Reinhard Höpfner

Vorlesung Stochastik I

Kapitel VI: Schwache Konvergenz

Sommersemester 2016

Institut für Mathematik, Johannes Gutenberg Universität Mainz

October 5, 2016

Übersicht zu Kapitel VI :

A. Wahrscheinlichkeitsmasse auf metrischen Räumen, schwache Konvergenz, stochastische Konvergenz auf separablen metrischen Räumen

B. Schwache Konvergenz auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ – zwei wichtige Sätze

***C. Verallgemeinerungen (ohne Beweise)**

A. Wahrscheinlichkeitsmasse auf metrischen Räumen, schwache Konvergenz, stochastische Konvergenz

In diesem Teilkapitel sei (E, d) ein metrischer Raum; $\mathcal{O}$ bezeichne das System der offenen und $\mathcal{F}$ das System der abgeschlossenen Teilmengen von E. $\mathcal{B}(E)$ ist die Borelsche σ-Algebra auf E. Wir schreiben $\mathcal{C}_b(E)$ für die Klasse aller stetigen und beschränkten Funktionen $f : E \to I\!R$, und $\mathcal{C}_u(E) \subset \mathcal{C}_b(E)$ für die Teilklasse der auf E gleichmässig stetigen beschränkten Funktionen. Q bzw. $(Q_n)_{n\geq 1}$ bezeichnet stets Wahrscheinlichkeitsmasse bzw. Folgen von Wahrscheinlichkeitsmassen auf $(E, \mathcal{B}(E))$. Eine klassische Referenz für schwache Konvergenz von Wahrscheinlichkeitsmassen auf metrischen Räumen ist das Buch von Billingsley (1968).

6.1 Hilfssatz: a) Sei Q beliebig. Dann gibt es für jedes $A \in \mathcal{B}(E)$ und jedes $\varepsilon > 0$ ein $F_\varepsilon \in \mathcal{F}$ und ein $G_\varepsilon \in \mathcal{O}$ so dass

$$F_\varepsilon \subset A \subset G_\varepsilon \quad \text{und} \quad Q(G_\varepsilon \setminus F_\varepsilon) \;<\; \varepsilon \;.$$

b) Q ist eindeutig bestimmt durch die Werte der Integrale

$$\int f\, dQ \;, \quad f \in \mathcal{C}_u(E) \;.$$

Beweis: 1) Betrachte das System $\mathcal{H}$ aller Mengen $A \in \mathcal{B}(E)$ mit der Eigenschaft

$$\text{für jedes } \varepsilon > 0 \text{ gibt es } F_\varepsilon \in \mathcal{F},\, G_\varepsilon \in \mathcal{O} \text{ mit } F_\varepsilon \subset A \subset G_\varepsilon \text{ und } Q(G_\varepsilon \setminus F_\varepsilon) < \varepsilon \;.$$

Dann ist $\mathcal{H}$ eine Sub-σ-Algebra von $\mathcal{B}(E)$: *i)* man hat $E \in \mathcal{H}$ und $\emptyset \in \mathcal{H}$, denn E, $\emptyset$ sind zugleich offen und abgeschlossen in (E, d) (cf. Dieudonné 1969 S. 33, 37); *ii)* $\mathcal{H}$ ist stabil unter der Bildung von Komplementen; *iii)* $\mathcal{H}$ ist stabil unter Bildung von abzählbaren Vereinigungen:
zu A_n, $n \geq 1$ in $\mathcal{H}$ und $\varepsilon > 0$ wählt man zuerst $F_{n,\varepsilon} \in \mathcal{F}$ und $G_{n,\varepsilon} \in \mathcal{G}$ mit Eigenschaft

$$F_{n,\varepsilon} \subset A_n \subset G_{n,\varepsilon} \quad \text{und} \quad Q(G_{n,\varepsilon} \setminus F_{n,\varepsilon}) \;<\; \varepsilon 2^{-(n+1)} \;\;,\;\; n \in I\!N$$

und setzt $\widetilde{F} := \bigcup_{n=1}^{\infty} F_{n,\varepsilon} \in \mathcal{B}(E)$. Stetigkeit in $\emptyset$ von Wahrscheinlichkeitsmassen liefert ein $N \in I\!N$ mit

$$Q\left(\widetilde{F} \setminus \bigcup_{n=1}^{N} F_{n,\varepsilon}\right) \;<\; \frac{\varepsilon}{2} \;.$$

Nun betrachtet man $F := \bigcup_{n=1}^{N} F_{n,\varepsilon} \in \mathcal{F}$ und $G := \bigcup_{n=1}^{\infty} G_{n,\varepsilon} \in \mathcal{G}$ (endliche Vereinigungen abgeschlossener Mengen sind abgeschlossen, beliebige Vereinigungen offener Mengen offen) und hat

$$F \subset \bigcup_{n=1}^{\infty} A_n \;\subset G \quad \text{und} \quad Q(G \setminus F) \;<\; Q(G \setminus \widetilde{F}) + Q(\widetilde{F} \setminus F) \;<\; \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \;.$$

Dabei war $\varepsilon > 0$ beliebig, also ist $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{H}$ nachgewiesen.
Weiter ist $\mathcal{F}$ ein Teilsystem von $\mathcal{H}$: für $A := F$ abgeschlossen setzt man $F_\varepsilon := F$ und gewinnt eine offene Obermenge G_ε mit $Q(G_\varepsilon \setminus F) < \varepsilon$ aus absteigender Stetigkeit des Wahrscheinlichkeitsmasses Q entlang offener Umgebungen $\{x \in E : d(x, F) < \delta\} \downarrow F$ für $\delta \downarrow 0$.
Mit $\mathcal{O}$ ist auch $\mathcal{F}$ ein Erzeuger der Borel-σ-Algebra $\mathcal{B}(E)$: also gilt $\mathcal{H} = \mathcal{B}(E)$. Das zeigt a).
2) $\mathcal{F}$ ist ein durchschnittsstabiler Erzeuger von $\mathcal{B}(E)$, welcher den vollen Raum E enthält: nach Eindeutigkeitssatz 1.13 ist Q also eindeutig bestimmt durch die Werte $Q(F)$, $F \in \mathcal{F}$. Mit

$$\varphi(t) \;:=\; \begin{cases} 1 & t \leq 0 \\ 1-t & 0 < t < 1 \\ 0 & t \geq 1 \end{cases}$$

sind für jedes $F \in \mathcal{F}$ die Funktionen

$$\varphi_{n,F}(x) \;:=\; \varphi(\, n\, d(x, F)\,)\;, \quad n \geq 1\;,\; x \in E$$

beschränkt und gleichmässig stetig auf E. Für $(\varphi_{n,F})_n \subset \mathcal{C}_u(E)$ gilt

$$\varphi_{n,F}(x) \;\downarrow\; 1_F(x)\;, \quad n \to \infty$$

für alle $x \in E$, was mit dominierter Konvergenz 2.17

$$\int \varphi_{n,F}\, dQ \quad\downarrow\quad \int 1_F\, dQ \;=\; Q(F)$$

impliziert. Also sind Wahrscheinlichkeitsmasse Q auf $(E, \mathcal{B}(E))$ schon durch ihre Integrale

$$\int f\, dQ\;, \quad f \in \mathcal{C}_u(E)$$

eindeutig bestimmt. Das ist b). □

6.1' Hilfssatz: Sei (E, d) ein vollständig separabler metrischer Raum. Dann ist Q auf $(E, \mathcal{B}(E))$ *kompakt approximierbar*, d.h.: zu jedem $A \in \mathcal{B}(E)$ und zu jedem $\varepsilon > 0$ existiert ein Kompaktum K_ε mit der Eigenschaft

$$K_\varepsilon \subset A\;, \quad Q(A \setminus K_\varepsilon) < \varepsilon\;.$$

Beweis: 1) Wir zeigen zuerst: zu jedem $\varepsilon > 0$ existiert ein Kompaktum K so dass $Q(K^c) < \varepsilon$.
Da (E, d) separabel, existiert zu jedem n eine offene Überdeckung von E durch abzählbar viele offene Kugeln $B_{n,1}, B_{n,2}, \ldots$ mit Radius $\frac{1}{n}$. Mit aufsteigender Stetigkeit von Q gibt es zu jedem $\varepsilon > 0$ ein $k(n, \varepsilon)$ so dass gilt

$$G_n \;:=\; \bigcup_{j=1}^{k(n,\varepsilon)} B_{n,j} \quad \text{erfüllt} \quad Q(G_n) \;>\; 1 - 2^{-n}\varepsilon\;.$$

Für $G := \bigcap_n G_n$ gilt damit

$$Q(G^c) \;=\; Q\left(\bigcup_n G_n^c\right) \;<\; \varepsilon \; .$$

Nach Konstruktion findet man zu jedem $\delta > 0$ eine endliche Kollektion offener δ-Kugeln, die G überdeckt. Damit ist G präkompakt (siehe z.B. Dieudonné 1969, Kapitel 3.16 und 3.17). Da (E,d) vollständig, ist G damit auch relativkompakt (Dieudonné 1969, 3.17.5), d.h. der Abschluss $K := \overline{G}$ von G ist kompakt. Damit ist ein Kompaktum K mit $Q(K^c) < \varepsilon$ gefunden.

2) Sei nun $A \in \mathcal{B}(E)$ beliebig. Nach a) existiert eine aufsteigende Folge von Kompakta $(K_n)_n$ mit $Q(K_n) \uparrow 1$. Zu $\varepsilon > 0$ gibt es nach 6.1 a) eine abgeschlossene Menge $F_\varepsilon \subset A$ so dass $Q(A \setminus F_\varepsilon) < \frac{1}{2}\varepsilon$. Als abgeschlossene Teilmenge eines Kompaktums ist $F_\varepsilon \cap K_n$ kompakt, für alle n. Für hinreichend grosses n ist dann $K_\varepsilon := F_\varepsilon \cap K_n$ eine kompakte Teilmenge von A mit $Q(A \setminus K_\varepsilon) < \varepsilon$. $\square$

6.2 Definition: Zusammen mit Q und $(Q_n)_{n\geq 1}$ betrachte Zufallsvariable mit Werten in $(E, \mathcal{B}(E))$

$$X_n : (\Omega_n, \mathcal{A}_n) \longrightarrow (E, \mathcal{B}(E)) \; , \; n \geq 1 \; , \quad X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{B}(E))$$

die auf irgendwelchen $(\Omega_n, \mathcal{A}_n, P_n)$, $n \geq 1$, $(\Omega, \mathcal{A}, P)$ definiert sind.

a) Die Folge $(Q_n)_n$ *konvergiert schwach in* E gegen Q

$$Q_n \quad \stackrel{w}{\longrightarrow} \quad Q \quad \text{für } n \to \infty$$

falls gilt

$$\int f \, dQ_n \;\stackrel{n\to\infty}{\longrightarrow}\; \int f \, dQ \qquad \text{für jedes } f \in \mathcal{C}_b(E) \; .$$

b) Die Folge $(X_n)_{n\geq 1}$ *konvergiert schwach in* E (oder: *der Verteilung nach*) gegen X

$$X_n \;\stackrel{\mathcal{L}}{\longrightarrow}\; X \; , \quad n \to \infty$$

falls gilt

$$\mathcal{L}(X_n | P_n) =: Q_n \quad \stackrel{w}{\longrightarrow} \quad Q := \mathcal{L}(X|P) \quad \text{für } n \to \infty \; ;$$

nach a) ist dies äquivalent zu

$$E_{P_n}(f(X_n)) = \int f \, dQ_n \quad \stackrel{n\to\infty}{\longrightarrow} \quad \int f \, dQ = E_P(f(X)) \qquad \forall \, f \in \mathcal{C}_b(E) \; .$$

6.3 Bemerkung: a) In 6.2 b) braucht man von Zufallsvariablen X_n, $n \geq 1$, X mit Werten in demselben metrischen Raum (E, d) (versehen mit seiner Borel-σ-Algebra) nur noch die Verteilungen: es ist daher irrelevant, auf welchem Wahrscheinlichkeitsraum jede einzelne der betrachteten Zufallsvariablen lebt.

b) Sei (E', d') ein anderer metrischer Raum, sei $h : E \longrightarrow E'$ stetig. Aus schwacher Konvergenz

$$(*) \qquad X_n \xrightarrow{\mathcal{L}} X \quad \text{für } n \to \infty$$

(mit den Bezeichnungen und Voraussetzungen aus 6.2) folgt sofort

$$(**) \qquad Y_n := h \circ X_n \xrightarrow{\mathcal{L}} h \circ X =: Y \quad \text{für } n \to \infty \ :$$

zum Nachweis von $(**)$ ist zu zeigen

$$E_{P_n}(f(h \circ X_n)) \longrightarrow E_P(f(h \circ X)) \quad \text{für } n \to \infty$$

für alle $f \in \mathcal{C}_b(E')$, was wegen $f \circ h \in \mathcal{C}_b(E)$ ein Spezialfall von $(*)$ ist.
Diese Eigenschaft der schwachen Konvergenz, sich auf beliebige 'stetige Funktionale' der betrachteten Zufallsvariablen fortzusetzen, ist extrem wichtig und wird in 6.4" unten allgemeiner formuliert werden.

Nach der Art einer Garderobe, aus der man sich passende Kleider herausgreift, bietet der folgende Satz ('Portmanteau Theorem') eine reichliche Auswahl nützlicher Charakterisierungen von schwacher Konvergenz.

6.4 Hauptsatz: Sei (E, d) ein metrischer Raum, seien Q_n, $n \geq 1$, Q Wahrscheinlichkeitsmasse auf $(E, \mathcal{B}(E))$. Dann sind die folgenden Aussagen gleichwertig:

i) $Q_n \xrightarrow{w} Q$ für $n \to \infty$,

ii) für jedes $f \in \mathcal{C}_u(E)$ gilt $\int f\, dQ_n \overset{n\to\infty}{\longrightarrow} \int f\, dQ$,

iii) für jede abgeschlossene Teilmenge $F \subset E$ gilt $\limsup_{n\to\infty} Q_n(F) \leq Q(F)$,

iv) für jede offene Teilmenge $G \subset E$ gilt $\liminf_{n\to\infty} Q_n(G) \geq Q(G)$,

v) für jedes Q-randlose $H \in \mathcal{B}(E)$ gilt $\lim_{n\to\infty} Q_n(H) = Q(H)$.

Eine Menge $H \in \mathcal{B}(E)$ heisst dabei *Q-randlos* falls $F \setminus G$ eine Q-Nullmenge ist, wobei F den Abschluss und G die Menge der inneren Punkte von H bezeichnet.

Beweis: 1) Direkt aus der Definition folgt i)$\Longrightarrow$ii).

2) Zeige ii)$\Longrightarrow$iii): Für eine abgeschlossene Teilmenge $F \subset E$ betrachte die Funktionenfolge $(\varphi_{m,F})_m$ aus dem Beweis von 6.1; wegen $\int_E \varphi_{m,F} dQ \downarrow Q(F)$ für $m \to \infty$ gibt es zu jedem $\varepsilon > 0$ ein m mit $\int_E \varphi_{m,F} dQ < Q(F) + \varepsilon$. Wegen $\varphi_{m,F} \in \mathcal{C}_u(E)$ hat man für festes und hinreichend grosses m unter ii)

$$\limsup_{n\to\infty} Q_n(F) \;\leq\; \lim_{n\to\infty} \int_E \varphi_{m,F}\, dQ_n \;=\; \int_E \varphi_{m,F}\, dQ \;<\; Q(F) + \varepsilon \,.$$

Hierbei ist $\varepsilon > 0$ beliebig, also gilt $\limsup_{n\to\infty} Q_n(F) \leq Q(F)$.

3) Da Komplemente offener Mengen abgeschlossen sind, gilt iii)$\Longrightarrow$iv).

4) Zeige iv)$\Longrightarrow$i): Betrachte eine Funktion $f \in \mathcal{C}_b(E)$ fest. Dabei reicht es, die Aussage i) im Spezialfall $(0,1)$-wertiger Funktionen f zu beweisen (sonst wähle M mit $\sup |f| < M < \infty$ und betrachte anstelle von f die $(0,1)$-wertige Funktion $\widetilde{f} := \frac{1}{2M}(f + M)$).
Sei also $f : E \to (0,1)$ stetig. Dann hat man einerseits die Abschätzung

$$\left| f \;-\; \sum_{k=1}^{2^m} \frac{k}{2^m}\, 1_{\{\, \frac{k-1}{2^m} < f \leq \frac{k}{2^m}\,\}} \right| \;\leq\; \frac{1}{2^m}$$

auf E, und andererseits durch Umsummieren (genau wie in Schritt 0) des Beweises von 5.14)

$$\sum_{k=1}^{2^m} k\, 1_{\{\, \frac{k-1}{2^m} < f \leq \frac{k}{2^m}\,\}}(y) \;=\; \sum_{k=1}^{2^m} 1_{\{\, \frac{k-1}{2^m} < f\,\}}(y) \;=\; \sum_{k=1}^{2^m} 1_{G_{k,m}}(y)$$

mit offenen Mengen $G_{k,m} := \{\frac{k-1}{2^m} < f\}$, wegen der Stetigkeit von f; nach iv) gilt für $G_{k,m} \in \mathcal{O}$

$$\liminf_{n\to\infty} \int_E \left(\sum_{k=1}^{2^m} \frac{1}{2^m}\, 1_{G_{k,m}} \right) dQ_n \;\geq\; \int_E \left(\sum_{k=1}^{2^m} \frac{1}{2^m}\, 1_{G_{k,m}} \right) dQ$$

für beliebiges festes m. Zusammen ergibt sich daraus

$$(*) \qquad \liminf_{n\to\infty} \int_E f\, dQ_n \;\geq\; \int_E f\, dQ \,.$$

Mit derselben Überlegung für die $(0,1)$-wertige Funktion $1-f$ anstelle von f erhält man

$$(**) \qquad \liminf_{n\to\infty} \int_E (1-f)\, dQ_n \;\geq\; \int_E (1-f)\, dQ$$

oder äquivalent

$$\limsup_{n\to\infty} \int_E f\, dQ_n \;\leq\; \int_E f\, dQ \,.$$

Also liefern $(*)$ und $(**)$ zusammen die gewünschte Aussage i):

$$\lim_{n\to\infty} \int_E f\, dQ_n \;=\; \int_E f\, dQ \,.$$

5) Zeige iii)+iv)$\Longrightarrow$v): Für Q-randlose Mengen $H \in \mathcal{B}(E)$ folgt aus iii)+iv) angewandt auf den Abschluss F und das Innere G der Menge H

$$Q(G) \;\leq\; \liminf_{n\to\infty} Q_n(G) \;\leq\; \limsup_{n\to\infty} Q_n(F) \;\leq\; Q(F)$$

wobei $Q(F) = Q(G)$ wegen Q-Randlosigkeit: also steht überall '='.

6) Zeige v)$\Longrightarrow$iii): Sei $F \subset E$ abgeschlossen, setze $F_\delta := \{x \in E : d(x,F) \leq \delta\}$: dann ist mit absteigender Stetigkeit von Q die Funktion

$$(+) \qquad [0,\infty) \ni \delta \longrightarrow Q\left(F_\delta\right) \in [\,Q(F), 1\,]$$

rechtsstetig und nichtfallend, hat also höchstens abzählbar viele Sprungstellen. Folglich kann man – diese abzählbar vielen Punkte vermeidend – eine Folge $\delta_k \downarrow 0$ auswählen, so dass alle δ_k Stetigkeitsstellen der Funktion (+) sind: für diese δ_k sind die Mengen F_{δ_k} Q-randlos. Unter v) gilt für beliebiges festes k

$$\lim_{n\to\infty} Q_n\left(F_{\delta_k}\right) \;=\; Q\left(F_{\delta_k}\right) \;.$$

Sei $\varepsilon > 0$ beliebig, wähle k so dass $Q(F_{\delta_k}) < Q(F) + \varepsilon$, dann gilt

$$\limsup_{n\to\infty} Q_n(F) \;\;\leq\;\; \lim_{n\to\infty} Q_n\left(F_{\delta_k}\right) \;\;\leq\;\; Q(F) + \varepsilon \,.$$

Damit ist v)$\Longrightarrow$iii) gezeigt, und Satz 6.4 vollständig bewiesen. $\square$

Was bedeutet schwache Konvergenz insbesondere im Spezialfall $E = I\!R$, $d(x,y) = |x-y|$?

6.4' Satz: Seien Q_n, $n \geq 1$, Q Wahrscheinlichkeitsmasse auf $(I\!R, \mathcal{B}(I\!R))$, seien F_n, $n \geq 1$, F die zugehörigen Verteilungsfunktionen. Sei D_F die Menge aller Unstetigkeitsstellen von F. Gleichwertig:

i) es gilt $Q_n \xrightarrow{w} Q$ (schwache Konvergenz in $I\!R$ für $n \to \infty$) ;

ii) für alle $t \in I\!R \setminus D_F$ gilt $F_n(t) \longrightarrow F(t)$ für $n \to \infty$.

Beweis: Die Aussage i)$\Longrightarrow$ii) folgt aus 6.4 v), da für $t \notin D_F$ die Halbachse $(-\infty, t]$ eine Q-randlose Menge ist. Wir zeigen ii)$\Longrightarrow$i) in drei Schritten:

1) Nach 1.19 ist F als Verteilungsfunktion rechtsstetig, nichtfallend und $[0,1]$-wertig, damit ist die Menge D_F der Sprungsstellen von F höchstens abzählbar.

2) Betrachte die Klasse $\widetilde{\mathcal{C}}_2 := \{(a,b] : a,b \in I\!R \setminus D_F\,,\ a < b\}$. Nach Definition von $\widetilde{\mathcal{C}}_2$ gilt für $(a,b] \in \widetilde{\mathcal{C}}_2$

$$Q(\{b\} \cup \{a\}) \;=\; [F(b) - F(b^-)] + [F(a) - F(a^-)] \;=\; 0\,,$$

damit liefert Voraussetzung ii)

$$Q_n((a,b]) \;=\; F_n(b)-F_n(a) \;\longrightarrow\; F(b)-F(a) \;=\; Q((a,b])$$

für alle $(a,b] \in \widetilde{\mathcal{C}}_2$.

3) Wegen Separabilität von $I\!R$ und Abzählbarkeit von D_F kann ähnlich wie in 1.6' a) jede offene Menge $G \subset I\!R$ als abzählbare Vereinigung von Mengen aus $\widetilde{\mathcal{C}}_2$ geschrieben werden. Für die Vereinigung jeweils endlich vieler Mengen $A_1, \ldots, A_m$ aus $\widetilde{\mathcal{C}}_2$ impliziert Schritt 2) mit $\cap$-Stabilität von $\widetilde{\mathcal{C}}_2$ und Einschluss-Ausschluss-Formel (vgl. Stochastik-Einführungsvorlesung), dass

$$Q_n\left(\bigcup_{j=1}^m A_j\right) \;=\; \sum_j Q_n(A_j) \;-\; \sum_{j_1<j_2} Q_n(A_{j_1}\cap A_{j_2}) \;+\; \sum_{j_1<j_2<j_3} Q_n(A_{j_1}\cap A_{j_2}\cap A_{j_3}) \;\pm \ldots$$

für $n \to \infty$ gegen

$$Q\left(\bigcup_{j=1}^m A_j\right) \;=\; \sum_j Q(A_j) \;-\; \sum_{j_1<j_2} Q(A_{j_1}\cap A_{j_2}) \;+\; \sum_{j_1<j_2<j_3} Q(A_{j_1}\cap A_{j_2}\cap A_{j_3}) \;\pm \ldots .$$

konvergiert. Ist nun $G = \bigcup_{j=1}^{\infty} A_j$ eine Darstellung der offenen Menge G mit $A_j \in \widetilde{\mathcal{C}}_2$, so wählt man m gross genug für $Q(\bigcup_{j=1}^{m} A_j) > Q(G) - \varepsilon$, und hat

$$\liminf_{n\to\infty} Q_n(G) \;\geq\; \liminf_{n\to\infty} Q_n(\bigcup_{j=1}^m A_j) \;=\; Q(\bigcup_{j=1}^m A_j) \;>\; Q(G)-\varepsilon$$

und damit –da $\varepsilon > 0$ beliebig war– die notwendige und hinreichende Bedingung 6.4 iv) für schwache Konvergenz $Q_n \xrightarrow{w} Q$ in $I\!R$. $\square$

Wir werden einen anderen Beweis für Satz 6.4 unten in 6.12 sehen. Der folgende Satz, bereits angekündigt in 6.3 b), ist das zweite Hauptsatz dieses Teilkapitels.

6.4" Satz (continuous mapping theorem): Betrachte metrische Räume (E,d), (E',d'), versehen mit ihren σ-Algebren $\mathcal{B}(E)$, $\mathcal{B}(E')$. Seien Q_n, $n \geq 1$, Q Wahrscheinlichkeitsmasse auf $(E, \mathcal{B}(E))$. Sei $h : (E,\mathcal{B}(E)) \longrightarrow (E', \mathcal{B}(E'))$ messbar, und bezeichne D_h die Menge aller Unstetigkeitsstellen von h. Dann gilt $D_h \in \mathcal{B}(E)$. Unter der Voraussetzung

$$Q_n \;\xrightarrow{w}\; Q \quad (\text{schwach in } E, \text{ für } n\to\infty)\,, \quad Q(D_h)=0$$

gilt für die Bildmasse $Q_n^h = \mathcal{L}(h|Q_n)$, $n \geq 1$, $Q^h = \mathcal{L}(h|Q)$:

$$Q_n^h \;\xrightarrow{w}\; Q^h \quad (\text{schwach in } E', \text{ für } n\to\infty)\,.$$

Beweis: 1) Wir zeigen zuerst $D_h \in \mathcal{B}(E)$. Die Menge D_h aller Unstetigkeitsstellen von $h(\cdot)$ ist

$$D_h \ := \ \Big\{ x \in E : \text{ es gibt Folgen } v_n \to x, \, \widetilde{v}_n \to x \text{ in } E \text{ so dass } \liminf_n d'(h(v_n), h(\widetilde{v}_n)) > 0 \Big\}$$

als Teilmenge von E. Betrachte für $\eta > 0$ Mengen

$$D(\eta) \ := \ \{ x \in E : \text{ in beliebiger Nähe von } x \text{ gibt es Punkte } v, \widetilde{v} \text{ mit } d'(h(v), h(\widetilde{v})) > \eta \}$$

und deren δ-Umgebungen in E

$$A(\eta, \delta) \ := \ \bigcup_{z \in D(\eta)} B_\delta(z) \ = \ \{ y \in E : \ \text{ es gibt einen Punkt } x \in D(\eta) \text{ mit } d(y, x) < \delta \}$$

für $\delta > 0$. Beliebige Vereinigungen offener Mengen sind offen, also ist $A(\eta, \delta)$ offen in E. Damit gilt $A(\eta, \delta) \in \mathcal{B}(E)$. Folglich sind auch

$$\bigcap_k A(\eta, \frac{1}{k}) \ = \ \{ y \in E : \text{ in beliebiger Nähe von } y \text{ gibt es Punkte } v, \widetilde{v} \text{ mit } d'(h(v), h(\widetilde{v})) > \eta \} \ = \ D(\eta)$$

und damit $D_h \ = \bigcup_m D(\frac{1}{m})$ Borelmengen.

2) Unter der Voraussetzung $Q(D_h) = 0$ zeige schwache Konvergenz der Bildmasse mit Portmanteau. Sei F' eine abgeschlossene Menge in E'. Betrachte das Urbild $A := h^{-1}(F') \in \mathcal{B}(E)$. Häufungspunkte von A sind entweder Stetigkeitsstellen von h – dann in A enthalten, da F' abgeschlossen ist – oder Unstetigkeitstellen von h, also gilt für den Abschluss $F := \overline{h^{-1}(F')}$ von A in E

$$F \ \subset \ A \cup D_h \, .$$

Ist nun D_h eine Q-Nullmenge, erhält man aus schwacher Konvergenz $Q_n \xrightarrow{w} Q$ in E mit 6.4 iii)

$$\begin{aligned}
\limsup_{n \to \infty} \ Q_n^h(F') \ &= \ \limsup_{n \to \infty} \ Q_n(A) \\
&\leq \ \limsup_{n \to \infty} \ Q_n(F) \ \leq \ Q(F) \\
&\leq \ Q(A) + Q(D_h) \ = \ Q(A) \ = \ Q^h(F') \, .
\end{aligned}$$

Wieder mit 6.4 iii) ist dies die schwache Konvergenz $Q_n^h \xrightarrow{w} Q^h$ in E'. □

Setzen wir zusätzlich Separabilität des metrischen Raumes (E, d) voraus, steht auch der Begriff der stochastischen Konvergenz für Zufallsvariablen mit Werten in $(E, \mathcal{B}(E))$ zur Verfügung.

6.5 Hilfssatz: Ist (E, d) ein *separabler* metrischer Raum und sind X, Y Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(E, \mathcal{B}(E))$, so ist $d(X, Y)$ eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$.

Beweis: Ist (E, d) ein separabler metrischer Raum, so gilt $\mathcal{B}(E \times E) = \mathcal{B}(E) \otimes \mathcal{B}(E)$ wie in 4.14. Als stetige Abbildung von $E \times E$ nach $\mathbb{R}$ ist $(x, y) \to d(x, y)$ messbar bezüglich $\mathcal{B}(E \times E)$. Jedes Paar (X, Y) ist eine messbare Abbildung von $(\Omega, \mathcal{A})$ in den Produktraum $(E \times E, \mathcal{B}(E) \otimes \mathcal{B}(E))$. Unter Separabilität wird also $d(X, Y)$ eine $\mathcal{A}$-messbare Abbildung. $\square$

6.5' Definition: Sei (E, d) ein separabler metrischer Raum, seien X_n, Y_n, $n \geq 1$, X Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in $(E, \mathcal{B}(E))$.

a) Die Folge $(X_n)_n$ konvergiert *P-stochastisch gegen X* falls

$$\lim_{n\to\infty} P\,(\,d(X_n, X) > \varepsilon\,) \;=\; 0 \;\text{ für jedes feste } \varepsilon > 0\;.$$

b) Die Folge $(X_n)_n$ konvergiert *P-fast sicher gegen X* falls

$$\text{für } P\text{-fast alle } \omega \in \Omega \text{ gilt} \quad \lim_{n\to\infty} d(X_n, X)(\omega) \;=\; 0\;.$$

c) Die Folgen $(X_n)_n$ und $(Y_n)_n$ heissen *stochastisch äquivalent* für $n \to \infty$ falls

$$\lim_{n\to\infty} P\,(\,d(X_n, Y_n) > \varepsilon\,) \;=\; 0 \;\text{ für jedes feste } \varepsilon > 0\;.$$

Man sieht analog zum Beweis von 2.22, dass P-fast sichere Konvergenz P-stochastische Konvergenz impliziert; auch bleibt das Teilfolgenkriterium aus 2.23 in separablen metrischen Räumen gültig.

6.6 Satz: Sei E ein separabler metrischer Raum, seien X_n, $n \geq 1$, X Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(E, \mathcal{B}(E)$. Dann gilt für $n \to \infty$:

$$X_n \longrightarrow X \quad P\text{-stochastisch} \quad\Longrightarrow\quad X_n \xrightarrow{\mathcal{L}} X\;.$$

Beweis: Wir benutzen die Charakterisierung 6.4 ii) der schwachen Konvergenz in E. Sei $f : E \to \mathbb{R}$ gleichmässig stetig und beschränkt, sei M eine obere Schranke für $|f|$, sei $\varepsilon > 0$ beliebig. Wegen gleichmässiger Stetigkeit von f gibt es zu $\varepsilon > 0$ ein $\delta = \delta(\varepsilon) > 0$ so dass

$$x, x' \in E\;,\; d(x, x') \leq \delta \quad\Longrightarrow\quad |f(x) - f(x')| \leq \varepsilon\;.$$

Kombiniert mit der vorausgesetzten stochastischen Konvergenz ergibt sich

$$\begin{aligned} \limsup_{n\to\infty} |E(f(X_n)) - E(f(X))| &\leq \limsup_{n\to\infty} E\,(|f(X_n) - f(X)|) \\ &\leq \varepsilon + 2\,M \lim_{n\to\infty} P\,(\,d(X_n, X) > \delta\,) \quad = \; \varepsilon\;. \end{aligned}$$

Dies gilt für jedes $\varepsilon > 0$. Mit Portmanteau 6.4 ii) ist die Behauptung bewiesen. $\square$

6.6' Hilfssatz ('Lemma von Slutsky'): Sei E ein separabler metrischer Raum, seien X_n, Y_n, $n \geq 1$, X Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(E, \mathcal{B}(E))$. Sind $(X_n)_n$ und $(Y_n)_n$ stochastisch äquivalent, so

$$Y_n \xrightarrow{\mathcal{L}} X \quad \text{für } n \to \infty \quad \Longleftrightarrow \quad X_n \xrightarrow{\mathcal{L}} X \quad \text{für } n \to \infty \, .$$

Beweis: Zu $f \in \mathcal{C}_u(E)$ wähle M, $\varepsilon > 0$, $\delta > 0$ wie im Beweis von Satz 6.6 und schreibe

$$|E(f(X_n)) - E(f(Y_n))| \;\leq\; \varepsilon \,+\, 2\,M \cdot P\,(\,d(X_n, Y_n) > \delta\,) \quad , \quad n \geq 1 \, .$$

Unter stochastischer Äquivalenz der Folgen $(X_n)_n$, $(Y_n)_n$ nach 6.5' c) sind daher für jedes $f \in \mathcal{C}_u(E)$ die Aussagen $E(f(X_n)) \to E(f(X))$ und $E(f(Y_n)) \to E(f(X))$ für $n \to \infty$ äquivalent. □

6.7 Beispiel: Versehe den Grundraum $(\Omega, \mathcal{A}) := (I\!R, \mathcal{B}(I\!R))$ mit der auf $(0,1)$ konzentrierten Gleichverteilung $P := \mathcal{R}(0,1)$. Die Zufallsvariablen

$$X_n := \begin{cases} 1_{(0,\frac{1}{2})}, & \text{falls } n \text{ gerade} \\ 1_{(\frac{1}{2},1)}, & \text{falls } n \text{ ungerade} \end{cases}$$

besitzen für alle $n \in I\!N$ dieselbe Binomialverteilung $\mathcal{B}(1, \frac{1}{2})$, sie konvergieren also trivialerweise der Verteilung nach. Sie konvergieren aber nicht P-stochastisch. Wir zeigen dies indirekt: angenommen, es könne auf $(\Omega, \mathcal{A})$ eine $\{0,1\}$-wertige Zufallsvariable $X : \omega \to X(\omega)$ mit der Eigenschaft

$$\text{für jedes } \varepsilon > 0 \text{ fest:} \quad \lim_{n\to\infty} P\,(\,|X_n - X| > \varepsilon\,) \;=\; 0$$

definiert werden. Betrachtet man insbesondere die Teilfolge $(n_k)_k$ der geraden Zahlen, erhält man

$$P\left(|1_{(0,\frac{1}{2})} - X| > \varepsilon\right) \;=\; 0 \quad \text{für jedes } \varepsilon > 0 \text{ fest}$$

und damit $X = 1_{(0,\frac{1}{2})}$ P-fast sicher. Betrachtet man dagegen die Teilfolge $(n_k)_k$ der ungeraden Zahlen, führt dasselbe Argument auf $X = 1_{(\frac{1}{2},1)}$ P-fast sicher. Dies liefert einen Widerspruch, also war die Annahme absurd. □

Das Beispiel 6.7 zeigt, dass eine Umkehrung der Aussage von Satz 6.6 gilt im allgemeinen nicht möglich ist. Jedoch gilt im Spezialfall einer *deterministischen* Limesvariablen:

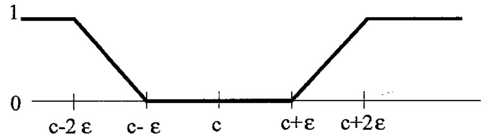
6.8 Hilfssatz: Ist (E, d) ein separabler metrischer Raum und $X \equiv x_0$ P-fast sicher konstant, so

$$X_n \longrightarrow x_0 \quad P\text{-stochastisch} \quad \Longleftrightarrow \quad X_n \xrightarrow{\mathcal{L}} x_0 \, .$$

Beweis: Sei $\varepsilon > 0$ fest. Mit φ wie im Beweis von 6.1 definiert man Funktionen $f_\varepsilon \in \mathcal{C}_u(E)$ durch

$$F_\varepsilon := \{x \in E : d(x, x_0) \leq \varepsilon\} \,, \quad f_\varepsilon(x) \;:=\; 1 - \varphi\left(\frac{1}{\varepsilon}\, d(x, F_\varepsilon)\right) \,, \; x \in E$$

(leicht zu malen im Spezialfall



$E = I\!R$ und $x_0 = c$). Aus $X_n \xrightarrow{\mathcal{L}} X$ und $X \equiv x_0$ P-fast sicher folgt für jedes feste $\varepsilon > 0$

$$E(f_\varepsilon(X_n)) \quad \overset{n\to\infty}{\longrightarrow} \quad E(f_\varepsilon(X)) \;=\; f_\varepsilon(x_0) = 0$$

für $n \to \infty$, und damit weiter

$$P(\, d(X_n, X) > 2\varepsilon\,) \;\leq\; \int_E f_\varepsilon(X_n)\, dP \quad \longrightarrow \quad 0\,.$$

Da $\varepsilon > 0$ beliebig war, ist dies die P-stochastische Konvergenz der $(X_n)_n$ gegen $X \equiv x_0$. $\square$

B. Schwache Konvergenz auf $(I\!R, \mathcal{B}(I\!R))$ – zwei wichtige Sätze

6.9 Bemerkung: Kombinieren wir für Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(I\!R, \mathcal{B}(I\!R))$ das Tableau der Konvergenzarten aus Kapitel II.C mit dem in Teilkapitel A eingeführten Begriff der Verteilungskonvergenz, so ergibt sich folgendes Bild:

$$\begin{array}{ccccc}
P\text{-fs Konvergenz} & & \begin{pmatrix}\text{gleichgr. int.}\\ \rightsquigarrow\end{pmatrix} & & L^p\text{-Konv., } p \geq 1 \\
& \searrow & & \swarrow & \\
& & P\text{-stoch. Konv.} & & \\
& & \Downarrow & & \\
& & \text{Verteilungskonvergenz in } (I\!R, \mathcal{B}(I\!R)). & &
\end{array}$$

Es gibt eine Brücke von der Verteilungskonvergenz zurück zur fast sicheren Konvergenz, unter dem Stichwort 'verteilungsgleiche Ersetzung'. Dies wird in den mit * versehenen Nummern 6.10 – 6.13 behandelt. Sollte man dieses Thematik überspringen wollen, startet man die Lektüre dieses Teilkapitels mit 6.14.

***6.10 Lemma ('Verteilungsgleiche Ersetzung'):** Sei Q ein beliebiges Wahrscheinlichkeitsmass auf $(I\!R, \mathcal{B}(I\!R))$, sei F die zu Q gehörende Verteilungsfunktion.

a) Es gibt zwei Möglichkeiten, eine 'Inverse' zu F zu definieren:

$$\begin{aligned} G(u) &:= \inf\{t \in I\!R : F(t) \geq u\} \; , \quad 0 < u < 1 \; , \\ F^{-1}(u) &:= \inf\{t \in I\!R : F(t) > u\} \; , \quad 0 < u < 1 \; . \end{aligned}$$

Dabei sind beide Funktionen $G \, , \, F^{-1} : (0,1) \to I\!R$ messbar und nichtfallend, und es gilt

$$G \text{ ist linksstetig, } F^{-1} \text{ ist rechtsstetig .}$$

b) Auf $(\widehat{\Omega}, \widehat{\mathcal{A}}, \widehat{P}) := ((0,1), \mathcal{B}((0,1)), \lambda\!\!\lambda|_{(0,1)})$ mit der kanonischen Variable $\widehat{U} := \mathrm{id}|_{(0,1)}$ liefert sowohl

$$\widehat{X} \;:=\; G(\widehat{U}) \;:\; (\widehat{\Omega}, \widehat{\mathcal{A}}) \longrightarrow (I\!R, \mathcal{B}(I\!R))$$

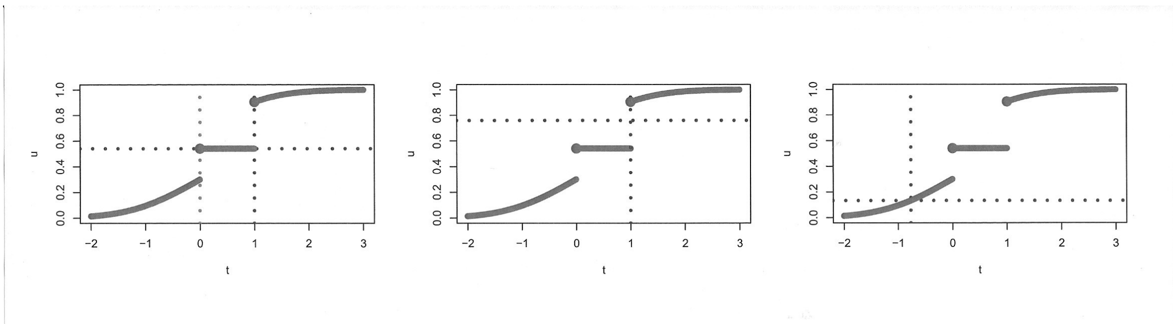
als auch

$$\widehat{X}' \;:=\; F^{-1}(\widehat{U}) \;:\; (\widehat{\Omega}, \widehat{\mathcal{A}}) \longrightarrow (I\!R, \mathcal{B}(I\!R))$$

eine Zufallsvariable mit Verteilung Q und Verteilungsfunktion F:

$$\mathcal{L}(\widehat{X}|\widehat{P}) \;=\; Q \;=\; \mathcal{L}(\widehat{X}'|\widehat{P}) \; .$$

Beweis: 1) Klar sind sowohl G als auch F^{-1} nichtfallende Funktionen $(0,1) \to I\!R$. Für Levels $u \in (0,1)$, die von F in streng monotonem Wachstum – auch durch Sprünge – überquert werden, fallen beide Definitionen einer Inversen zusammen; ein Unterschied entsteht dagegen für diejenigen Levels u, die 'flats' von F entsprechen.



i) Die Funktion F^{-1} ist rechtsstetig: betrachte Folgen $u_n \downarrow u$ in $(0,1)$. Da F^{-1} nichtfallend, gilt

$$\beta := F^{-1}(u) \;\leq\; \lim_{n \to \infty} F^{-1}(u_n) =: \gamma \; .$$

Angenommen, $\beta < \gamma$. Nach Definition von $F^{-1}(u) = \inf\{t : F(t) > u\}$ ist dann der Funktionswert $F\left(\frac{\beta+\gamma}{2}\right)$ notwendig $> u$, und damit $> u_n$ für schliesslich alle n. Das liefert aber den Widerspruch $F^{-1}(u_n) \leq \frac{\beta+\gamma}{2} < \gamma$ für schliesslich alle n. Also gilt $\beta = \gamma$.

ii) Die Funktion G ist linksstetig: betrachte Folgen $u_n \uparrow u$ in $(0,1)$. Da G nichtfallend, hat man

$$\beta := \lim_{n\to\infty} G(u_n) \ \le \ G(u) =: \gamma \ .$$

Nach Definition von $G(v) = \inf\{t : F(t) \ge v\}$ hat die Verteilungsfunktion F zur Zeit β alle Levels u_n bereits erreicht

$$F(\beta) \ge u_n \ , \quad n \ge 1$$

was wegen $\sup_n u_n = u$ auch

$$F(\beta) \ge u$$

und damit wieder nach Definition von G

$$G(u) \le \beta \quad \Longrightarrow \quad \beta = \gamma$$

impliziert. Damit ist G linksstetig.

iii) Als nichtfallende Funktionen $(0,1) \to I\!R$ sind F^{-1} und G notwendig $\mathcal{B}(0,1)$ - $\mathcal{B}(I\!R)$ - messbar, da Urbilder von Halbachsen stets Halbachsen sind.

2) Die Menge $D_F := \{t \in I\!R : F(t^-) < F(t)\}$ aller Sprungstellen der Verteilungsfunktion F ist höchstens abzählbar: als nichtfallende Funktion kann $F : I\!R \to [0,1]$ nicht mehr als k Sprünge mit Sprunghöhe $\ge \frac{1}{k}$ aufweisen, für jedes $k \in I\!N$.

3) Die Menge K_F der 'flats' von F

$$K_F := \{u \in (0,1) : \text{ es gibt } -\infty < \underline{t} < \overline{t} < +\infty \text{ so dass } F(\underline{t}) = F(\overline{t}) = u\}$$

ist höchstens abzählbar: zu jedem $u \in K_F$ findet man eine rationale Zahl $r = r(u) \in \mathbb{Q}$ mit $F(r) = u$. Damit wird K_F bijektiv auf eine Teilmenge von $\mathbb{Q}$ abgebildet. Für $u \in K_F$ nimmt F auf dem halboffenen Intervall

$$I(u) \ := \ [\, \min\{t : F(t) = u\}, \ \sup\{t : F(t) = u\}\,) \ \neq \emptyset$$

den Wert u an (das Intervall enthält seinen linken Randpunkt wegen der Rechtsstetigkeit von F).

4) Die zwei Funktionen G, $F^{-1} : (0,1) \to I\!R$ erfüllen $G \le F$ und stimmen $\lambda\!\!\lambda$-fast sicher überein: nach Definition hat man für alle $u \in (0,1)$

$$G(u) = \inf\{t \in I\!R : F(t) \ge u\} \ \le \ \inf\{t \in I\!R : F(t) > u\} = F^{-1}(u)$$

wobei

$$G(u) < F^{-1}(u) \quad \Longrightarrow \quad \emptyset \neq [G(u), F^{-1}(u)) \ = \ I(u) \quad \Longrightarrow \quad u \in K_F \ ;$$

aber K_F ist nach Schritt 3) höchstens abzählbar.

5) Direkt aus der Definition von $G(\cdot)$ folgen die beiden wichtigen Aussagen

$$(6.10') \qquad F \circ G(u) \; \geq \; u \quad \text{für alle } u \in (0,1)$$

(wegen Rechtsstetigkeit von F; diese Ungleichung ist strikt falls F das Level u mit einem Sprung überquert) und

$$(6.10'') \qquad G \circ F(t) \; \leq \; t \quad \text{für alle } t \in I\!R \text{ mit } \quad 0 < F(t) < 1$$

(die Ungleichung ist strikt falls das Level $u := F(t)$ bereits vor der Zeit t erreicht wird).

6) Wir kommen zum Abschluss des Beweises und zeigen zuerst

$$(*) \qquad \{u \in (0,1) : G(u) \leq t\} \; = \; (0, F(t)] \cap (0,1) \; , \quad t \in I\!R \; :$$

im Fall $F(t) = 1$ hat die Verteilungsfunktion F zur Zeit t bereits alle Levels $u \in (0,1)$ überquert, also gilt $G(u) \leq t$ für alle $u \in (0,1)$, und auf beiden Seiten von $(*)$ steht das Intervall $(0,1)$; im Fall $F(t) < 1$ hat man einerseits wegen (6.10') und Monotonie von F die Implikation

$$G(u) \leq t \quad \Longrightarrow \quad u \leq F(G(u)) \leq F(t)$$

und damit die Inklusion '$\subset$' von $(*)$, andererseits wegen Monotonie von G und (6.10") auch

$$u \leq F(t) \quad \Longrightarrow \quad G(u) \leq G(F(t)) \leq t$$

und damit die Inklusion '$\supset$' von $(*)$.
Zum Beweis der Aussage b) des Satzes betrachtet man nun auf

$$(\widehat{\Omega}, \widehat{\mathcal{A}}, \widehat{P}) \; := \; \Big((0,1) \, , \, \mathcal{B}((0,1)) \, , \, \lambda\!\!\lambda|_{(0,1)} \Big)$$

die kanonische Variable $\widehat{U}$, definiert durch $\widehat{U}(u) := u$ für alle $0 < u < 1$. Auf beiden Seiten von $(*)$ steht dann genau das Ereignis

$$\{G(\widehat{U}) \leq t\} \; ,$$

und Anwendung von $\lambda\!\!\lambda$ auf die rechte Seite von $(*)$ zeigt, dass die Zufallsvariable $\widehat{X} := G(\widehat{U})$ unter $\widehat{P}$ die Verteilungsfunktion F besitzt. Also gilt

$$\mathcal{L}(\widehat{X} | \widehat{P}) \; = \; Q \; .$$

Da die Funktionen G und F^{-1} nach Schritt 4) $\lambda\!\!\lambda$-fast sicher übereinstimmen, stimmen die Zufallsvariablen $\widehat{X} = G(\widehat{U})$ und $\widehat{X}' := F^{-1}(\widehat{U})$ $\widehat{P}$-fast sicher überein. Insbesondere besitzen damit $\widehat{X}$ und $\widehat{X}'$

dieselbe Verteilung unter $\widehat{P}$. $\square$

*6.11 **Satz:** Betrachte auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ Wahrscheinlichkeitsmasse Q_n, $n \geq 1$, Q und die zugehörigen Verteilungsfunktionen F_n, $n \geq 1$, F. Sei D_F die Menge aller Unstetigkeitsstellen von F. Gilt

$$(*) \qquad F_n(t) \stackrel{n\to\infty}{\longrightarrow} F(t) \quad \text{für alle } t \in \mathbb{R} \setminus D_F \ ,$$

so gibt es einen Wahrscheinlichkeitsraum $(\widehat{\Omega}, \widehat{\mathcal{A}}, \widehat{P})$ und darauf reellwertige Zufallsvariablen $\widehat{X}_n$, $n \geq 1$, $\widehat{X}$ mit den Eigenschaften

$$\begin{cases} Q_n = \mathcal{L}(\widehat{X}_n | \widehat{P}) \ , \ n \geq 1 \ , \quad Q = \mathcal{L}(\widehat{X} | P) \ , \\ \widehat{X}_n \longrightarrow \widehat{X} \quad \widehat{P}\text{-fast sicher für } n \to \infty \ . \end{cases}$$

Bemerkung: In Satz 6.11 wird *keine* Aussage gemacht über eine a priori gegebene Folge von Zufallsvariablen $(X_n)_n$ mit Verteilungsfunktionen F_n, $n \geq 1$. Hier werden *gegebene* Verteilungsfunktionen F_n, $n \geq 1$, F *mit der Eigenschaft* $(*)$ 'künstlich' durch Zufallsvariable unterlegt, die – definiert auf einem geeigneten Wahrscheinlichkeitsraum – fast sicher konvergieren.

Beweis: Wir betrachten die zu F_n, $n \geq 1$, F gehörenden linksstetigen Inversen $(0,1) \to \mathbb{R}$

$$G_n(u) := \inf\{t \in \mathbb{R} : F_n(t) \geq u\} \ , \ n \geq 1 \ , \quad G(u) := \inf\{t \in \mathbb{R} : F(t) \geq u\}$$

und definieren auf $(\widehat{\Omega}, \widehat{\mathcal{A}}, \widehat{P}) := ((0,1), \mathcal{B}((0,1)), \lambda\!\!\lambda|_{(0,1)})$ Zufallsvariable

$$\widehat{X}_n := G_n(\widehat{U}) \ , \ n \geq 1 \ , \quad \widehat{X} := G(\widehat{U}) \ ,$$

wobei $\widehat{U}$ die kanonische Variable auf $(\widehat{\Omega}, \widehat{\mathcal{A}})$ bezeichnet. Nach 6.10 besitzen die $\widehat{X}_n$, $n \geq 1$, $\widehat{X}$ die vorgegebenen Verteilungen Q_n, $n \geq 1$, Q. Zu zeigen bleibt

$$(+) \qquad G_n(u) \stackrel{n\to\infty}{\longrightarrow} G(u) \quad \text{für } \lambda\!\!\lambda\text{-fast alle } u \in (0,1) \ ;$$

denn $(+)$ schreibt sich um in

$$\widehat{X}_n \to \widehat{X} \quad \widehat{P}\text{-fast sicher} \ , \quad n \to \infty$$

und die Behauptung des Satzes ist bewiesen.

Zum Beweis von $(+)$ fixieren wir ein $u \in (0,1)$ so dass $u \notin K_F$ (die Menge K_F aller 'flats' von F ist höchstens abzählbar, Beweisschritt 3) in 6.10). Für $\varepsilon > 0$ beliebig wähle $t_i \in \mathbb{R}$ so dass

$$t_1 < G(u) < t_2 \ , \quad |t_i - G(u)| < \varepsilon \ , \quad t_i \notin D_F \ ,$$

(die letzte Bedingung ist leicht zu erfüllen, da die Menge D_F der Unstetigkeitsstellen von F höchstens abzählbar ist). Nach Definition von G muss gelten

$$F(t_1) \; < \; u \; ;$$

da u kein 'flat' von F ist (wegen $u \notin K_F$), muss ebenfalls gelten

$$F(t_2) \; > \; u \; .$$

An den Stellen $t_i \notin D_F$ kann man die Voraussetzung $(*)$ des Satzes ausnutzen

$$\lim_{n\to\infty} F_n(t_1) \; = \; F(t_1) \; < \; u \; < \; F(t_2) \; = \; \lim_{n\to\infty} F_n(t_2)$$

und hat damit

$$F_n(t_1) \; < \; u \; < \; F_n(t_2) \quad \text{für schliesslich alle } n \; .$$

Nach Definition von G_n und wegen (6.10") bedeutet das

$$t_1 \; < \; G_n(u) \; \leq \; t_2 \quad \text{für schliesslich alle } n \; .$$

Wegen $t_i \in U_\varepsilon(G(u))$ hat man also gezeigt:

$$G_n(u) \; \in \; U_\varepsilon(G(u)) \quad \text{für schliesslich alle } n \; .$$

Da $\varepsilon > 0$ beliebig war, hat man hiermit $(+)$

$$G_n(u) \stackrel{n\to\infty}{\longrightarrow} G(u) \quad \forall\, u \notin K_F$$

nachgewiesen und den Beweis des Satzes abgeschlossen. □

Mit verteilungsgleicher Ersetzung kann man nun die Aussage des Satzes 6.4' neu interpretieren:

***6.12 Alternativer Beweis für Satz 6.4':** Wir folgern aus Satz 6.11: sind Q_n, $n \geq 1$, Q Wahrscheinlichkeitsmasse auf $(I\!R, \mathcal{B}(I\!R))$ und F_n, $n \geq 1$, F die zugehörigen Verteilungsfunktionen, ist D_F die Menge aller Unstetigkeitsstellen von F, so sind die folgenden Aussagen i) und ii) gleichwertig:

i) es gilt $Q_n \stackrel{w}{\to} Q$ für $n \to \infty$;

ii) für jedes $t \in I\!R \setminus D_F$ gilt $\lim\limits_{n\to\infty} F_n(t) \; = \; F(t)$.

Beweisrichtung ii) $\Longrightarrow$ i): Unter ii) existiert nach 6.11 ein Wahrscheinlichkeitsraum $(\widehat{\Omega}, \widehat{\mathcal{A}}, \widehat{P})$ und Zufallsvariable $\widehat{X}_n$, $n \geq 1$, $\widehat{X}$ auf $(\widehat{\Omega}, \widehat{\mathcal{A}})$ mit Werten in $(I\!R, \mathcal{B}(I\!R))$ so dass gilt

$$\begin{cases} Q_n = \mathcal{L}(\widehat{X}_n | \widehat{P}) \; , \; n \geq 1 \; , \;\; Q = \mathcal{L}(\widehat{X} | \widehat{P}) \; , \\ \widehat{X}_n \to \widehat{X} \quad \widehat{P}\text{-fast sicher für } n \to \infty \; . \end{cases}$$

Für beliebiges $f \in \mathcal{C}_b(\mathbb{R})$ zeigt dann dominierte Konvergenz

$$\int_{\mathbb{R}} f\, dQ_n = \int_{\widehat{\Omega}} f(\widehat{X}_n)\, d\widehat{P} \quad \overset{n\to\infty}{\longrightarrow} \quad \int_{\widehat{\Omega}} f(\widehat{X})\, d\widehat{P} = \int_{\mathbb{R}} f\, dQ \; .$$

Das aber ist schwache Konvergenz $Q_n \overset{w}{\to} Q$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Beweisrichtung i) $\Longrightarrow$ ii): Die zu beweisende Behauptung folgt aus Portmanteau 6.4 v): für festes $t \notin D_F$ ist die Halbgerade $H := (\infty, t]$ eine Q-randlose Menge, also impliziert i) sofort

$$F_n(t) = Q_n(H) \;\longrightarrow\; Q(H) = F(t)\;, \quad n \to \infty\; . \qquad \square$$

Mit dem eben gegebenen Beweis kann man 6.11 und 6.12 auch so zusammenfassen:

*__6.13 Folgerung:__ Mit den Voraussetzungen und Bezeichnungen aus 6.12 sind gleichwertig:

i) für $n \to \infty$ gilt $Q_n \overset{w}{\to} Q$ (schwache Konvergenz in $\mathbb{R}$) ;

ii) es gibt einen Wahrscheinlichkeitsraum $(\widehat{\Omega}, \widehat{\mathcal{A}}, \widehat{P})$ und Zufallsvariable $\widehat{X}_n$, $n \geq 1$, $\widehat{X}$ auf $(\widehat{\Omega}, \widehat{\mathcal{A}})$ mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ so dass gilt

$$\begin{cases} Q_n = \mathcal{L}(\widehat{X}_n|\widehat{P})\;,\; n \geq 1\;,\;\; Q = \mathcal{L}(\widehat{X}|P)\;, \\ \widehat{X}_n \;\longrightarrow\; \widehat{X} \quad \widehat{P}\text{-fast sicher für } n \to \infty\; . \end{cases}$$

Für den zweiten Themenbereich dieses Teilkapitels, Straffheit von Wahrscheinlichkeitsmassen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und Helly'scher Auswahlsatz, braucht man die mit * versehenen Nummern 6.10 – 6.13 nicht. Wir bereiten das zweite Hauptergebnis des Teilkapitels (6.16 und 6.17 unten) vor.

6.14 Definition: Eine Familie $\mathcal{Q}$ von Wahrscheinlichkeitsmassen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heisst *straff*, wenn es zu jedem $\varepsilon > 0$ ein $K = K(\varepsilon) < \infty$ gibt so dass gilt

$$\sup_{Q \in \mathcal{Q}} Q(\,[-K, K]^c\,) \;<\; \varepsilon\; .$$

Für ein einzelnes Wahrscheinlichkeitsmass Q auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ findet man man, z.B. durch $K_n := [-n, n]$, $n \geq 1$, eine aufsteigende Folge kompakter Mengen mit $Q(K_n) \uparrow 1$ für $n \to \infty$. Insbesondere gibt es zu beliebig kleinem $\varepsilon > 0$ eine Konstante $K = K(\varepsilon) < \infty$ so dass gilt:

$$Q(\,[-K, K]^c\,) \;<\; \varepsilon\; .$$

Damit ist auch ist jede *endliche* Familie $\mathcal{Q}$ von Wahrscheinlichkeitsmassen in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ straff.

6.15 Hilfssatz: Betrachte Wahrscheinlichkeitsmasse Q_n, $n \geq 1$, Q auf $(I\!R, \mathcal{B}(I\!R))$. Aus schwacher Konvergenz $Q_n \stackrel{w}{\to} Q$ in $I\!R$ für $n \to \infty$ folgt Straffheit der Folge $(Q_n)_n$.

Beweis: Assoziiere Verteilungsfunktionen F_n, $n \geq 1$, F zu Q_n, $n \geq 1$, Q und wähle zu $\varepsilon > 0$ beliebig klein ein $K = K(\varepsilon)$ so dass gilt

$$\pm K \notin D_F \ , \quad F(-K) + (1 - F(K)) = Q\left([-K, +K]^c\right) \ < \ \varepsilon \ .$$

Für festes $n \in I\!N$ hat man

$$Q_n\left([-K, +K]^c\right) \ \leq \ Q_n\left((-K, +K]^c\right) \ = \ F_n(-K) + (1 - F_n(K))$$

während Satz 6.4' wegen $\pm K \notin D_F$ die Konvergenz

$$\lim_{n \to \infty} \left(F_n(-K) + (1 - F_n(K))\right) \ = \ Q\left([-K, +K]^c\right) \ < \ \varepsilon$$

impliziert. Folglich gibt es ein $n_0 = n_0(\varepsilon)$ so dass

$$Q_n\left([-K, +K]^c\right) \ < \ \varepsilon \quad \text{für alle } n > n_0$$

gilt. Endliche Familien von Wahrscheinlichkeitsmassen auf $(I\!R, \mathcal{B}(I\!R))$ sind stets straff, also erreicht man durch eventuell notwendiges Vergrössern von K auch

$$Q_n\left([-K, +K]^c\right) \ < \ \varepsilon \quad \text{für } 1 \leq n \leq n_0 \ .$$

Dies liefert die gewünschte Aussage $\sup\limits_{n \geq 1} Q_n\left([-K, +K]^c\right) \ < \ \varepsilon$. $\square$

Damit können wir einen wichtigen Auswahlsatz (eine 'griffige' Formulierung wird in 6.17 nachgeliefert) auf $(I\!R, \mathcal{B}(I\!R))$ beweisen. Dies ist das zweite Hauptergebnis dieses Teilkapitels.

6.16 Helly'scher Auswahlsatz: Sei $(F_n)_n$ eine beliebige Folge von Verteilungsfunktionen auf $I\!R$.
a) Dann gibt es Teilfolgen $(n_k)_k$ und zugehörige Limesfunktionen $F : I\!R \to [0,1]$, nichtfallend und rechtsstetig, so dass gilt

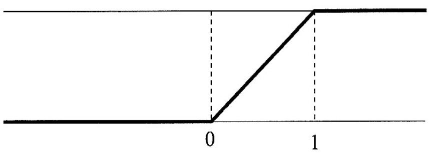
$$\text{für jede Stetigkeitstelle } t \text{ von } F(\cdot) \ : \ F_{n_k}(t) \quad \stackrel{k \to \infty}{\longrightarrow} \quad F(t) \ .$$

b) Unter der Straffheitsbedingung

$$(S) \qquad \forall\, \varepsilon > 0 \ \exists\, K = K(\varepsilon) < \infty \ \text{ so dass } \ \limsup_n [F_n(-K) + (1 - F_n(K))] \ < \ \varepsilon$$

ist jede Limesfunktion F in a) Verteilungsfunktion eines Wahrscheinlichkeitsmasses auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Bemerkung: i) Bezeichnet Q_n das zu F_n assoziierte Wahrscheinlichkeitsmass auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so zeigen die Argumente aus dem Beweis von 6.15: Bedingung (S) in b) ist äquivalent zur Straffheit der Familie $\{Q_n : n \in \mathbb{N}\}$ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ gemäss Definition 6.14.
ii) Das Beispiel: $F_n := G(\cdot \pm n)$, mit

G=  VF der Gleichverteilung auf (0,1);

zeigt, dass in a) die (rechtsstetigen und nichtfallenden) Funktionen $F \equiv 0$ oder $F \equiv 1$ als Limesfunktionen vorkommen können; Bedingung (S) in b) schliesst dies aus.

Beweis: 1) Sei $\{r_1, r_2, \ldots\}$ eine Abzählung von $\mathbb{Q}$. Betrachte zuerst den Punkt r_1. Da $[0,1]$ kompakt ist in $\mathbb{R}$, kann man aus der Folge $(F_n(r_1))_n \subset [0,1]$ konvergente Teilfolgen auswählen: sei $(n^1_\ell)_\ell$ eine Teilfolge der natürlichen Zahlen und $G(r_1)$ ein Element aus $[0,1]$ so dass

$$F_{n^1_\ell}(r_1) \quad \overset{\ell\to\infty}{\longrightarrow} \quad G(r_1) \ \in [0,1] \ .$$

Betrachte dann den Punkt r_2: zu $(n^1_\ell)_\ell$ gibt es eine weitere Teilfolge $(n^2_\ell)_\ell \subset (n^1_\ell)_\ell$ und ein Element $G(r_2)$ aus $[0,1]$ so dass (mit demselben Argument wie eben)

$$\text{für } r \in \{r_1, r_2\} \text{ gilt } : \quad F_{n^2_\ell}(r) \quad \overset{\ell\to\infty}{\longrightarrow} \quad G(r) \ \in [0,1] \ .$$

Sukzessiv in $k \in \mathbb{N}$ fortschreitend wählt man so Teilfolgen $(n^k_\ell)_\ell \subset (n^{k-1}_\ell)_\ell$ und Limespunkte $G(r_k)$ in $[0,1]$ mit der Eigenschaft

$$\text{für } r \in \{r_1, r_2, \ldots, r_k\} \text{ gilt } : \quad F_{n^k_\ell}(r) \quad \overset{\ell\to\infty}{\longrightarrow} \quad G(r)$$

(beachte dabei: man kann die ersten Folgenglieder von $(n^k_\ell)_\ell$ als Teilfolge von $(n^{k-1}_\ell)_\ell$ stets so festlegen, dass $n^{k-1}_{k-1} < n^k_k$ sichergestellt ist). Schliesslich betrachtet man die Diagonalfolge $(n^\ell_\ell)_\ell$ und hat

$$(\diamond) \qquad \text{für alle } r \in \mathbb{Q} \ : \quad F_{n^\ell_\ell}(r) \quad \overset{\ell\to\infty}{\longrightarrow} \quad G(r) \ .$$

Mit allen $F_{n^\ell_\ell}$, $\ell \geq 1$, ist die Funktion $G : \mathbb{Q} \to [0,1]$ nichtfallend. Setzt man nun

$$(\circ) \qquad F(t) \ := \ \inf\{G(r) : r \in \mathbb{Q}, \ r > t\} \ , \quad t \in \mathbb{R} \ ,$$

so definiert dies eine nichtfallende Funktion $F : \mathbb{R} \to [0,1]$.

2) Wir zeigen: die in ($\circ$) definierte Funktion F ist rechtsstetig. Sei dazu $t \in \mathbb{R}$ fest, sei $(t_n)_n$ eine Folge in $\mathbb{R}$ mit $t_n \downarrow t$; zu zeigen ist $F(t_n) \longrightarrow F(t)$. Betrachte zuerst rationale Folgen $(r_n)_n$ mit $r_n \downarrow t$. Da $G : \mathbb{Q} \to [0,1]$ nichtfallend, liefert ($\circ$)

$$(\circ\circ) \qquad \lim_{n\to\infty} G(r_n) \;=\; F(t)$$

unabhängig von der speziellen Wahl der rationalen Folge $(r_n)_n$. 'Begleitet' man nun die gegebene reelle Folge $t_n \downarrow t$ durch eine rationale Folge im Sinne von

$$(r_n)_n \subset \mathbb{Q} \;, \quad t_n < r_n \;\text{ für alle } n \;, \quad r_n \downarrow t \;,$$

so erhält man –da F nichtfallend ist und da ($\circ$) gilt– wegen ($\circ\circ$)

$$F(t) \;\le\; F(t_n) \;\le\; G(r_n) \;\longrightarrow\; F(t) \;, \quad n \to \infty$$

und damit die Rechtsstetigkeit von F.

3) Wir zeigen: für die in Schritt 1) gewählte Diagonalfolge $(n_\ell^\ell)_\ell$ gilt

$$(+) \qquad \lim_{\ell\to\infty} F_{n_\ell^\ell}(t) \;=\; F(t) \quad \text{für jedes } \; t \notin D_F \;;$$

dabei bezeichnet D_F die Menge der Unstetigkeitstellen von F. Alle Unstetigkeitsstellen von F sind Sprungstellen, und D_F ist höchstens abzählbar (vgl. Beweis 6.4', oder Beweisschritt 2) in 6.10). Sei $t \notin D_F$ fest, sei $\varepsilon > 0$ beliebig klein. Zu t und ε gibt es dann $\underline{r}, \overline{r}$ so dass

$$\underline{r}, \overline{r} \in \mathbb{Q} \;, \quad \underline{r} < t < \overline{r} \quad , \quad G(\underline{r})\,,\, G(\overline{r}) \;\in\; B_\varepsilon(F(t)) \;:$$

die Möglichkeit, $\overline{r}$ so zu wählen, folgt aus ($\circ\circ$); wegen Stetigkeit von F in t gibt es ein $\underline{t} < t$ mit $F(\underline{t}) > F(t) - \varepsilon$, und zu $\underline{t}$ wählt man mit ($\circ\circ$) ein rationales $\underline{r} \in (\underline{t}, t)$ mit der gewünschten Eigenschaft. Nach ($\diamond$) in Schritt 1) gilt entlang der Diagonalfolge $(n_\ell^\ell)_\ell$:

$$\begin{aligned}
\limsup_{\ell\to\infty} F_{n_\ell^\ell}(t) \;&\le\; \lim_{\ell\to\infty} F_{n_\ell^\ell}(\overline{r}) \;=\; G(\overline{r}) \;<\; F(t) + \varepsilon \\
\liminf_{\ell\to\infty} F_{n_\ell^\ell}(t) \;&\ge\; \lim_{\ell\to\infty} F_{n_\ell^\ell}(\underline{r}) \;=\; G(\underline{r}) \;>\; F(t) - \varepsilon \;.
\end{aligned}$$

Da $\varepsilon > 0$ beliebig war, ist hiermit (+) bewiesen, und Behauptung a) des Satzes gezeigt.

4) Beachte, dass jede in a) realisierbare Limesfunktion $F : \mathbb{R} \to [0,1]$ nach 1) - 3) zwar nichtfallend und rechtstetig ist, aber i.a. noch nicht die Verteilungsfunktion eines Wahrscheinlichkeitsmasses; dazu bleibt noch zu zeigen

$$(++) \qquad \lim_{x\to-\infty} F(x) = 0 \quad , \quad \lim_{x\to+\infty} F(x) = 1 \;.$$

Sei für jedes $\varepsilon > 0$ ein $K = K(\varepsilon)$ wie in der Straffheitsbedingung (S) gegeben. Für jede Teilfolge $(n_k)_k$ und jede Limesfunktion F nach Teil a) des Satzes mit

$$(+++) \qquad \lim_{k\to\infty} F_{n_k}(t) \;=\; F(t), \quad t \notin D_F$$

vergrössert man $K(\varepsilon)$ zu einem $\widetilde{K} > K(\varepsilon)$ mit der Eigenschaft $\pm\widetilde{K} \notin D_F$ (dies ist leicht möglich, denn D_F ist höchstens abzählbar). Dann gilt aber nach (+++) und (S)

$$\begin{aligned} F(-\widetilde{K}) + (1 - F(\widetilde{K})) \;&=\; \lim_{k\to\infty} \Big(F_{n_k}(-\widetilde{K}) + (1 - F_{n_k}(\widetilde{K}))\Big) \\ &\le\; \limsup_{n\to\infty} \Big(F_n(\widetilde{K}) + (1 - F_n(\widetilde{K}))\Big) \;<\; \varepsilon\,. \end{aligned}$$

Da $\varepsilon > 0$ beliebig war, hat F die Eigenschaft (++). Dies schliesst den Beweis des Satzes ab. □

Insbesondere haben wir eben bewiesen, dass unter der Straffheitsbedingung (S) aus 6.16 b) ein beliebiger Häufungspunkt F der Folge $(F_n)_n$ (im Sinne von (+++), entlang geeigneter Teilfolgen $(n_k)_k$ von $\mathbb{N}$) notwendig Verteilungsfunktion eines Wahrscheinlichkeitsmasses Q auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist. Mit 6.12 und 6.14 kann man den Satz 6.16 nun wie folgt aussprechen:

6.17 Folgerung (Auswahlsatz): Sei $(Q_n)_n$ eine straffe Folge von Wahrscheinlichkeitsmassen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann gibt es zu jeder Teilfolge $(n_k)_k$ eine weitere Teilfolge $(n_{k_\ell})_\ell$ und ein Wahrscheinlichkeitsmass Q auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so dass schwache Konvergenz in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ gilt:

$$Q_{n_{k_\ell}} \;\xrightarrow{w}\; Q \quad \text{für } \ell \to \infty\,.$$

*C. Verallgemeinerungen (ohne Beweise)

In diesem Teilkapitel geben wir ohne Beweise (aber mit Literaturverweisen) Verallgemeinerungen der in Teilkapitel B im Fall $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ bewiesenen Aussagen an, für Wahrscheinlichkeitsmasse auf (bzw. für Zufallsvariable mit Werten in) vollständig separablen metrischen Räumen.

Zuerst sind nach 6.1' Wahrscheinlichkeitsmasse auf vollständig separablen metrischen Räumen kompakt approximierbar, und man definiert analog zu 6.14:

6.18 Definition: Sei (E, d) separabel und vollständig. Eine Familie $\mathcal{Q}$ von Wahrscheinlichkeitsmassen auf $(E, \mathcal{B}(E))$ heisst *straff*, falls es zu jedem $\varepsilon > 0$ ein Kompaktum $K = K(\varepsilon)$ gibt mit

$$\sup_{Q\in\mathcal{Q}} Q(\,K^c\,) \;<\; \varepsilon\,.$$

Wegen 6.1' sind insbesondere endliche Familien von Wahrscheinlichkeitsmassen unter den Voraussetzungen aus 6.18 straff.

Um Straffheit nachzuweisen, muss man natürlich Kompakta in E (oder relativkompakte Teilmengen von E) beschreiben können (ein Beispiel: der Funktionenraum $E := C([0,1])$ versehen mit der Metrik $d(.,.)$ der gleichmässigen Konvergenz ist vollständig und separabel, und der Satz von Arcela-Ascoli charakterisiert die relativkompakten Mengen). Eine 'klassische' Referenz in diesem Zusammenhang ist das Buch von Billingsley (1968); weitere für die Stochastik wichtige Beispiele – etwa der Raum der càdlàg-Funktionen $[0,\infty) \to I\!R^d$ als Pfadraum für $I\!R^d$-wertige stochastische Prozesse mit Sprüngen – findet man in Jacod und Shiryaev (1987, Kapitel VI.1).

Kolmogorov und Prohorov haben um 1950 eine allgemeine Fassung des Auswahlsatzes 6.17 vorgelegt (siehe Billingsley 1968, S. 35–37):

6.19 Satz: Sei (E,d) separabel und vollständig. Sei $(Q_n)_n$ eine straffe Folge von Wahrscheinlichkeitsmassen auf $(E,\mathcal{B}(E))$. Dann gibt es zu jeder Teilfolge $(n_k)_k$ eine weitere Teilfolge $(n_{k_\ell})_\ell$ und ein Wahrscheinlichkeitsmass Q auf $(E,\mathcal{B}(E))$, so dass schwache Konvergenz in E gilt:

$$Q_{n_{k_\ell}} \quad \xrightarrow{w} \quad Q \quad \text{für } \ell \to \infty \; .$$

'Verteilungsgleiche Ersetzung' wie in 6.13 ist allgemein in vollständigen separablen metrischen Räumen möglich: für Wahrscheinlichkeitsmasse auf vollständigen separablen metrischen Räumen kann schwache Konvergenz wie in 6.13 durch Zufallsvariable unterlegt werden, welche – definiert auf einem geeigneten $(\widehat{\Omega},\widehat{\mathcal{A}},\widehat{P})$ – sogar fast sicher konvergieren. Dies braucht jedoch einen grundsätzlich anderen Beweisansatz (die in 6.10+6.11 bereitgestellten Methoden basieren auf der Verteilungsfunktion, für die es im allgemeinen keine Entsprechung gibt). Bei Ikeda und Watanabe (1989, S. 9–10) findet man eine allgemeine Formulierung:

6.20 Satz: Sei (E,d) separabel und vollständig, seien Q_n, $n \ge 1$, Q Wahrscheinlichkeitsmasse auf $(E,\mathcal{B}(E))$. Gilt

$$Q_n \overset{w}{\to} Q \;\; \text{(schwache Konvergenz in } E\text{)}$$

so gibt es einen Wahrscheinlichkeitsraum $(\widehat{\Omega},\widehat{\mathcal{A}},\widehat{P})$ und Zufallsvariable $\widehat{X}_n$, $n \ge 1$, $\widehat{X}$ auf $(\widehat{\Omega},\widehat{\mathcal{A}})$ mit Werten in $(E,\mathcal{B}(E))$ so dass

$$\begin{cases} Q_n = \mathcal{L}(\widehat{X}_n|\widehat{P}) \; , \; n \ge 1 \; , \quad Q = \mathcal{L}(\widehat{X}|P) \; , \\ \widehat{X}_n \;\longrightarrow\; \widehat{X} \quad \widehat{P}\text{-fast sicher für } n \to \infty \; . \end{cases}$$

Eine zu 6.11 analoge Charakterisierung der schwachen Konvergenz gilt in Euklidischen Räumen beliebiger Dimension $d \geq 1$; die Beweise (siehe Billingsley 1968, S. 17–18) sind analog zum eindimensionalen Fall aus Teilkapitel B:

6.21 Satz: Seien Q_n, $n \geq 1$, Q Wahrscheinlichkeitsmasse auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)$, $d \geq 1$, dann ist

$$Q_n \overset{w}{\to} Q \quad \text{(schwache Konvergenz in } \mathbb{R}^d\text{) für } n \to \infty$$

äquivalent zu Konvergenz der zugehörigen Verteilungsfunktionen an allen Stellen $t \in \mathbb{R}^d$, für die das d-dimensionale Intervall $(-\infty, t]$ Q-randlos ist.

6.21' Bemerkung: Der Helly'sche Auswahlsatz 6.16 gilt analog in beliebiger Dimension $d \geq 1$, mit (bis auf Notationen) demselben Beweis (siehe Billingsley 1968, S. 227).

Wir schliessen mit einem Hinweis, dass Straffheit in $\mathbb{R}^d$ nichts anderes ist als komponentenweise Straffheit aller Marginalien:

6.22 Bemerkung: Sei $(Q_n)_n$ eine Folge von Wahrscheinlichkeitsmassen auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. $(Q_n)_n$ ist straff in $\mathbb{R}^d$ im Sinne von 6.18, wenn es für jedes $\varepsilon > 0$ ein $K = K(\varepsilon) < \infty$ gibt mit der Eigenschaft

$$\sup_{n \geq 1} Q_n \left((\mathop{\times}_{i=1}^{d} [-K, K])^c \right) \; < \; \varepsilon \, .$$

Schreibt man Q_i für die Marginalverteilungen von Q, d.h. für die Bildmasse von Q unter den Projektionsabbildungen $\pi_i : (x_1, \ldots, x_d) \longrightarrow x_i$, so gilt für jedes $1 \leq j \leq d$

$$\{x : |x_j| > K\} \; \subset \; \{x : \max_{1 \leq i \leq d} |x_i| > K\} \; \subset \; \{x : |x_i| > K \text{ für ein } 1 \leq i \leq d\}$$

und damit

$$Q_n^j \left([-K, K]^c \right) \; \leq \; Q_n \left((\mathop{\times}_{i=1}^{d} [-K, K])^c \right) \; \leq \; \sum_{i=1}^{d} Q_n^i \left([-K, K]^c \right) \, .$$

Folglich ist Straffheit in $\mathbb{R}$ der Folgen $(Q_n^i)_n$ aller Marginalverteilungen, $1 \leq i \leq d$, dasselbe wie Straffheit von $(Q_n)_n$ in $\mathbb{R}^d$. $\square$