) $n_\varepsilon \in \mathbb{N}$, so dass für alle $n \geq n_\varepsilon$ und $I \subset \{1, \ldots, r\}$

$$L_n\Big(\bigcup_{i \in I} \overline{U_\varepsilon(x_i)}\Big) \leq \mu\Big(\bigcup_{i \in I} \overline{U_\varepsilon(x_i)}\Big) + \varepsilon \quad \text{sowie} \quad L_n\Big(\bigcup_{i=1}^{r} U_\varepsilon(x_i)\Big) \geq 1 - \varepsilon \quad .$$

Wie oben impliziert dies $d_P(L_n, \mu) \leq 2\varepsilon$ für $n \geq n_\varepsilon$, und es folgt die Behauptung. □

I.3 Relative Kompaktheit in $\mathcal{P}(S)$

Vorbereitend behandeln wir Kompaktheit in metrischen Räumen. Eine Borelsche Teilmenge K eines metrischen Raumes S heißt *relativ kompakt*, falls der topologische Abschluss $\overline{K}$ von K kompakt ist, falls also jede offene Überdeckung von $\overline{K}$ eine endliche Überdeckung enthält.

Proposition. Eine Teilmenge K in einem metrischen Raum S ist genau dann relativ kompakt, wenn jede Folge (x_n) in K eine konvergente Teilfolge besitzt.

Beweis. „$\Rightarrow$“: Sei (x_n) eine Folge in K. Wir nehmen an, dass sie keine Häufungspunkte hat. Dann gibt es zu jedem $y \in S$ eine Umgebung U_y von y, die nur endliche viele Folgenglieder enthält. Diese Umgebungen überdecken ganz S, daher überdecken wegen der Kompaktheit von $\overline{K}$ schon endlich viele dieser Umgebungen K. Also besteht die gesamte Folge nur aus endlich vielen Gliedern, ein Widerspruch. Es gibt also ein Häufungspunkt x der Folge, d.h. jede Umgebung $U_\varepsilon(x)$ enthält unendlich viele Folgenglieder. Ist daher $\varepsilon_1 > \varepsilon_2 > \ldots > 0$ eine Nullfolge, so gibt es eine Teilfolge $1 \leq n_1 < n_2 < \cdots$, so dass $x_{n_k} \in U_{\varepsilon_k}(x)$. Es folgt $x_{n_k} \to x$.

„$\Leftarrow$“: Sei $(O_i)_{i \in I}$ eine offene Überdeckung von $\overline{K}$. Wir zeigen zuerst, dass es ein $\varepsilon > 0$ gibt, so dass für jedes $x \in K$ ein $i \in I$ gibt mit $U_\varepsilon(x) \subset O_i$. Andernfalls gibt es eine Folge (x_n) in K mit der Eigenschaft, dass $U_{1/n}(x_n) \not\subset U_i$ für alle $i \in I$. Nach Annahme gibt es eine konvergente Teilfolge x_{n_k} mit Grenzwert $x \in \overline{K}$. Folglich gibt es ein $\eta > 0$ und ein $i \in I$, so dass $U_\eta(x) \subset O_i$. Für ausreichend großes k gilt nun $U_{1/n_k}(x_{n_k}) \subset U_\eta(x)$ und damit $U_{1/n_k}(x_{n_k}) \subset U_i$, ein Widerspruch.

Wir konstruieren nun aus (O_i) eine endliche Überdeckung von $\overline{K}$. Zunächst wählen wir $x_1 \in K$ und $i_1 \in I$ so dass $U_\varepsilon(x_1) \subset O_{i_1}$. Seien nun $x_1, \ldots, x_n \in K$ und $i_1, \ldots, i_n \in I$ schon gewählt, mit $U_\varepsilon(x_i) \subset O_i$ und $d(x_i, x_j) \geq \varepsilon/2$ für alle $i \neq j$. Falls $\overline{K} \subset O_1 \cup \cdots \cup O_n$, so ist die endliche Überdeckung gefunden. Andernfalls gibt es ein $x \in \overline{K}$ mit $x \notin O_1 \cup \cdots \cup O_n$. Es folgt $d(x, x_i) \geq \varepsilon$ für $i \leq n$. Wähle nun $x_{n+1} \in K$ mit $d(x_{n+1}, x) < \varepsilon/2$ und $i_{n+1} \in I$ mit $U_\varepsilon(x_{n+1}) \subset O_{i_{n+1}}$. Dann gilt auch $d(x_i, x_{n+1}) \geq \varepsilon/2$. Diese Konstruktion muss nach endlich vielen Schritten abbrechen und zu einer endlicher Überdeckung führen. Andernfalls entsteht nämlich eine Folge $x_1, x_2, \ldots$ in K mit $d(x_i, x_j) \geq \varepsilon/2$, die also keine konvergente Teilfolge enthält. Dies ist ein Widerspruch. □

Wir gehen nun der für viele Konvergenzbeweise wichtigen Frage nach, wann man aus einer Folge von W-Maßen eine schwach konvergente Teilfolge aussondern kann.

Satz 1.6. (Prohorov) *Eine Menge $\mathcal{K} \subset \mathcal{P}(S)$ ist relativ kompakt, falls für alle $\varepsilon > 0$ ein relativ kompaktes $K \subset S$ existiert, so dass*

$$\sup_{\mu \in \mathcal{K}} \mu(K^c) \leq \varepsilon \quad .$$

Man nennt Mengen $\mathcal{K}$ von W-Maßen, die dieser Bedingung genügen, auch *straff (tight)*.

Beweis. Sei (μ_n) eine Folge in $\mathcal{K}$. Wir wollen aus ihr eine konvergente Teilfolge aussondern. Dazu wählen wir nach Annahme relativ kompakte Teilmengen $K_1 \subset K_2 \subset \ldots$, so dass $\mu_n(K_j^c) \leq 1/j$ für alle n und j, und bilden das System kompakter Mengen

$$\mathcal{H} := \Big\{ \bigcup_{i=1}^{r} \overline{K}_{j_i} \cap \overline{U_{\varepsilon_i}(x_i)} \ : \ r, j_i \in \mathbb{N}, \ \varepsilon_i \in \mathbb{Q}^+, \ x_i \in Q \Big\} \quad .$$

$\mathcal{H}$ ist abzählbar, wir können daher per Diagonalverfahren eine Teilfolge $(\mu_{n'})$ von (μ_n) auswählen, so dass der Limes

$$\alpha(H) = \lim_{n'} \mu_{n'}(H)$$

für alle $H \in \mathcal{H}$ existiert. Die Additivität und Subadditivität von $\mu_{n'}$ überträgt sich im Grenzübergang auf α (i.Allg. jedoch nicht die σ-Additivität). In Anbetracht von Satz 1.1 kann man nicht erwarten, dass α auf $\mathcal{H}$ bereits das gesuchte Grenzmaß ist, dies erfordert eine weitere Approximationsschritt. Dazu setzen wir für offenes $O \subset S$

$$\mu^*(O) := \sup\{\alpha(H) : H \in \mathcal{H}, \ H \subset O\} \quad ,$$

insbesondere $\mu^*(\emptyset) = 0$, und für beliebiges $B \subset S$

$$\mu^*(B) := \inf\{\mu^*(O) : O \text{ offen}, \ O \supset B\} \quad .$$

Wir zeigen, dass μ^* ein äußeres Maß ist. Nur die für äußere Maße geforderte Subadditivität ist hier nicht offenbar. Zunächst zeigen wir sie für offene Mengen.

Seien also $O_1, O_2, \ldots$ offene Mengen und sei $H \in \mathcal{H}$ derart, dass $H \subset \bigcup_{n=1}^{\infty} O_n$. Es langt zu zeigen, dass es Mengen $H_1, \ldots, H_m \in \mathcal{H}$ mit $H \subset H_1 \cup \cdots \cup H_m$ und $H_n \subset O_n$, $n = 1, \ldots, m$ gibt. Dann folgt

$$\alpha(H) \leq \alpha(H_1) + \cdots + \alpha(H_m) \leq \mu^*(O_1) + \cdots + \mu^*(O_m) \leq \sum_{n=1}^{\infty} \mu^*(O_n)$$

und nach Definition von μ^*

$$\mu^*\Big(\bigcup_{n=1}^{\infty} O_n\Big) \leq \sum_{n=1}^{\infty} \mu^*(O_n) \quad .$$

Zur Konstruktion der $H_1, \ldots, H_m$ bemerken wir: H wird überdeckt von all den Umgebungen $U_\varepsilon(x)$, $\varepsilon \in \mathbb{Q}^+$, $x \in Q$, für die $\overline{U_\varepsilon(x)}$ in einem O_n enthalten ist. H ist kompakt, daher überdecken bereits endlich viele $U_{\varepsilon_1}(x_1), \ldots, U_{\varepsilon_r}(x_r)$ dieser Umgebungen H. Wir setzen

$$H_n := \overline{K}_j \cap \bigcup_{i=1}^{r} \big\{ \overline{U_{\varepsilon_i}(x_i)} : \overline{U_{\varepsilon_i}(x_i)} \subset O_n \big\} \ , \quad n \geq 1 \quad ,$$

und wählen dabei j so, dass $H \subset \overline{K}_j$. Dann ist $H_n \in \mathcal{H}$, $H_n \subset O_n$ und $H \subset H_1 \cup \ldots \cup H_m$ für ein $m \geq 1$.

Für beliebige $B_n \subset S$ gibt es offene $O_n \subset B_n$, so dass $\mu^*(O_n) \leq \mu^*(B_n) + \varepsilon 2^{-n}$. Es folgt

$$\mu^*\Big(\bigcup_{n=1}^{\infty} B_n\Big) \leq \mu^*\Big(\bigcup_{n=1}^{\infty} O_n\Big) \leq \sum_{n=1}^{\infty} \mu^*(O_n) \leq \sum_{n=1}^{\infty} \mu^*(B_n) + \varepsilon \quad .$$

Mit $\varepsilon \to 0$ erkennen wir, dass μ^* in der Tat äußeres Maß ist.

Als nächstes zeigen wir, dass jedes offene $V \subset S$ in der σ-Algebra $\mathfrak{A}^*$ der μ^*-messbaren Mengen liegt. Sei $B \subset S$ beliebig und O offene Obermenge von B mit $\mu^*(O) \leq \mu^*(B) + \varepsilon$. Dann gibt es ein $H_1 \in \mathcal{H}$ mit $H_1 \subset O \cap V$ und $\mu^*(O \cap V) \leq \alpha(H_1) + \varepsilon$ und ein $H_2 \in \mathcal{H}$ mit $H_2 \subset O \cap H_1^c$ und $\mu^*(O \cap H_1^c) \leq \alpha(H_2) + \varepsilon$. H_1 und H_2 sind disjunkt, $O \cap V^c \subset O \cap H_1^c$ und $H_1 \cup H_2 \subset O$, daher folgt

$$\begin{aligned}\mu^*(B) + \varepsilon \geq \mu^*(O) &\geq \alpha(H_1 \cup H_2) \geq \alpha(H_1) + \alpha(H_2) \\ &\geq \mu^*(O \cap V) + \mu^*(O \cap V^c) - 2\varepsilon \geq \mu^*(B \cap V) + \mu^*(B \cap V^c) - 2\varepsilon \geq \mu^*(B) - 2\varepsilon \quad .\end{aligned}$$

Mit $\varepsilon \to 0$ folgt $\mu^*(B) = \mu^*(B \cap V) + \mu^*(B \cap V^c)$, V ist also, wie behauptet, Element von $\mathfrak{A}^*$.

Die σ-Algebra $\mathfrak{A}^*$ umfasst daher die Borel-σ-Algebra $\mathscr{B}$, und nach dem Satz von Caratheodory ist die Einschränkung von μ^* auf $\mathscr{B}$ ein Maß. Wir bezeichnen sie mit μ. Nach Konstruktion gilt $\mu(S) \leq 1$. μ ist sogar W-Maß: die K_j gehören alle zu $\mathcal{H}$ (jede kompakte Menge wird von endlich vielen ε-Umgbungen überdeckt), und damit folgt

$$\mu(S) \geq \alpha(K_j) = \lim_{n'} \mu_{n'}(K_j) \geq 1 - \frac{1}{j}$$

für alle $j \in \mathbb{N}$. Schließlich ist μ der schwache Limes von $(\mu_{n'})$, denn für offenes O, $H \in \mathcal{H}$, $H \subset O$ gilt

$$\alpha(H) = \lim_{n'} \mu_{n'}(H) \leq \liminf_{n'} \mu_{n'}(O) \quad ,$$

also $\mu(O) \leq \liminf_{n'} \mu_{n'}(O)$. □

In vollständigen metrischen Räumen ist Straffheit auch eine notwendige Bedingung. Zum Beweis benötigen wir folgende Charakterisierung relativ kompakter Mengen.

Proposition 1.7. *Sei der metrische Raum S vollständig. Dann ist $K \subset S$ genau dann relativ kompakt, wenn K total beschränkt ist, d.h. wenn für alle $\varepsilon > 0$ Elemente $x_1, \ldots, x_k$ von S existieren, so dass*

$$K \subset \bigcup_{i=1}^{k} U_\varepsilon(x_i).$$

Beweis. Sei K totalbeschränkt und $(y_n)_n$ Folge in K. Wir konstruieren eine Teilfolge, die Cauchy-Folge ist und also nach Annahme konvergiert. Dazu wählen wir eine Nullfolge $(\varepsilon_i)_i$ positiver Zahlen. Nach Annahme gibt es endlich viele ε_1-Umgebungen, die K überdecken. Eine dieser Umgebungen enthält dann eine unendliche Teilfolge $(y_{n1})_n$ von $(y_n)_n$, folglich erhalten wir für diese Teilfolge $d(y_{i1}, y_{j1}) \leq 2\varepsilon_1$ für alle i, j. Weiter gibt es endlich viele ε_2-Umgebungen, die K überdecken, von denen eine wiederum eine unendliche Teilfolge $(y_{n2})_n$ von $(y_{n1})_n$ enthält. Sie erfüllt also $d(y_{i2}, y_{j2}) \leq 2\varepsilon_2$ für alle i, j. Diese Prozedur lässt sich ad infinitum fortsetzen und dann die Diagonalfolge $(z_n)_n = (y_{nn})_n$ bilden. Nach Konstruktion ist $z_m, z_{m+1}, \ldots$ Teilfolge von $(y_{nm})_n$, daher gilt $d(z_m, z_n) \leq 2\varepsilon_m$ für alle $m < n$. Folglich handelt es sich um eine Cauchy-Folge. – Umgekehrt bilden alle ϵ-Umgebungen eine offene Überdeckung von K. Ist K relativ kompakt, so enthält die Überdeckung bekanntlich eine endliche Überdeckung. Folglich ist dann K total beschränkt. □

Satz 1.8. *Sei S vollständig und $\mathcal{K} \subset \mathcal{P}(S)$. Dann ist äquivalent:*

i) $\mathcal{K}$ ist relativ kompakt;

ii) Zu jedem $\varepsilon > 0$ gibt es endlich viele $x_1, \ldots, x_r \in S$, so dass für alle $\mu \in \mathcal{K}$ gilt

$$\mu\Big(\bigcup_{i=1}^{r} U_\varepsilon(x_i)\Big) \geq 1 - \varepsilon$$

iii) Zu jedem $\varepsilon > 0$ gibt es ein relativ kompaktes $K \subset S$, so dass für alle $\mu \in \mathcal{K}$ gilt

$$\mu(K) \geq 1 - \varepsilon \quad .$$

Beweis.

i) $\Rightarrow$ ii) Wir wählen $x_1, x_2, \ldots$ als dichte Folge in S. Falls ii) nicht zuträfe, gäbe es ein $\varepsilon > 0$ und zu jedem $n \in \mathbb{N}$ ein $\mu_n \in \mathcal{K}$ mit $\mu_n\big(\bigcup_{i=1}^n U_\varepsilon(x_i)\big) < 1 - \varepsilon$. Da K relativ kompakt ist, enthält dann (μ_n) eine gegen ein W-Maß μ schwach konvergente Teilfolge $(\mu_{n'})$. Es folgt für alle r

$$\mu\Big(\bigcup_{i=1}^r U_\varepsilon(x_i)\Big) \le \liminf \mu_{n'}\Big(\bigcup_{i=1}^r U_\varepsilon(x_i)\Big) \le 1 - \varepsilon$$

und mit $r \to \infty$ die Ungleichung $\mu(S) \le 1 - \varepsilon$, ein Widerspruch.

ii) $\Rightarrow$ iii) Sei $\varepsilon > 0$. Wir wählen $\varepsilon_j > 0$, so dass $\varepsilon = \varepsilon_1 + \varepsilon_2 + \cdots$. Es gibt dann $x_{j1}, \ldots, x_{jr_j}$, so dass für alle $\mu \in \mathcal{K}$

$$\mu\Big(\bigcup_{i=1}^{r_j} U_{\varepsilon_j}(x_{ji})\Big) \ge 1 - \varepsilon_j \quad .$$

Dann gilt für

$$K := \bigcap_{j=1}^{\infty} \bigcup_{i=1}^{r_j} U_{\varepsilon_j}(x_{ji}) \quad .$$

$\mu(K^c) \le \varepsilon_1 + \varepsilon_2 + \cdots = \varepsilon$ für alle $\mu \in \mathcal{K}$. Außerdem ist K total beschränkt und deswegen nach Proposition 1.7 relativ kompakt.

iii) $\Rightarrow$ i) Dies ist der Inhalt von Satz 1.6. □

Korollar 1.9. *$\mathcal{P}(S)$ ist vollständig, sofern dies für S zutrifft.*

Beweis. Sei (μ_n) Cauchy-Folge in $\mathcal{P}(S)$. Für jedes $\varepsilon > 0$ gibt es dann ein m, so dass für $n \ge m$ die Ungleichung $d_P(\mu_m, \mu_n) < \varepsilon/2$ gilt. Es folgt

$$\mu_m\Big(\bigcup_{i=1}^r \overline{U_{\varepsilon/2}(x_i)}\Big) \le \mu_n\Big(\bigcup_{i=1}^r U_\varepsilon(x_i)\Big) + \frac{\varepsilon}{2} \quad ,$$

wobei wir über r und $x_1, x_2, \ldots$ noch beliebig verfügen können. Wir wählen (x_n) als dichte Folge in S und r so groß, dass

$$\mu_m\Big(\bigcup_{i=1}^r \overline{U_{\varepsilon/2}(x_i)}\Big) \ge 1 - \frac{\varepsilon}{2} \quad ,$$

so dass

$$\mu_n\Big(\bigcup_{i=1}^r U_\varepsilon(x_i)\Big) \ge 1 - \varepsilon$$

für alle $n \ge m$. Indem wir r weiter vergrößern, erreichen wir, dass diese Ungleichung für alle $n \ge 1$ gilt. Nach Satz 1.8 enthält damit (μ_n) eine konvergente Teilfolge. Deren Limes ist dann Grenzwert der Gesamtfolge, die nach Voraussetzung ja Cauchy-Folge ist. □

Ausfgabe. Ist $\mathcal{P}(S)$ vollständig, so auch S.

Aufgabe. Jedes endliche Maß μ auf einem vollständigen, separablen metrischen Raum S ist „von innen regulär“, d.h. zu jedem $\varepsilon > 0$ gibt es ein kompaktes $K \subset S$, so dass $\mu(K^c) \le \varepsilon$.

Der Satz von de Finetti

Mit Hilfe des Satzes von Prohorov beweisen wir nun einen bekannten Satz über Folgen von Zufallsvariablen mit austauschbaren Verteilungen. Sei $Z_1, Z_2, \ldots$ eine unendliche Folge von Zufallsvariablen mit einem metrischen Wertebereich S, der vollständig und separabel sei. Man sagt, dass die Folge eine austauschbare Verteilung besitzt, falls für alle natürlichen Zahlen k und alle Permutationen $\pi(1), \ldots, \pi(k)$ der Zahlen $1, \ldots, k$

$$(Z_{\pi(1)}, \ldots, Z_{\pi(k)}) \stackrel{v}{=} (Z_1, \ldots, Z_k)$$

gilt ($\stackrel{v}{=}$ bedeutet Gleichheit in Verteilung). Der Satz von de Finetti besagt dann, das $Z_1, Z_2, \ldots$ in Verteilung einer Folge von unabhängigen, identisch verteilten Zufallsvariablen gleicht, die ihrerseits eine zufällige Verteilung L haben. Genauer:

Behauptung. *Es gibt eine Zufallsvariable L mit Werten in $\mathcal{P}(S)$, so dass gilt*

$$\mathbf{E}\big[\varphi_1(Z_1) \cdots \varphi_k(Z_k)\big] = \mathbf{E}\Big[\int \varphi_1 \, dL \cdots \int \varphi_k \, dL\Big]$$

für alle k und alle beschränkten stetigen Abbildungen $\varphi_1, \ldots, \varphi_k$ von S nach $\mathbb{R}$.
Beweis. Wir leiten erst eine Version dieser Formel mit der empirischen Verteilung

$$L_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i}$$

ab. Für $n \geq k$ gilt unter Beachtung der Austauschbarkeit

$$\mathbf{E}\big[\varphi_1(Z_1) \cdots \varphi_k(Z_k)\big] = \frac{1}{n(n-1)\cdots(n-k+1)} \sideset{}{^*}\sum_{1 \leq i_1, \ldots, i_k \leq n} \mathbf{E}\big[\varphi_1(Z_{i_1}) \cdots \varphi_k(Z_{i_k})\big] \quad ,$$

wobei * bedeutet, dass nur über paarweise verschiedene $i_1, \ldots, i_k$ summiert wird. Dadurch bleiben größenordnungsmäßig $O(n^{k-1})$ Summanden unberücksichtigt, daher folgt

$$\mathbf{E}\big[\varphi_1(Z_1) \cdots \varphi_k(Z_k)\big] = n^{-k}\, \mathbf{E}\Big[\sum_{1 \leq i_1, \ldots, i_k \leq n} \varphi_1(Z_{i_1}) \cdots \varphi_k(Z_{i_k})\Big] + O(n^{-1}) \quad .$$

Wegen $\int \varphi \, dL_n = n^{-1} \sum_{i=1}^{n} \varphi(Z_i)$ können wir das auch als

$$\mathbf{E}\big[\varphi_1(Z_1) \cdots \varphi_k(Z_k)\big] = \mathbf{E}\Big[\int \varphi_1 \, dL_n \cdots \int \varphi_k \, dL_n\Big] + O(n^{-1})$$

schreiben. Unsere Behauptung folgt, wenn wir zeigen können, dass L_n in Verteilung gegen eine Zufallsvariable L konvergiert, denn $\mu \mapsto \int \varphi_1 \, d\mu \cdots \int \varphi_k \, d\mu$ ist für stetige, beschränkte $\varphi_1, \ldots, \varphi_k$ eine stetige, beschränkte Abbildung von $\mathcal{P}(S)$ nach $\mathbb{R}$. Wir brauchen den Grenzübergang sogar nur entlang irgendeiner Teilfolge zu vollziehen, deswegen langt es, die relative Kompaktheit der Folge der Verteilungen von L_n nachzuweisen. Nun gilt für relativ kompaktes $K \subset S$ und $\varepsilon > 0$ nach der Markov-Ungleichung und wegen der Austauschbarkeit

$$\mathbf{P}\{L_n(K^c) > \varepsilon\} \leq \frac{1}{\varepsilon}\mathbf{E}\big[L_n(K^c)\big] = \frac{1}{\varepsilon}\mathbf{E}\Big[\frac{1}{n}\sum_{i=1}^{n} I_{\{Z_i \in K^c\}}\Big] = \frac{1}{\varepsilon}\mathbf{P}\{Z_1 \in K^c\} \quad ,$$

und wir erhalten

$$\mathbf{P}\{L_n(K^c) > \varepsilon\} \leq \varepsilon \quad ,$$

wenn wir K in S nur ausreichend groß wählen (innere Regularität der Verteilung von Z_1). Die Behauptung folgt also mit Hilfe des folgenden Kriteriums.

Behauptung. *$\mathscr{K} \subset \mathcal{P}(\mathcal{P}(S))$ ist relativ kompakt, falls es für alle $\varepsilon > 0$ ein relativ kompaktes $K = K_\varepsilon \subset S$ gibt, so dass für alle $\mathscr{P} \in \mathscr{K}$ gilt*

$$\mathscr{P}\big(\{\mu \in \mathcal{P}(S) : \mu(K^c) > \varepsilon\}\big) \leq \varepsilon \quad .$$

Für $\mathcal{K} := \bigcap_{i=1}^\infty \{\mu \in \mathcal{P}(S) : \mu(K^c_{\varepsilon 2^{-i}}) \leq \varepsilon 2^{-i}\}$ gilt dann nämlich

$$\mathscr{P}\big(\mathcal{K}^c\big) \leq \sum_{i=1}^\infty \mathscr{P}\big(\mu(K^c_{\varepsilon 2^{-i}}) > \varepsilon 2^{-i}\big) \leq \sum_{i=1}^\infty \varepsilon 2^{-i} = \varepsilon$$

für $\mathscr{P} \in \mathscr{K}$. Außerdem ist $\mathcal{K}$ straff, also nach dem Satz von Prohorov relativ kompakt. Daher ist $\mathscr{K}$ straff und nach dem Satz von Prohorov relativ kompakt. – Übrigens ist die angegebene Bedingung im Fall von Vollständigkeit auch notwendig (Übung). □

I.4 Der Raum $D_S[0,\infty)$

Wir wenden uns nun dem Fall der Verteilungskonvergenz

$$X^n \xrightarrow{v} X$$

von S-wertigen stochastischen Prozessen

$$X^n = (X^n_t)_{t\geq 0}\,, \quad X = (X_t)_{t\geq 0}$$

zu. Die Idee ist, X^n und X als Zufallsvariable mit Werten in einem Funktionenraum aufzufassen. Um den Anschluss an die bisherige Theorie herzustellen, benötigen wir passende Funktionenräume, die sich in geeigneter Weise metrisieren lassen. Dies führt dann zu folgender Strategie des Nachweises der Verteilungskonvergenz:

- Erstens zeigt man Konvergenz der endlichdimensionalen Verteilungen, damit sind die Verteilungen von $\big(X^n_{t_1}, \ldots, X^n_{t_k}\big)$ mit $0 \leq t_1 < \cdots < t_k$ gemeint. Dann kann es höchstens eine Grenzprozess X geben.

- Zweitens zeigt man relative Kompaktheit der Verteilungen von X^n.

Die Strategie folgt der elementaren Tatsache, dass eine Folge in einem metrischen Raum genau dann konvergiert, wenn sie relativ kompakt ist und höchstens einen Häufungspunkt besitzt. In vielen Fällen ist der Raum $C_S = C_S[0,\infty)$ aller stetigen Funktionen $f : \mathbb{R}_+ \to S$, versehen mit einer Supremumsmetrik geeignet. Für stochastische Prozesse mit unstetigen Pfaden ist dieser Rahmen jedoch im Allgemeinen zu eng. Hier ist es in aller Regel ausreichend, für f nur Sprungstellen als Unstetigkeitsstellen zuzulassen.

Definition. *Eine Funktion $f : \mathbb{R}_+ \to S$ heißt **cadlag-Funktion** (continue à droite, limites à gauche) falls für alle $t \geq 0$ die rechts- und linksseitigen Limiten*

$$f(t+) = \lim_{s\downarrow t} f(s) \quad , \quad f(t-) = \lim_{s\uparrow t} f(s)$$

existieren (mit $f(0-) := f(0)$), und falls für alle $t \geq 0$

$$f(t+) = f(t)$$

gilt. Der Raum aller cadlag-Funktionen mit Werten in S wird mit $D_S = D_S[0,\infty)$ bezeichnet.

Die folgende Eigenschaft einer cadlag-Funktion f wird für uns wichtig sein: Für alle $\eta > 0$ gibt es eine divergente Folge reeller Zahlen $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$, so dass für alle $i \geq 1$

$$s, t \in [\tau_{i-1}, \tau_i) \;\Rightarrow\; d(f(s), f(t)) \leq \eta \;. \tag{I.1}$$

Die τ_i lassen sich induktiv konstruieren, etwa $\tau_i := (\tau_{i-1} + 1) \wedge \inf\{t > \tau_{i-1} : d(f(t), f(\tau_{i-1})) > \eta/2\}$. Wegen der Rechtsstetigkeit von f bilden die τ_i eine strikt wachsende Folge. Wäre $\tau := \sup_i \tau_i$ endlich, so hätte man einerseits $d(f(\tau_i), f(\tau_{i-1})) \geq \eta/2$ für alle i und andererseits mit linksseitigem Limes $\lim_n d(f(\tau_i), f(\tau_{i-1})) = d(f(\tau-), f(\tau-)) = 0$, ein Widerspruch. Also ist die Folge (τ_i) divergent.

Diese Folge (τ_i) enthält insbesondere alle Sprungstellen s von f mit $d(f(s-), f(s)) > \eta$. Wir erkennen also, dass sich bei einer cadlag-Funktion f große Sprünge nicht häufen können, und folgern: *Eine cadlag-Funktion hat höchstens abzählbar viele Unstetigkeitsstellen.*

Wir wollen D_S metrisieren. Eine Supremumsmetrik, wie man sie in $C_S[0,\infty)$ benutzt, ist ungeeignet, denn D_S wäre dann nicht separabel. In der sup-Metrik haben nämlich die überabzählbar vielen Funktionen

$$f_u(t) = \begin{cases} x & \text{für } t < u \\ y & \text{für } t \geq u \end{cases} \quad , \quad u > 0 \quad .$$

alle den Abstand $d(x, y)$ voneinander, gleichgültig, wie nahe die Sprungstellen beieinander liegen. Bei der Wahl der Metrik ist also darauf zu achten, dass benachbarte Sprungstellen von $f, g \in D_S$ den Abstand zwischen f und g nicht über Gebühr vergrößern. Skorohod folgend werden dazu, bevor zwischen den Funktionen der Supremumsabstand gebildet wird, zunächst benachbarte Sprungstellen von f und g übereinander geschoben, die Definitionsbereiche von f und g sozusagen aneinander angeglichen. Dies führt zu dem folgenden Konvergenzbegriff.

Definition. *Wir sagen, eine Folge (f_n) in D_S **konvergiert per Angleichen** gegen $f \in D_S$, falls es für alle $t > 0$ stetige, strikt wachsende Abbildungen $\alpha_n : \mathbb{R}_+ \to \mathbb{R}_+$ mit $\alpha_n(0) = 0$ gibt, so dass gilt*

$$\sup_{s \leq t} |\alpha_n(s) - s| \to 0 \quad \text{und} \quad \sup_{s \leq t} d\big(f(\alpha_n(s)), f_n(s)\big) \to 0 \quad .$$

Bemerkung. Wenn f stetig ist, gilt auch $\sup_{s \leq t} d(f(\alpha_n(s)), f(s)) \to 0$ für alle $t > 0$. Dies folgt, weil f dann auf kompakten Intervallen gleichmäßig stetig ist. Für stetiges f läuft unser Konvergenzbegriff also auf die Bedingung $\sup_{s \leq t} d(f(s), f_n(s)) \to 0$ hinaus, auf die lokal gleichmäßige punktweise Konvergenz von f_n gegen f. Ein Angleichen der Definitionsbereiche erübrigt sich.

Die Konvergenz per Angleichen lässt sich metrisieren. Dazu bezeichnen wir für $\eta > 0$ mit $A(\eta)$ die Menge aller stetigen Bijektionen $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$, die $\sup_{t \geq 0} |\alpha(t) - t| \leq \eta$ erfüllen. Offenbar gilt

$$\alpha \in A(\eta) \Rightarrow \alpha^{-1} \in A(\eta) \;\; ; \qquad \alpha_1 \in A(\eta_1), \alpha_2 \in A(\eta_2) \Rightarrow \alpha_1 \circ \alpha_2 \in A(\eta_1 + \eta_2) \quad .$$

Für $f, g \in D_S$ setzen wir

$$d_S(f, g) := \inf \Big\{ \eta > 0 : \sup_{\max(t, \alpha(t)) \leq 1/\eta} d\big(f(\alpha(t)), g(t)\big) \leq \eta \text{ für ein } \alpha \in A(\eta) \Big\} \;. \tag{I.2}$$

Wegen $d \leq 1$ gilt offenbar auch $d_S \leq 1$, weil wir nur den Fall $\eta \leq 1$ im Blick haben müssen. Die Forderung $\max(t, \alpha(t)) \leq 1/\eta$ an t benötigen wir, um d_S zur Metrik zu machen, sie ist stärker als $t \leq \eta^{-1}$ und (wegen $\sup_{t \geq 0} |\alpha(t) - t| \leq \eta \leq 1$) schwächer als $t \leq \eta^{-1} - 1$.

Proposition 1.10. *Mit* (I.2) *ist eine Metrik d_S auf D_S gegeben, die die Konvergenz per Angleichen metrisiert.*

Wir bezeichnen d_S als die **Skorohod-Metrik**.

Beweis. Gilt $d_S(f,g)=0$, so gibt es Bijektionen α_n mit $\alpha_n(t)\to t$ und $d\big(f(\alpha_n(t)),g(t)\big)\to 0$ für alle $t\geq 0$. Für einen Stetigkeitspunkt t von g folgt $f(t)=g(t)$. Da die Stetigkeitspunkte liegen dicht, folgt $f=g$.

Symmetrie: $d_S(f,g)=d_S(g,f)$ ergibt sich, indem man in der Definition $t=\alpha^{-1}(s)$ setzt und beachtet, dass mit α auch α^{-1} zu $A(\eta)$ gehört.

Zur $\triangle$-Ungleichung $d_S(f,h)\leq d_S(f,g)+d_S(g,h)$: Offenbar gilt $d_S(f,h)\leq 1$, deswegen können wir uns auf den Fall $d_S(f,g)+d_S(g,h)<1$ beschränken. Für $\eta_1=d_S(f,g)+\varepsilon,\ \eta_2=d_S(g,h)+\varepsilon$ mit $\varepsilon>0$ können wir dann $\eta_1+\eta_2\leq 1$ annehmen. Seien $\alpha_i\in A(\eta_i)$, $i=1,2$, dazu passende Bijektionen gemäß der Definition von d_S.

Wir betrachten erst den Fall $d_S(f,g)\leq d_S(g,h)$ und damit $\eta_1\leq\eta_2$. Falls nun $\alpha_1\circ\alpha_2(t)\leq 1/(\eta_1+\eta_2)$ gilt, folgt wegen $\alpha_1\in A(\eta_1)$ die Ungleichung $\alpha_2(t)\leq\alpha_1(\alpha_2(t))+\eta_1\leq(\eta_1+\eta_2)^{-1}+\eta_1$. Dieser Ausdruck ist in η_1 fallend, solange $\eta_1+\eta_2\leq 1$ gilt. Daher folgt $\alpha_2(t)\leq 1/\eta_2$ und wegen $\eta_1\leq\eta_2$ auch $\alpha_2(t)\leq 1/\eta_1$. Außerdem haben wir $\alpha_1(\alpha_2(t))\leq 1/\eta_1$. Nehmen wir noch die Bedingung $t\leq 1/(\eta_1+\eta_2)$ hinzu, so gilt auch $t\leq 1/\eta_2$. Aus $\max(t,\alpha_1\circ\alpha_2(t))\leq 1/(\eta_1+\eta_2)$ folgt also

$$d(f(\alpha_1\circ\alpha_2)(t),h(t))\leq d(f(\alpha_1(\alpha_2(t))),g(\alpha_2(t))+d(g(\alpha_2(t)),h(t))\leq\eta_1+\eta_2$$

und mit $\varepsilon\to 0$ $d_S(f,h)\leq d_S(f,g)+d_S(g,h)$. Im konträren Fall $d_S(h,g)\leq d_S(g,f)$ erhalten wir entsprechend die Abschätzung $d_S(h,f)\leq d_S(h,g)+d_S(g,f)$. Beidesmal ergibt sich die $\triangle$-Ungleichung.

Die Konvergenzaussage der Proposition folgt direkt aus den Definitionen. □

Proposition 1.11. *Ist S mit der Metrik d ein separabler Raum, so auch D_S mit der Metrik d_S.*

Beweis. Es bezeichne Q eine abzählbare, dichte Teilmenge von S. Wir zeigen, dass die abzählbar vielen Treppenfunktionen der Gestalt $g=\sum_{i=1}^k q_i 1_{[t_{i-1},t_i)}$ mit $q_i\in Q$ und rationalen Zahlen t_i eine dichte Teilmenge des D_S bilden. Sei $f\in D_S$ und sei $\eta>0$. Wie oben seien $0=\tau_0<\tau_1<\cdots<\tau_k$ reelle Zahlen mit $\tau_k>2/\eta$ und $d(f(s),f(t))\leq\eta/2$, sofern $s,t\in[\tau_{i-1},\tau_i)$ für ein $i\leq k$. Wähle $q_i\in Q$ mit $d(q_i,f(\tau_{i-1}))\leq\eta/2$ und rationale $0=t_0<t_1<\cdots<t_k$ mit $|t_i-\tau_i|\leq\eta$. Definiere die Bijektion α durch

$$\alpha(t_i)=\tau_i\ ,\qquad i\leq k\ ,$$

dazwischen wird α linear interpoliert und oberhalb ρ_k mit Steigung 1 fortgesetzt. Wegen $|\alpha(t_i)-t_i|\leq\eta$ folgt $\alpha\in A(\eta)$. Weiter gehören für $t\leq 1/\eta$ nach Konstruktion $\alpha(t)$ und t immer demselben Intervall $[\tau_{i-1},\tau_i)$ an, was

$$d(f(t),g(\alpha(t)))\leq d(f(t),f(\tau_{i-1}))+d(f(\tau_{i-1}),g(\alpha(t)))\leq\frac{\eta}{2}+d\big(f(\tau_{i-1}),q_i\big)\leq\eta$$

impliziert. Insgesamt erhalten wir $d_S(f,g)\leq\eta$, wie zu beweisen war. □

Ein wesentlicher Nachteil der Skorohod-Metrik ist, dass sie zu keiner Vollständigkeit des metrischen Raumes führt:

Aufgabe. Die stetigen Funktionen f_n, $n\geq 1$, seien gegeben durch $f_n(t)=0$ für $t\leq 1-n^{-1}$, durch $f_n(t)=1$ für $t\geq 1$ und durch lineare Interpolation auf dem Intervall $[1-n^{-1},1]$. Zeigen Sie, dass die Folge (f_n) bzgl. der Metrik d_S eine Cauchy-Folge ist, jedoch im Raum D_S nicht konvergiert.

Anschaulich gesprochen konvergiert in diesem Beispiel die Folge (f_n) gegen die Funktion $f(t)=0$ für $t<1$ und $f(t)=1$ für $t\geq 1$. Dabei entsteht erst im Limes der Sprung an der Stelle 1, der im Grenzprozess nicht mehr durch Angleichen aufgefangen werden kann.

Für Vollständigkeit benötigt man deshalb eine andere Metrik d_B, die auf dieses Szenario sensibel reagiert. Billingsley folgend erreicht man dies, indem man für die angleichenden Bijektionen nicht nur Konvergenz gegen die Identität fordert, sondern auch Konvergenz ihrer Steigungen gegen 1. In diesem theoretischen Kontext hat also die Skorohod-Metrik d_S Mängel. Das bedeutet aber nicht, dass sie nutzlos wäre. So werden wir sie wesentlich im Beweis von Korollar 2.9 einsetzen, dort wäre sie kaum durch die nun zu behandelnde Metrik d_B zu ersetzen.

Bei der Konstruktion von d_B gehen wir nach demselben Schema wie bei d_S vor. Für $\eta > 0$ bezeichne $B(\eta)$ die Menge aller stetigen, strikt monotonen $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ mit $\beta(0) = 0$ und

$$\Big| \log \frac{\beta(t) - \beta(s)}{t - s} \Big| \le \eta \qquad \text{für alle } s \ne t \; .$$

Es ist dann β eine stetige Bijektion, die der Wachstumsbedingung

$$e^{-\eta} t \le \beta(t) \le e^{\eta} t \qquad \text{für alle } t \ge 0$$

genügt. Es gilt

$$\beta \in B(\eta) \Rightarrow \beta^{-1} \in B(\eta) \;\; ; \qquad \beta_1 \in B(\eta_1), \beta_2 \in B(\eta_2) \Rightarrow \beta_1 \circ \beta_2 \in B(\eta_1 + \eta_2) \;\; .$$

Für $f, g \in D_S$ setzen wir nun

$$d_B(f,g) := \inf \Big\{ \eta > 0 : \sup_{\max(t,\beta(t)) \le 1/\eta} d\big(f(\beta(t)), g(t)\big) \le \eta \text{ für ein } \beta \in B(\eta) \Big\} \; .$$

Proposition 1.12. *Auch mit d_B ist eine Metrik auf D_S gegeben. Die Metriken d_S und d_B sind äquivalent, d. h. beide ergeben dasselbe System von offenen Mengen.*

Wir bezeichnen d_B als die **Billingsley-Metrik**.

Beweis. Der Beweis der Eigenschaften einer Metrik verläuft wie bei der Skorohod-Metrik, abgesehen von der dort befindlichen Abschätzung von $\alpha_2(t)$. Hier verfahren wir wir folgt: Sei $\beta_1 \circ \beta_2(t) \le 1/(\eta_1 + \eta_2)$. Dann folgt $\beta_2(t) \le e^{\eta_1} \beta_1(\beta_2(t)) \le e^{\eta_1}/(\eta_1 + \eta_2)$. Dieser Ausdruck ist fallend in η_1 für $\eta_1 + \eta_2 \le 1$. Unter dieser Annahme folgt also $\beta_2(t) \le 1/\eta_2$, was den Beweis ergänzt.

Die Äquivalenz beider Metriken beweisen wir, indem wir zeigen, dass jede in d_S konvergente Folge (f_n) mit Grenzwert f auch in d_B gegen f konvergiert, und umgekehrt (siehe die folgende Aufgabe).

Sei also (f_n) in d_S gegen f konvergent. Dann hat man Bijektionen α_n, so dass $\alpha_n(t) \to t$ und $d_S(f_n(\alpha_n(t), f(t)) \to 0$ für alle $t \ge 0$ gilt (gleichmäßig auf kompakten Intervallen). Wie oben gezeigt, gibt es zu vorgegebenem $\eta > 0$ reelle Zahlen $0 = \tau_0 < \tau_1 < \cdots < \tau_k$ mit $\tau_k > 1/\eta$, so dass $|f(s) - f(t)| \le \eta/2$, sofern $s, t \in [\tau_{i-1}, \tau_i)$ für ein $i \le k$. Wir definieren Bijektionen β_n durch

$$\beta_n(\tau_j) = \alpha_n(\tau_j), \quad j \le k \;\; ,$$

dawischen wird β_n linear interpoliert und oberhalb von τ_k erhält β_n die Steigung 1. Offenbar konvergieren dann die Steigungen von β_n überall gegen 1, und es folgt $\beta_n \in B(\eta)$ für ausreichend großes n. Außerdem gehören $\alpha_n(t)$ und $\beta_n(t)$ für alle $t \le 1/\eta$ beide zu demselben Intervall $[\tau_{i-1}, \tau_i)$, $i \le k$, was $d(f(\alpha_n(t)), f(\beta_n(t))) \le \eta/2$ nach sich zieht. Es folgt für ausreichend großes n

$$d(f(\beta_n(t)), f_n(t)) \le d(f(\alpha_n(t)), f_n(t)) + \eta/2 \le \eta \; ,$$

gleichmäßig für $t \le 1/\eta$. Für alle $\eta > 0$ gibt es also ein $n_0 \in \mathbb{N}$, so dass $d_B(f_n, f) \le \eta$ für $n \ge n_0$. Mit anderen Worten: $d_B(f_n, f)$ konvergiert gegen 0.

In dieser Überlegung können wir ohne weiteres die Rollen von α_n und β_n vertauschen, d. h. aus der Konvergenz in d_B folgt auch diejenige in d_S (hier kann man sogar α_n als β_n wählen). □

Aufgabe. Zeigen Sie: Zwei Metriken d_S und d_B auf einem Raum S induzieren genau dann dasselbe System von offenen Mengen, wenn ihre konvergenten Folgen (samt Grenzwert) übereinstimmen.

Satz 1.13. *Ist S mit der Metrik d ein vollständiger Raum, so auch D_S mit der Metrik d_B.*

Beweis. Sei (f_n) Cauchy-Folge in D_S. Zur Konstruktion ihres Grenzwertes können wir ohne Einschränkung $d_B(f_{n-1}, f_n) < e^{-n}$ annehmen (denn es langt, eine konvergente Teilfolge zu konstruieren). Dann gibt es $\beta_n \in B(e^{-n})$, so dass $d(f_{n-1}(s), f_n(\beta_n(s))) \leq e^{-n}$ für $s \leq e^{n-1}$. Es folgt für $m < n$ und $t > 0$

$$\begin{aligned}&\big|\log \beta_m \circ \cdots \circ \beta_1(t) - \log \beta_n \circ \cdots \circ \beta_1(t)\big| \\ &\qquad \leq \sum_{k=m+1}^{n} \Big|\log \frac{\beta_k(\beta_{k-1} \circ \cdots \circ \beta_1(t))}{\beta_{k-1} \circ \cdots \circ \beta_1(t)}\Big| \leq \sum_{k=m+1}^{n} e^{-k} \leq e^{-m} \quad .\end{aligned}$$

$\log \beta_n \circ \cdots \circ \beta_1(t)$, $n \geq 1$, ist also Cauchy-Folge in $\mathbb{R}$, folglich existiert für alle $t \geq 0$ der Limes

$$\beta(t) := \lim_n \beta_n \circ \cdots \circ \beta_1(t) \quad .$$

Aus $\beta_n \circ \cdots \circ \beta_1 \in B(e^{-n} + \cdots + e^{-1}) \subset B(1)$ folgt durch Grenzübergang $\beta \in B(1)$ und $\beta \circ \beta_1^{-1} \circ \cdots \circ \beta_k^{-1} = \lim_n \beta_{k+1} \circ \cdots \circ \beta_n \in B(e^{-k})$. Durch Übergang zu inversen Abbildungen folgt $\gamma_k := \beta_k \circ \cdots \circ \beta_1 \circ \beta^{-1} \in B(e^{-k})$. Insbesondere folgt aus $t \leq e^{k-2}$ für $s := \gamma_{k-1}(t)$ die Abschätzung $s \leq e^{e^{-k+1}} t \leq e^{k-1}$ und damit

$$d\big(f_{k-1}(\gamma_{k-1}(t)), f_k(\gamma_k(t))\big) = d\big(f_{k-1}(s), f_k(\beta_k(s))\big) \leq e^{-k} \quad .$$

Also ist $f_k(\gamma_k(t))$, $k \geq 1$ für alle $t \geq 0$ Cauchy-Folge in S, deren Limes wir mit $f(t)$ bezeichnen. Für $t \leq e^{m-1}$ folgt weiter

$$\begin{aligned}d\big(f_m(\gamma_m(t), f(t))\big) &= \lim_n d\big(f_m(\gamma_m(t)), f_n(\gamma_n(t))\big) \\ &\leq \sum_{k=m+1}^{\infty} d\big(f_{k-1}(\gamma_{k-1}(t)), f_k(\gamma_k(t))\big) \leq e^{-m} \quad .\end{aligned}$$

Die Konvergenz ist insbesondere gleichmäßig auf beschränkten Intervallen. Daraus folgt, dass mit den $f_m \circ \gamma_m$ auch f zu D_S gehört, und außerdem $d_B(f, f_m) \leq e^{-m+1}$ gilt. f ist also Grenzwert von f_n. □

$C_S[0,\infty)$ als Teilraum von $D_S[0,\infty)$

C_S ist in D_S abgeschlossen: Ist (f_n) eine Folge stetiger Funktionen, die durch Angleichen gegen $f \in D_S$ konvergiert, so gilt $f_n \circ \alpha_n(t) \to f(t)$ für geeignete α_n. Da $f_n \circ \alpha_n$ stetig ist und die Konvergenz gleichmäßig auf Kompakta stattfindet, ist dann auch f stetig.

Wir haben bereits bemerkt, dass auf C_S ein Angleichen nicht nötig ist. Wenn wir also innerhalb C_S zu der Metrik

$$d_C(f,g) := \inf\{\varepsilon > 0 \,:\, \sup_{t \leq 1/\varepsilon} d\big(f(t), g(t)\big) \leq \varepsilon\}$$

übergehen, bleibt die Konvergenz von Folgen davon unberührt. Da C_S mit der Metrik d_C bekanntlich eine vollständiger Raum ist, haben d_B und d_C in C_S auch dieselben Cauchy-Folgen (nämlich die konvergenten Folgen).

Der Stetigkeitsmodul

Für spätere Zwecke gehen wir nun noch etwas technischer auf die Aussage (I.1) ein. Sie beinhaltet die lokal gültige gleichmäßige Stetigkeit einer cadlag-Funktion, die nur durch Unstetigkeitsstellen gestört wird. Je dominanter diese auftreten, um so enger werden die Partitionen $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$. Dies quantifiziert die folgende Begriffsbildung.

Definition. *Seien $u, \delta > 0$. Der **Stetigkeitsmodul** $\omega_{u,\delta}(f)$ einer Funktion $f : [0,\infty) \to S$ ist das Infimum aller $\varepsilon > 0$, für die es Zahlen $0 = \tau_0 < \tau_1 < \cdots < \tau_k$ mit $\tau_k > u$ gibt, so dass für alle $i \leq k$*

$$\tau_i > \tau_{i-1} + \delta \quad \textit{sowie} \quad s,t \in [\tau_{i-1}, \tau_i) \Rightarrow d\big(f(s), f(t)\big) \leq \varepsilon \quad .$$

Eine Funktion $f \in D_S$ mit kleinem Stetigkeitsmodul kann große Sprungstellen haben. Diese müssen dann aber in den Partition $\tau_0 < \cdots < \tau_k$ untergebracht werden und dürfen daher nicht zu nahe beieinander liegen. Offenbar ist $\omega_{u,\delta}(f)$ monoton fallend in δ, auch besagt (I.1), dass $\omega_{u,\delta}(f)$ für eine cadlag-Funktion f beliebig klein gemacht werden kann. Wir haben also die folgende Aussage.

Lemma 1.14. *Für alle $f \in D_S$ und alle $u > 0$ gilt $\lim_{\delta\to 0}\omega_{u,\delta}(f) = 0$.*

Aufgabe. Erfüllt ein Funktion $f : [0,\infty) \to S$ für alle $u > 0$ die Eigenschaft $\lim_{\delta\to 0}\omega_{u,\delta}(f) = 0$, so ist sie cadlag.
Hinweis: Betrachten Sie der Reihe nach die Fälle, dass an einer Stelle $t \geq 0$ die Grenzwerte $f(t-)$ oder $f(t+)$ nicht existieren, oder dass $f(t) \neq f(t+)$ gilt.

Mit dem Stetigkeitsmodul erfasst man, wie sich eine cadlag-Funktion in der Skorohod- oder Billingsley-Metrik durch Treppenfunktionen approximieren lässt.

Beispiel. Sei $f \in D_S$ und für $n \in \mathbb{N}$, $u > 1/n$

$$f_{u,n} := \sum_{1\leq k\leq un} f\Big(\frac{k}{n}\Big) 1_{[\frac{k-1}{n},\frac{k}{n})} \ . \tag{I.3}$$

Wir wollen den Abstand zwischen f und $f_{u,n}$ abschätzen. Dazu wählen wir $\delta > 1/n$ und $\varepsilon > \omega_{u,\delta}(f)$. Sei $0 = \tau_0 < \cdots < \tau_k$ passend zu ε gewählte Partiton. Wir definieren $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ durch

$$\alpha(\tau_i) := \frac{j}{n}, \text{ falls } \frac{j-1}{n} < \tau_i \leq \frac{j}{n} \ , \qquad i \leq k \ ,$$

und lineare Interpolation dazwischen ($\alpha(s)$ erhält Steigung 1 für $s \geq \tau_k$). Wegen $\delta > 1/n$ haben wir $\alpha(\tau_{i-1}) < \alpha(\tau_i)$, α ist also strikt monoton. Es folgt

$$s < \tau_i \quad \Leftrightarrow \quad \alpha(s) < \frac{j}{n} \quad \Leftrightarrow \quad [n\alpha(s)] \leq j-1 \quad \Leftrightarrow \quad \frac{1}{n}\big[n\alpha(s)\big] < \tau_i \ ,$$

d.h. s und $\frac{1}{n}\big[n\alpha(s)\big]$ liegen immer in demselben Intervall $[\tau_{i-1},\tau_i)$. Für $s \leq u$ folgt

$$d\big(f(s), f_{u,n}(\alpha(s))\big) = d\big(f(s), f([n\alpha(s)]/n)\big) \leq \varepsilon \ .$$

Weiter gilt $|\alpha(\tau_i) - \tau_i| \leq 1/n$ und folglich $\sup_t |\alpha(t) - t| \leq 1/n$ sowie $\alpha(u - n^{-1}) \leq u$. Insgesamt folgt $d_S(f, f_{u,n}) \leq \max(\varepsilon, (u - n^{-1})^{-1}, n^{-1})$ und mit $\varepsilon \to \omega_{u,\delta}(f)$

$$d_S(f, f_{u,n}) \leq \max(\omega_{u,\delta}(f), (u - n^{-1})^{-1}, n^{-1}) \ . \tag{I.4}$$

Außerdem gilt nach Konstruktion $\tau_i - \tau_{i-1} - \frac{1}{n} \leq \alpha(\tau_i) - \alpha(\tau_{i-1}) \leq \tau_i - \tau_{i-1} + \frac{1}{n}$ und damit

$$\log\Big(\frac{\delta - \frac{1}{n}}{\delta}\Big) \leq \log\Big(\frac{\alpha(u) - \alpha(v)}{u - v}\Big) \leq \log\Big(\frac{\delta + \frac{1}{n}}{\delta}\Big) \leq \log\Big(\frac{\delta}{\delta - \frac{1}{n}}\Big) \ .$$

Wir können die Bijektion α auch für den Billingsley-Abstand nutzen und erhalten

$$d_B(f_{u,n}, f) \leq \max\big(\omega_{u,\delta}(f), (u - n^{-1})^{-1}, -\log\big(1 - (\delta n)^{-1}\big)\big) \ . \tag{I.5}$$

Die Abschätzungen (I.4) und (I.5) zusammen mit Lemma 1.14 ermöglichen die folgende Aussage.

Lemma 1.15. *Für $u_n \to \infty$ gilt $d_S(f, f_{u_n,n}) \to 0$ und $d_B(f, f_{u_n,n}) \to 0$.*

I.5 Schwache Konvergenz von W-Maßen auf $D_S[0,\infty)$

Möchte man eine Familie (X_t) von S-wertigen Zufallsvariablen zu einer einzige Zufallsgröße X mit Werten im Raum D_S der cadlag-Pfade zusammenfassen, so stellt sich die Frage nach einer geeigneten σ-Algebra auf D_S. Eine natürliche Wahl ist die von den Projektionsabbildungen

$$\pi_t : D_S \to S \ , \quad \pi_t(f) := f(t)$$

erzeugte σ-Algebra, also die kleinste σ-Algebra, für die alle π_t, $t \geq 0$ messbare Abbildungen werden. Dann lassen sich die Ereignisse $\{X \in \pi^{-1}(B)\} = \{\pi_t(X) \in B\} = \{X_t \in B\}$ für alle Borel-Mengen $B \subset S$ bilden, und die Mengen $\pi_t^{-1}(B)$ geben einen übersichtlichen Erzeuger für die σ-Algebra her. Es ist nicht selbstverständlich, dass man damit gerade die von der Skorohod-Metrik stammenden Borel-σ-Algebra in D_S erhält (keine der Projektionsabbildungen ist bzgl. der Skorohod-Metrik stetig!).

Proposition 1.16. *Die Borel-σ-Algebra in D_S wird von den Projektionsabbildungen $\pi_t, t \geq 0$, erzeugt.*

Beweis. Sei $\sigma(\pi)$ die von den Projektionen erzeugte σ-Algebra. Wir zeigen zuerst $\sigma(\pi) \subset \mathscr{B}_{D_S}$. Dazu stellen wir fest: $d_S(f_n, f) \to 0$ impliziert $f_n(t) \to f(t)$ für fast alle $t \geq 0$ (nämlich den Stetigkeitspunkten von f). Ist also $\varphi \in C_b(S)$ und $\varepsilon > 0$, so folgt für $n \to \infty$ nach dem Satz von der dominierten Konvergenz

$$\int_t^{t+\varepsilon} \varphi\big(f_n(s)\big) ds \to \int_t^{t+\varepsilon} \varphi\big(f(s)\big) ds \quad .$$

Daher ist das Funktional $f \mapsto \int_t^{t+\varepsilon} \varphi\big(\pi_s(f)\big) ds$ stetig und damit Borel-messbar. Für $\varepsilon \to 0$ folgt $\varepsilon^{-1} \int_t^{t+\varepsilon} \varphi\big(\pi_s(f)\big) ds \to \varphi\big(\pi_t(f)\big)$ aufgrund der Rechtsstetigkeit von f, also ist $\varphi \circ \pi_t$ Borel-messbar. Schließlich gibt es für offenes $O \subset S$ Funktionen $\varphi_n \in C_b(S)$, die punktweise gegen 1_O konvergieren. Damit ist $1_O \circ \pi_t$ Borel-messbar, d.h. $\{\pi_t \in O\}$ ist eine Borel-Menge in D_S. π_t ist also Borel-messbar, und es gilt $\sigma(\pi) \subset \mathscr{B}_{D_S}$.

Umgekehrt: Sei $g \in D_S$ fest gewählt und $\varphi : S^{n^2+1} \to \mathbb{R}$ def. als

$$\varphi(x_0, \ldots, x_{n^2}) = d_S\Big(g, \sum_{k=0}^{n^2} x_k \cdot 1_{[\frac{k}{n}, \frac{k+1}{n})}\Big) \quad .$$

φ ist stetig und also $\psi_n := \varphi \circ (\pi_0, \pi_{n^{-1}}, \ldots, \pi_{n-n^{-1}}, \pi_n)$ messbar bezüglich $\sigma(\pi)$. Unter Nutzung der Notation in (I.3) und mithilfe von Lemma 1.15 haben wir

$$\psi_n(f) = d_S(g, f_{n^2,n}) \to d_S(g, f) \quad ,$$

damit ist auch $f \mapsto d_S(f, g)$ eine $\sigma(\pi)$-messbare Abbildung. Folglich ist $U_\varepsilon(g) = \{f : d_S(f, g) < \varepsilon\}$ ein Element von $\sigma(\pi)$. Wegen der Separabilität von D_S ist jede offene Teilmenge von D_S als abzählbare Vereinigung solcher Umgebungen Element von $\sigma(\pi)$. Also gilt auch $\mathscr{B}_{D_S} \subset \sigma(\pi)$. $\square$

Für die Betrachtung schwacher Konvergenz auf dem Raum der cadlag-Funktionen sind diejenigen Projektionsabbildungen geeignet, die fast sicher stetig sind. Es stellt sich heraus, dass dies meist der Fall ist: Für $\mu \in \mathcal{P}(D_S)$ sei

$$L(\mu) := \{t \geq 0 : \pi_t \text{ ist } \mu\text{-fast sicher stetig}\} \quad .$$

Proposition 1.17. *Das Komplement von $L(\mu)$ in $\mathbb{R}_+$ ist für alle $\mu \in \mathcal{P}(D_S)$ abzählbar.*

Beweis. Aus $d_B(f, f_n) \to 0$ folgt nach Definition der Konvergenz per Angleichen, dass $f_n(t)$ für festes t höchstens die Häufungspunkte $f(t)$ oder $f(t-)$ hat. Also langt es zu zeigen, dass die Gleichung $\mu(\pi_t \neq \pi_{t-}) = 0$ für höchstens abzählbar viele $t \geq 0$ nicht gültig ist.
Sei (t_n) eine beschränkte Folge positiver Zahlen, die alle voneinander verschieden sind, und sei $\varepsilon > 0$. Die Menge $\{f \in D_S : d(f(t_m), f(t_m-)) \geq \varepsilon \text{ für ein } m \geq n\}$ geht dann für $n \to \infty$ über in $\{f : d(f(t_m), f(t_m-)) \geq \varepsilon\ \infty\text{-oft}\} = \emptyset$. Es folgt $\mu\big(d(\pi_{t_m}, \pi_{t_m-}) \geq \varepsilon \text{ für ein } m \geq n\big) \to 0$ und damit $\mu\big(d(\pi_{t_n}, \pi_{t_n-}) \geq \varepsilon\big) \to 0$. Es kann zu vorgegebenem t also höchstens endlich viele $s \leq t$ mit $\mu\big(d(\pi_s, \pi_{s-}) \geq \varepsilon\big) \geq \varepsilon$. Mit $\varepsilon \to 0$ folgt die Behauptung. $\square$

Wir können nun schwache Konvergenz auf dem Raum der cadlag-Funktionen wie folgt charakterisieren.

Proposition 1.18. *Seien $\mu, \mu_n \in \mathcal{P}\big(D_S\big)$. Dann ist äquivalent:*

i) $\mu_n \xrightarrow{s} \mu$,

ii) Die Folge (μ_n) ist relativ kompakt und für alle $t_1, \ldots, t_k \in L(\mu)$ gilt

$$(\pi_{t_1}, \ldots, \pi_{t_k})(\mu_n) \xrightarrow{s} (\pi_{t_1}, \ldots, \pi_{t_k})(\mu) \quad .$$

Beweis. Nur der Schluss ii) $\Rightarrow$ i) bedarf einer Begründung. Nehmen wir an, dass μ_n nicht gegen μ konvergiert. Wegen der relativen Kompaktheit von (μ_n) gibt es dann eine konvergente Teilfolge $(\mu_{n'})$ mit Limes $\nu \neq \mu$. Folglich konvergiert $(\pi_{t_1},\dots,\pi_{t_k})(\mu_{n'})$ für $t_1,\dots,t_k \in L(\nu)$ schwach gegen $(\pi_{t_1},\dots,\pi_{t_k})(\nu)$. Nach Annahme folgt $(\pi_{t_1},\dots,\pi_{t_k})(\nu) = (\pi_{t_1},\dots,\pi_{t_k})(\mu)$ für $t_1,\dots,t_k \in L(\nu)\cap L(\mu)$. Nach der letzten Proposition ist $L(\nu)\cap L(\mu)$ dicht in $\mathbb{R}_+$, außerdem konvergiert π_s punktweise gegen π_t auf dem Raum der cadlag-Funktionen, falls s von rechts gegen t strebt. Deswegen überträgt sich die Gleichung $(\pi_{t_1},\dots,\pi_{t_k})(\nu) = (\pi_{t_1},\dots,\pi_{t_k})(\mu)$ durch einen Grenzübergang auf alle $t_1,\dots,t_k \geq 0$. ν und μ haben also dieselben endlichdimensionalen Verteilungen und sind dann bekanntlich gleich, im Widerspruch zu unserer Annahme. □

Vielleicht noch übersichtlicher ist folgende Variante, formuliert für stochastische Prozesse. Der Beweis ist identisch.

Proposition 1.19. *Seien $X = (X_t)$ und $X^n = (X^n_t)$ stochastische Prozesse mit Pfaden im Raum der cadlag-Funktionen. Dann ist äquivalent:*

i) X^n ist in Verteilung gegen X konvergent,

ii) Die Folge der Verteilungen von X^n ist relativ kompakt und es gibt eine Teilmenge L von $\mathbb{R}_+$ mit abzählbarem Komplement, so dass $(X^n_{t_1},\dots,X^n_{t_k})$ für alle $t_1,\dots,t_k \in L$ in Verteilung gegen $(X_{t_1},\dots,X_{t_k})$ konvergiert.

Es bleibt, relative Kompaktheit in $\mathcal{P}\big(D_S\big)$ zu charakterisieren. Dies gelingt mithilfe der Stetigkeitsmodule von cadlag-Funktionen.

Satz 1.20. *Sei S vollständig. Dann ist $\mathcal{K} \subset \mathcal{P}\big(D_S\big)$ genau dann relativ kompakt, wenn gilt:*

i) $\{\pi_t(\mu) : \mu \in \mathcal{K}\}$ ist für alle $t \geq 0$ relativ kompakte Teilmenge von $\mathcal{P}(S)$.

ii) Für alle $\varepsilon, t > 0$ gibt es ein $\delta > 0$, so dass $\mu\big(\{f : \omega_{t,\delta}(f) \geq \varepsilon\}\big) \leq \varepsilon$ für alle $\mu \in \mathcal{K}$.

Zunächst zeigen wir, dass der Ausdruck in Aussage ii) wohldefiniert ist.

Lemma 1.21. *$\{f : \omega_{u,\delta}(f) < \varepsilon\}$ ist offen in D_S, so dass $w_{u,\delta} : D_S \to \mathbb{R}$ eine Borel-messbare Funktion ist.*

Beweis. Sei $\omega_{u,\delta}(f) < \varepsilon$. Dann gibt es ein $\varepsilon' < \varepsilon$ und $0 = \tau_0 < \tau_1 < \dots < \tau_k$ mit $\tau_i - \tau_{i-1} > \delta$, $\tau_k > u$ und $d\big(f(s), f(t)\big) \leq \varepsilon'$ für $s,t \in [\tau_{i-1}, \tau_i)$. Sei weiter $d_S(g,f) < \eta$ und $\alpha \in A(\eta)$ so, dass $d\big(f(s), g(\alpha(s))\big) \leq \eta$ für alle $t \leq \eta^{-1} - 1$. Für $\tau'_i = \alpha(\tau_i)$ gilt dann $\tau'_k \geq \tau_k - \eta > u$ und $\tau'_i - \tau'_{i-1} \geq \tau_i - \tau_{i-1} - 2\eta > \delta$, falls η genügend klein ist. Für $s', t' \in \big[\tau'_{i-1}, \tau'_i\big)$ und $s := \alpha^{-1}(s'), t := \alpha^{-1}(t')$ folgt $s,t \in [\tau_{i-1}, \tau_i)$, also

$$d\big(g(s'), g(t')\big) \leq d\big(g(\alpha(s)), f(s)\big) + d\big(f(s), f(t)\big) + d\big(f(t), g(\alpha(t))\big) \leq 2\eta + \varepsilon' < \varepsilon$$

falls η klein genug ist. Insgesamt gilt $\omega_{u,\delta}(g) < \varepsilon$ und damit $U_\eta(f) \subset \{\omega_{u,\delta} < \varepsilon\}$. □

Beweis von Satz 1.20. Die Bedingungen sind hinreichend: Wir machen von Satz 1.8 ii) Gebrauch, arbeiten also mit der Billingsley-Metrik d_B. Sei $\varepsilon > 0$ und $u > 2/\varepsilon$. Nach Voraussetzung gibt es ein $\delta > 0$, so dass $\mu\left(\omega_{u,\delta} \geq \varepsilon/2\right) \leq \varepsilon/2$ für alle $\mu \in \mathcal{K}$. Wir setzen

$$G := \{f : \omega_{u,\delta}(f) < \varepsilon/2\} \cap \bigcap_{\frac{k}{n} \leq t} \left\{ f : f\Big(\frac{k}{n}\Big) \in U_{\varepsilon/2}(x) \text{ für ein } x \in E \right\} \quad ,$$

wobei E eine (noch genauer zu spezifizierende) endliche Teilmenge von S sei. n wählen wir nach (I.5) so groß, dass für alle $f \in G$

$$d_B\Big(f, \sum_{\frac{k}{n} \leq u} f\Big(\frac{k}{n}\Big) 1_{\left[\frac{k}{n}, \frac{k+1}{n}\right)}\Big) < \frac{\varepsilon}{2}$$

gilt. Für jedes $f \in G$ gibt es daher $x_k \in E$, so dass

$$d_B\Big(f, \sum_{k \le un} x_k 1_{[\frac{k}{n}, \frac{k+1}{n})}\Big) < \varepsilon \quad .$$

Mit anderen Worten: G wird im Raum der cadlag-Funktionen durch die endlich vielen ε-Umgebungen der Funktionen $\sum_{k \le un} x_k 1_{[k/n,(k+1)/n)}$ mit $x_k \in E$ überdeckt. Außerdem können wir nach Voraussetzung i) und Satz 1.8 ii) die endliche Menge E so groß wählen, dass für alle $\mu \in \mathcal{K}$ gilt:

$$\mu\Big(\Big\{f : f\Big(\frac{k}{n}\Big) \in U_{\varepsilon/2}(x) \text{ für ein } x \in E\Big\}\Big) = \pi_{k/n}(\mu)\Big(\bigcup_{x \in E} U_{\varepsilon/2}(x)\Big) \ge 1 - \frac{\varepsilon}{2(un+1)}$$

für alle $k \le un$. Es folgt $\mu(G^c) \le \varepsilon$ für alle $\mu \in \mathcal{K}$, so dass $\mathcal{K}$ nach Satz 1.8 ii) relativ kompakt ist.

Die Bedingungen sind auch notwendig:
Zu i): Ist $\mathcal{K}$ relativ kompakt, so gibt es nach Satz 1.8 ii) zu vorgegebenen $\varepsilon > 0$ Funktionen $f_1, \dots, f_r \in D_S$, so dass $\mu\big(\bigcup_{i=1}^r U_\varepsilon(f_i)\big) \ge 1 - \varepsilon$ für alle $\mu \in \mathcal{K}$. Da die f_i cadlag-Funktionen sind, sind zu vorgegebenem t die Mengen $\{f_i(s) : te^{-\varepsilon} \le s \le te^{\varepsilon}\}$ relativ kompakt in S und werden daher alle durch endlich viele Umgebungen $U_\varepsilon(y_j), j = 1, \dots, k$ überdeckt. Gilt nun $t < 1/\varepsilon$ und $d_B(g, f_i) < \varepsilon$, so folgt $d\big(g(t), f_i(\beta(t))\big) < \varepsilon$ für ein $\beta \in B(\varepsilon)$, insbesondere $te^{-\varepsilon} \le \beta(t) \le te^{\varepsilon}$. Es folgt $g(t) \in \bigcup_j U_{2\varepsilon}(y_j)$ und damit

$$\pi_t(\mu)\Big(\bigcup_{j=1}^k U_{2\varepsilon}(y_j)\Big) = \mu\Big(\pi_t^{-1}\Big(\bigcup_{j=1}^k U_{2\varepsilon}(y_j)\Big)\Big) \ge \mu\Big(\bigcup_{i=1}^r U_\varepsilon(f_i)\Big) \ge 1 - \varepsilon \quad .$$

Nach Satz 1.8 folgt also i).
Zu ii): Falls ii) nicht zutrifft, gibt es $\varepsilon, u > 0$ und $\mu_n \in \mathcal{K}$, so dass $\mu_n\left(\omega_{u,1/n} \ge \varepsilon\right) > \varepsilon$ für alle n. (μ_n) enthält nach Voraussetzung eine konvergente Teilfolge $\mu_{n'}$ mit Limes μ. Da nach Lemma 1.21 die Menge $\big\{f : \omega_{u,\delta}(f) \ge \varepsilon\big\}$ abgeschlossen ist, folgt für alle $\delta > 0$

$$\mu\left(\omega_{u,\delta} \ge \varepsilon\right) \ge \limsup_{n'} \mu_{n'}(\omega_{u,\delta} \ge \varepsilon) \ge \limsup_{n'} \mu_{n'}(\omega_{u,1/n'} \ge \varepsilon) \ge \varepsilon \quad .$$

Nach Lemma 1.14 gilt andererseits $\{\omega_{u,\delta} \ge \varepsilon\} \downarrow \emptyset$ für $\delta \downarrow 0$, so dass sich ein Widerspruch ergibt. □

Aufgabe. Sei μ_n auf dem Raum der cadlag-Funktionen schwach gegen μ konvergent. Dann gilt $\mu\big(C_S\big) = 1$ genau dann, wenn für alle $t, \delta > 0$

$$\mu_n\big(\big\{f : d(f(s), f(s-)) \ge \delta \text{ für ein } s \le t\big\}\big) \to 0 \quad .$$

Hinweis. $\big\{f : d(f(s), f(s-)) > \delta \text{ für ein } s < t\big\}$ und $\big\{f : d(f(s), f(s-)) \ge \delta \text{ für ein } s \le t\big\}$ sind offen bzw. abgeschlossen in D_S.

I.6 Der Satz von Donsker und Aldous' Kriterium

Der funktionale Grenzwertsatz von Donsker verallgemeinert den zentralen Grenzwertssatz für Summen unabhängiger Zufallsvariabler. Er besagt, dass man aus einer Irrfahrt durch Reskalieren im Limes eine Brown'sche Bewegung erhält.

Seien $Z_1, Z_2, \dots$ unabhängige, identisch verteilte, reellwertige Zufallsvariablen mit Erwartungswert 0 und endlicher Varianz σ^2. Mittels $S_0 := 0$ und $S_n := Z_1 + \dots + Z_n$ konstruieren wir Prozesse Y^n mit Pfaden in $D_S[0, \infty)$ gemäß

$$Y_t^n := \frac{1}{\sigma\sqrt{n}} S_{[nt]} \quad , \quad t \ge 0 \quad .$$

Satz 1.22. *Y^n konvergiert für $n \to \infty$ in Verteilung gegen eine standard Brown'sche Bewegung W.*

Beweis. Für $s < t$ ist $Y^n_t - Y^n_s$ Summe von $[nt] - [ns]$ unabhängigen Zufallsvariablen und hat folglich Varianz $\frac{1}{n}([nt]-[ns])$. Nach dem zentralen Grenzwertsatz gilt daher $Y^n_t - Y^n_s \xrightarrow{v} W_t - W_s$. Da Y^n wie W unabhängige Zuwächse hat, folgt nach Korollar 1.4 $(Y^n_{t_2} - Y^n_{t_1}, \ldots, Y^n_{t_k} - Y^n_{t_{k-1}}) \xrightarrow{v} (W_{t_2} - W_{t_1}, \ldots, W_{t_k} - W_{t_{k-1}})$ bzw. $\left(Y^n_{t_1} \ldots, Y^n_{t_k}\right) \xrightarrow{v} (W_{t_1}, \ldots, W_{t_k})$ für $0 \le t_1 < \cdots < t_k$. Nachzuweisen bleibt daher nach Proposition 1.19 die relative Kompaktheit der Verteilungen von Y^n. Wir machen dazu Gebrauch von der leicht zu überprüfenden Tatsache, dass S_n und $S_n^2 - \sigma^2 n$ Martingale sind. Also sind auch Y^n_t und $(Y^n_t)^2 - \frac{1}{n}[nt]$ Martingale, und zwar bzgl. der von den Y^n erzeugten Filtrierungen $\mathscr{F}^n$ (d.h. $\mathscr{F}^n_t$ sei das von den $Y^n_s, s \le t$ erzeugte Ereignisfeld). Sei $t, \delta, \theta > 0$ und τ eine beschränkte $\mathscr{F}^n$-Stoppzeit. Dann folgt nach dem Satz vom Stoppen von Martingalen

$$\begin{aligned}\mathbf{P}\{|Y^n_{\tau+\theta} - Y^n_\tau| \ge \eta\} &\le \eta^{-2}\mathbf{E}\left(Y^n_{\tau+\theta} - Y^n_\tau\right)^2 \\ &= \eta^{-2}\left(\mathbf{E}Y^2_{\tau+\theta} - \mathbf{E}Y^2_\tau\right) = \eta^{-2}\mathbf{E}\Big[\frac{1}{n}[n(\tau+\theta)] - \frac{1}{n}[n\tau]\Big] \le \eta^{-2}\Big(\theta + \frac{1}{n}\Big) \quad .\end{aligned}$$

Die Behauptung folgt daher aus Satz 1.20 und dem nachfolgenden Kriterium von Aldous. □

Das Kriterium von Aldous gibt für die Forderung an Stetigkeitsmodule aus Satz 1.20 eine wahrscheinlichkeitstheoretisch handhabbare hinreichende Bedingung. Im folgenden seien $X^n, n \in \mathbb{N}, S$-wertige stochastische Prozesse auf W-Räumen $(\mathscr{A}^n, \mathbf{P}^n)$ und $\mathscr{F}^n$ die durch

$$\mathscr{F}^n_t = \sigma\left(X^n_s, s \le t\right)$$

gegebene Filtrierung in $\mathscr{A}^n$.

Satz 1.23. *Gilt für alle $t > 0$, für jede Folge $\tau_n \le t$ von $\mathscr{F}^n$-Stoppzeiten und für jede Zahlenfolge $\theta_n \downarrow 0$*

$$\mathbf{P}^n\{d(X^n_{\tau_n+\theta_n}, X^n_{\tau_n}) \ge \eta\} \to 0 \qquad \text{für alle } \eta > 0 \quad ,$$

so gibt es für alle $t, \varepsilon > 0$ ein $\delta > 0$, so dass für alle n gilt

$$\mathbf{P}^n\{\omega_{t,\delta}(X^n) > \varepsilon\} \le \varepsilon \quad .$$

Beweis. i) Im ersten Schritt formen wir die Annahme des Satzes geeignet um. Seien $\delta, \eta > 0$ und $\tau \le \tau'$ zwei Stoppzeiten für den Prozess X (wir unterdrücken den Index n). Sei weiter U eine von X unabhängige, uniform auf $[0,1]$ verteilte Zufallsvariable. Dann gilt

$$\begin{aligned}&\mathbf{P}\{d(X_\tau, X_{\tau'}) \ge \eta, \tau' \le \tau+\delta, \tau \le t\} \\ &\quad\le \mathbf{P}\{d(X_\tau, X_{\tau'+\delta U}) \ge \eta/2, \tau' \le \tau+\delta, \tau \le t\} + \mathbf{P}\{d(X_{\tau'}, X_{\tau'+\delta U}) \ge \eta/2, \tau' \le t+\delta\} \\ &\quad= \mathbf{E}\Big[I_{\{\tau' \le \tau+\delta, \tau \le t\}} \int_0^1 I_{\{d(X_\tau, X_{\tau'+\delta\theta}) \ge \eta/2\}}\, d\theta\Big] + \mathbf{E}\Big[I_{\{\tau' \le t+\delta\}} \int_0^1 I_{\{d(X_{\tau'}, X_{\tau'+\delta\theta}) \ge \eta/2\}}\, d\theta\Big] \\ &\quad\le \mathbf{E}\Big[I_{\{\tau \le t\}} \int_0^2 I_{\{d(X_\tau, X_{\tau+\delta\theta}) \ge \eta/2\}}\, d\theta\Big] + \mathbf{E}\Big[I_{\{\tau' \le t+\delta\}} \int_0^1 I_{\{d(X_{\tau'}, X_{\tau'+\delta\theta}) \ge \eta/2\}}\, d\theta\Big] \\ &\quad\le \int_0^2 \big(\mathbf{P}\{d(X_{\tau\wedge t}, X_{\tau\wedge t+\delta\theta}) \ge \eta/2\} + \mathbf{P}\{d(X_{\tau'\wedge(t+\delta)}, X_{\tau'\wedge(t+\delta)+\delta\theta}) \ge \eta/2\}\big)\, d\theta \ .\end{aligned}$$

Nach Annahme folgt also für jede Zahlenfolge $\delta_n \to 0$ und beliebige $\mathscr{F}^n$-Stoppzeiten $\tau'_n \ge \tau_n$

$$\mathbf{P}^n\{d(X^n_{\tau_n}, X^n_{\tau'_n}) \ge \eta, \tau'_n \le \tau_n + \delta_n, \tau_n \le t\} \to 0 \quad .$$

ii) Sei nun $\varepsilon, t > 0$. Wir betrachten die $\mathscr{F}^n$-Stoppzeiten $0 = \tau^n_0 < \tau^n_1 < \cdots$, gegeben als

$$\tau^n_i := \inf\left\{s > \tau^n_{i-1} : d\big(X^n_s, X^n_{\tau^n_{i-1}}\big) > \varepsilon/2\right\} \quad ,$$

also $d\big(X^n_{\tau^n_{i-1}}, X^n_{\tau^n_i}\big) \geq \varepsilon/2$. Für die Ereignisse

$$A_{k,n,\delta} := \{\tau^n_k > t\} \cap \bigcap_{i=1}^k \{\tau^n_i > \tau^n_{i-1} + \delta \text{ oder } \tau^n_{i-1} > t\}$$

gilt dann $A_{k,n,\delta} \subset \{\omega_{t,\delta}(X^n) \leq \varepsilon\}$, so dass es langt, $\mathbf{P}^n\{A^c_{k,n,\delta}\} \leq \varepsilon$ zu zeigen, sofern δ ausreichend klein ist, wobei wir für die Wahl von k freie Hand haben. Nun sei nach Teil i) des Beweises $\delta > 0$ so klein gewählt, dass für genügend großes n und alle i

$$\mathbf{P}^n\{\tau^n_i \leq \tau^n_{i-1} + \delta, \tau^n_k \leq t\} \leq \mathbf{P}^n\{d(X^n_{\tau^n_{i-1}}, X^n_{\tau^n_i}) \geq \varepsilon/2, \tau^n_i \leq \tau^n_{i-1} + \delta, \tau^n_{i-1} \leq t\} \leq \varepsilon/4 \quad .$$

Dann folgt

$$\begin{aligned} t \cdot \mathbf{P}^n\{\tau^n_k \leq t\} &\geq \mathbf{E}^n\big[\tau^n_k ; \tau^n_k \leq t\big] = \sum_{i=1}^k \mathbf{E}^n\big[\tau^n_i - \tau^n_{i-1} ; \tau^n_k \leq t\big] \\ &\geq \delta \sum_{i=1}^k \mathbf{P}^n\{\tau^n_i - \tau^n_{i-1} > \delta, \tau^n_k \leq t\} \\ &= \delta \sum_{i=1}^k \Big(\mathbf{P}^n\{\tau^n_k \leq t\} - \mathbf{P}^n\{\tau^n_i \leq \tau^n_{i-1} + \delta, \tau^n_k \leq t\}\Big) \\ &\geq k\delta\Big(\mathbf{P}^n\{\tau^n_k \leq t\} - \frac{\varepsilon}{4}\Big) \end{aligned}$$

und damit für $k \geq 2t/\delta$

$$\mathbf{P}^n\{\tau^n_k \leq t\} \leq \frac{\varepsilon}{2} \quad .$$

Nach Teil i) des Beweises erreichen wir durch weiteres Verkleinern von δ, dass für alle i und genügend großes n

$$\mathbf{P}^n\{\tau^n_{i-1} \leq t, \tau^n_i \leq \tau^n_{i-1} + \delta\} \leq \frac{\varepsilon}{2k} \quad .$$

Es folgt $\mathbf{P}^n\{A^c_{n,k}\} \leq \varepsilon$ und damit $\mathbf{P}^n\{\omega_{t,\delta}(X^n) > \varepsilon\} \leq \varepsilon$ für genügend großes n. Indem wir δ weiter verkleinern, folgt die Behauptung schießlich für alle n. □

Kapitel II

Martingalprobleme

Literatur :

Dawson,D.: *Measure Valued Markov Processes*, in École d' Été de Probabilités de Saint-Flour XXI, 1993, LNM 1541.
Ethier,S., Kurtz,T.: *Markov Processes, Characterization and Convergence*, Wiley, 1986.
Rogers,C., Williams,D.: *Diffusions, Markov Processes, and Martingals*, Wiley, Vol.1, 1994, Vol.2, 1987.
Stroock,D., Varadhan,S.: *Multidimensional diffusion processes*, Springer 1979.

Markov'sche Prozesse lassen sich auf verschiedene Weise charakterisieren. Der klassische Weg stellt die (starke) Markov'sche Eigenschaft an den Anfang und in den Mittelpunkt. Man kann als Ausgangspunkt aber auch Martingaleigenschaften von Markov'schen Prozessen wählen. Am Ende erweisen sich (zumindest für Prozesse vom Feller'schen Typ) beide Zugänge als verschiedene Seiten ein und derselben Medaille, dennoch ist es nicht belanglos, wie man an die Sache herangeht. Der klassische Weg baut auf der Theorie der Halbgruppen von Operatoren auf (man vgl. etwa Rogers & Williams, Vol.1). Geht man dagegen von Martingaleigenschaften aus, so steht einem nicht nur die Martingaltheorie zur Verfügung, auch die Theorie der schwachen Konvergenz lässt sich effektiv einsetzen. Diese Vorgehensweise geht auf Stroock und Varadhan zurück, die so sehr allgemein multidimensionale Diffusionsprozesse haben konstruieren können. Heute ist diese Strategie unter dem Namen „Lösen von Martingalproblemen“ bekannt und erweist sich als besonders erfolgreich bei der Konstruktion von Markov'schen Prozessen mit eher unübersichtlichen Zustandsräumen.

II.1 Die infinitesimale Beschreibung stochastischer Prozesse durch Martingalprobleme

Einfache stochastische Prozesse wie die Brown'sche Bewegung lassen sich durch ihre endlichdimensionalen Verteilungen charakterisieren. Für Markov'sche Prozesse $X = (X_t)$ ist das allerdings nur ausnahmsweise möglich, die Verteilungen von $(X_{t_1}, \dots, X_{t_k})$ sind meistens nicht direkt zugänglich. Man versucht stattdessen, X lokal, d.h. durch Angabe der Verteilung ihrer infinitesimalen Zuwächse $X_{t+dt} - X_t$, zu charakterisieren.
Diese Vorgehensweise ist aus der Theorie der gewöhnlichen Differentialgleichung wohlbekannt. Sei etwa $X_t, t \geq 0$, eine differenzierbare Abbildung in den $\mathbb{R}^n$. Handelt es sich um einen Phasenfluss (dies ist in etwa das deterministische Analogon eines Markov'schen Prozesses, vgl. V.I.Arnold, *Gewöhnliche Differentialgleichungen*), so kann man ihn durch eine Differentialgleichung der Gestalt

$$X_{t+h} - X_t \ = \ b(X_t) \cdot h + o(h), \quad h \to 0$$

mit einem Vektorfeld $\mathbb{R}^n \ni x \longmapsto b(x) \in \mathbb{R}^n$ beschreiben. Gibt man sich also den Wert von X_t vor, so kennt auch denjenigen von X_{t+h} für „infinitesimales" h. Dagegen ist nur ausnahmsweise der Funktionsverlauf von X_t angebbar, und das Langzeitverhalten von X_t ist häufig schwer zu erfassen. (Wie üblich fassen wir t als Zeitparameter auf.) Dabei stellt sich die Frage, ob die Differentialgleichung den Phasenfluss bei gegebenen Startwert X_0 eindeutig festlegt. Dies hängt wesentlich von der Glattheit des Vektorfeldes $x \mapsto b(x)$ ab und ist nicht von vornherein gewährleistet.

Beispiel : Die Differentialgleichung $\frac{d}{dt}X_t = \sqrt{X_t}$ hat in $\mathbb{R}_1$ zwei Lösungen mit $X_0 = 0\ : X \equiv 0$ und $X_t = t^2/4$.

Auch für stochastische Prozesse ist die Frage, ob infinitesimale Bedingungen ihn in Verteilung eindeutig charakterisieren, von besonderer Bedeutung. Sei also X ein stochastischer Prozess auf einem W-Raum $(\mathfrak{A}, P)$, adaptiert an eine Filtration $\mathscr{F}$ und mit Pfaden in $D_S[0,\infty)$, wobei S ein separabler metrischer Raum sei. Ähnlich wie wir im deterministischen Fall den Zuwachs $X_{t+h} - X_t$ bei Vorgabe des Wertes von X_t infinitesimal angegeben haben, wollen wir um die Verteilung von $X_{t+h} - X_t$ bei Kenntnis der Vorgeschichte bis zum Zeipunkt t festlegen. Mathematisch gesprochen geht es um die bedingte Verteilung von $X_{t+h} - X_t$, gegeben $\mathscr{F}_t$. Man kann sie beschreiben durch die bedingten Erwartungen $\mathbf{E}\big[f(X_{t+h}) - f(X_t)|\mathscr{F}_t\big]$, wobei f eine genügend große Klasse $\mathcal{D}$ von beschränkten messbaren Funktionen auf S durchlaufe.

Wir machen nun den **Ansatz**, dass es für jedes $f \in \mathcal{D}$ eine Funktion $Gf : X \to \mathbb{R}$ gibt, so dass

$$\mathbf{E}\big[f(X_{t+h}) - f(X_t)|\mathcal{F}_t\big] \;=\; h \cdot Gf(X_t) + o(h) \quad \text{für } h \to 0 \;\; .$$

In diesem Ansatz stecken zwei Annahmen: Die wesentliche ist, dass $\mathbf{E}\big[f(X_{t+h}) - f(X_t)|\mathscr{F}_t\big]$ allein von X_t abhängt und von der weiteren Vorgeschichte unbeeinflusst ist. Diese „Gedächtnislosigkeit" der Prozesse wird später (Abschnitt II) die Markov'sche Eigenschaft zur Konsequenz haben. Die andere Annahme ist, dass Gf nicht von t abhängt, dass wir es also mit einer zeitlich homogenen Situation zu tun haben. Diese Einschränkung lässt sich ohne weiteres beheben.

Wir geben nun unserem Ansatz eine mathematisch griffigere Gestalt. Seien $t, h > 0$. Dann gilt für $t_k = t + \frac{k}{n}h$.

$$\begin{aligned}
\mathbf{E}\big[f(X_{t+h}) - f(X_t)|\mathscr{F}_t\big] \;&=\; \sum_{k=1}^{n} \mathbf{E}\Big[\mathbf{E}\big[f\left(X_{t_k}\right) - f\left(X_{t_{k-1}}\right)|\mathscr{F}_{t_{k-1}}\big]\Big|\mathscr{F}_t\Big] \\
&=\; \mathbf{E}\Big[\frac{h}{n}\sum_{k=1}^{n}\left(Gf\left(X_{t_{k-1}}\right) + o(h)\right)\Big|\mathscr{F}_t\Big] \;\; .
\end{aligned}$$

Der Grenzübergang $n \to \infty$ liefert (wir klären hier nicht, wann er zulässig ist)

$$\mathbf{E}\big[f(X_{t+h}) - f(X_t)|\mathscr{F}_t\big] \;=\; \mathbf{E}\Big[\int_t^{t+h} Gf\left(X_s\right) ds\big|\mathscr{F}_t\Big] \;\; ,$$

mit anderen Worten, der durch

$$M_f(t) = f(X_t) - \int_0^t Gf(X_s)ds$$

definierte Prozess M_f ist ein Martingal.

Dies eröffnet die Perspektive, auf der Theorie der Martingale aufzubauen. Dazu ist es günstig, dass die Martingale M_f die gleichen Pfadeigenschaften wie X haben. Um dies zu garantieren, nehmen wir an, dass $\mathcal{D}$ nur stetige Funktionen enthält.

Wegen der Linearität bedingter Erwartungen ist es keine Einschränkung $\mathcal{D}$ als Untervektorraum von $C_b(S)$ zu wählen. Die Zuordnung $f \mapsto Gf$ liefert dann eine lineare Abbildung

$$G : \mathcal{D} \to M_b(S)$$

in den Raum $M_b(S)$ der beschränkten Borel-Abbildungen $f : S \to \mathbb{R}$.

Es hat sich der Sprachgebrauch „X löst das Martingalproblem $(G, \mathcal{D})$" eingebürgert. Dazu zunächst ein paar Beispiele.

Beispiele.

1. Für eine **Phasenfluss** X im $\mathbb{R}^n$, beschrieben durch die Gleichung $dX = b(X)dt$, gilt für stetig differenzierbares f nach der Kettenregel

$$df(X) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(X) \cdot b_i(X)dt = \operatorname{grad} f(X) \cdot b(X)dt$$

   X ist also in trivialer Weise Lösung des Martingalproblems, das sich aus

$$Gf = b \cdot \operatorname{grad} f$$

   und (etwa) der Menge $\mathcal{D}$ aller stetig differenzierbaren f mit kompakten Träger zusammensetzt.

2. **Sprungprozesse.** Sei X ein Prozess, dessen Pfade konstant bis auf Sprungstellen seien. Genauer: Nimmt X_t den Wert x an, so gibt es im Zeitintervall $[t, t+h]$ mit Wahrscheinlichkeit $\lambda(x) \cdot h + o(h)$ einen nach dem W-Maß $p(x, dy)$ verteilten Sprung. Die Wahrscheinlichkeit für mehr als einen Sprung sei von der Größenordnung $o(h)$. Dies bedeutet, dass

$$\begin{aligned}\mathbf{E}\big[f(X_{t+h})|\mathscr{F}_t\big] &= \big(1 - h\lambda(X_t)\big)f(X_t) + h\lambda(X_t)\int p(X_t, dy)f(y) + o(h)\\ &= f(X_t) + h\int \lambda(X_t, dy)(f(y) - f(x)) + o(h)\end{aligned}$$

   mit $\lambda(x, dy) := \lambda(x)p(x, dy)$. Man kann daher Sprungprozesse als Lösungen von Martingalproblemen mit

$$Gf(x) = \int \lambda(x, dy)(f(y) - f(x))$$

   erhalten. Hier kann man $\mathcal{D} = C_b(S)$ wählen.

   Wichtig ist der Spezialfall, dass S abzählbar ist. Dann betrachtet man auf S die diskrete Metrik, bei der alle Funktionen $f : S \to \mathbb{R}$ stetig sind. Hier ist das Maß $\lambda(x, dy)$ durch seine Gewichte $\lambda(x, y)$ gegeben, mit $\sum_{y \in S} \lambda(x, y) < \infty$ und

$$Gf(x) = \sum_{y \in S} \lambda(x, y)(f(y) - f(x)) \ .$$

3. Für eine **standard Brown'sche Bewegung** $W = (W_t)_{t \geq 0}$ in $\mathbb{R}$ (der Startpunkt x darf beliebig sein) und für beschränktes, 2-mal stetig differenzierbares f mit beschränkter 2ter Ableitung

$$f(W_t) - \frac{1}{2}\int_0^t f''(W_s)\, ds$$

   ein Martingal. Dies folgt etwa aus der Itô-Formel. W löst also das Martingalproblem zu

$$Gf := \frac{1}{2} f''$$

   und der Menge $\mathcal{D}$ aller beschränkten, 2-mal stetig differenzierbaren f mit beschränkter zweiter Ableitung.

4. Ist W standard Brown'sche Bewegung mit Startpunkt $x \geq 0$, so bezeichnet man $B = |W|$ als eine **in Null reflektierte Brown'sche Bewegung** . Der Wertebereich ist $S = \mathbb{R}_+$. Ist $f\colon \mathbb{R}_+ \to \mathbb{R}$ 2-mal stetig differenzierbar, so kann man sie, sofern $f'(0) = 0$, durch $g(x) = f(|x|)$ zu einer 2-mal stetig differenzierbaren Funktion auf $\mathbb{R}$ fortsetzen (Übung!). Dann gilt $f''(x) = g''(|x|)$ und

$$f(B_t) - \frac{1}{2}\int_0^t f''(B_s)\, ds = g(W_t) - \frac{1}{2}\int_0^t g''(W_s)\ ds,$$

d.h. auch für die reflektierte Brown'sche Bewegung ist

$$Gf(x) \ := \ \frac{1}{2} f''(x)\ , \quad x \geq 0$$

die richtige Wahl, nun aber mit dem Definitionsbereich

$$\mathcal{D} = \left\{ f \in C^2(\mathbb{R}_+)\colon f, f'' \text{ sind beschränkt, } f'(0) = 0 \right\} \ .$$

Das Randverhalten von B in 0 drückt sich in $\mathcal{D}$ aus, und zwar in der Bedingung $f'(0) = 0$: Die kräftigen Oszillationen von B, die in der Null nur in positive Richtung wirken, dürfen sich dort in $f(B_t)$ nicht zu stark bemerkbar machen.

5. Ist W standard Brown'sche Bewegung mit Startpunkt $x \geq 0$ und

$$\tau := \inf\{t \geq 0 : W_t = 0\}$$

der Zeitpunkt des ersten Erreichens von Null, so heißt

$$X_t := W_{t\wedge\tau}\ , \quad t \geq 0\ ,$$

Brown'sche Bewegung mit Absorbtion in Null. Ist also $f : \mathbb{R}_+ \to \mathbb{R}$ beschränkt und zweimal stetig differenzierbar mit beschränkter 2ter Ableitung, so folgt nach dem Satz vom optionalen Stoppen von Martingalen, dass

$$f(X_t) - \frac{1}{2}\int_0^{t\wedge\tau} f''(X_s)\, ds = f(W_{t\wedge\tau}) - \frac{1}{2}\int_0^{t\wedge\tau} f''(W_s)\, ds$$

ein Martingal ist. Gilt nun noch $f''(0) = 0$, so ist dieser Ausdruck gleich

$$f(X_t) - \frac{1}{2}\int_0^t f''(X_s)\, ds\ .$$

Wiederum ist

$$Gf := \frac{1}{2} f''$$

für ein Martingalproblem die richtige Wahl, nun aber ist

$$\mathcal{D} = \left\{ f \in C^2(\mathbb{R}_+)\colon f, f'' \text{ sind beschränkt, } f''(0) = 0 \right\} \ .$$

die geeignete Wahl des Definitionsbereiches. □

Allgemein geht es um die Frage, ob sich ein durch einen Operator $G\colon \mathcal{D} \to M_b(S)$ gestelltes „Martingalproblem" lösen lässt, und welche Eigenschaften die Lösungen haben. Dabei besteht eine Lösung aus einem stochastischen Prozess X und einem W-Maß $\mathbf{P}$, das für die Martingaleigenschaft als Referenzmaß dient.

Definition. *Sei S ein separabler metrischer Raum.*

a) *Ein* **Martingalproblem** *besteht aus einem Vektorraum $\mathcal{D} \subset C_b(S)$ und einem linearen Operator $G\colon \mathcal{D} \to M_b(S)$. S* heißt der **Zustandsraum** und G der **Generator** oder **infinitesimale Erzeuger** des Martingalproblems.

b) *Sei X ein stochastischer Prozess mit Pfaden in $D_S[0,\infty)$ und $\boldsymbol{P}$ ein W-Maß (beides auf einem gemeinsamen Ereignisfeld). Wir sagen, das Paar $(X, \boldsymbol{P})$ löst das Martingalproblem $(G, \mathcal{D})$, falls für alle $f \in \mathcal{D}$*

$$M_f(t) \ := \ f(X_t) - \int_0^t Gf(X_s)ds \ , \quad t \geq 0$$

ein $\boldsymbol{P}$-Martingal ist, bzgl. der von X induzierten Filtrierung $\mathscr{F}_t^X := \sigma(X_s, s \leq t)$, $t \geq 0$ (es macht keinen Unterschied, wenn man die Filtrierung noch vervollständigt). Ist zusätzlich eine andere Filtration $\mathscr{F}$ gegeben, an die X adaptiert ist, so nennt man $(X, \boldsymbol{P})$ Lösung bzgl. $\mathscr{F}$, falls M_f für alle $f \in \mathcal{D}$ ein $\mathscr{F}$-Martingal ist. Die **Anfangsverteilung** *von $(X, \boldsymbol{P})$ ist die Verteilung μ von X_0:*

$$\mu(B) \ := \ \mathbf{P}\{X_0 \in B\} \quad \textit{für alle Borel'schen } B \subset S \ .$$

c) *Ein Martingalproblem heißt* **eindeutig** *, falls zwei Lösungen $(X, \boldsymbol{P})$ und $(Y, \boldsymbol{Q})$ mit der gleichen Anfangsverteilung insgesamt in Verteilung übereinstimmen:*

$$\mathbf{P}\{X \in \tilde{B}\} \ = \ \mathbf{Q}\{Y \in \tilde{B}\} \quad \textit{für alle Borel'schen } \tilde{B} \subset D_S[0,\infty) \ .$$

Ein Martingalproblem heißt **wohlgestellt***, falls es eindeutig ist und falls für jedes W-Maß $\mu \in \mathcal{P}(S)$ eine Lösung existiert, die μ als Anfangsverteilung besitzt.*

Für ein wohlgestelltes Martingalproblem bestimmt die Anfangsverteilung μ oder der Startpunkt x (also $\mu = \delta_x$) die Verteilung einer Lösung (X, P). Man hält dies fest, indem man dann für Wahrscheinlichkeiten und Erwartungswerte $\mathbf{P}_\mu\{\cdot\}$, $\mathbf{E}_\mu[\cdot]$ bzw. $\mathbf{P}_x\{\cdot\}$, $\mathbf{E}_x[\cdot]$ schreibt.

Bemerkungen. (i) Die Annahme $\mathcal{D} \subset C_b(S)$ treffen wir hier der Einfachheit halber, dann hat M_f von vornherein f.s. cadlag-Pfade (in der Literatur werden auch allgemeinere Martingalprobleme behandelt).

(ii) Sei $M = (M_t)_{t \geq 0}$ ein $\mathscr{F}$-Martingal mit cádlág-Pfaden. Dann kann man immer zu der Filtration $\mathscr{F}_+ = (\mathscr{F}_{t+})_{t \geq 0}$ übergehen, gegeben durch

$$\mathscr{F}_{t+} := \bigcap_{u > t} \mathscr{F}_u \ ,$$

ohne die Martingaleigenschaft zu verlieren. Nach dem Konvergenzsatz für Rückwärtsmartingale gilt nämlich für $s < t$

$$\mathbf{E}[M_t \mid \mathscr{F}_{s+}] = \lim_{u \downarrow s} \mathbf{E}[M_t \mid \mathscr{F}_u] = \lim_{u \downarrow s} M_u = M_s \text{ f.s.}$$

Beispiel: Interagierende Teilchensysteme. Anschaulich stellt man sich vor, dass jeder Punkt in $L = \mathbb{Z}^d$ einen von zwei Ausprägungen (0/1, besetzt/unbesetzt, gesund/infiziert,...) annimmt. Diese Konfiguration von Ausprägungen quer durch L ändert sich mit der Zeit, aber nicht unabhängig von Punkt zu Punkt, sondern gesteuert durch lokale Beeinflussung zwischen benachbarten Punkten. Den gesamten Zustand können wir als Element

$$x = \big(x(i)\big)_{i \in L} \ \text{ aus der Menge } S := \{0,1\}^L$$

auffassen, mit der Ausprägung $x(i)$ gleich 0 oder 1 an der Stelle $i \in L$. Wir betrachten hier allein den Fall, dass bei einer Zustandsänderung sich x nur an einer einzigen Stelle j verändert („geflippt“ wird). Dann geht x in $x^j = \big(x^j(i)\big)_{i \in L}$ über mit

$$x^j(i) := \begin{cases} 1 - x(i) \ , & \text{falls } i = j \ , \\ x(i) \ , & \text{falls } i \neq j \ . \end{cases}$$

Für $x \in S$ setzen wir

$$T_x := \{i \in L : x(i) = 1\} \ .$$

1. Der einfachere Fall eines solchen interagierenden Teilchensystems ist, dass sich immer nur endlich viele Punkte in L im Zustand 1 befinden. Dann kann man sich auf den abzählbaren Zustandsraum

   $$S' := \{x \in S : T_x \text{ ist endlich}\}$$

   zurückziehen, und man hat es mit einem Sprungprozess wie auch schon im obigen Beispiel 2 zu tun (die Metrik ist die diskrete). Der Generator hat hier die Gestalt

   $$Gf(x) = \sum_{j \in L} \lambda(x, x^j)(f(x^j) - f(x)) \ .$$

   Ein interessantes Beispiel ist der **Kontaktprozess** mit Ansteckungsrate $\lambda > 0$ und den Sprungraten

   $$\lambda(x, x^j) = \begin{cases} 1 \ , & \text{falls } x(j) = 1 \ , \\ \lambda \cdot \#\{k \in L : k \sim j, x(k) = 1\} \ , & \text{falls } x(j) = 0 \ . \end{cases}$$

   Dabei bedeutet $k \sim j$, dass k und j zwei unmittelbar benachbarte Punkte in L sind. Die Interpretation ist: Jeder Punkt in L kann gesund oder infiziert sein. Infizierte Punkte gesunden mit Rate 1; gesunde Punkte infizieren sich von Nachbarpunkten mit Rate λ, die Infektionsrate ist also proportional zu der Anzahl der infizierten Nachbarn.

   Ein anderes bekanntes Beispiel ist das **Wählermodell** mit Raten

   $$\lambda(x, x^j) = \#\{k \in L : k \sim j \text{ und } x(k) \neq x(j)\} \ .$$

   Hier übernimmt j die „Meinung" $x(k)$ eines Nachbarn k, die Rate ist proportional zu der Anzahl der Nachbarn, die anderer Meinung als j sind.

2. Bei interagierenden Teilchensystemen denkt man aber insbesondere auch an den Fall, dass der Zustandsraum die gesamte Menge $S = \{0,1\}^L$ ist, dass nun also auch unendlich viele Stellen i die Ausprägung $x(i) = 1$ haben dürfen. S wird dann mit der Produkttopologie versehen, die von der Metrik

   $$d(x, y) = \sum_{k=1}^{\infty} |x(i_k) - y(i_k)| 2^{-k}$$

   erzeugt wird, dabei bezeichnet $i_1, i_2, \ldots$ irgendeine Aufzählung aller Elemente von L). Nach dem Satz von Tychonov ist S dann eine kompakte Menge. Es gilt

   $$d(x_n, x) \to 0 \quad \Leftrightarrow \quad \forall i \in L \ \exists m \in \mathbb{N} \ \forall n \geq m : x_n(i) = x(i) \ .$$

   Daher ist jede Funktion $f : S \to \mathbb{R}$, für die $f(x)$ nur von endlich vielen $x(i)$ abhängt, eine stetige, beschränkte Funktion. Anders ausgedrückt: Für den Vektorraum

   $$\mathcal{D} := \big\{ f : S \to \mathbb{R} : \exists F \subset L \text{ endlich } \forall x, y \in S : x|_F = y|_F \Rightarrow f(x) = f(y) \big\}$$

   gilt

   $$\mathcal{D} \subset C_b(S) \ .$$

   $\mathcal{D}$ ist ein geeigneter Definitionsbereich für G. Nach dem Satz von Stone-Weierstrass ist nämlich $\mathcal{D}$ dicht in $C_b(S)$ bzgl. gleichmäßiger Konvergenz. Im Übrigen ist der Generator G durch dieselbe Formel wie oben im Fall 1 gegeben.

   In einem endlichen Teilfenster $F \subset L$ wird man in Modellen wie dem Kontaktprozess oder dem Wählermodell in endlicher Zeit immer noch nur endlich viele Zustandsänderungen beobachten. Global gesehen folgen die Zustandssprünge nun aber typischerweise nicht mehr in in diskreten Schritten aufeinander, sie bilden in der Zeit eine zwar abzählbare, aber dichte Teilmenge. □

Häufig verzichtet man bei der Angabe einer Lösung $(X, \mathbf{P})$ eines Martingalproblems auf die explizite Angabe der zugehörigen W-Maßes $\mathbf{P}$. Dies ist immer dann möglich, wenn (wie etwa bei einer Brown'schen Bewegung) klar ist, was die Verteilung von X ist. Ob nämlich $(X, \mathbf{P})$ Lösung einer Martingalproblems ist, hängt einzig von der Verteilung L_x von X (bzgl. $\mathbf{P}$) ab. Dies ist Konsequenz der folgenden Proposition.

Proposition 2.1. *Ein Martingalproblem $(G, \mathcal{D})$ wird genau dann von $(X, \boldsymbol{P})$ gelöst, falls für alle $f \in \mathcal{D}$, alle messbaren beschränkten $g : S^k \to \mathbb{R}$ $(k \in \mathbb{N})$ und alle $0 \le t_1 \le \ldots \le t_k \le t$ und $h > 0$ gilt*

$$\int \eta(X) d\boldsymbol{P} \;=\; 0$$

mit $\eta = \eta_{f,g,t_1,\ldots,t_k,t,h} : D_S[0,\infty) \to \mathbb{R}$, gegeben durch

$$\eta(X) \;:=\; \left(f(X_{t+h}) - f(X_t) - \int_t^{t+h} Gf(X_s)ds\right) g(X_{t_1}, \ldots, X_{t_k}) \quad .$$

Die Aussage bleibt bestehen, wenn man sich auf stetige $g : S^k \to \mathbb{R}$ beschränkt.

Da nach der Transformationsformel

$$\int \eta \circ X \, d\mathbf{P} \;=\; \int \eta(x) \, \mathbf{P}\{X \in dx\} \;=\; \int \eta \circ id(x) \, \mathbf{P}\{X \in dx\}$$

bedeutet dies also: $(X, \mathbf{P})$ löst genau dann das Martingalproblem, wenn dies für die Verteilung $L_X(dx) = \mathbf{P}\{X \in dx\}$ von X zutrifft in dem Sinne, dass (id, L_X) das Martingalproblem löst. Hier ist $id = (\pi_t)$ als der stochastische Prozess der Projektionsabbildungen aufzufassen.

Beweis. Ist $(X, \mathbf{P})$ Lösung, so gilt

$$\begin{aligned} \mathbf{E}\eta(X) \;&=\; \mathbf{E}\Big[g\left(X_{t_1}, \ldots, X_{t_k}\right) \cdot \mathbf{E}\big[f\left(X_{t+h}\right) - f\left(X_t\right) - \int_t^{t+h} Gf\left(X_s\right) ds \big| \mathscr{F}_t^X\big]\Big] \\ &=\; \mathbf{E}\Big[g\left(X_{t_1}, \ldots, X_{t_k}\right)\Big(\mathbf{E}\big[M_f(t+h)\big|\mathscr{F}_t^X\big] - M_f(t)\Big)\Big] \;=\; 0 \quad . \end{aligned}$$

Umgekehrt nehmen wir an, dass die Bedingung des Satzes für alle stetigen $g(x_1, \ldots, x_n)$ gilt. Durch ein Approximationsargument folgt, dass dann auch

$$\mathbf{E}\big[\big(M_f(t+h) - M_f(t)\big) I_{\{(X_{t_1},\ldots,X_{t_k}) \in O\}}\,\big] \;=\; 0$$

für alle offenen $O \subset S^k$ gilt, also

$$\mathbf{E}\Big[(M_f(t+h) - M_f(t))^+ I_{\{(X_{t_1},\ldots,X_{t_k}) \in O\}}\Big] \;=\; \mathbf{E}\Big[(M_f(t+h) - M_f(t))^- \, I_{\{(X_{t_1},\ldots,X_{t_k}) \in O\}}\Big]$$

Die Ereignisse $\big\{(X_{t_1}, \ldots, X_{t_k}) \in O\big\}$, $O \subset S^k$ offen, $t_1 \le \ldots \le t_k \le t$ bilden nun aber einen $\cap$-stabilen Erzeuger des von $X_s, s \le t$ erzeugten Ereignisfeldes $\mathscr{F}_t^X$. Nach dem Eindeutigkeitssatz für endliche Maße folgt daher

$$\mathbf{E}\big[(M_f(t+h) - M_f(t))^+ \cdot I_A\big] \;=\; \mathbf{E}\big[(M_f(t+h) - M_f(t))^- \cdot I_A\big]$$

bzw.

$$\mathbf{E}\big[(M_f(t+h) - M_f(t)) \cdot I_A\big] \;=\; 0$$

für alle $A \in \mathscr{F}_t^X$. M_f ist damit Martingal. □

II.2 Lösungseigenschaften von Martingalproblemen

Ein Vorteil der Definition einer Lösung eines Martingalproblems ist, dass sofort viele Martingale zur Verfügung stehen. Die folgende Proposition erweitert noch die Möglichkeiten.

Proposition 2.2. *Sei X Lösung des Martingalproblems $(G, \mathcal{D})$ und sei $q : \mathbb{R}_+ \to \mathbb{R}$ stetig differenzierbar. Dann ist auch für $f \in \mathcal{D}$*

$$M_{q,f}(t) := q(t)f(X_t) - \int_0^t (q(s)Gf(X_s) + q'(s)f(X_s))\, ds \ , \quad t \geq 0 \ ,$$

ein Martingal.

Beweis. Nach dem Satz von Fubini gilt

$$\begin{aligned}\int_0^t q(s)Gf(X_s)\, ds &= \int_0^t q(t)Gf(X_s)\, ds - \int_0^t \int_s^t q'(u)\, du\, Gf(X_s)\, ds \\ &= q(t)\int_0^t Gf(X_s)\, ds - \int_0^t q'(u) \int_0^u Gf(X_s)\, dsdu \\ &= q(t)\int_0^t Gf(X_s)\, ds - \int_0^t q'(u)(f(X_u) - M_f(u))\, du\end{aligned}$$

und folglich

$$M_{q,f}(t) = q(t)M_f(t) - \int_0^t q'(u)M_f(u)\, du$$

Da M_f ein Martingal ist, folgt für $s \leq t$

$$\mathbf{E}[M_{q,f}(t) \mid \mathscr{F}_s] = q(t)M_f(s) - M_f(s)\int_s^t q'(u)\, du - \int_0^s q'(u)M_f(u)\, du = M_{q,f}(s) \text{ f.s.}$$

Dies ist die Behauptung. □

Die Martingale ermöglichen es u.a., den Satz vom optionalen Stoppen ins Spiel zu bringen. Wir betrachten zwei Anwendungen. Die erste besagt kurz gesprochen, dass man Sprünge in den Pfaden nicht vorhersehen kann. Wir sagen, dass eine Stoppzeit T **angekündigt** (**prävisibel**) ist, falls es eine unendliche Folge $T_1 < T_2 < \cdots$ von Stoppzeiten gibt mit $T_n \uparrow T$ f.s. Mit T ist auch $T \wedge c$ angekündigt, eine geeignete Stoppfolge ist $T_n \wedge (c - n^{-1})$, $n \geq 1$.

Satz 2.3: Quasi-Linksstetigkeit. *Sei $\mathcal{D}$ dicht in $C_b(S)$. Dann gilt für jede angekündigte Stoppzeit T auf dem Ereignis $\{T < \infty\}$*

$$X_{T-} = X_T \text{ f.s.}$$

Insbesondere gilt $X_{t-} = X_t$ f.s. für alle reellen $t \geq 0$.

Beweis. Ohne Einschränkung sei $T \leq c$ f.s. (sonst gehe man zu $T \wedge c$ über und lasse dann $c \to \infty$ gehen). Sei $T_1 < T_2 < \cdots$ f.s. eine T ankündigende Stoppfolge und sei $\mathscr{F}_{T-}$ die von $\bigcup_{n\geq 1} \mathscr{F}_{T_n}$ erzeugte σ-Algebra. Nach dem Martingalkonvergenzsatz gilt dann

$$\mathbf{E}[f(X_T) \mid \mathscr{F}_{T_n}] \to \mathbf{E}[f(X_T) \mid \mathscr{F}_{T-}] \text{ f.s.}$$

Für $f \in \mathcal{D}$ gilt andererseits nach dem Stoppsatz für Martingale

$$\mathbf{E}[f(X_T) \mid \mathscr{F}_{T_n}] = f(X_{T_n}) + \mathbf{E}\Big[\int_{T_n}^T Gf(X_s)\, ds \mid \mathscr{F}_{T_n}\Big] \text{ f.s.}$$

Wir dünnen nun die Stoppfolge soweit aus, dass zusätzlich

$$\mathbf{E}\Big[\int_{T_n}^T |Gf(X_s)|\,ds\Big] \le \frac{1}{n^3}$$

gilt. Dann folgt nach der Markov-Ungleichung

$$\begin{aligned}\mathbf{P}\Big\{\Big|\mathbf{E}\Big[\int_{T_n}^T Gf(X_s)\,ds \mid \mathscr{F}_{T_n}\Big]\Big| \ge \frac{1}{n}\Big\} &\le n\mathbf{E}\Big[\Big|\mathbf{E}\Big[\int_{T_n}^T Gf(X_s)\,ds \mid \mathscr{F}_{T_n}\Big]\Big|\Big] \\ \le n\mathbf{E}\Big[\mathbf{E}\Big[\int_{T_n}^T |Gf(X_s)|\,ds \mid \mathscr{F}_{T_n}\Big]\Big] &\le n\mathbf{E}\Big[\int_{T_n}^T |Gf(X_s)|\,ds\Big] \le \frac{1}{n^2}\ .\end{aligned}$$

Nach dem Borel-Cantelli Lemma folgt

$$\mathbf{E}\Big[\int_{T_n}^T Gf(X_s)\,ds \mid \mathscr{F}_{T_n}\Big] \to 0 \text{ f.s.}$$

Außerdem gilt $f(X_{T_n}) \to f(X_{T-})$ f.s., und wir erhalten insgesamt

$$f(X_{T-}) = \mathbf{E}[f(X_T) \mid \mathscr{F}_{T-}] \text{ f.s.}$$

Da $\mathcal{D}$ dicht in $C_b(S)$, gilt diese Aussage sogar für alle $f \in C_b(S)$. Es folgt

$$\mathbf{E}[(f(X_T) - F(X_{T-}))^2 \mid \mathscr{F}_{T-}] = \mathbf{E}[f^2(X_T) - 2f(X_T)f(X_{T-}) + f^2(X_{T-}) \mid \mathscr{F}_{T-}] = 0 \text{ f.s.}$$

und $\mathbf{E}[(f(X_T) - F(X_{T-}))^2] = 0$ bzw. $f(X_T) = f(X_{T-})$ f.s. Da es nun eine abzählbare, separierende Teilmenge von stetigen, beschränkten Funktionen gibt, folgt auch $X_T = X_{T-}$ f.s. □

Als nächstes behandeln wir die starke Markoff'sche Eigenschaft. Sei nun $(G, \mathcal{D})$ ein wohlgestelltes Martingalproblem. Bezeichne $\widetilde{\mathscr{B}}$ die σ–Algebra der Borel–Mengen $\tilde{B}$ im Pfadraum $D_S[0,\infty)$. Dann ist die Abbildung von $S \times \widetilde{\mathscr{B}}$, gegeben durch

$$(x, \tilde{B}) \ \mapsto\ \Gamma(x, \tilde{B}) := \mathbf{Q}_x\{Y \in \tilde{B}\}$$

wohldefiniert, dabei bezeichne $(Y, \mathbf{Q}_x)$ eine Lösung des Martingalproblems zur Anfangsverteilung $\mu = \delta_x$ (ohne Einschränkung können wir später annehmen, dass Y nicht vom Startwert x abhängt).

Definition. *Ein wohlgestelltes Martingalproblem heißt* **regulär**, *falls $x \mapsto \Gamma(x, \tilde{B})$ messbar ist für alle $\tilde{B} \in \widetilde{\mathscr{B}}$, falls also Γ ein Übergangskern ist.*

Bemerkung. Typischerweise ist die Menge $\mathcal{L}$ der Verteilungen L von Lösungen (X, P) eines Martingalproblems $(G, \mathcal{D})$ eine Borelmenge im polnischen Raum $\mathcal{P}(D_S[0,\infty))$ aller W-Maße auf den cadlag-Funktionen (versehen mit der Prohorov-Metrik). Die Abbildung $L \mapsto \mu$, die der Verteilung $L \in \mathcal{P}(D_S[0,\infty))$ ihre Anfangsverteilung $\mu \in \mathcal{P}(S)$ zuordnet, ist stetig und damit Borelmessbar. Ist das Martingalproblem nun wohlgestellt, so ergibt dies eine Bijektion von $\mathcal{L}$ nach $\mathcal{P}(S)$. Nach einem bemerkenswerten Satz von Kuratowski ist in dieser Konstellation (eine bijektive, messbare Abbildung von einer Borelmenge in einem polnischen Raum auf einen polnischen Raum) auch die Umkehrabbildung Borelmessbar. Dies impliziert aber, dass das Martingalproblem regulär ist.

Es folgt das Hauptresultat über reguläre Martingalprobleme, die **starke Markov'sche Eigenschaft**.

Satz 2.4. *Sei $(X, \boldsymbol{P})$ beliebige Lösung eines regulären Martingalproblems $(G, \mathcal{D})$ bzgl. der Filtration $\mathscr{F}$ und sei τ eine endliche $\mathscr{F}$-Stoppzeit. Dann gilt für Borelsches $\tilde{B} \subset D_S[0, \infty)$*

$$\mathbf{P}\{X_{\tau+\cdot} \in \tilde{B} | \mathscr{F}_\tau\} \;=\; \Gamma(X_\tau, \tilde{B}) \quad \mathbf{P}\text{-fast sicher.}$$

Beweis. Ohne Einschränkung der Allgemeinheit können wir τ als beschränkt annehmen (sonst betrachte man $\tau \wedge c$ und führe ein Grenzübergang $c \to \infty$ durch. Dies ermöglich uns, den Satz vom optionalen Stoppen für Martingale zu benutzen.

Sei $F \in \mathscr{F}_\tau$ und $\mathbf{P}\{F\} > 0$. Dann wird durch die oben betrachteten Lösungen $(Y, \mathbf{Q}_x)$ gemäß

$$\mathbf{P}_1\{X \in \tilde{B}\} \;:=\; \frac{1}{\mathbf{P}(F)} \mathbf{E}\left[I_F \cdot \mathbf{Q}_{X_\tau}\{Y \in \tilde{B}\}\right]$$

ein W-Maß auf dem von X erzeugten Ereignisfeld definiert. Es folgt

$$\mathbf{E}_1 \eta(X) = \frac{1}{\mathbf{P}\{F\}} \mathbf{E}\Big[I_F \int \eta(Y)\, d\mathbf{Q}_{X_\tau}\Big] \;=\; 0 \quad ,$$

nach Proposition 2.1. erfüllt mit $(Y, \mathbf{P}_x)$ also auch $(X, \mathbf{P}_1)$ das Martingalproblem $(G, \mathcal{D})$. Zweitens betrachten wir das W-Maß

$$\mathbf{P}_2\{X \in \tilde{B}\} \;:=\; \frac{1}{\mathbf{P}\{F\}} \mathbf{E}\big[I_F \cdot I_{\{X_{\tau+\cdot} \in \tilde{B}\}}\big]$$

Nach dem Satz vom optionalen Stoppen folgt unter den Bedingungen von Proposition 2.1.

$$\begin{aligned} \mathbf{E}_2 \eta(X) \;&=\; \frac{1}{\mathbf{P}\{F\}} \mathbf{E}\big[I_F \cdot \eta(X_{\tau+\cdot})\big] \;=\; \frac{1}{\mathbf{P}\{F\}} \mathbf{E}\big[\mathbf{E}[\ldots | \mathscr{F}_{\tau+t}]\big] \\ &=\; \frac{1}{\mathbf{P}\{F\}} \mathbf{E}\big[I_F \cdot g\left(X_{\tau+t_1}, \ldots, X_{\tau+t_k}\right) \cdot \mathbf{E}[M_f(\tau+t+h) - M_f(\tau+t) | \mathscr{F}_{\tau+t}]\big] \;=\; 0 \quad . \end{aligned}$$

X ist also bzgl. $\mathbf{P}_1$ wie $\mathbf{P}_2$ Lösung des Martingalproblems. Außerdem stimmen die Anfangsverteilungen überein,

$$\mathbf{P}_1\{X_0 \in B\} \;=\; \frac{1}{\mathbf{P}\{F\}} \mathbf{E}\left[I_F \cdot \mathbf{Q}_{X_\tau}\{Y_0 \in B\}\right] \;=\; \frac{1}{\mathbf{P}\{F\}} \mathbf{E}\big[I_F \cdot I_{\{X_\tau \in B\}}\big] \;=\; \mathbf{P}_2\{X_0 \in B\} \quad .$$

Daher stimmen $(X, \mathbf{P}_1)$ und $(X, \mathbf{P}_2)$ insgesamt in Verteilung überein. Es folgt

$$\frac{1}{\mathbf{P}\{F\}} \mathbf{E}[I_F \cdot \Gamma(X_\tau, \tilde{B})] \;=\; \frac{1}{\mathbf{P}\{F\}} \mathbf{E}\big[I_F \cdot I_{\{X_{\tau+\cdot} \in \tilde{B}\}}\big] \quad .$$

Nach Definition der bedingten Erwartung ist dies gerade die Behauptung. □

Beispiel: Blumenthal's 01-Gesetz. Sei X Lösung eines regulären Martingalproblems und sei $A \in \mathscr{F}_{0+}$. Dann folgt für alle $x \in S$

$$\mathbf{P}_x\{A\} = 0 \text{ oder } \mathbf{P}_x\{A\} = 1 \; .$$

Zum Beweis gehen wir zur Filtration $\mathscr{F}_+$ über. Es gilt

$$\mathbf{P}_x\{A\} = \mathbf{E}_x[I_A^2] = \mathbf{E}_x[\mathbf{E}[I_A^2 \mid \mathscr{F}_{0+}]] = \mathbf{E}_x[I_A \mathbf{E}[I_A \mid \mathscr{F}_{0+}]]$$

Die (starke) Markoff'sche Eigenschaft, angewandt auf die Stoppzeit T, ergibt $\mathbf{E}[I_A \mid \mathscr{F}_{0+}] = \mathbf{E}_{X_T}[I_A] = \mathbf{E}_x[I_A]$ und es folgt

$$\mathbf{P}_x\{A\} = \mathbf{E}_x[I_A]^2 = \mathbf{P}_x\{A\}^2 \; .$$

II.3 Eindeutigkeit und Dualität von Martingalproblemen

Für die Frage der Eindeutigkeit eines Martingalproblems ist das folgende Resultat wesentlich.

Proposition 2.5. *Sei $(G, \mathcal{D})$ ein Martingalproblem mit der Eigenschaft, dass für zwei Lösungen $(X, \boldsymbol{P})$ und $(Y, \boldsymbol{Q})$ mit der gleichen Anfangsverteilung immer gilt*

$$\boldsymbol{P}\{X_t \in B\} \;=\; \boldsymbol{Q}\{Y_t \in B\} \quad \forall B \in \mathscr{B}_s \quad \forall t > 0 \quad .$$

Dann ist $(G, \mathcal{D})$ ein eindeutiges Martingalproblem.

Beweis. Zu zeigen ist $\mathbf{P}\{X_{s_1} \in B_1, \ldots, X_{s_j} \in B_j\} = \mathbf{Q}\{Y_{s_1} \in B_1, \ldots, Y_{s_j} \in B_j\}$ für alle Zeitpunkte $0 \le s_1 \le \cdots \le s_j$ und alle Borel'schen Mengen $B_1, \ldots, B_j \subset S$. Wir führen den Beweis per Induktion nach j. Für $j = 1$ ist dies gerade die Voraussetzung. Für den Induktionsschritt von j nach $j+1$ definieren wir den stochastischen Prozess $\bar{X}$ durch

$$\bar{X}_t \;:=\; X_{t+s_j} \;, \quad t \ge 0$$

und das W-Mass $\bar{\mathbf{P}}$ durch

$$\bar{\mathbf{P}}\{A\} \;:=\; \frac{\mathbf{P}\{X_{s_1} \in B_1, \ldots, X_{s_j} \in B_j, A\}}{\mathbf{P}\{X_{s_1} \in B_1, \ldots, X_{s_j} \in B_j\}} \;.$$

Für $t_1, \ldots, t_k \le t$ gilt dann

$$\begin{aligned} &\bar{\mathbf{E}}\Big[\Big(f\left(\bar{X}_{t+h}\right) - f\left(\bar{X}_t\right) - \int_t^{t+h} Gf\left(\bar{X}_u\right) du\Big) g\left(\bar{X}_{t_1}, \ldots, \bar{X}_{t_k}\right)\Big] \\ &\quad = \frac{1}{\mathbf{P}\{X_{s_1} \in B_1, \ldots\}} \mathbf{E}\Big[\Big(f\left(X_{t+h+s_j}\right) - f\left(X_{t+s_j}\right) - \int_{t+s_j}^{t+h+s_j} Gf\left(X_u\right) du\Big) \\ &\qquad\qquad \times g\left(X_{t_1+s_j}, \ldots, X_{t_k+s_j}\right) I_{\{X_{s_1} \in B_1, \ldots, X_{s_j} \in B_j\}}\Big] \end{aligned}$$

Nach Proposition 2.1 löst damit mit $(X, \mathbf{P})$ auch $(\bar{X}, \bar{\mathbf{P}})$ das Martingalproblem. Ähnliches gilt für $(\bar{Y}, \bar{\mathbf{Q}})$, das wir analog aus $(Y, \mathbf{Q})$ gewinnen. Nach Induktionsannahme haben aber $(\bar{X}, \bar{\mathbf{P}})$ und $(\bar{Y}, \bar{\mathbf{Q}})$ dieselbe Startverteilung:

$$\bar{\mathbf{P}}\{\bar{X}_0 \in B\} \;=\; \frac{\mathbf{P}\{X_{s_1} \in B_1, \ldots, X_{s_{j-1}} \in B_{j-1}, X_{s_j} \in B_j \cap B\}}{\mathbf{P}\{X_{s_1} \in B_1, \ldots, X_{s_j} \in B_j\}} \;=\; \bar{\mathbf{Q}}\{\bar{Y}_0 \in B\} \;.$$

Für $s_{j+1} \ge s_j$ folgt also nach Annahme $\bar{\mathbf{P}}\{\bar{X}_{s_{j+1}-s_j} \in B_{j+1}\} = \bar{\mathbf{Q}}\{\bar{Y}_{s_{j+1}-s_j} \in B_{j+1}\}$, und das heißt

$$\mathbf{P}\{X_{s_1} \in B_1, \ldots, X_{s_{j+1}} \in B_{j+1}\} \;=\; \mathbf{Q}\{Y_{s_1} \in B_1, \ldots, Y_{s_{j+1}} \in B_{j+1}\} \;.$$

□

Für die Eindeutigkeit brauchen also nur die 1-dimensionalen Verteilungen von X betrachtet werden. Eine dafür gebräuchliche Methode besteht in einer Dualitätsbetrachtung: Seien $(G, \mathcal{D})$ und $(G', \mathcal{D}')$ zwei Martingalprobleme auf den Zustandsräumen S und S' und sei

$$h : S \times S' \to \mathbb{R}$$

eine messbare Funktion mit der Eigenschaft, dass für alle $x \in S$, $y \in S'$

$$h(\cdot, y) \in \mathcal{D} \;\text{ und }\; h(x, \cdot) \in \mathcal{D}' \;.$$

Dann heißen die beiden Martingalprobleme **dual** bzgl. h, falls für zwei Lösungen X und Y beider Probleme, die zudem noch unabhängig sind, und für alle $t > 0$

$$\mathbf{E}[h(X_t, Y_0)] = \mathbf{E}[h(X_0, Y_t)]$$

gilt (vorausgesetzt, beide Erwartungswerte sind wohldefiniert).

Unter den vielen Anwendungen von Dualität ist hier die folgende von Interesse: Wenn es für das Martingalproblem $(G', \mathcal{D}')$ und für jedes $y \in S'$ eine Lösung Y mit $Y_0 = y$ f.s. existiert, dann ist für jede Lösung X des Problems $(G, \mathcal{D})$ der Ausdruck $E[h(X_t, y)]$ durch y und die Anfangsverteilung von X festgelegt, eben (da man X und Y o.E. als unabhängig annehmen darf) durch die Gleichung

$$\mathbf{E}[h(X_t, y)] = \mathbf{E}[h(X_0, Y_t)] \ .$$

Ist außerdem die Familie $f(\cdot, y)$, $y \in S'$, von Funktionen in $\mathcal{D}$ ausreichend groß, so ist damit auch die Verteilung von X_t für alle $t > 0$ eindeutig bestimmt, so dass nach der vorigen Proposition das Martingalproblem $(G, \mathcal{D})$ eindeutig ist. Auf diesem Wege werden wir später Eindeutigkeitsfragen bei interagierenden Teilchensysteme behandeln.

Zuvor klären wir, wie man Dualität anhand der Generatoren erkennt. Dazu zuächst eine heuristische Überlegung. Für $0 < s < t$ und ausreichend kleinem $h > 0$ gilt

$$\begin{aligned}
&\mathbf{E}[h(X_{s+h}, Y_{t-(s+h)})] - \mathbf{E}[h(X_s, Y_{t-s}] \\
&\qquad = \mathbf{E}[h(X_{s+h}, Y_{t-(s+h)}) - h(X_s, Y_{t-s-h})] - \mathbf{E}[h(X_s, Y_{t-s} - h(X_s, Y_{t-s-h})] \\
&\qquad = \mathbf{E}\Big[\int_s^{s+h} Gh(X_u, Y_{t-s-h})\, du - \int_{t-s-h}^{t-s} G'h(X_s, Y_v)\, dv\Big] \\
&\qquad = \mathbf{E}\Big[\int_s^{s+h} \big(Gh(X_u, Y_{t-s-h}) - G'h(X_s, Y_{t-u})\big)\, du\Big]
\end{aligned}$$

(Dabei steht $Gh(x, y)$ für $Gf_y(x)$ und $G'h(x, y)$ für $G'f_x(y)$ mit $f_x(y) = f_y(x) = h(x, y)$.) Gilt nun für alle $x \in S, y \in S'$

$$Gh(x, y) = G'h(x, y) \ ,$$

so ist zu erwarten, dass $\mathbf{E}[h(X_{s+h}, Y_{t-(s+h)})] - \mathbf{E}[h(X_s, Y_{t-s}] = o(h)$ für $h \to 0$ folgt, bzw.

$$\frac{d}{ds}\mathbf{E}[h(X_s, Y_{t-s})] = 0 \ .$$

Dies impliziert aber die Dualitätsgleichung.

Proposition 2.6. *Seien $(G, \mathcal{D})$ und $(G', \mathcal{D}')$ zwei Martingalprobleme mit Zustandsräumen S und S' und sei die beschränkte, messbare Funktion $f : S \times S' \to \mathbb{R}$ derart, dass für alle $x \in S$, $y \in S'$*

$$h(\cdot, y) \in \mathcal{D} \quad \textit{und} \quad h(x, \cdot) \in \mathcal{D}' \ .$$

Gilt dann

$$Gh(x, y) = G'h(x, y)$$

für alle $x \in S$, $y \in S'$, so folgt für zwei unabhängige Lösungen X und Y beider Probleme

$$\mathbf{E}[h(X_t, Y_0)] = \mathbf{E}[h(X_0, Y_t)] \ .$$

Beweis. Wir zeigen zuächst, dass $\mathbf{E}[h(X_t, Y_0) - h(X_0, Y_t)]$ stetig in t ist. Bezeichnet μ und ν die Anfangsverteilungen von X und Y, so folgt aufgrund von Unabhängigkeit

$$\begin{aligned}
&\mathbf{E}[h(X_t, Y_0) - h(X_0, Y_t)] \\
&\qquad = \mathbf{E}[h(X_t, Y_0) - h(X_0, Y_0)] - \mathbf{E}[h(X_0, Y_t) - h(X_0, Y_0)] \\
&\qquad = \int_{S'} \mathbf{E}[h(X_t, y) - h(X_0, y)]\ \nu(dy) - \int_S \mathbf{E}[h(x, Y_t) - h(x, Y_0)]\ \mu(dx) \ .
\end{aligned}$$

Weil X und Y Lösungen der Martingalprobleme sind, folgt

$$\begin{aligned}\mathbf{E}[h(X_t,&Y_0)-h(X_0,Y_t)]\\ &=\int_{S'}\int_0^t \mathbf{E}[Gh(X_s,y)]\,ds\,\nu(dy)-\int_S\int_0^t \mathbf{E}[G'h(x,Y_s)]\,ds\,\mu(dx)\\ &=\int_0^t \mathbf{E}[Gh(X_s,Y_0)-G'h(X_0,Y_s)]\ ds\ .\end{aligned}$$

Dies ergibt die behauptete Stetigkeit.

Ähnlich folgt für $u\ge 0$ unter Beachtung von $\int_0^u h(X_t,Y_{u-t})\,du=\int_0^u h(X_{u-t},Y_t)\,dt$

$$\begin{aligned}\int_0^u \mathbf{E}[h(X_t,&Y_0)-h(X_0,Y_t)]\,du\\ &=\int_0^u \mathbf{E}[h(X_t,Y_0)-h(X_t,Y_{u-t})]\,du-\int_0^u \mathbf{E}[h(X_0,Y_t)-h(X_{u-t},Y_t)]\,dt\\ &=-\int_0^u\int_t^u \mathbf{E}[G'h(X_t,Y_{s-t})]\,dsdt+\int_0^u\int_t^u \mathbf{E}[Gh(X_{s-t},Y_t)]\,dsdt\\ &=-\int_0^u\int_0^s \mathbf{E}[G'h(X_t,Y_{s-t})]\,dtds+\int_0^u\int_0^s \mathbf{E}[Gh(X_{s-t},Y_t)]\,dtds\\ &=\int_0^u\int_0^s \big(\mathbf{E}[Gh(X_t,Y_{s-t})]-\mathbf{E}[G'h(X_t,Y_{s-t})]\big)\,dtds\ .\end{aligned}$$

Nach Annahme verschwindet dieser Ausdruck. Differenzieren wir nun nach u, so folgt die Behauptung unter Benutzung der oben bewiesenen Stetigkeit. $\square$.

Beispiel: Selbstdualität des Kontaktprozesses. Wir erinnern:

$$T_x:=\{i\in L: x(i)=1\}\ \text{ für } x\in S=\{0,1\}^L\ .$$

Für $x,y\in S$ setzen wir

$$h(x,y):=\begin{cases}1\ , & \text{falls } T_x\cap T_y=\emptyset\ ,\\ 0 & \text{sonst}\ .\end{cases}$$

Es gilt

$$Gh(x,y)=\sum_{\substack{j\\ x(j)=1}}(h(x^j,y)-h(x,y))+\lambda\sum_{\substack{j\\ x(j)=0}}\sum_{\substack{k\sim j\\ x(k)=1}}(h(x^j,y)-h(x,y))\ .$$

Die Aussage

$$T_x\cap T_y=\emptyset\quad\Leftrightarrow\quad T_{x^j}\cap T_y=\emptyset$$

ist im Fall $x(j)=1$ nur in dem Ausnahmefall $T_x\cap T_y=\{j\}$ ungültig, und im Fall $x(j)=0$ nur ungültig, sofern $T_x\cap T_y=\emptyset$ und $y(j)=1$ gilt. Es folgt

$$\sum_{\substack{j\\ x(j)=1}}(h(x^j,y)-h(x,y))=\begin{cases}1\ , & \text{falls } \#T_x\cap T_y=1\ ,\\ 0 & \text{sonst}\ ,\end{cases}$$

sowie

$$\sum_{\substack{(j,k)\\ j\sim k\\ x(j)=0,x(k)=1}}(h(x^j,y)-h(x,y))=\begin{cases}-\#\{(j,k): k\sim j, y(j)=x(k)=1\}\ , & \text{falls } T_x\cap T_y=\emptyset\ ,\\ 0 & \text{sonst}\ .\end{cases}$$

Diese Darstellungen sind völlig symmetrisch in x und y, wir erhalten also

$$Gh(x,y)=G'h(x,y)\quad \text{mit } G'=G\ .$$

Nun ist $h(x,y)$ nur von den Werten $x(i)$ mit $i \in T_y$ abhängig. Ist also T_y endlich, so ist $h(\cdot,y)$ stetig, genauer gilt $h(\cdot,y) \in \mathcal{D}$ und

$$\mathcal{D} = \mathrm{Lin}\{h(\cdot,y) : y \in S'\} \quad \text{mit } S' = \{y \in S : T_y \text{ ist endlich}\} \ .$$

Deswegen betrachten wir die beiden Martingalprobleme, die zu den Kontaktprozessen mit Zustandsräumen S und S' gehören. Entsprechend präzisieren wir den Definitionsbereich von h:

$$h : S \times S' \to \mathbb{R} \ .$$

Dann sind zwei Lösungen X und Y nach der Proposition 2.6 zueinander dual:

$$\mathbf{E}[h(X_t, y) = \mathbf{E}[h(X_0, Y_t)]$$

oder auch

$$\mathbf{P}(X_t \equiv 0 \text{ auf } T_{Y_0}) = \mathbf{P}(Y_t \equiv 0 \text{ auf } T_{X_0}) \ .$$

Insbesondere ist durch Y die Erwartungswerte $\mathbf{E}[f(X_t)]$ für alle $f \in \mathcal{D}$ festgelegt, und damit die Verteilung von X_t, denn $\mathcal{D}$ ist dicht in $C_b(S)$.

Damit wird die Frage der Eindeutigkeit des Martingalproblems für den Kontaktprozess zurückgespielt auf die Existenz einer Lösung im Fall endlich vieler Infizierter. □

Beispiel: Wählermodell. Sei S und S' wie im letzten Beispiel. Hier betrachten wir die Funktion

$$h : S \times S' \to \mathbb{R} \ ,$$

gegeben durch

$$h(x,y) := \prod_{\substack{i \\ y(i)=1}} x(i) \ ,$$

d.h. $h(x,y) = 1$, falls $T_y \subset T_x$. Erneut ist $h(x,y)$ nur von den $x(i)$ mit $i \in T_y$ abhängig, es gilt also wieder $h(\cdot,y) \in \mathcal{D}$ für $y \in S'$.

Nun gilt

$$Gh(x,y) = \sum_j \sum_{\substack{k \\ k \sim j \\ x(k) \neq x(j)}} (h(x^j, y) - h(x,y)) \ .$$

Für $y(j) = 0$ ist $h(x^j,y) = h(x,y)$, also

$$\begin{aligned}
Gh(x,y) &= \sum_{\substack{j \sim k \\ y(j)=1}} |x(j) - x(k)|(h(x^j,y) - h(x,y)) \\
&= \sum_{\substack{j \sim k \\ y(j)=1}} |x(k) - x(j)|(1 - 2x(j)) \prod_{\substack{i \neq j \\ y(i)=1}} x(i) \\
&= \sum_{\substack{j \sim k \\ y(j)=1}} (x(k) - x(j)) \prod_{\substack{i \neq j \\ y(i)=1}} x(i) \\
&= \sum_{\substack{j \sim k \\ y(j)=1}} \Big(x(k) \prod_{\substack{i \neq j \\ y(i)=1}} x(i) - h(x,y)\Big) \ .
\end{aligned}$$

Definieren wir $y^{(j,k)} \in S'$ durch die Eigenschaft

$$T_{y^{(j,k)}} = (T_y \setminus \{j\}) \cup \{k\}$$

so können wir

$$Gh(x,y) = \sum_{\substack{j \sim k \\ y(j)=1}} (h(x, y^{(j,k)}) - h(x,y)) = G'h(x,y)$$

schreiben, mit dem Generator

$$G'f(y) := \sum_{j\sim k} \lambda'(y, y^{(j,k)})(f(y^{(j,k)}) - f(y)) \ , \quad \lambda'(y, y^{(j,k)}) := y(j)$$

auf dem Zustandsraum S'. Dieses Modell heißt *verschmelzende Irrfahrten* auf L, man kann es sich so veranschaulichen: In den Punkten von T_y sitzen Teilchen, die sich unabhängig voneinander bewegen und in L zu einer rein zufälligen Nachbarstelle springen. Sitzt dort schon ein Teilchen, so verschmelzen beide, d.h. sie bewegen sich in Zukunft gemeinsam fort. Die Dualitätsrelation $\mathbf{E}[h(X_t, Y_0)] = \mathbf{E}[h(X_0, Y_t)]$ lautet hier

$$\mathbf{P}(X_t \equiv 1 \text{ auf } T_{Y_0}) = \mathbf{P}(X_0 \equiv 1 \text{ auf } T_{Y_t}) \ . \qquad \square$$

II.4 Schwache Konvergenz von Martingalproblemen

In diesem Abschnitt betrachten wir Martingalprobleme $(G, \mathcal{D})$, bei denen der Bildbereich von G aus stetigen Funktionen besteht. Solche Martingalprobleme vertragen sich gut mit dem Konzept schwacher Konvergenz von Prozessen.

Proposition 2.7. *Seien $G_n : \mathcal{D} \to M_b(S)$ und $G : \mathcal{D} \to C_b(S)$ lineare Operatoren. Seien weiter $(X^n, \boldsymbol{P}^n)$ Lösungen der Martingalprobleme $(G_n, \mathcal{D})$, die in Verteilung gegen $(X, \boldsymbol{P})$ konvergieren, also*

$$\lim_{n\to\infty} \int h(X^n) d\boldsymbol{P}^n \ = \ \int h(X) d\boldsymbol{P}$$

für alle stetigen, beschränkten Abbildungen $h : D_S[0,\infty) \to \mathbb{R}$. Dann löst $(X, \boldsymbol{P})$ das Martingalproblem $(G, \mathcal{D})$, sofern eine der beiden folgenden Bedingungen erfüllt ist:

i) $\lim_{n\to\infty} \sup_{x\in S} |G_n f(x) - Gf(x)| \ = \ 0$ *für alle $f \in \mathcal{D}$.*

ii) Für alle kompakten $K \subset S$ und alle $f \in \mathcal{D}$ gilt

$$\lim_{n\to\infty} \sup_{x\in K} |G_n f(x) - Gf(x)| \ = \ 0$$

und für alle $\varepsilon, t > 0$ gibt es eine kompakte Menge $K \subset S$ mit

$$\boldsymbol{P}^n\{X_s^n \in K \text{ für alle } s \le t\} \ \ge \ 1 - \varepsilon.$$

für alle $n \ge 1$.

Beweis. Sei $f \in \mathcal{D}, g \in C_b(S^k)$. Dann ist die Abbildung $\int_t^{t+h} Gf \circ \pi_s \, ds$ von $D_S[0,\infty)$ nach $\mathbb{R}$ stetig (vergleiche den Beweis von Proposition 1.16), und die Abbildung

$$\eta \ := \ \Big(f \circ \pi_{t+h} - f \circ \pi_t - \int_t^{t+h} Gf \circ \pi_s \, ds\Big) \cdot g \circ (\pi_{t_1}, \dots, \pi_{t_k})$$

von $D_S[0,\infty)$ nach S ist (bezüglich der Verteilung von X) fast sicher stetig, falls dies für $\pi_{t_1}, \dots, \pi_{t_k}, \pi_t$ und π_{t+h} gilt. Da X^n in Verteilung gegen X konvergiert, folgt gemäß Satz 1.1 iv) $\mathbf{E}[\eta(X)] \ = \ \lim_n \mathbf{E}^n[\eta(X^n)]$ bzw. nach Annahme i) oder ii)

$$\mathbf{E}[\eta(X)] \ = \ \lim_n \mathbf{E}^n[\eta_n(X^n)]$$

mit

$$\eta_n \ := \ \Big(f \circ \pi_{t+h} - f \circ \pi_t - \int_t^{t+h} G_n f \circ \pi_s \, ds\Big) \cdot g \circ (\pi_{t_1}, \dots, \pi_{t_k}) \ .$$

Nach Proposition 2.1 ergibt sich unter der zusätzlichen Annahme $t_1 \le \cdots \le t_k \le t$ die Gleichung $\mathbf{E}^n[\eta_n(X^n)] = 0$, also $\mathbf{E}[\eta(X)] = 0$. Da aber diejenigen t dicht liegen, für die π_t fast sicher stetig bzgl. der Verteilung von X ist (Proposition 1.17), und weil X f.s. cadlag-Pfade besitzt, folgt $\mathbf{E}[\eta(X)] = 0$ für alle $0 \le t_1 \le \cdots \le t_k \le t$ und für alle $h > 0$. Die Behauptung folgt daher aus Proposition 2.1. $\square$

Wir werden später Lösungen von Martingalproblemen per Grenzübergang konstruieren. Der folgende Satz ist also nicht nur als Konvergenzresultat zu verstehen, er enthält gleichzeitig eine Existenzaussage für das Martingalproblem $(G, \mathcal{D})$.

Satz 2.8. *Sei S vollständig, seien $G : \mathcal{D} \to C_b(S)$, $G_n : \mathcal{D} \to M_b(S)$, $n \in \mathbb{N}$, linear und seien $(X^n, \boldsymbol{P}^n)$ Lösungen der Martingalprobleme $(G_n, \mathcal{D})$. Es gelte:*

i) *Das Martingalproblem $(G, \mathcal{D})$ ist eindeutig,*

ii) *$\mathcal{D}(K)$ ist dicht in $C_b(K)$ für alle kompakten $K \subset S$, mit $\mathcal{D}(K) := \{f|_K : f \in \mathcal{D}\}$,*

iii) $\sup_{x \in K} \left| G_n f(x) - Gf(x) \right| \to 0$ *für alle kompakten $K \subset S$ und alle $f \in \mathcal{D}$,*

iv) *Für alle $\varepsilon, t > 0$ gibt es ein kompaktes $K \subset S$, so dass für alle n*

$$\boldsymbol{P}^n\{X_s^n \in K \text{ für alle } s \le t\} \;\ge\; 1 - \varepsilon.$$

Konvergiert dann X_0^n in Verteilung gegen ein $\mu \in \mathcal{P}(S)$, so konvergiert $(X^n, \boldsymbol{P}^n)$ in Verteilung gegen eine Lösung $(X, \boldsymbol{P})$ des Martingalproblems $(G, \mathcal{D})$ mit Anfangsverteilung μ.

Beweis. Zunächst haben die Verteilungen von $(X^n, \mathbf{P}^n)$ bzgl. schwacher Konvergenz höchstens einen Häufungspunkt: Konvergiert $(X^n, \mathbf{P}^n)$ entlang einer Teilfolge (n') gegen $(X, \mathbf{P})$, so löst $(X, \mathbf{P})$ nach Proposition 2.7 ii) das Martingalproblem $(G, \mathcal{D})$. Außerdem konvergiert $X_0^{n'}$ in Verteilung gegen X_0, da π_0 eine stetige Abbildung von D_S nach S ist. Daher hat X_0 die Verteilung μ. Da $(G, \mathcal{D})$ eindeutig ist, ist die Verteilung von $(X, \mathbf{P})$ eindeutig festgelegt.
Es bleibt daher die relative Kompaktheit der Verteilungen von $(X^n, \mathbf{P}^n)$ nachzuweisen. Wir wollen dazu die beiden Bedingungen aus Satz 1.20 nachprüfen. Die erste ist klar: Nach Voraussetzung *iii)* gilt $\mathbf{P}^n\left(X_t^n \in K\right) \ge 1 - \varepsilon$ für alle n, daher sind für alle $t \ge 0$ die Verteilungen von $X_t^n = \pi_t(X^n)$ relativ kompakt (vgl. Satz 1.8). Daher genügt es, die Bedingung aus Satz 1.23 nachzuprüfen. Zunächst gilt (den Index n unterdrücken wir erst einmal) nach dem Satz vom optionalen Stoppen von Martingalen für Stoppzeiten $\tau \le t$ und $\theta > 0$

$$\begin{aligned}\mathbf{P}\{|M_f(\tau+\theta) - M_f(\tau)| \ge \eta\} &\le \eta^{-2}\mathbf{E}[(M_f(\tau+\theta) - M_f(\tau))^2] \\ &= \eta^{-2}\mathbf{E}[M_f(\tau+\theta)^2 - M_f(\tau)^2] \;.\end{aligned}$$

Um den Erwartungswert abzuschätzen, wählen wir $m \in \mathbb{N}$ und setzen $t_i := \tau \wedge \frac{i}{m}$ für $i = 0, 1, \ldots$. Wegen $M_f(\tau)^2 - M_f(0)^2 = \sum_{i \le tm}(M_f(t_{i+1}) + M_f(t_i))(M_f(t_{i+1}) - M_f(t_i))$ gilt

$$\begin{aligned}&\mathbf{E}[M_f(\tau)^2] - \mathbf{E}[M_f(0)^2] \\ &\quad = \sum_{i \le tm} \mathbf{E}\Big[\Big(f(X_{t_{i+1}}) - \int_0^{t_{i+1}} Gf(X_s)\,ds + f(X_{t_i}) - \int_0^{t_i} Gf(X_s)\,ds\Big)(M_f(t_{i+1}) - M_f(t_i))\Big]\end{aligned}$$

Da nun mit M_f auch der gestoppte Prozess $M_f(\tau \wedge t)$, $t \ge 0$, ein Martingal ist, folgt

$$\begin{aligned}&\mathbf{E}[M_f(\tau)^2 - M_f(0)^2] \\ &\quad = \sum_{i \le tm} \mathbf{E}\Big[\Big(f(X_{t_{i+1}}) + f(X_{t_i}) - \int_{t_i}^{t_{i+1}} Gf(X_s)\,ds\Big)(M_f(t_{i+1}) - M_f(t_i))\Big] \\ &\quad = \sum_{i \le tm} \mathbf{E}\Big[\Big(f(X_{t_{i+1}}) + f(X_{t_i}) - \int_{t_i}^{t_{i+1}} Gf(X_s)\,ds\Big)\Big(f(X_{t_{i+1}}) - f(X_{t_i}) - \int_{t_i}^{t_{i+1}} Gf(X_s)\,ds\Big)\Big] \\ &\quad = \sum_{i \le tm} \mathbf{E}\Big[f(X_{t_{i+1}})^2 - f(X_{t_i})^2 - 2f(X_{t_{i+1}})\int_{t_i}^{t_{i+1}} Gf(X_s)\,ds + \Big(\int_{t_i}^{t_{i+1}} Gf(X_s)\,ds\Big)^2\Big] \\ &\quad = \sum_{i \le tm} \mathbf{E}\Big[f(X_{t_{i+1}})^2 - f(X_{t_i})^2 - 2\int_{t_i}^{t_{i+1}} f(X_s)Gf(X_s)\,ds + R_{i,m}\Big] \;,\end{aligned}$$

wobei für die Restterme $R_{i,m}$ Abschätzung

$$
\begin{aligned}
|\sum_{i\le tm} R_{i,m}| &= O\Big(\sum_{i\le tm}\Big(\int_{t_i}^{t_{i+1}} |f(X_s)-f(X_{t_{i+1}})|ds + \Big(\int_{t_i}^{t_{i+1}} Gf(X_s)\,ds\Big)^2\Big)\Big)\\
&= O\Big(\int_0^t |f(X_s)-f(X_{\lceil sm\rceil/m})|\,ds + \frac{1}{m}\Big)
\end{aligned}
$$

gilt. Für $m\to\infty$ konvergiert der Integrand punktweise gegen 0, und es folgt mit dominierter Konvergenz

$$
\mathbf{E}[M_f(\tau)^2 - M_f(0)^2] = \mathbf{E}\Big[f(X_\tau)^2 - f(X_0)^2 - 2\int_0^\tau f(X_s)Gf(X_s)\,ds\Big] + o(1)\ ,
$$

wobei der Restterm mit $m\to\infty$ verschwindet. Eine entsprechende Überlegung gilt für die Stoppzeit $\tau+\theta$, und wir erhalten

$$
\begin{aligned}
\mathbf{E}[M_f(\tau+\theta)^2 - M_f(\tau)^2] &= \mathbf{E}\Big[f(X_{\tau+\theta})^2 - f(X_\tau)^2 - 2\int_\tau^{\tau+\theta} f(X_s)Gf(X_s)\,ds\Big]\\
&= \mathbf{E}\Big[f(X_{\tau+\theta})^2 - g(X_{\tau+\theta}) - f(X_\tau)^2 + g(X_\tau) - \int_\tau^{\tau+\theta}(2f(X_s)Gf(X_s) - Gg(X_s))\,ds\Big]
\end{aligned}
$$

für beliebiges $g\in\mathcal{D}$. Nehmen wir nun noch an, dass der gesamte Pfad von X bis zum Zeitpunkt $\tau+\theta$ f.s. innerhalb einer kompakten Menge K verläuft, dann können wir nach Bedingung ii) zu vorgegebenem $\varepsilon>0$ die Funktion $g=g_{\varepsilon,K}$ so wählen, dass $\big|(f^2-g)(x)\big|\le\varepsilon/2$ für alle $x\in K$ gilt. Es folgt

$$
\mathbf{E}[M_f(\tau+\theta)^2 - M_f(\tau)^2]\ \le\ \varepsilon + \theta\sup_{x\in K}(2|fGf| + |Gg|)\quad .
$$

Sei nun $\tau_n\le t$ eine Folge von Stoppzeiten für die Prozesse X^n und θ_n Nullfolge, so dass sich die X^n bis zum Zeitpunkt $\tau_n+\theta_n$ innerhalb einer (von n unabhängigen) kompakten Menge K befinden. Dann erhalten wir insgesamt für $M_f^n = f(X^n) - \int G_n f(X^n)ds$

$$
\mathbf{P}^n\{\big|M_f^n(\tau_n+\theta_n) - M_f^n(\tau_n)\big|\ge\eta\}\ \le\ \frac{\varepsilon}{\eta^2} + \frac{1}{\eta^2}\theta_n\sup_{x\in K}(2|fG_nf| + |G_ng|)\ .
$$

Da G_nf und G_ng nach Annahme iii) auf K gleichmäßig gegen Gf und Gg konvergieren, gibt es ein $c>0$, so dass

$$
\mathbf{P}^n\{|M_f^n(\tau_n+\theta_n) - M_f^n(\tau_n)|\ge\eta\}\ \le\ \frac{\varepsilon}{\eta^2} + \frac{c}{\eta^2}\theta_n\quad .
$$

Die Grenzübergänge $n\to\infty$ und dann $\varepsilon\to 0$ führen zu

$$
\mathbf{P}^n\{|M_f^n(\tau_n+\theta_n) - M_f^n(\tau_n)|\ge\eta\}\ \to\ 0\quad .
$$

Aus der Abschätzung

$$
\begin{aligned}
&\mathbf{P}^n\{|f(X^n_{\tau_n+\theta_n}) - f(X^n_{\tau_n})|\ge\eta\}\\
&\qquad\le\ \mathbf{P}^n\Big\{|M_f^n(\tau_n+\theta_n) - M_f^n(\tau_n)|\ge\eta - \int_{\tau_n}^{\tau_n+\theta_n}\big|G_nf(X_s^n)\big|\,ds\Big\}\\
&\qquad\le\ \mathbf{P}^n\{|M_f^n(\tau_n+\theta_n) - M_f^n(\tau_n)|\ge\eta - c\theta_n\}
\end{aligned}
$$

folgt damit

$$
\mathbf{P}^n\{|f(X^n_{\tau_n+\theta_n}) - f(X^n_{\tau_n})|\ge\eta\}\ \to\ 0\quad .
$$

Da $\mathcal{D}(K)$ dicht in $C_b(K)$ ist, gilt diese Aussage sogar für alle $f\in C_b(S)$ (man wähle zu $f\in C_b(S)$ ein $g\in\mathcal{D}(K)$ mit $\sup_{x\in K}|f(x)-g(x)|\le\eta/4$).

Nun gibt es endlich viele $x_1, \dots, x_r \in K$, so dass $K \subset \bigcup_{i=1}^r U_{\eta/4}(x_i)$. Ist daher $y \in K$, $z \in S$, so folgt bei passender Wahl von $j = j_y$

$$d(y,z) \ \le\ d(y,x_j) + d(x_j,z) \ \le\ 2d(y,x_j) + |d(y,x_j) - d(z,x_j)| \ \le\ \eta/2 + \max_i |f_i(y) - f_i(z)|$$

mit $f_i(x) := d(x, x_i)$, also $f_i \in C_b(S)$, und damit im Limes $n \to \infty$

$$\mathbf{P}^n\{d(X^n_{\tau_n+\theta_n}, X^n_{\tau_n}) \ge \eta\} \ \le \sum_{i=1}^r \mathbf{P}^n\{|f_i(X^n_{\tau_n+\theta_n}) - f_i(X^n_{\tau_n})| \ge \eta/2\} \to 0 \quad .$$

Sei schließlich $\tau_n \le t$ eine beliebige Folge von Stoppzeiten. Mit σ_n bezeichnen wir die Austrittszeit von X^n aus der kompakten Menge K. Nach Annahme iv) können wir zu vorgegebenem $\varepsilon > 0$ das Kompaktum K so wählen, dass $\mathbf{P}\{\sigma_n \le t + \theta_n\} \le \varepsilon$ für alle n. Es folgt

$$\mathbf{P}^n\{d(X^n_{\tau_n+\theta_n}, X^n_{\tau_n}) \ge \eta\} \le \varepsilon + \mathbf{P}^n\{d(X^n_{\tau_n\wedge\sigma_n+\theta_n}, X^n_{\tau_n\wedge\sigma_n}) \ge \eta\} \ .$$

Von den Wahrscheinlichkeiten rechterhand haben wir gezeigt, dass sie eine Nullfolge bilden. Mit $\varepsilon \to 0$ folgt also wie gewünscht $\mathbf{P}^n\{d(X^n_{\tau_n+\theta_n}, X^n_{\tau_n}) \ge \eta\} \ \to\ 0.$ □

Bemerkung. In Anwendungen dieses Satzes liegt häufig folgende allgemeinere Situation vor: Neben der linearen Abbildung $G : \mathcal{D} \to C_b(S)$ hat man eine Folge von linearen Abbildungen $G_n : \mathcal{D}_n \to M_b(S_n)$, $n \ge 1$, mit Borelmengen $S_n \subset S$ und Definitionsbereichen $\mathcal{D}_n \subset C_b(S_n)$. (Man beachte, dass mit S in kanonischer Weise auch die S_n separable metrische Räume sind.) Die Prozesse X^n haben Werte in S_n und lösen die Martingalprobleme $(G_n, \mathcal{D}_n)$. Der Zusammenhang zwischen den Definitionsbereichen $\mathcal{D}$ und $\mathcal{D}_n$ ist durch die Forderung

$$\{f|_{S_n} : f \in \mathcal{D}\} \subset \mathcal{D}_n \tag{II.1}$$

gegeben, wobei $f|_{S_n}$ die Einschränkung von f auf S_n bezeichne. Die Annahme iii) des Satzes 2.8 ersetzen wir durch

iii') $\sup_{x \in K \cap S_n} |G_n f_n - Gf(x)| \to 0$ *für alle kompakten* $K \subset S$ *und alle* $f \in \mathcal{D}$,

mit $f_n := f|_{S_n}$, der Einschränkung von f auf S_n. Dann bleibt die Aussage von Satz 2.8 gültig.

Diese Aussage scheint viel allgemeiner zu sein, lässt sich jedoch unmittelbar auf Satz 2.8 zurückführen. Dazu ersetzen wir die G_n durch lineare Abbildungen $G'_n : \mathcal{D} \to M_b(S)$, gemäß

$$G'_n f(x) := \begin{cases} G_n f_n(x) & \text{für } x \in S_n \ , \\ Gf(x) & \text{für } x \notin S_n \ . \end{cases}$$

Ersetzen wir nun überall in Satz 2.8 G_n durch G'_n, so sind offenbar alle Annahmen des Satzes erfüllt, und die Behauptung folgt.

Der folgende Satz ist eine Variante von Satz 2.8, bei dem die approximierenden Prozesse aus Markov'schen Ketten gewonnen sind. Zur Erinnerung: Eine Markov'sche Kette $Z(0), Z(1), Z(2), \dots$ mit Werten in S und Übergangswahrscheinlichkeiten $P(z,B)$, $z \in S$, $B \subset S$ Borelsch, erfüllt die charakteristische Eigenschaft

$$\mathbf{P}\{Z(k+1) \in B | Z(0), \dots, Z(k)\} \ = \ P(Z(k), B) \quad \mathbf{P}\text{-fast sicher}$$

bzw.

$$\mathbf{E}\big[f(Z(k+1)\big|Z(0), \dots, Z(k)\big] \ = \ Pf(Z(k)) \quad \mathbf{P}\text{-fast sicher}$$

für $f \in M_b(S)$, mit

$$Pf(z) \ = \ \int P(z,dx) f(x) \quad .$$

Korollar 2.9. *Sei S vollständig, seien $G : \mathcal{D} \to C_b(S)$ und $G_n : \mathcal{D} \to M_b(S)$, $n \in \mathbb{N}$, linear und sei $Z^n = (Z^n(k))_{k\geq 0}$, $n \geq 1$, eine Folge von Markov'schen Ketten mit Werten in S und Übergangswahrscheinlichkeiten $P_n(z, B)$. Schließlich sei $(b_n)_{n\geq 1}$ eine divergente Folge positiver Zahlen. Es gelte:*

i) *Das Martingalproblem $(G, \mathcal{D})$ ist eindeutig,*

ii) *$\mathcal{D}(K)$ ist dicht in $C_b(K)$ für alle kompakten $K \subset S$,*

iii) $\sup_{x\in K} |b_n(P_n f(x) - f(x)) - Gf(x)| \to 0$ *für alle $f \in \mathcal{D}$ und alle kompakten $K \subset S$,*

iv) *für alle $\varepsilon, t > 0$ gibt es ein kompaktes $K \subset S$, so dass für alle n*

$$\boldsymbol{P}^n\{Z^n_k \in K \text{ für alle } k \leq tb_n\} \;\geq\; 1 - \varepsilon \ .$$

Falls dann die Verteilungen von Z^n_0 gegen ein W-Maß μ konvergieren, so konvergieren die durch

$$X^n_t \;:=\; Z^n([b_n t])\,, \quad t \geq 0 \ ,$$

gegebenen Prozesse X^n in Verteilung gegen eine Lösung des Martingalproblems $(G, \mathcal{D})$ mit Anfangsverteilung μ.

Beweis. Wir führen die Aussage durch „Poissonisieren“ auf Satz 2.8 zurück. Dazu sei $(N_t)_{t\geq 0}$ ein standard Poissonprozess, unabhängig von den Prozessen Z^n. Wir setzen

$$Y^n_t := Z^n(N_{b_n t}) \ , \quad t \geq 0 \ .$$

Dann erfüllt Y^n das Martingalproblem $(G_n, \mathcal{C}_b(S))$ mit

$$G_n f := b_n(Pf - f) \ , \ f \in \mathcal{C}_b(S) \ ,$$

und folglich auch das Martingalproblem $(G_n, \mathcal{D})$. Nach den Voraussetzungen sind alle Annahmen von Satz 2.8 erfüllt. Also konvergiert Y_n in Verteilung gegen eine Lösung $(X, \mathbf{P})$ des Martingalproblems $(G, \mathcal{D})$ mit Startverteilung μ.

Weiter haben wir $d_S(Y_n, X_n) \to 0$ in Wahrscheinlichkeit (in der Skorohod-Metrik d_S). Betrachten wir nämlich $\alpha_n(t) := N(b_n t)/b_n$, so gilt nach dem Gesetz der großen Zahlen $\sup_{t\leq u} |\alpha_n(t) - t| \to 0$ f.s. für alle $u > 0$, sowie $X^n_{\alpha_n(t)} = Z^n(N_{b_n t}) = Y^n_t$ für alle $t \geq 0$.

Wir greifen nun auf Satz 1.1 zurück. Sei $F \subset D_S$ eine abgeschlossene Menge im Skorohod-Raum und $\varepsilon > 0$. Dann gilt $\limsup_n \mathbf{P}^n\{Y^n \in F^\varepsilon\} \leq \mathbf{P}\{X \in F^\varepsilon\}$ aufgrund der Verteilungskonvergenz von Y^n gegen X, wobei F^ε hier die abgeschlossene ε-Umgebung von F in der Metrik d_S bezeichne. Es folgt

$$\limsup_n \mathbf{P}^n\{X^n \in F\} \leq \limsup_n \mathbf{P}^n\{Y^n \in F^\varepsilon\} + \lim_n \mathbf{P}^n\{d_S(X^n, Y^n) \geq \varepsilon\} \leq \mathbf{P}\{X \in F^\varepsilon\} \ .$$

Mit $\varepsilon \to 0$ folgt $\limsup_n \mathbf{P}^n\{X^n \in F\} \leq \mathbf{P}\{X \in F\}$ und damit nach Satz 1.1 die Behauptung. □

Bemerkung. Die Verallgemeinerung von Satz 2.8, wie in der anschließenden Bemerkung ausgeführt, bleibt entsprechend angepasst auch hier gültig. Eine Annahme wie in (II.1) erübrigt sich, sie ist automatisch erfüllt. Allein die Bedingung iii) des Korollars 2.9 muss umformuliert werden, sie lautet nun

iii') $\sup_{x\in K\cap S_n} |b_n(P_n f_n(x) - f(x)) - Gf(x)| \to 0$ für alle $f \in \mathcal{D}$ und alle kompakten $K \subset S$,

mit $f_n := f|_{S_n}$. Für den Beweis beziehen wir uns nicht mehr auf Satz 2.8, sondern eben auf die nachfolgende Bemerkung.

Feller'sche Martingalprobleme

Als Anwendung des Satzes charakterisieren wir die Martingalprobleme, deren Lösung stetig von der Anfangsverteilung abhängen.

Definition. *Ein Martingalproblem heißt* **Feller'sch**, *falls gilt: Sind* $(X^n, \boldsymbol{P}^n)$ *Lösungen mit Anfangsverteilungen* δ_{x_n}*, und gilt* $x_n \to x$ *in* S*, so konvergiert* $(X^n, \boldsymbol{P}^n)$ *in Verteilung gegen eine Lösung* $(X, \boldsymbol{P})$ *mit Anfangsverteilung* δ_x*.*

Natürlich muss dazu das Martingalproblem mindestens eindeutig sein. Der nächste Satz sagt, dass ein eindeutiges Martingalproblem in wichtigen Fällen Feller'sch ist, insbesondere, wenn S kompakt ist.

Korollar 2.10. *Sei* S *vollständig,* $\mathcal{D}$ *dicht in* $C_b(S)$*,* $G : \mathcal{D} \to C_b(S)$ *linear und das Martingalproblem* $(G, \mathcal{D})$ *eindeutig. Dann ist* $(G, \mathcal{D})$ *Feller'sch, wenn gilt: Ist* $(X^n, \boldsymbol{P}^n)$ *eine Folge von Lösungen mit Anfangsverteilungen* δ_{x_n}*, und ist* x_n *in* S *konvergent, so gibt es zu jedem* $\varepsilon, t > 0$ *ein kompaktes* $K \subset S$, so dass für alle n

$$\mathbf{P}^n\{X_s^n \in K \text{ für alle } s \leq t\} \ \geq \ 1 - \varepsilon \quad .$$

Die Bedingung ist auch notwendig (Übung). Der Beweis folgt unmittelbar aus Satz 2.8 mit der Wahl $G_n = G$.

Bemerkung: Jedes wohlgestellte Feller'sche Martingalproblem ist regulär. Der Raum aller beschränkten, Borel-messbaren $\varphi : D_S \to \mathbb{R}$ für die $x \mapsto \int \varphi(Y) d\mathbf{Q}_x$ ebenfalls Borel-meßbar ist, ist nämlich ein unter monotoner Konvergenz abgeschlossener Vektorraum, und für stetiges φ ist $x \mapsto \int \varphi(Y) d\mathbf{Q}_x$ bereits stetig.

II.5 Zwei Beispiele

A. Endlicher Zustandsraum

Sei S endlich. Mit der diskreten Metrik $d(x, y) = 1 - \delta_{xy}$ wird S zum kompakten Raum, und jedes $f : S \to \mathbb{R}$ ist stetig. Wir wählen $\mathcal{D} = C_b(S)$, dann lässt sich jede lineare Abbildung $G : \mathcal{D} \to C_b(S)$ durch eine Matrix $Q = (Q_{xy})$ beschreiben:

$$Gf(x) = \sum_y Q_{xy} f(y) \quad .$$

Dann ist $(G, \mathcal{D})$ ein eindeutiges Martingalproblem: Sei $(X, \mathbf{P})$ Lösung. Für $z \in S$ und $f = 1_{\{z\}}$ folgt

$$M_f(t) = I_{\{X_t = z\}} - \int_0^t \sum_x I_{\{X_s = x\}} Q_{xz} \, ds \quad .$$

Wegen $\mathbf{E} M_f(t) = \mathbf{E} M_f(0)$ folgt

$$\mathbf{P}\{X_t = z\} \ = \ \mathbf{P}\{X_0 = z\} + \sum_x Q_{xz} \int_0^t \mathbf{P}\{X_s = x\} \, ds \quad .$$

Durch Differenzieren geht sie in ein System linearer Differentialgleichungen für $\mathbf{P}\{X_t = z\}$ über, das bekanntlich bei gegebenen Anfangswahrscheinlichkeiten $\mathbf{P}\{X_0 = z\}$ eindeutig lösbar ist. Die Eindeutigkeit folgt daher nach Proposition 2.5.

Damit $(G, \mathcal{D})$ lösbar wird, muss Q ein paar Eigenschaften erfüllen. Wir leiten sie aus dem Sachverhalt ab, dass nach Proposition 2.2

$$e^{-\lambda t} f(X_t) - \int_0^t \big(Gf(X_s) - \lambda f(X_s)\big) e^{-\lambda s}\, ds \ , \quad \lambda \geq 0 \ ,$$

ein Martingal ist. Nach dem Satz vom optionalen Stoppen, angewandt auf den Zeitpunkt $\tau = \inf\{t > 0 : X_t \neq X_0\}$ des ersten Sprunges von X, folgt

$$\mathbf{E}\Big[e^{-\lambda\tau} f(X_\tau)\Big| X_0 = x\Big] - (Gf(x) - \lambda f(x))\, \mathbf{E}\Big[\frac{1-e^{-\lambda\tau}}{\lambda}\Big| X_0 = x\Big] \ = \ f(x) \quad .$$

Für $f = 1_{\{x\}}$ gilt $\mathbf{E}[e^{-\lambda\tau} f(X_\tau)\big|X_0 = x] = 0$ und $Gf(x) = Q_{xx}$, und man erhält

$$(\lambda - Q_{xx})\mathbf{E}\Big[\frac{1-e^{-\lambda\tau}}{\lambda}\Big| X_0 = x\Big] = 1$$

bzw.

$$\mathbf{E}\Big[e^{-\lambda\tau}\Big| X_0 = x\Big] \ = \ \frac{-Q_{xx}}{\lambda - Q_{xx}} \ , \quad \lambda \geq 0 \quad .$$

Für ausreichend großes λ ist der Nenner positiv. Daher muss jedenfalls $Q_{xx} \ \leq \ 0$ für alle $x \in S$ gelten. Im Fall $Q_{xx} = 0$ ist $\mathbf{E}[e^{-\lambda\tau}|X_0 = x] = 0$ und folglich $\tau = \infty$ f.s. Der Zustand x wird dann nie verlassen:

$$\mathbf{P}\{\tau = \infty \,|\, X_0 = x\} \ = \ 0 \qquad \text{für } Q_{xx} = 0 \quad .$$

Gilt $Q_{xx} < 0$, so ist $\frac{-Q_{xx}}{\lambda - Q_{xx}}$ die Laplace-Transformierte einer Exponentialverteilung mit Parameter $-Q_{xx}$. Gegeben den Startwert $X_0 = x$ ist also τ exponentialverteilt.
Wählen wir weiter $f = 1_{\{y\}}$ mit $y \neq x$, so folgt

$$\mathbf{E}\big[e^{-\lambda\tau} \cdot I_{\{X_\tau = y\}}|X_0 = x\big] \ = \ Q_{xy}\frac{1 - \mathbf{E}[e^{-\lambda\tau}|X_0 = x]}{\lambda}$$

und falls $Q_{xx} < 0$

$$\mathbf{E}\big[e^{-\lambda\tau} I_{\{X_\tau = y\}}|X_0 = x\big] \ = \ \frac{Q_{xy}}{-Q_{xx}}\mathbf{E}\big[e^{-\lambda\tau}|X_0 = x\big] \quad .$$

Für $\lambda = 0$ folgt

$$\mathbf{P}\{X_\tau = y|X_0 = x\} \ = \ \frac{Q_{xy}}{-Q_{xx}} \ , \quad y \neq x \quad .$$

Die Q_{xy} sind also, passend normiert, die Sprungwahrscheinlichkeiten. Daher muss $Q_{xy} \geq 0$ sowie $\sum_{y \neq x} \frac{Q_{xy}}{-Q_{xx}} = 1$ bzw. $\sum_y Q_{xy} \ = \ 0$ gelten.
Schießlich folgt für alle x, y.

$$\mathbf{E}\big[e^{-\lambda\tau} I_{\{X_\tau = y\}}\big|X_0 = x\big] \ = \ \mathbf{P}\{X_\tau = y|X_0 = x\} \cdot \mathbf{E}\big[e^{-\lambda\tau}|X_0 = x\big] \quad .$$

Diese Gleichung bedeutet: Gegeben $X_0 = x$ sind τ und X_τ voneinander unabhängig.

Insgesamt ergibt sich folgendes Bild: Zunächst ist X_0 nach der Anfangsverteilung zu wählen. Ist $X_0 = x$, so findet der erste Sprung nach einer exponentiellen Zeit τ mit Rate $\lambda(x) = -Q_{xx}$ statt. Die Verteilung des Sprunges ist unabhängig von τ gemäß den Übergangswahsscheinlichkeiten $-Q_{xy}/Q_{xx}$. Aufgrund der starken Markov'schen Eigenschaft können wir nun X_τ als neuen Startpunkt begreifen. Bis zum nächsten Sprung vergeht also wieder eine exponentielle Wartezeit, und deren Parameter wie die Sprungwahrscheinlichkeiten sind nun von X_τ abhängig. Dies setzt sich ad infinitum fort.

Für Q haben wir folgende Eigenschaften festgestellt:

$$Q_{xx} \leq 0 \, , \ Q_{xy} \geq 0 \quad \text{für } x \neq y \, , \quad \sum_y Q_{xy} = 0 \quad ,$$

man spricht dann von einer Q-Matrix. Wir zeigen abschließend, dass unter diesen Bedingungen das Martingalproblem lösbar ist. Dazu definieren wir die Matrizen

$$P_n \ := \ E + \frac{1}{n}Q \ \ ,$$

wobei E für die Einheitsmatrix steht. Es ist dann P_n für genügend großes n eine stochastische Matrix. Da S kompakt ist, sind die Bedingungen von Satz 2.9. in trivialer Weise erfüllt; wir können also eine Lösung der Martingalprobleme $(G, \mathcal{D})$ als Grenzwert von Markov'schen Ketten gewinnen. □

B. Der Fleming-Viot-Prozess

Der Moran Prozess. Dies ist ein Modell für die Änderungen, die eine Population durch genetische Mutation und Vererbung erfährt. Der Typenraum T (Typ = Farbe eines Individuums oder Allele eines Gens) wird als endlich angenommen, die Populationsgröße sei N. Das Modell berücksichtigt zwei Mechanismen für Änderungen des Typs:

i) **Mutationen** Ein Individuum ändert seinen Typ mit Rate $\alpha(x, y)$ von x nach y.

ii) **Vererbung** (sampling-replacement) Ein Individuum ändert mit Rate β seinen Typ in denjenigen eines rein zufällig aus der Population ausgewählten Individuums.

Es ist bequem, hier auch $y = x$ zuzulassen, und auch die Möglichkeit, dass das Individuum sich selbst wählt. Dann gibt es keine Veränderung. Es geht um die relativen Häufigkeiten der Typen in der Population, daher ist es natürlich, die Zustände im Raum der Häufigkeitsverteilungen

$$S_N \ := \ \Big\{ \frac{1}{N} \sum_{j=1}^{N} \delta_{x_j} \ : \ x_j \in T \Big\} \ \subset \ \mathcal{P}(T)$$

zu wählen. S_N ist endlich, daher kann der Prozess

$$X_t \ = \ \text{relative Häufigkeitsverteilung der Typen zur Zeit } t$$

nach dem Muster des letzten Abschnitts durch eine Q-Matrix $(Q(\mu, \nu))$ beschrieben werden. Für $\mu \in S$ ist dann die Rate einer Veränderung durch Mutation oder Vererbung von x nach y gegeben durch

$$Q\Big(\mu, \mu + \frac{\delta_y}{N} - \frac{\delta_x}{N}\Big) \ := \ N\alpha(x, y)\mu(\{x\}) + N\beta\mu(\{x\})\mu(\{y\}) \quad \text{falls } y \neq x \ \ .$$

Alle anderen Raten verschwinden, und $Q(\mu, \mu)$ ist so zu wählen, dass eine Q-Matrix entsteht. Eine Lösung (X_t) des zugehörigen Martingalprolems heißt Moran-Prozess und beschreibt die zeitliche Entwicklung der Typen in der Population.

Der Fleming-Viot Prozess Der Fleming-Viot-Prozess entsteht aus dem Moran-Prozess beim Grenzübergang $N \to \infty$. Es gibt verschiedene Varianten. Ein wesentlicher Aspekt des hier betrachteten Modells ist, dass bei jeder Mutation ein neuer Typ entsteht. Dazu muss der Typenraum überabzählbar sein; wir wählen ihn als das Intervall $[0, 1]$.

Darin eingebettet betrachten wir die endlichen Typenräume $T_N := \{\frac{j}{N} \ : \ j = 1, \ldots, N\}$ und die zugehörigen Moran-Prozesse mit Raten

$$\alpha_N(x, y) \ := \ \frac{1}{N} \,, \quad \beta_N \ := \ N \ \ .$$

(Mutationen treten also mit einer Gesamtrate $\sum_{x,y} N\alpha(x, y)\mu(\{x\}) = N$ ein, Vererbungen mit der Rate $\sum_{x,y} N\beta_N\mu\big(\{x\}\big)\mu\big(\{y\}\big) = N^2$.) Dann besitzen die Moran-Prozesse mit Typenräumen T_N im Grenzwert $N \to \infty$ einen nicht trivialen Grenzwert. Wir fassen sie nun als Prozesse auf dem (mit dem Prohorov-Abstand metrisierten) Raum

$$S \ := \ \mathcal{P}([0, 1])$$

aller Wahrscheinlichkeitsmaße auf $[0,1]$ auf, der alle S_N umfasst, und wollen auf diesem Raum die Konvergenz der Martingalprobleme beweisen. Als Definitionsbereich $\mathcal{D}$ wählen wir die Menge aller Abbildungen $f(\mu)$ der Gestalt

$$f(\mu) \;=\; p\Big(\int \varphi_1 d\mu, \dots, \int \varphi_r d\mu\Big) \;.$$

Dabei sei $r \in \mathbb{N}$, $\varphi_1, \dots, \varphi_r \in C_b([0,1])$ und $p(u_1, \dots, u_r)$ ein Polynom in r reellen Variablen. $\mathcal{D}$ ist eine Algebra, die die Punkte trennt (zu jedem Paar μ, ν gibt es ein stetiges, beschränktes φ mit $\int \varphi d\mu \neq \int \varphi d\nu$). Nach dem Satz von Stone-Weierstraß ist daher $\mathcal{D}$ dicht in $C_b(S)$.

Die Generatoren G_N für die Moran-Prozesse sind nun

$$\begin{aligned} G_N f(\mu) \;&=\; \sum_{\nu \in S_n} Q_N(\mu,\nu) f(\nu) = \sum_{\nu} Q_N(\mu,\nu)\big(f(\nu) - f(\mu)\big) \\ &=\; \sum_{x,y \in T_N} \mu\big(\{x\}\big)\Big(f\Big(\mu + \frac{\delta_y}{N} - \frac{\delta_x}{N}\Big) - f(\mu)\Big) \\ &\quad + N^2 \sum_{x,y \in T_N} \mu\big(\{x\}\big)\mu\big(\{y\}\big)\Big(f\Big(\mu + \frac{\delta_y}{N} - \frac{\delta_x}{N}\Big) - f(\mu)\Big) \quad . \end{aligned}$$

Für $f \in \mathcal{D}$ gilt nach der Taylor-Formel

$$\begin{aligned} f\Big(\mu + \frac{\delta_y - \delta_x}{N}\Big) - f(\mu) &= p\Big(\int \varphi_1 d\mu + \frac{\varphi_1(y) - \varphi_1(x)}{N}, \dots\Big) - p\Big(\int \varphi_1 d\mu, \dots\Big) \\ = \; \sum_i f_i(\mu) \frac{\varphi_i(y) - \varphi_i(x)}{N} &+ \frac{1}{2} \sum_{i,j} f_{ij}(\mu) \frac{\varphi_i(y) - \varphi_i(x)}{N} \frac{\varphi_j(y) - \varphi_j(x)}{N} + O\left(N^{-3}\right) \end{aligned}$$

mit

$$f_i(\mu) \;:=\; \frac{\partial p}{\partial u_i}\Big(\int \varphi_1 d\mu, \dots\Big) \;, \quad f_{ij}(\mu) \;:=\; \frac{\partial^2 p}{\partial u_i \partial u_j}\Big(\int \varphi_1 d\mu, \dots\Big) \quad .$$

Eingesetzt gibt dies

$$\begin{aligned} G_N f(\mu) \;&=\; \frac{1}{N} \sum_i f_i(\mu) \sum_{x,y} \big(\varphi_i(y) - \varphi_i(x)\big)\mu\big(\{x\}\big) \\ &\quad + \frac{1}{2} \sum_{i,j} f_{ij}(\mu) \sum_{x,y} \left(\varphi_i(y) - \varphi_i(x)\right)\left(\varphi_j(y) - \varphi_j(x)\right) \mu\big(\{x\}\big)\mu\big(\{y\}\big) + O\left(N^{-1}\right) \\ &=\; \sum_i f_i(\mu)\Big(\frac{1}{N} \sum_y \varphi_i(y) - \int \varphi_i d\mu\Big) \\ &\quad + \sum_{i,j} f_{ij}(\mu)\Big(\int \varphi_i \varphi_j d\mu - \int \varphi_i d\mu \int \varphi_j d\mu\Big) + O\left(N^{-1}\right) \end{aligned}$$

Es konvergiert also punktweise

$$G_N f \;\to\; Gf$$

wobei $G : \mathcal{D} \to C_b\big(S\big)$ definiert ist als

$$Gf \;:=\; \sum_i f_i(\mu)\Big(\int_0^1 \varphi_i(x)\, dx - \int \varphi_i d\mu\Big) + \sum_{i,j} f_{ij}(\mu)\Big(\int \varphi_i \varphi_j d\mu - \int \varphi_i d\mu \int \varphi_j d\mu\Big) \;.$$

Man prüft leicht nach, dass diese Konvergenz für alle $f \in \mathcal{D}$ gleichmäßig ist. Da mit $[0,1]$ auch $\mathcal{P}([0,1])$ kompakt ist (vgl. Satz 1.8), bleibt, um Satz 2.8. anwenden zu können, nur noch der Nachweis, dass das Martingalproblem $(G, \mathcal{D})$ eindeutig ist.

Sei dazu $\mathcal{D}_r$ die Menge aller $f \in \mathcal{D}$ von der Gestalt

$$f(\mu) \;=\; \int \varphi_1 d\mu \cdots \int \varphi_r d\mu \;,$$

also $p(u_1, \ldots, u_r) = u_1 \cdots u_r$. Für $f \in \mathcal{D}_r$ gilt

$$f(\mu) \;=\; f_i(\mu) \int \varphi_i d\mu \;=\; f_{ij}(\mu) \int \varphi_i d\mu \int \varphi_j d\mu \qquad \text{für } i \neq j$$

und $f_{ii}(\mu) = 0$, also

$$M_f(t) \;=\; f(X_t) - \int_0^t \Big(\sum_i f_i(X_s) \int_0^1 \varphi_i(x)\, dx + \sum_{i \neq j} f_{ij}(X_s) \int \varphi_i \varphi_j dX_s - r^2 f(X_s) \Big)\, ds \quad .$$

Aus $\mathbf{E} M_f(t) = \mathbf{E} M_f(0)$ ergibt sich durch Differenzieren nach t die lineare Differentialgleichung

$$\frac{d}{dt} \mathbf{E} f(X_t) \;=\; \sum_i \mathbf{E} f_i(X_t) \int_0^1 \varphi_i(x)\, dx + \sum_{i \neq j} \mathbf{E} \left(f_{ij}(X_t) \cdot \int \varphi_i \varphi_j dX_t \right) - r \mathbf{E} f(X_t) \quad .$$

Es ist $f_i \in \mathcal{D}_{r-1}$, $f_{ij} \in \mathcal{D}_{r-2}$, also $f_{ij}(\mu) \int \varphi_i \varphi_j d\mu$ aus $\mathcal{D}_{r-1}$. dies legt es nahe, die Gleichungen induktiv nach r zu lösen. Sei die Verteilung von X_0 gegeben. Für $r = 1$ ist $f_i \equiv 1$, $f_{ij} \equiv 0$, die Lösung der Differentialgleichung ist dann bekanntlich eindeutig durch ihren Anfangswert $\mathbf{E} f(X_0)$ festgelegt. Beim Schritt von $r-1$ nach r liegen $\mathbf{E} f_i(X_t)$ und $\mathbf{E}\left(f_{ij}(X_t) \int \varphi_i \varphi_j dX_t\right)$ schon fest, so dass erneut $\mathbf{E} f(X_0)$ die Lösung der Differentialgleichung eindeutig bestimmt. Da $\mathcal{D}$ von den $\mathcal{D}_r$ linear aufgespannt wird, ist $\mathbf{E} f(X_t)$ für alle $f \in \mathcal{D}$ eindeutig, und da $\mathcal{D}$ dicht ist, gilt dies schließlich für alle $f \in C_b(S)$. Daher bestimmt die Verteilung von X_0 eindeutig diejenige von X_t, und nach Proposition 2.5. ist das Martingalproblem eindeutig. Wegen Satz 2.8. können wir also feststellen, dass das Martingalproblem $(G, \mathcal{D})$ wohlgestellt ist.

Definition. *Eine Lösung des Martingalproblems $(G, \mathcal{D})$ heißt **Fleming-Viot-Prozess**.*

II.6 Halbgruppen von Operatoren

Bei Lösungen $(X, \mathbf{P})$ eines regulären Martingalproblems mit Werten in S handelt es sich, wie wir gesehen haben, um Markovprozesse, um Prozesse, die die Markovsche Eigenschaft erfüllen mit Übergangswahrscheinlichkeiten, gegeben durch Kerne $\Gamma_t = (\Gamma_t(x, dy))$, $t \geq 0$. Es gilt also für $t, h \geq 0$

$$\mathbf{P}(X_{t+h} \in B \mid \mathscr{F}_t) = \Gamma_h(X_t, B) \text{ f.s.}$$

Wir betrachten nun umgekehrt die Fragestellung, wie man für einen solchen Markovprozess X, der f.s. cadlag-Pfade besitzt, ein Martingalproblem generieren kann. Dazu betrachten wir die Halbgruppe $P = (P_t)_{t \geq 0}$ von linearen Operatoren

$$P_t : M_b(S) \to M_b(S) \;,$$

gegeben durch

$$P_t f(x) := \mathbf{E}_x[f(X_t)] = \int \Gamma_t(x, dy) f(y) \;.$$

Die Halbgruppeneigenschaft lautet

$$P_{s+t} = P_s P_t \quad \text{sowie} \quad P_0 = Id \;,$$

sie ergibt sich aus der Markoveigenschaft:

$$P_{s+t} f(x) = \mathbf{E}_x[f(X_{s+t})] = \mathbf{E}_x[\mathbf{E}_x[X_{s+t} \mid \mathscr{F}_s]] = \mathbf{E}_x[\mathbf{E}_{X_s}[f(X_t))]] = \mathbf{E}_x[P_t f(X_s)] = P_s(P_t f)(x) \;.$$

Es handelt sich um eine *positive* Halbgruppe, d.h.

$$f \geq 0 \quad \Rightarrow \quad P_t f \geq 0 \; ,$$

und um eine Halbgruppe von *Kontraktionen*, d.h.

$$\|P_t f\| \leq \|f\| \; , \quad \text{mit } \|f\| := \sup_x |f(x)| \; .$$

Um zusätzliche Glattheit zu erreichen, schränken wir die Operatoren ein auf den Vektorraum

$$\mathcal{V} := \{g \in M_b(S) : \text{ die Funktion } t \mapsto P_t g(x) \text{ ist stetig für alle } x \in S\} \; .$$

Mit $g \in \mathcal{V}$ gilt nämlich aufgrund der Halbgruppeneigenschaft auch $P_t g \in \mathcal{V}$. Für $g \in C_b(S)$ folgt für rechts- und linksseitige Limiten

$$\lim_{s \downarrow t} P_s g(x) = P_t g(x) \quad \text{ und } \quad \lim_{s \uparrow t} P_s g(x) = \mathbf{E}_x[g(X_{t-})] \; ,$$

da X f.s. cadlag-Pfade besitzt. Daher gilt $C_b(S) \subset \mathcal{V}$, falls $X_t = X_{t-}$ $\mathbf{P}_x$-f.s. Dies ist eine sehr schwache Bedingung, die alle vernünftigen Markovprozesse erfüllen. Wir betrachten daher in diesem Abschnitt Markovprozesse mit der Eigenschaft

$$\mathbf{P}_x(X_{t-} \neq X_t) = 0 \quad \text{für alle } x \in S, t > 0 \; . \tag{*}$$

Bemerkung: Feller'sche Prozesse. Gilt $P_t f \in C_b(S)$ für alle $f \in C_b(S)$, $t \geq 0$, so kann (und wird) man in den gesamten Überlegungen den Vektorraum $\mathcal{V}$ durch $C_b(S)$ ersetzen.

Für $g \in \mathcal{V}$ ist also $P_t g(x)$ Borel-messbar in x und stetig in t. Dies impliziert, dass die Funktion $(t,x) \mapsto P_t g(x)$ Borel-messbar ist. Wir können also Integrale der Gestalt $\int_0^\infty q(u) P_u g(x)\, du$ bilden, mit einer integrierbaren Funktion $q(u), u \geq 0$, und die resultierende Funktion in x ist Borel-messbar. Mit dem Satz von Fubini folgt

$$P_s \int_0^\infty q(u) P_u g(x)\, du = \int_0^\infty q(u) P_{s+u} g(x)\, du \; .$$

Der Integrand ist durch die integrierbare Funktion $|q(t)| \cdot \|g\|$ dominiert, indem wir also einen Grenzübergang $s \to t$ durchführen, erkennen wir, dass auch die Funktion $x \mapsto \int_0^\infty q(t) P_t g(x)\, dt$ zu $\mathcal{V}$ gehört. Insgesamt erhalten wir einen linearen Operator

$$\int_0^\infty q(u) P_u \, du : \mathcal{V} \to \mathcal{V} \; , \; \text{ mit } \Big(\int_0^\infty q(u) P_u \, du \Big) g(x) := \int_0^\infty q(u) P_u g(x)\, du \; .$$

Wir definieren nun den infinitesimalen Erzeuger der Halbgruppe.

Wir sagen, dass Funktionen $g_h \in \mathcal{V}$, $h > 0$, für $h \to 0$ *beschränkt konvergieren* mit Grenzwert $g \in M_b(S)$, falls die Funktionen punktweise gegen g konvergieren und falls es eine Konstante $c < \infty$ gibt, so dass $\sup |g_h| \leq c$ für ausreichend kleines h gilt. Wir schreiben dann

$$g_h \xrightarrow{b} g \;\; \text{für } h \to 0.$$

Definition. *Der* Generator (infinitesimale Erzeuger) *der Halbgruppe $P = (P_t)_{t \geq 0}$ ist definiert als der lineare Operator $A : \mathcal{D}_A \to \mathcal{V}$ mit Definitionsbereich*

$$\mathcal{D}_A := \big\{ f \in \mathcal{V} : \text{es gibt ein } g \in \mathcal{V} \text{ mit } \tfrac{1}{h}(P_h f - f) \xrightarrow{b} g \text{ für } h \to 0 \big\}$$

und

$$Af(x) := \lim_{h \to 0} \frac{1}{h} \big(P_h f(x) - f(x) \big) \; , \quad x \in S \; .$$

Der Operator A erfüllt das *Maximumprinzip*, d.h. für $f \in \mathcal{D}_A$ und $x_0 \in S$ gilt

$$f(x) \le f(x_0) \text{ für alle } x \in S \quad \Rightarrow \quad Af(x_0) \le 0 \ .$$

Der Beweis ergibt sich aus $P_h f(x_0) = \mathbf{E}_{x_0}[f(X_h)] \le f(x_0)$.

Wir werden klären, wann der infinitesimale Erzeuger die Halbgruppe eindeutig festlegt.

Proposition 2.11. *Für $f \in \mathcal{D}_A$ ist $P_t f \in \mathcal{D}_A$, und es gilt*

$$\frac{d}{dt} P_t f(x) = AP_t f = P_t Af \ .$$

Beweis. Es gilt

$$\frac{1}{h}(P_h P_t f(x) - P_t f(x)) = \mathbf{E}_x\big[\frac{P_h f(X_t) - f(X_t)}{h}\big] \ .$$

Wegen $f \in \mathcal{D}_A$ gilt für kleine $h > 0$ und ein $c > 0$

$$\Big|\frac{P_h f(X_t) - f(X_t)}{h}\Big| \le c \ \text{ und damit } \ \Big|\frac{1}{h}(P_h P_t f(x) - P_t f(x))\Big| \le c \ .$$

Mit dominierter Konvergenz folgt für $h \to 0$

$$\frac{1}{h}(P_h P_t f(x) - P_t f(x)) \xrightarrow{b} \mathbf{E}_x[Af(X_t)] = P_t Af(x) \ .$$

Wegen $Af \in \mathcal{V}$ gilt auch $P_t Af \in \mathcal{V}$. Es folgt also $P_t f \in \mathcal{D}_A$ und

$$AP_t f(x) = P_t Af(x) \ .$$

Gehen wir weiter in

$$\int_0^t \frac{1}{h}(P_{u+h} f(x) - P_u f(x))\, du = \frac{1}{h}\int_t^{t+h} P_u f(x)\, du - \frac{1}{h}\int_0^h P_u f(x)\, du$$

zum Grenzwert $h \to 0$ über, so folgt

$$\int_0^t P_u Af(x)\, du = P_t f(x) - f(x) \ ,$$

rechts aufgrund der Stetigkeit von $P_t f(x)$, und links mittels dominierter Konvergenz, dabei beachte man, dass der linke Integrand dem Absolutbetrag nach für kleine $h > 0$ durch c beschränkt ist. Die Behauptung folgt nun durch Differenzieren nach t, da $P_t Af(x)$ in t stetig ist. □

Wir können nun ein neues Martingalproblem formulieren.

Proposition 2.12. *Sei $f \in \mathcal{D}_A$. Dann ist durch*

$$M_f(t) := f(X_t) - \int_0^t Af(X_u)\, du \ , \quad t \ge 0 \ ,$$

ein Martingal M_f gegeben.

Beweis. Aufgrund der Markoveigenschaft gilt für $s \le t$

$$\begin{aligned}\mathbf{E}[M_f(t) \mid \mathscr{F}_s] &= \mathbf{E}[f(X_t) \mid \mathscr{F}_s] - \int_0^s Af(X_u)\, du - \int_s^t \mathbf{E}[Af(X_u) \mid \mathscr{F}_s] \\ &= P_{t-s} f(X_s) - \int_0^s Af(X_u)\, du - \int_s^t P_{u-s} Af(X_s)\, du \text{ f.s.}\end{aligned}$$

Nach Proposition 2.11 gilt $\int_s^t P_{u-s}Af(X_s)\,du = P_{t-s}f(X_s) - f(X_s)$, und es folgt die Behauptung. □

Es handelt sich sogar typischerweise um ein eindeutiges Martingalproblem $(A, \mathcal{D}_A)$. Dazu überzeugen wir uns erst einmal, dass der Definitionsbereich von A ausreichend groß ist. Wir konstruieren seine Elemente mithilfe der *Resolventen* $R_\lambda : \mathcal{V} \to \mathcal{V}$ von A, den linearen Operatoren

$$R_\lambda := \int_0^\infty e^{-\lambda u} P_u\,du \ , \quad \lambda > 0 \ .$$

Proposition 2.13. *Für $\lambda > 0$ gilt $R_\lambda = (\lambda - A)^{-1}$. Genauer:*

(i) *Für $g \in \mathcal{V}$ ist $R_\lambda g \in \mathcal{D}_A$ und $(\lambda - A)R_\lambda g = g$.*

(ii) *Für $f \in \mathcal{D}_A$ ist $Af \in \mathcal{V}$ und $R_\lambda(\lambda - A)f = f$.*

Beweis. (i) Für $g \in \mathcal{V}$ gilt

$$\begin{aligned}\frac{1}{h}(P_h R_\lambda g(x) - R_\lambda g(x)) &= \frac{1}{h}\int_0^\infty e^{-\lambda u} P_h P_u g(x)\,du - \frac{1}{h}\int_0^\infty e^{-\lambda u} P_u g(x)\,du \\ &= \frac{e^{\lambda h} - 1}{h}\int_0^\infty e^{-\lambda u} P_u g(x)\,du - \frac{e^{\lambda h}}{h}\int_0^h e^{-\lambda u} P_u g(x)\,du \ .\end{aligned}$$

Für $h \to 0$ konvergiert dieser Ausdruck gegen $\lambda R_\lambda g(x) - g(x)$. Auch folgt

$$\Big|\frac{1}{h}(P_h R_\lambda g(x) - R_\lambda g(x))\Big| \le \frac{e^{\lambda h} - 1}{\lambda h}\|g\| + e^{\lambda h}\|g\| \ ,$$

es handelt sich also um beschränkte Konvergenz. Schließlich ist mit g auch $R_\lambda g$ eine Element von $\mathcal{V}$, daher folgt insgesamt $R_\lambda g \in \mathcal{D}_A$ und $AR_\lambda g = \lambda R_\lambda g - g$ bzw.

$$(\lambda - A)R_\lambda g = g \ .$$

(ii) Umgekehrt gilt für $f \in \mathcal{D}_A$ definitionsgemäß $Af \in \mathcal{V}$. Wie im Beweis von Proposition 2.11 gilt $|h^{-1}(P_h P_u f(x) - P_u f(x)| \le c$ für alle $u \ge 0$ und ausrechend kleines h. Nach dem Satz von der dominierten Konvergenz folgt also

$$R_\lambda Af(x) = \lim_{h\to 0}\int_0^\infty e^{-\lambda u}\frac{P_h P_u f(x) - P_u f(x)}{h}\,du \ .$$

Indem wir P_h aus dem Integral ziehen, folgt

$$R_\lambda Af(x) = \lim_{h\to 0}\frac{P_h - Id}{h}\int_0^\infty e^{-\lambda u} P_u f(x)\,du \ .$$

Aus $f \in \mathcal{D}_A \subset \mathcal{V}$ folgt nach (i) $R_\lambda f \in \mathcal{D}_A$ und daher

$$R_\lambda Af(x) = AR_\lambda f(x)$$

bzw. mit (i) $R_\lambda(\lambda - A)f(x) = (\lambda - A)R_\lambda f(x) = f(x)$. □

Proposition 2.14. *Seien P, Q die Halbgruppen zweier Markovprozesse, die (*) erfüllen. Haben Sie dann denselben infinitesimalen Erzeuger A, so gilt $P = Q$.*

Beweis. Nach Proposition 2.13 gilt auch für die zugehörigen Resolventen Gleichheit, d.h. für $f \in C_b(S)$, $x \in S$

$$\int_0^\infty e^{-\lambda t} P_t f(x)\,dt = \int_0^\infty e^{-\lambda t} Q_t f(x)\,dt \ , \quad \lambda > 0 \ .$$

Da $P_t f(x)$, $Q_t f(x)$ in t stetige Funktionen sind, folgt nach dem Eindeutigkeitssatz für Laplacetransformierte $P_t f(x) = Q_t f(x)$. Dies impliziert $P_t = Q_t$ für $t \ge 0$. □

Wir kehren nun zum Martingalproblem $(A, \mathcal{D}_A)$ zurück und untersuchen dessen Eindeutigkeit. Aus der Eigenschaft, dass

$$M_f(t) = f(X_t) - \int_0^t Af(X_s)\, ds \ , \quad t \geq 0 \ ,$$

ein Martingal ist, folgt

$$\mathbf{E}[f(X_t)] = \mathbf{E}[f(X_0)] + \int_0^t \mathbf{E}[Af(X_s)]\, ds \ , \quad f \in \mathcal{D}_A \ .$$

Diese Gleichung verallgemeinert die Gleichung $P_t f(x) = f(x) + \int_0^t P_s Af(x)\, ds$ aus Proposition 2.11 auf beliebige Startverteilungen, sie heißt Vorwärtsgleichung. Aufgefasst als eine Gleichung für die Verteilung μ_t von X_t lautet sie bei einer Startverteilung μ

$$\mu_t f = \mu f + \int_0^t \mu_s(Af)\, ds \ , \quad f \in \mathcal{D}_A$$

mit $\mu f := \int f\, d\mu$. In Verallgemeinerung von Proposition 2.14 zeigen wir, dass ihre Lösung eindeutig durch μ bestimmt ist. Es folgt nämlich

$$\begin{aligned} \lambda \int_0^\infty e^{-\lambda t} \mu_t f\, dt &= \lambda \int_0^\infty e^{-\lambda t} \Big(\mu f + \int_0^t \mu_s(Af) \Big)\, ds\, dt \\ &= \mu f + \lambda \int_0^\infty \mu_s(Af) \int_s^\infty e^{-\lambda t}\, dt\, ds \\ &= \mu f + \int_0^\infty e^{-\lambda s} \mu_s(Af)\, ds \ , \end{aligned}$$

bzw. $\int_0^\infty e^{-\lambda t} \mu_t(\lambda f - Af)\, dt = \mu f$, $f \in \mathcal{D}_A$, oder nach Proposition 2.13

$$\int_0^\infty e^{-\lambda t} \mu_t g\, dt = \mu(R_\lambda g) \ , \quad g \in \mathcal{V} \ .$$

Nach dem Eindeutigkeitssatz für die Laplacetransformation ist also $\mu_t g$ wegen der Stetigkeit für alle $t \geq 0$ festgelegt. Gilt (*) und also $C_b(S) \subset \mathcal{V}$, so folgt, dass alle μ_t durch μ bestimmt sind.

Als Anwendung charakterisieren wir stationäre Verteilungen. Eine Anfangsverteilung μ von X heißt *stationär*, falls die Verteilung von X_t unabhängig von $t \geq 0$ ist.

Proposition 2.15. *Das W-Maß μ ist genau dann eine stationäre Verteilung von X, wenn*

$$\int Af\, d\mu = 0 \text{ für alle } f \in \mathcal{D}_A \ .$$

Beweis. Für stationäres μ gilt $\int P_t f\, d\mu = \mathbf{E}_\mu[f(X_t)] = \mathbf{E}_\mu[f(X_0)] = \int f\, d\mu$. Für $f \in \mathcal{D}_A$ daher durch beschränkte Konvergenz $\int Af\, d\mu = 0$. Umgekehrt hat unter der angegebenen Bedingung die Vorwärtsgleichung die Lösung $\mu_t = \mu$, und deren Eindeutigkeit ergibt die Stationarität. $\square$