4 Schwache Konvergenz (Version 16.1.2018)

4.1 Grundlagen

In diesem Kapitel sei (E,d) ein metrischer Raum. Die Borel-σ-Algebra auf E bezeichnen wir mit $\mathcal{B}(E)$, die Menge aller Wahrscheinlichkeitsmaße auf $(E,\mathcal{B}(E))$ mit $\mathcal{M}_1(E)$ und die Menge aller beschränkten stetigen Funktionen von E nach $\mathbb{R}$ mit $C_b(E)$. Der topologische *Rand* der Menge $A \subset E$ wird (wie üblich) mit ∂A bezeichnet. Zunächst zeigen wir, dass zwei Wahrscheinlichkeitsmaße gleich sind, wenn ihre Integrale über alle $f \in C_b(E)$ übereinstimmen (man beachte die Analogie zum Eindeutigkeitssatz für charakteristische Funktionen).

Proposition 4.1. *Seien $\mu,\nu \in \mathcal{M}_1(E)$. Gilt $\int f\,\mathrm{d}\mu = \int f\,\mathrm{d}\nu$ für alle $f \in C_b(E)$, so folgt $\mu = \nu$.*

Beweis. Sei F eine abgeschlossene Teilmenge von E. Für $m \in \mathbb{N}$ sei $f_m \in C_b(E)$ definiert durch $f_m(x) = (1 - md(x,F))^+$. Dann gilt mit dominierter Konvergenz

$$\mu(F) = \lim_{m\to\infty} \int f_m\,\mathrm{d}\mu = \lim_{m\to\infty} \int f_m\,\mathrm{d}\nu = \nu(F).$$

Da das System der abgeschlossenen Teilmengen von E ein durchschnittsstabiler Erzeuger von $\mathcal{B}(E)$ ist, folgt die Behauptung. $\square$

Definition 4.2. (a) Seien $\mu,\mu_n \in \mathcal{M}_1(E)$ für $n \in \mathbb{N}$. Die Folge (μ_n) konvergiert *schwach* gegen μ (Notation: $\mu_n \Rightarrow \mu$), wenn $\lim_{n\to\infty} \int f\mathrm{d}\mu_n = \int f\mathrm{d}\mu$ für alle $f \in C_b(E)$ gilt.

(b) Es seien X_n, $n \in \mathbb{N}$ und X $(E,\mathcal{B}(E))$-wertige Zufallsgrößen auf den Wahrscheinlichkeitsräumen $(\Omega_n,\mathcal{F}_n,\mathbb{P}_n)$ bzw. $(\Omega,\mathcal{F},\mathbb{P})$. Wenn $\mathbb{P}_n X_n^{-1} \Rightarrow \mathbb{P}X^{-1}$, so sagt man, die Folge $(X_n)_{n\in\mathbb{N}}$ konvergiert *in Verteilung* gegen X und schreibt oft $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$.

Bemerkung 4.3. Proposition 4.1 zeigt, dass schwache Grenzwerte eindeutig sind.

Der folgende Satz wir oft *Portmanteau-Theorem* genannt.

Satz 4.4. *Seien $\mu_n,\mu \in \mathcal{M}_1(E)$ für $n \in \mathbb{N}$. Die folgenden Aussagen sind äquivalent:*

(a) $\mu_n \Rightarrow \mu$.

(b) $\lim_{n\to\infty} \int f\mathrm{d}\mu_n = \int f\mathrm{d}\mu$ *für alle Lipschitz-stetigen $f \in C_b(E)$.*

(c) $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$ *für jede abgeschlossene Menge $F \subset E$.*

(d) $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ *für jede offene Menge* $G \subset E$.

(e) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *für alle* $A \in \mathcal{B}(E)$ *mit* $\mu(\partial A) = 0$.

Beweis.

(a)⇒(b) Das ist trivial.

(b)⇒(c) Sei F abgeschlossen und $f_m(x) := (1 - md(x, F))^+$, $m \in \mathbb{N}$. Dann ist f_m Lipschitz-stetig und beschränkt und

$$\limsup_{n\to\infty} \mu_n(F) \leq \lim_{n\to\infty} \int f_m \mathrm{d}\mu_n = \int f_m \mathrm{d}\mu.$$

Es gilt $f_m \downarrow 1_F$ für $m \to \infty$ und mit dominierter Konvergenz folgt $\lim_{m\to\infty} \int f_m \mathrm{d}\mu = \mu(F)$. Also folgt (c).

(c)⇔(d) Dies gilt, da das Komplement einer Menge genau dann offen ist, wenn die Menge abgeschlossen ist.

(c)+(d)⇒(e) Sei $A \in \mathcal{B}(E)$ mit $\mu(\partial A) = 0$. Sei $\mathring{A}$ das Innere von A und $\bar{A}$ der Abschluss von A. Dann gelten

$$\limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\bar{A}) \leq \mu(\bar{A})$$
$$\liminf_{n\to\infty} \mu_n(A) \geq \limsup_{n\to\infty} \mu_n(\mathring{A}) \geq \mu(\mathring{A}).$$

Nun gilt $0 = \mu(\partial A) = \mu(\bar{A}) - \mu(\mathring{A})$ und somit die Behauptung.

(e)⇒(c) Sei $F \subset E$ abgeschlossen und $F^\delta := \{x \in E : d(x, F) \leq \delta\}$ für $\delta > 0$. Man sieht leicht, dass $\partial(F^\delta) \subset \{x : d(x, F) = \delta\}$ (die Mengen sind nicht immer gleich!). Die Mengen $\partial(F^\delta)$ sind daher paarweise disjunkt (und als abgeschlossene Mengen messbar). Es gibt daher nur abzählbar viele $\delta > 0$ mit $\mu(\partial(F^\delta)) > 0$ und somit eine fallende Nullfolge $\delta_k > 0$ mit $\mu(\partial(F^{\delta_k})) = 0$ für alle $k \in \mathbb{N}$. Somit folgt $\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n(F^{\delta_k}) = \mu(F^{\delta_k})$ für alle $k \in \mathbb{N}$. Da $F^{\delta_k} \downarrow F$ (da F abgeschlossen ist), folgt mit dominierter Konvergenz $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$.

(c)⇒ (a) Sei $f \in C_b(E)$. Dann existieren $a > 0$ und $b \in \mathbb{R}$, so dass $\bar{f}(x) := af(x) + b \in (0, 1)$ für alle $x \in E$ gilt. Die Menge $F_i^{(k)} := \{x : \bar{f}(x) \geq i/k\}$ ist abgeschlossen für $k \in \mathbb{N}$ und $0 \leq i \leq k$. Für $\nu \in \mathcal{M}_1(E)$ und alle $k \in \mathbb{N}$ gilt daher

$$\sum_{i=1}^{k} \frac{i-1}{k} \big(\nu(F_{i-1}^{(k)}) - \nu(F_i^{(k)})\big) \leq \int \bar{f}\, \mathrm{d}\nu \leq \sum_{i=1}^{k} \frac{i}{k} \big(\nu(F_{i-1}^{(k)}) - \nu(F_i^{(k)})\big),$$

und daher (wegen $F_k^{(k)} = \emptyset$)

$$\frac{1}{k}\sum_{i=1}^{k-1} \nu(F_i^{(k)}) \leq \int \bar{f}\,\mathrm{d}\nu \leq \frac{1}{k}\sum_{i=0}^{k-1} \nu(F_i^{(k)}).$$

Daher folgt mit (c)

$$\limsup_{n\to\infty} \int \bar{f}\,\mathrm{d}\mu_n \leq \int \bar{f}\,\mathrm{d}\mu + 1/k.$$

Da k beliebig war und $a > 0$, folgt

$$\limsup_{n\to\infty} \int f\,\mathrm{d}\mu_n \leq \int f\,\mathrm{d}\mu.$$

Wendet man dieselbe Ungleichung auf $-f$ an, so folgt die Behauptung.

□

Das folgende Lemma liefert ein nützliches hinreichendes Kriterium für schwache Konvergenz.

Lemma 4.5. *Seien $\mu_n, \mu \in \mathcal{M}_1(E)$ für $n \in \mathbb{N}$. Sei $\mathcal{U}$ ein durchschnittsstabiles System in $\mathcal{B}(E)$ mit der Eigenschaft, dass jede offene Teilmenge von E als abzählbare Vereinigung von Mengen in $\mathcal{U}$ dargestellt werden kann. Gilt $\lim_{n\to\infty} \mu_n(U) = \mu(U)$ für alle $U \in \mathcal{U}$, so folgt $\mu_n \Rightarrow \mu$.*

Beweis. Für $m \in \mathbb{N}$ und $A_1, ..., A_m \in \mathcal{U}$ gilt

$$\begin{aligned}\mu_n\Big(\bigcup_{j=1}^m A_j\Big) &= \sum_{k=1}^m (-1)^{k+1} \sum_{j_1<j_2<...<j_k} \mu_n\big(A_{j_1} \cap A_{j_2} \cap ... \cap A_{j_k}\big)\\ &\to \sum_{k=1}^m (-1)^{k+1} \sum_{j_1<j_2<...<j_k} \mu\big(A_{j_1} \cap A_{j_2} \cap ... \cap A_{j_k}\big) = \mu\Big(\bigcup_{j=1}^m A_j\Big).\end{aligned}$$

Sei $G \subset E$ offen und $\varepsilon > 0$. Dann existieren nach Voraussetzung $m \in \mathbb{N}$ und $A_1, ..., A_m \in \mathcal{U}$ mit $\mu(G) \leq \mu(\bigcup_{j=1}^m A_j) + \varepsilon$ und $\bigcup_{j=1}^m A_j \subset G$. Somit ist

$$\liminf_{n\to\infty} \mu_n(G) \geq \liminf_{n\to\infty} \mu_n\Big(\bigcup_{j=1}^m A_j\Big) = \mu\Big(\bigcup_{j=1}^m A_j\Big) \geq \mu(G) - \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, folgt die Aussage aus Satz 4.4(d). □

Warnung 4.6. Aus der Voraussetzung $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ für alle A in einem durchschnittsstabilen Erzeuger $\mathcal{U}$ von $\mathcal{B}(E)$ folgt *nicht* $\mu_n \Rightarrow \mu$. Ein Beispiel ist $E = [0, 1)$, $\mu_n = \delta_{1-\frac{1}{n}}$, $\mu = \delta_0$ und $\mathcal{U} = \{[a, b) : 0 < a \leq b < 1\}$.

Satz 4.7. *Seien $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R})$ für $n \in \mathbb{N}$ mit zugehörigen Verteilungsfunktionen F_n bzw. F. Es gilt $\mu_n \Rightarrow \mu$ genau dann wenn für jedes $t \in \mathbb{R}$, in dem F stetig ist, $\lim_{n\to\infty} F_n(t) = F(t)$ ist.*

Beweis.

"$\Rightarrow$" Es gelte $\mu_n \Rightarrow \mu$. Ist F stetig in $t \in \mathbb{R}$, so gilt $\mu(\partial(-\infty, t]) = \mu(\{t\}) = 0$ und die Behauptung folgt aus Teil (e) von Satz 4.4.

"$\Leftarrow$" Es gelte $\lim_{n\to\infty} F_n(t) = F(t)$ für alle $t \in D := \{x \in \mathbb{R} : F \text{ ist stetig in } x\}$. Dann ist D^c abzählbar und somit D dicht in $\mathbb{R}$. Das System $\mathcal{U} := \{(a, b] : a \le b;\ a, b \in D\}$ ist durchschnittsstabil und jedes offene Intervall lässt sich als abzählbare Vereinigung von Elementen aus $\mathcal{U}$ schreiben und da jede offene Menge in $\mathbb{R}$ eine abzählbare Vereinigung von offenen Intervallen ist, erfüllt $\mathcal{U}$ die (erste) Bedingung von Lemma 4.5. Es gilt

$$\lim_{n\to\infty} \mu_n((a, b]) = \lim_{n\to\infty} \big(F_n(b) - F_n(a)\big) = F(b) - F(a) = \mu((a, b]) \text{ für } (a, b] \in \mathcal{U}.$$

Mit Lemma 4.5 folgt $\mu_n \Rightarrow \mu$.

□

Lemma 4.8. *Seien (E_1, d_1) und (E_2, d_2) metrische Räume und $h : E_1 \to E_2$ stetig. Weiter seien μ_n, $n \in \mathbb{N}$ und μ Wahrscheinlichkeitsmaße auf $(E_1, \mathcal{B}(E_1))$ mit $\mu_n \Rightarrow \mu$. Dann folgt $\mu_n h^{-1} \Rightarrow \mu h^{-1}$ auf $(E_2, \mathcal{B}(E_2))$.*

Beweis. Ist $f \in C_b(E_2)$, so folgt $f \circ h \in C_b(E_1)$ und daher (nach dem Satz über die Integration bezüglich Bildmaßen)

$$\lim_{n\to\infty} \int f \,\mathrm{d}(\mu_n h^{-1}) = \lim_{n\to\infty} \int f \circ h \,\mathrm{d}\mu_n = \int f \circ h \,\mathrm{d}\mu = \int f \,\mathrm{d}(\mu h^{-1}).$$

□

Für einige Anwendungen ist es nützlich, das letzte Lemma etwas zu verallgemeinern.

Lemma 4.9. *Seien (E_1, d_1) und (E_2, d_2) metrische Räume und $h : E_1 \to E_2$ messbar. Dann ist $D_h := \{x \in E_1 : h \text{ ist nicht stetig in } x\} \in \mathcal{B}(E_1)$.*

Beweis. Für $m, n \in \mathbb{N}$ sei

$$A_{m,n} := \{x \in E_1 :\ \text{es gibt } y, z \in E_1 \text{ mit } d_1(x, y) < \frac{1}{m},\ d_1(x, z) < \frac{1}{m} \text{ und } d_2(h(y), h(z)) \ge \frac{1}{n}\}.$$

Dann ist $A_{m,n}$ offen und $D_h = \bigcup_n \bigcap_m A_{m,n} \in \mathcal{B}(E_1)$. □

Satz 4.10. *Seien* (E_1, d_1) *und* (E_2, d_2) *metrische Räume und* $h : E_1 \to E_2$ *messbar. Weiter seien* μ_n, $n \in \mathbb{N}$ *und* μ *Wahrscheinlichkeitsmaße auf* $(E_1, \mathcal{B}(E_1))$ *mit* $\mu_n \Rightarrow \mu$ *und* $\mu(D_h) = 0$. *Dann folgt* $\mu_n h^{-1} \Rightarrow \mu h^{-1}$ *auf* $(E_2, \mathcal{B}(E_2))$.

Beweis. Sei $F \subset E_2$ abgeschlossen. Dann gilt

$$\limsup_{n\to\infty} \mu_n(h^{-1}(F)) \le \limsup_{n\to\infty} \mu_n(\overline{h^{-1}(F)}) \le \mu(\overline{h^{-1}(F)}).$$

Es ist $\overline{h^{-1}(F)} \subset h^{-1}(F) \cup D_h$. Wegen $\mu(D_h) = 0$ folgt $\mu(\overline{h^{-1}(F)}) = \mu(h^{-1}(F))$ und mit Satz 4.4(c) die Behauptung. □

4.2 Der Satz von Prohorov

Definition 4.11. (a) Die Teilmenge $\Gamma \subset \mathcal{M}_1(E)$ heißt (sequentiell) *relativ kompakt*, wenn jede Folge (μ_n) in Γ eine schwach konvergente Teilfolge hat (der Grenzwert muss nicht in Γ liegen).

(b) Die Teilmenge $\Gamma \subset \mathcal{M}_1(E)$ heißt *straff*, wenn für jedes $\varepsilon > 0$ eine kompakte Menge K_ε existiert, so dass $\mu(K_\varepsilon) \ge 1 - \varepsilon$ für jedes $\mu \in \Gamma$ ist.

Bemerkung 4.12. (a) Ist $\Gamma \subset \mathcal{M}_1(E)$ straff, dann auch jede Teilmenge von Γ.

(b) Ist E kompakt, dann ist $\mathcal{M}_1(E)$ straff.

(c) $\mathcal{M}_1(\mathbb{R})$ ist nicht straff.

(d) Ist E σ-kompakt, das heißt es existiert eine Folge K_n kompakter Mengen mit $E = \bigcup_{n=1}^\infty K_n$, dann ist jede einelementige Menge (und somit jede endliche Menge) $\Gamma \subset \mathcal{M}_1(E)$ straff.

Wir werden folgendes Lemma benötigen, welches direkt aus der Definition der schwachen Konvergenz folgt.

Lemma 4.13. *Seien* $\mu, \mu_n \in \mathcal{M}_1(E)$ *für* $n \in \mathbb{N}$. *Dann gilt* $\mu_n \Rightarrow \mu$ *genau dann, wenn jede Teilfolge* (μ'_n) *von* (μ_n) *eine weitere Teilfolge* $(\mu''_n)_n$ *hat mit* $\mu''_n \Rightarrow \mu$.

Nun folgt einer der wichtigsten Sätze der Wahrscheinlichkeitstheorie. Die wichtigere Richtung sagt, dass Straffheit eine hinreichende Bedingung für die relative Kompaktheit ist. In vielen Anwendungen ist die Straffheit vergleichsweise einfach zu zeigen. Wir werden eine Anwendung beim Beweis des Donskerschen Invarianzprinzips sehen.

Satz 4.14 (Satz von Prohorov)**.** *Sei* (E, d) *vollständig und separabel. Dann ist* $\Gamma \subset \mathcal{M}_1(E)$ *genau dann relativ kompakt, wenn* Γ *straff ist.*

Die Richtung "$\Leftarrow$" gilt dabei sogar allgemein auf metrischen Räumen. Bevor wir die Aussage beweisen zeigen wir folgendes Lemma. Dabei erinnnern wir daran, dass eine Teilmenge A eines metrischen Raums (E, d) *total beschränkt* heißt, wenn es zu jedem $\varepsilon > 0$ endlich viele Mengen mit Durchmesser kleiner ε gibt, die A überdecken.

Lemma 4.15. *Sei (E, d) vollständig, dann ist $A \subset E$ genau dann kompakt, wenn A abgeschlossen und total beschänkt ist.*

Beweis. Die Richtung "$\Rightarrow$" ist klar. Wir zeigen die Umkehrung. Dazu sei (x_n) eine Folge in A. Zu jedem $m \in \mathbb{N}$ betrachten wir eine endliche Überdeckung von A aus Mengen mit Durchmesser $< 1/m$. Dann exisiert zu jedem m eine Teilfolge von (x_n), deren Elemente alle bis auf endlich viele in *einer* Menge der zu m gehörigen Überdeckung liegen. Mit dem üblichen Diagonalfolgenargument konstruiert man eine Teilfolge, für die das für *jedes* m gilt. Diese ist eine Cauchyfolge in E und konvergiert daher gegen ein $x \in E$. Da A abgeschlossen ist, gilt $x \in A$. Daher ist A kompakt. $\square$

Bemerkung 4.16. Ohne die Voraussetzung "vollständig" ist die Aussage des Lemmas falsch. Sei zum Beispiel $E = \mathbb{Q}$ mit der euklidischen Metrik und $A = (-\sqrt{2}, \sqrt{2}) \cap \mathbb{Q}$. Dann ist A abgeschlossen (!) und total beschränkt aber nicht kompakt.

Nun zeigen wir die Richtung "$\Rightarrow$" im Satz von Prohorov. Diese ist einfacher als die Umkehrung aber auch nicht ganz so wichtig. Den Beweis der Richtung "$\Leftarrow$" diskutieren wir im Spezialfall $E = \mathbb{R}$ im Anschluss daran und im allgemeinen Fall am Ende des Kapitels.

Beweis von Satz 4.14 "$\Rightarrow$". Wir zeigen, dass aus relativer Kompaktheit die Straffheit folgt. Sei also $\Gamma \subset \mathcal{M}_1(E)$ relativ kompakt und $\varepsilon > 0$. Weiter sei $\{s_j : j \in \mathbb{N}\}$ eine abzählbare, dichte Teilmenge von E und $B_k(s_j)$ die offene Kugel mit Radius $1/k$ und Mittelpunkt s_j, $j, k \in \mathbb{N}$. Sei

$$G_{k,n} = \bigcup_{j=1}^{n} B_k(s_j),\ k, n \in \mathbb{N}.$$

Wir behaupten, dass für jedes k gilt, dass $\lim_{n\to\infty} \inf_{\mu\in\Gamma} \mu(G_{k,n}) = 1$. Angenommen, dies gilt nicht für ein bestimmtes k. Dann existiert eine aufsteigende Folge (n_j) von Zahlen in $\mathbb{N}$ und $\mu_j \in \Gamma$ mit $\lim_{j\to\infty} \mu_j(G_{k,n_j}) = c < 1$. Wegen schwacher relativer Kompaktheit besitzt die Folge (μ_j) eine Teilfolge, die gegen ein $\mu \in \mathcal{M}_1(E)$ schwach konvergiert. Mit Satz 4.4 folgt dann (wegen der Monotonie der Mengen $G_{k,n}$ als Funktion von n) $\mu(G_{k,n}) \le c$ für alle $n \in \mathbb{N}$. Da die Mengen $G_{k,n}$, $n \in \mathbb{N}$ für jedes feste k den Raum E überdecken, gilt aber $1 = \mu(E) = \lim_{n\to\infty} \mu(G_{k,n}) \leq c < 1$, was ein Widerspruch ist. Also gilt $\lim_{n\to\infty} \inf_{\mu\in\Gamma} \mu(G_{k,n}) = 1$ für alle $k \in \mathbb{N}$. Somit existiert zu jedem k ein n_k mit

$$\mu(G_{k,n_k}) \geq 1 - \varepsilon/2^k$$

für alle $\mu \in \Gamma$. Wir setzen nun

$$K := \bigcap_{k=1}^{\infty} \bar{G}_{k,n_k}.$$

K ist abgeschlossen und total beschränkt und daher nach Lemma 4.15 kompakt. Weiter gilt $\mu(K) \geq 1 - \varepsilon$ für jedes $\mu \in \Gamma$. Da $\varepsilon > 0$ beliebig war, folgt die Straffheit von Γ. $\square$

Beweis von Satz 4.14 "$\Leftarrow$" im Fall $E = \mathbb{R}$. Dieser Spezialfall ist einfacher als der allgemeine Fall, da wir hier mit Verteilungsfunktionen arbeiten können. Sei $\Gamma \subset \mathcal{M}_1(\mathbb{R})$ straff und (μ_n) eine Folge in Γ. Wir bezeichnen die zu μ_n gehörige Verteilungsfunktion mit F_n. Mit dem üblichen Diagonalfolgenargument finden wir eine Teilfolge (μ_{n_k}) so dass $\tilde{F}(q) := \lim_{k\to\infty} F_{n_k}(q)$ für alle $q \in \mathbb{Q}$ existiert. Wenn $\tilde{F}$ sich zu einer Verteilungsfunktion fortsetzen lässt, dann ist das zu $\tilde{F}$ gehörige Wahrscheinlichkeitsmaß sicher ein heißer Kandidat für den schwachen Grenzwert der Teilfolge und der Beweis ist komplett. Es ist aber leider nicht ganz klar, dass sich $\tilde{F}$ so fortsetzen lässt und daher definieren wir einen besseren Kandidaten, nämlich

$$F(x) := \inf\{\tilde{F}(q) : q > x, \, q \in \mathbb{Q}\}, \; x \in \mathbb{R}.$$

Die Funktion $x \mapsto F(x)$ ist offensichtlich monoton wachsend, rechtsseitig stetig und nimmt Werte im Intervall $[0,1]$ an. Um zu sehen, dass $\lim_{x\to\infty} F(x) = 1$ und $\lim_{x\to-\infty} F(x) = 0$ gilt, brauchen wir die Straffheitsvoraussetzung. Diese impliziert, dass zu jedem $\varepsilon > 0$ ein $M > 0$ existiert, so dass $\mu([-M,M]^c) < \varepsilon$ für alle $\mu \in \Gamma$ gilt. Daraus folgt insbesondere $F_n(M) \geq 1 - \varepsilon$ für alle $n \in \mathbb{N}$ und damit $F(M) \geq 1 - \varepsilon$ und, da $\varepsilon > 0$ beliebig ist, folgt $\lim_{x\to\infty} F(x) = 1$. Ebenso folgt $F_n(-M-1) \leq \varepsilon$ für alle $n \in \mathbb{N}$ und damit $F(-M-2) \leq \varepsilon$ und, da $\varepsilon > 0$ beliebig ist, folgt $\lim_{x\to-\infty} F(x) = 0$. Also ist F eine Verteilungsfunktion. Sei μ das zugehörige Wahrscheinlichkeitsmaß auf $\mathbb{R}$. Wir zeigen, dass $\mu_{n_k} \Rightarrow \mu$ gilt. Dazu verwenden wir Satz 4.7. Sei also $x \in \mathbb{R}$ eine Stetigkeitsstelle von F und $\varepsilon > 0$. Dann existieren rationale Zahlen q_1, q_2 und q_3, so dass $q_1 < q_2 < x < q_3$ und $F(q_3) - F(q_1) < \varepsilon$. Nun gilt für jedes $k \in \mathbb{N}$

$$F_{n_k}(q_2) \leq F_{n_k}(x) \leq F_{n_k}(q_3),$$

woraus man nach Grenzwertbildung $k \to \infty$ sofort

$$F(q_1) \leq \tilde{F}(q_2) \leq \liminf_{k\to\infty} F_{n_k}(x) \leq \limsup_{k\to\infty} F_{n_k}(x) \leq \tilde{F}(q_3) \leq F(q_3)$$

erhält. Da $\varepsilon > 0$ beliebig war, folgt also $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ und somit ist die Aussage gezeigt. $\square$

Nun skizzieren wir den Beweis im allgemeinen Fall. Details finden sich in der Monographie "Wahrscheinlichkeitstheorie" von Achim Klenke (S. 153ff).

Beweisskizze von Satz 4.14 "$\Leftarrow$" im allgemeinen Fall. Sei (E, d) ein metrischer Raum (nicht notwendig separabel oder vollständig) und $\Gamma \subset \mathcal{M}_1(E)$ straff. Dann existiert eine aufsteigende Folge kompakter Mengen $K_1 \subset K_2 \subset ...$ von E so dass $\mu(K_m) \geq 1 - 1/m$ für alle $\mu \in \Gamma$ und $m \in \mathbb{N}$ gilt. Setze $E' := \bigcup_{m\in\mathbb{N}} K_m$. Dann ist E' mit der auf E' eingeschränkten Metrik d ein σ-kompakter metrischer Raum und damit insbesondere separabel. Da jede Funktion $f \in C_b(E)$ durch Einschränkung auf E' zu einer stetigen und beschränkten Funktion auf E' wird, genügt es, die Aussage für (E', d) zu zeigen. Zur Vereinfachung der Notation nehmen wir nun oBdA an, dass $E' = E$ gilt.

Aufgrund der Separabilität von (E, d) existiert eine abzählbare Basis der Topologie von E, das heißt eine abzählbare Menge $\mathcal{U}$ von offenen Mengen in E, so dass sich jede offene Menge in E als Vereinigung von Mengen in $\mathcal{U}$ darstellen lässt. Setze $\mathcal{C}' := \{\bar{U} \cap K_n : U \in \mathcal{U}, n \in \mathbb{N}\}$ und

$$\mathcal{C} := \Big\{ \bigcup_{m=1}^{N} C_m : N \in \mathbb{N} \text{ und } C_1, ..., C_N \in \mathcal{C}' \Big\}.$$

Sei nun $(\mu_n)_{n\in\mathbb{N}}$ eine Folge in Γ. Mit dem üblichen Diagonalfolgenargument finden wir eine Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$, für die der Grenzwert

$$\alpha(C) := \lim_{k\to\infty} \mu_{n_k}(C)$$

für jedes $C \in \mathcal{C}$ existiert. *Wenn* es ein $\mu \in \mathcal{M}_1(E)$ gibt mit

$$\mu(G) = \sup\{\alpha(C) : C \in \mathcal{C} \text{ mit } C \subset G\} \text{ für } G \subset E \text{ offen},$$

dann ist μ eindeutig und es gilt für jedes offene G und jedes $C \in \mathcal{C}$ mit $C \subset G$

$$\alpha(C) = \lim_{k\to\infty} \mu_{n_k}(C) \leq \liminf_{k\to\infty} \mu_{n_k}(G)$$

und somit $\mu(G) \leq \liminf_{k\to\infty} \mu_{n_k}(G)$. Aus Satz 4.4(d) folgt nun $\mu_{n_k} \Rightarrow \mu$ und damit die Behauptung. Es bleibt also nur zu zeigen, dass ein solches $\mu \in \mathcal{M}_1(E)$ existiert. Dies ist ein Maßfortsetzungsproblem, das man ähnlich wie beim Satz von Carathéodory (positiv) lösen kann. Den etwas länglichen Beweis findet man in der zitierten Monographie von Klenke. $\square$

4.3 Der Zentrale Grenzwertsatz

Unser nächstes Ziel ist der Beweis des mehrdimensionalen Zentralen Grenzwertsatzes. Wir hatten diesen im eindimensionalen Fall bereits in Wahrscheinlichkeitstheorie 1 bewiesen und zwar ohne die Verwendung von charakteristischen Funktionen. Wir formulieren die Aussage nun noch einmal und skizzieren einen alternativen Beweis mit charakteristischen Funktionen und Straffheit und zwar ohne Benutzung des (nicht ganz einfachen) Konvergenzsatzes für charakteristische Funktionenen. Danach zeigen wir den Satz von Cramér-Wold und als Folgerung den mehrdimensionalen Zentralen Grenzwertsatz. Wie vorher bezeichnen wir die charakteristische Funktion eines Wahrscheinlichkeitsmaßes ν auf $\mathbb{R}^d$ mit $\hat{\nu}$.

Proposition 4.17. *Seien μ_n, μ Wahrscheinlichkeitsmaße auf $\mathbb{R}^d$ für $n \in \mathbb{N}$.*

a) Gilt $\mu_n \Rightarrow \mu$, so folgt $\hat{\mu}_n(x) \to \hat{\mu}(x)$ für alle $x \in \mathbb{R}^d$.

b) Ist $\{\mu_n\}_{n\in\mathbb{N}}$ straff und gilt $\hat{\mu}_n(x) \to \hat{\mu}(x)$ für alle $x \in \mathbb{R}^d$, dann folgt $\mu_n \Rightarrow \mu$.

Beweis.

a) Dies folgt aus der Definition, da cos und sin beschränkte stetige Funktionen sind.

b) Aus der Voraussetzung und dem Satz von Prohorov folgt, dass zu jeder Teilfolge von (μ_n) ein Maß $\nu \in \mathcal{M}_1(\mathbb{R})$ und eine weitere Teilfolge existieren, so dass diese schwach gegen ν konvergiert und daher nach a) auch die charakteristischen Funktionen der Teilfolge punktweise gegen $\hat{\nu}$ konvergieren. Da nach Voraussetzung aber $\hat{\mu}_n(x) \to \hat{\mu}(x)$ gilt, folgt $\hat{\nu}(x) = \hat{\mu}(x)$ für alle $x \in \mathbb{R}^d$. Aus dem Eindeutigkeitssatz 2.2 folgt somit $\nu = \mu$ und daher aus Lemma 4.13 die Behauptung $\mu_n \Rightarrow \mu$.

□

Bemerkung 4.18. Die Aussage b) in Proposition 4.17 gilt sogar ohne die Straffheitsvoraussetzung, ist dann aber schwieriger zu zeigen (siehe Klenke, Satz 15.23).

Satz 4.19 (Eindimensionaler Zentraler Grenzwertsatz)**.** *Seien X_1, X_2, ... unabhängige und identisch verteilte Zufallsgrößen mit $0 < \mathbb{V}(X_1) < \infty$. Dann gilt für $S_n := \sum_{i=1}^n X_i$*

$$\mu_n := \mathcal{L}\Big(\frac{S_n - n\mathbb{E}X_1}{\sqrt{n\mathbb{V}(X_1)}}\Big) \Rightarrow \mathcal{N}(0,1) =: \mu.$$

Beweis. Wir verifizieren die Voraussetzungen von Proposition 4.17b). Die Straffheit folgt, da für jedes $M > 0$ und $n \in \mathbb{N}$ mit der Tschebychevungleichung gilt

$$\mu_n\big([-M,M]^c\big) = \mathbb{P}\Big(\Big|\frac{S_n - n\mathbb{E}X_1}{\sqrt{n\mathbb{V}(X_1)}}\Big| > M\Big) \le \frac{1}{M^2}.$$

Die punktweise Konvergenz von $\hat{\mu}_n(t)$ gegen $\exp\{-t^2/2\}$ verifiziert man leicht. Da wir bereits wissen, dass dies die charakteristische Funktion von $\mu = \mathcal{N}(0,1)$ ist, folgt die Behauptung. $\square$

Satz 4.20 (Satz von Cramér-Wold)**.** *Sei $d \in \mathbb{N}$. Für $x, y \in \mathbb{R}^d$ sei $\pi_x(y) := \langle x, y\rangle$, wobei $\langle ., .\rangle$ das Standardskalarprodukt auf $\mathbb{R}^d$ ist.*
Seien μ und μ_n Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}^d)$. Dann gilt $\mu_n \Rightarrow \mu$ genau dann wenn $\mu_n\pi_x^{-1} \Rightarrow \mu\pi_x^{-1}$ für alle $x \in \mathbb{R}^d$ gilt.

Beweis. Da $\pi_x : \mathbb{R}^d \to \mathbb{R}$ stetig ist, folgt die Richtung "$\Rightarrow$" aus Lemma 4.8.
Zum Beweis der Umkehrung betrachten wir zunächst die Projektionen $\pi_j := \pi_{e_j}$, $1 \le j \le d$ auf die Einheitsvektoren $e_j \in \mathbb{R}^d$. Da für jedes $j \in \{1, ..., d\}$ die Folge $\mu_n\pi_j^{-1}$ schwach konvergiert, ist die Folge nach Satz 4.14 straff auf $\mathbb{R}$, das heißt zu $\varepsilon > 0$ existiert eine kompakte Menge $K_j \subset \mathbb{R}$ mit $\mu_n(\pi_j^{-1}(K_j)) \ge 1 - \frac{\varepsilon}{d}$ für alle n und j. Die Menge $K := K_1 \times ... \times K_d$ ist kompakt in $\mathbb{R}^d$ und

$$\mu_n(K^c) \le \sum_{j=1}^{d} \mu_n(\pi_j^{-1}(K_j^c)) \le \varepsilon.$$

Also ist die Folge μ_n, $n \in \mathbb{N}$ straff. Die Behauptung folgt nun aus Proposition 4.17, sobald wir die punktweise Konvergenz von $\hat{\mu}_n$ gegen $\hat{\mu}$ haben. Sei dazu $x \in \mathbb{R}^d$. Dann gilt

$$\hat{\mu}_n(x) = \int \mathrm{e}^{it}\, \mu_n\pi_x^{-1}(\mathrm{d}t) \to \int \mathrm{e}^{it}\, \mu\pi_x^{-1}(\mathrm{d}t) = \hat{\mu}(x)$$

und die Aussage folgt. $\square$

Warnung 4.21. Im letzten Satz genügt es für die Konvergenz $\mu_n \Rightarrow \mu$ nicht, dass $\mu_n\pi_j^{-1} \Rightarrow \mu\pi_j^{-1}$ für alle $j \in \{1, ..., d\}$ gilt. Als Beispiel nehme man $d = 2$, $\mu = \frac{1}{2}\delta_{(-1,-1)} + \frac{1}{2}\delta_{(1,1)}$ und $\mu_n = \frac{1}{2}\delta_{(-1,1)} + \frac{1}{2}\delta_{(1,-1)}$.

Satz 4.22 (Mehrdimensionaler Zentraler Grenzwertsatz)**.** *Sei (X_n) eine Folge unabhängiger, identisch verteilter $\mathbb{R}^d$-wertiger Zufallsgrößen mit $\mathbf{E}|X_1|^2 < \infty$. Sei $a = \mathbf{E}X_1 \in \mathbb{R}^d$ und Σ die Kovarianzmatrix von X_1. Dann gilt*

$$\mu_n := \mathcal{L}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)\Big) \Rightarrow \mathcal{N}(0, \Sigma) =: \mu.$$

Beweis. Sei $T_n := \sum_{i=1}^n (X_i - a)/\sqrt{n}$. Wegen Satz 4.20 genügt es zu zeigen, dass für jedes $x \in \mathbb{R}^d$ gilt: $\mathcal{L}(\langle x, T_n\rangle) \Rightarrow \mu\pi_x^{-1}$. Nun ist $\mu\pi_x^{-1} = \mathcal{N}(0, x^T\Sigma x)$ und

$$\langle x, T_n\rangle = \sum_{i=1}^{n} \big(\langle x, X_i\rangle - \langle x, a\rangle\big)/\sqrt{n}.$$

Aus dem eindimensionalen Zentralen Grenzwertsatz 4.19 folgt nun leicht die Behauptung. $\square$

4.4 Konvergenz in Wahrscheinlichkeit

Um Konvergenz in Wahrscheinlichkeit definieren zu können, werden wir in Kürze Ausdrücke der Form $d(X_1(\omega),\, X_2(\omega))$ für E-wertige Zufallsgrößen X_1, X_2 betrachten und es stellt sich dabei die Frage, ob die Abbildung $\omega \mapsto d(X_1(\omega),\, X_2(\omega))$ messbar ist. Allgemein muss das nicht sein. Im folgenden Lemma werden wir aber hinreichende Bedingungen dafür formulieren.

Definition 4.23. Seien (E_1, d_1) und (E_2, d_2) metrische Räume. Dann definieren wir auf $E_1 \times E_2$ die Metrik $\tilde{d}((x_1, x_2), (y_1, y_2)) := d_1(x_1, y_1) + d_2(x_2, y_2)$.

Lemma 4.24. *Seien (E_1, d_1) und (E_2, d_2) separable metrische Räume. Dann gilt $\mathcal{B}(E_1 \times E_2) = \mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$.*

Beweis.

"$\subseteq$" Da (E_1, d_1) und (E_2, d_2) separabel sind existieren abzählbar dichte Teilmengen $\{u_1, ...\}$ von E_1 und $\{v_1, ...\}$ von E_2. Sei $U_{1,m,i} := \{x \in E_1 : d_1(x, u_i) < 1/m\}$ und $U_{2,m,i} := \{x \in E_2 : d_2(x, v_i) < 1/m\}$, $i, m \in \mathbb{N}$. Dann ist $\mathcal{U}_1 := \{U_{1,m,i};\, m, i \in \mathbb{N}\}$ eine abzählbare Basis der Topologie von (E_1, d_1) und $\mathcal{U}_2 := \{U_{2,m,i};\, m, i \in \mathbb{N}\}$ eine abzählbare Basis der Topologie von (E_2, d_2). Damit ist $\mathcal{V} := \{U \times \tilde{U} : U \in \mathcal{U}_1,\, \tilde{U} \in \mathcal{U}_2\}$ eine abzählbare Basis der Topologie von $E_1 \times E_2$. Es gilt $\mathcal{V} \subseteq \mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$ und somit $\mathcal{B}(E_1 \times E_2) \subseteq \mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$.

"$\supseteq$" Diese Richtung gilt sogar ohne Separabilität: Sei

$$\begin{aligned} M_1 :=\ & \{A \in \mathcal{B}(E_1) :\ A \times E_2 \in \mathcal{B}(E_1 \times E_2)\} \\ M_2 :=\ & \{B \in \mathcal{B}(E_2) :\ E_1 \times B \in \mathcal{B}(E_1 \times E_2)\}. \end{aligned}$$

Dann ist M_1 eine Teil-σ-Algebra von $\mathcal{B}(E_1)$, welche alle offenen Mengen in E_1 enthält und somit gilt $M_1 = \mathcal{B}(E_1)$. Ebenso folgt $M_2 = \mathcal{B}(E_2)$. Daher gilt für $A \in \mathcal{B}(E_1)$ und $B \in \mathcal{B}(E_2)$

$$A \times B = (A \times E_2) \cap (E_1 \times B) \in \mathcal{B}(E_1 \times B_2)$$

und die Behauptung folgt.

$\square$

Ist (E, d) ein separabler metrischer Raum, so zeigt das Lemma, dass für E-wertige Zufallsvariablen X, Y die Abbildung $\omega \mapsto d(X(\omega), Y(\omega))$ $(\mathcal{F}, \mathcal{B})$-messbar ist. Daher ist die folgende Definition sinnvoll.

Definition 4.25. Sei (E, d) ein separabler metrischer Raum und X und X_n Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in (E, d) für $n \in \mathbb{N}$. Wir sagen, die Folge $(X_n)_{n \in \mathbb{N}}$ *konvergiert in Wahrscheinlichkeit* gegen X, wenn $d(X_n, X)$ in Wahrscheinlichkeit gegen Null konvergiert.

Satz 4.26. *Konvergiert X_n in Wahrscheinlichkeit gegen X, so folgt $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$.*

Beweis. Sei $\mu_n := \mathcal{L}(X_n)$ und $\mu := \mathcal{L}(X)$. Weiter sei $F \subset E$ abgeschlossen und F^ε die abgeschlossenene ε-Umgebung von F. Dann gilt

$$\mu_n(F) = \mathbb{P}(X_n \in F) = \mathbb{P}(X_n \in F, X \in F^\varepsilon) + \mathbb{P}(X_n \in F, X \notin F^\varepsilon) \leq \mu(F^\varepsilon) + \mathbb{P}(d(X_n, X) > \varepsilon).$$

Nun folgt die Behauptung mit Satz 4.4(c). □

Das folgende Lemma wird sich als nützlich herausstellen.

Lemma 4.27. *Sei (E, d) ein separabler metrischer Raum und X_n und Y_n Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in (E, d) für $n \in \mathbb{N}$. Gelten $\mathcal{L}(X_n) \Rightarrow \mu$ und $d(X_n, Y_n) \to 0$ in Wahrscheinlichkeit, so folgt $\mathcal{L}(Y_n) \Rightarrow \mu$.*

Beweis. Sei $F \subset E$ abgeschlossen und F^ε wie im letzten Beweis definiert. Dann gilt

$$\mathbb{P}(Y_n \in F) = \mathbb{P}(Y_n \in F, X_n \in F^\varepsilon) + \mathbb{P}(Y_n \in F, X_n \notin F^\varepsilon) \leq \mathbb{P}(X_n \in F^\varepsilon) + \mathbb{P}(d(X_n, Y_n) > \varepsilon)$$

und die Behauptung folgt wiederum mit Satz 4.4(c). □

4.5 Der Satz von Donsker

Sei nun $E = C := C([0,1])$ der Raum der reellen stetigen Funktionen auf dem Intervall $[0,1]$. Wir betrachten auf C die Metrik $d(f, g) := \|f - g\|_\infty := \sup_{t\in[0,1]} |f(t) - g(t)|$. Dann ist (C, d) ein separabler vollständiger metrischer Raum. Zunächst zeigen wir, dass die Borel-σ-Algebra $\mathcal{B}(C)$ mit der durch die Projektionen erzeugten σ-Algebra auf C übereinstimmt. Dabei sei für $m \in \mathbb{N}$ und $0 \leq t_1 < t_2 < ... < t_m \leq 1$ die Projektionsabbildung $\pi_{t_1,...,t_m} : C \to \mathbb{R}^m$ definiert als $\pi_{t_1,...,t_m}(f) = (f(t_1), ..., f(t_m))$. Man beachte, dass diese Abbildung stetig (und daher messbar) ist.

Lemma 4.28. *Es gilt*

$$\mathcal{B}(C) = \bigvee_{t\in[0,1]} \pi_t^{-1}(\mathcal{B}).$$

Beweis. Sei $\mathcal{B}' := \bigvee_{t\in[0,1]} \pi_t^{-1}(\mathcal{B})$. Wir zeigen $\mathcal{B}(C) = \mathcal{B}'$.

"$\supseteq$" Da π_t stetig ist, ist für offenes $U \subset \mathbb{R}$ auch $\pi^{-1}(U)$ offen in C und liegt daher in $\mathcal{B}(C)$. Also folgt $\mathcal{B}' \subseteq \mathcal{B}(C)$.

"$\subseteq$" Für $f \in C$ und $\varepsilon > 0$ sei $B_\varepsilon(f) := \{g \in C : d(f, g) \leq \varepsilon\}$. Dann ist

$$B_\varepsilon(f) = \bigcap_{t\in[0,1]\cap\mathbb{Q}} \{g \in C : |g(t) - f(t)| \leq \varepsilon\} = \bigcap_{t\in[0,1]\cap\mathbb{Q}} \pi_t^{-1}\big([f(t) - \varepsilon, f(t) + \varepsilon]\big) \in \mathcal{B}'.$$

Da C separabel ist, ist jede offene Menge in C eine abzählbare Vereinigung von solchen Kugeln und es folgt $\mathcal{B}(C) \subseteq \mathcal{B}'$.

□

Wir wollen nun den Satz von Donsker formulieren. Dazu betrachten wir eine Folge von unabhängigen und identisch verteilten reellwertigen Zufallsgrößen X_1, X_2, ... mit $\mathbb{E}X_1 = 0$ und $\mathbb{V}(X_1) = 1$. Sei $S_n := \sum_{i=1}^n X_i$, $n \in \mathbb{N}_0$. Für $n \in \mathbb{N}$ und $\omega \in \Omega$ definieren wir die Abbildung $Y_n(\omega, .) : [0,1] \to \mathbb{R}$ durch

$$Y_n\big(\omega, \frac{k}{n}\big) := \frac{1}{\sqrt{n}} S_k(\omega),\ k = 0, 1, ..., n$$

und interpolieren linear auf den Intervallen $[k/n, (k+1)/n]$ für $k = 0, ..., n-1$, das heißt

$$Y_n(t) = \frac{1}{\sqrt{n}}\big(S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1}\big),\ t \in [0,1].$$

Wir können Y_n als Abbildung von Ω nach C auffassen. Für festes $t \in [0,1]$ ist $Y_n(t)$ offensichtlich messbar und daher ist Y_n nach Lemma 4.28 eine $(C, \mathcal{B}(C))$-wertige Zufallsgröße. Wir bezeichnen die Verteilung von Y_n mit μ_n (wobei wir beachten müssen, dass μ_n natürlich von der Verteilung von X_1 abhängt).

Satz 4.29 (Satz von Donsker)**.** *Es existiert ein Wahrscheinlichkeitsmaß μ auf $(C, \mathcal{B}(C))$, so dass $\mu_n \Rightarrow \mu$. Das Maß μ ist charakterisiert durch die Eigenschaft, dass für alle $0 \le t_1 < ... < t_m \le 1$ gilt $\mu\pi^{-1}_{t_1,...,t_m} = \mathcal{N}(0, (t_i \wedge t_j)_{ij})$. Insbesondere hängt μ also nicht von der Verteilung von X_1 ab. Das Maß μ wird als* Wienermaß *bezeichnet und $(C, \mathcal{B}(C), \mu)$ als* Wienerraum.

Bevor wir eine Beweisskizze geben, formulieren wir eine sehr nützliche Folgerung, die als Donskersches Invarianzprinzip bezeichnet wird, und die direkt aus Satz 4.29 und Satz 4.10 folgt.

Satz 4.30. *Ist $h : C \to \mathbb{R}$ messbar mit $\mu(D_h) = 0$ (D_h wurde in Lemma 4.9 definiert), so gilt $\mathcal{L}(h(Y_n)) \Rightarrow \mu h^{-1}$.*

Beweisskizze von Satz 4.29. Man zeigt die folgenden Aussagen:

i) Für alle $0 \le t_1 < ... < t_m \le 1$ gilt $\mu_n \pi^{-1}_{t_1,...,t_m} \Rightarrow \mathcal{N}(0, (t_i \wedge t_j)_{ij})$.

ii) Die Menge $\{\mu_n\}$ ist straff auf $(C, \mathcal{B}(C))$.

Angenommen, man hat beide Aussagen gezeigt, dann folgt die Aussage des Satzes sofort: wegen der Staffheit und des Satzes von Prohorov besitzt jede Teilfolge von (μ_n) eine weitere Teilfolge, die gegen ein Wahrscheinlichkeitsmaß ν konvergiert (wobei ν von der Teilfolge abhängt). Aus Lemma 4.8 folgt, dass die endlich dimensionalen Verteilungen der letzten Teilfolge gegen die von ν schwach konvergieren. Wegen (i) folgt aber $\nu\pi^{-1}_{t_1,...,t_m} = \mathcal{N}(0, (t_i \wedge t_j)_{ij})$ für jedes $m \in \mathbb{N}$ und $0 \le t_1 < ... < t_m \le 1$. Da die endlich dimensionalen Verteilungen die Verteilung

festlegt ist ν eindeutig bestimmt und hängt damit nicht von den gewählten Teilfolgen ab. Setzen wir nun $\mu := \nu$, so folgt die Aussage des Satzes aus Lemma 4.13.
Die Beweise der Aussagen i) und ii) sind leider nicht ganz so einfach. Wir werden in der Vorlesung kurz darauf eingehen. An dieser Stelle nur soviel: i) folgt aus dem mehrdimensionalen zentralen Grenzwertsatz und für ii) verifiziert man ein Straffheitskriterium auf dem Raum $\mathcal{M}_1(C)$, welches man wiederum aus den Satz von Arzela-Ascoli herleitet, welcher ein notwendiges und hinreichendes Kriterium für die relative Kompaktheit einer Teilmenge von C enthält. Übrigens ist die Aussage i) im Spezialfall $m = 1$ und $t_1 = 1$ exakt die Aussage des eindimensionalen zentralen Grenzwertsatzes. $\square$

Warnung 4.31. Allgemein folgt für eine Folge (μ_n) von Wahrscheinlichkeitsmaßen auf C und einem weiteren $\mu \in \mathcal{M}_1(C)$ aus der Konvergenz der endlich dimensionalen Verteilungen *nicht* die schwache Konvergenz $\mu_n \Rightarrow \mu$. Es ist durchaus möglich, dass $\{\mu_n\}$ nicht straff ist. Ein einfaches Beispiel ist das folgende: für $n \in \mathbb{N}$ sei $g_n \in C$ die “Dreiecksfunktion” $g_n(t) = n^2 t$ für $0 \leq t \leq \frac{1}{2n}$, $g_n(t) = n - n^2 t$ für $\frac{1}{2n} \leq t \leq \frac{1}{n}$ und $g_n(t) = 0$ für $t \in [\frac{1}{n}, 1]$. Setze $\mu_n := \delta_{g_n}$ für $n \in \mathbb{N}$ und $\mu := \delta_0$. Dann konvergieren die endlich dimensionalen Verteilungen der μ_n gegen die von μ, aber es gilt *nicht* $\mu_n \Rightarrow \mu$. Dies sieht man zum Beispiel so: Sei $g(f) := \sup_{t\in[0,1]}(|f(t)| \wedge 1)$ für $f \in C$. Dann gilt $g \in C_b(C)$ und $\int g \, \mathrm{d}\mu_n = 1$ für alle $n \in \mathbb{N}$, aber $\int g \, \mathrm{d}\mu = 0$.

Nun präsentieren wir noch zwei Anwendungen des Donskerschen Invarianzprinzips 4.30.

Beispiel 4.32. Wir wenden Satz 4.30 an auf die Funktion $h(f) := \sup_{0\leq t\leq 1} f(t)$ für $f \in C$. Offensichtlich ist $h : C \to \mathbb{R}$ stetig. Wir betrachten zunächst die symmetrische Irrfahrt $S_n := \sum_{i=1}^n X_i$ mit unabhängigen und identisch verteilten Zufallsgrößen $X_1, ...$ mit $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ und definieren $M_n := \max_{0\leq i\leq n} S_n$, $n \in \mathbb{N}_0$. Wir berechnen gleich die Verteilung von M_n und leiten damit eine Formel für die Verteilungsfunktion von μh^{-1} her. Aus dem Invarianzprinzip von Donsker folgt dann aber auch, dass für *jede* Folge von unabhängigen und identisch verteilten Zufallsgrößen $Y_1, ...$ mit $\mathbb{E}Y_1 = 0$ und $\mathbb{V}(Y_1) = 1$ die asymptotische Verteilung des Maximums der zugehörigen Partialsummen dieselbe ist wie für die symmetrische Irrfahrt.

Ist nun wie oben $S_n = \sum_{i=1}^n X_i$, $n \in \mathbb{N}_0$ die symmetrische Irrfahrt auf $\mathbb{Z}$, so gilt für $k, l, n \in \mathbb{N}_0$

$$\mathbb{P}(S_n = l + k) = \mathbb{P}(S_n = l + k,\ M_n \geq k) = \mathbb{P}(M_n \geq k,\ S_n = k - l),$$

wobei die letzte Gleichheit eine Anwendung des *Spiegelungsprinzips* ist. Damit folgt

$$\begin{aligned}
\mathbb{P}(M_n \geq k) &= \sum_{l=-\infty}^{\infty} \mathbb{P}(M_n \geq k,\, S_n = l+k) \\
&= \sum_{l=-\infty}^{-1} \mathbb{P}(M_n \geq k,\, S_n = k+l) + \sum_{l=1}^{\infty} \mathbb{P}(S_n = k+l) + \mathbb{P}(S_n = k) \\
&= 2\mathbb{P}(S_n > k) + \mathbb{P}(S_n = k).
\end{aligned}$$

Insbesondere folgt also

$$\mathbb{P}(M_n \geq k) \leq 2P(S_n \geq k), \quad \mathbb{P}(M_n \geq k) \geq 2P(S_n > k).$$

Sei Φ die Verteilungsfunktion der Standardnormalverteilung. Weiter seien $t, \varepsilon > 0$. Ist $n \in \mathbb{N}$ hinreichend groß, so existieren $k_1 = k_1(n)$ und $k_2 = k_2(n)$ in $\mathbb{N}_0$ mit $t - \varepsilon < k_1/\sqrt{n} \leq t < k_2/\sqrt{n} \leq t + \varepsilon$. Mit dem zentralen Grenzwertsatz folgt

$$\begin{aligned}
\liminf_{n\to\infty} \mathbb{P}\Big(\frac{M_n}{\sqrt{n}} > t\Big) &\geq \liminf_{n\to\infty} \mathbb{P}\Big(\frac{M_n}{\sqrt{n}} \geq \frac{k_2}{\sqrt{n}}\Big) \\
&\geq 2 \liminf_{n\to\infty} \mathbb{P}\Big(\frac{S_n}{\sqrt{n}} > \frac{k_2}{\sqrt{n}}\Big) \\
&\geq 2 \lim_{n\to\infty} \mathbb{P}\Big(\frac{S_n}{\sqrt{n}} > t + \varepsilon\Big) \\
&= 2\big(1 - \Phi(t+\varepsilon)\big).
\end{aligned}$$

Analog folgt

$$\limsup_{n\to\infty} \mathbb{P}\Big(\frac{M_n}{\sqrt{n}} > t\Big) \leq 2\big(1 - \Phi(t-\varepsilon)\big).$$

Da $\varepsilon > 0$ beliebig ist und Φ stetig ist, folgt

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{M_n}{\sqrt{n}} \leq t\Big) = 2\Phi(t) - 1.$$

Für $t \leq 0$ gilt natürlich $\lim_{n\to\infty} \mathbb{P}\Big(\frac{M_n}{\sqrt{n}} \leq t\Big) = 0$. Damit haben wir für obige Funktion h und das Wienermaß μ gezeigt:

- μh^{-1} hat die Verteilungsfunktion $t \mapsto \max\{2\Phi(t) - 1,\, 0\}$.
- Für jede Folge von u.i.v. Zufallsgrößen $X_1, \ldots$ mit $\mathbb{E}X_1 = 0$ und $\mathbb{V}(X_1) = 1$ gilt für alle $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{0\leq i\leq n} \frac{S_i}{\sqrt{n}} \leq t\Big) = \max\{2\Phi(t) - 1,\, 0\}.$$

Beispiel 4.33. Wir betrachten nun die Abbildung $g : C \to \mathbb{R}$, die definiert ist durch

$$g(f) = \lambda(\{t \in [0,1] : f(t) \geq 0\}),$$

wobei λ das Lebesguemaß auf $[0,1]$ sei. Nun ist g leider nicht stetig, zum Beispiel nicht im Punkt $f \equiv 0$, so dass wir erst überprüfen müssen, ob die Voraussetzungen von Satz 4.30 für die Abbildung g erfüllt sind. Wir zeigen, dazu

(a) g ist $(\mathcal{B}(C), \mathcal{B})$-messbar.

(b) $\mu(D_g) = 0$.

Um diese Aussagen zu zeigen, definieren wir die Abbildung $\psi : C \times [0,1] \to \mathbb{R}$ durch $\psi(f,t) = f(t)$. Offensichtlich ist ψ stetig und damit $(\mathcal{B}(C \times [0,1]), \mathcal{B})$-messbar. Da sowohl C als auch $[0,1]$ separable metrische Räume sind, folgt aus Lemma 4.24, dass $\mathcal{B}(C \times [0,1]) = \mathcal{B}(C) \otimes \mathcal{B}([0,1])$ gilt. Damit ist ψ $\mathcal{B}(C) \otimes \mathcal{B}([0,1])$-messbar.

Beweis von (a): Es gilt $A := \{(f,t) \in C \times [0,1] : f(t) \geq 0\} = \psi^{-1}([0,\infty)) \in \mathcal{B}(C) \otimes \mathcal{B}([0,1])$. Für $f \in C$ ist

$$g(f) = \int_0^1 \mathbb{1}_A(f,t) \, \lambda(\mathrm{d}t).$$

Aus dem Satz von Fubini folgt, dass g $(\mathcal{B}(C), \mathcal{B})$-messbar ist.

Beweis von (b): Diese Aussage werden wir in der Vorlesung zeigen.

Um Satz 4.30 anwenden zu können, benötigen wir das folgende bemerkenswerte *Arcus-Sinus-Gesetz*. Sei S_n, $n \in \mathbb{N}_0$ die symmetrische Irrfahrt auf $\mathbb{Z}$ und $\pi_m := \big|\{k \in \{0,...,m-1\} : S_k \geq 0, S_{k+1} \geq 0\}\big|$ die Anzahl der Segmente der Irrfahrt bis m, die oberhalb der x-Achse liegen. Dann gilt für $0 \leq a < b \leq 1$

$$\lim_{n\to\infty} \mathbb{P}\Big(a \leq \frac{\pi_n}{n} \leq b\Big) = \int_a^b \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}} \, \mathrm{d}x.$$

Der Integrand ist die Dichte der sogenannten *Arkus-Sinus Verteilung*, die ihren Namen deshalb trägt, weil die zugehörige Verteilungsfunktion eine Arkus-Sinus Funktion ist. Ein Beweis der Aussage findet sich zum Beispiel in der Monographie *Probability: Theory and Examples* von R. Durrett.
Nun gilt $\mathcal{L}\big(\frac{\pi_n}{n}\big) = \mu_n g^{-1}$ und daher folgt mit dem Satz von Donsker

- μg^{-1} hat die Verteilungsfunktion $t \mapsto \frac{2}{\pi} \arcsin \sqrt{t}$, $0 \leq t \leq 1$.
- Für jede Folge von u.i.v. Zufallsgrößen $X_1, ...$ mit $\mathbb{E}X_1 = 0$ und $\mathbb{V}(X_1) = 1$ gilt für alle $t \in [0,1]$

$$\lim_{n\to\infty} \mathbb{P}\Big(g(Y_n(.)) \leq t\Big) = \frac{2}{\pi} \arcsin \sqrt{t}.$$