Empirische Prozesse

Lutz Dümbgen
Universität Bern

Januar 2024

Vorwort

Dieses Skriptum erstellte ich für eine gleichnamige vierstündige Kursusvorlesung im Sommersemester 1997 in Heidelberg. Es ging hervor aus meiner Ausarbeitung einer Spezialvorlesung "Empirische Prozesse" von D.W. Müller im Sommersemester 1988, welche sich an Pollard (1984) anlehnte. Seitdem verwendete ich das Skriptum in anderen Kursen (Lübeck, Bern, Göttingen) und ergänzte es um Teile, die meines Erachtens für Studierende der Mathematischen Statistik wissenswert sind.

Insbesondere wird die allgemeine Theorie der Konvergenz in Verteilung komplett behandelt, wobei ich aber bewusst auf die Einführung von "messbaren Einhüllenden" verzichtete, da dies meines Erachtens die Theorie unnötig kompliziert macht.

Bei Ulrich Erlenmaier, Sandra Freitag, Markus Fuchs, Martin Kania, Dirk Klingbiel, Alexandre Mösching, Johannes Schmidt-Hieber, Wiebke Schreiber, Dominic Schuhmacher, Philip Stange, Kaspar Stucki und Uwe Wehrspohn möchte ich mich für gute Fragen, zahlreiche Korrekturen und Verbesserungsvorschläge bedanken.

Für Fehlermeldungen und Kommentare bin ich jederzeit dankbar.

Bern, im Januar 2024, *Lutz Dümbgen*

Inhalt

Kapitel 1

Einführung

Die Theorie der empirischen Prozesse verbindet Resultate und Techniken aus der Wahrscheinlichkeitstheorie, insbesondere für Gaußsche Prozesse und Zufallselemente mit Werten in Banachräumen, mit Fragestellungen der nichtparametrischen Statistik. In diesem Kapitel wollen wir zwei wichtige Klassen von Prozessen vorstellen und den abstrakten Rahmen einführen, in welchem wir diese Prozesse behandeln werden.

1.1 Empirische Prozesse

Seien $X_1, X_2, \ldots, X_n$ stochastisch unabhängige, identisch verteilte Zufallsvariablen mit Werten in einem messbaren Raum $(\mathcal{X}, \mathcal{B})$ und Verteilung P. Ein Schätzer für P ist die empirische Verteilung

$$\hat{P}_n \ := \ \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

Dabei bezeichnet δ_x das Dirac-Maß im Punkt x. Für eine messbare Menge $D \subset \mathcal{X}$ ist also

$$\hat{P}_n(D) \ = \ \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \in D\}$$

ein Schätzer für $P(D)$, und für eine bezüglich P integrierbare Funktion $f : \mathcal{X} \to \mathbb{R}$ ist

$$\hat{P}_n(f) := \int f \, d\hat{P}_n \ = \ \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

ein Schätzer für das Integral

$$P(f) \ := \ \int f \, dP.$$

Die Frage ist nun, wie präzise dieser Schätzer $\hat{P}_n$ ist.

Konsistenz. Bekanntlich gilt für jede feste Funktion f aus $\mathcal{L}^1(P)$, der Menge aller bezüglich P integrierbaren Funktionen:

$$\mathbb{E}\big|\hat{P}_n(f) - P(f)\big| \;\to\; 0 \quad \text{und} \quad \hat{P}_n(f) \;\to\; P(f) \text{ fast sicher}$$

nach dem schwachen bzw. starken Gesetz der großen Zahlen. Sofern nichts anderes gesagt wird, beziehen sich asymptotische Aussagen immer auf den Fall, dass $n \to \infty$.

Aber viele statistische Verfahren basieren auf $\hat{P}_n(f)$ für unendlich viele Funktionen f, und die punktweise Konvergenz ist nicht ausreichend. Sei daher $\mathcal{F}$ eine Familie von Funktionen $f \in \mathcal{L}^1(P)$. Eine der zentralen Fragen, mit denen wir uns beschäftigen werden, ist: Unter welchen Voraussetzungen an P und $\mathcal{F}$ gilt:

$$\|\hat{P}_n - P\|_{\mathcal{F}} \;:=\; \sup_{f \in \mathcal{F}} \big|\hat{P}_n(f) - P(f)\big| \;\to\; 0$$

im Mittel bzw. fast sicher? Da man Mengen mit ihren Indikatorfunktionen identifizieren kann, formulieren wir viele Resultate für Funktionenfamilien, was den Spezialfall von Mengenfamilien beinhaltet.

Beispiel 1.1 (Data-Mining). Angenommen eine Handelskette möchte ihre “typischen Kunden” genauer untersuchen und beschreiben. Zu diesem Zweck werden die Kassenzettel von n Kunden ausgewertet. Für den i–ten Kunden enthalte $X_i = (X_i(j))_{j=1}^q$ seine absoluten oder relativen Ausgaben für Produkte aus q verschiedenen Kategorien. Nun betrachtet man alle möglichen Kombinationen der q Variablen. Genauer gesagt, berechnet man $\hat{P}_n(D)$ für verschiedenste Rechtecke

$$D \;=\; D_1 \times D_2 \times \cdots \times D_q$$

in $[0,\infty)^q$; das heißt, $D_1, D_2, \ldots, D_q$ sind Intervalle in $[0,\infty)$. Zum Beispiel ist der relative Anteil von Kunden, welche höchstens den Betrag a für Waren aus Kategorie 1 und 2 aber nichts in Kategorie 3 ausgaben, gleich $\hat{P}_n(D)$ mit $D = [0,a] \times [0,a] \times \{0\} \times [0,\infty)^{q-3}$.

Nun betrachten wir die n untersuchten Kunden als zufällige Stichprobe aus der Population aller Kunden, und P sei die Verteilung aller Kundenprofile in $[0,\infty)^q$. Dann stellt sich die Frage, wie groß die maximale Abweichung $\|\hat{P}_n - P\|_{\mathcal{D}}$ ist, wenn $\mathcal{D}$ die Familie aller Rechtecke in $[0,\infty)^q$ ist.

Beispiel 1.2 (Projection Pursuit). Sei $\mathcal{X} = \mathbb{R}^q$ mit $q >> 2$. Um den hochdimensionalen Datensatz $(X_i)_{i=1}^n$ zu analysieren, betrachtet man mitunter Scatterplots von beliebig vielen Projektionen auf den $\mathbb{R}^2$, wobei die Daten möglicherweise erst noch affin linear transformiert werden. Folglich betrachten wir für diverse Paare (v,w) von linear unabhängigen Vektoren $v, w \in \mathbb{R}^q$ einen Scatterplot der Punkte $(v^\top X_i, w^\top X_i) \in \mathbb{R} \times \mathbb{R}$. Wenn nun für ein bestimmtes Paar (v,w) eine interessante Struktur sichtbar ist, zum Beispiel zwei deutlich getrennte Punktwolken, dann fragt man sich natürlich, ob diese real ist oder ein Artefakt aufgrund von Stichprobenfehlern. Hierzu könnte man $\|\hat{P}_n - P\|_{\mathcal{D}}$ analysieren, wobei $\mathcal{D}$ aus allen Mengen der Form

$$\Big\{x \in \mathbb{R}^q : (v^\top x, w^\top x) \in B\Big\}$$

mit Vektoren $v, w \in \mathbb{R}^q$ und einem Rechteck $B \subset \mathbb{R} \times \mathbb{R}$ besteht.

Asymptotische Normalität. Für jede Funktion f aus $\mathcal{L}^2(P)$, dem Raum aller bezüglich P quadratintegrierbaren Funktionen, folgt aus dem Zentralen Grenzwertsatz, dass

$$\mathcal{L}\big(\sqrt{n}(\hat{P}_n - P)(f)\big) \;\to_w\; \mathcal{N}\big(0, P(f^2) - P(f)^2\big).$$

Dabei steht '$\to_w$' für schwache Konvergenz (siehe auch Kapitel 7), und $\mathcal{N}(\mu, \sigma^2)$ ist die Normalverteilung mit Mittelwert μ und Varianz σ^2. Eine naheliegende Frage ist nun, unter welchen Voraussetzungen an P und eine Familie $\mathcal{F} \subset \mathcal{L}^2(P)$ ein Wahrscheinlichkeitsmaß L existiert, so dass

$$\mathcal{L}\big(\sqrt{n}\|\hat{P}_n - P\|_{\mathcal{F}}\big) \;\to_w\; L\,?$$

Wenn ja, wie kann man L charakterisieren? Gilt eine analoge Aussage für andere Funktionale von $\sqrt{n}(\hat{P}_n - P)$ wie beispielsweise

$$n \int (\hat{P}_n - P)(f)^2 \, M(df)$$

mit einem gegebenem Maß M auf $\mathcal{F}$?

Beispiel 1.3 (Goodness-of-Fit-Tests). Um die Nullhypothese, dass $P = P_o$, versus die Alternativhypothese, dass $P \neq P_o$, zu testen, kann man für eine hinreichend große Familie $\mathcal{F}$ von Funktionen die Testgröße $T := \sqrt{n}\|\hat{P}_n - P_o\|_{\mathcal{F}}$ oder $T := n\int(\hat{P}_n - P_o)(f)^2\, M(df)$ berechnen. Dabei ist M ein gewisses Maß auf der Menge $\mathcal{F}$. Die Frage ist dann, ob T im Falle von $P = P_o$ eine Grenzverteilung hat, und wie man deren Quantile bestimmen kann. Ferner möchte man wissen, wie sich T im Falle von $P \neq P_o$ verhält.

Bezeichnung. Stochastische Prozesse (d.h. zufällige Funktionen) wie $\big(\hat{P}_n(f)\big)_{f\in\mathcal{F}}$ oder die zentrierte Variante $\big((\hat{P}_n - P)(f)\big)_{f\in\mathcal{F}}$ oder die standardisierte Variante $\big(\sqrt{n}(\hat{P}_n - P)(f)\big)_{f\in\mathcal{F}}$ nennt man *empirische Prozesse* indiziert durch die Funktionenklasse $\mathcal{F}$. Ein wichtiger Spezialfall sind *mengenindizierte empirische Prozesse* $\big(\hat{P}_n(D)\big)_{D\in\mathcal{D}}$, wobei $\mathcal{D}$ eine Familie messbarer Teilmengen von $\mathcal{X}$ ist.

1.2 Partialsummenprozesse

Eine zweite wichtige Klasse von stochastischen Prozessen sind *Partialsummenprozesse*. Dabei betrachtet man feste Punkte $x_{n1}, x_{n2}, \ldots, x_{nn} \in \mathcal{X}$ und unabhängige, identisch verteilte Zufallsvariablen $Y_{n1}, Y_{n2}, \ldots, Y_{nn} \in \mathbb{R}$ mit Erwartungswerten $\mu_{n1}, \mu_{n2}, \ldots, \mu_{nn}$. Für eine Familie $\mathcal{F}$ von Funktionen auf $\mathcal{X}$ sei dann $\hat{S}_n = \big(\hat{S}_n(f)\big)_{f\in\mathcal{F}}$ definiert durch

$$\hat{S}_n(f) \;:=\; c_n \sum_{i=1}^{n} f(x_{ni}) Y_{ni}$$

mit einer Normierungskonstanten $c_n > 0$. Die Frage ist nun, wie sehr sich $\hat{S}_n$ und $\mathbb{E}\,\hat{S}_n$ unterscheiden.

Beispiel 1.4 (Bildverarbeitung). Die x_{ni} seien Punkte auf einem regelmäßigen Gitter im $\mathbb{R}^2$, und

$$Y_{ni} \ = \ \beta(x_{ni}) + \epsilon_{ni},$$

wobei $\beta : \mathbb{R}^2 \to \mathbb{R}$ eine unbekannte Bildfunktion ist, und $\epsilon_{n1}, \epsilon_{n2}, \ldots, \epsilon_{nn}$ sind unabhängige Zufallsvariablen (Rauschen) mit Mittelwert Null.

Um das Rauschen zu eliminieren oder bestimmte Eigenschaften der Bildfunktion β zu extrahieren, wendet man oft einen linearen Filter an. Das heißt, man berechnet für eine gewisse Filterfunktion $F : \mathbb{R}^2 \to \mathbb{R}$ und verschiedene Punkte $t \in \mathbb{R}^2$ Summen der Form

$$c_n \sum_{i=1}^{n} F(x_{ni} - t) Y_{ni}.$$

Je nach Gestalt von F schätzt man mit dieser Summe zum Beispiel den Wert von β oder eine ihrer Ableitungen an der Stelle t.

1.3 Abstrakter Rahmen

Man kann sowohl empirische Prozesse als auch Partialsummenprozesse in einen abstrakten Rahmen einbetten. Hierzu seien $\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn}$ unabhängige stochastische Prozesse auf einer beliebigen Indexmenge $\mathcal{T}$. Dann betrachten wir den stochastischen Prozess

$$Z_n \ := \ \sum_{i=1}^{n} \phi_{ni}$$

auf $\mathcal{T}$. Die in Abschnitt 1.1 gestellten Fragen lauten nun:

Uniforme Konsistenz? Unter welchen Voraussetzungen konvergiert $\|Z_n - \mathbb{E}\, Z_n\|$ im Mittel oder in Wahrscheinlichkeit gegen Null?

Dabei ist $\mathbb{E}\, Z_n = \big(\mathbb{E}\, Z_n(t)\big)_{t \in \mathcal{T}}$ der punktweise Erwartungswert von Z_n, und für eine Funktion $x : \mathcal{T} \to \mathbb{R}$ setzen wir $\|x\| = \|x\|_{\mathcal{T}} := \sup_{t \in \mathcal{T}} |x(t)|$.

Grenzverteilungen? Unter welchen Voraussetzungen konvergiert die Verteilung eines gegebenen Funktionals $H(Z_n)$ schwach gegen ein Wahrscheinlichkeitsmaß? Wie kann man diese Grenzverteilung charakterisieren?

In den Abschnitten 1.1 bzw. 1.2 ist $\mathcal{T} = \mathcal{F}$ und

$$\phi_{ni}(f) \ := \ \frac{1}{n} f(X_i) \quad \text{oder} \quad \phi_{ni}(f) \ := \ \frac{1}{\sqrt{n}} (f(X_i) - P(f))$$

bzw.

$$\phi_{ni}(f) \ := \ c_n f(x_{ni}) Y_{ni}.$$

Kapitel 2

Uniforme Konsistenz mittels Approximationen

Bevor wir uns in dem abstrakten Rahmen aus Abschnitt 1.3 bewegen, behandeln wir in diesem und dem nächsten Kapitel einige klassische Resultate für geschätzte Verteilungen.

Im Folgenden sei P ein festes Wahrscheinlichkeitsmaß auf einem messbaren Raum $(\mathcal{X}, \mathcal{B})$, und seien $\mathcal{G}$ und $\mathcal{F}$ Teilmengen von $\mathcal{L}^1(P)$. Ferner sei $(\hat{P}_n)_n$ irgendeine Folge von zufälligen Wahrscheinlichkeitsmaßen auf $\mathcal{X}$, so dass $\int h\, d\hat{P}_n$ für alle $h \in \mathcal{G} \cup \mathcal{F}$ eine reellwertige Zufallsvariable ist.

Genau gesagt, betrachten wir einen Wahrscheinlichkeitsraum $(\Omega_n, \mathcal{A}_n, \mathrm{I\!P}_n)$ und eine Abbildung

$$\Omega_n \times \mathcal{B} \ni (\omega, B) \ \mapsto\ \hat{P}_n(B \,|\, \omega),$$

so dass $\hat{P}_n(\cdot \,|\, \omega)$ für jedes feste $\omega \in \Omega_n$ ein Wahrscheinlichkeitsmaß auf $(\mathcal{X}, \mathcal{B})$ darstellt, und wir nehmen an, dass $\omega \mapsto \int h(x)\, \hat{P}_n(dx \,|\, \omega)$ für jedes $h \in \mathcal{G} \cup \mathcal{F}$ eine reellwertige und messbare Funktion definiert. Wie üblich unterschlagen wir aber das Argument ω und schreiben einfach $\int h\, d\hat{P}_n$ anstelle von $\int h(x)\, \hat{P}_n(dx \,|\, \omega)$ oder $\int h\, d\hat{P}_n(\cdot \,|\, \omega)$.

Im Folgenden nehmen wir stets an, dass folgende Bedingung erfüllt ist:

(A) Für jedes feste $g \in \mathcal{G}$ gilt:

$$\int g\, d\hat{P}_n \ \to_* \ \int g\, dP.$$

Dabei steht "$\to_*$" entweder für Konvergenz in Wahrscheinlichkeit oder für fast sichere Konvergenz. Im letzteren Falle unterstellen wir, dass alle $\hat{P}_n$ auf ein und demselben Wahrscheinlichkeitsraum definiert sind.

Wenn beispielsweise $\hat{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ mit unabhängigen, identisch verteilten Zufallsvariablen $X_1, X_2, X_3, \ldots$ mit Verteilung P, dann ist Bedingung (A) erfüllt mit $\mathcal{G} = \mathcal{L}^1(P)$ und fast sicherer Konvergenz.

Die Frage ist nun, unter welchen Voraussetzungen an $\mathcal{G}$ und $\mathcal{F}$ aus Bedingung (A) folgt, dass

$$\|\hat{P}_n - P\|_{\mathcal{F}} \ \to_* \ 0. \tag{2.1}$$

Wir ignorieren hier die Frage der Messbarkeit von $\|\hat{P}_n - P\|_{\mathcal{F}}$. In unseren Beweisen werden wir $\|\hat{P}_n - P\|_{\mathcal{F}}$ durch Zufallsvariablen Y_n nach oben abschätzen, so dass $Y_n \to_* 0$.

2.1 Bracketing

Der folgende Satz liefert ein einfaches, aber oft erfolgreiches Kriterium, um (2.1) nachzuweisen. Dabei wird die Funktionenklasse $\mathcal{F}$ durch endliche Teilmengen von $\mathcal{G}$ approximiert.

Satz 2.1. *Angenommen, zu jedem $\epsilon > 0$ existieren $m \in \mathbb{N}$ und Funktionen $g_1, h_1, g_2, h_2, \ldots, g_m, h_m$ in $\mathcal{G}$ mit folgenden Eigenschaften:*

- *Für alle $i \le m$ ist $g_i \le h_i$ (punktweise) und $P(h_i - g_i) \le \epsilon$;*
- *zu jedem $f \in \mathcal{F}$ existiert ein $j \le m$ mit $g_j \le f \le h_j$.*

Dann impliziert (A) auch (2.1).

Anmerkung 2.2. Die obigen Paare (g_i, h_i) sind sogenannte “Brackets” (Klammern), daher der Name “Bracketing”.

Beweis von Satz 2.1. Für beliebiges festes $\epsilon > 0$ und Funktionen $g_1, h_1, g_2, h_2, \ldots, g_m, h_m$ mit den genannten Eigenschaften gilt: Ist $f \in \mathcal{F}$ mit $g_j \le f \le h_j$, so ist

$$\begin{aligned}
(\hat{P}_n - P)(f) \ &\le \ \hat{P}_n(h_j) - P(g_j) \ = \ (\hat{P}_n - P)(h_j) + P(h_j - g_j) \ \le \ (\hat{P}_n - P)(h_j) + \epsilon, \\
(\hat{P}_n - P)(f) \ &\ge \ \hat{P}_n(g_j) - P(h_j) \ = \ (\hat{P}_n - P)(g_j) - P(h_j - g_j) \ \ge \ (\hat{P}_n - P)(g_j) - \epsilon.
\end{aligned}$$

Mit $\mathcal{G}_\epsilon := \{g_1, \ldots, g_m, h_1, \ldots, h_m\}$ ist folglich

$$\|\hat{P}_n - P\|_{\mathcal{F}} \ \le \ \|\hat{P}_n - P\|_{\mathcal{G}_\epsilon} + \epsilon \ \to_* \ \epsilon.$$

Da $\epsilon > 0$ beliebig klein ist, impliziert dies (2.1). Man kann nämlich auch schreiben

$$\|\hat{P}_n - P\|_{\mathcal{F}} \ \le \ Y_n \ := \ \inf_{k \in \mathbb{N}} \bigl(\|\hat{P}_n - P\|_{\mathcal{G}_{1/k}} + 1/k\bigr),$$

und $Y_n \to_* 0$. □

Beispiel 2.3 (Satz von Glivenko-Cantelli für Verteilungsfunktionen). Es sei $\mathcal{X} = \mathbb{R}$ und $\mathcal{G} = \bigl\{1_{(-\infty,t]} : -\infty < t < \infty\bigr\} \cup \bigl\{1_{(-\infty,t)} : -\infty < t \le \infty\bigr\}$. Ferner sei $F(t) := P((-\infty, t])$ sowie $\hat{F}_n(t) := \hat{P}_n((-\infty, t])$ für $t \in \mathbb{R}$. Dann folgt aus (A), dass

$$\|\hat{F}_n - F\|_{\mathbb{R}} \ \to_* \ 0.$$

Beweis. Die Supremumsnorm $\|\hat{F}_n - F\|_{\mathbb{R}}$ ist gleich $\|\hat{P}_n - P\|_{\mathcal{F}}$ mit der Familie $\mathcal{F} := \big\{1_{(-\infty,t]} : t \in \mathbb{R}\big\}$. Für festes $\epsilon > 0$ sei $k := \lceil 1/\epsilon \rceil$ und

$$t_j \ := \ \min\big\{t \in \mathbb{R} : F(t) \geq j/k\big\} \quad (1 \leq j < k).$$

Ferner seien $t_0 := -\infty$ und $t_k := \infty$. Nun betrachten wir die $m = 2k - 1$ Brackets

$$(g_{2j-1}, h_{2j-1}) \ := \ \big(1_{(-\infty,t_{j-1}]}, 1_{(-\infty,t_j)}\big) \quad (1 \leq j \leq k)$$

sowie

$$(g_{2j}, h_{2j}) \ := \ \big(1_{(-\infty,t_j]}, (-\infty, t_j]\big) \quad (1 \leq j < k).$$

Dann ist $P(h_\ell - g_\ell) \leq \epsilon$ für $1 \leq \ell \leq m$, und zu jeder Funktion $f \in \mathcal{F}$ existiert ein Index $\ell \in \{1, \ldots, m\}$ mit $g_\ell \leq f \leq h_\ell$. □

Beispiel 2.4. Sei $\mathcal{X} = \mathbb{N}$, und sei $\mathcal{G} = \big\{1_{\{k\}} : k \in \mathbb{N}\big\}$. Dann folgt aus (A), dass

$$\|\hat{P}_n - P\|_{\mathcal{P}(\mathbb{N})} \ \to_* \ 0,$$

wobei $\mathcal{P}(\mathbb{N})$ die Menge aller Teilmengen von $\mathbb{N}$ bezeichnet.

Beweis. Man kann sich leicht davon überzeugen, dass (A) auch für die größere Funktionenfamilie $\bar{\mathcal{G}} := \big\{1_D : D \subset \mathbb{N},\ \#D < \infty \text{ oder } \#D^c < \infty\big\}$ erfüllt ist, wobei $D^c := \mathbb{N} \setminus D$. Nun wählen wir für festes $\epsilon > 0$ eine natürliche Zahl n_o, so dass für $S_o := \{1, \ldots, n_o\}$ gilt:

$$P(S_o^c) \leq \epsilon.$$

Dann gilt für beliebige Mengen $D \subset \mathbb{N}$:

$$1_{D \cap S_o} \ \leq \ 1_D \ \leq \ 1_{(D \cap S_o) \cup S_o^c}.$$

Dies führt zu $m = 2^{n_o}$ Brackets $(1_{D_o}, 1_{D_o \cup S_o^c})$, $D_o \subset S_o$, welche die (Un-)Gleichung

$$P(1_{D_o \cup S_o^c} - 1_{D_o}) \ = \ P(S_o^c) \ \leq \ \epsilon$$

erfüllen. □

2.2 Endlichdimensionale Approximationen

Ein etwas anderes aber ähnliches Kriterium wie Satz 2.1 liefert der folgende Satz:

Satz 2.5. *Angenommen, für jedes $\epsilon > 0$ existieren $m \in \mathbb{N}$ und Funktionen $g_1, g_2, \ldots, g_m, h$ in $\mathcal{G}$ mit folgenden Eigenschaften: $P(h) \leq \epsilon$, und für jedes $f \in \mathcal{F}$ existieren Zahlen $\lambda_1, \lambda_2, \ldots, \lambda_m \in [-1, 1]$ derart, dass*

$$\Big| f - \sum_{j=1}^{m} \lambda_j g_j \Big| \ \leq \ h.$$

Dann impliziert (A) auch (2.1).

Beweis von Satz 2.5. Für festes $\epsilon > 0$ seien $g_1, \ldots, g_m$ und h Funktionen mit den besagten Eigenschaften. Für $f \in \mathcal{F}$ und geeignete Zahlen $\lambda_1, \ldots, \lambda_m \in [-1, 1]$ ist dann

$$\begin{aligned}\big|(\hat{P}_n - P)(f)\big| &\le \Big|(\hat{P}_n - P)\Big(f - \sum_{j=1}^m \lambda_j g_j\Big)\Big| + \sum_{j=1}^m |\lambda_j| \big|(\hat{P}_n - P)(g_j)\big| \\ &\le (\hat{P}_n + P)(h) + \sum_{j=1}^m \big|(\hat{P}_n - P)(g_j)\big| \\ &\le 2\epsilon + \big|(\hat{P}_n - P)(h)\big| + \sum_{j=1}^m \big|(\hat{P}_n - P)(g_j)\big|.\end{aligned}$$

Mit $\mathcal{G}_\epsilon := \{g_1, \ldots, g_m, h\}$ ist folglich

$$\|\hat{P}_n - P\|_{\mathcal{F}} \le 2\epsilon + (m+1)\|\hat{P}_n - P\|_{\mathcal{G}_\epsilon} \to_* 2\epsilon.$$

Für $\epsilon \downarrow 0$ ergibt sich die Behauptung. □

Beispiel 2.6 (Schwache Konvergenz und Bounded-Lipschitz-Distanz). Sei $(\mathcal{X}, d)$ ein metrischer Raum. Im Folgenden betrachten wir die folgenden Teilmengen von $\mathbb{R}^{\mathcal{X}}$:

$$\begin{aligned}\mathcal{C}(\mathcal{X}) &:= \big\{f \in \mathbb{R}^{\mathcal{X}} : f \text{ stetig}\big\}, \\ \mathcal{C}_{\mathrm{b}}(\mathcal{X}) &:= \big\{f \in \mathcal{C}(\mathcal{X}) : \|f\|_{\mathcal{X}} < \infty\big\}, \\ \mathcal{C}_{\mathrm{Lip}}(\mathcal{X}) &:= \big\{f \in \mathbb{R}^{\mathcal{X}} : \|f\|_{\mathrm{Lip}} < \infty\big\}, \\ \mathcal{C}_{\mathrm{b,Lip}}(\mathcal{X}) &:= \big\{f \in \mathcal{C}_{\mathrm{Lip}}(\mathcal{X}) : \|f\|_{\mathcal{X}} < \infty\big\}\end{aligned}$$

mit der Seminorm

$$\|f\|_{\mathrm{Lip}} := \sup_{x,y \in \mathcal{X} : x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}.$$

Angenommen, (A) ist erfüllt mit $\mathcal{G} = \mathcal{C}_{\mathrm{b,Lip}}(\mathcal{X})$. Dann ergibt sich aus den Aufgaben 2.2 und 2.3, dass (A) auch mit $\mathcal{G} = \mathcal{C}_{\mathrm{b}}(\mathcal{X})$ erfüllt ist.

Satz 2.7 (Bounded-Lipschitz-Distanz). *Sei $(\mathcal{X}, d)$ ein separabler metrischer Raum, und sei*

$$\mathcal{F}_{BL} := \big\{f \in \mathbb{R}^{\mathcal{X}} : \|f\|_{\mathcal{X}} \le 1, \|f\|_{\mathrm{Lip}} \le 1\big\}.$$

Aus Bedingung (A) mit $\mathcal{G} = \mathcal{C}_{\mathrm{b,Lip}}(\mathcal{X})$ folgt, dass auch

$$\|\hat{P}_n - P\|_{\mathcal{F}_{BL}} \to_* 0.$$

Beweis von Satz 2.7. Sei $\{x_j : j \in \mathbb{N}\}$ eine dichte Teilmenge von $\mathcal{X}$. Für ein festes $\delta > 0$ definieren wir

$$h_{m,\delta} := \max_{j=1,\ldots,m} \big(1 - d(\cdot, x_j)/\delta\big)^+, \quad m \in \mathbb{N},$$

und $h_{0,\delta}(\cdot) := 0$. Dann gilt für die Funktionen $g_{j,\delta} := h_{j,\delta} - h_{j-1,\delta}$, $j \in \mathbb{N}$:

$$0 \le g_{j,\delta} \le \big(1 - d(\cdot, x_j)/\delta\big)^+, \quad \sum_{j=1}^m g_{j,\delta} = h_{m,\delta} \uparrow 1 \quad (m \to \infty).$$

Für eine beliebige Funktion $f \in \mathcal{F}_{BL}$ und $x \in \mathcal{X}$ ist nun

$$\begin{aligned}\Big|f(x) - \sum_{j=1}^m f(x_j) g_{j,\delta}(x)\Big| &\le \big|f(x) - f(x) h_{m,\delta}(x)\big| + \sum_{j=1}^m \big|f(x) - f(x_j)\big| g_{j,\delta}(x) \\ &\le 1 - h_{m,\delta}(x) + \sum_{j=1}^m \delta g_{j,\delta}(x) \\ &\le 1 - h_{m,\delta}(x) + \delta.\end{aligned}$$

Sei also $\epsilon > 0$ beliebig vorgegeben. Setzen wir $\delta := \epsilon/2$ und wählen wir $m \in \mathbb{N}$ derart, dass $P(1 - h_{m,\delta}) \le \delta$, dann erfüllen die Funktionen $g_j := g_{j,\delta}$ und $h := 1 - h_{m,\delta} + \delta$ die in Satz 2.5 genannten Bedingungen, wobei $\lambda_j := f(x_j)$. □

Korollar 2.8 (Topsoe). *Es sei $(\mathcal{X}, d)$ ein separabler metrischer Raum, und $\mathcal{D}$ sei eine Familie von Borelmengen $D \subset \mathcal{X}$ mit folgender Eigenschaft:*

$$\sup_{A \in \mathcal{D} \cup \mathcal{D}^c} P(U_\epsilon(A) \setminus A) \ \to \ 0 \quad \text{für } \epsilon \downarrow 0.$$

Dabei ist $\mathcal{D}^c := \{D^c : D \in \mathcal{D}\}$, und $U_\epsilon(A)$ bezeichnet die Menge aller $x \in \mathcal{X}$ mit $d(x, A) := \inf_{y \in A} d(x, y) < \epsilon$. Unter dieser Voraussetzung an P impliziert Bedingung (A) mit der Familie $\mathcal{G} = \mathcal{C}_{\mathrm{b,Lip}}(\mathcal{X})$, dass

$$\|\hat{P}_n - P\|_{\mathcal{D}} \ \to_* \ 0.$$

Beweis von Korollar 2.8. Für jede nichtleere Menge $A \subset \mathcal{X}$ ist $d(\cdot, A)$ lipschitzstetig mit Lipschitzkonstante 1. Für festes $\epsilon \in (0, 1]$ betrachten wir nun die Funktionenfamilie

$$\mathcal{F}_\epsilon \ := \ \Big\{\big(1 - d(\cdot, A)/\epsilon\big)^+ : \emptyset \neq A \subset \mathcal{X}\Big\}.$$

Dann besteht $\mathcal{F}_\epsilon$ aus lipschitzstetigen Funktionen mit Lipschitzkonstante ϵ^{-1} und Wertebereich $[0, 1] \subset [-\epsilon^{-1}, \epsilon^{-1}]$. Hieraus ergibt sich, dass

$$\|\hat{P}_n - P\|_{\mathcal{F}_\epsilon} \ \le \ \epsilon^{-1} \|\hat{P}_n - P\|_{\mathcal{F}_{BL}} \ \to_* \ 0.$$

Andererseits gelten für eine beliebige Menge $D \in \mathcal{D}$ mit $\emptyset \neq D \neq \mathcal{X}$ und die Funktionen

$$\begin{aligned} g_\epsilon &:= 1 - \big(1 - d(\cdot, D^c)/\epsilon\big)^+, \\ h_\epsilon &:= \big(1 - d(\cdot, D)/\epsilon\big)^+ \end{aligned}$$

folgende Tatsachen:

$$0 \ \le \ g_\epsilon \ \le \ 1_D \ \le \ h_\epsilon \ \le \ 1,$$

und

$$\{h_\epsilon - 1_D > 0\} \ \subset \ U_\epsilon(D) \setminus D, \quad \{1_D - g_\epsilon > 0\} \ \subset \ U_\epsilon(D^c) \setminus D^c.$$

Folglich ist

$$\|\hat{P}_n - P\|_{\mathcal{D}} \ \le \ \epsilon^{-1} \|\hat{P}_n - P\|_{\mathcal{F}_{BL}} + \sup_{A \in \mathcal{D} \cup \mathcal{D}^c} P(U_\epsilon(A) \setminus A) \ \to_* \ \sup_{A \in \mathcal{D} \cup \mathcal{D}^c} P(U_\epsilon(A) \setminus A).$$

Nach Voraussetzung konvergiert die rechte Seite gegen 0 für $\epsilon \downarrow 0$, und hieraus ergibt sich die Behauptung, dass $\|\hat{P}_n - P\|_{\mathcal{D}} \to_* 0$. □

Beispiel 2.9 (Satz von Glivenko-Cantelli für konvexe Mengen). Sei $\mathcal{D}$ die Menge aller konvexen Borelmengen $D \subset \mathbb{R}^q$. Falls

$$P(\partial D) \;=\; 0 \quad \text{für alle } D \in \mathcal{D},$$

kann man mit Hilfe von Blaschkes Auswahlsatz sogar zeigen, dass die Voraussetzung von Korollar 2.8 erfüllt ist. Folglich impliziert dann Bedingung (A) mit $\mathcal{G} = \mathcal{C}_{\mathrm{b,Lip}}(\mathcal{X})$, dass $\|\hat{P}_n - P\|_{\mathcal{D}} \to_* 0$.

Die Bedingung, dass $P(\partial D) = 0$ für konvexen Borelmengen $D \subset \mathbb{R}^d$, ist beispielsweise erfüllt, wenn P eine Lebesgue-Dichtefunktion hat.

Wir geben noch ein einfaches Gegenbeispiel, welches zeigt, dass diese Bedingung an P und $\mathcal{D}$ essentiell ist: Sei P die Gleichverteilung auf der (euklidischen) Einheitssphäre $\{x \in \mathbb{R}^2 : \|x\| = 1\}$, und sei $\hat{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ mit stochastisch unabhängigen Zufallsvariablen $X_1, X_2, X_3, \ldots$ mit Verteilung P. Für die konvexe Hülle D_n von $\{X_1, X_2, \ldots, X_n\}$ gelten dann die Gleichungen

$$\hat{P}_n(D_n) \;=\; 1 \quad \text{und} \quad P(D_n) \;=\; 0,$$

also $\|\hat{P}_n - P\|_{\mathcal{D}} \equiv 1$.

2.3 Aufgaben

Aufgabe 2.1. Seien $X_1, X_2, X_3, \ldots$ stochastisch unabhängige Zufallsvariablen mit Verteilung P auf $\mathbb{N}$, und sei $\hat{P}_n := n^{-1} \sum_{i=1}^n \delta_{X_i}$. Unter welcher Voraussetzung an P existiert eine Konstante $C(P) < \infty$ derart, dass

$$\mathbb{E}\,\|\hat{P}_n - P\|_{\mathcal{D}} \;\le\; C(P) n^{-1/2} \quad \text{für alle } n \in \mathbb{N}?$$

Vorschlag: Zeigen Sie zunächst, dass für zwei beliebige Wahrscheinlichkeitsverteilungen P, Q auf $\mathbb{N}$ gilt:

$$\|Q - P\|_{\mathcal{D}} \;=\; \frac{1}{2} \sum_{k \in \mathbb{N}} \big| P(\{k\}) - Q(\{k\}) \big|.$$

Aufgabe 2.2. Sei $(\mathcal{X}, d)$ ein metrischer Raum, und sei $f \in \mathbb{R}^{\mathcal{X}}$, so dass $f \ge h_o$ für eine Funktion $h_o \in \mathbb{R}^{\mathcal{X}}$ mit $L_o := \|h_o\|_{\mathrm{Lip}} < \infty$.

(a) Zeigen Sie, dass

$$f_L(x) \;:=\; \inf_{y \in \mathcal{X}} \big(f(y) + L\, d(x, y) \big)$$

für $L \geq L_o$ eine Funktion $f_L \in \mathbb{R}^{\mathcal{X}}$ mit folgenden Eigenschaften definiert:

$$\begin{aligned}
&\inf_{z\in\mathcal{X}} f(z) \ \leq\ f_L \ \leq\ f,\\
&f_L \ \geq\ h_o,\\
&f_L \ \geq\ h \quad \text{für jede Funktion } h : \mathcal{X} \to \mathbb{R} \text{ mit } h \leq f, \|h\|_{\mathrm{Lip}} \leq L,\\
&\|f_L\|_{\mathrm{Lip}} \ \leq\ L.
\end{aligned}$$

(b) Zeigen Sie, dass für jedes $x \in \mathcal{X}$ gilt: $f_L(x)$ ist monoton wachsend in L, und

$$\lim_{L\to\infty} f_L(x) \ =\ \liminf_{y\to x} f(y).$$

Aufgabe 2.3. Sei $(\mathcal{X}, d)$ ein metrischer Raum, und sei $f \in \mathcal{C}_{\mathrm{b}}(\mathcal{X})$. Angenommen, $(\hat{P}_n)_n$ erfüllt Bedingung (A) mit $\mathcal{G} = \mathcal{C}_{\mathrm{b,Lip}}(\mathcal{X})$. Zeigen Sie, dass dann auch Bedingung (A) mit $\mathcal{G} = \mathcal{C}_{\mathrm{b}}(\mathcal{X})$ erfüllt ist.

Tipp: Betrachten Sie

$$\begin{aligned}
\underline{f}_L(x) \ &:=\ \inf_{y\in\mathcal{X}}\bigl(f(y) + L\,d(x,y)\bigr),\\
\overline{f}_L(x) \ &:=\ \sup_{y\in\mathcal{X}}\bigl(f(y) - L\,d(x,y)\bigr)
\end{aligned}$$

und verwenden Sie Aufgabe 2.2.

Aufgabe 2.4. Sei $(\mathcal{X}, d)$ ein separabler metrischer Raum, und sei

$$\mathcal{F}_L \ :=\ \bigl\{f \in \mathbb{R}^{\mathcal{X}} : \|f\|_{\mathrm{Lip}} \leq 1\bigr\},$$

die Familie aller lipschitzstetigen Funktionen $f : \mathcal{X} \to \mathbb{R}$ mit Lipschitzkonstante 1. Für ein $x_o \in \mathcal{X}$ sei

$$P(d(\cdot, x_o)) \ <\ \infty.$$

Angenommen, (A) ist erfüllt mit

$$\mathcal{G} \ :=\ \mathcal{C}_{\mathrm{b,Lip}}(\mathcal{X}) \cup \{d(\cdot, x_o)\}.$$

Zeigen Sie, dass dann auch

$$\|\hat{P}_n - P\|_{\mathcal{F}_L} \ \to_*\ 0.$$

Kapitel 3

Symmetrisierungen

In diesem Kapitel geht es darum, für stochastische Prozesse Z auf einer Menge $\mathcal{T}$ die Supremumsnorm $\|Z\| = \|Z\|_{\mathcal{T}}$ nach oben abzuschätzen, indem man Z mit geeigneten Prozessen $\tilde{Z}$ vergleicht, deren Verteilung man mitunter besser handhaben kann.

Um sicherzustellen, dass Zufallselemente wie $\|Z\|_{\mathcal{T}}$ messbar sind, nehmen wir stets an, dass $\mathcal{T}$ abzählbar ist. In der Regel kann man sich mit Hilfe eines einfachen Approximationsargumentes auf diesen Fall zurückziehen.

3.1 Die erste Symmetrisierung

Oftmals ist $\|Z - Z'\|$ einfacher zu behandeln als $\|Z\|$, wenn Z' eine *unabhängige Kopie* von Z ist. Falls eine solche Kopie nicht existiert, kann man den ursprünglichen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit $Z : \Omega \to \mathbb{R}^{\mathcal{T}}$ in den Produktraum $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, \mathbb{P} \otimes \mathbb{P})$ einbetten vermöge

$$Z(t)(\omega_1, \omega_2) := Z(t)(\omega_1), \quad Z'(t)(\omega_1, \omega_2) := Z(t)(\omega_2).$$

Das nächste Lemma beschreibt zwei von vielen Möglichkeiten, wie man von $\|Z - Z'\|$ auf $\|Z\|$ schließen kann, wobei Z' nicht unbedingt die gleiche Verteilung wie Z haben muss.

Lemma 3.1. *Seien Z, Z' unabhängige stochastische Prozesse auf $\mathcal{T}$.*

(a) *Angenommen, es existieren Konstanten $\delta, \beta > 0$, so dass*

$$\mathbb{P}\big(|Z'(t)| \le \delta\big) \ \ge \ \beta$$

für beliebige $t \in \mathcal{T}$. Dann ist

$$\mathbb{P}\big(\|Z\| > \eta + \delta\big) \ \le \ \beta^{-1}\, \mathbb{P}\big(\|Z - Z'\| > \eta\big) \quad \text{für alle } \eta \ge 0.$$

(b) *Angenommen, $\mathbb{E}\, Z'$ existiert in $\mathbb{R}^{\mathcal{T}}$. Dann ist*

$$\mathbb{E}\, \Psi(\|Z - \mathbb{E}\, Z'\|) \ \le \ \mathbb{E}\, \Psi(\|Z - Z'\|)$$

für jede konvexe und monoton wachsende Funktion $\Psi : [0, \infty) \to [0, \infty)$.

Beweis von Lemma 3.1. Für Teil (a) sei ohne Einschränkung $\mathcal{T} \subset \mathbb{N}$, und sei

$$\tau = \tau(Z) \ := \ \begin{cases} \min\bigl\{t \in \mathcal{T} : |Z(t)| > \eta + \delta\bigr\} & \text{falls } \|Z\| > \eta + \delta, \\ \min(\mathcal{T}) & \text{sonst.} \end{cases}$$

Dann ist τ eine $\mathcal{T}$-wertige Zufallsvariable, so dass $\|Z\| > \eta+\delta$ genau dann, wenn $|Z(\tau)| > \eta+\delta$. Nun ist

$$\begin{aligned} \mathbb{P}(\|Z\| > \eta + \delta) \ &= \ \mathbb{P}(|Z(\tau)| > \eta + \delta) \\ &= \ \mathbb{E}\bigl(1\{|Z(\tau)| > \eta + \delta\}\bigr) \\ &\le \ \mathbb{E}\bigl(1\{|Z(\tau)| > \eta + \delta\}\, \beta^{-1}\, \mathbb{P}(|Z'(\tau)| \le \delta \,|\, Z)\bigr) \\ &= \ \beta^{-1}\, \mathbb{E}\bigl(\mathbb{P}\bigl(|Z(\tau)| > \eta + \delta, |Z'(\tau)| \le \delta \bigm| Z\bigr)\bigr) \\ &= \ \beta^{-1}\, \mathbb{P}\bigl(|Z(\tau)| > \eta + \delta, |Z'(\tau)| \le \delta\bigr) \\ &\le \ \beta^{-1}\, \mathbb{P}\bigl(\bigl|Z(\tau) - Z'(\tau)\bigr| > \eta\bigr) \\ &\le \ \beta^{-1}\, \mathbb{P}(\|Z - Z'\| > \eta). \end{aligned}$$

Für den Beweis von Teil (b) benötigt man die Jensensche Ungleichung: Für ein Wahrscheinlichkeitmaß Q auf $\mathbb{R}$ mit $\int |x|\, Q(dx) \ < \ \infty$ und eine konvexe Funktion $h : \mathbb{R} \to \mathbb{R}$ ist stets $h\bigl(\int x\, Q(dx)\bigr) \le \int h(x)\, Q(dx)$.

Vermöge $\Psi(-r) := \Psi(r)$ für $r > 0$ wird Ψ eine gerade, konvexe Funktion auf $\mathbb{R}$, und

$$\begin{aligned} \mathbb{E}\, \Psi(\|Z - \mathbb{E}\, Z'\|) \ &= \ \mathbb{E} \sup_{t \in \mathcal{T}} \Psi\bigl(Z(t) - \mathbb{E}\, Z'(t)\bigr) \\ &= \ \mathbb{E} \sup_{t \in \mathcal{T}} \Psi\Bigl(\mathbb{E}\bigl(Z(t) - Z'(t) \bigm| Z\bigr)\Bigr) \\ &\le \ \mathbb{E} \sup_{t \in \mathcal{T}} \mathbb{E}\Bigl(\Psi\bigl(Z(t) - Z'(t)\bigr) \Bigm| Z\Bigr) \\ &\le \ \mathbb{E}\, \mathbb{E}\Bigl(\sup_{t \in \mathcal{T}} \Psi\bigl(Z(t) - Z'(t)\bigr) \Bigm| Z\Bigr) \\ &= \ \mathbb{E}\, \Psi(\|Z - Z'\|). \end{aligned}$$

□

Erste Anwendung auf empirische Verteilungsfunktionen

Seien $X_1, X_2, \ldots, X_n$ stochastisch unabhängige, reellwertige Zufallsvariablen mit Verteilungsfunktion F, und sei $\hat{F}_n$ die entsprechende empirische Verteilungsfunktion,

$$\hat{F}_n(t) \ = \ \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \le t\}.$$

P. Massart (1990) bewies folgende Ungleichung für $\|\hat{F}_n - F\| = \|\hat{F}_n - F\|_{\mathbb{R}}$:

$$\mathbb{P}\bigl(\|\hat{F}_n - F\| \ge \eta\bigr) \ \le \ 2 \exp(-2n\eta^2) \quad \text{für beliebige } \eta \ge 0.$$

Der Beweis dieser Ungleichung ist sehr technisch. In diesem Abschnitt werden wir mit Hilfe der ersten Symmetrisierung eine schwächere, doch ähnliche Ungleichung beweisen:

$$\mathbb{P}\bigl(\|\hat{F}_n - F\| \ge \eta\bigr) \ \le \ e^2 \exp(-n\eta^2) \quad \text{für beliebige } \eta \ge 0. \tag{3.1}$$

Quantiltransformation. Mit Hilfe eines Standardtricks können wir uns auf einen Spezialfall zurückziehen. Sei nämlich F^{-1} die Quantilfunktion von F auf $(0,1)$, das heißt,

$$F^{-1}(u) \; := \; \min\bigl\{t \in \mathbb{R} : F(t) \ge u\bigr\}.$$

Dann gilt für beliebige $u \in (0,1)$ und $t \in \mathbb{R}$:

$$F^{-1}(u) \;\le\; t \quad \text{genau dann, wenn} \quad u \;\le\; F(t).$$

Hieraus ergeben sich folgende Tatsachen: Seien $U_1, U_2, U_3, \ldots$ stochastisch unabhängige, nach $\mathcal{U}[0,1]$ verteilte Zufallsvariablen. Dann sind die Variablen $X_i := F^{-1}(U_i)$ stochastisch unabhängig und nach F verteilt. Definiert man

$$\hat{G}_n(v) \; := \; \frac{1}{n}\sum_{i=1}^{n} 1\{U_i \le v\}$$

und $G(v) := v$ für $v \in [0,1]$, dann gilt mit den speziell konstruierten Zufallsvariablen X_i die Darstellung:

$$\bigl(\hat{F}_n(t)\bigr)_{t\in\mathbb{R}} \;=\; \bigl(\hat{G}_n(F(t))\bigr)_{t\in\mathbb{R}}.$$

Insbesondere ist

$$\|\hat{F}_n - F\| \;\le\; \|\hat{G}_n - G\|_{[0,1]}$$

mit Gleichheit, falls F stetig ist.

Symmetrisierung von $\hat{G}_n - G$ und Spiegelungsprinzip. Mit

$$\hat{G}_n'(v) \; := \; \frac{1}{n}\sum_{i=n+1}^{2n} 1\{U_i \le v\}$$

ergibt sich aus Lemma 3.1 (b), dass

$$\mathbb{E}\,\Psi\bigl(\|\hat{G}_n - G\|_{[0,1]}\bigr) \;\le\; \mathbb{E}\,\Psi\bigl(\|\hat{G}_n - \hat{G}_n'\|_{[0,1]}\bigr)$$

für jede monoton wachsende und konvexe Funktion $\Psi : [0,\infty) \to [0,\infty)$. Dabei nutzen wir aus, dass die Funktionen $\hat{G}_n, \hat{G}_n'$ rechtsseitig stetig sind, so dass die hier auftretenden Suprema über $[0,1]$ durch Suprema über die abzählbare Menge $[0,1] \cap \mathbb{Q}$ ersetzt werden können.

Aus Symmetriegründen ist

$$\mathbb{P}\bigl(\|\hat{G}_n - \hat{G}_n'\|_{[0,1]} \ge \eta\bigr) \;\le\; 2\,\mathbb{P}\Bigl(\sup_{v\in[0,1]} (\hat{G}_n - \hat{G}_n')(v) \ge \eta\Bigr) \quad \text{für beliebige } \eta \ge 0.$$

Daher untersuchen wir nun die Verteilung von $\sup_{[0,1]}(\hat{G}_n - \hat{G}_n')$. Zu diesem Zweck betrachten wir die Ordnungsstatistiken $0 < U_{(1)} < U_{(2)} < \cdots < U_{(2n)} < 1$ der Zufallsvariablen $U_1, U_2, \ldots, U_{2n}$. Die Differenz $\hat{G}_n - \hat{G}_n'$ ist eine Treppenfunktion, welche auf $[0, U_{(1)})$ und

$[U_{(2n)}, 1]$ den Wert Null annimmt und auf den Intervallen $[U_{(k)}, U_{(k+1)})$, $1 \leq k < 2n$, konstant ist. An der Stelle $U_{(k)}$, $1 \leq k \leq 2n$, ändert sich ihr Wert um $+1/n$ oder $-1/n$, je nachdem, ob $U_{(k)}$ in der Menge $\{U_1, \ldots, U_n\}$ oder $\{U_{n+1}, \ldots, U_{2n}\}$ liegt. Nun betrachten wir das zufällige Tupel $W = (W_k)_{k=0}^{2n} \in \mathbb{Z}^{\{0,1,\ldots,2n\}}$ mit $W_k := n(\hat{G}_n - \hat{G}'_n)(U_{(k)})$, wobei $U_{(0)} := 0$. Dieses Tupel liegt in der Menge

$$\mathcal{W} := \Big\{ w \in \mathbb{Z}^{\{0,1,\ldots,2n\}} : w_0 = 0, |w_k - w_{k-1}| = 1 \text{ für } 1 \leq k \leq 2n \Big\}$$

und erfüllt die Gleichung $W_{2n} = 0$. Am besten stellt man sich ein Tupel $w \in \mathcal{W}$ als “Zickzackpfad”, der die Punkte $(0, w_0)$, $(1, w_1)$, …, $(2n, w_{2n})$ miteinander verbindet, vor; siehe Abbildung 3.1. Aus Symmetriegründen ist W uniform verteilt auf der Menge $\{w \in \mathcal{W} : w_{2n} = 0\}$. Folglich gilt für $\ell \in \{0, 1, \ldots, n\}$:

$$\begin{aligned}
\mathbb{P}\Big(\sup_{[0,1]} (\hat{G}_n - \hat{G}'_n) \geq \frac{\ell}{n} \Big) &= \mathbb{P}\Big(\max_{k=0,1,\ldots,2n} W_k \geq \ell \Big) \\
&= \frac{\#\Big\{ w \in \mathcal{W} : \max_k w_k \geq \ell, w_{2n} = 0 \Big\}}{\#\big\{ w \in \mathcal{W} : w_{2n} = 0 \big\}}.
\end{aligned}$$

Um den Zähler auf der rechten Seite zu bestimmen, verwenden wir nun das Spiegelungsprinzip: Zu jedem Pfad $w \in \mathcal{W}$ definieren wir seine Spiegelung $\tilde{w}$ wie folgt:

$$\tilde{w}_k := \begin{cases} w_k & \text{falls } \max_{j \leq k} w_j < \ell, \\ \ell - (w_k - \ell) = 2\ell - w_k & \text{falls } \max_{j \leq k} w_j \geq \ell. \end{cases}$$

Diese Abbildung $w \mapsto \tilde{w}$ ist eine bijektive Abbildung von $\mathcal{W}$ nach $\mathcal{W}$, und zwar ist $\tilde{\tilde{w}} = w$. Illustriert wird sie in Abbildung 3.1 für $n = 50$ und $\ell = 7$. Ein Pfad $w \in \mathcal{W}$ erfüllt die Bedingungen $\max_{k \leq 2n} w_k \geq \ell$ und $w_{2n} = 0$ genau dann, wenn $\tilde{w}_{2n} = 2\ell$. Daher ist

$$\mathbb{P}\Big(\sup_{[0,1]} (\hat{G}_n - \hat{G}'_n) \geq \frac{\ell}{n} \Big) = \frac{\#\Big\{ w \in \mathcal{W} : w_{2n} = 2\ell \Big\}}{\#\big\{ w \in \mathcal{W} : w_{2n} = 0 \big\}} = \binom{2n}{n+\ell} \Big/ \binom{2n}{n}. \tag{3.2}$$

Denn $w \in \mathcal{W}$ erfüllt die Bedingung $w_{2n} = 2\ell$ genau dann, wenn $w_k - w_{k-1} = 1$ für genau $n + \ell$ “Zeitpunkte” $k \in \{1, 2, \ldots, 2n\}$.

Eine Abschätzung. Die in (3.2) gegebene Formel ist für unsere Zwecke noch zu unhandlich. Wir verwenden daher folgende Abschätzung für $\ell \in \{1, 2, \ldots, n\}$:

$$\binom{2n}{n+\ell} \Big/ \binom{2n}{n} = \frac{n!n!}{(n+\ell)!(n-\ell)!} = \prod_{i=0}^{\ell-1} \frac{n-i}{n+1+i} = (1 + \ell/n)^{-1} \prod_{i=0}^{\ell-1} \frac{n-i}{n+i}$$

ist gleich

$$\begin{aligned}
\exp\Big(-\log(1 + \ell/n) + \sum_{i=1}^{\ell-1} \log\Big(\frac{1 - i/n}{1 + i/n} \Big) \Big) &\leq \exp\Big(-\log(1 + \ell/n) - 2 \sum_{i=1}^{\ell-1} i/n \Big) \\
&= \exp\big(\ell/n - \log(1 + \ell/n) - \ell^2/n \big) \\
&\leq \exp\big(1 - \log(2) - \ell^2/n \big).
\end{aligned}$$

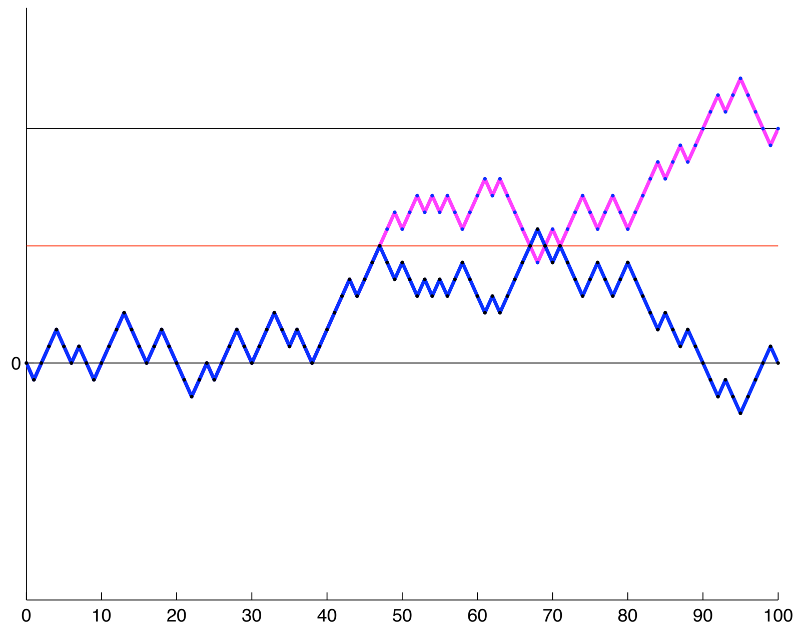

Abbildung 3.1: Ein Pfad $w \in \mathcal{W}$ und seine Spiegelung $\tilde{w}$ an einer Horizontalen in Höhe ℓ.

Dabei verwendeten wir die Abschätzung $\log((1-x)/(1+x)) \leq -2x$ für $x \in [0,1)$ sowie die Tatsache, dass $t - \log(1+t)$ monoton wachsend in $t \in [0,1]$ ist. Da $\sup_{[0,1]}(\hat{G}_n - \hat{G}'_n)$ nur Werte in $\{0, 1/n, 2/n, \ldots, 1\}$ annehmen kann, ergibt sich hieraus die allgemeine Ungleichung

$$\mathbb{P}\big(\|\hat{G}_n - \hat{G}'_n\|_{[0,1]} \geq \eta\big) \ \leq \ \exp(1 - n\eta^2) \quad \text{für beliebige } \eta \geq 0. \tag{3.3}$$

Zurück zu $\hat{F}_n$. Für den maximalen Fehler $\|\hat{F}_n - F\| = \|\hat{F}_n - F\|_{\mathbb{R}}$ wissen wir, dass

$$\mathbb{E}\,\Psi(\|\hat{F}_n - F\|) \ \leq \ \mathbb{E}\,\Psi\big(\|\hat{G}_n - \hat{G}'_n\|_{[0,1]}\big)$$

für monoton wachsende und konvexe Funktionen $\Psi : [0,\infty) \to [0,\infty)$. Ferner erfüllt die Zufallsvariable $\|\hat{G}_n - \hat{G}'_n\|_{[0,1]}$ die Exponentialungleichung (3.3). Aus nachfolgendem Lemma 3.2 ergibt sich dann die Ungleichung (3.1).

Lemma 3.2 (Kemperman). *Seien Y, Z nichtnegative Zufallsvariablen mit folgenden Eigenschaften: Für beliebige monoton wachsende und konvexe Funktionen $\Psi : [0,\infty) \to [0,\infty)$ sei*

$$\mathbb{E}\,\Psi(Y) \ \leq \ \mathbb{E}\,\Psi(Z).$$

Ferner sei

$$\mathbb{P}(Z \geq \eta) \ \leq \ D_1 \exp(-D_2\eta^2) \quad \textit{für beliebige } \eta \geq 0$$

mit gewissen Konstanten $D_1 \geq 1$ und $D_2 > 0$. Dann ist

$$\mathbb{P}(Y \geq \eta) \ < \ D_1 e \exp(-D_2\eta^2) \quad \textit{für beliebige } \eta \geq 0.$$

Beweis von Lemma 3.2. Für $\eta = 0$ ist die Behauptung trivial, also sei $\eta > 0$. Für jede Zahl $0 \le \eta_o < \eta$ folgt aus der Markov-Ungleichung, dass

$$\mathbb{P}(Y \ge \eta) \ \le \ \frac{\mathbb{E}((Y - \eta_o)^+)}{\eta - \eta_o}.$$

Doch $y \mapsto (y - \eta_o)^+$ is monoton wachsend und konvex, weshalb

$$\begin{aligned}
\mathbb{P}(Y \ge \eta) \ &\le \ \frac{\mathbb{E}((Z - \eta_o)^+)}{\eta - \eta_o} \\
&= \ \frac{1}{\eta - \eta_o} \int_0^\infty \mathbb{P}((Z - \eta_o)^+ \ge t)\, dt \\
&= \ \frac{1}{\eta - \eta_o} \int_{\eta_o}^\infty \mathbb{P}(Z \ge s)\, ds \\
&\le \ \frac{D_1}{\eta - \eta_o} \int_{\eta_o}^\infty \exp(-D_2 s^2)\, ds \\
&= \ \frac{D_1}{\eta - \eta_o} \int_{\eta_o}^\infty \exp\bigl(-D_2(\eta + (s - \eta))^2\bigr)\, ds \\
&< \ \frac{D_1 \exp(-D_2\eta^2)}{\eta - \eta_o} \int_{\eta_o}^\infty \exp\bigl(-2D_2\eta(s - \eta)\bigr)\, ds \\
&= \ D_1 \exp(-D_2\eta^2)\, \frac{\exp(2D_2\eta(\eta - \eta_o))}{2D_2\eta(\eta - \eta_o)}.
\end{aligned}$$

Da die Funktion $0 < t \mapsto e^t/t$ an der Stelle $t = 1$ ihren minimalen Wert e annimmt, ergibt sich für ein geeignetes η_o die Ungleichung

$$\mathbb{P}(Y \ge \eta) \ < \ D_1 e \exp(-D_2\eta^2),$$

sofern $2D_2\eta^2 \ge 1$. Doch im Falle von $2D_2\eta^2 \le 1$ ist

$$D_1 e \exp(-D_2\eta^2) \ \ge \ D_1 e^{1/2} \ > \ 1.$$

Daher gilt die besagte Ungleichung für alle $\eta \ge 0$. □

3.2 Die zweite Symmetrisierung

Nun betrachten wir den speziellen Prozess $Z_n = \sum_{i=1}^n \phi_{ni}$ aus Abschnitt 1.3. Seien

$$\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn}, \quad \phi'_{n1}, \phi'_{n2}, \ldots, \phi'_{nn}, \quad \xi_1, \xi_2, \ldots, \xi_n$$

stochastisch unabhängig, wobei

$$\mathcal{L}(\phi'_{ni}) \ = \ \mathcal{L}(\phi_{ni}) \quad \text{und} \quad \mathcal{L}(\xi_i) \ = \ \mathcal{U}\{-1, 1\}.$$

Die Zufallsfolge $(\xi_1, \xi_2, \ldots, \xi_n)$ ist eine sogenannte *Rademacher-Folge*.

Die erste Symmetrisierung, angewandt auf $Z = Z_n - \mathbb{E}\, Z_n$ und $Z' = Z_n' - \mathbb{E}\, Z_n' = Z_n' - \mathbb{E}\, Z_n$ mit $Z_n' := \sum_{i=1}^n \phi_{ni}'$, führt zu dem Prozess

$$Z_n - Z_n' \;=\; \sum_{i=1}^n (\phi_{ni} - \phi_{ni}').$$

Da für alle $i \le n$ die Prozesse ϕ_{ni} und ϕ_{ni}' identisch verteilt sind, ist mit

$$(\phi_{ni}'', \phi_{ni}''') \;:=\; \begin{cases} (\phi_{ni}, \phi_{ni}') & \text{falls } \xi_i = 1, \\ (\phi_{ni}', \phi_{ni}) & \text{falls } \xi_i = -1, \end{cases}$$

das Tupel $(\phi_{n1}'', \ldots, \phi_{nn}'', \phi_{n1}''', \ldots, \phi_{nn}''')$ genauso verteilt wie $(\phi_{n1}, \ldots, \phi_{nn}, \phi_{n1}', \ldots, \phi_{nn}')$. Deshalb ist

$$Z_n - Z_n' \;=_{\mathcal{L}}\; \sum_{i=1}^n (\phi_{ni}'' - \phi_{ni}''') \;=\; \sum_{i=1}^n \xi_i(\phi_{ni} - \phi_{ni}') \;=\; Z_n^o - \tilde{Z}_n^o,$$

wobei

$$Z_n^o \;:=\; \sum_{i=1}^n \xi_i \phi_{ni} \quad \text{und} \quad \tilde{Z}_n^o \;:=\; \sum_{i=1}^n \xi_i \phi_{ni}'.$$

Diese Prozesse Z_n^o und $\tilde{Z}_n^o$ sind zwar nicht unabhängig, aber identisch verteilt. Hieraus ergeben sich folgende Ungleichungen.

Lemma 3.3. ***(a)*** *Für beliebige $\eta \ge 0$ ist*

$$\mathbb{P}\big(\|Z_n - Z_n'\| > \eta\big) \;\le\; 2\,\mathbb{P}\big(\|Z_n^o\| > \eta/2\big).$$

(b) *Für beliebige konvexe und monoton wachsende Funktionen $\Psi : [0,\infty) \to [0,\infty)$ ist*

$$\mathbb{E}\,\Psi\big(\|Z_n - Z_n'\|\big) \;\le\; \mathbb{E}\,\Psi\big(2\|Z_n^o\|\big).$$

Beweis von Lemma 3.3.. Da $\|Z_n - Z_n'\| =_{\mathcal{L}} \|Z_n^o - \tilde{Z}_n^o\| \le \|Z_n^o\| + \|\tilde{Z}_n^o\|$, ist

$$\begin{aligned} \mathbb{P}\big(\|Z_n - Z_n'\| \ge \eta\big) &\le \mathbb{P}\big(\|Z_n^o\| \ge \eta/2\big) + \mathbb{P}\big(\|\tilde{Z}_n^o\| \ge \eta/2\big) \\ &= 2\,\mathbb{P}\big(\|Z_n^o\| \ge \eta/2\big). \end{aligned}$$

Ferner ist

$$\begin{aligned} \mathbb{E}\,\Psi\big(\|Z_n - Z_n'\|\big) &\le \mathbb{E}\,\Psi\Big(\frac{1}{2}\big(2\|Z_n^o\| + 2\|\tilde{Z}_n^o\|\big)\Big) \\ &\le \frac{1}{2}\,\mathbb{E}\,\Psi(2\|Z_n^o\|) + \frac{1}{2}\,\mathbb{E}\,\Psi(2\|\tilde{Z}_n^o\|) \\ &= \mathbb{E}\,\Psi(2\|Z_n^o\|). \end{aligned}$$

□

Nach der zweiten Symmetrisierung kann man mit der *bedingten Verteilung von Z_n^o gegeben* $\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn}$ arbeiten. Das heißt, die ϕ_{ni} werden als feste Funktionen auf $\mathcal{T}$ betrachtet, und nur noch die ξ_i sind zufällig.

Zweite Anwendung auf empirische Verteilungsfunktionen

Mit Hilfe beider Symmetrisierungen kann man auch exponentielle Ungleichungen für empirische Verteilungsfunktionen von unabhängigen Zufallsvariablen $X_1, X_2, \ldots, X_n$ mit beliebigen Verteilungsfunktionen $F_1, F_2, \ldots, F_n$ beweisen. Hier hat $\hat{F}_n(x)$ den Erwartungswert

$$\bar{F}_n(x) := \frac{1}{n}\sum_{i=1}^n F_i(x).$$

Wir wissen, dass

$$\mathbb{E}\,\Psi(\|\hat{F}_n - \bar{F}_n\|) \;\le\; \mathbb{E}\,\Psi(2\|\hat{F}_n^o\|)$$

für monoton wachsende und konvexe Funktionen $\Psi : [0,\infty) \to [0,\infty)$. Wie wir gleich zeigen werden, ist

$$\mathbb{P}(2\|\hat{F}_n^o\| \ge \eta) \;\le\; 4\exp(-n\eta^2/8) \tag{3.4}$$

für beliebige $\eta \ge 0$. Zusammen mit Lemma 3.2 ergibt sich dann, dass

$$\mathbb{P}(\|\hat{F}_n - \bar{F}_n\| \ge \;\eta) \;\le\; 4e\exp(-n\eta^2/8) \quad \text{für beliebige } \eta \ge 0.$$

Beweis von (3.4)**.** Es genügt zu zeigen, dass

$$\mathbb{P}(2\|\hat{F}_n^o\| \ge \eta \,|\, X_1, \ldots, X_n) \;\le\; 4\exp(-n\eta^2/8)$$

für beliebige $\eta > 0$. Hier kann man leicht zeigen, dass mit den Ordnungsstatistiken $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ der Beobachtungen $X_1, \ldots, X_n$ gilt:

$$\begin{aligned}\mathbb{P}(2\|\hat{F}_n^o\| \ge \eta \,|\, X_1, \ldots, X_n) \;&=\; \mathbb{P}\Big(\sup_{x\in\mathbb{R}}\Big|\sum_{i=1}^n 1\{X_{(i)} \le x\}\xi_i\Big| \ge n\eta/2 \,\Big|\, X_1, \ldots, X_n\Big) \\ &\le\; \mathbb{P}\Big(\max_{k=0,1,\ldots,n} |W_k| \ge n\eta/2\Big),\end{aligned}$$

wobei $W_0 := 0$ und $W_k := \sum_{i=1}^k \xi_i$ für $1 \le k \le n$. Mit Hilfe des Spiegelungsprinzips kann man zeigen, dass

$$\mathbb{P}\Big(\max_{k=0,1,\ldots,n} |W_k| \ge n\eta/2\Big) \;\le\; 4\,\mathbb{P}(W_n \ge n\eta/2),$$

siehe auch Aufgabe 3.5. Aus der Hoeffding-Ungleichung, die in Kapitel 4 hergeleitet wird, ergibt sich, dass $\mathbb{P}(W_n \ge \gamma) \le \exp(-\gamma^2/(2n))$ für beliebige $\gamma > 0$. Mit $\gamma := n\eta/2$ ergibt sich dann die Behauptung. □

3.3 Entsymmetrisierung

Eine naheliegende Frage ist, ob die Abschätzungen bei den zwei Symmetrisierungen sehr konservativ sind. Mit anderen Worten, haben $\|Z_n - \mathbb{E}\,Z_n\|$ und $\|Z_n^o\|$ ähnliche Eigenschaften? Zumindest im Falle von Prozessen ϕ_{ni} mit identischen Erwartungswerten kann man folgende Aussagen machen.

Lemma 3.4. *Angenommen, die Prozesse $\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn}$ haben ein und denselben Erwartungswert in $\mathbb{R}^{\mathcal{T}}$, also $\mathbb{E}\,\phi_{ni} = \mathbb{E}\,Z_n/n$. Dann gilt für jede konvexe und monoton wachsende Funktion $\Psi : [0,\infty) \to [0,\infty)$ sowie eine beliebige Zahl $0 < \lambda < 1$:*

$$
\begin{aligned}
\mathbb{E}\,\Psi\left(\frac{1}{2}\,\|Z_n - \mathbb{E}\,Z_n\|\right) \;&\leq\; \mathbb{E}\,\Psi\Big(\|Z_n^o\|\Big) \\
&\leq\; \lambda\,\mathbb{E}\,\Psi\left(\frac{2}{\lambda}\,\|Z_n - \mathbb{E}\,Z_n\|\right) + (1-\lambda)\,\mathbb{E}\,\Psi\left(\frac{1}{1-\lambda}\left|\frac{\sum_{i=1}^n \xi_i}{n}\right|\,\|\,\mathbb{E}\,Z_n\|\right).
\end{aligned}
$$

Speziell für $\Psi := \mathrm{id}$ ergibt sich eine einfache Ungleichungskette.

Korollar 3.5.

$$
\begin{aligned}
\frac{1}{2}\,\mathbb{E}\,\|Z_n - \mathbb{E}\,Z_n\| \;\leq\; \mathbb{E}\,\|Z_n^o\| \quad &\leq \quad 2\,\mathbb{E}\,\|Z_n - \mathbb{E}\,Z_n\| + \mathbb{E}\Big|\sum_{i=1}^n \xi_i/n\Big|\,\|\,\mathbb{E}\,Z_n\| \\
&\leq \quad 2\,\mathbb{E}\,\|Z_n - \mathbb{E}\,Z_n\| + \frac{1}{\sqrt{n}}\|\,\mathbb{E}\,Z_n\|.
\end{aligned}
$$

Falls also $\|\,\mathbb{E}\,Z_n\| = o(\sqrt{n})$, dann ist $\mathbb{E}\,\|Z_n - \mathbb{E}\,Z_n\| = o(1)$ genau dann, wenn $\mathbb{E}\,\|Z_n^o\| = o(1)$.

Beweis von Lemma 3.4.. Die erste Ungleichung folgt aus den Symmetrisierungslemmata 3.1 und 3.3 mit $\tilde{\psi}(r) := \psi(r/2)$ an Stelle von $\psi(r)$. Für die zweite Ungleichung seien $I := \{i \leq n : \xi_i = 1\}$ und $J := \{j \leq n : \xi_j = -1\} = \{1, \ldots, n\} \setminus I$. Dann ist

$$
\begin{aligned}
Z_n^o \;&=\; \sum_{i=1}^n \xi_i(\phi_{ni} - \mathbb{E}\,\phi_{n1}) + \sum_{i=1}^n \xi_i\,\mathbb{E}\,\phi_{n1} \\
&=\; \sum_{i\in I}(\phi_{ni} - \mathbb{E}\,\phi_{n1}) - \sum_{j\in J}(\phi_{nj} - \mathbb{E}\,\phi_{n1}) + Y_n\,\mathbb{E}\,Z_n \\
&=\; Z_{(I)} - Z_{(J)} + Y_n\,\mathbb{E}\,Z_n,
\end{aligned}
$$

wobei $Y_n := \sum_{i=1}^n \xi_i/n$ und $Z_{(M)} := \sum_{i\in M}(\phi_{ni} - \mathbb{E}\,\phi_{n1})$. Doch $Z_n - \mathbb{E}\,Z_n = Z_{(I)} + Z_{(J)}$, und $Z_{(I)}, Z_{(J)}$ sind stochastisch unabhängig und zentriert, gegeben I. Man kann also das Symmetrisierungslemma 3.1 auf die bedingte Verteilung von $(Z, Z') = (Z_{(I)}, -Z_{(J)})$ beziehungsweise $(Z, Z') = (Z_{(J)}, -Z_{(I)})$ gegeben I anwenden und erhält

$$
\begin{aligned}
\mathbb{E}\,\Psi(\|Z_n^o\|) \;&\leq\; \frac{\lambda}{2}\,\mathbb{E}\,\Psi\left(\frac{2}{\lambda}\|Z_{(I)}\|\right) + \frac{\lambda}{2}\,\mathbb{E}\,\Psi\left(\frac{2}{\lambda}\|Z_{(J)}\|\right) + (1-\lambda)\,\mathbb{E}\,\Psi\Big(\frac{|Y_n|}{1-\lambda}\|\,\mathbb{E}\,Z_n\|\Big) \\
&\leq\; \frac{\lambda}{2}\,\mathbb{E}\,\Psi\left(\frac{2}{\lambda}\|Z_{(I)} + Z_{(J)}\|\right) + \frac{\lambda}{2}\,\mathbb{E}\,\Psi\left(\frac{2}{\lambda}\|Z_{(J)} + Z_{(I)}\|\right) \\
&\qquad + (1-\lambda)\,\mathbb{E}\,\Psi\left(\frac{|Y_n|}{1-\lambda}\|\,\mathbb{E}\,Z_n\|\right) \\
&=\; \lambda\,\mathbb{E}\,\Psi\left(\frac{2}{\lambda}\|Z_n - \mathbb{E}\,Z_n\|\right) + (1-\lambda)\,\mathbb{E}\,\Psi\left(\frac{|Y_n|}{1-\lambda}\|\,\mathbb{E}\,Z_n\|\right).
\end{aligned}
$$

□

3.4 Mehr Details zum uniformen empirischen Prozess

Wegen der besonderen Bedeutung von empirischen Verteilungsfunktionen stellen wir noch weitere Überlegungen hierzu an, auch wenn diese nichts mit Symmetrisierungen zu tun haben. Wir betrachten stochastisch unabhängige Zufallsvariablen $U_1, U_2, \ldots, U_n$ mit uniformer Verteilung auf $[0,1]$. Ordnet man diese der Größe nach, so erhält man die n Ordnungsstatistiken $U_{(n,1)} < U_{(n,2)} < \cdots < U_{(n,n)}$. Man kann leicht zeigen, dass die Verteilung des Vektors $(U_{(n,i)})_{1\le i\le n}$ die Dichtefunktion

$$u \ \mapsto \ f(u) \ := \ n!\,1\{0 < u_1 < u_2 < \cdots < u_n < 1\}$$

bezüglich des Lebesguemaßes auf $\mathbb{R}^n$ hat. Eine andere, bisweilen sehr nützliche Darstellung ist die folgende: Seien $Y_1, Y_2, Y_3, \ldots$ stochastisch unabhängige, standardexponentialverteilte Zufallsvariablen; das heißt, $\mathbb{P}(Y_i > t) = e^{-t}$ für $t \ge 0$. Dann ist

$$\big(U_{(n,k)}\big)_{k=1}^n \ =_{\mathcal{L}} \ \left(\frac{\sum_{i=1}^k Y_i}{\sum_{j=1}^{n+1} Y_j}\right)_{k=1}^n . \tag{3.5}$$

Insbesondere gilt für den Vektor der “Spacings” $U_{(n,k)} - U_{(n,k-1)}$, $1 \le k \le n+1$:

$$\big(U_{(n,k)} - U_{(n,k-1)}\big)_{k=1}^{n+1} \ =_{\mathcal{L}} \ \left(\frac{Y_k}{\sum_{j=1}^{n+1} Y_j}\right)_{k=1}^{n+1}, \tag{3.6}$$

wobei $U_{(n,0)} := 0$ und $U_{(n,n+1)} := 1$. Hieraus kann man beispielsweise ableiten, dass

$$\mathbb{P}\Big(\min_{k=1,2,\ldots,n+1} \big(U_{(n,k)} - U_{(n,k-1)}\big) > \frac{t}{n^2}\Big) \ \to \ \exp(-t) \quad \text{für alle } t > 0, \tag{3.7}$$

und dass

$$\max_{1\le k\le n+1} \big(U_{(n,k)} - U_{(n,k-1)}\big)\frac{n}{\log n} \ \to_p \ 1. \tag{3.8}$$

Beweis von (3.5).. Es genügt zu zeigen, dass der Vektor $\big(U_{(n,k)} - U_{(n,k-1)}\big)_{k=1}^n$ genauso verteilt ist wie $(Y_k/Y_+)_{k=1}^n$, wobei $Y_+ := \sum_{i=1}^{n+1} Y_i$. Die lineare Abbildung

$$\mathbb{R}^n \ni u \ \mapsto \ (u_1, u_2 - u_1, u_3 - u_2, \ldots, u_n - u_{n-1}) \in \mathbb{R}^n$$

hat Determinante 1. Daher ist der Vektor $\big(U_{(n,k)} - U_{(n,k-1)}\big)_{k=1}^n$ nach der Dichtefunktion

$$x \ \mapsto \ g(x) \ := \ n!\,1\{x \in \Sigma\}$$

auf $(0,1)^n$ verteilt, wobei

$$\Sigma \ := \ \Big\{x \in (0,1)^n : \sum_{i=1}^n x_i < 1\Big\}.$$

Der Vektor $Y := (Y_i)_{i=1}^{n+1}$ ist nach der Dichtefunktion

$$y \ \mapsto \ h(y) \ := \ \exp(-y_+)$$

auf $(0,\infty)^{n+1}$ verteilt. Die Abbildung $T : (0,\infty)^{n+1} \to \Sigma \times (0,\infty)$ mit

$$T(y) \;:=\; \Big(\frac{y_1}{y_+}, \frac{y_2}{y_+}, \ldots, \frac{y_n}{y_+}, \, y_+\Big)$$

ist bijektiv. Ihre Umkehrabbildung ist gegeben durch

$$T^{-1}(x) \;=\; x_{n+1} \cdot \Big(x_1, x_2, \ldots, x_n, \, 1 - \sum_{i=1}^n x_i\Big).$$

Insbesondere sind T und T^{-1} stetig differenzierbar, und

$$DT^{-1}(x) \;=\; \left(\begin{array}{cccc|c} x_{n+1} & 0 & \cdots & 0 & x_1 \\ 0 & x_{n+1} & \ddots & \vdots & x_2 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & x_{n+1} & x_n \\ \hline -x_{n+1} & -x_{n+1} & \cdots & -x_{n+1} & 1 - \sum_{i=1}^n x_i \end{array}\right).$$

Addiert man nun die ersten n Zeilen zur letzten, dann zeigt sich, dass

$$\det(DT^{-1}(x)) \;=\; x_{n+1}^n.$$

Folglich ist der Zufallsvektor $T(Y) \in \Sigma \times (0,\infty)$ nach der Lebesgue-Dichtefunktion

$$x \;\mapsto\; \tilde{g}(x) \;:=\; |\det(DT^{-1}(x))| \, h(T^{-1}(x)) \;=\; x_{n+1}^n \exp(-x_{n+1})$$

verteilt. Doch man kann schreiben

$$\tilde{g}(x) \;=\; g(x_1, x_2, \ldots, x_n) \cdot \frac{x_{n+1}^n \exp(-x_{n+1})}{n!}.$$

Dies zeigt, dass der Zufallsvektor $(Y_k/Y_+)_{k=1}^n$ und die Zufallsvariable Y_+ stochastisch unabhängig sind. Dabei ist $(Y_k/Y_+)_{k=1}^n$ nach der Dichtefunktion g verteilt, und Y_+ ist gammaverteilt mit Parameter $n+1$; siehe auch Aufgabe 3.7. $\square$

3.5 Aufgaben

Aufgabe 3.1. Warum wird eine Verteilung P auf $\mathbb{R}$ durch ihre Verteilungsfunktion F eindeutig charakterisiert?

Aufgabe 3.2. Zeigen Sie, dass eine Verteilungsfunktion F rechtsseitig stetig und die entsprechende Quantilsfunktion F^{-1} linksseitig stetig ist.

Aufgabe 3.3. Zeigen Sie, dass im Falle einer stetigen Verteilungsfunktion F gilt:

$$\begin{aligned} \int_{\mathbb{R}} h(\hat{F}_n(t), F(t)) \, F(dt) \quad &=_{\mathcal{L}} \quad \int_{[0,1]} h(\hat{G}_n(u), u) \, du, \\ \int_{\mathbb{R}} h(\hat{F}_n(t), F(t)) \, \hat{F}_n(dt) \quad &=_{\mathcal{L}} \quad \int_{[0,1]} h(\hat{G}_n(u), u) \, \hat{G}_n(du) \end{aligned}$$

für messbare Funktionen $h : [0,1]^2 \to [0,\infty)$.

Aufgabe 3.4. Berechnen Sie im Falle einer stetigen Verteilungsfunktion F den Grenzwert von $\mathbb{P}\big(\sqrt{n}\sup_{\mathbb{R}}(\hat{F}_n - \hat{F}'_n) \geq \eta\big)$ für festes $\eta \geq 0$.

Aufgabe 3.5. Seien $W_0 := 0$ und $W_k := \sum_{i=1}^k \xi_i$ für $k \in \{1, \ldots, n\}$ mit einer Rademacherfolge $\xi_1, \ldots, \xi_n$. Zeigen Sie mithilfe des Spiegelungsprinzips, dass

$$\mathbb{P}\Big(\max_{0 \leq k \leq n} W_k \geq b\Big) \ = \ \mathbb{P}(W_n \geq b) + \mathbb{P}(W_n < b)$$

für beliebige Zahlen $b \in \mathbb{N}$.

Aufgabe 3.6. Beweisen Sie die Aussagen (3.7) und (3.8) über minimale und maximale Spacings.

Aufgabe 3.7 (Gamma- und Beta-Verteilungen). Die Gamma-Verteilung mit Parameter $a > 0$ ist definiert als die Verteilung mit Dichtefunktion

$$(0, \infty) \ni x \ \mapsto \ \frac{x^{a-1}\exp(-x)}{\Gamma(a)}.$$

Dabei ist $\Gamma(a) := \int_0^\infty x^{a-1}\exp(x)\,dx$.

Die Beta-Verteilung mit Parametern $a, b > 0$ ist definiert als die Verteilung mit Dichtefunktion

$$(0, 1) \ni x \ \mapsto \ \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)}.$$

Dabei ist $B(a, b) := \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$.

(a) Seien X und Y stochastisch unabhängig und gammaverteilt mit Parameter a bzw. b. Zeigen Sie, dass $X/(X+Y)$ und $X+Y$ stochastisch unabhängig sind, wobei $X/(X+Y)$ betaverteilt ist mit Parametern a und b, während $X+Y$ gammaverteilt ist mit Parameter $a+b$. Betrachten Sie hierfür die Transformation

$$(0, \infty) \times (0, \infty) \ni (x, y) \ \mapsto \ T(x, y) \ := \ \Big(\frac{x}{x+y}, x+y\Big) \in (0, 1) \times (0, \infty).$$

Zeigen Sie ferner, dass

$$B(a, b) \ = \ \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{3.9}$$

(b) Berechnen Sie mit Hilfe der Gleichung (3.9) sowie der bekannten Formel $\Gamma(a+1) = a\Gamma(a)$ den Erwartungswert und die Varianz einer betaverteilten Zufallsvariable.

(c) Zeigen Sie, dass die Ordnungsstatistik $U_{(n,i)}$ betaverteilt ist mit Parametern i und $n+1-i$. Bestimmen Sie ihren Erwartungswert und ihre Varianz.

Aufgabe 3.8. Sei F eine stetige Verteilungsfunktion auf $\mathbb{R}$.

(a) Berechnen Sie den Erwartungswert von

$$n\int(\hat{F}_n - F)^2\,dF.$$

(b) Berechnen Sie den Erwartungswert von

$$\sum_{i=1}^{n}\Big(\tilde{F}_n(X_i) - F(X_i)\Big)^2,$$

wobei $\tilde{F}_n := n(n+1)^{-1}\hat{F}_n$.

Hinweis: Betrachten Sie die Ordnungstatistiken $X_{(n,i)} =_{\mathcal{L}} F^{-1}(U_{(n,i)})$. Verwenden Sie nun Aufgabe 3.7 (c).

Kapitel 4

Univariate Exponentialungleichungen

In diesem Kapitel geht es darum, Wahrscheinlichkeiten von Abweichungen einer reellwertigen Zufallsvariable von ihrem Erwartungswert abzuschätzen.

4.1 Die allgemeine Vorgehensweise

Zunächst sei an die Markov-Ungleichung erinnert, wonach $\mathbb{P}(\pm Y \geq \eta) \leq \mathbb{E}(Y^{\pm})/\eta$ für reelle Zufallsvariablen Y und Konstanten $\eta > 0$. Diese Schranke ist oft sehr konservativ. Ersetzt man jedoch $\pm Y$ und η durch $e^{\pm rY}$ und $e^{r\eta}$ mit $r > 0$, also

$$\mathbb{P}(\pm Y \geq \eta) \ = \ \mathbb{P}\big(\exp(\pm rY) \geq \exp(r\eta)\big) \ \leq \ \frac{\mathbb{E}\exp(\pm rY)}{\exp(r\eta)},$$

dann ist die resultierende Schranke oftmals erstaunlich präzise, wenn man $r > 0$ geeignet wählt.

Dies demonstrieren wir zunächst für eine standardnormalverteilte Zufallsvariable Y. Bekanntlich ist

$$\mathbb{P}(\pm Y \geq \eta) \ = \ \frac{\exp(-\eta^2/2)}{\sqrt{2\pi}\eta}(1 + o(1)) \ = \ \exp(-\eta^2/2 + o(\eta^2)) \quad (\eta \to \infty).$$

Andererseits ist

$$\mathbb{P}(\pm Y \geq \eta) \ \leq \ \inf_{r>0} \mathbb{E}\exp(\pm rY)/\exp(r\eta) \ = \ \inf_{r>0} \exp(r^2/2 - r\eta) \ = \ \exp(-\eta^2/2)$$

für beliebige $\eta \geq 0$.

Andere Beispiele werden in den Aufgaben 4.3 und 4.4 behandelt. Dabei hilft das folgende allgemeine Resultat.

Lemma 4.1. *Sei Y eine reellwertige Zufallsvariable. Mit $\kappa(t) := \log \mathbb{E}\exp(tY) \in (-\infty, \infty]$ sei $\kappa < \infty$ in einer Umgebung von 0. Dann gilt für beliebige $y \in \mathbb{R}$,*

$$\begin{aligned} \mathbb{P}(Y \geq y) \ &\leq \ \exp(-K(y)) \quad \textit{für } y \geq \mathbb{E}(Y), \\ \mathbb{P}(Y \leq y) \ &\geq \ \exp(-K(y)) \quad \textit{für } y \leq \mathbb{E}(Y), \end{aligned}$$

wobei

$$K(y) \ := \ \sup_{t \in \mathbb{R}} \, (ty - \kappa(t)).$$

Die Theorie der *Großen Abweichungen (large deviations)* erklärt genauer, weshalb dieses Lemma zu recht präzisen Abschätzungen führt, sobald Y ein Stichprobenmittelwert ist.

Beweis von Lemma 4.1. Wir setzen folgende Tatsachen als bekannt voraus: Ist μ ein Maß auf $\mathbb{R}$ und $g : \mathbb{R} \to \mathbb{R}$ messbar, dann ist die Menge Θ aller $t \in \mathbb{R}$ mit $\int e^{tx}|g(x)| \, \mu(dx) < \infty$ ein Intervall. Auf Θ definiert $\tilde{g}(t) := \int e^{tx} g(x) \, \mu(dx)$ eine stetige Funktion, und auf dem Inneren von Θ ist $\tilde{g}$ unendlich oft differenzierbar mit k-ter Ableitung $\tilde{g}^{(k)}(t) = \int e^{tx} x^k g(x) \, \mu(dx)$. Wendet man dies auf die Funktion κ an, also $\mu = \mathcal{L}(Y)$ und $g \equiv 1$, dann ergeben sich folgende Aussagen, siehe auch Aufgabe 4.2: $\kappa : \mathbb{R} \to (-\infty, \infty]$ ist auf $\Theta = \{\kappa < \infty\}$ stetig und auf dem Inneren von Θ unendlich oft differenzierbar. Für einen inneren Punkt t von Θ gelten die Formeln

$$\kappa'(t) \ = \ \mathbb{E}(Y_t) \quad \text{und} \quad \kappa''(t) \ = \ \mathrm{Var}(Y_t).$$

Dabei ist Y_t eine Zufallsvariable mit Dichtefunktion f_t bezüglich $\mathcal{L}(Y)$, nämlich $f_t(x) = e^{tx - \kappa(t)}$. Insbesondere ist κ eine konvexe Funktion mit $\kappa(0) = 0$, $\kappa'(0) = \mathbb{E}(Y)$ und $\kappa''(0) = \mathrm{Var}(Y)$.

Nun zur eigentlichen Aussage: Für $y \in \mathbb{R}$ ergibt sich aus der Markov-Ungleichung, dass

$$\mathbb{P}(Y \geq y) \ \leq \ \inf_{t \geq 0} \mathbb{E} \exp(tY - ty) \ = \ \inf_{t \geq 0} \exp(\kappa(t) - ty) \ = \ \exp\Bigl(- \sup_{t \geq 0} \, (ty - \kappa(t))\Bigr).$$

Doch $t \mapsto ty - \kappa(t)$ ist eine konkave Funktion mit Ableitung $y - \mathbb{E}(Y)$ an der Stelle $t = 0$. Wenn also $y \geq \mathbb{E}(Y)$, dann ist $ty - \kappa(t) \leq 0 = 0y - \kappa(0)$ für alle $t \leq 0$, so dass $\sup_{t \geq 0}(ty - \kappa(t)) = \sup_{t \in \mathbb{R}}(ty - \kappa(t)) = K(y)$.

Analog ist

$$\mathbb{P}(Y \leq y) \ \leq \ \inf_{t \leq 0} \mathbb{E} \exp(tY - ty) \ = \ \inf_{t \leq 0} \exp(\kappa(t) - ty) \ = \ \exp\Bigl(- \sup_{t \leq 0} \, (ty - \kappa(t))\Bigr),$$

und im Falle von $y \leq \mathbb{E}(Y)$ ist $\sup_{t \leq 0}(ty - \kappa(t)) = \sup_{t \in \mathbb{R}}(ty - \kappa(t)) = K(y)$. $\square$

4.2 Die Ungleichungen von Hoeffding und Bennett

Satz 4.2 (Hoeffding 1963). *Seien $Y_1, Y_2, \ldots, Y_n$ stochastisch unabhängige, reellwertige Zufallsvariablen derart, dass*

$$\mathbb{E}(Y_i) \ = \ 0 \quad \textit{und} \quad a_i \ \leq \ Y_i \ \leq \ b_i$$

mit Konstanten $a_i < b_i$. Dann ist

$$\mathbb{P}\Bigl(\sum_{i=1}^n Y_i \geq \eta\Bigr) \ \leq \ \exp\Bigl(- \frac{2\eta^2}{\sum_{i=1}^n (b_i - a_i)^2}\Bigr) \quad \textit{für alle } \eta > 0.$$

Korollar 4.3. *Für beliebige reelle Zahlen $c_1, c_2, \ldots, c_n$ und $\eta \geq 0$ ist*

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n \xi_i c_i\Big| \geq \eta\Big) \;\leq\; 2\exp\Big(-\frac{\eta^2}{2\sum_{i=1}^n c_i^2}\Big).$$

Dabei ist $\xi_1, \xi_2, \ldots, \xi_n$ eine Rademacherfolge wie in Kapitel 3. □

Das Korollar folgt aus Hoeffdings Ungleichung, angewandt auf $Y_i := \pm\xi_i c_i$ und $(a_i, b_i) := (-c_i, c_i)$. Man beachte, dass $\sum_{i=1}^n c_i^2$ gleich der Varianz von $\sum_{i=1}^n \xi_i c_i$ ist. Hier ist eine allgemeine Exponentialungleichung, welche explizit Varianzen berücksichtigt:

Satz 4.4 (Bennett 1962). *Seien $Y_1, Y_2, \ldots, Y_n$ unabhängige, reellwertige Zufallsvariablen, so dass*

$$\mathbb{E}(Y_i) \;=\; 0 \quad \textit{und} \quad Y_i \;\leq\; M \;<\; \infty \quad \textit{für } 1 \leq i \leq n.$$

Ferner sei $\sum_{i=1}^n \mathrm{Var}(Y_i) \leq V < \infty$. Dann ist

$$\mathbb{P}\Big(\sum_{i=1}^n Y_i \geq \eta\Big) \;\leq\; \exp\Big(-\frac{\eta^2}{2V}\, B\Big(\frac{M\eta}{V}\Big)\Big) \quad \textit{für beliebige } \eta > 0,$$

wobei

$$B(x) \;:=\; \frac{(1+x)\log(1+x) - x}{x^2/2}.$$

Korollar 4.5. *Seien $Y_1, Y_2, \ldots, Y_n$ unabhängige, identisch verteilte Zufallsvariablen mit*

$$\mathbb{E}(Y_i) \;=\; 0, \quad Y_i \;\leq\; M \;<\; \infty \quad \textit{und} \quad \mathrm{Var}(Y_i) \;\leq\; \sigma^2 \quad \textit{für } 1 \leq i \leq n.$$

Dann ist

$$\mathbb{P}\Big(\frac{1}{\sqrt{n\sigma^2}}\sum_{i=1}^n Y_i \geq \eta\Big) \;\leq\; \exp\Big(-\frac{\eta^2}{2}\, B\Big(\frac{\eta M}{\sqrt{n\sigma^2}}\Big)\Big) \quad \textit{für beliebige } \eta > 0.$$

Wie wir gleich sehen werden, ist $\lim_{x\to 0+} B(x) = 1$. Korollar 4.5 zeigt also, dass für die Summe $(n\sigma^2)^{-1/2}\sum_{i=1}^n Y_i$ ähnliche Ungleichungen gelten wie für eine standardnormalverteilte Zufallsgröße.

Für die nachfolgenden Überlegungen und Beweise erinnern wir an eine besonders nützliche Version der Taylor-Entwicklung: Sei I ein offenes Intervall mit $0 \in I$, und sei $f : I \to \mathbb{R}$ eine zweimal stetig differenzierbare Funktion. Dann gilt für beliebige $x \in I$ die Formel

$$\begin{aligned} f(x) \;&=\; f(0) + f'(x)x + \int_0^x (x-s) f''(s)\, ds \\ &=\; f(0) + f'(0)x + x^2 \int_0^1 (1-u) f''(xu)\, du. \end{aligned} \tag{4.1}$$

Anmerkung 4.6. Die Funktion $B(\cdot)$ in Satz 4.4 kann man man für beliebige $x > -1$ definieren, wobei $B(0) = 1$, und es gilt die Darstellung

$$B(x) \;=\; 2\int_0^1 \frac{1-u}{1+xu}\, du.$$

Dies ergibt sich aus (4.1) mit $f(x) := (1+x)\log(1+x) - x$ für $x > -1$ und den Tatsachen, dass $f(0) = 0$, $f'(x) = \log(1+x)$, $f'(0) = 0$ und $f''(x) = 1/(1+x)$. Aus dieser Integraldarstellung ergeben sich folgende Eigenschaften:

- $B(x)$ ist stetig und monoton fallend in $x > -1$.
- $xB(x)$ ist monoton wachsend in $x \geq 0$.

Ausserdem gilt die Entwicklung

$$B(x) = (2 + o(1))\log(x)/x \quad \text{für } x \to \infty.$$

Die obige Integraldarstellung von $B(x)$ kann man auch deuten als

$$B(x) \ = \ \mathbb{E}\Big(\frac{1}{1+xU}\Big)$$

mit einer Zufallsvariable $U \sim \mathrm{Beta}(1,2)$. Da $\mathbb{E}(U) = 1/3$, ergibt sich aus der Konvexität von $u \mapsto 1/(1+xu)$ und der Jensen-Ungleichung, dass

$$B(x) \ \geq \ \frac{1}{1+x/3}.$$

Setzt man dies in Korollar 4.5 ein, so erhält man die *Bernstein-Ungleichung*.

Eine wesentliche Zutat für die Hoeffding-Ungleichung ist folgende Momentenungleichung:

Lemma 4.7 (Hoeffding 1963). *Sei Y eine reellwertige Zufallsvariable mit $\mathbb{E}(Y) = 0$ und $Y \in [a, b]$ für Konstanten $a \leq b$. Dann ist*

$$\log \mathbb{E}\exp(tY) \ \leq \ \frac{t^2(b-a)^2}{8} \quad \textit{für beliebige } t \in \mathbb{R}.$$

Beweis von Lemma 4.7. Ohne Einschränkung sei $a < 0 < b$. Denn $a > 0$ oder $b < 0$ ergäbe einen Widerspruch zu $\mathbb{E}(Y) = 0$, und $a = 0$ oder $b = 0$ würde implizieren, dass $Y = 0$ fast sicher.

Aus der Konvexität der Exponentialfunktion folgt, dass

$$\begin{aligned} \exp(tY) \ &= \ \exp\Big(\frac{b-Y}{b-a}ta + \frac{Y-a}{b-a}tb\Big) \\ &\leq \ \frac{b-Y}{b-a}\exp(ta) + \frac{Y-a}{b-a}\exp(tb), \end{aligned}$$

und wegen $\mathbb{E}(Y) = 0$ ist

$$\begin{aligned} \log \mathbb{E}\exp(tY) \ &\leq \ \log\Big(\frac{b}{b-a}\exp(ta) \ + \ \frac{-a}{b-a}\exp(tb)\Big) \\ &= \ -u\lambda + \log(1 - \lambda + \lambda e^u) \\ &=: \ L(u), \end{aligned}$$

wobei $\lambda := -a/(b-a) \in (0,1)$ und $u := t(b-a)$. Nun ist aber

$$\begin{aligned} L(0) &= 0, \\ L'(u) &= -\lambda + \frac{\lambda}{(1-\lambda)e^{-u}+\lambda} \quad \text{und} \quad L'(0) \ = \ 0, \\ L''(u) &= \frac{\lambda e^u}{1-\lambda+\lambda e^u}\Big(1 - \frac{\lambda e^u}{1-\lambda+\lambda e^u}\Big) \ \le \ 1/4 \end{aligned}$$

wegen der bekannten Ungleichung $x(1-x) \le 1/4$ für beliebige $x \in \mathbb{R}$. Folglich ist nach der Taylorschen Formel

$$L(u) \ = \ L(0) + uL'(0) + u^2 L''(u^*)/2 \ = \ u^2 L''(u^*)/2 \ \le \ u^2/8 = t^2(b-a)^2/8$$

mit einer geeigneten Zwischenstelle $u^* \in \mathbb{R}$. □

Beweis der Hoeffding-Ungleichung. Nach der Markov-Ungleichung ist

$$\mathbb{P}\Big(\pm\sum_{i=1}^n Y_i \ge \eta\Big) \ \le \ \exp(-t\eta)\,\mathbb{E}\exp\Big(\pm t\sum_{i=1}^n Y_i\Big) \ = \ \exp(-t\eta)\prod_{i=1}^n \mathbb{E}\exp(\pm tY_i)$$

für beliebige $t > 0$. Nach Lemma 4.7 ist die rechte Seite nicht größer als

$$\exp(-t\eta)\prod_{i=1}^n \exp\big(t^2(b_i-a_i)^2/8\big) \ = \ \exp\Big(-t\eta + t^2\sum_{i=1}^n (b_i-a_i)^2/8\Big).$$

Für $t = 4\eta \Big/ \sum_{i=1}^n (b_i-a_i)^2$ ergibt sich dann die gewünschte Abschätzung. □

Beweis der Bennett-Ungleichung. Indem man $Y_1, \ldots, Y_n$, V und η durch $Y_1/M, \ldots, Y_n/M$, V/M^2 und η/M ersetzt, kann man die Behauptung auf den Fall $M = 1$ zurückführen. Sei also $Y_i \le 1$ für $1 \le i \le n$. Wie wir später zeigen werden, ist

$$\mathbb{E}\exp\Big(t\sum_{i=1}^n Y_i\Big) \ \le \ \exp\big(V(e^t-1-t)\big) \quad \text{für alle } t > 0. \tag{4.2}$$

Nun folgt aus der Markov-Ungleichung, dass

$$\begin{aligned} \mathbb{P}\Big(\sum_{i=1}^n Y_i \ge \eta\Big) &\le \inf_{t>0} \exp\big(V(e^t-1-t) - t\eta\big) \\ &= \exp\big(V(e^{t_o}-1-t_o) - t_o\eta\big) \qquad \big(\text{mit } t_o = \log(1+\eta/V)\big) \\ &= \exp\Big(-\frac{\eta^2}{2V}B\Big(\frac{\eta}{V}\Big)\Big). \end{aligned}$$

Beweis von (4.2). Sei $h(x) := e^x - 1 - x$. Da $\mathbb{E}\exp\big(\sum_{i=1}^n tY_i\big) = \prod_{i=1}^n \mathbb{E}\exp(tY_i)$, genügt es zu zeigen, dass für beliebige $t > 0$ und Zufallsvariablen Y mit $\mathbb{E}(Y) = 0$, $Y \le 1$ und $\mathbb{E}(Y^2) \le \sigma^2 < \infty$ gilt:

$$\mathbb{E}\exp(tY) \ \le \ \exp(\sigma^2 h(t)). \tag{4.3}$$

Aus $\mathbb{E}(Y) = 0$ folgt, dass

$$\mathbb{E}\exp(tY) \; = \; 1 + \mathbb{E}\, h(tY).$$

Nun ist $h(0) = 0$, $h'(x) = e^x - 1$, $h'(0) = 0$ und $h''(x) = e^x$. Daher ergibt sich aus (4.1) die Darstellung

$$h(x) \; = \; x^2 g(x) \quad \text{mit} \quad g(x) \; := \; \int_0^1 (1-u)e^{xu}\, du.$$

Offensichtlich ist g monoton wachsend, so dass

$$\mathbb{E}\, h(tY) \; = \; \mathbb{E}(t^2Y^2 g(tY)) \; \le \; \mathbb{E}(t^2 Y^2 g(t)) \; = \; \mathbb{E}(Y^2)h(t) \; \le \; \sigma^2 h(t).$$

Nun folgt die behauptete Ungleichung (4.3) aus der Tatsache, dass $1 + z \le \exp(z)$ für $z \in \mathbb{R}$. $\square$

4.3 Massenkonzentration

Die vorangehenden Exponentialungleichungen kann man unter anderem dazu benutzen, die Abweichungen eines Stichprobenmittelwertes von seinem Erwartungswert abzuschätzen. Inzwischen gibt es jedoch eine umfangreiche Literatur über das allgemeinere Phänomen der *Massenkonzentration* (concentration of measure). Hier ist eines der ersten Resultate in diese Richtung:

Satz 4.8 (Azuma–McDiarmid). *Seien $X_1, X_2, \ldots, X_n$ stochastisch unabhängige Zufallsvariablen mit Werten in $(\mathcal{X}, \mathcal{B})$, und sei $H : \mathcal{X}^n \to \mathbb{R}$ eine messbare Funktion derart, dass*

$$\big|H(x) - H(y)\big| \; \le \; \sum_{i=1}^n 1\{x_i \ne y_i\} c_i \quad \text{für beliebige } x, y \in \mathcal{X}^n \tag{4.4}$$

mit gewissen Konstanten $c_1, c_2, \ldots, c_n > 0$. Dann erfüllt $Y := H(X_1, \ldots, X_n)$ die Ungleichungen

$$\mathbb{P}\big(\pm(Y - \mathbb{E}\, Y) \ge \eta\big) \; \le \; 2\exp\Big(-\frac{\eta^2}{2\sum_{i=1}^n c_i^2}\Big)$$

für beliebige $\eta \ge 0$.

Anmerkung 4.9. Hier ist eine von vielen möglichen Anwendungen des Satzes 4.8: Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien stochastisch unabhängig mit Verteilung P auf $(\mathcal{X}, \mathcal{B})$. Für beliebige Familien $\mathcal{F}$ von messbaren Funktionen $f : \mathcal{X} \to [0,1]$ gilt dann:

$$\mathbb{P}\Big(\big|\|\hat{P}_n - P\|_{\mathcal{F}} - \mathbb{E}\,\|\hat{P}_n - P\|_{\mathcal{F}}\big| \ge \eta\Big) \; \le \; 2\exp(-n\eta^2/2) \quad \text{für beliebige } \eta \ge 0,$$

sofern $\|\hat{P}_n - P\|_{\mathcal{F}}$ messbar ist. Die Standardabweichung von $\|\hat{P}_n - P\|_{\mathcal{F}}$ ist also stets von der Größenordnung $O(n^{-1/2})$, selbst wenn $\mathbb{E}\,\|\hat{P}_n - P\|_{\mathcal{F}}$ langsamer oder gar nicht gegen Null konvergiert. Dies folgt aus Satz 4.8, angewandt auf $H(X_1, X_2, \ldots, X_n) := \|\hat{P}_n - P\|_{\mathcal{F}}$, denn diese Funktion erfüllt Ungleichung (4.4) mit $c_i := 1/n$.

Beweis von Satz 4.8. Für ganze Zahlen a, b sei $X_{a:b} := (X_i)_{i=a}^b$, falls $1 \le a \le b \le n$, und ansonsten $X_{a:b} := ()$. Dann definieren wir die Zufallsvariablen

$$Y_k \ := \ \mathbb{E}\big(Y \,\big|\, X_{1:k}\big) \quad \text{für } 0 \le k \le n,$$

also $Y_n = Y = H(X_{1:n})$, $Y_0 = \mathbb{E}\, H(X_{1:n})$ und

$$Y_k \ = \ \int H(X_{1:k}, z)\, P_{k+1:n}(dz) \quad \text{für } 0 \le k < n,$$

wobei $P_{a:b} := X_{a:b}$. Für $1 \le k \le n$ gilt:

$$\mathbb{E}(Y_k \,|\, X_{1:k-1}) \ = \ Y_{k-1} \quad \text{und} \quad |Y_k - Y_{k-1}| \ \le \ c_k.$$

Denn mit $P_k := P_{k:k}$ gilt:

$$\begin{aligned} Y_n \ &= \ H(X_{1:n-1}, X_n), \\ Y_{n-1} \ &= \ \int\!\!\int H(X_{1:n-1}, y)\, P_n(dy), \end{aligned}$$

also

$$\begin{aligned} Y_n - Y_{n-1} \ &= \ \int\!\!\int \big(H(X_{1:n-1}, X_n) - H(X_{1:n-1}, y)\big)\, P_n(dy) \\ &\in \ [-c_n, c_n], \end{aligned}$$

und für $1 \le k < n$ ist

$$\begin{aligned} Y_k \ &= \ \int H(X_{1:k-1}, X_k, z)\, P_{k+1:n}(dz), \\ Y_{k-1} \ &= \ \int\!\!\int H(X_{1:k-1}, y, z)\, P_{k+1:n}(dz)\, P_k(dy), \end{aligned}$$

also

$$\begin{aligned} Y_k - Y_{k-1} \ &= \ \int\!\!\int \big(H(X_{1:k-1}, X_k, z) - H(X_{1:k-1}, y, z)\big)\, P_{k+1:n}(dz)\, P_k(dy) \\ &\in \ [-c_k, c_k]. \end{aligned}$$

Hieraus folgt, dass

$$\begin{aligned} \mathbb{E}(e^{t(Y_k - Y_0)}) &= \ \mathbb{E}\big(e^{t(Y_{k-1} - Y_0)} e^{t(Y_k - Y_{k-1})}\big) \\ &= \ \mathbb{E}\,\mathbb{E}\big(e^{t(Y_{k-1} - Y_0)} e^{t(Y_k - Y_{k-1})} \,\big|\, X_{1:k-1}\big) \\ &= \ \mathbb{E}\big(e^{t(Y_{k-1} - Y_0)}\, \mathbb{E}\big(e^{t(Y_k - Y_{k-1})} \,\big|\, X_{1:k-1}\big)\big) \\ &\le \ \mathbb{E}(e^{t(Y_{k-1} - Y_0)}) \cdot e^{t^2 c_k^2/2} \end{aligned}$$

gemäß Lemma 4.7, angewandt auf die bedingte Verteilung von $Y_k - Y_{k-1}$, gegeben $X_{1:k-1}$. Wendet man dies induktiv für $k = n, n-1, n-2, \ldots, 2, 1$ an, so ergibt sich die Ungleichung

$$\mathbb{E}(e^{t(Y_n - Y_0)}) \ \le \ \exp\Big(t^2 \sum_{k=1}^n c_k^2/2\Big),$$

und man kann wie im Beweis von Satz 4.2 argumentieren. $\square$

4.4 Aufgaben

Aufgabe 4.1. Sei Y standardnormalverteilt. Zeigen Sie, dass

$$\mathbb{P}(Y \geq \eta) \ \leq \ \exp(-\eta^2/2)/(2+\eta)$$

für beliebige $\eta \geq 0$.

Aufgabe 4.2. Sei Y eine reellwertige Zufallsvariable mit $\mathbb{E}\exp(tY) < \infty$ für alle t in einer Nullumgebung.

(a) Sei $\kappa(t) := \log \mathbb{E}\exp(tY)$. Zeigen Sie, dass κ auf dem Intervall $\{\kappa < \infty\}$ stetig und auf seinem Inneren unendlich oft differenzierbar ist. Zeigen Sie, dass für einen inneren Punkt t von $\{\kappa < \infty\}$ gilt:

$$\kappa'(t) \ = \ \mathbb{E}(Y_t) \quad \text{und} \quad \kappa''(t) \ = \ \mathrm{Var}(Y_t).$$

Dabei ist Y_t eine Zufallsvariable mit Dichtefunktion f_t bezüglich $\mathcal{L}(Y)$, nämlich $f_t(x) = e^{tx-\kappa(t)}$.

(b) Sei $\mathbb{E}(Y) = 0$, und seien $Y_1, Y_2, \ldots, Y_n$ unabhängige Kopien von Y mit arithmetischem Mittelwert $\bar{Y}_n$. Zeigen Sie, dass

$$\frac{\log \mathbb{P}(\pm\bar{Y}_n \geq \epsilon)}{n} \ \leq \ -K(\pm\epsilon) \quad \text{für alle } \epsilon > 0,$$

wobei

$$K(y) \ := \ \sup_{t\in\mathbb{R}}(ty - \kappa(t)).$$

Zeigen Sie ferner, dass

$$K(y) \ = \ \frac{y^2}{2\mathrm{Var}(Y)} + O(y^3) \quad \text{für } y \to 0,$$

sofern $\mathrm{Var}(Y) > 0$.

Aufgabe 4.3 (Chernoff-Ungleichungen).

(a) Zeigen Sie, dass für $p \in (0,1)$ und $q \in [0,1]$ folgende Ungleichung gilt:

$$\Psi(p,q) \ := \ q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p} \ \geq \ 2(p-q)^2.$$

(b) Sei $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ mit stochastisch unabhängigen Zufallsvariablen $X_i \in [0,1]$, wobei $p := \mathbb{E}(\bar{X}) \in (0,1)$. Zeigen Sie, dass

$$\left.\begin{array}{l} \mathbb{P}(\bar{X} \leq q) \\ \mathbb{P}(\bar{X} \geq q) \end{array}\right\} \ \leq \ \exp\bigl(-n\Psi(p,q)\bigr) \quad \text{falls} \ \left\{\begin{array}{l} q \leq p, \\ q \geq p. \end{array}\right.$$

Aufgabe 4.4. Berechnen Sie Exponentialungleichungen für $\mathbb{P}(Y \geq y)$, wenn $y \geq \mathbb{E}(Y)$, und $\mathbb{P}(Y \leq y)$, wenn $y \leq \mathbb{E}(Y)$, im Falle einer Gamma- beziehungsweise Poisson-verteilten Zufallsvariable Y.

Aufgabe 4.5. Berechnen Sie Exponentialungleichungen für $\mathbb{P}(Y \geq y)$, wenn $y \geq \mathbb{E}(Y)$, und $\mathbb{P}(Y \leq y)$, wenn $y \leq \mathbb{E}(Y)$, im Falle einer Beta-verteilten Zufallsvariable Y.

Hinweis: Stellen Sie $Y \sim \mathrm{Beta}(a, b)$ dar als $V/(V + W)$ mit stochastisch unabängigen Zufallsvariablen $V \sim \mathrm{Gamma}(a)$ und $W \sim \mathrm{Gamma}(b)$. Schreiben Sie dann Ungleichungen der Form $Y \geq r$ oder $Y \leq r$ als $(1 - r)V - rW \geq 0$ beziehungsweise $(1 - r)V - rW \leq 0$.

Aufgabe 4.6 (Eine weitere Ungleichung von Hoeffding). Wie in Aufgabe 4.3 sei $\bar{X}$ der Mittelwert von stochastisch unabhängigen Zufallsvariablen $X_1, X_2, \ldots, X_n \in [0, 1]$, und sei $p := \mathbb{E}(\bar{X}) \in (0, 1)$. Ferner sei $Y \sim \mathrm{Bin}(n, p)$. Zeigen Sie, dass

$$\mathbb{E}\,\psi(\bar{X}) \;\leq\; \mathbb{E}\,\psi(Y/n)$$

für beliebige konvexe Funktionen $\psi : \mathbb{R} \to \mathbb{R}$.

Vorschlag: Zeigen Sie zunächst durch Bedingen auf alle bis auf einen Summanden X_i, dass es genügt, den Fall von Bernoulli-Variablen $X_i \sim \mathrm{Bin}(1, p_i)$ zu betrachten. Zeigen Sie dann, dass $\mathbb{E}\,\psi(\bar{X})$ zunimmt, wenn man zwei beliebige Wahrscheinlichkeiten p_i und p_j durch ihren Mittelwert $(p_i + p_j)/2$ ersetzt.

Kapitel 5

Mengenindizierte empirische Prozesse

Nun untersuchen wir $(\hat{P}_n(D))_{D\in\mathcal{D}}$ mit einer Familie $\mathcal{D}$ von messbaren Teilmengen von $\mathcal{X}$. Wir setzen generell voraus, dass $\mathcal{D}$ *punktweise separabel* im folgenden Sinne ist. Es gibt eine abzählbare Teilmenge $\mathcal{D}_o$ von $\mathcal{D}$, so dass für alle $D \in \mathcal{D}$ eine Folge $(D_k)_k$ in $\mathcal{D}_o$ existiert mit

$$\lim_{k\to\infty} 1_{D_k}(x) \;=\; 1_D(x) \quad \text{für alle } x \in \mathcal{X}.$$

Unter dieser Bedingung ist

$$\|\nu\|_{\mathcal{D}} \;=\; \|\nu\|_{\mathcal{D}_o}$$

für jedes endliche signierte Maß ν auf $\mathcal{X}$. Schreibt man nämlich $\nu = \nu^+ - \nu^-$ mit endlichen Maßen ν^+, ν^-, dann folgt aus dem Satz von der majorisierten Konvergenz, dass

$$|\nu(D) - \nu(D_k)| \;\le\; (\nu^+ + \nu^-)\big(|1_D - 1_{D_k}|\big) \;\to\; 0 \quad (k \to \infty).$$

Daher kann man die Resultate aus Kapitel 3, welche eine abzählbare Indexmenge voraussetzen, anwenden.

5.1 Glivenko-Cantelli-Klassen

Das erste Ziel ist, notwendige und hinreichende Bedingungen zu finden, unter welchen $\|\hat{P}_n - P\|_{\mathcal{D}} \to_p 0$. In letzterem Falle nennt man $\mathcal{D}$ eine *Glivenko-Cantelli-Klasse für P*. Zunächst halten wir fest, dass folgende drei Aussagen äquivalent sind:

(5.1) $$\|\hat{P}_n - P\|_{\mathcal{D}} \;\to_p\; 0;$$

(5.2) $$\mathbb{E}\,\|\hat{P}_n - P\|_{\mathcal{D}} \;\to\; 0;$$

(5.3) $$\mathbb{E}\,\|\hat{P}_n^o\|_{\mathcal{D}} \;\to\; 0.$$

Dabei ist $\hat{P}_n^o := n^{-1}\sum_{i=1}^n \xi_i\delta_{X_i}$ mit einer von $(X_i)_i$ unabhängigen Rademacherfolge $(\xi_i)_i$. Die Äquivalenz von (5.1) und (5.2) folgt aus der Tatsache, dass $\|\hat{P}_n - P\|_{\mathcal{D}}$ stets kleiner oder gleich

Eins ist. Denn für beliebige Zufallsvariablen Y mit Werten in $[0, 1]$ und $0 < \epsilon \leq 1$ gelten die Ungleichungen

$$\mathbb{P}(Y \geq \epsilon) \;\leq\; \mathbb{E}(Y)/\epsilon \quad \text{und} \quad \mathbb{E}(Y) \;\leq\; \epsilon + (1-\epsilon)\,\mathbb{P}(Y \geq \epsilon).$$

Die Äquivalenz von (5.2) und (5.3) ergibt sich aus den Symmetrisierungsungleichungen in Korollar 3.5, wonach

$$\mathbb{E}\,\|\hat{P}_n - P\|_{\mathcal{D}}/2 \;\leq\; \mathbb{E}\,\|\hat{P}_n^o\|_{\mathcal{D}} \;\leq\; 2\,\mathbb{E}\,\|\hat{P}_n - P\|_{\mathcal{D}} + n^{-1/2}.$$

Zunächst beweisen wir eine hinreichende Bedingung, die auf der folgenden Zufallsgröße beruht:

$$\hat{\Delta}_n \;:=\; \#\{D \cap \hat{\mathcal{X}}_n : D \in \mathcal{D}\},$$

wobei

$$\hat{\mathcal{X}}_n \;:=\; \{X_1, X_2, \ldots, X_n\}.$$

Diese Zahl $\hat{\Delta}_n$ ist stets kleiner oder gleich 2^n.

Satz 5.1. *Konvergiert $\log(\hat{\Delta}_n)/n$ in Wahrscheinlichkeit gegen Null, so ist die Mengenfamilie $\mathcal{D}$ eine Glivenko-Cantelli-Klasse für P.*

Beweis. Es genügt zu zeigen, dass (5.3) erfüllt ist, sofern $\log(\hat{\Delta}_n)/n \to_p 0$. Wegen $\|\hat{P}_n^o\|_{\mathcal{D}} \leq 1$ ist nur zu zeigen, dass

$$\mathbb{P}\big(\|\hat{P}_n^o\|_{\mathcal{D}} \geq \epsilon\big) \;\to\; 0$$

für eine beliebige feste Zahl $\epsilon > 0$. Wegen

$$\mathbb{P}\big(\|\hat{P}_n^o\|_{\mathcal{D}} \geq \epsilon\big) \;=\; \mathbb{E}\,\mathbb{P}\big(\|\hat{P}_n^o\|_{\mathcal{D}} \geq \epsilon \,\big|\, X_1, \ldots, X_n\big)$$

genügt es zu zeigen, dass

$$\mathbb{P}\big(\|\hat{P}_n^o\|_{\mathcal{D}} \geq \epsilon \,\big|\, X_1, \ldots, X_n\big) \;\to_p\; 0.$$

Nun ist

$$\|\hat{P}_n^o\|_{\mathcal{D}} \;=\; \max_{a \in \hat{A}_n} \Big| \frac{1}{n} \sum_{i=1}^n \xi_i a_i \Big|$$

mit der zufälligen Menge $\hat{A}_n := \big\{(1_D(X_i))_{i=1}^n : D \in \mathcal{D}\big\}$, die aus $\hat{\Delta}_n$ Vektoren in $\{0,1\}^n$ besteht. Ferner folgt aus Korollar 4.3, dass

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n \xi_i a_i\Big| \geq \epsilon \,\Big|\, X_1, \ldots, X_n\Big) \;\leq\; 2\exp\Big(-\frac{n\epsilon^2}{2}\Big) \quad \text{für alle } a \in \{0,1\}^n.$$

Folglich ist

$$\begin{aligned} \mathbb{P}\big(\|\hat{P}_n^o\|_{\mathcal{D}} \geq \epsilon \,\big|\, X_1, \ldots, X_n\big) &\leq \sum_{a \in \hat{A}_n} \mathbb{P}\Big(\Big| n^{-1} \sum_{i=1}^n \xi_i a_i \Big| \geq \epsilon \,\Big|\, X_1, \ldots, X_n\Big) \\ &\leq 2\hat{\Delta}_n \exp\Big(-\frac{n\epsilon^2}{2}\Big) \\ &= 2\exp\Big(-n\Big(\frac{\epsilon^2}{2} - \frac{\log \hat{\Delta}_n}{n}\Big)\Big) \\ &\to_p 0. \end{aligned}$$

□

Beispiel 5.2 (Rechtecke im $\mathbb{R}^d$). Sei $\mathcal{D}$ die Menge aller achsenparallelen Rechtecke im $\mathbb{R}^d$. Dann ist

$$\hat{\Delta}_n \leq 1 + \Big(\frac{n(n+1)}{2}\Big)^d.$$

Denn jede nichtleere Menge $D \cap \hat{\mathcal{X}}_n$ mit $D \in \mathcal{D}$ lässt sich schreiben als

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d] \cap \hat{\mathcal{X}}_n$$

mit Zahlen $a_j \leq b_j$ aus der Menge $\{X_1(j), \ldots, X_n(j)\}$. Dabei bezeichnet $X_i(j)$ die j-te Komponente von X_i. Nach Satz 5.1 ist also $\mathcal{D}$ eine Glivenko-Cantelli-Klasse für beliebige Wahrscheinlichkeitsmaße P.

Beispiel 5.3 (Halbräume im $\mathbb{R}^2$). Sei $\mathcal{D}$ die Menge aller abgeschlossenen Halbräume im $\mathbb{R}^2$, also aller Mengen der Form $\{x \in \mathbb{R}^2 : u^\top x \leq r\}$ mit einem Vektor $u \in \mathbb{R}^2 \setminus \{0\}$ und einer reellen Zahl r. Auch diese Familie ist eine Glivenko-Cantelli-Klasse für beliebige Verteilungen P. Denn im Falle von $n \geq 2$ ist stets

$$\hat{\Delta}_n \leq 2n(n-1).$$

Um dies nachzuweisen, betrachten wir einen beliebigen Halbraum $D = \{x : u^\top x \leq r\}$. Die Zahl r können wir so wählen, dass $u^\top X_i = r$ für mindestens einen Index i. Falls dies für genau einen Index i der Fall ist, können wir den Richtungsvektor u noch gegen den Uhrzeigersinn drehen, bis erstmalig $u^\top X_j = r$ für einen weiteren Index $j \neq i$. Die ursprüngliche Menge $D \cap \hat{\mathcal{X}}_n$ entspricht jetzt der Menge $\tilde{D} \cap \hat{\mathcal{X}}_n$ mit einem nicht notwendig abgeschlossenen Halbraum $\tilde{D}$ der folgenden Form: Sein Rand wird von zwei verschiedenen Punkten X_k, X_ℓ aus $\hat{\mathcal{X}}_n$ aufgespannt, und $(\partial\tilde{D}) \cap \tilde{D}$ ist von der Form

$$\{(1-t)X_k + tX_\ell : t \leq 0\} \quad \text{oder} \quad \{(1-t)X_k + tX_\ell : t \in \mathbb{R}\}.$$

Ferner liegt das Innere von $\tilde{D}$ auf der rechten Seite von der Verbindungsgerade von X_k und X_ℓ, wenn man in Richtung $X_\ell - X_k$ schaut. Nun gibt es maximal $n(n-1)$ Möglichkeiten ein Paar (X_k, X_ℓ) von zwei verschiedenen Datenpunkten auszuwählen, und dann muss man sich noch für eine von zwei Versionen von $(\partial\tilde{D}) \cap \tilde{D}$ entscheiden. Also ist $\hat{\Delta}_n \leq 2n(n-1)$.

Die Bedingung an $\hat{\Delta}_n$ in Satz 5.1 ist nicht nur hinreichend sondern auch notwendig für (5.1–5.3). Um dies zu beweisen, führen wir noch eine weitere Zufallsgröße ein, nämlich

$$\hat{V}_n \ := \ \max\bigl\{\#S : S \subset \hat{\mathcal{X}}_n, \#\{D \cap S : D \in \mathcal{D}\} = 2^{\#S}\bigr\}.$$

Dann kann man Satz 5.1 wie folgt abrunden.

Satz 5.4 (Vapnik–Červonenkis, Steele). *Die Mengenfamilie $\mathcal{D}$ ist eine Glivenko-Cantelli-Klasse für P genau dann, wenn eine der beiden folgenden Aussagen zutrifft, die ihrerseits äquivalent sind:*

$$\frac{\log \hat{\Delta}_n}{n} \ \to_p \ 0; \tag{5.4}$$

$$\frac{\hat{V}_n}{n} \ \to_p \ 0. \tag{5.5}$$

Beweis von Satz 5.4. Aus Satz 5.1 wissen wir bereits, dass $\mathcal{D}$ eine Glivenko-Cantelli-Klasse für P ist, sofern (5.4) gilt. Im nachfolgenden Abschnitt werden wir zeigen, dass

$$\log(2)\,\frac{\hat{V}_n}{n} \ \le \ \frac{\log \hat{\Delta}_n}{n} \ \le \ \frac{\hat{V}_n}{n}\log\Bigl(e\,\frac{n}{\hat{V}_n}\Bigr);$$

siehe Korollar 5.9. Diese Ungleichungen implizieren die Äquivalenz von (5.4) und (5.5). Zu zeigen bleibt also nur noch, dass (5.3) auch Aussage (5.5) impliziert. Dazu sei $\hat{I}_n$ eine zufällige Teilmenge von $\{1, \ldots, n\}$ mit

$$\#\hat{I}_n \ = \ \hat{V}_n \quad \text{und} \quad \#\bigl\{\{X_i : i \in \hat{I}_n\} \cap D : D \in \mathcal{D}\bigr\} \ = \ 2^{\hat{V}_n}.$$

Insbesondere sind alle Punkte X_i mit $i \in \hat{I}_n$ paarweise verschieden. Ferner definieren wir

$$Z \ := \ \frac{1}{n}\sum_{i \in \hat{I}_n} \xi_i \delta_{X_i} \quad \text{und} \quad Z' \ := \ -\frac{1}{n}\sum_{i \in \{1,\ldots,n\}\setminus \hat{I}_n} \xi_i \delta_{X_i}.$$

Nach Konstruktion der Menge $\hat{I}_n$ ist

$$\|Z\|_{\mathcal{D}} \ = \ \max_{M \subset \hat{I}_n} \Bigl|\frac{1}{n}\sum_{i \in M} \xi_i\Bigr| \ \ge \ \frac{\hat{V}_n}{2n}.$$

Ferner sind Z und Z' stochastisch unabhängig und zentriert, gegeben $X_1, \ldots, X_n$. Zusammen mit Symmetrisierungslemma 3.1 (b) folgt also, dass

$$\begin{aligned}
\mathbb{E}\Bigl(\frac{\hat{V}_n}{2n}\Bigr) \ &\le \ \mathbb{E}\,\|Z\|_{\mathcal{D}} \\
&= \ \mathbb{E}\,\mathbb{E}\bigl(\|Z\|_{\mathcal{D}} \bigm| X_1, \ldots, X_n\bigr) \\
&\le \ \mathbb{E}\,\mathbb{E}\bigl(\|Z - Z'\|_{\mathcal{D}} \bigm| X_1, \ldots, X_n\bigr) \\
&= \ \mathbb{E}\,\|\hat{P}_n^o\|.
\end{aligned}$$

□

5.2 Vapnik-Červonenkis-Theorie

In diesem Abschnitt beschäftigen wir uns systematisch mit den Zahlen $\hat{\Delta}_n$ und $\hat{V}_n$. Dazu betrachten wir eine beliebige endliche Teilmenge E von $\mathcal{X}$ und definieren

$$\Delta(E) = \Delta(E, \mathcal{D}) \ := \ \#(\mathcal{D} \cap E),$$

wobei $\mathcal{D} \cap E := \{D \cap E : D \in \mathcal{D}\}$. Die Familie $\mathcal{D}$ *zerlegt* E *vollständig*, falls $\mathcal{D} \cap E = \{A : A \subset E\}$, was äquivalent ist zu $\Delta(E) = 2^{\#E}$. Ferner sei

$$\begin{aligned} V(E) = V(E, \mathcal{D}) \ &:= \ \max\bigl\{\#B : B \subset E, B \text{ wird von } \mathcal{D} \text{ vollständig zerlegt}\bigr\} \\ &= \ \max\bigl\{\#B : B \subset E, \Delta(B) = 2^{\#B}\bigr\}. \end{aligned}$$

Beispiel 5.5. Sei $\mathcal{X} = \mathbb{R}$ und $\mathcal{D} = \bigl\{(-\infty, t] : t \in \mathbb{R}\bigr\}$. Für jede nichtleere endliche Menge $E \subset \mathbb{R}$ ist

$$\Delta(E) \ = \ \#E + 1 \quad \text{und} \quad V(E) \ = \ 1.$$

Beispiel 5.6. Sei $\mathcal{X} = \mathbb{R}^2$, und $\mathcal{D}$ sei die Menge aller abgeschlossenen Halbräume im $\mathbb{R}^2$; siehe auch Beispiel 5.3. Hier kann man sich schnell an Hand von Skizzen davon überzeugen, dass

$$V(E) \ \le \ 3.$$

Einen formalen Beweis werden wir später liefern.

Beispiel 5.7. Sei $\mathcal{X} = \mathbb{R}^d$ und $\mathcal{D} = \bigl\{\text{konvexe Mengen } D \subset \mathbb{R}^d\bigr\}$. Ist speziell E eine Teilmenge der Euklidischen Einheitssphäre $\{x \in \mathbb{R}^d : x^\top x = 1\}$, dann ist

$$\Delta(E) \ = \ 2^{\#E} \quad \text{und somit} \quad V(E) \ = \ \#E.$$

Denn für $A \subset E$ ist $A = \mathrm{conv}(A) \cap E$, wobei $\mathrm{conv}(A)$ die konvexe Hülle von A bezeichnet.

Der nachfolgende Satz und sein Korollar zeigen, dass die Funktionen $\Delta(\cdot)$ und $V(\cdot)$ in engem Zusammenhang stehen.

Satz 5.8 (Vapnik–Červonenkis). *Für jede endliche Teilmenge E von $\mathcal{X}$ ist*

$$\Delta(E) \ \le \ \sum_{i=0}^{V(E)} \binom{\#E}{i}.$$

Korollar 5.9. *Für endliche Mengen $E \subset \mathcal{X}$ ist stets*

$$\log(2)\, \frac{V(E)}{\#E} \ \le \ \frac{\log \Delta(E)}{\#E} \ \le \ \frac{V(E)}{\#E} \log\Bigl(e\, \frac{\#E}{V(E)}\Bigr).$$

Satz 5.8 wurde mehrfach unabhängig bewiesen und publiziert, von Vapnik–Červonenkis (1971), von Sauer (1972) und von Shelah (1972), weshalb er in der Kombinatorik unter dem Namen Sauer–Shelah–Lemma bekannt ist. Einen direkten Beweis findet man in der Monographie von Pollard (1984). Hier präsentieren wir stattdessen den Beweis einer noch allgemeineren Aussage von Pajor (1985).

Satz 5.10 (Pajor). *Für jede endliche Menge $E \subset \mathcal{X}$ existieren mindestens $\Delta(E, \mathcal{D})$ verschiedene Mengen $S \subset E$, die von $\mathcal{D}$ vollständig zerlegt werden.*

Dass Satz 5.8 aus Satz 5.10 folgt, kann man wie folgt sehen: Wenn $\mathcal{D}$ keine Menge $B \subset E$ mit $\#B > V(E)$ Elementen vollständig zerlegt, dann haben alle Mengen S in Satz 5.10 höchstens $V(E)$ Elemente. Doch es gibt genau

$$\sum_{i=0}^{V(E)} \binom{\#E}{i}$$

Teilmengen von E mit höchstens $V(E)$ Elementen, also ist diese Zahl eine obere Schranke für $\Delta(E)$.

Beweis von Satz 5.10. Wir beweisen die Aussage durch vollständige Induktion nach $\Delta(E, \mathcal{D})$.

Induktionsanfang: Die Menge $S_1 := \emptyset$ erfüllt stets die Gleichung

$$\Delta(S_1, \mathcal{D}) \;=\; 1 \;=\; 2^{\#S_1}.$$

Im Falle von $\Delta(E, \mathcal{D}) = 1$, ist die Aussage also wahr.

Induktionsschritt: Angenommen, für ein gegebenes $k \in \mathbb{N}$ gilt die Aussage für alle Mengenfamilien $\tilde{\mathcal{D}}$, so dass $\Delta(E, \tilde{\mathcal{D}}) \le k$. Nun sei $\mathcal{D}$ derart, dass $\Delta(E, \mathcal{D}) = k + 1$. Dann existieren Mengen $D_1, D_2 \in \mathcal{D}$, so dass $D_1 \cap E \ne D_2 \cap E$. Insbesondere existiert ein Punkt $x \in E$, so dass

$$\begin{aligned} x \;&\in\; D \cap E \quad \text{für mindestens ein } D \in \mathcal{D}, \\ x \;&\notin\; D' \cap E \quad \text{für mindestens ein } D' \in \mathcal{D}. \end{aligned}$$

Definieren wir also

$$\mathcal{D}_x \;:=\; \{D \in \mathcal{D} \colon x \in D\} \quad \text{und} \quad \mathcal{D}'_x \;:=\; \{D' \in \mathcal{D} \colon x \notin D\},$$

dann ist

$$\Delta(E, \mathcal{D}_x) \;\le\; k, \quad \Delta(E, \mathcal{D}'_x) \;\le\; k \quad \text{und} \quad \Delta(E, \mathcal{D}_x) + \Delta(E, \mathcal{D}'_x) \;=\; k + 1.$$

Wir wollen zeigen, dass die Mengenfamilie

$$\mathcal{S} \;:=\; \bigl\{S \subset E \colon \Delta(S, \mathcal{D}) = 2^{\#S}\bigr\}$$

mindestens $\Delta(E, \mathcal{D})$ verschiedene Mengen enthält. Zu diesem Zweck zerlegen wir $\mathcal{S}$ in die vier disjunkten Teilfamilien

$$\begin{aligned} \mathcal{S}_{11} \;&:=\; \bigl\{S \subset E \colon \Delta(S, \mathcal{D}_x) = 2^{\#S} = \Delta(S, \mathcal{D}'_x)\bigr\}, \\ \mathcal{S}_{10} \;&:=\; \bigl\{S \subset E \colon \Delta(S, \mathcal{D}_x) = 2^{\#S} > \Delta(S, \mathcal{D}'_x)\bigr\}, \\ \mathcal{S}_{01} \;&:=\; \bigl\{S \subset E \colon \Delta(S, \mathcal{D}_x) < 2^{\#S} = \Delta(S, \mathcal{D}'_x)\bigr\}, \\ \mathcal{S}_{00} \;&:=\; \bigl\{S \subset E \colon \Delta(S, \mathcal{D}) = 2^{\#S} > \max\{\Delta(S, \mathcal{D}_x), \Delta(S, \mathcal{D}'_x)\}\bigr\}. \end{aligned}$$

Nach Voraussetzung ist

$$k \;\ge\; \begin{cases} \Delta(E,\mathcal{D}_x) \;=\; \#\mathcal{S}_{11} + \#\mathcal{S}_{10}, \\ \Delta(E,\mathcal{D}'_x) \;=\; \#\mathcal{S}_{11} + \#\mathcal{S}_{01}. \end{cases}$$

Es genügt also zu zeigen, dass

$$\#\mathcal{S}_{11} \;\le\; \#\mathcal{S}_{00}.$$

Zu diesem Zweck sei $S \in \mathcal{S}_{11}$. Dann ist sicher $x \notin S$, denn sonst wäre $S \neq S \cap D'$ für alle $D' \in \mathcal{D}'_x$. Daher definiert

$$\mathcal{S}_{11} \ni S \;\mapsto\; S \cup \{x\}$$

eine injektive Abbildung, und es genügt zu zeigen, dass $S \cup \{x\} \in \mathcal{S}_{00}$ für $S \in \mathcal{S}_{11}$. Für jede Menge $B \subset S \cup \{x\}$ gibt es nach Voraussetzung Mengen $D \in \mathcal{D}_x$ und $D' \in \mathcal{D}'_x$ so dass

$$B \cap S \;=\; D \cap S \;=\; D' \cap S.$$

Also ist

$$B \;=\; \begin{cases} (S \cup \{x\}) \cap D & \text{falls } x \in B, \\ (S \cup \{x\}) \cap D' & \text{falls } x \notin B. \end{cases}$$

Dies zeigt, dass $S \cup \{x\}$ von $\mathcal{D}$ vollständig zerlegt wird. Doch

$$\begin{aligned} \mathcal{D}_x \cap (S \cup \{x\}) \;&=\; \{A \cup \{x\} \colon A \subset S\}, \quad \text{also} \quad \Delta(S \cup \{x\}, \mathcal{D}_x) \;=\; 2^{\#S} \;<\; 2^{\#(S \cup \{x\})}, \\ \mathcal{D}'_x \cap (S \cup \{x\}) \;&=\; \{A \colon A \subset S\}, \qquad\quad\; \text{also} \quad \Delta(S \cup \{x\}, \mathcal{D}_x) \;=\; 2^{\#S} \;<\; 2^{\#(S \cup \{x\})}, \end{aligned}$$

so dass tatsächlich $S \cup \{x\} \in \mathcal{S}_{00}$. □

Beweis von Korollar 5.9. Im Falle von $V(E) = 0$ besteht $\mathcal{D} \cap E$ nur aus einer Menge, und die behaupteten Ungleichungen sind Gleichungen, wenn man $0 \cdot \log(e/0)$ als $\lim_{x \downarrow 0} x \log(e/x) = 0$ interpretiert.

Sei also $k := V(E) \ge 1$, und wir schreiben $n := \#E$. Falls $B \subset E$ von $\mathcal{D}$ vollständig zerlegt wird, ist $\Delta(E) = \#(\mathcal{D} \cap E) \ge \#(\mathcal{D} \cap B) = 2^{\#B}$. Dies zeigt, dass $\Delta(E) \ge 2^{V(E)}$, woraus die erste behauptete Ungleichung folgt.

Da $x \log(e/x)$ monoton wachsend in $x \in [0,1]$ ist, gilt im Falle von $k/n \ge 1/2$ die Ungleichung

$$\frac{k}{n} \log\Bigl(e \frac{n}{k}\Bigr) \;\ge\; \frac{1}{2} \log(2e) \;>\; \log(2) \;=\; \frac{\log(2^n)}{n} \;\ge\; \frac{\log \Delta(E)}{n}.$$

Es genügt daher, den Fall $0 < k/n < 1/2$ zu untersuchen. Nach Satz 5.8 ist

$$\begin{aligned}
\frac{\log \Delta(E)}{n} &\leq \frac{1}{n} \log\Big(\sum_{i\leq k} \binom{n}{i}\Big) \\
&\leq \frac{1}{n} \log\Big(\sum_{i\leq n} \binom{n}{i} e^{\lambda(k-i)}\Big) \qquad [\text{für beliebige } \lambda \geq 0] \\
&= \frac{1}{n} \log\big(e^{\lambda k}(1+e^{-\lambda})^n\big) \\
&= \lambda \frac{k}{n} + \log(1+e^{-\lambda}) \\
&= H\Big(\frac{k}{n}\Big) \qquad \Big[\text{für } \lambda = \log\Big(\frac{n}{k} - 1\Big) > 0\Big],
\end{aligned}$$

wobei $H(x) := -x\log x - (1-x)\log(1-x)$ für $0 < x < 1$. Doch

$$-(1-x)\log(1-x) \;=\; (1-x)\sum_{k=1}^{\infty} \frac{x^k}{k} \;\leq\; (1-x)\sum_{k=1}^{\infty} x^k \;=\; x,$$

weshalb $H(x) \leq x - x\log(x) = x\log(e/x)$, also $\log(\Delta(E))/n \leq (k/n)\log(en/k)$. $\square$

5.3 Vapnik-Červonenkis-Klassen

Die Familie $\mathcal{D}$ heißt *Vapnik-Červonenkis-Klasse* (VC-Klasse), falls es eine ganze Zahl $V \geq 0$ gibt, so dass $\mathcal{D}$ keine $(V+1)$-elementige Teilmenge von $\mathcal{X}$ zerlegt. In Formeln: $\mathcal{D}$ ist eine VC-Klasse, falls

$$V(\mathcal{D}) \;:=\; \sup\big\{\#A : A \subset \mathcal{X} \text{ endlich}, \Delta(A,\mathcal{D}) = 2^{\#A}\big\}$$

endlich ist. Mitunter nennt man diese Zahl $V(\mathcal{D})$ den *Vapnik-Červonenkis-Index* oder die *Vapnik-Červonenkis-Dimension* von $\mathcal{D}$. Für manche Autoren (z.B. van der Vaart und Wellner 1996) ist auch $V(\mathcal{D}) + 1$ die Vapnik-Červonenkis-Dimension von $\mathcal{D}$. Offensichtlich ist

$$V(E,\mathcal{D}) \;\leq\; V(\mathcal{D})$$

für beliebige endliche Mengen $E \subset \mathcal{X}$.

Hier ist eine andere Charakterisierung der VC-Eigenschaft.

Lemma 5.11. *Eine Mengenfamilie $\mathcal{D}$ ist genau dann eine Vapnik-Červonenkis-Klasse, wenn es ein reelles Polynom p gibt, so dass*

$$\Delta(E,\mathcal{D}) \;\leq\; p(\#E)$$

für beliebige endliche Teilmengen E von $\mathcal{X}$. (Aus diesem Grunde spricht man auch von Mengenfamilien mit polynomialer Diskrimination.)

Beweis. Ist $V := V(\mathcal{D}) < \infty$, dann folgt aus Satz 5.8, dass $\Delta(E) \le p(\#E)$ mit dem Polynom

$$p(n) \ := \ \sum_{i=0}^{V} \binom{n}{i}.$$

(Zunächst gilt dies nur für den Fall $\#E \ge V$. Doch $p(n) = 2^n$ für $0 \le n \le V$, also gilt die Gleichung $\Delta(E) \le p(\#E)$ auch im Falle von $\#E < V$.) Andererseits gibt es zu einem beliebigen Polynom p eine ganze Zahl $V \ge 0$ mit $2^n > p(n)$ für alle $n > V$. Falls also $\Delta(E, \mathcal{D}) \le p(\#E)$ für beliebige endliche Mengen $E \subset \mathcal{X}$, so ist sicher $V(\mathcal{D}) \le V$. □

Das folgende Lemma ist oft nützlich, wenn man nachweisen will, dass ein bestimmtes Mengensystem eine VC-Klasse ist, indem man es durch endlich viele Boolesche Operationen aus einfacheren Mengenfamilien aufbaut. Den Beweis überlassen wir dem Leser als Übungsaufgabe.

Lemma 5.12. *Seien $\mathcal{C}, \mathcal{D}$ Familien von Teilmengen von $\mathcal{X}$. Dann ist*

$$\begin{aligned}\Delta\big(E, \{\mathcal{X} \setminus D : D \in \mathcal{D}\}\big) \ &= \ \Delta(E, \mathcal{D}),\\ \Delta\big(E, \{C \star D : C \in \mathcal{C}, D \in \mathcal{D}\}\big) \ &\le \ \Delta(E, \mathcal{C})\Delta(E, \mathcal{D})\end{aligned}$$

für beliebige endliche Teilmengen E von $\mathcal{X}$. Dabei bezeichnet '$\star$' eine der Booleschen Verknüpfungen '$\cap$', '$\cup$' oder '$\setminus$'. □

Hier ist ein anderes Kriterium für die VC-Eigenschaft.

Satz 5.13. *Sei $\mathcal{G}$ ein endlichdimensionaler Vektorraum von reellen Funktionen auf $\mathcal{X}$. Dann ist*

$$\mathcal{D} \ := \ \big\{\{x \in \mathcal{X} : g(x) \ge 0\} : g \in \mathcal{G}\big\}$$

eine VC-Klasse mit VC-Index $V(\mathcal{D}) \le \dim(\mathcal{G})$.

Beweis von Satz 5.13. Sei $E \subset \mathcal{X}$ mit $\#E > \dim(\mathcal{G})$. Die Menge

$$\big\{(g(x))_{x \in E} : g \in \mathcal{G}\big\}$$

ist ein Untervektorraum von $\mathbb{R}^E$ mit Dimension $d \le \dim(\mathcal{G}) < \dim(\mathbb{R}^E) = \#E$. Daher existiert ein $h = \big(h(x)\big)_{x \in E}$ in $\mathbb{R}^E$ derart, dass

$$\sum_{x \in E} h(x)g(x) \ = \ 0 \quad \text{für alle } g \in \mathcal{G}$$

und $h(x) \ne 0$ für mindestens ein $x \in E$. Multipliziert man h mit -1, falls nötig, dann ist

$$A \ := \ \{x \in E : h(x) \ge 0\} \ \ne \ E.$$

Doch $A \ne \{g \ge 0\} \cap E$ für beliebige $g \in \mathcal{G}$. Denn anderenfalls wäre

$$\sum_{x \in E} h(x)g(x) \ = \ \sum_{x \in A} \underbrace{h(x)g(x)}_{\ge 0} + \sum_{x \in E \setminus A} \underbrace{h(x)g(x)}_{> 0} \ > \ 0.$$

Also wird E von $\mathcal{D}$ nicht vollständig zerlegt. □

Beispiel 5.14 (Halbräume und Kegelschnitte). Sei $\mathcal{G}$ die Menge aller affin linearen Funktionen

$$x \mapsto g(x) = \lambda_0 + \sum_{i=1}^{d} \lambda_i x_i \quad (\lambda_0, \ldots, \lambda_d \in \mathbb{R})$$

auf $\mathbb{R}^d$. Dann ist $\big\{\{g \geq 0\} : g \in \mathcal{G}\big\}$ die Menge aller abgeschlossenen Halbräume (inklusive $\mathbb{R}^d$ und $\emptyset$) und stellt eine VC-Klasse mit Index $V(\mathcal{D}) \leq d + 1$.

Betrachtet man stattdessen die Menge $\mathcal{G}$ aller Funktionen

$$x \mapsto g(x) = \lambda_0 + \sum_{i=1}^{d} \lambda_i x_i + \sum_{i,j=1}^{d} \lambda_{ij} x_i x_j$$

mit reellen Koeffizienten $\lambda_i, \lambda_{ij} = \lambda_{ji}$, dann ist $\big\{\{g \geq 0\} : g \in \mathcal{G}\big\}$ die Menge aller abgeschlossenen Kegelschnitte (Ellipsen, Hyperboloide, Halbräume, Paraboloide) und zerlegt keine Teilmenge von $\mathbb{R}^d$ mit mehr als $(d + 1)(d + 2)/2$ Elementen.

Hier ist noch eine Anmerkung zur punktweisen Separabilität von $\mathcal{D}$:

Lemma 5.15. *Seien $\mathcal{G}$ und $\mathcal{D}$ wie in Satz 5.13. Ferner enthalte $\mathcal{G}$ eine strikt positive Funktion. Dann ist $\mathcal{D}$ punktweise separabel.*

Beweis von Lemma 5.15. Sei $\mathcal{G} = \mathrm{span}(v_0, v_1, v_2, \ldots, v_d)$ mit $v_0 > 0$. Die Menge $\mathcal{G}_o$ aller Linearkombinationen von $v_0, v_1, v_2, \ldots, v_d$ mit rationalen Koeffizienten ist abzählbar und liefert eine Menge $\mathcal{D}_o := \big\{\{g \geq 0\} : g \in \mathcal{G}_o\big\}$ mit den gewünschten Eigenschaften. Seien nämlich $\lambda_0, \lambda_1, \ldots, \lambda_d$ beliebige reelle Zahlen. Dann wähle man Folgen $(\lambda_{ik})_k$ von rationalen Zahlen, so dass

$$\lim_{k\to\infty} k|\lambda_{ik} - \lambda_i| \;=\; 0$$

für $0 \leq i \leq d$. Mit

$$g(x) \;:=\; \sum_{i=0}^{d} \lambda_i v_i(x) \quad \text{und} \quad g_k(x) \;:=\; \frac{v_0(x)}{k} + \sum_{i=0}^{d} \lambda_{ik} v_i(x)$$

konvergiert dann $(g_k)_k$ punktweise gegen g, und für jedes $x \in \mathcal{X}$ gibt es eine Zahl $k_o(x)$, so dass $g_k(x) > g(x)$ falls $k \geq k_o(x)$. Dies impliziert, dass

$$\lim_{k\to\infty} 1\{g_k(x) \geq 0\} \;=\; 1\{g(x) \geq 0\} \quad \text{für alle } x \in \mathcal{X}.$$

□

5.4 Aufgaben

Aufgabe 5.1. Sei $\mathcal{X} = \mathbb{N}$ und $\mathcal{D}$ die Menge aller Teilmengen von $\mathbb{N}$. Beweisen Sie die uniforme Konsistenz von $\hat{P}_n$ auf $\mathcal{D}$ mit Hilfe des Satzes 5.4.

Hinweis: Zeigen Sie, dass $\mathbb{E}(\hat{V}_n/n) \leq \sum_{k\in\mathbb{N}} \min(P\{k\}, 1/n)$.

Aufgabe 5.2. Beweisen Sie Lemma 5.12.

Aufgabe 5.3. Für $m, d \in \mathbb{N}$ sei $\mathcal{D}_{m,d}$ die Menge aller konvexen Hüllen von bis zu m Punkten in $\mathbb{R}^d$. Zeigen Sie, dass $\mathcal{D}_{m,d}$ eine Vapnik-Červonenkis-Klasse ist. Geben Sie auch eine explizite Schranke für $V(\mathcal{D}_{m,d})$ an. Für welche Kombinationen (m, d) ist $\mathcal{D}_{m,d}$ punktweise separabel?

Aufgabe 5.4. Seien $\mathcal{C}$ und $\mathcal{D}$ VC-Klassen über Mengen $\mathcal{X}$ bzw. $\mathcal{Y}$. Zeigen Sie, dass $\mathcal{C} \times \mathcal{D} := \{C \times D : C \in \mathcal{C}, D \in \mathcal{D}\}$ eine VC-Klasse über $\mathcal{X} \times \mathcal{Y}$ ist.

Aufgabe 5.5. Sei $\mathcal{D}$ die Familie aller Mengen der Form

$$\big\{(x, y) \in \mathbb{R} \times \mathbb{R} : y \leq h(x)\big\}$$

mit einer monoton wachsenden Funktion $h : \mathbb{R} \to \mathbb{R}$. Zeigen Sie, dass $V(\mathcal{D}) = \infty$.

Kapitel 6

Abstrakte Gesetze der grossen Zahlen

In diesem Kapitel werden die Resultate für mengenindizierte empirische Prozesse auf den abstrakten Rahmen von Abschnitt 1.3 übertragen. Hierbei spielt die Approximation der Indexmenge $\mathcal{T}$ durch endliche Teilmengen eine zentrale Rolle. Dabei werden wir auch verstehen, weshalb man von der Vapnik-Červonenkis-*Dimension* spricht.

6.1 Überdeckungszahlen

Sei $(\mathcal{T}, \rho)$ ein pseudometrischer Raum. Das heißt, ρ ist eine Funktion auf $\mathcal{T} \times \mathcal{T}$, so dass für beliebige $s, t, u \in \mathcal{T}$ gilt:

$$\begin{aligned} \rho(s,t) &= \rho(t,s), \\ \rho(s,t) &\geq 0 \quad \text{mit Gleichheit, falls } s = t, \\ \rho(s,t) &\leq \rho(s,u) + \rho(u,t). \end{aligned}$$

Für $\epsilon > 0$ definieren wir nun die *Überdeckungszahl*

$$N(\epsilon) = N(\epsilon, \mathcal{T}, \rho) \;:=\; \min\bigl\{\#\mathcal{T}_o \,:\, \mathcal{T}_o \subset \mathcal{T}, \rho(t, \mathcal{T}_o) \leq \epsilon \text{ für alle } t \in \mathcal{T}\bigr\}.$$

Dabei ist

$$\rho(t, \mathcal{T}_o) \;:=\; \inf_{s \in \mathcal{T}_o} \rho(t,s).$$

Mit anderen Worten, $N(\epsilon, \mathcal{T}, \rho)$ ist die kleinste Anzahl von abgeschlossenen Kugeln mit Radius ϵ bezüglich ρ, welche die Menge $\mathcal{T}$ überdecken. Mitunter betrachtet man auch die *Kapazitätszahl*

$$D(\epsilon) = D(\epsilon, \mathcal{T}, \rho) \;:=\; \max\bigl\{\#\mathcal{T}_o \,:\, \mathcal{T}_o \subset \mathcal{T}, \rho(s,t) > \epsilon \text{ für verschiedene } s, t \in \mathcal{T}_o\bigr\}.$$

Falls $\mathcal{T}$ Teilmenge eines normierten Raumes $(\mathbb{V}, |\cdot|)$ und ρ die durch die Norm $|\cdot|$ induzierte Metrik ist, schreiben wir auch $N(\epsilon, \mathcal{T}, |\cdot|)$ und $D(\epsilon, \mathcal{T}, |\cdot|)$ an Stelle von $N(\epsilon, \mathcal{T}, \rho)$ bzw. $D(\epsilon, \mathcal{T}, \rho)$.

Zwischen Überdeckungs- und Kapazitätszahlen besteht folgender Zusammenhang:

Lemma 6.1.
$$N(\epsilon) \ \le\ D(\epsilon) \ \le\ N(\epsilon/2).$$

Aufgrund dieser Ungleichungen kann man mit Überdeckungs- oder Kapazitätszahlen arbeiten, je nachdem, was gerade praktischer ist.

Beweis von Lemma 6.1. Sei $\mathcal{T}_o$ eine maximale Teilmenge von $\mathcal{T}$, so dass

(6.1) $$\rho(s,t) > \epsilon \text{ für verschiedene Punkte } s,t \in \mathcal{T}_o.$$

Dann ist $\rho(u,\mathcal{T}_o) \le \epsilon$ für alle $u \in \mathcal{T}$. Denn anderenfalls gäbe es einen Punkt $u_o \in \mathcal{T}$ mit $\rho(u_o,\mathcal{T}_o) > \epsilon$. Doch dann könnte man $\mathcal{T}_o$ durch die größere Menge $\mathcal{T}_o \cup \{u_o\}$ ersetzen, ohne (6.1) zu verletzen, was der Maximalität von $\mathcal{T}_o$ widerspricht. Dies zeigt, dass $N(\epsilon) \le \#\mathcal{T}_o = D(\epsilon)$.

Nun sei $\mathcal{T}_*$ eine minimale endliche Teilmenge von $\mathcal{T}$, so dass $\rho(s,\mathcal{T}_*) \le \epsilon/2$ für alle $s \in \mathcal{T}$. Ferner sei $\mathcal{T}_o$ eine Menge von Punkten aus $\mathcal{T}$, die paarweise echt größeren Abstand als ϵ haben. Nach der Voraussetzung an $\mathcal{T}_*$ kann man jedem Punkt $t_o \in \mathcal{T}_o$ einen Punkt $t_* \in \mathcal{T}_*$ zuordnen, so dass $\rho(t_o,t_*) \le \epsilon/2$. Aus der Voraussetzung an $\mathcal{T}_o$ und der Dreiecksungleichung folgt außerdem, dass diese Zuordnung injektiv sein muss. Denn sonst gäbe es zwei verschiedene Punkte $s,t \in \mathcal{T}_o$ und ein $t_* \in \mathcal{T}_*$ mit $\rho(s,t) \le \rho(s,t_*) + \rho(t_*,t) \le \epsilon/2 + \epsilon/2 = \epsilon$. Dies zeigt, dass $\#\mathcal{T}_o \le \#\mathcal{T}_* = N(\epsilon/2)$, weshalb $D(\epsilon) \le N(\epsilon/2)$. □

Nun betrachten wir ein einfaches, aber wichtiges Beispiel für die Abschätzung von Kapazitätszahlen.

Lemma 6.2. *Sei* $(\mathbb{V}, |\cdot|)$ *ein endlichdimensionaler normierter Vektorraum über* $\mathbb{R}$, *und sei* $B := \{t \in \mathbb{V} : |t| \le 1\}$ *die entsprechende abgeschlossene Einheitskugel. Dann ist*
$$\Big(\frac{C(B)}{\epsilon}\Big)^{\dim(\mathbb{V})} \ \le\ D(\epsilon, B, |\cdot|) \ \le\ \Big(1 + \frac{2}{\epsilon}\Big)^{\dim(\mathbb{V})}$$
für beliebige $\epsilon > 0$ *mit einer Konstante* $C(B) > 0$. *Insbesondere ist*
$$\lim_{\epsilon\downarrow 0} \frac{\log D(\epsilon, B, |\cdot|)}{\log(1/\epsilon)} \ =\ \lim_{\epsilon\downarrow 0} \frac{\log N(\epsilon, B, |\cdot|)}{\log(1/\epsilon)} \ =\ \dim(\mathbb{V}).$$

Allgemein nennt man die Abbildung
$$\epsilon \ \mapsto\ \log N(\epsilon, \mathcal{T}, \rho) \quad \text{oder} \quad \epsilon \ \mapsto\ \log D(\epsilon, \mathcal{T}, \rho)$$
die *metrische Entropie* von $(\mathcal{T}, \rho)$, und
$$\limsup_{\epsilon\downarrow 0} \frac{\log N(\epsilon, \mathcal{T}, \rho)}{\log(1/\epsilon)} \ =\ \limsup_{\epsilon\downarrow 0} \frac{\log D(\epsilon, \mathcal{T}, \rho)}{\log(1/\epsilon)}$$
ist die *metrische Dimension* von $(\mathcal{T}, \rho)$.

Beweis von Lemma 6.2. Ohne Einschränkung sei $\mathbb{V} = \mathbb{R}^d$. Sei $B_o \subset B$ derart, dass $\rho(s,t) > \epsilon$ für verschiedene $s, t \in B_o$. Dann sind die Kugeln $t + (\epsilon/2)B$ mit Mittelpunkten $t \in B_o$ paarweise disjunkt und in der Menge $(1 + \epsilon/2)B$ enthalten. Folglich ist

$$\mathrm{Leb}((1+\epsilon/2)B) \;\ge\; \sum_{t \in B_o} \mathrm{Leb}(t + (\epsilon/2)B),$$

also

$$(1+\epsilon/2)^d \,\mathrm{Leb}(B) \;\ge\; \#B_o \cdot (\epsilon/2)^d \,\mathrm{Leb}(B),$$

weshalb $\#B_o \le (1 + 2/\epsilon)^d$. Dies zeigt, dass $D(\epsilon, B, |\cdot|) \le (1 + 2/\epsilon)^d$.

Andererseits gibt es strikt positive Konstanten $c_1 = c_1(B)$ und $c_2 = c_2(B)$ derart, dass

$$[-c_1, c_1]^d \;\subset\; B \;\subset\; [-c_2, c_2]^d.$$

Für $\delta > 0$ ist dann

$$B_o \;:=\; \bigl\{(-c_1 + \delta k_i)_{i=1}^d : k \in \{0, 1, \ldots, \lfloor 2c_1/\delta \rfloor\}^d\bigr\}$$

eine Teilmenge von B derart, dass $s - t \not\in (-\delta, \delta)^d = (\delta/c_2)(-c_2, c_2)^d$, also $|s - t| \ge \delta/c_2$ für zwei verschiedene Punkte $s, t \in B_o$. Ferner ist $\#B_o \ge (2c_1/\delta)^d = \bigl(C(B)/(\delta/c_2)\bigr)^d$ mit $C(B) := 2c_1/c_2$. Dies zeigt, dass $D(\epsilon, B, |\cdot|) \ge (C(B)/\epsilon)^d$ für alle $\epsilon > 0$.

Aus den beiden Ungleichungen für $D(\epsilon, B, |\cdot|)$ und Lemma 6.1 ergibt sich die Behauptung über den Grenzwert von $\log N(\epsilon)/\log(1/\epsilon)$ bzw. $\log D(\epsilon)/\log(1/\epsilon)$. □

6.2 Funktionenklassen

Sei $\mathcal{F}$ eine Familie von messbaren Funktionen auf $\mathcal{X}$, und sei M ein Maß auf $\mathcal{X}$. Nun betrachten wir folgende Pseudometrik ρ_M auf $\mathcal{F}$:

$$\rho_M(f, g) \;:=\; \int |f - g| \, dM.$$

Wir nehmen an, dass $\mathcal{F}$ eine *messbare Einhüllende* F besitzt. Das heißt, F ist eine messbare Funktion auf $\mathcal{X}$, so dass

$$|f| \;\le\; F \quad \text{für alle } f \in \mathcal{F}.$$

Ferner nehmen wir an, dass $M(F) = \int M \, dF$ strikt positiv und endlich ist. Zur Abschätzung der metrischen Entropie von $(\mathcal{F}, \rho_M)$ betrachten wir für eine Funktion $f \in \mathcal{F}$ ihren *Subgraphen*

$$\mathrm{sgr}(f) \;:=\; \bigl\{(x, r) \in \mathcal{X} \times \mathbb{R} : r \le f(x)\bigr\}.$$

Wenn die Mengenfamilie $\mathrm{sgr}(\mathcal{F}) := \{\mathrm{sgr}(f) : f \in \mathcal{F}\}$ eine Vapnik-Červonenkis-Klasse ist, kann man die Kapazitätszahlen $D(\epsilon, \mathcal{F}, \rho_M)$ konkret nach oben abschätzen, fast unabhängig vom Maß M.

Satz 6.3 (Dudley, Pollard). *Angenommen,* $\mathrm{sgr}(\mathcal{F})$ *ist eine VC-Klasse und* $0 < M(F) < \infty$. *Dann ist*

$$\log D\bigl(M(F)\epsilon, \mathcal{F}, \rho_M\bigr) \ \le\ V(\mathrm{sgr}(\mathcal{F}))\bigl(\log(e/\epsilon) + \log\log(e/\epsilon) + 3\bigr)$$

für beliebige $\epsilon \in (0, 1]$.

Anmerkung 6.4. Dieser Satz zeigt, dass

$$\limsup_{\epsilon \downarrow 0} \frac{\log D(\epsilon, \mathcal{F}, \rho_M)}{\log(1/\epsilon)} \ \le\ V(\mathrm{sgr}(\mathcal{F})).$$

Vergleicht man dies mit Lemma 6.2, dann wird verständlich, warum man auch von der VC-*Dimension* $V(\mathrm{sgr}(\mathcal{F}))$ spricht.

Anmerkung 6.5. Für eine Mengenfamilie $\mathcal{D}$ sei $\tilde{\mathcal{D}}$ die Familie der entsprechenden Indikatorfunktionen. Dann ist $\mathcal{D}$ eine VC-Klasse genau dann, wenn $\mathrm{sgr}(\tilde{\mathcal{D}})$ eine VC-Klasse ist. Genauer gesagt, ist $V(\mathcal{D}) = V(\mathrm{sgr}(\tilde{\mathcal{D}}))$.

Beweis von Satz 6.3. Sei $V := V(\mathrm{sgr}(\mathcal{F})) \ge 1$, denn anderenfalls besteht $\mathcal{F}$ ohnehin nur aus einer Funktion. Ferner sei $M(F) = 1$, denn anderenfalls könnten wir einfach M durch $M(F)^{-1}M$ ersetzen. Sei $\mathcal{F}_o$ eine beliebige Teilmenge von $\mathcal{F}$ mit $m := \#\mathcal{F}_o < \infty$ und

$$\rho_M(f, g) \ >\ \epsilon \quad \text{für verschiedene } f, g \in \mathcal{F}_o.$$

Nun konstruieren wir zufällige endliche Teilmengen von $\mathcal{X} \times \mathbb{R}$ mit Hilfe von unabhängigen und identisch verteilten Zufallsvariablen $(X_1, R_1), (X_2, R_2), (X_3, R_3), \ldots$ mit Werten in $\mathcal{X} \times \mathbb{R}$, wobei

$$\mathbb{P}(X_1 \in B) \ =\ \int_B F\, dM \quad \text{und} \quad \mathcal{L}(R_1 \,|\, X_1) \ =\ \mathcal{U}\bigl[-F(X_1), F(X_1)\bigr].$$

Dann definieren wir $E_n := \bigl\{(X_i, R_i) : 1 \le i \le n\bigr\}$, also $\#E_n \le n$ mit Gleichheit fast sicher. Nun zeigen wir, dass $\Delta(E_n, \mathrm{sgr}(\mathcal{F})) \ge m$ mit strikt positiver Wahrscheinlichkeit, falls n hinreichend groß ist. Für beliebige $f, g \in \mathcal{F}_o$ mit $f \not\equiv g$ gilt:

$$\begin{aligned}
&\mathbb{P}\bigl(\mathrm{sgr}(f) \cap E_n = \mathrm{sgr}(g) \cap E_n\bigr) \\
&\quad = \ \mathbb{P}\bigl((X_i, R_i) \notin \mathrm{sgr}(f) \triangle \mathrm{sgr}(g) \text{ für alle } i \le n\bigr) \\
&\quad = \ \mathbb{P}\bigl((X_1, R_1) \notin \mathrm{sgr}(f) \triangle \mathrm{sgr}(g)\bigr)^n \\
&\quad = \ \Bigl(1 - \mathbb{P}\bigl((X_1, R_1) \in \mathrm{sgr}(f) \triangle \mathrm{sgr}(g)\bigr)\Bigr)^n \\
&\quad = \ \Bigl(1 - \mathbb{E}\,\mathbb{P}\Bigl(R_1 \in [-\infty, f(X_1)] \triangle [-\infty, g(X_1)] \Bigm| X_1\Bigr)\Bigr)^n \\
&\quad = \ \Bigl(1 - \mathbb{E}\, \frac{|f(X_1) - g(X_1)|}{2F(X_1)}\Bigr)^n \\
&\quad = \ \bigl(1 - \rho_M(f, g)/2\bigr)^n \\
&\quad < \ (1 - \epsilon/2)^n \\
&\quad \le \ \exp(-n\epsilon/2).
\end{aligned}$$

Folglich ist

$$\mathbb{P}\big(\Delta(E_n, \mathrm{sgr}(\mathcal{F}_o)) < m\big) \ \le\ \binom{m}{2}\exp(-n\epsilon/2) \ <\ \exp(2\log m - n\epsilon/2 - \log 2) \ \le\ 1,$$

falls n größer oder gleich

$$\frac{4\log m}{\epsilon} - \frac{2\log 2}{\epsilon} \ \le\ n_o := \Big\lfloor \frac{4\log m}{\epsilon} \Big\rfloor$$

ist. Es existiert also eine Teilmenge E von $\mathcal{X} \times \mathbb{R}$ mit höchstens n_o Elementen derart, dass $\Delta\big(E, \mathrm{sgr}(\mathcal{F})\big) \ge m$. Nach Korollar 5.9 ist

$$\frac{\log m}{n_o} \ \le\ \frac{V}{n_o}\log\Big(e\frac{n_o}{V}\Big),$$

also

$$\frac{\log m}{V} \ \le\ \log\Big(e\frac{4\log m}{V\epsilon}\Big) \ =\ \log\frac{\log m}{V} + \log 4 + \log\frac{e}{\epsilon}. \tag{6.2}$$

Zu zeigen ist nun, dass $y := \log(m)/V$ nicht größer ist als $x + \log x + 3$, wobei $x := \log(e/\epsilon) \ge 1$. Ungleichung (6.2) kann man umformen zu

$$y - \log y \ \le\ x + \log 4. \tag{6.3}$$

Angenommen $y > x + \log x + 3$. Dann folgte aus Ungleichung (6.3) und der strikten Isotonie von $y \mapsto y - \log y$ auf $[1, \infty)$, dass

$$\begin{aligned} x + \log 4 >\ & x + \log x + 3 - \log(x + \log x + 3) \\ =\ & x + 3 - \log(1 + (\log x + 3)/x) \\ \ge\ & x + 3 - \log 4, \end{aligned}$$

denn $x \mapsto (\log x + 3)/x$ ist monoton fallend auf $[1, \infty)$. Dies ergäbe aber die Ungleichung $3 < 2\log 4$ beziehungsweise $\exp(3/4) < 2$, welche definitiv falsch ist. $\square$

Uniforme Überdeckungszahlen. Um die Abhängigkeit von M loszuwerden, definieren wir die *uniformen Überdeckungszahlen*

$$N(\epsilon, \mathcal{F}) \ :=\ \sup\big\{N\big(M(F)\epsilon, \mathcal{F}, \rho_M\big) : M \text{ ein Maß auf } \mathcal{X}, 0 < M(F) < \infty\big\}.$$

Für ein beliebiges Maß M mit $0 < M(F) < \infty$ ist also

$$N(\epsilon, \mathcal{F}, \rho_M) \ \le\ N\Big(\frac{\epsilon}{M(F)}, \mathcal{F}\Big).$$

Aus Satz 6.3 folgt nun, dass

$$\log N(\epsilon, \mathcal{F}) \ \le\ V(\mathrm{sgr}(\mathcal{F}))\big(\log(e/\epsilon) + \log\log(e/\epsilon) + 3\big) \quad \text{für } 0 < \epsilon \le 1.$$

Im Falle einer VC-Klasse $\mathrm{sgr}(\mathcal{F})$ kann man also die uniformen Überdeckungszahlen explizit nach oben abschätzen.

Kombinationen von Funktionenklassen und andere Metriken. Oftmals kann man eine Funktionenklasse durch elementare Rechenoperationen aus einfacheren Funktionenklassen aufbauen. Die Überdeckungszahlen lassen sich dann wie folgt abschätzen:

Lemma 6.6. *Seien $\mathcal{F}$ und $\mathcal{G}$ Familien messbarer Funktionen auf $\mathcal{X}$ mit messbaren Einhüllenden F beziehungsweise G, und sei M ein Maß auf $\mathcal{X}$. Für beliebige $u > 0$ gilt:*

$$N(u, \mathcal{F} \star \mathcal{G}, \rho_M) \ \le \ N(u/2, \mathcal{F}, \rho_M) N(u/2, \mathcal{G}, \rho_M).$$

Dabei ist $\mathcal{F} \star \mathcal{G} := \{f \star g : f \in \mathcal{F}, g \in \mathcal{G}\}$, und $f \star g$ bezeichnet punktweises Maximum, Minimum, Summe oder Differenz. Ferner ist

$$N(u, \mathcal{F}\mathcal{G}, \rho_M) \ \le \ N(u/2, \mathcal{F}, \rho_{GM}) N(u/2, \mathcal{G}, \rho_{FM})$$

mit der Familie $\mathcal{F}\mathcal{G} := \{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$. Dabei ist $HM(B) := \int_B H \, dM$.

Später werden wir anstelle von ρ_M auch andere Pseudometriken betrachten. Dabei ist folgende Ungleichung von Nutzen:

Lemma 6.7. *Für $1 \le q < \infty$ ist*

$$N(u, \mathcal{F}, \rho_{M,q}) \ \le \ N\big(u^q, \mathcal{F}, \rho_{(2F)^{q-1}M}\big),$$

wobei $\rho_{M,q}(f, g) := \big(\int |f - g|^q \, dM\big)^{1/q}$.

Die Beweise beider Lemmata überlassen wir dem Leser als Übungsaufgabe.

6.3 Uniforme Konsistenz allgemein

Nun geben wir Bedingungen an, unter welchen $\mathbb{E}\,\|Z_n - \mathbb{E}\,Z_n\|$ gegen Null konvergiert. Mit Hilfe der zwei Symmetrisierungen wird dies zurückgeführt auf die Behauptung, dass $\mathbb{E}\,\|Z_n^o\| \to 0$, wobei $Z_n^o := \sum_{i=1}^n \xi_i \phi_{ni}$ wie in Kapitel 3. Die Zufallsvariable $\|Z_n^o\|$ wird durch $\|Z_n^o\|_{\hat{\mathcal{T}}_n}$ approximiert, wobei $\hat{\mathcal{T}}_n$ eine zufällige, in der Regel endliche Teilmenge von $\mathcal{T}$ ist. Für zwei beliebige Punkte $s, t \in \mathcal{T}$ ist stets

$$\big|Z_n^o(s) - Z_n^o(t)\big| \ \le \ \hat{\rho}_n(s, t),$$

wobei

$$\hat{\rho}_n(s, t) = \hat{\rho}_n(s, t \,|\, \boldsymbol{\phi}_n) \ := \ \sum_{i=1}^n \big|\phi_{ni}(s) - \phi_{ni}(t)\big|$$

mit $\boldsymbol{\phi}_n = (\phi_{ni})_{i=1}^n$. Dies definiert eine Pseudometrik $\hat{\rho}_n$ auf $\mathcal{T}$. Wenn nun

$$\hat{\rho}_n(t, \hat{\mathcal{T}}_n) \ \le \ \delta \quad \text{für alle } t \in \mathcal{T}$$

mit einer positiven Zahl δ, so ist

$$\|Z_n^o\|_{\mathcal{T}} \ \le \ \delta + \|Z_n^o\|_{\hat{\mathcal{T}}_n}.$$

Der Term $\|Z_n^o\|_{\hat{\mathcal{T}}_n}$ wird dann mithilfe von Hoeffdings Ungleichung weiter abgeschätzt.

Satz 6.8. *Der Erwartungswert von $\|Z_n - \mathbb{E}\, Z_n\|$ konvergiert gegen Null, falls für eine Nullfolge $(\delta_n)_n$ folgende drei Bedingungen erfüllt sind:*

$$\mathbb{E} \sum_{i=1}^n \|\phi_{ni}\| \;=\; O(1), \tag{6.4}$$

$$\mathbb{E} \sum_{i=1}^n 1\big\{\|\phi_{ni}\| > \delta_n\big\}\|\phi_{ni}\| \;\to\; 0, \tag{6.5}$$

$$\log N(\epsilon, \mathcal{T}, \hat{\rho}_n) \;=\; o_p(\delta_n^{-1}) \quad \textit{für beliebige } \epsilon > 0. \tag{6.6}$$

Dieser Satz beinhaltet zwei Spezialfälle:

Korollar 6.9. $\mathbb{E}\,\|Z_n - \mathbb{E}\, Z_n\|$ *konvergiert gegen Null, falls gilt:*

$$\lim_{n\to\infty, K\to\infty} \mathbb{E} \sum_{i=1}^n 1\Big\{\|\phi_{ni}\| > \frac{K}{n}\Big\}\|\phi_{ni}\| \;=\; 0,$$

$$\log N(\epsilon, \mathcal{T}, \hat{\rho}_n) \;=\; o_p(n) \quad \textit{für beliebige } \epsilon > 0.$$

Korollar 6.10. $\mathbb{E}\,\|Z_n - \mathbb{E}\, Z_n\|$ *konvergiert gegen Null, falls für beliebige $\delta, \epsilon > 0$ gilt:*

$$\mathbb{E} \sum_{i=1}^n \|\phi_{ni}\| \;=\; O(1),$$

$$\mathbb{E} \sum_{i=1}^n 1\big\{\|\phi_{ni}\| > \delta\big\}\|\phi_{ni}\| \;\to\; 0,$$

$$\log N(\epsilon, \mathcal{T}, \hat{\rho}_n) \;=\; O_p(1).$$

Beweis von Satz 6.8. Seien

$$\tilde{\phi}_{ni} \;:=\; 1\big\{\|\phi_{ni}\| \le \delta_n\big\}\phi_{ni} \quad \text{und} \quad \tilde{Z}_n \;:=\; \sum_{i=1}^n \tilde{\phi}_{ni}.$$

Dann ist

$$\begin{aligned}
\mathbb{E}\big\|(Z_n - \mathbb{E}\, Z_n) - (\tilde{Z}_n - \mathbb{E}\, \tilde{Z}_n)\big\| \;&\le\; 2\,\mathbb{E}\,\|Z_n - \tilde{Z}_n\| \\
&=\; 2\,\mathbb{E}\Big\|\sum_{i=1}^n 1\big\{\|\phi_{ni}\| > \delta_n\big\}\phi_{ni}\Big\| \\
&\le\; 2\,\mathbb{E} \sum_{i=1}^n 1\big\{\|\phi_{ni}\| > \delta_n\big\}\|\phi_{ni}\| \\
&\to\; 0
\end{aligned}$$

nach (6.5). Außerdem ist $\|\tilde{\phi}_{ni}\| \le \|\phi_{ni}\|$ und

$$\hat{\rho}_n(s, t \,|\, \tilde{\boldsymbol{\phi}}_n) \;\le\; \hat{\rho}_n(s, t \,|\, \boldsymbol{\phi}_n)$$

für alle $s, t \in \mathcal{T}$. Deshalb kann man ohne Einschränkung der Allgemeinheit annehmen, dass stets

$$\|\phi_{ni}\| \ \leq \ \delta_n.$$

Nach Symmetrisierungslemma 3.1 (b) ist $\mathbb{E}\,\|Z_n - \mathbb{E}\,Z_n\| \leq 2\,\mathbb{E}\,\|Z_n^o\|$. Ferner ist

$$\begin{aligned}
\mathbb{E}(\|Z_n^o\|^2) &\leq \mathbb{E}\sum_{i,j=1}^n \|\phi_{ni}\|\|\phi_{nj}\| \\
&= \sum_{i=1}^n \mathbb{E}(\|\phi_{ni}\|^2) + \sum_{i,j=1}^n 1\{i \neq j\}\,\mathbb{E}\,\|\phi_{ni}\|\,\mathbb{E}\,\|\phi_{nj}\| \\
&\leq \delta_n\,\mathbb{E}\sum_{i=1}^n \|\phi_{ni}\| + \Big(\mathbb{E}\sum_{i=1}^n \|\phi_{ni}\|\Big)^2 \\
&= O(1).
\end{aligned}$$

Für beliebige $\epsilon > 0$ ist somit

$$\mathbb{E}\,\|Z_n^o\| \ \leq \ \epsilon + \mathbb{E}\big(1\{\|Z_n^o\| > \epsilon\}\|Z_n^o\|\big) \ \leq \ \epsilon + \sqrt{\mathbb{P}(\|Z_n^o\| > \epsilon)\,\mathbb{E}(\|Z_n^o\|^2)} \ = \ \epsilon + o(1),$$

sofern wir zeigen können, dass $\mathbb{P}(\|Z_n^o\| > \epsilon)$ gegen Null konvergiert. Zu diesem Zweck sei $\hat{\mathcal{T}}_n = \hat{\mathcal{T}}_n(\boldsymbol{\phi}_n)$ eine zufällige Teilmenge von $\mathcal{T}$ mit $N(\epsilon/2, \mathcal{T}, \hat{\rho}_n)$ Elementen, so dass $\hat{\rho}_n(t, \hat{\mathcal{T}}_n) \leq \epsilon/2$ für alle $t \in \mathcal{T}$. Dann ist $\|Z_n^o\| \leq \epsilon/2 + \|Z_n^o\|_{\hat{\mathcal{T}}_n}$, wie schon am Anfang dieses Abschnitts gezeigt wurde. Aus Korollar 4.3 folgt, dass

$$\mathbb{P}\big(|Z_n^o(t)| \geq \epsilon/2 \,\big|\, \boldsymbol{\phi}_n\big) \ \leq \ 2\exp\Big(-\frac{\epsilon^2}{8\sum_{i=1}^n \phi_{ni}(t)^2}\Big) \ \leq \ 2\exp\Big(-\frac{\epsilon^2}{8\delta_n S_n}\Big),$$

wobei $S_n := \sum_{i=1}^n \|\phi_{ni}\|$ stochastisch beschränkt ist nach Voraussetzung (6.4). Folglich ist

$$\begin{aligned}
\mathbb{P}\big(\|Z_n^o\| \geq \epsilon \,\big|\, \boldsymbol{\phi}_n\big) &\leq \mathbb{P}\big(\|Z_n^o\|_{\hat{\mathcal{T}}_n} \geq \epsilon/2 \,\big|\, \boldsymbol{\phi}_n\big) \\
&\leq \sum_{t \in \hat{\mathcal{T}}_n} \mathbb{P}\big(|Z_n^o(t)| \geq \epsilon/2 \,\big|\, \boldsymbol{\phi}_n\big) \\
&\leq 2N(\epsilon/2, \mathcal{T}, \hat{\rho}_n)\exp\Big(-\frac{\epsilon^2}{8\delta_n S_n}\Big) \\
&= 2\exp\Big(-\frac{1}{\delta_n S_n}\Big(\frac{\epsilon^2}{8} - \delta_n S_n \log N(\epsilon/2, \mathcal{T}, \hat{\rho}_n)\Big)\Big) \\
&= 2\exp\Big(-\frac{1}{\delta_n S_n}\Big(\frac{\epsilon^2}{8} + o(1)\Big)\Big) \\
&\to_p 0.
\end{aligned}$$

□

6.4 Zufällige signierte Masse

Sei $\mathcal{F}$ eine Familie von messbaren Funktionen auf $\mathcal{X}$. Ähnlich wie im vorigen Kapitel nehmen wir an, dass $\mathcal{F}$ *punktweise separabel* ist. Das bedeutet, es gibt eine abzählbare Teilmenge $\mathcal{F}_o$ von $\mathcal{F}$, so dass für alle $f \in \mathcal{F}$ eine Folge $(f_k)_k$ in $\mathcal{F}_o$ existiert mit

$$\lim_{k\to\infty} f_k(x) \ = \ f(x) \quad \text{für alle } x \in \mathcal{X}.$$

Unter dieser Bedingung ist die Einhüllende

$$F(x) \ := \ \sup_{f \in \mathcal{F}} |f(x)| \ = \ \sup_{f \in \mathcal{F}_o} |f(x)|$$

von $\mathcal{F}$ ebenfalls messbar, und für jedes signierte Maß γ auf $\mathcal{X}$ mit $|\gamma|(F) < \infty$ ist

$$\|\gamma\|_{\mathcal{F}} \ = \ \|\gamma\|_{\mathcal{F}_o} \ \le \ |\gamma|(F).$$

Hier ist $|\gamma| := \gamma^+ + \gamma^-$ die Totalvariation von γ, wobei γ^+ und γ^- den Positiv- beziehungsweise Negativteil von $\gamma = \gamma^+ - \gamma^-$ bezeichnen.

Nun sei $Z_n = \sum_{i=1}^n \phi_{ni}$ mit unabhängigen stochastischen Prozessen ϕ_{ni} auf $\mathcal{F}$. Angenommen, $\phi_{ni}(f) = \int f \, d\gamma_{ni}$ mit zufälligen signierten Maßen γ_{ni} auf $\mathcal{X}$, wobei auch $|\gamma_{ni}|(F)$ messbar und fast sicher endlich sei. Der empirische Prozess $\hat{P}_n$ und der Partialsummenprozess $\hat{S}_n$ sind von dieser Form, nämlich

$$\gamma_{ni} \ = \ |\gamma_{ni}| \ = \ \frac{1}{n}\delta_{X_i} \quad \text{und} \quad |\gamma_{ni}|(F) \ = \ \frac{F(X_i)}{n},$$

beziehungsweise

$$\gamma_{ni} \ = \ c_n Y_{ni} \delta_{x_{ni}}, \quad |\gamma_{ni}| \ = \ c_n |Y_{ni}| \delta_{x_{ni}} \quad \text{und} \quad |\gamma_{ni}|(F) \ = \ c_n |Y_{ni}| F(x_{ni}).$$

Nun betrachten wir nochmals die Gesetze der großen Zahlen im vorigen Abschnitt. Da $\|\phi_{ni}\|_{\mathcal{F}} \le |\gamma_{ni}|(F)$, ist Bedingung (6.4) sicher erfüllt, falls

$$\mathbb{E}\, \hat{M}_n(F) \ = \ O(1). \tag{6.7}$$

Dabei ist $\hat{M}_n$ das zufällige Maß

$$\hat{M}_n \ := \ \sum_{i=1}^n |\gamma_{ni}|.$$

Die zweite Bedingung von Korollar 6.10 ist erfüllt, wenn

$$\mathbb{E} \sum_{i=1}^n 1\big\{|\gamma_{ni}|(F) > \delta\big\} \, |\gamma_{ni}|(F) \ \to \ 0 \quad \text{für beliebige } \delta > 0. \tag{6.8}$$

Was die dritte Bedingung von Korollar 6.10 anbelangt, so ist

$$\hat{\rho}_n(f,g) \ \le \ \int |f-g| \, d\hat{M}_n \ = \ \rho_{\hat{M}_n}(f,g).$$

Folglich ist

$$N(u, \mathcal{F}, \hat{\rho}_n) \ \le \ N(u, \mathcal{F}, \rho_{\hat{M}_n}) \ \le \ N\Big(\frac{u}{\hat{M}_n(F)}, \mathcal{F}\Big)$$

mit den uniformen Überdeckungszahlen $N(\epsilon, \mathcal{F})$ aus Abschnitt 6.2. Im Falle von (6.7) muss man also nur noch sicherstellen, dass

$$N(\epsilon, \mathcal{F}) \ < \ \infty \quad \text{für beliebige } \epsilon > 0. \tag{6.9}$$

Im Falle von $V(\mathrm{sgr}(\mathcal{F})) < \infty$ ist diese Bedingung erfüllt. Speziell für empirische Prozesse und Partialsummenprozesse ergeben sich folgende Resultate:

Satz 6.11. *Im Falle von (6.9) konvergiert* $\mathbb{E}\,\|\hat{P}_n - P\|_{\mathcal{F}}$ *gegen Null, sofern* $P(F) < \infty$.

Satz 6.12. *Sei* $\hat{S}_n(f) = n^{-1}\sum_{i=1}^n f(x_{ni})Y_{ni}$ *mit unabhängigen Zufallsvariablen* Y_{ni}*, welche zentriert und gleichgradig integrierbar sind. Das heißt,* $\mathbb{E}(Y_{ni}) = 0$*, und*

$$\max_{1\le i\le n} \mathbb{E}\big(1\{|Y_{ni}| > K\}|Y_{ni}|\big) \ \to\ 0 \quad (\min(n,K)\to\infty).$$

Im Falle von (6.9) konvergiert dann $\mathbb{E}\,\|\hat{S}_n\|_{\mathcal{F}}$ *gegen Null, sofern*

$$n^{-1}\sum_{i=1}^n F(x_{ni}) \ =\ O(1),$$

$$n^{-1}\sum_{i=1}^n 1\big\{F(x_{ni}) > n\delta\big\}F(x_{ni}) \ \to\ 0 \quad \textit{für beliebige } \delta > 0.$$

Beweis von Satz 6.11. Hier ist $\hat{M}_n = \hat{P}_n$, so dass $\mathbb{E}\,\hat{M}_n(F) = P(F)$ konstant in n und endlich ist. Dies liefert Bedingung (6.7). Ferner ist

$$\mathbb{E}\sum_{i=1}^n 1\big\{|\gamma_{ni}|(F) \ge \delta\big\}\,|\gamma_{ni}|(F) \ =\ P\big(1\{F > n\delta\}F\big),$$

und dies konvergiert für beliebige $\delta > 0$ gegen Null. Also ist auch Bedingung (6.8) erfüllt. □

Beweis von Satz 6.12. Hier ist $|\gamma_{ni}|(F) = n^{-1}F(x_{ni})|Y_{ni}|$. Daher folgt Bedingung (6.7) aus

$$\mathbb{E}\,\hat{M}_n(F) \ =\ n^{-1}\sum_{i=1}^n F(x_{ni})\,\mathbb{E}\,|Y_{ni}| \ \le\ n^{-1}\sum_{i=1}^n F(x_{ni}) \max_{1\le j\le n} \mathbb{E}\,|Y_{nj}| \ =\ O(1).$$

Ferner ist

$$\begin{aligned}
&\mathbb{E}\sum_{i=1}^n 1\big\{|\gamma_{ni}|(F) \ge \delta\big\}\,|\gamma_{ni}|(F) \\
&\le\ n^{-1}\sum_{i=1}^n F(x_{ni})\,\mathbb{E}\big(1\{|Y_{ni}| > K\}|Y_{ni}|\big) + Kn^{-1}\sum_{i=1}^n 1\big\{F(x_{ni}) > n\delta/K\big\}F(x_{ni}) \\
&=\ O(1)\cdot \max_{1\le j\le n} \mathbb{E}\big(1\{|Y_{nj}| > K\}|Y_{nj}|\big) + o(1)
\end{aligned}$$

für beliebige feste $K > 0$, und der limes superior der rechten Seite wird nach Voraussetzung beliebig klein, wenn man K hinreichend groß wählt. Hieraus ergibt sich Bedingung (6.8). □

Beispiel 6.13 (Familien $\mathcal{F}$ für Kerndichte-, M- und S-Schätzer). Sei $H : \mathbb{R} \to \mathbb{R}$ monoton wachsend oder monoton fallend. Mit $\mathcal{X} := \mathbb{R}^d$ betrachten wir

$$\begin{aligned}
\mathcal{F}_1 \ &:=\ \Big\{H(b|\cdot - \theta|^2) : b > 0, \theta \in \mathbb{R}^d\Big\},\\
\mathcal{F}_2 \ &:=\ \Big\{x \mapsto H((x-\theta)^\top B(x-\theta)) : B \in \mathbb{R}^{d\times d}, \theta \in \mathbb{R}^d\Big\}.
\end{aligned}$$

Dann ist $\mathrm{sgr}(\mathcal{F}_j)$ eine VC-Klasse mit Index $V(\mathrm{sgr}(\mathcal{F}_1)) \leq d + 3$ und $V(\mathrm{sgr}(\mathcal{F}_2)) \leq d^2/2 + 3d/2 + 2$. Dies kann man besonders leicht einsehen, wenn H eine isotone Bijektion von $\mathbb{R}$ ist. Dann ist nämlich

$$\begin{aligned} \mathrm{sgr}\Big(H(b|\cdot - \theta|^2)\Big) &= \Big\{(x,r) \in \mathbb{R}^d \times \mathbb{R} : b|\theta|^2 - 2b\theta^\top x + bx^\top x - H^{-1}(r) \geq 0\Big\} \\ &= \{g \geq 0\} \end{aligned}$$

mit einer Funktion g in dem Vektorraum $\mathcal{G}$, welcher von den $d + 3$ Funktionen

$$\begin{aligned} v_0(x,r) &:= 1, \\ v_i(x,r) &:= x_i \quad \text{für } 1 \leq i \leq d, \\ v_{d+1}(x,r) &:= H^{-1}(r), \\ v_{d+2}(x,r) &:= x^\top x \end{aligned}$$

aufgespannt wird. Analog ist

$$\mathrm{sgr}\Big(x \mapsto H((x-\theta)^\top B(x-\theta))\Big) = \{g \geq 0\}$$

mit einer Funktion g aus dem Vektorraum $\mathcal{G}$, der von obigen Funktionen $v_1, \ldots, v_{d+1}$ sowie den Funktionen

$$v_{ij}(x,r) := x_i x_j \quad (1 \leq i \leq j \leq d)$$

aufgespannt wird. Nun folgt die Behauptung aus Satz 5.13.

Im allgemeinen Fall muss man einen geeigneten Ersatz für die Funktion H^{-1} finden. Sei E eine endliche Teilmenge von $\mathbb{R}^d \times \mathbb{R}$, welche von $\mathrm{sgr}(\mathcal{F}_1)$ vollständig zerlegt wird. Zu $A \subset E$ existieren also $b_A > 0$ und $\theta_A \in \mathbb{R}^d$ mit

$$A = E \cap \mathrm{sgr}\Big(H(b_A|\cdot - \theta_A|^2)\Big).$$

Nun definieren wir die endliche Menge $T := \Big\{b_A|x - \theta_A|^2 : (x,r) \in E, A \subset E\Big\}$. Dann ist

$$\begin{aligned} E \cap \mathrm{sgr}\Big(H(b_A|\cdot - \theta_A|^2)\Big) &= \Big\{(x,r) \in E : r \leq H(b_A|x - \theta_A|^2)\Big\} \\ &= \Big\{(x,r) \in E : b_A\theta_A^\top\theta_A - 2b_A\theta_A^\top x + b_A x^\top x - \overline{H}(r) \geq 0\Big\} \\ &= \Big\{(x,r) \in E : g(x,r) \geq 0\Big\} \end{aligned}$$

mit einer Funktion $g \in \mathcal{G} := \mathrm{span}\{v_0, v_1, v_2, \ldots, v_{d+2}\}$, wobei nun

$$v_{d+2}(x,r) := \overline{H}(r) := \min\Big\{t \in T : H(t) \geq r\Big\}.$$

Analog lässt sich die Familie $\mathcal{F}_2$ behandeln.

Reduktion auf Mengenfamilien. Mitunter kann man sich komplizierte Abschätzungen von Überdeckungszahlen mit folgendem Trick ersparen: Sei $\mathcal{F}$ eine Familie von messbaren Funktionen auf $\mathcal{X}$ mit beschränktem Wertebereich $[0, c]$. Dann gilt für jedes endliche signierte Maß γ auf $\mathcal{X}$ die Ungleichung

$$\|\gamma\|_{\mathcal{F}} \ \le\ c\|\gamma\|_{\mathcal{D}(\mathcal{F})}$$

mit der Mengenfamilie

$$\mathcal{D}(\mathcal{F}) \ :=\ \bigl\{\{f \ge t\} : f \in \mathcal{F}, t \in \mathbb{R}\bigr\}.$$

Denn für $f \in \mathcal{F}$ ist

$$\Bigl|\int f\, d\gamma\Bigr| \ =\ \Bigl|\int_0^c \gamma(\{f \ge t\})\, dt\Bigr| \ \le\ \int_0^c \bigl|\gamma(\{f \ge t\})\bigr|\, dt \ \le\ c\|\gamma\|_{\mathcal{D}}.$$

Beispiel (Isotone Funktionen) Sei $\mathcal{F}$ die Menge aller monoton wachsenden Funktionen $f : \mathbb{R} \to [0, 1]$. Dann besteht $\mathcal{D}(\mathcal{F})$ aus allen offenen und abgeschlossenen Halbgeraden (r, ∞) bzw. $[r, \infty)$ sowie $\emptyset$ und $\mathbb{R}$. Folglich ist

$$\|\gamma\|_{\mathcal{F}} \ \le\ \sup_{r \in \mathbb{R}} \bigl|\gamma((r, \infty))\bigr|.$$

Symmetrische konvexe Hüllen. Ein anderer oft hilfreicher Trick ist die Darstellung einer Funktionenfamilie als symmetrische konvexe Hülle einer einfacheren Funktionenklasse. Genauer gesagt, sei $\mathcal{F}$ eine Funktionenfamilie mit messbarer Einhüllender F, und sei $\mathcal{G}(\mathcal{F})$ die Menge aller punktweisen Limites von Funktionenfolgen in

$$\Bigl\{\sum_{i=1}^{\infty} \lambda_i f_i \ :\ f_i \in \mathcal{F}, \lambda_i \in \mathbb{R}, \sum_{i=1}^{\infty} |\lambda_i| \le 1\Bigr\}.$$

Ist nun γ ein signiertes Maß auf $\mathcal{X}$ mit $|\gamma|(F) < \infty$, dann ist

$$\|\gamma\|_{\mathcal{G}(\mathcal{F})} \ =\ \|\gamma\|_{\mathcal{F}}.$$

Beispiel (Funktionen mit beschränkter Totalvariation) Sei $\mathcal{F}$ die Familie aller Indikatorfunktionen von Halbgeraden $[r, \infty)$, $r \in \mathbb{R}$. Dann ist $\mathcal{G}(\mathcal{F})$ die Menge aller Funktionen $g : \mathbb{R} \to \mathbb{R}$ mit Grenzwert $g(-\infty) = 0$ und

$$\mathrm{TV}(g) := \sup\Bigl\{\sum_{i=1}^{m} \bigl|g(x_i) - g(x_{i-1})\bigr| : m \in \mathbb{N}, x_0 < x_1 < \cdots < x_m\Bigr\} \ \le\ 1.$$

Dies ergibt sich im Wesentlichen aus Aufgabe 6.6.

6.5 Verfeinerungen

Das Gesetz der Großen Zahlen, Satz 6.8, kann man noch verfeinern, sofern Bedingung (6.5) durch eine stärkere Bedingung ersetzt wird. Im Beweis von Satz 6.8 tauchte in einer entscheidenden Ungleichung die Größe

$$\exp\Bigl(-\frac{\epsilon^2}{8\sum_{i=1}^n \phi_{ni}(t)^2}\Bigr)$$

auf, und die Summe wurde durch $\max_i \|\phi_{ni}\| \sum_j \|\phi_{nj}\|$ nach oben abgeschätzt. Der entscheidende Schritt ist nun, diese grobe Abschätzung zu verbessern.

Lemma 6.14. *Angenommen $\|\phi_{ni}\| \le \delta_n$ für alle i. Ferner sei*

$$V_n \ := \ \sup_{t\in\mathcal{T}} \mathbb{E} \sum_{i=1}^n \phi_{ni}(t)^2.$$

Dann ist

$$\mathbb{P}\Big(\sup_{t\in\mathcal{T}} \sum_{i=1}^n \phi_{ni}(t)^2 \ge 8\tau\Big) \ \le \ \frac{2}{1 - V_n/\tau} \mathbb{E}\min\Big(N\Big(\frac{\tau}{16\delta_n}, \mathcal{T}, \hat\rho_n\Big) \exp\Big(-\frac{\tau}{8\delta_n^2}\Big), 1\Big)$$

für beliebige $\tau > V_n$.

Satz 6.15. *Unter den Voraussetzungen von Lemma 6.14 ist*

$$\begin{aligned}\mathbb{P}\big(\|Z_n - \mathbb{E}\, Z_n\| \ge 3\eta\big) \quad &\le \quad 16\, \mathbb{E}\min\Big(N\Big(\frac{\tau}{16\delta_n}, \mathcal{T}, \hat\rho_n\Big) \exp\Big(-\frac{\tau}{8\delta_n^2}\Big), 1\Big) \\ &\quad + 8\, \mathbb{E}\min\Big(N\Big(\frac{\eta}{2}, \mathcal{T}, \hat\rho_n\Big) \exp\Big(-\frac{\eta^2}{64\tau}\Big), 1\Big)\end{aligned}$$

für beliebige $\eta \ge \sqrt{2V_n}$ und $\tau \ge 2V_n$.

Korollar 6.16. *Die Voraussetzungen von Lemma 6.14 seien erfüllt mit $\delta_n := 1/n$, und für beliebige $n \in \mathbb{N}$, $\epsilon \in (0,1]$ sei*

$$N(\epsilon, \mathcal{T}, \hat\rho_n) \ \le \ A\epsilon^{-B} \quad \textit{fast sicher}$$

mit Konstanten $A, B > 0$. Für hinreichend großes $\kappa > 0$ ist dann

$$\sum_{n=1}^\infty \mathbb{P}\Big(\|Z_n - \mathbb{E}\, Z_n\| \ge \kappa \max\Big(\sqrt{\log(n) V_n}, \frac{\log n}{n}\Big)\Big) \ < \ \infty.$$

Beweis von Lemma 6.14. Die Grundidee des Beweises ist LeCams “square root trick”. Seien

$$\begin{aligned} I \quad &:= \quad \big\{i \in \{1,2,\ldots,n\} : \xi_i = 1\big\}, \\ J \quad &:= \quad \{1,2,\ldots,n\} \setminus I, \\ S_K \quad &:= \quad \sum_{i\in K} \phi_{ni}^2 \quad \text{für } K \subset \{1,2,\ldots,n\}, \\ S \quad &:= \quad S_{\{1,2,\ldots,n\}} \ = \ S_I + S_J. \end{aligned}$$

Dann sind S_I, S_J identisch verteilt, weshalb

$$\mathbb{P}\big(\|S\| \ge 8\tau\big) \ \le \ 2\,\mathbb{P}\big(\|S_I\| \ge 4\tau\big) \ = \ 2\,\mathbb{P}\Big(\big\|\sqrt{S_I}\big\| \ge 2\sqrt{\tau}\Big).$$

Doch für beliebige $t \in \mathcal{T}$ ist

$$\mathbb{P}\big(\sqrt{S_I(t)} \ge \sqrt{\tau} \,\big|\, \xi_1, \ldots, \xi_n\big) \ \le \ \frac{\mathbb{E}\big(S_I(t) \,\big|\, \xi_1, \ldots, \xi_n\big)}{\tau} \ \le \ \frac{\mathbb{E}\, S(t)}{\tau} \ \le \ \frac{V_n}{\tau},$$

und $\sqrt{S_I}, \sqrt{S_J}$ sind stochastisch unabhängig, gegeben $\xi_1, \ldots, \xi_n$. Wendet man Symmetrisierungslemma 3.1 (a) auf diese bedingte Verteilung an, so folgt, dass

$$\mathbb{P}\Big(\|\sqrt{S_I}\| \geq 2\sqrt{\tau}\Big) \;\leq\; \frac{1}{1 - V/\tau}\,\mathbb{P}\Big(\|\sqrt{S_I} - \sqrt{S_J}\| \geq \sqrt{\tau}\Big).$$

Für beliebige $s, t \in \mathcal{T}$ ist

$$\begin{aligned}
&\Big|\big(\sqrt{S_I(s)} - \sqrt{S_J(s)}\big) - \big(\sqrt{S_I(t)} - \sqrt{S_J(t)}\big)\Big| \\
&\leq\; \big|\sqrt{S_I(s)} - \sqrt{S_I(t)}\big| + \big|\sqrt{S_J(s)} - \sqrt{S_J(t)}\big| \\
&\leq\; \Big(\sum_{i \in I}\big(\phi_{ni}(s) - \phi_{ni}(t)\big)^2\Big)^{1/2} + \Big(\sum_{i \in J}\big(\phi_{ni}(s) - \phi_{ni}(t)\big)^2\Big)^{1/2} \\
&\leq\; \Big(2\sum_{i=1}^{n}\big(\phi_{ni}(s) - \phi_{ni}(t)\big)^2\Big)^{1/2} \\
&\leq\; \sqrt{4\delta_n \hat{\rho}_n(s,t)}.
\end{aligned}$$

Ist also $\hat{\mathcal{T}}_n = \hat{\mathcal{T}}_n(\phi_{n1}, \ldots, \phi_{nn})$ eine Teilmenge von $\mathcal{T}$ mit

$$\#\hat{\mathcal{T}}_n \;=\; N\Big(\frac{\tau}{16\delta_n}, \mathcal{T}, \hat{\rho}_n\Big) \quad \text{und} \quad \sup_{t \in \mathcal{T}} \hat{\rho}_n(t, \hat{\mathcal{T}}_n) \;\leq\; \frac{\tau}{16\delta_n},$$

so ist

$$\begin{aligned}
&\mathbb{P}\Big(\|\sqrt{S_I} - \sqrt{S_J}\| \geq \sqrt{\tau}\,\Big|\,\phi_{n1}, \ldots, \phi_{nn}\Big) \\
&\leq\; \mathbb{P}\Big(\|\sqrt{S_I} - \sqrt{S_J}\|_{\hat{\mathcal{T}}_n} \geq \frac{\sqrt{\tau}}{2}\,\Big|\,\phi_{n1}, \ldots, \phi_{nn}\Big) \\
&\leq\; \sum_{t \in \hat{\mathcal{T}}_n} \mathbb{P}\Big(\Big|\sqrt{S_I(t)} - \sqrt{S_J(t)}\Big| \geq \frac{\sqrt{\tau}}{2}\,\Big|\,\phi_{n1}, \ldots, \phi_{nn}\Big) \\
&\leq\; \sum_{t \in \hat{\mathcal{T}}_n} \mathbb{P}\Big(\Big|S_I(t) - S_J(t)\Big| \geq \frac{(\sqrt{S_I(t)} + \sqrt{S_J(t)})\sqrt{\tau}}{2}\,\Big|\,\phi_{n1}, \ldots, \phi_{nn}\Big) \\
&\leq\; \sum_{t \in \hat{\mathcal{T}}_n} \mathbb{P}\Big(\Big|S_I(t) - S_J(t)\Big| \geq \frac{\sqrt{S(t)\tau}}{2}\,\Big|\,\phi_{n1}, \ldots, \phi_{nn}\Big) \\
&=\; \sum_{t \in \hat{\mathcal{T}}_n} \mathbb{P}\Big(\Big|\sum_{i=1}^{n}\xi_i \phi_{ni}(t)^2\Big| \geq \frac{\sqrt{S(t)\tau}}{2}\,\Big|\,\phi_{n1}, \ldots, \phi_{nn}\Big) \\
&\leq\; 2\sum_{t \in \hat{\mathcal{T}}_n} \exp\Big(-\frac{S(t)\tau}{8\sum_{i=1}^{n}\phi_{ni}(t)^4}\Big) \\
&\leq\; 2\,\#\hat{\mathcal{T}}_n \exp\Big(-\frac{\tau}{8\delta_n^2}\Big).
\end{aligned}$$

□

Beweis von Satz 6.15. Da $\mathrm{Var}[Z_n(t)] \leq V_n$ für beliebige $t \in \mathcal{T}$, folgt aus Symmetrisierungslemma 3.1 (a), dass

$$\begin{aligned}
\mathbb{P}\big(\|Z_n - \mathbb{E}\,Z_n\| \geq 3\eta\big) \;&\leq\; \frac{1}{1 - V_n/\eta^2}\,\mathbb{P}\big(\|Z_n - Z_n'\| \geq 2\eta\big) \\
&\leq\; \frac{2}{1 - V_n/\eta^2}\,\mathbb{P}\big(\|Z_n^o\| \geq \eta\big). \qquad (6.10)
\end{aligned}$$

Doch mit $X_n := \sup_{t\in\mathcal{T}} \sum_{i=1}^n \phi_{ni}(t)^2$ kann man wie im Beweis von Satz 6.8 zeigen, dass

$$\mathbb{P}\Bigl(\|Z_n^o\| \ge \eta \,\Big|\, \phi_{n1},\ldots,\phi_{nn}\Bigr) \;\le\; 2N\Bigl(\frac{\eta}{2},\mathcal{T},\hat\rho_n\Bigr)\exp\Bigl(-\frac{\eta^2}{8X_n}\Bigr).$$

Zusammen mit Lemma 6.14 ergibt dies die Ungleichungen

$$\begin{aligned}
\mathbb{P}\bigl(\|Z_n^o\| \ge \eta\bigr) \;&\le\; \mathbb{P}(X_n \ge 8\tau) \\
&\quad + 2\,\mathbb{E}\min\Bigl(N\Bigl(\frac{\eta}{2},\mathcal{T},\hat\rho_n\Bigr)\exp\Bigl(-\frac{\eta^2}{64\tau}\Bigr),1\Bigr) \\
&\le\; \frac{4}{1-V_n/\tau}\,\mathbb{E}\min\Bigl(N\Bigl(\frac{\tau}{16\delta_n},\mathcal{T},\hat\rho_n\Bigr)\exp\Bigl(-\frac{\tau}{8\delta_n^2}\Bigr),1\Bigr) \\
&\quad + 2\,\mathbb{E}\min\Bigl(N\Bigl(\frac{\eta}{2},\mathcal{T},\hat\rho_n\Bigr)\exp\Bigl(-\frac{\eta^2}{64\tau}\Bigr),1\Bigr).
\end{aligned}$$

Setzt man dies in (6.10) ein und setzt voraus, dass $\min(\eta^2,\tau) \ge 2V_n$, ergibt sich die Behauptung. □

Beweis von Korollar 6.16. Die Behauptung folgt aus Satz 6.15 vermöge

$$\begin{aligned}
\tau = \tau_n \;&:=\; C_\tau \max\Bigl(V_n, \frac{\log n}{n^2}\Bigr), \\
\eta = \eta_n \;&:=\; \sqrt{C_\eta \log(n)\tau_n} \;=\; \sqrt{C_\tau C_\eta}\,\max\Bigl(\sqrt{\log(n)V_n}, \frac{\log n}{n}\Bigr)
\end{aligned}$$

mit hinreichend großen Konstanten $C_\tau, C_\eta > 0$. Denn für hinreichend großes n ist $\eta_n^2 \wedge \tau_n \ge 2V_n$, und

$$\begin{aligned}
\log N\Bigl(\frac{n\tau_n}{16},\mathcal{T},\hat\rho_n\Bigr) - \frac{n^2\tau_n}{8} \;&\le\; \text{const.} - B\log\Bigl(\frac{\log n}{n}\Bigr) - \frac{C_\tau \log n}{8} \\
&\le\; \text{const.} + \Bigl(B - \frac{C_\tau}{8}\Bigr)\log n, \\
\log N\Bigl(\frac{\eta_n}{2},\mathcal{T},\hat\rho_n\Bigr) - \frac{\eta_n^2}{64\tau_n} \;&\le\; \text{const.} - B\log\Bigl(\frac{\log n}{n}\Bigr) - \frac{C_\eta \log n}{64} \\
&\le\; \text{const.} + \Bigl(B - \frac{C_\eta}{64}\Bigr)\log n.
\end{aligned}$$

□

6.6 Anwendung auf Dichteschätzung

Sei P ein Wahrscheinlichkeitmaß auf $\mathbb{R}^d$ mit Dichte p bezüglich des Lebesgue-Maßes. Ein *Kernschätzer* für $p(x)$ ist

$$\hat p_n(x) \;:=\; \frac{1}{n}\sum_{i=1}^n k_n(x - X_i) \;=\; \int k_n(x-y)\hat P_n(dy),$$

wobei k_n eine willkürlich gewählte Funktion ist, so dass $\int k_n(x)\,dx = 1$. Mit anderen Worten, $\hat p_n$ ist die Dichte der Faltung von $\hat P_n$ mit dem signierten Maß $k_n(x)\,dx$. Mit

$$\overline{p}_n(x) \;:=\; \mathbb{E}\,\hat p_n(x)$$

ist

$$\hat{p}_n - p \;=\; (\overline{p}_n - p) + (\hat{p}_n - \overline{p}_n) \;=:\; \text{“Bias”} + \text{“Fluktuation”}.$$

Speziell sei

$$k_n(x) \;:=\; \frac{1}{h_n^d} k\Big(\frac{x}{h_n}\Big)$$

mit einer *Bandbreite* $h_n > 0$. Je kleiner die Bandbreite h_n, desto kleiner wird im allgemeinen der Bias sein, aber desto größer ist die Fluktuation.

Abschätzung des Bias. Angenommen, p ist Lipschitz-stetig, also

(6.11) $$|p(x) - p(y)| \;\le\; L|x-y| \quad \text{für beliebige } x, y \in \mathbb{R}^d$$

mit einer Konstanten $L < \infty$. Ferner sei

(6.12) $$\|k\| \;\le\; 1, \qquad \int k(z)\,dz \;=\; 1 \quad \text{und} \quad \int |z||k(z)|\,dz \;<\; \infty,$$

wobei $\|\cdot\|$ die Supremumsnorm über $\mathcal{T} := \mathbb{R}^d$ bezeichnet. Dann ist

$$\overline{p}_n(x) - p(x) \;=\; \frac{1}{h_n^d} \int k\Big(\frac{x-y}{h_n}\Big)\Big(p(y) - p(x)\Big)\,dy \;=\; \int k(z)\big(p(x - h_n z) - p(x)\big)\,dy,$$

so dass

$$\|\overline{p}_n - p\| \;\le\; L \int |x||k(x)|\,dx\; h_n \;=\; O(h_n).$$

Abschätzung der Fluktuation. Angenommen,

(6.13) $$\big\{\mathrm{sgr}(k(\cdot - t)) : t \in \mathbb{R}^d\big\} \text{ ist eine VC-Klasse;}$$

siehe Beispiel 6.13. Nun definieren wir

$$\phi_{ni}(x) \;:=\; \frac{1}{n} k\Big(\frac{x - X_i}{h_n}\Big).$$

Dann ist $\hat{p}_n = h_n^{-d} Z_n$, $\|\phi_{ni}\| \le 1/n$, und

$$\begin{aligned}
\mathbb{E} \sum_{i=1}^n \phi_{ni}(x)^2 \;&=\; \frac{1}{n}\,\mathbb{E}\, k\Big(\frac{x - X_1}{h_n}\Big)^2 \\
&=\; \frac{1}{n} \int k\Big(\frac{x-y}{h_n}\Big)^2 p(y)\,dy \\
&\le\; \int |k(z)|\,dz\,\|p\|\,\frac{h_n^d}{n} \\
&=\; O\Big(\frac{h_n^d}{n}\Big).
\end{aligned}$$

Man beachte, dass aus (6.11) die Endlichkeit von $\|p\|$ folgt. Gemäß Satz 6.3 sind die Bedingungen von Korollar 6.16 erfüllt, und zwar ist

$$\|\hat{p}_n - \overline{p}_n\| \;=\; \frac{1}{h_n^d} O\Big(\max\Big(\sqrt{h_n^d \frac{\log n}{n}}, \frac{\log n}{n}\Big)\Big) \quad \text{fast sicher.}$$

Dabei verwendeten wir das Borel-Cantelli-Lemma.

Nun versuchen wir Bias und Fluktuation zu balancieren. Die Gleichung

$$h_n \;=\; \frac{1}{h_n^d}\sqrt{h_n^d\frac{\log n}{n}}$$

impliziert, dass $h_n = \big(\log(n)/n\big)^{1/(d+2)}$, und es ergibt sich folgendes Resultat:

Korollar 6.17. *Unter den Bedingungen (6.11), (6.12) und (6.13) ist*

$$\|\hat{p}_n - p\| \;=\; O\Big(\Big(\frac{\log n}{n}\Big)^{1/(d+2)}\Big) \quad \textit{fast sicher},$$

falls

$$h_n \;=\; c_n\Big(\frac{\log n}{n}\Big)^{1/(d+2)}$$

mit einer Folge $(c_n)_n$ in $(0,\infty)$, die von 0 und ∞ weg beschränkt ist.

6.7 Aufgaben

Aufgabe 6.1. Beweisen Sie Lemma 6.6.

Aufgabe 6.2. Beweisen Sie Lemma 6.7.

Aufgabe 6.3. Sei $\mathcal{F}$ die Menge aller Funktionen f auf $[0,1]$, so dass $|f| \le 1$ und $|f(x)-f(y)| \le |x-y|$ für beliebige $x, y \in [0,1]$, und sei $\rho(f,g) := \|f-g\|_{[0,1]}$. Zeigen Sie, dass

$$\log N(\epsilon, \mathcal{F}, \rho) \;\le\; A\epsilon^{-1} \quad \text{für } 0 < \epsilon \le 1$$

mit einer Konstanten $A > 0$.

Zusatz: Für $\beta > 0$ sei $\mathcal{F}_{(\beta)}$ die Menge aller k-mal stetig differenzierbaren Funktionen f auf $[0,1]$, wobei $k = k(\beta) := \lceil\beta\rceil - 1$, so dass gilt:

$$\begin{aligned} |f^{(\ell)}| \;&\le\; 1 \quad \text{für } 0 \le \ell \le k, \\ |f^{(k)}(x) - f^{(k)}(y)| \;&\le\; |x-y|^{\beta-k} \quad \text{für beliebige } x, y \in [0,1]. \end{aligned}$$

Dabei ist $f^{(\ell)}$ die ℓ-te Ableitung von f mit $f^{(0)} := f$. Zeigen Sie, dass

$$\log N(\epsilon, \mathcal{F}_{(\beta)}, \rho) \;\le\; A\epsilon^{-1/\beta} \quad \text{für } 0 < \epsilon \le 1$$

mit einer Konstanten $A > 0$.

Aufgabe 6.4. Sei $\mathcal{D}$ eine Familie von Teilmengen von $\mathcal{X}$, und sei $\tilde{\mathcal{D}}$ die Menge der entsprechenden Indikatorfunktionen. Zeigen Sie, dass $V(\mathcal{D}) = V(\operatorname{sgr}(\tilde{\mathcal{D}}))$.

Aufgabe 6.5. Sei $\mathcal{F}$ die Familie aller monoton wachsenden Funktionen $f : \mathbb{R} \to [0,1]$. Bestimmen Sie eine Konstante $C > 0$ derart, dass

$$\log N(\epsilon, \mathcal{F}) \;\le\; C\epsilon^{-1} \quad \text{für } 0 < \epsilon \le 1.$$

Leiten Sie ein analoges Ergebnis für die Menge $\mathcal{F}_R$ aller Funktionen $f : \mathbb{R} \to [0, R]$ mit Totalvariation $\mathrm{TV}(f, \mathbb{R}) \le R$ her, siehe auch nachfolgende Aufgabe 6.6.

Sei $F : \mathbb{R} \to [0, \infty)$ messbar, und sei P ein Wahrscheinlichkeitsmaß mit $P(F^\gamma) < \infty$ für ein $\gamma > 1$. Bestimmen Sie eine sinnvolle Schranke für $\log N(\epsilon, \mathcal{F}, \rho_P)$, $0 < \epsilon \le 1$, wobei $\mathcal{F}$ die Menge aller monoton wachsenden Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit $|f| \le F$ bezeichnet.

Aufgabe 6.6 (Totalvariation von Funktionen). Für ein Intervall $J \subset \mathbb{R}$ und eine Funktion $f : J \to \mathbb{R}$ definieren wir

$$\begin{aligned} \mathrm{TV}(f, J) \ &:= \ \sup\Bigl\{ \sum_{i=1}^m \bigl| f(x_i) - f(x_{i-1}) \bigr| : \\ &\qquad\qquad m \in \mathbb{N};\ x_0, x_1, \ldots, x_m \in J;\ x_0 < x_1 < \cdots < x_m \Bigr\}, \\ \mathrm{TV}^{(\pm)}(f, J) \ &:= \ \sup\Bigl\{ \sum_{i=1}^m \bigl(f(x_i) - f(x_{i-1}) \bigr)^{\pm} : \\ &\qquad\qquad m \in \mathbb{N};\ x_0, x_1, \ldots, x_m \in J;\ x_0 < x_1 < \cdots < x_m \Bigr\}. \end{aligned}$$

(a) Zeigen Sie, dass

$$\mathrm{TV}(f, J) \ = \ \mathrm{TV}^{(+)}(f, J) + \mathrm{TV}^{(-)}(f, J).$$

(b) Zeigen Sie, dass

$$T(f, J) \ = \ T(f, J \cap (-\infty, b]) + T(f, J \cap [b, \infty)) \quad \text{für beliebige } b \in \mathbb{R},$$

wobei $T(f, \cdot)$ für $\mathrm{TV}(f, \cdot)$ oder $\mathrm{TV}^{(\pm)}(f, \cdot)$ steht, und $T(f, \emptyset) := 0$.

(c) Sei $f : \mathbb{R} \to \mathbb{R}$ mit $\mathrm{TV}(f, \mathbb{R}) < \infty$. Zeigen Sie, dass Grenzwerte $f(\pm\infty)$ existieren. Zeigen Sie ferner, dass

$$\lim_{b \to -\infty} T\bigl(f, (-\infty, b]\bigr) \ = \ \lim_{b \to \infty} T\bigl(f, [b, \infty)\bigr) \ = \ 0.$$

(d) Sei $f : \mathbb{R} \to \mathbb{R}$ mit Grenzwert $f(-\infty) = 0$ und $\mathrm{TV}(f, \mathbb{R}) < \infty$. Zeigen Sie, dass

$$f(x) \ = \ \mathrm{TV}^{(+)}\bigl(f, (-\infty, x]\bigr) - \mathrm{TV}^{(-)}\bigl(f, (-\infty, x]\bigr) \quad \text{für alle } x \in \mathbb{R}.$$

Das heißt, f lässt als Differenz zweier monoton wachsender Funktionen f_1, f_2 darstellen, und deren Grenzwerte erfüllen die Gleichungen $f_1(-\infty) = f_2(-\infty) = 0$ sowie $f_1(\infty) + f_2(\infty) = \mathrm{TV}(f, \mathbb{R})$.

Aufgabe 6.7. Angenommen, der Kern k in Abschnitt 6.6 erfüllt folgende Bedingungen:

$$\|k\| \ \le \ 1, \qquad \int k(z)\, dz = 1, \qquad \int z k(z)\, dz \ = \ 0 \quad \text{und} \quad \int |z|^2 |k(z)|\, dz \ < \ \infty.$$

Welche Konvergenzrate kann man für $\|\hat{p}_n - p\|$ erzielen, wenn man voraussetzt, dass p differenzierbar ist mit Lipschitz-stetiger Ableitung $\nabla p : \mathbb{R}^d \to \mathbb{R}^d$? Wie muss man die Bandweite wählen, um diese zu erreichen?

Kapitel 7

Konvergenz in Verteilung

In diesem Kapitel behandeln wir die allgemeine Theorie der Verteilungskonvergenz beziehungsweise der schwachen Konvergenz. Im Hinblick auf empirische Prozesse betrachten wir von Anfang an auch nicht-messbare Funktionen auf Wahrscheinlichkeitsräumen. Dieser Zugang geht auf Jørgen Hoffmann-Jørgensen zurück und wurde unter anderem von Richard Dudley weiterentwickelt.

7.1 Äussere Erwartungswerte

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein beliebiger Wahrscheinlichkeitsraum, und im Folgenden seien X, Y, Z generische beschränkte Funktionen von Ω nach $\mathbb{R}$. Der *äußere Erwartungswert* von Y ist definiert als

$$\mathbb{E}^*(Y) \ := \ \inf\bigl\{\mathbb{E}(X) : X \text{ messbar}, X \geq Y\bigr\}.$$

Dies ist eine Verallgemeinerung der *äußeren Wahrscheinlichkeit*

$$\mathbb{P}^*(B) \ := \ \inf\bigl\{\mathbb{P}(A) : A \in \mathcal{A}, A \supset B\bigr\}$$

einer Menge $B \subset \Omega$, da $\mathbb{P}^*(B) = \mathbb{E}^*(1_B)$. Denn offensichtlich ist $\mathbb{E}^*(1_B) \leq \mathbb{P}^*(B)$, und die umgekehrte Ungleichung folgt aus der Tatsache, dass für messbare Funktionen X mit $X \geq 1_B$ stets $X \geq 1\{X \geq 1\}$, und $B \subset \{X \geq 1\} \in \mathcal{A}$.

Analog definiert man den *inneren Erwartungswert* von Y als

$$\mathbb{E}_*(Y) \ := \ \sup\bigl\{\mathbb{E}(X) : X \text{ messbar}, X \leq Y\bigr\} \ = \ -\,\mathbb{E}^*(-Y),$$

und die *innere Wahrscheinlichkeit* von B ist gleich

$$\mathbb{P}_*(B) \ := \ \sup\bigl\{\mathbb{P}(A) : A \in \mathcal{A}, A \subset B\bigr\} \ = \ \begin{cases} 1 - \mathbb{P}^*(\Omega \setminus B), \\ \quad \mathbb{E}_*(1_B). \end{cases}$$

Die folgenden Eigenschaften der äußeren Erwartung kann man leicht nachweisen.

Lemma 7.1. *Das Funktional $Y \mapsto \mathbb{E}^*(Y)$ ist sublinear, das heißt,*

$$\begin{aligned} \mathbb{E}^*(rY) &= r\,\mathbb{E}^*(Y) \quad \textit{für beliebige } r \geq 0, \\ \mathbb{E}^*(Y+Z) &\leq \mathbb{E}^*(Y) + \mathbb{E}^*(Z). \end{aligned}$$

Insbesondere ist $\mathbb{E}^*(Y) + \mathbb{E}^*(-Y) \geq 0$ mit Gleichheit, falls Y messbar ist. Ferner ist

$$\begin{aligned} \mathbb{E}^*(Y) &\leq \mathbb{E}^*(Z) \quad \textit{falls } Y \leq Z, \\ \mathbb{E}^*(r+Y) &= r + \mathbb{E}^*(Y) \quad \textit{für beliebige } r \in \mathbb{R}. \end{aligned}$$

□

Lemma 7.2. *Im Falle $Y \geq 0$ ist*

$$\mathbb{E}^*(Y) \;=\; \int_0^\infty \mathbb{P}^*(Y \geq r)\,dr \;=\; \int_0^\infty \mathbb{P}^*(Y > r)\,dr.$$

Beweis. Sei $X : \Omega \to \mathbb{R}$ messbar mit $X \geq Y$. Für jede Zahl $\delta > 0$ ist

$$Y \;\leq\; \delta \sum_{i=0}^\infty 1\{Y > \delta i\} \quad \text{und} \quad X \;\geq\; \delta \sum_{i=1}^\infty 1\{X \geq \delta i\}.$$

Folglich ist

$$\mathbb{E}^*(Y) \;\leq\; \delta \sum_{i=0}^\infty \mathbb{P}^*(Y > \delta i) \;\leq\; \delta \sum_{i=1}^\infty \mathbb{P}^*(Y > \delta i) + \delta \;\leq\; \int_0^\infty \mathbb{P}^*(Y > r)\,dr + \delta$$

und

$$\mathbb{E}(X) \;\geq\; \delta \sum_{i=1}^\infty \mathbb{P}(X \geq \delta i) \;\geq\; \delta \sum_{i=1}^\infty \mathbb{P}^*(Y \geq \delta i) \;\geq\; \int_0^\infty \mathbb{P}^*(Y \geq r)\,dr - \delta.$$

Für $\delta \downarrow 0$ und $\mathbb{E}(X) \downarrow \mathbb{E}^*(Y)$ folgt die Behauptung. □

7.2 Definition und Eigenschaften der Verteilungskonvergenz

Sei $\mathbb{M}$ eine Menge, versehen mit einer Metrik d und der entsprechenden Borel-σ-Algebra $\mathcal{B}(\mathbb{M})$. Zunächst führen wir einige Bezeichnungen ein. Es seien

$$\begin{aligned} \mathcal{C}(\mathbb{M}) &:= \big\{\text{stetige Funktionen auf } \mathbb{M}\big\}, \\ \mathcal{C}_{\mathrm{b}}(\mathbb{M}) &:= \big\{\text{beschränkte, stetige Funktionen auf } \mathbb{M}\big\}, \\ \mathcal{C}_{\mathrm{Lip}}(\mathbb{M}) &:= \big\{\text{Lipschitz-stetige Funktionen auf } \mathbb{M}\big\}, \\ \mathcal{C}_{\mathrm{Lip}}(\mathbb{M}, [0,1]) &:= \big\{f \in \mathcal{C}_{\mathrm{Lip}}(\mathbb{M}) : 0 \leq f \leq 1\big\}. \end{aligned}$$

Streng genommen, hängen all diese Funktionenklassen von der Metrik d oder zumindest von der hiervon induzierten Topologie ab, doch wird diese Abhängigkeit nicht hervorgehoben. Desweiteren seien für $x \in \mathbb{M}$, $\epsilon \geq 0$ und $\emptyset \neq S \subset \mathbb{M}$

$$\begin{aligned} d(x,S) &:= \inf_{y \in S} d(x,y) \quad \text{für } x \in \mathbb{M}, S \subset \mathbb{M}, \\ U(S,\epsilon) &:= \big\{y \in \mathbb{M} : d(y,S) < \epsilon\big\}, \\ B(S,\epsilon) &:= \big\{y \in \mathbb{M} : d(y,S) \leq \epsilon\big\}. \end{aligned}$$

Bekanntlich ist $d(\cdot, S)$ Lipschitzstetig mit Konstante Eins. Daher sind die Mengen $U(S,\epsilon)$ und $B(S,\epsilon)$ offen bzw. abgeschlossen. Speziell ist $U(x,\epsilon) := U(\{x\},\epsilon)$ und $B(x,\epsilon) := B(\{x\},\epsilon)$ die offene bzw. abgeschlossene Kugel um x mit Radius ϵ.

Ein $\mathbb{M}$-*Zufallselement* ist eine Abbildung von einem Wahrscheinlichkeitsraum nach $\mathbb{M}$. Eine messbare Abbildung von einem Wahrscheinlichkeitsraum nach $\mathbb{M}$ heißt $\mathbb{M}$-*Zufallsvariable*. Im folgenden sei stets $(Z_n)_n$ eine Folge von $\mathbb{M}$-Zufallselementen, und Z sei eine $\mathbb{M}$-Zufallsvariable.

Definition 7.3. Die Folge $(Z_n)_n$ *konvergiert in Verteilung gegen* Z, falls

$$\lim_{n\to\infty} \mathbb{E}^* f(Z_n) \;=\; \mathbb{E} f(Z) \quad \text{für beliebige } f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M}).$$

In Symbolen: $Z_n \to_{\mathcal{L}} Z$.

Die Limesvariable Z ist in Verteilung eindeutig, das heißt: Ist $\tilde{Z}$ eine weitere $\mathbb{M}$-Zufallsvariable, so dass $Z_n \to_{\mathcal{L}} \tilde{Z}$, dann ist $\mathcal{L}(\tilde{Z}) = \mathcal{L}(Z)$. Äquivalent zu $Z_n \to_{\mathcal{L}} Z$ ist, dass $\mathbb{E}_* f(Z_n) \to \mathbb{E} f(Z)$ für alle $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M})$, denn $\mathbb{E}(-f)(Z)$ ist gleich $-\mathbb{E} f(Z)$. Insbesondere hat $(Z_n)_n$ dann folgende Eigenschaft:

Definition 7.4. Die Folge $(Z_n)_n$ heißt *asymptotisch messbar*, falls

$$\lim_{n\to\infty} \big(\mathbb{E}^* f(Z_n) - \mathbb{E}_* f(Z_n)\big) \;=\; 0 \quad \text{für beliebige } f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M}).$$

Man kann die Verteilungskonvergenz noch auf andere Arten charakterisieren, indem man nur Teilklassen von $\mathcal{C}_{\mathrm{b}}(\mathbb{M})$, bestimmte Borelmengen oder halbstetige Funktionen betrachtet. Eine Funktion $f : \mathbb{M} \to \mathbb{R}$ heißt *halbstetig von unten*, falls für beliebige $r \in \mathbb{R}$ die Menge $\{f > r\}$ offen ist. Sie heißt *halbstetig von oben*, wenn $-f$ halbstetig von unten ist.

Satz 7.5 (Portmanteau-Theorem). *Die folgenden Aussagen über* $(Z_n)_n$ *und* Z *sind äquivalent:*

$$Z_n \;\to_{\mathcal{L}}\; Z; \tag{7.1}$$

$$\lim_{n\to\infty} \mathbb{E}^* f(Z_n) \;=\; \mathbb{E} f(Z) \quad \textit{für alle } f \in \mathcal{C}_{\mathrm{Lip}}(\mathbb{M},[0,1]); \tag{7.2}$$

$$\limsup_{n\to\infty} \mathbb{P}^*(Z_n \in A) \;\le\; \mathbb{P}(Z \in A) \quad \textit{für alle abgeschlossenen } A \subset \mathbb{M}; \tag{7.3}$$

$$\liminf_{n\to\infty} \mathbb{P}_*(Z_n \in U) \;\ge\; \mathbb{P}(Z \in U) \quad \textit{für alle offenen } U \subset \mathbb{M}; \tag{7.4}$$

$$\limsup_{n\to\infty} \mathbb{E}^* f(Z_n) \;\le\; \mathbb{E} f(Z) \quad \textit{für alle von oben halbstetigen, beschränkten } f : \mathbb{M} \to \mathbb{R}; \tag{7.5}$$

$$\liminf_{n\to\infty} \mathbb{E}_* f(Z_n) \;\ge\; \mathbb{E} f(Z) \quad \textit{für alle von unten halbstetigen, beschränkten } f : \mathbb{M} \to \mathbb{R}; \tag{7.6}$$

$$\lim_{n\to\infty} \mathbb{P}^*(Z_n \in B) \;=\; \lim_{n\to\infty} \mathbb{P}_*(Z_n \in B) \;=\; \mathbb{P}(Z \in B) \quad \textit{für alle } B \in \mathcal{B}(\mathbb{M}) \textit{ mit } \mathbb{P}(Z \in \partial B) = 0. \tag{7.7}$$

Beweis des Portmanteau-Theorems. Offensichtlich folgt (7.2) aus (7.1). Die Äquivalenz von (7.3) und (7.4) beziehungsweise von (7.5) und (7.6) ist leicht nachzuweisen. Ebenfalls trivial ist die Tatsache, dass (7.1) aus (7.5-7.6) folgt.

Angenommen es gilt (7.2). Für eine beliebige abgeschlossene Menge $A \subset \mathcal{X}$ und $R > 0$ sei $f_R(x) := \max\bigl(1 - R\, d(x, A), 0\bigr)$. Dann ist $1_A \le f_R \in \mathcal{C}_{\mathrm{Lip}}(\mathbb{M}, [0,1])$, und

$$\limsup_{n\to\infty} \mathbb{P}^*(Z_n \in A) \;\le\; \lim_{n\to\infty} \mathbb{E}^* f_R(Z_n) \;=\; \mathbb{E} f_R(Z).$$

Doch für $R \uparrow \infty$ konvergiert die rechte Seite gegen $\mathbb{P}(Z \in A)$, und somit folgt (7.3). Bei diesem Beweisschritt hängt man über den "Kleiderbügel" 1_A den "Mantel" f_R, was eine von mehreren Erklärungen für den Namen "Port(e)manteau-Theorem" ist.

Angenommen es gilt (7.3). Ist f eine von oben halbstetige Funktion auf $\mathbb{M}$ mit Werten in $[a, b]$, dann ergibt sich aus Lemma 7.1, Lemma 7.2 und dem Lemma von Fatou, dass

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{E}^* f(Z_n) &= \limsup_{n\to\infty}\Bigl(a + \int_a^b \mathbb{P}^*\bigl(f(Z_n) \ge t\bigr)\, dt\Bigr) \\
&= a + \limsup_{n\to\infty} \int_a^b \mathbb{P}^*\bigl(Z_n \in \{f \ge t\}\bigr)\, dt \\
&\le a + \int_a^b \limsup_{n\to\infty} \mathbb{P}^*\bigl(Z_n \in \{f \ge t\}\bigr)\, dt \\
&\le a + \int_a^b \mathbb{P}\bigl(Z \in \{f \ge t\}\bigr)\, dt \\
&= \mathbb{E} f(Z),
\end{aligned}$$

also Aussage (7.6).

Zu zeigen bleibt die Äquivalenz von (7.3-7.4) und (7.7). Mit dem Abschluss $\overline{B}$ und dem Inneren B^o von $B \in \mathcal{B}(\mathbb{M})$ folgt aus den Aussagen (7.3-7.4), dass

$$\left.\begin{array}{lllll}
\limsup\limits_{n\to\infty} \mathbb{P}^*(Z_n \in B) & \le & \limsup\limits_{n\to\infty} \mathbb{P}^*(Z_n \in \overline{B}) & \le & \mathbb{P}(Z \in \overline{B}) \\
\liminf\limits_{n\to\infty} \mathbb{P}_*(Z_n \in B) & \ge & \liminf\limits_{n\to\infty} \mathbb{P}_*(Z_n \in B^o) & \ge & \mathbb{P}(Z \in B^o)
\end{array}\right\} \;=\; \mathbb{P}(Z \in B),$$

sofern $\mathbb{P}(Z \in \partial B) = 0$. Umgekehrt gelte Aussage (7.7). Für beliebige abgeschlossene Mengen $A \subset \mathbb{M}$ sind die Mengen $\partial B(A, \epsilon) \subset \bigl\{d(\cdot, A) = \epsilon\bigr\}$, $\epsilon > 0$, paarweise disjunkt. Folglich ist $\mathbb{P}\bigl(Z \in \partial B(A, \epsilon)\bigr) = 0$ für alle bis auf abzählbar viele $\epsilon > 0$. Für solche $\epsilon > 0$ ist

$$\limsup_{n\to\infty} \mathbb{P}^*(Z_n \in A) \;\le\; \limsup_{n\to\infty} \mathbb{P}^*\bigl(Z_n \in B(A, \epsilon)\bigr) \;=\; \mathbb{P}\bigl(Z \in B(A, \epsilon)\bigr),$$

und für $\epsilon \downarrow 0$ ergibt sich Aussage (7.3). $\square$

Als Korollar zum Portmanteau-Theorem ergeben sich zwei Resultate, die in vielen statistischen Anwendungen von Nutzen sind:

Korollar 7.6 (Slutzkis Lemma, I). *Für $n \in \mathbb{N}$ sei $\tilde{Z}_n$ ein $\mathbb{M}$-Zufallselement, das auf dem gleichen Wahrscheinlichkeitsraum wie Z_n definiert ist. Angenommen $(Z_n)_n$ konvergiert in Verteilung gegen Z, und*

$$d(Z_n, \tilde{Z}_n) \;\to_p\; 0;$$

das heißt, $\mathbb{P}^*\bigl(d(Z_n, \tilde{Z}_n) > \epsilon\bigr) \to 0$ für beliebige $\epsilon > 0$. Dann konvergiert auch $\tilde{Z}_n$ in Verteilung gegen Z.

Korollar 7.7 (Slutzkis Lemma, II). *Sei* $(\mathbb{M}, |\cdot|)$ *ein normierter Raum. Für* $n \in \mathbb{N}$ *sei* $L_n : \mathbb{M} \to \mathbb{M}$ *ein zufälliger linearer Operator, definiert auf dem gleichen Wahrscheinlichkeitsraum wie* Z_n*, so dass*

$$\|L_n - \mathrm{Id}\| = \sup_{x \in \mathbb{M} : |x| \le 1} |L_n x - x| \ \to_p \ 0.$$

Dann folgt aus $Z_n \to_{\mathcal{L}} Z$*, dass auch* $L_n Z_n \to_{\mathcal{L}} Z$.

Beweis von Korollar 7.6. Für beliebige $f \in \mathcal{C}_{\mathrm{Lip}}(\mathbb{M}, [0,1])$ mit Lipschitzkonstante c und $\epsilon > 0$ ist

$$\begin{aligned}
&\big|\mathbb{E}^* f(Z_n) - \mathbb{E}^* f(\tilde{Z}_n)\big| \\
&\le \ \mathbb{E}^*\big|f(Z_n) - f(\tilde{Z}_n)\big| \\
&\le \ \mathbb{E}^*\, 1\big\{d(Z_n, \tilde{Z}_n) \le \epsilon\big\}\big|f(Z_n) - f(\tilde{Z}_n)\big| + \mathbb{P}^*\big(d(Z_n, \tilde{Z}_n) > \epsilon\big) \\
&\le \ c\epsilon + o(1).
\end{aligned}$$

□

Beweis von Korollar 7.7. Es genügt zu zeigen, dass $|L_n Z_n - Z_n| \to_p 0$. Doch für jedes feste $\epsilon > 0$ und beliebige $\delta > 0$ ist

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}^*\big(|L_n Z_n - Z_n| \ge \epsilon\big) \ &\le \ \limsup_{n\to\infty} \mathbb{P}^*(\|L_n - \mathrm{Id}\| \ge \delta) + \limsup_{n\to\infty} \mathbb{P}^*(|Z_n| \ge \epsilon/\delta) \\
&\le \ 0 + \mathbb{P}(|Z| \ge \epsilon/\delta),
\end{aligned}$$

und die rechte Seite konvergiert gegen Null für $\delta \downarrow 0$. □

7.3 Stetige und asymptotisch stetige Abbildungen

In diesem Abschnitt seien $(\mathbb{M}, d)$ und $(\tilde{\mathbb{M}}, \tilde{d})$ zwei metrische Räume, und $(Z_n)_n$ sei stets eine Folge von $\mathbb{M}$-Zufallselementen, die in Verteilung gegen eine $\mathbb{M}$-wertige Zufallsvariable Z konvergiert. Eine erste Frage ist, für welche Abbildungen $H : \mathbb{M} \to \tilde{\mathbb{M}}$ folgt, dass $(H(Z_n))_n \to_{\mathcal{L}} H(Z)$. Man kann leicht nachweisen, dass alle stetigen Abbildungen H dies leisten. Diese Aussage lässt sich noch etwas verallgemeinern:

Satz 7.8 (von der stetigen Abbildung). *Die Menge* $\mathbb{M}_o$ *aller Stetigkeitsstellen von* H *ist eine Borelmenge in* $\mathbb{M}$*. Falls* Z *nur Werte in* $\mathbb{M}_o$ *annimmt, konvergiert* $(H(Z_n))_n$ *in Verteilung gegen* $H(Z)$.

Beweis des Satzes von der stetigen Abbildung. Wir beweisen nur die Aussage, dass $\mathbb{M}_o$ eine Borelmenge ist, denn die Konvergenz in Verteilung von $H(Z_n)$ ergibt sich aus dem allgemeineren Satz von Rubin (Satz 7.9).

Ein Punkt $x \in \mathbb{M}$ gehört zu $\mathbb{M}_o$ genau dann, wenn zu jedem $k \in \mathbb{N}$ ein $\ell \in \mathbb{N}$ existiert derart, dass $\tilde{d}(H(y), H(z)) \le k^{-1}$ für alle $y, z \in U(x, \ell^{-1})$. Mit anderen Worten, x gehört nicht zu $\mathbb{M}_o$, falls für ein $k \in \mathbb{N}$ und beliebige $\ell \in \mathbb{N}$ zwei Punkte $y, z \in \mathbb{M}$ existieren derart, dass

$\tilde{d}(H(y), H(z)) > k^{-1}$ aber $x \in U(y, \ell^{-1}) \cap U(z, \ell^{-1})$. Somit lässt sich $\mathbb{M} \setminus \mathbb{M}_o$ schreiben als $\bigcup_{k\in\mathbb{N}} \bigcap_{\ell\in\mathbb{N}} W_{k,\ell}$ mit den offenen Mengen

$$W_{k,\ell} \ := \bigcup_{y,z\in\mathbb{M}\,:\,\tilde{d}(H(y),H(z))>k^{-1}} U(y, \ell^{-1}) \cap U(z, \ell^{-1}).$$

□

Satz 7.8 ist ein Spezialfall eines noch allgemeineren Resultates über Folgen von Abbildungen:

Satz 7.9 (Rubin). *Für $n \in \mathbb{N}$ sei H_n eine Abbildung von $\mathbb{M}$ nach $\tilde{\mathbb{M}}$. Angenommen, für alle Punkte x in einer Borelmenge $\mathbb{M}_o \subset \mathbb{M}$ existiert*

$$H(x) \ := \lim_{n\to\infty, y\to x} H_n(y).$$

Dann ist $H : \mathbb{M}_o \to \tilde{\mathbb{M}}$ stetig. Falls Z nur Werte in $\mathbb{M}_o$ annimmt, konvergiert $(H_n(Z_n))_n$ in Verteilung gegen $H(Z)$.

Anmerkung 7.10. Ist $(\tilde{\mathbb{M}}, \tilde{d})$ vollständig, so ist die Menge $\mathbb{M}_o$ aller $x \in \mathbb{M}$ derart, dass

$$H(x) \ := \lim_{n\to\infty, y\to x} H_n(y)$$

existiert, eine Borelmenge in $(\mathbb{M}, d)$. Denn ein Punkt x gehört zu $\mathbb{M} \setminus \mathbb{M}_o$ genau dann, wenn

$$\lim_{\ell\to\infty} \sup_{m,n\geq\ell;\, y,z\in U(x,\ell^{-1})} \tilde{d}(H_m(y), H_n(z)) \ > \ 0.$$

Folglich kann man $\mathbb{M} \setminus \mathbb{M}_o$ darstellen als $\bigcup_{k\in\mathbb{N}} \bigcap_{\ell\in\mathbb{N}} W_{k,\ell}$ mit den offenen Mengen

$$W_{k,\ell} \ := \bigcup_{m,n\geq\ell;\, y,z\in\mathbb{M}\,:\,\tilde{d}(H_m(y),H_n(z))>k^{-1}} U(y, \ell^{-1}) \cap U(z, \ell^{-1}).$$

Beweis von Satz 7.9. Für beliebiges festes $\epsilon > 0$ ist

$$\mathbb{M}_o \ = \ \bigcup_{\ell\in\mathbb{N}} V_{\epsilon,\ell},$$

wobei

$$V_{\epsilon,\ell} \ := \ \Big\{ x \in \mathbb{M}_o : \tilde{d}(H(x), H_n(y)) \leq \epsilon \text{ für alle } n \geq \ell, y \in U(x, \ell^{-1}) \Big\}.$$

Insbesondere ist $\tilde{d}(H(x), H(y)) \leq \epsilon$, falls $x \in V_{\epsilon,\ell}$ und $y \in \mathbb{M}_o \cap U(x, \ell^{-1})$. Dies impliziert die Stetigkeit von H. Ferner ergibt sich für beliebige abgeschlossene Mengen $\tilde{A} \subset \tilde{\mathbb{M}}$ und $\ell \in \mathbb{N}$, dass

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}^*\big(H_n(Z_n) \in \tilde{A}\big) \\
&\quad\leq \ \limsup_{n\to\infty} \mathbb{P}^*\big(Z_n \notin U(V_{\epsilon,\ell}, \ell^{-1})\big) + \limsup_{n\to\infty} \mathbb{P}^*\big(Z_n \in U(V_{\epsilon,\ell}, \ell^{-1}) \text{ und } H_n(Z_n) \in \tilde{A}\big) \\
&\quad\leq \ \mathbb{P}(Z \notin V_{\epsilon,\ell}) + \limsup_{n\to\infty} \mathbb{P}^*\big(Z_n \in U(V_{\epsilon,\ell}, \ell^{-1}) \text{ und } H_n(Z_n) \in \tilde{A}\big) \\
&\quad\leq \ \mathbb{P}(Z \notin V_{\epsilon,\ell}) + \limsup_{n\to\infty} \mathbb{P}^*\Big(Z_n \in B\big(H^{-1}(B(\tilde{A}, \epsilon)), \ell^{-1}\big)\Big) \\
&\quad\leq \ \mathbb{P}(Z \notin V_{\epsilon,\ell}) + \mathbb{P}\Big(Z \in B\big(H^{-1}(B(\tilde{A}, \epsilon)), \ell^{-1}\big)\Big).
\end{aligned}$$

Die zweite und vierte Ungleichung folgen aus dem Portmanteau-Theorem. Die dritte Ungleichung ergibt sich wie folgt: Ist $Z_n \in U(V_{\epsilon,\ell}, \ell^{-1})$ und $H_n(Z_n) \in \tilde{A}$, dann existiert ein $x \in V_{\epsilon,\ell}$ mit $Z_n \in U(x, \ell^{-1})$. Insbesondere ist $\tilde{d}(H(x), H_n(Z_n)) \le \epsilon$, sobald $n \ge \ell$, also $H(x) \in B(\tilde{A}, \epsilon)$ und $Z_n \in U\big(H^{-1}(B(\tilde{A}, \epsilon)), \ell^{-1}\big)$.

Doch wegen der Stetigkeit von H ist $H^{-1}(B(\tilde{A}, \epsilon)) = A_\epsilon \cap \mathbb{M}_o$ für eine abgeschlossene Menge $A_\epsilon \subset \mathbb{M}$. Folglich ist

$$\begin{aligned}
\mathbb{P}(Z \notin V_{\epsilon,\ell}) + \mathbb{P}\Big(Z \in B\big(H^{-1}(B(\tilde{A}, \epsilon)), \ell^{-1}\big)\Big) \ &\le\ \mathbb{P}(Z \notin V_{\epsilon,\ell}) + \mathbb{P}\big(Z \in B(A_\epsilon, \ell^{-1})\big) \\
&\to\ \mathbb{P}(Z \in A_\epsilon) \quad (\ell \to \infty) \\
&=\ \mathbb{P}(Z \in A_\epsilon \cap \mathbb{M}_o) \\
&=\ \mathbb{P}\big(H(Z) \in B(\tilde{A}, \epsilon)\big) \\
&\to\ \mathbb{P}\big(H(Z) \in \tilde{A}\big) \quad (\epsilon \downarrow 0). \qquad \square
\end{aligned}$$

Hadamard-differenzierbare Funktionale. Eine typische Anwendungssituation von Satz 7.9 sind Hadamard-differenzierbare Funktionale. Seien $\mathbb{M}$ und $\tilde{\mathbb{M}}$ normierte Vektorräume; die Norm bezeichnen wir in beiden Fällen mit $\|\cdot\|$. Eine Abbildung $T : \mathbb{M} \to \tilde{\mathbb{M}}$ heißt *Gateaux-differenzierbar an der Stelle* $x_o \in \mathbb{M}$, falls es einen linearen Operator $DT(x_o, \cdot) : \mathbb{M} \to \tilde{\mathbb{M}}$ gibt derart, dass

$$\lim_{t \downarrow 0} t^{-1}\big(T(x_o + tv) - T(x_o)\big) - DT(x_o, v) \ = \ 0$$

für beliebige $v \in \mathbb{M}$. Die Abbildung heißt *Fréchet-differenzierbar an der Stelle* $x_o \in \mathbb{M}$, falls die lineare Abbildung $DT(x_o, \cdot)$ sogar stetig ist und

$$\lim_{t \downarrow 0} \sup_{v \in B} \Big\| t^{-1}\big(T(x_o + tv) - T(x_o)\big) - DT(x_o, v) \Big\| \ = \ 0$$

für jede beschränkte Teilmenge B von $\mathbb{M}$. Man kann leicht zeigen, dass dies gleichbedeutend mit folgender Aussage ist:

$$\lim_{\mathbb{M} \setminus \{0\} \ni v \to 0} \|v\|^{-1} \big\| T(x_o + v) - T(x_o) - DT(x_o, v) \big\| \ = \ 0$$

Nun gibt es noch eine Abschwächung der Fréchet-Differenzierbarkeit, die in einigen statistischen Anwendungen nachweisbar und ausreichend ist: Die Abbildung T heißt *Hadamard-differenzierbar an der Stelle* $x_o \in \mathbb{M}$, falls es eine stetige lineare Abbildung $DT(x_o, \cdot) : \mathbb{M} \to \tilde{\mathbb{M}}$ gibt derart, dass

$$\lim_{t \downarrow 0} \sup_{v \in K} \Big\| t^{-1}\big(T(x_o + tv) - T(x_o)\big) - DT(x_o, v) \Big\| \ = \ 0$$

für jede *kompakte* Teilmenge K von $\mathbb{M}$. In Aufgabe 7.1 wird gezeigt, dass diese Aussage gleichbedeutend ist mit folgender Charakterisierung: Es existiert eine lineare Abbildung $DT(x_o, \cdot) : \mathbb{M} \to \tilde{\mathbb{M}}$ derart, dass

$$\lim_{\lambda \to \infty, w \to v} \lambda\big(T(x_o + \lambda^{-1} w) - T(x_o)\big) \ = \ DT(x_o, v) \quad \text{für beliebige } v \in \mathbb{M}.$$

Der Spezialfall, dass $\mathbb{M} = \ell_\infty(\mathcal{T})$. Seien $\mathbb{M}$ und $\mathbb{M}_o \subset \mathbb{M}$ abgeschlossene lineare Teilräume von $\ell_\infty(\mathcal{T})$, versehen mit der Supremumsnorm $\|\cdot\| = \|\cdot\|_{\mathcal{T}}$. Für ein festes $F \in \ell_\infty(\mathcal{T})$ sei T ein reellwertiges Funktional auf dem affinen Raum $F + \mathbb{M}$. Ferner sei $(\hat{F}_n)_n$ eine Folge von $(F + \mathbb{M})$-Zufallselementen, so dass gilt:

$$Z_n := \sqrt{n}(\hat{F}_n - F) \to_{\mathcal{L}} Z$$

mit einer $\mathbb{M}_o$-Zufallsvariablen Z. Oftmals existiert der Grenzwert

$$H(x) := \lim_{\lambda\to\infty, y\to x} \lambda\left(T(F + \lambda^{-1}y) - T(F)\right)$$

für beliebige $x \in \mathbb{M}_o$, wobei λ eine positive reelle Zahl und y eine Funktion aus $\mathbb{M}$ bezeichnen. Ist H linear auf $\mathbb{M}_o$, dann nennt man T *Hadamard-differenzierbar (kompakt differenzierbar) im Punkt F tangential zu* $\mathbb{M}_o$. Mit $H_n(y) := \sqrt{n}\big(T(F + y/\sqrt{n}) - T(F)\big)$ folgt dann aus Satz 7.9, dass

$$\sqrt{n}\big(T(\hat{F}_n) - T(F)\big) = H_n(Z_n) \to_{\mathcal{L}} H(Z).$$

Beispiel 7.11. Sei $T(G) := \sup_{t\in\mathcal{T}} G(t)$. Ferner sei ρ eine Metrik auf $\mathcal{T}$. Angenommen, F erfüllt die folgenden zwei Bedingungen:

$$\begin{aligned} K &:= \big\{t \in \mathcal{T} : F(t) = T(F)\big\} \quad \text{ist nichtleer und kompakt bezüglich } \rho;\\ \sup_{t\in\mathcal{T}:\rho(t,K)\ge\delta} F(t) &< T(F) \quad \text{für beliebige } \delta > 0. \end{aligned}$$

Dann ist

$$\lim_{\lambda\to\infty, y\to x} \lambda\left(T(F + \lambda^{-1}y) - T(F)\right) = \max_{t\in K} x(t)$$

für beliebige Funktionen $x \in \ell_\infty(\mathcal{T})$, welche in jedem Punkt $t \in K$ stetig sind. Die Menge $\mathbb{M}_o$ all dieser Funktionen ist ein abgeschlossener linearer Teilraum von $\mathbb{M}$; siehe Aufgabe 7.2.

Beweis. Für beliebige $\delta > 0$ gilt:

$$\begin{aligned}
\lambda\Big(\sup_{t\in\mathcal{T}:\rho(t,K)\ge\delta}(F + \lambda^{-1}y)(t) - T(F)\Big) &\le \lambda\Big(\sup_{t\in\mathcal{T}:\rho(t,K)\ge\delta} F(t) - T(F)\Big) + \|y\| \\
&\to -\infty \quad \text{für } \lambda\to\infty \text{ und } y\to x,\\
\lambda\Big(\sup_{t\in\mathcal{T}:\rho(t,K)<\delta}(F + \lambda^{-1}y)(t) - T(F)\Big) &\le \sup_{t\in\mathcal{T}:\rho(t,K)<\delta} y(t)\\
&\to \max_{t\in K} x(t) \quad \text{für } y\to x \text{ und } \delta\downarrow 0,\\
\lambda\Big(\sup_{t\in\mathcal{T}}(F + \lambda^{-1}y)(t) - T(F)\Big) &\ge \lambda\Big(\sup_{t\in K}(F + \lambda^{-1}y)(t) - T(F)\Big)\\
&= \sup_{t\in K} y(t)\\
&\to \max_{t\in K} x(t) \quad \text{für } y\to x.
\end{aligned}$$

□

Beispiel 7.12. Sei $\mathcal{T} = \mathbb{R}$, sei $\mathbb{M}$ die Menge aller $x \in \ell_\infty(\mathbb{R})$, so dass $x(r) \to 0$ für $|r| \to \infty$, und sei F eine feste Verteilungsfunktion. Für $0 < \beta < 1$ und $G \in F + \mathbb{M}$ sei

$$T(G) \ := \ \inf\bigl\{r \in \mathbb{R} : G(r) \geq \beta\bigr\},$$

das "β-Quantil" von G. Angenommen,

$$F(r) \ = \ \beta + c(r - T(F)) + o(|r - T(F)|) \quad \text{für } r \to T(F)$$

mit $c > 0$. Dann ist

$$\lim_{\lambda\to\infty, y\to x} \lambda\bigl(T(F + \lambda^{-1}y) - T(F)\bigr) \ = \ -\frac{x(T(F))}{c}$$

für beliebige Funktionen $x \in \mathbb{M}$, welche in $T(F)$ stetig sind.

Beweis. Für beliebige $\delta > \|x\|/c$ gilt:

$$\begin{aligned}
\lambda\Bigl(\sup_{r\leq T(F)-\lambda^{-1}\delta}(F + \lambda^{-1}y)(r) - \beta\Bigr) \ &\leq \ \lambda\Bigl(F\bigl(T(F) - \lambda^{-1}\delta\bigr) - \beta\Bigr) + \|y\| \\
&\to \ -c\delta + \|x\| \quad (\lambda \to \infty, y \to x) \\
&< \ 0, \\
\lambda\Bigl(\inf_{r\geq T(F)+\lambda^{-1}\delta}(F + \lambda^{-1}y)(r) - \beta\Bigr) \ &\geq \ \lambda\Bigl(F\bigl(T(F) + \lambda^{-1}\delta\bigr) - \beta\Bigr) - \|y\| \\
&\to \ c\delta - \|x\| \quad (\lambda \to \infty, y \to x) \\
&> \ 0,
\end{aligned}$$

und

$$\lambda\Bigl((F + \lambda^{-1}y)\bigl(T(F) + \lambda^{-1}s\bigr) - \beta\Bigr) \ \to \ cs + x(T(F)) \quad (\lambda \to \infty, y \to x)$$

gleichmäßig in $s \in [-\delta, \delta]$. □

Beispiel 7.13. (Die Länge des "Shorth", Grübel 1988) Seien $\mathcal{T} = \mathbb{R}$, F und $\mathbb{M}$ wie in Beispiel 7.12. Für $G \in F + \mathbb{M}$ sei nun

$$T(G) \ := \ \inf\bigl\{u - t : -\infty < t < u < \infty, G(u) - G(t) \geq 1/2\bigr\},$$

ein robustes "Skalenfunktional". Ein Intervall $[t, u]$ mit $G(u) - G(t) \geq 1/2$ und $u - t = T(G)$ ist ein "shorth" (*short*est interval with probability at least one *h*alf) von G. Angenommen, F hat eine stetige Dichte f derart, dass $\{f > 0\} = (a, b)$. Ferner sei f streng monoton wachsend auf $(a, m(F)]$ und streng monoton fallend auf $[m(F), b)$ für ein $m(F) \in (a, b)$. Für jedes feste $\delta \in (0, b - a)$ hat dann die Funktion $r \mapsto F(r + \delta) - F(r - \delta)$ ein eindeutiges Maximum an der Stelle r_δ, so dass $f(r_\delta + \delta) = f(r_\delta - \delta)$. Insbesondere ist $T(F) = 2\delta_o$, wobei $r_o = r_{\delta_o}$ und $\delta_o > 0$ eindeutig charakterisiert werden durch

$$f(r_o + \delta_o) \ = \ f(r_o - \delta_o) \quad \text{und} \quad F(r_o + \delta_o) - F(r_o - \delta_o) \ = \ 1/2.$$

Hier kann man zeigen, dass

$$\lim_{\lambda\to\infty, y\to x} \lambda\bigl(T(F + \lambda^{-1}y) - T(F)\bigr) \ = \ \frac{x(r_o - \delta_o) - x(r_o + \delta_o)}{f(r_o + \delta_o)} \tag{7.8}$$

für beliebige $x \in \mathbb{M}$, welche in $r_o \pm \delta_o$ stetig sind (Aufgabe 7.4).

7.4 Gleichmässige Konvergenz

Sei $\mathcal{F}$ eine Familie von messbaren, beschränkten Funktionen f auf $\mathbb{M}$. Der Satz von der stetigen Abbildung impliziert, dass $\mathbb{E}^* f(Z_n) \to \mathbb{E} f(Z)$, falls $Z_n \to_{\mathcal{L}} Z$ und f $\mathcal{L}(Z)$-fast überall stetig ist. Das kann man auch wie folgt formulieren:

$$\lim_{\delta \downarrow 0} \mathbb{P}\Big(\sup_{x,y \in U(Z,\delta)} (f(x) - f(y)) > \epsilon \Big) \; = \; 0 \quad \text{für beliebige } \epsilon > 0.$$

(Als Funktion von z ist $\sup_{x,y \in U(z,\delta)}(f(x) - f(y))$ von unten halbstetig.) Nun geht es um die Frage, unter welchen Voraussetzungen $\mathbb{E}^* f(Z_n) - \mathbb{E} f(Z)$ gleichmäßig in $f \in \mathcal{F}$ gegen Null konvergiert.

Satz 7.14 (Billingsley, Topsøe 1967). *Angenommen,* $\mathbb{P}(Z \in \mathbb{M}_o) = 1$ *für eine separable Borelmenge* $\mathbb{M}_o \subset \mathbb{M}$. *Folgende zwei Bedingungen an* $\mathcal{F}$ *und* $\mathcal{L}(Z)$ *sind äquivalent:*

(7.9)
$$\sup_{f \in \mathcal{F}} \big| \mathbb{E}^* f(Z_n) - \mathbb{E} f(Z) \big| \; \to \; 0 \quad \text{wann immer } Z_n \to_{\mathcal{L}} Z.$$

(7.10)
$$\sup_{f \in \mathcal{F}} \sup_{x,y \in \mathbb{M}} (f(x) - f(y)) \; < \; \infty \quad \text{und}$$
$$\lim_{\delta \downarrow 0} \sup_{f \in \mathcal{F}} \mathbb{P}\Big(\sup_{x,y \in U(Z,\delta)} (f(x) - f(y)) > \epsilon \Big) \; = \; 0 \quad \text{für beliebige } \epsilon > 0.$$

Dieser Satz verallgemeinert folgendes Resultat:

Korollar 7.15 (Rao 1962). *Angenommen,* $\mathbb{P}(Z \in \mathbb{M}_o) = 1$ *für eine separable Borelmenge* $\mathbb{M}_o \subset \mathbb{M}$. *Sei* $\sup_{f \in \mathcal{F}} \sup_{x,y \in \mathbb{M}} (f(x) - f(y)) < \infty$, *und* $\mathcal{F}$ *sei in jedem Punkt* $x \in \mathbb{M}_o$ *gleichgradig stetig. Das heißt,*

$$\lim_{\delta \downarrow 0} \sup_{f \in \mathcal{F}, y \in U(x,\delta)} \big| f(y) - f(x) \big| \; = \; 0 \quad \text{für beliebige } x \in \mathbb{M}_o.$$

Dann folgt aus $Z_n \to_{\mathcal{L}} Z$, *dass*

$$\sup_{f \in \mathcal{F}} \big| \mathbb{E}^* f(Z_n) - \mathbb{E} f(Z) \big| \; \to \; 0.$$

Ein spezielles Beispiel für $\mathcal{F}$ in Korollar 7.15 ist die Klasse aller Funktionen f auf $\mathbb{M}$ mit $|f(x)| \leq 1$ und $|f(x) - f(y)| \leq d(x,y)$ für alle $x, y \in \mathbb{M}$.

Korollar 7.16. *Angenommen,* $\mathbb{P}(Z \in \mathbb{M}_o) = 1$ *für eine separable Borelmenge* $\mathbb{M}_o \subset \mathbb{M}$. *Dann folgt aus* $Z_n \to_{\mathcal{L}} Z$, *dass*

$$\sup_{A \subset \mathbb{M}} \big(\mathbb{P}^*(Z_n \in A) - \mathbb{P}(Z \in U(A,\epsilon)) \big)^+ \; \to \; 0 \quad \text{für beliebige } \epsilon > 0.$$

Korollar 7.17. *Angenommen,* $\mathbb{M} = \mathbb{R}^d$ *und*

$$\mathbb{P}(Z \in \partial H) \; = \; 0 \quad \text{für beliebige Halbräume } H \subset \mathbb{R}^d.$$

Dann folgt aus $Z_n \to_{\mathcal{L}} Z$, dass

$$\sup_{\text{Halbräume } H \subset \mathbb{R}^d} \big| \mathbb{P}^*(Z_n \in H) - \mathbb{P}(Z \in H) \big| \;\to\; 0.$$

Beweis von Satz 7.14. Zunächst weisen wir die Notwendigkeit von (7.10) für (7.9) nach. Angenommen die erste Teilbedingung von (7.10) ist nicht erfüllt. Dann existieren Folgen $(f_n)_n$ in $\mathcal{F}$ und $(x_{1n})_n$, $(x_{2n})_n$ in $\mathbb{M}$ mit

$$1 \;\le\; r_n \;:=\; f_n(x_{1n}) - f_n(x_{2n}) \;\to\; \infty.$$

Nun seien $(Z_n^{(1)})_n$ und $(Z_n^{(2)})_n$ Folgen von Zufallsvariablen mit

$$\mathcal{L}(Z_n^{(j)}) \;:=\; \Big(1 - \frac{1}{\sqrt{r_n}}\Big)\mathcal{L}(Z) + \frac{1}{\sqrt{r_n}}\delta_{x_{jn}}.$$

Beide Folgen konvergieren in Verteilung gegen Z, aber

$$\mathbb{E}\, f_n(Z_n^{(1)}) - \mathbb{E}\, f_n(Z_n^{(2)}) \;=\; \sqrt{r_n} \;\to\; \infty.$$

Folglich kann dann auch (7.9) nicht erfüllt sein. Angenommen, die zweite Teilbedingung von (7.10) ist nicht erfüllt. Dann existiert ein $\epsilon_o > 0$ und eine Folge $(f_n)_n$ in $\mathcal{F}$ mit

$$\mathbb{P}\Big(\sup_{x,y \in U(Z,1/n)} (f_n(x) - f_n(y)) > \epsilon_o\Big) \;\ge\; \epsilon_o \quad \text{für alle } n.$$

Nun seien $X_n^{(1)}, X_n^{(2)} : \mathbb{M} \to \mathbb{M}$ Abbildungen mit

$$X_n^{(j)}(z) \in U(z, n^{-1}) \quad \text{und} \quad f_n(X_n^{(1)}(z)) - f_n(X_n^{(2)}(z)) \;\ge\; \frac{1}{2} \sup_{x,y \in U(z,1/n)} (f_n(x) - f_n(y))$$

für beliebige $z \in \mathbb{M}$. Dann sind $Z_n^{(j)} := X_n^{(j)}(Z)$ $\mathbb{M}$-Zufallselemente mit $d(Z_n^{(j)}, Z) < n^{-1}$, also $Z_n^{(j)} \to_{\mathcal{L}} Z$ nach Korollar 7.6. Aber Aussage (7.9) kann nicht erfüllt sein, da

$$\begin{aligned}
\mathbb{E}^* f_n(Z_n^{(1)}) - \mathbb{E}^* f_n(Z_n^{(2)}) \;&\ge\; \mathbb{E}_*\big(f_n(Z_n^{(1)}) - \mathbb{E}^* f_n(Z_n^{(2)})\big) \\
&\ge\; \frac{\epsilon_o}{2}\, \mathbb{P}\Big(\sup_{x,y \in U(Z,1/n)} (f_n(x) - f_n(y)) > \epsilon_o\Big) \\
&\ge\; \frac{\epsilon_o^2}{2}.
\end{aligned}$$

Somit ist Bedingung (7.10) notwendig für (7.9).

Eine Frage, der wir hier nicht nachgehen, ist: Kann man erreichen, dass die $X_n^{(j)}$ messbar sind? Dann wären die $Z_n^{(j)}$ sogar Zufallsvariablen.

Angenommen es gilt (7.10). Die Wahrheit von Aussage (7.9) bleibt unverändert, wenn man jede Funktion $f \in \mathcal{F}$ mit einer Konstanten $c > 0$ multipliziert und eine Konstante a_f addiert. Daher nehmen wir ohne Einschränkung an, dass $0 \le f \le 1$ für alle $f \in \mathcal{F}$. Für $\delta > 0$ gibt es paarweise disjunkte Borelmengen $B_{\delta,1}, B_{\delta,2}, B_{\delta,3}, \ldots$ mit folgenden Eigenschaften:

$$\sup_{x,y \in B_{\delta,k}} d(x,y) \;<\; \frac{\delta}{2}, \quad \mathbb{P}(Z \in \partial B_{\delta,k}) \;=\; 0 \quad \text{und} \quad \mathbb{M}_o \;\subset\; \bigcup_{\ell \in \mathbb{N}} B_{\delta,\ell}.$$

Sei nämlich $\{x_k : k \in \mathbb{N}\}$ eine dichte Teilmenge von $\mathbb{M}_o$. Für jedes k existiert ein $\delta_k \in [\delta/8, \delta/4)$ mit $\mathbb{P}(Z \in \partial B(x_k, \delta_k)) = 0$. Dann haben die Mengen $B_{\delta,1} := B(x_1, \delta_1)$ und $B_{\delta,k+1} := B(x_{k+1}, \delta_{k+1}) \setminus \bigcup_{\ell \le k} B(x_\ell, \delta_\ell)$ die gewünschten Eigenschaften. Nun gilt für beliebige $N \in \mathbb{N}$, $\epsilon > 0$ und $f \in \mathcal{F}$:

$$\begin{aligned}
\mathbb{E}^* f(Z_n) &\le \mathbb{P}^*\Big(Z_n \notin \bigcup_{k \le N} B_{\delta,k}\Big) + \sum_{k \le N} \mathbb{P}^*(Z_n \in B_{\delta,k}) \sup_{x \in B_{\delta,k}} f(x) \\
&\le \mathbb{P}^*\Big(Z_n \notin \bigcup_{k \le N} B_{\delta,k}\Big) + \sum_{k \le N} \big|\mathbb{P}^*(Z_n \in B_{\delta,k}) - \mathbb{P}(Z \in B_{\delta,k})\big| \\
&\qquad + \sum_{k \le N} \mathbb{P}(Z \in B_{\delta,k}) \sup_{x \in B_{\delta,k}} f(x).
\end{aligned}$$

Ferner ist

$$\begin{aligned}
\mathbb{E} f(Z) &\ge \sum_{k \le N} \mathbb{P}(Z \in B_{\delta,k}) \inf_{x \in B_{\delta,k}} f(x) \\
&= \sum_{k \le N} \mathbb{P}(Z \in B_{\delta,k}) \sup_{x \in B_{\delta,k}} f(x) - \sum_{k \le N} \mathbb{P}(Z \in B_{\delta,k}) \sup_{x,y \in B_{\delta,k}} (f(x) - f(y)) \\
&\ge \sum_{k \le N} \mathbb{P}(Z \in B_{\delta,k}) \sup_{x \in B_{\delta,k}} f(x) - \epsilon - \mathbb{P}\Big(\sup_{x,y \in U(Z,\delta)} (f(x) - f(y)) > \epsilon\Big) \\
&\ge \sum_{k \le N} \mathbb{P}(Z \in B_{\delta,k}) \sup_{x \in B_{\delta,k}} f(x) - \epsilon - \sup_{g \in \mathcal{F}} \mathbb{P}\Big(\sup_{x,y \in U(Z,\delta)} (g(x) - g(y)) > \epsilon\Big).
\end{aligned}$$

Denn für beliebige $z \in B_{\delta,k}$ ist $B_{\delta,k} \subset U(z, \delta)$, so dass aus $\sup_{x,y \in B_{\delta,k}} (f(x) - f(y)) > \epsilon$ folgt, dass auch $\sup_{x,y \in U(z,\delta)} (f(x) - f(y)) > \epsilon$. Falls also $Z_n \to_{\mathcal{L}} Z$, so folgt aus dem Portmanteau-Theorem, dass

$$\begin{aligned}
&\sup_{f \in \mathcal{F}} \big(\mathbb{E}^* f(Z_n) - \mathbb{E} f(Z)\big) \\
&\quad \le \mathbb{P}^*\Big(Z_n \notin \bigcup_{k \le N} B_{\delta,k}\Big) + \sum_{k \le N} \big|\mathbb{P}^*(Z_n \in B_{\delta,k}) - \mathbb{P}(Z \in B_{\delta,k})\big| \\
&\qquad + \epsilon + \sup_{g \in \mathcal{F}} \mathbb{P}\Big(\sup_{x,y \in U(Z,\delta)} (g(x) - g(y)) > \epsilon\Big) \\
&\quad \to \mathbb{P}\Big(Z \notin \bigcup_{k \le N} B_{\delta,k}\Big) + \epsilon + \sup_{f \in \mathcal{F}} \mathbb{P}\Big(\sup_{x,y \in U(Z,\delta)} (f(x) - f(y)) > \epsilon\Big) \\
&\quad \to 0 \quad (N \to \infty;\ \text{dann } \delta \downarrow 0;\ \text{dann } \epsilon \downarrow 0).
\end{aligned}$$

Ersetzt man $f \in \mathcal{F}$ durch $1 - f$, dann folgt analog, dass

$$\begin{aligned}
&\limsup_{n \to \infty} \sup_{f \in \mathcal{F}} \big(\mathbb{E} f(Z) - \mathbb{E}_* f(Z_n)\big) \\
&\quad = \limsup_{n \to \infty} \sup_{f \in \mathcal{F}} \big(\mathbb{E}^*(1-f)(Z_n) - \mathbb{E}(1-f)(Z)\big) \ \le\ 0.
\end{aligned}$$ □

Beweis von Korollar 7.15. Aus der gleichgradigen Stetigkeit von $\mathcal{F}$ in jedem Punkt von $\mathbb{M}_o$ folgt, dass

$$\mathbb{M}_o \subset \bigcup_{\delta > 0} \Big\{ z \in \mathbb{M} : \sup_{f \in \mathcal{F}, x,y \in U(z,\delta)} (f(x) - f(y)) \le \epsilon \Big\}$$

für beliebige $\epsilon > 0$, und dies impliziert den zweiten Teil von Bedingung (7.10) für beliebige $\mathbb{M}_o$-Zufallsvariablen Z. $\Box$

Beweis von Korollar 7.16.. Für beliebiges festes $\epsilon > 0$ und $A \subset \mathbb{M}$ ist

$$1_A \ \le \ f_A := \bigl(1 - d(\cdot, A)/\epsilon\bigr)^+ \ \le \ 1_{U(A,\epsilon)},$$

und f_A ist Lipschitz-stetig mit Konstante $1/\epsilon$. Folglich ist $\mathbb{P}^*(Z_n \in A) - \mathbb{P}(Z \in U(A,\epsilon)) \le \mathbb{E}^* f_A(Z_n) - \mathbb{E}\, f_A(Z)$, und die Behauptung folgt aus Korollar 7.15, angewandt auf $\mathcal{F} := \{f_A : A \subset \mathbb{M}\}$. $\Box$

Beweis von Korollar 7.17. Nach Korollar 7.16 genügt es zu zeigen, dass

$$\sup_{\text{Halbräume } H \subset \mathbb{R}^d} \mathbb{P}\bigl(Z \in U(H,\epsilon) \setminus H\bigr) \ \to \ 0 \quad \text{für } \epsilon \downarrow 0.$$

Diese Aussage ist Bestandteil von Aufgabe 7.5. $\Box$

7.5 Schwache Konvergenz (in Wahrscheinlichkeit)

Seien P und P_n, $n \in \mathbb{N}$, Wahrscheinlichkeitsmaße auf $\mathbb{M}$. Dann konvergiert die Folge $(P_n)_n$ *schwach* gegen P, falls

$$\lim_{n\to\infty} \int f\, dP_n \ = \ \int f\, dP \quad \text{für beliebige } f \in \mathcal{C}_b(\mathbb{M}).$$

In Symbolen: $P_n \to_w P$. Mit Zufallsvariablen $Z \sim P$ und $Z_n \sim P_n$ ist dies äquivalent zu $Z_n \to_{\mathcal{L}} Z$. Folglich kann man alle bisherigen Resultate über Verteilungskonvergenz umformulieren in Resultate über schwache Konvergenz.

Angenommen, man ersetzt die P_n durch zufällige Wahrscheinlichkeitsmaße $\tilde{P}_n$. Das bedeutet, jedes $\tilde{P}_n$ ist eine Abbildung von einem Wahrscheinlichkeitsraum in die Menge der Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{M})$. Eine naheliegende Verallgemeinerung der schwachen Konvergenz ist die *schwache Konvergenz in Wahrscheinlichkeit*: Die Folge $(\tilde{P}_n)_n$ konvergiert schwach gegen P in Wahrscheinlichkeit, wenn

$$\int f\, d\tilde{P}_n \ \to_p \ \int f\, dP \quad \text{für beliebige } f \in \mathcal{C}_b(\mathbb{M}).$$

In Symbolen: $\tilde{P}_n \to_{\mathrm{w,p}} P$. Man kann alle bisherigen Resultate auf zufällige Folgen $(\tilde{P}_n)_n$ übertragen.

Zum Beispiel besagt hier das Portmanteau-Theorem unter anderem, dass $\tilde{P}_n \to_{\mathrm{w,p}} P$ genau dann, wenn

$$\bigl(\tilde{P}_n(A) - P(A)\bigr)^+ \ \to_p \ 0 \quad \text{für beliebige abgeschlossene Mengen } A \subset \mathbb{M}.$$

Ferner folgt aus der schwachen Konvergenz von $(\tilde{P}_n)_n$ gegen P in Wahrscheinlichkeit, dass

$$\sup_{f \in \mathcal{F}} \Big| \int f \, d\tilde{P}_n - \int f \, dP \Big| \ \to_p \ 0,$$

vorausgesetzt, P ist auf einer separablen Borelmenge $\mathbb{M}_o \subset \mathbb{M}$ konzentriert, und die Funktionenfamilie $\mathcal{F}$ erfüllt Bedingung (7.10) mit $Z \sim P$.

Im Falle $\mathcal{X} = \mathbb{M}$ konvergiert $(\hat{P}_n)_n$ schwach gegen P in Wahrscheinlichkeit. Denn für beliebige Funktionen $f \in \mathcal{L}^2(P)$ ist

$$\mathbb{E}\, \hat{P}_n(f) \ = \ P(f) \quad \text{und} \quad \mathrm{Var}(\hat{P}_n(f)) \ = \ \big(P(f^2) - P(f)^2\big)/2 \ = \ O(1/n).$$

Insofern bietet Satz 7.14 eine andere Möglichkeit, um uniforme Konsistenz von $\hat{P}_n$ nachzuweisen.

7.6 Straffheit

Im folgenden sei $\mathbb{M}_o$ eine feste Borelmenge in $\mathbb{M}$.

Definition 7.18. Die Zufallsvariable Z heißt *straff auf* $\mathbb{M}_o$, wenn zu jedem $\epsilon > 0$ ein Kompaktum $K \subset \mathbb{M}_o$ existiert, so dass

$$\mathbb{P}(Z \notin K) \ \le \ \epsilon.$$

Mit anderen Worten, es gibt eine σ-kompakte Teilmenge von $\mathbb{M}_o$, die Z mit Wahrscheinlichkeit Eins enthält.

Definition 7.19. Die Folge $(Z_n)_n$ heißt *asymptotisch straff auf* $\mathbb{M}_o$, wenn zu jedem $\epsilon > 0$ ein Kompaktum $K \subset \mathbb{M}_o$ existiert, so dass

$$\limsup_{n \to \infty} \mathbb{P}^*(Z_n \notin U) \ \le \ \epsilon \quad \text{für alle offenen } U \supset K.$$

Im Falle von $\mathbb{M}_o = \mathbb{M}$ spricht man einfach von "Straffheit" anstelle von "Straffheit auf $\mathbb{M}_o$".

Satz 7.20 (Prohorov, LeCam). *Angenommen die Folge $(Z_n)_n$ ist sowohl asymptotisch messbar als auch asymptotisch straff auf $\mathbb{M}_o$. Dann existiert zu jeder Teilfolge von $(Z_n)_n$ eine Teilteilfolge, welche in Verteilung gegen eine auf $\mathbb{M}_o$ straffe Zufallsvariable konvergiert.*

Eine naheliegende Frage ist, ob Straffheit beziehungsweise asymptotische Straffheit sehr spezielle Eigenschaften sind. Aus dem Portmanteau-Theorem folgt direkt, dass die Folge $(Z_n)_n$ asymptotisch straff ist auf $\mathbb{M}_o$, sofern sie in Verteilung gegen eine auf $\mathbb{M}_o$ straffe Zufallsvariable konvergiert. Unter gewissen Voraussetzungen an den Teilraum $\mathbb{M}_o$ ist jede $\mathbb{M}_o$-Zufallsvariable notwendig straff, und es gibt auch ein einfacheres Kriterium für asymptotische Straffheit.

Satz 7.21 (Prohorov, LeCam). *Sei $\mathbb{M}_o$ separabel und vollständig bezüglich der Metrik d.*

(I) *Angenommen zu jeder Teilfolge von $(Z_n)_n$ existiert eine Teilteilfolge, welche in Verteilung gegen eine $\mathbb{M}_o$-Zufallsvariable konvergiert. Dann ist $(Z_n)_n$ asymptotisch straff auf $\mathbb{M}_o$. Sind darüber* *hinaus alle Z_n $\mathbb{M}_o$-Zufallsvariablen, dann gibt es sogar zu jedem $\epsilon > 0$ ein Kompaktum $K \subset \mathbb{M}_o$ mit*

$$\sup_{n\in\mathbb{N}} \mathbb{P}(Z_n \notin K) \ \le\ \epsilon.$$

***(II)** Angenommen zu jedem $\epsilon > 0$ existiert ein Kompaktum $K \subset \mathbb{M}_o$ mit*

$$\limsup_{n\to\infty} \mathbb{P}^*\bigl(Z_n \notin U(K,\epsilon)\bigr) \ \le\ \epsilon.$$

Dann ist die Folge $(Z_n)_n$ asymptotisch straff auf $\mathbb{M}_o$.

Beweis von Satz 7.20. Für $\ell \in \mathbb{N}$ sei $K_\ell \subset \mathbb{M}_o$ kompakt, so dass

$$\limsup_{n\to\infty} \mathbb{P}^*\{Z_n \notin U\} \ \le\ \ell^{-1} \quad \text{für alle offenen } U \supset K_\ell.$$

Erster Schritt: Zu jeder Teilfolge von $(Z_n)_n$ existiert eine Teilteilfolge $(Z_{n(k)})_k$, so dass für beliebige $f \in \mathcal{C}_\mathrm{b}(\mathbb{M})$ der Grenzwert

$$L(f) \ :=\ \lim_{k\to\infty} \mathbb{E}^* f(Z_{n(k)})$$

existiert.

Beweis: Bekanntlich ist $\bigl(\mathcal{C}_\mathrm{b}(\mathbb{M}), \|\cdot\|_{K_\ell}\bigr)$ separabel; siehe auch Abschnitt 7.7. Folglich kann man eine abzählbare Teilmenge $\mathcal{G}_\ell$ von $\mathcal{C}_\mathrm{b}(\mathbb{M})$ mit folgender Eigenschaft konstruieren: Für beliebige $f \in \mathcal{C}_\mathrm{b}(\mathbb{M})$ und $\epsilon > 0$ existiert ein $g \in \mathcal{G}_\ell$ mit

$$\|f - g\|_{K_\ell} \ \le\ \epsilon \quad \text{und} \quad \|g\|_\mathbb{M} \ =\ \|g\|_{K_\ell}.$$

Nach dem zweiten Cantorschen Diagonalverfahren existiert zu jeder Teilfolge von $(Z_n)_n$ eine Teilteilfolge $(Z_{n(k)})_k$, so dass die Folge $\bigl(\mathbb{E}^*\, g(Z_{n(k)})\bigr)_k$ für beliebige g in der abzählbaren Menge $\bigcup_{\ell=1}^\infty \mathcal{G}_\ell$ konvergiert. Doch dann konvergiert sogar $\bigl(\mathbb{E}^*\, f(Z_{n(k)})\bigr)_k$ für beliebige $f \in \mathcal{C}_\mathrm{b}(\mathbb{M})$. Denn zu beliebigem $\ell \in \mathbb{N}$ und $\epsilon > 0$ sei $g \in \mathcal{G}_\ell$ wie oben. Dann ist

$$\bigl|\mathbb{E}^*\, f(Z_n) - \mathbb{E}^*\, g(Z_n)\bigr| \ \le\ \|f - g\|_{U(K_\ell,\delta)} + (2\|f\|_\mathbb{M} + \epsilon)/\ell + o(1)$$

für beliebige $\delta > 0$. Wegen der Kompaktheit von K_ℓ konvergiert $\|f - g\|_{U(K_\ell,\delta)}$ gegen $\|f - g\|_{K_\ell}$ für $\delta \downarrow 0$. Folglich ist

$$\limsup_{j,k\to\infty} \Bigl|\mathbb{E}^*\, f(Z_{n(j)}) - \mathbb{E}^*\, f(Z_{n(k)})\Bigr| \ \le\ 2\epsilon + 2(2\|f\|_\mathbb{M} + \epsilon)/\ell.$$

Für $\ell \to \infty$ und $\epsilon \downarrow 0$ folgt, dass $\bigl(\mathbb{E}^*\, f(Z_{n(k)})\bigr)_k$ eine Cauchyfolge ist, also konvergiert.

Zweiter Schritt: Das Funktional $\mathcal{C}_\mathrm{b}(\mathbb{M}) \ni f \mapsto L(f)$ aus dem ersten Schritt ist eine Linearform, so dass gilt: $L(f) \ge 0$, falls $f \ge 0$, $L(1) = 1$, und

$$L(f) \ \le\ \|f\|_{K_\ell} + \|f\|_\mathbb{M}/\ell \quad \text{für alle } \ell \in \mathbb{N} \text{ und } f \in \mathcal{C}_\mathrm{b}(\mathbb{M}). \tag{7.11}$$

Beweis: Aus Lemma 7.1 folgt, dass L sublinear ist, das heißt, $L(rf) = rL(f)$ und $L(f+g) \le L(f) + L(g)$ für $r \ge 0$ und $f, g \in \mathcal{C}_b(\mathbb{M})$. Denn alle Funktionale $\mathbb{E}^* f(Z_n)$ haben diese Eigenschaft. Außerdem ist offensichtlich $L(f) \ge 0$, falls $f \ge 0$, und $L(1) = 1$. Doch die asymptotische Meßbarkeit von $(Z_n)_n$ impliziert, dass stets $L(f) = -L(-f)$. Zusammen mit der Sublinearität folgt daraus die Linearität von L. Aus der Konstruktion der Mengen K_ℓ folgt, dass stets $L(f) \le \|f\|_{K_\ell} + \|f\|_{\mathbb{M}}/\ell$ für alle $\ell \in \mathbb{N}$; siehe auch den Beweis des ersten Schrittes.

Dritter und letzter Schritt: Es existiert ein Wahrscheinlichkeitsmaß Q auf $\mathcal{B}(\mathbb{M})$, so dass

$$L(f) \;=\; \int f\,dQ \quad \text{für alle } f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M}),$$

und $Q(\mathbb{M} \setminus K_\ell) \le \ell^{-1}$ für alle $\ell \in \mathbb{N}$. Insbesondere konvergiert $(Z_{n_k})_k$ in Verteilung gegen eine $\mathbb{M}$-Zufallsvariable Z mit $\mathcal{L}(Z) = Q$, welche auf $\mathbb{M}_o$ straff ist.

Beweis: Aus (7.11) und dem Satz von Dini folgt, dass

$$L(f) \;=\; \lim_{m\to\infty} L(f_m)$$

für jede Folge $(f_m)_m$ in $\mathcal{C}_{\mathrm{b}}(\mathbb{M})$, welche punktweise monoton wachsend gegen $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M})$ konvergiert. Denn für beliebige $\ell \in \mathbb{N}$ ist

$$\begin{aligned}
\limsup_{m\to\infty} \big|L(f) - L(f_m)\big| \;&\le\; \limsup_{m\to\infty} \|f - f_m\|_{K_\ell} + \|f - f_1\|_{\mathbb{M}}/\ell \quad \text{(nach (7.11))} \\
&=\; \|f - f_1\|_{\mathbb{M}}/\ell \quad \text{(Satz von Dini)} \\
&\to\; 0 \quad \text{für } \ell \to \infty.
\end{aligned}$$

Nach dem Satz von Daniell-Port-Stone aus der Maßtheorie existiert ein Wahrscheinlichkeitsmaß Q auf $\mathcal{B}(\mathbb{M})$ mit $L(f) = \int f\,dQ$ für alle $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M})$; siehe auch die nachfolgende Beweisskizze. Insbesondere ist $Q(M \setminus K_\ell)$ gleich dem Grenzwert $\lim_{k\to\infty} L\big(\min(kd(\cdot, K_\ell), 1)\big) \le \ell^{-1}$. $\square$

Konstruktion von Q im Beweis von Satz 7.20. Für offene Mengen $U \subset \mathbb{M}$ sei

$$Q(U) \;:=\; \sup\big\{L(f) : f \in \mathcal{C}(\mathbb{M}), 0 \le f \le 1_U\big\}.$$

Für beliebige Mengen $S \subset \mathbb{M}$ definiert

$$Q(S) \;:=\; \inf\big\{Q(U) : U \supset S, U \text{ offen}\big\}$$

ein äußeres Wahrscheinlichkeitsmaß auf $\mathbb{M}$. Denn offensichtlich ist $Q(\emptyset) = 0$, $Q(\mathbb{M}) = 1$ und $Q(S) \le Q(T)$ für $S \subset T \subset \mathbb{M}$. Zu zeigen ist die σ-Subadditivität von Q. Sei $S := \bigcup_{k=1}^\infty S_k$ mit beliebigen Mengen $S_k \subset \mathbb{M}$. Für festes $\epsilon > 0$ wählen wir offene Mengen $U_k \supset S_k$ derart, dass $Q(U_k) \le Q(S_k) + 2^{-k}\epsilon$. Dann ist $U := \bigcup_{k=1}^\infty U_k$ eine offene Umgebung von S, also sicherlich $Q(S) \le Q(U)$. Für $k, N \in \mathbb{N}$ sei $h_{k,N} := \min\big(Nd(\cdot, \mathbb{M} \setminus U_k), 1\big)$, eine stetige Funktion mit $0 \le h_{k,N} \le 1_{U_k}$ und $h_{k,N} \uparrow 1_{U_k}$ für $N \to \infty$. Aus den Eigenschaften von L folgt daher, dass für jede stetige Funktion f mit $0 \le f \le 1_U$ gilt:

$$\begin{aligned} L(f) &= \lim_{N\to\infty} L\Big(\min\Big(f, \sum_{k=1}^N h_{k,N}\Big)\Big) \ \le\ \lim_{N\to\infty} \sum_{k=1}^N L(h_{k,N}) \ \le\ \sum_{k=1}^\infty Q(U_k) \\ &\le \sum_{k=1}^\infty Q(S_k) + \epsilon. \end{aligned}$$

Für $L(f) \uparrow Q(U)$ und $\epsilon \downarrow 0$ ergibt sich, dass $Q(S) \le \sum_{k=1}^\infty Q(S_k)$.

Nach dem Satz von Caratheodory aus der Maßtheorie genügt es nun zu zeigen, dass für eine beliebige abgeschlossene Menge $A \subset \mathbb{M}$ gilt:

$$Q(S \cap A) + Q(S \setminus A) \ \le\ Q(S) \quad \text{für alle } S \subset \mathbb{M}. \tag{7.12}$$

Denn die Familie aller Mengen $A \subset \mathbb{M}$ mit Eigenschaft (7.12) ist eine σ-Algebra über $\mathbb{M}$, und Q ist auf dieser Familie σ-additiv.

Zum Nachweis von (7.12) für eine abgeschlossene Menge A halten wir zunächst fest, dass $Q(S \cup T) = Q(S) + Q(T)$ für beliebige Mengen $S, T \subset \mathbb{M}$ mit strikt positivem Abstand $\inf\big\{d(x,y) : x \in S, y \in T\big\}$. Dies kann man leicht aus der Definition von Q ableiten. Nun betrachten wir die abgeschlossenen Mengen $B_k := \{x : d(x,A) \ge k^{-1}\}$ und die Mengen $C_k := B_k \setminus B_{k-1}$ mit $B_0 := \emptyset$. Zwei Mengen C_k und C_ℓ haben strikt positiven Abstand, wenn $|k - \ell| \ge 2$. Daher ist $\sum_{k=1}^\infty Q(C_k)$ gleich

$$\lim_{N\to\infty} \Big(\sum_{j=1}^N Q(C_{2j-1}) + \sum_{j=1}^N Q(C_{2j})\Big) \ =\ \lim_{N\to\infty} \Big(Q\Big(\bigcup_{j=1}^N C_{2j-1}\Big) + Q\Big(\bigcup_{j=1}^N C_{2j}\Big)\Big) \ \le\ 2,$$

also endlich. Dies impliziert, dass für beliebige Indizes $k_o \in \mathbb{N}$ gilt:

$$\begin{aligned} Q(S \cap A) + Q(S \setminus A) &= Q(S \cap A) + Q\Big(S \cap \Big(B_{k_o} \cup \bigcup_{k>k_o} C_k\Big)\Big) \\ &\le Q(S \cap A) + Q(S \cap B_{k_o}) + \sum_{k>k_o} Q(C_k) \\ &= Q\big(S \cap (A \cup B_{k_o})\big) + \sum_{k>k_o} Q(C_k) \\ &\le Q(S) + \sum_{k>k_o} Q(C_k), \end{aligned}$$

und für $k_o \to \infty$ ergibt sich (7.12).

Zu zeigen bleibt, dass tatsächlich $L(f) = \int f\, dQ$ für alle $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M})$. Hierzu genügt es, den Fall $f \ge 0$ zu betrachten. Für beliebige feste $\delta > 0$ ist $f = \delta \sum_{i=1}^\infty h_i$ mit

$$h_i(x) \ :=\ \big(\min(\delta^{-1} f(x), i) - i + 1\big)^+ \ \begin{cases} \le\ 1\big\{f(x) > \delta(i-1)\big\}, \\ \ge\ 1\big\{f(x) > \delta i\big\}, \end{cases}$$

und $h_i \equiv 0$ für hinreichend großes i. Folglich ist

$$L(f) = \delta \sum_{i=1}^{\infty} L(h_i) \begin{cases} \leq \ \delta \displaystyle\sum_{j=0}^{\infty} Q(\{f > \delta j\}) \ \leq \ \int_0^{\infty} Q(\{f > r\})\, dr + \delta \ = \ \int f\, dQ + \delta, \\ \geq \ \delta \displaystyle\sum_{i=1}^{\infty} Q(\{f > \delta i\}) \ \geq \ \int_0^{\infty} Q(\{f > r\})\, dr - \delta \ = \ \int f\, dQ - \delta, \end{cases}$$

denn $\int_{\delta z}^{\delta(z+1)} Q(\{f > r\})\, dr \leq \delta Q(\{f > \delta z\}) \leq \int_{\delta(z-1)}^{\delta z} Q(\{f > r\})\, dr$. Für $\delta \downarrow 0$ folgt die Behauptung. □

Beweis von Satz 7.21. Sei $\{x_i : i \in \mathbb{N}\}$ eine dichte Teilmenge von $\mathbb{M}_o$. Sowohl unter den Voraussetzungen von Teil (I) als auch unter denen von Teil (II) gibt es zu jedem $\delta > 0$ ein $N \in \mathbb{N}$ mit

$$\limsup_{n\to\infty} \mathbb{P}^*\Big(Z_n \notin \bigcup_{i=1}^{N} U(x_i, \delta)\Big) \ \leq \ \delta. \tag{7.13}$$

Nachweis von (7.13) unter den Voraussetzungen von Teil (I). Angenommen, es existieren ein $\delta_o > 0$ und Indizes $n(1) < n(2) < n(3) < \ldots$, so dass

$$\mathbb{P}^*\Big(Z_{n(k)} \notin \bigcup_{i=1}^{k} U(x_i, \delta_o)\Big) \ > \ \delta_o \quad \text{für alle } k \in \mathbb{N}.$$

Nach Voraussetzung kann man die Teilfolge $(Z_{n(k)})_k$ weiter ausdünnen, so dass für neue Indizes $n(1) < n(2) < n(3) < \cdots$ und $m(1) < m(2) < m(3) < \cdots$ gilt:

$$Z_{n(k)} \ \to_{\mathcal{L}} \ Z \quad (k \to \infty)$$

für eine $\mathbb{M}_o$-wertige Zufallsvariable Z, aber

$$\mathbb{P}^*\Big(Z_{n(k)} \notin \bigcup_{i=1}^{m(k)} U(x_i, \delta_o)\Big) \ > \ \delta_o \quad \text{für alle } k \in \mathbb{N}.$$

Nach dem Portmanteau-Theorem ist

$$\delta_o \ \leq \ \limsup_{k\to\infty} \mathbb{P}^*\Big(Z_{n(k)} \notin \bigcup_{i=1}^{m(k)} U(x_i, \delta_o)\Big) \ \leq \ \mathbb{P}\Big(Z \notin \bigcup_{k=1}^{N} U(x_i, \delta_o)\Big)$$

für jedes feste $N \in \mathbb{N}$. Doch wegen $\mathbb{M}_o \subset \bigcup_{k=1}^{\infty} U(x_i, \delta_o)$ ist die rechte Seite strikt kleiner als δ_o für hinreichend großes N, ein Widerspruch zu unserer Annahme.

Nachweis von (7.13) unter den Voraussetzungen von Teil (II). Für beliebiges $\delta > 0$ fixieren wir eine Zahl $0 < \epsilon < \delta$ und ein Kompaktum $K \subset \mathbb{M}$ derart, dass $\limsup_{n\to\infty} \mathbb{P}^*\big(Z_n \notin U(K, \epsilon)\big) \leq \epsilon$. Doch $K \subset \bigcup_{i=1}^{N} U(x_i, \delta - \epsilon)$ für hinreichend großes $N \in \mathbb{N}$, und dies impliziert, dass $U(K, \epsilon) \subset \bigcup_{i=1}^{N} U(x_i, \delta)$.

Anwendung von (7.13). Nun können wir die Teile (I) und (II) gemeinsam abhandeln. Für ein beliebiges festes $\epsilon > 0$ und für $k \in \mathbb{N}$ sei also $N(k) \in \mathbb{N}$ so gewählt, dass

$$\limsup_{n\to\infty} \mathbb{P}^*(Z_n \notin A_k) \ \le\ 2^{-k}\epsilon \quad \text{mit} \quad A_k \ :=\ \bigcup_{i=1}^{N(k)} B(x_i, 2^{-k}\epsilon).$$

Die Menge

$$K \ :=\ \bigcap_{k=1}^{\infty} A_k$$

ist abgeschlossen und nichtleer, denn alle A_k sind abgeschlossen, und $x_1 \in K$. Außerdem ist $K \subset \bigcap_{k\in\mathbb{N}} B(\mathbb{M}_o, 2^{-k}\epsilon) = \mathbb{M}_o$, da $\mathbb{M}_o$ vollständig und somit abgeschlossen ist. Die folgende Behauptung, die wir am Ende begründen, impliziert Kompaktheit von K:

(7.14) Zu jeder Folge $(y_m)_m$ von Punkten $y_m \in \bigcap_{k=1}^m A_k$ existiert eine Teilfolge, die gegen einen Punkt in K konvergiert.

Nun sei U eine beliebige offene Umgebung von K. Dann kann man aus (7.14) ableiten, dass $\bigcap_{k=1}^L A_k \subset U$ für ein hinreichend großes $L \in \mathbb{N}$. Folglich ist

$$\limsup_{n\to\infty} \mathbb{P}^*(Z_n \notin U) \ \le\ \sum_{k=1}^{L} \limsup_{n\to\infty} \mathbb{P}^*(Z_n \notin A_k) \ \le\ \epsilon.$$

In dem Spezialfall, dass in (I) alle Z_n $\mathbb{M}_o$-Zufallsvariablen sind, kann man die $N(k)$ so wählen, dass

$$\mathbb{P}(Z_n \notin A_k) \ \le\ 2^{-k}\epsilon \quad \text{für alle } n, k \in \mathbb{N},$$

und folglich ist

$$\mathbb{P}(Z_n \notin K) \ \le\ \epsilon \quad \text{für beliebige } n \in \mathbb{N}.$$

Beweis von (7.14): Aufgrund der Abgeschlossenheit aller A_k liegt jeder Häufungspunkt von $(y_m)_m$ in K. Man muss also nur nachweisen, dass es eine konvergente Teilfolge gibt. Sei $\mathbb{N}(0) := \mathbb{N}$. Für $k = 1, 2, 3, \ldots$ wählen wir nun induktiv $i(k) \in \{1, 2, \ldots, N(k)\}$, so dass die Menge

$$\mathbb{N}(k) \ :=\ \bigl\{m \in \mathbb{N}(k-1) : y_m \in B(x_{i(k)}, 2^{-k}\epsilon)\bigr\}$$

unendlich ist. Ferner sei $m(k)$ das k-tkleinste Element von $\mathbb{N}(k)$. Dies liefert eine Teilfolge $(y_{m(k)})_k$, so dass gilt:

$$y_{m(j)} \ \in\ B(x_{i(k)}, 2^{-k}\epsilon) \quad \text{für } j \ge k.$$

Insbesondere ist diese Folge eine Cauchy-Folge mit $d(y_{m(k)}, \mathbb{M}_o) \le 2^{-k}\epsilon$ für alle k. Wegen der Vollständigkeit von $\mathbb{M}_o$ konvergiert sie gegen einen Punkt $y \in \mathbb{M}_o$. $\square$

Straffheit und der Satz von Stone-Weierstrass. Schließlich sei noch erwähnt, dass man im Falle von Straffheit mithilfe des Satzes von Stone-Weierstrass die asymptotische Messbarkeit und Verteilungskonvergenz oft einfacher nachweisen kann. Auch die Gleichheit in Verteilung von $\mathbb{M}$-Zufallsvariablen wird dadurch vereinfacht.

Satz 7.22. *Sei $\mathcal{F}$ eine Teilfamilie von $\mathcal{C}_{\mathrm{b}}(\mathbb{M})$ mit folgenden Eigenschaften:*

Mit $f, g \in \mathcal{F}$ ist auch $fg \in \mathcal{F}$;
für beliebige verschiedene $x, y \in \mathbb{M}_o$ existiert ein $f \in \mathcal{F}$ mit $f(x) \neq f(y)$.

***(a)** Seien $Z, \tilde{Z}$ auf $\mathbb{M}_o$ straffe Zufallsvariablen mit*

$$\mathbb{E}\, f(Z) \;=\; \mathbb{E}\, f(\tilde{Z}) \quad \text{für alle } f \in \mathcal{F}. \tag{7.15}$$

Dann ist $\mathcal{L}(Z) = \mathcal{L}(\tilde{Z})$.

***(b)** Angenommen, die Folge $(Z_n)_{n\in\mathbb{N}}$ ist asymptotisch straff auf $\mathbb{M}_o$. Ferner sei*

$$\lim_{n\to\infty} \big(\mathbb{E}^*\, f(Z_n) - \mathbb{E}_*\, f(Z_n)\big) \;=\; 0 \quad \text{für alle } f \in \mathcal{F}. \tag{7.16}$$

Dann ist die Folge $(Z_n)_n$ asymptotisch messbar. Zusätzlich existiere

$$L(f) \;:=\; \lim_{n\to\infty} \mathbb{E}^*\, f(Z_n)$$

für alle $f \in \mathcal{F}$. Dann konvergiert $(Z_n)_n$ in Verteilung gegen eine auf $\mathbb{M}_o$ straffe Zufallsvariable Z.

Beweis. Sei $\mathcal{G}$ der von $\mathcal{F} \cup \{1\}$ aufgespannte Vektorraum, und sei $\overline{\mathcal{G}}$ sein Abschluss bezüglich $\|\cdot\|_{\mathbb{M}}$. Dann überträgt sich Eigenschaft (7.15) von $\mathcal{F}$ auf $\overline{\mathcal{G}}$. Aus Lemma 7.1 kann man ableiten, dass sich auch Eigenschaft (7.16) von $\mathcal{F}$ auf $\overline{\mathcal{G}}$ überträgt.

Zu beliebigem $\epsilon > 0$ wählen wir nun ein Kompaktum $K \subset \mathbb{M}_o$, so dass

$$\mathbb{P}(Z \notin K), \mathbb{P}(\tilde{Z} \notin K) \;\le\; \epsilon$$

in Teil (a), beziehungsweise

$$\limsup_{n\to\infty} \mathbb{P}^*\big(Z_n \notin U(K,\delta)\big) \;\le\; \epsilon \quad \text{für beliebige } \delta > 0$$

in Teil (b). Aus dem Satz von Stone-Weierstrass (Satz 7.34) folgt, dass zu jeder Funktion $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M})$ ein $g \in \overline{\mathcal{G}}$ existiert, so dass

$$\|f - g\|_K \;=\; 0 \quad \text{und} \quad \|g\|_{\mathbb{M}} \;\le\; \|f\|_{\mathbb{M}}.$$

Somit ist

$$\big|\mathbb{E}\, f(Z) - \mathbb{E}\, f(\tilde{Z})\big| \;\le\; \|f\|_{\mathbb{M}}\, \mathbb{P}(Z \notin K) + \|f\|_{\mathbb{M}}\, \mathbb{P}(\tilde{Z} \notin K) \;\le\; 2\|f\|_{\mathbb{M}}\, \epsilon,$$

und für $\delta > 0$ gilt:

$$\begin{aligned}
&\limsup_{n\to\infty} \big(\mathbb{E}^* f(Z_n) - \mathbb{E}_* f(Z_n)\big) \\
&\quad \le \quad 2\|f-g\|_{U(K,\delta)} + 2\|f\|_{\mathbb{M}} \limsup_{n\to\infty} \mathbb{P}^*\big(Z_n \notin U(K,\delta)\big) \\
&\quad \le \quad 2\|f-g\|_{U(K,\delta)} + 2\|f\|_{\mathbb{M}}\,\epsilon \\
&\quad \to \quad 2\|f\|_{\mathbb{M}}\,\epsilon \quad (\delta \downarrow 0).
\end{aligned}$$

Dies beweist Teil (a) und die asymptotische Messbarkeit von $(Z_n)_n$ in Teil (b).

Angenommen, in Teil (b) existiert der Grenzwert $L(f) := \lim_{n\to\infty} \mathbb{E}^* f(Z_n)$ für alle $f \in \mathcal{F}$. Aus Prohorovs Theorem und Teil (a) folgt die Existenz einer Zufallsvariablen $Z \in \mathbb{M}$, so dass zu jeder Teilfolge von $(Z_n)_n$ eine Teilteilfolge existiert, welche in Verteilung gegen Z konvergiert. Mit einem einfachen Widerspruchsbeweis zeigt man nun, dass dann auch die Folge $(Z_n)_n$ selbst in Verteilung gegen Z konvergiert. □

Dieser Satz vereinheitlicht eine ganze Reihe von Techniken, um schwache Konvergenz nachzuweisen. In manchen Fällen hat man noch das Glück, dass die Konvergenz von $\mathbb{E}^* f(Z_n)$ und $\mathbb{E}_* f(Z_n)$ für beliebige $f \in \mathcal{F}$ bereits die asymptotische Straffheit von $(Z_n)_n$ impliziert.

Beispiel 7.23 (Charakteristische Funktionen). Sei $\mathbb{M} = \mathbb{R}^d$ und $\mathcal{F}$ die Menge aller Funktionen

$$f(x) \ = \ \alpha \sin(v^\top x) + \beta \cos(v^\top x)$$

mit $\alpha, \beta \in \mathbb{R}$ und $v \in \mathbb{R}^d$. Diese Klasse $\mathcal{F}$ erfüllt die Voraussetzungen von Satz 7.22. Die Verteilung einer $\mathbb{R}^d$-wertigen Zufallsvariable Z wird also durch ihre *charakteristische Funktion*,

$$\mathbb{R}^d \ni v \ \mapsto \ \mathbb{E} \exp\big(\sqrt{-1}\, v^\top Z\big) \in \mathbb{C},$$

eindeutig festgelegt.

Betreffend Konvergenz in Verteilung betrachten wir der Einfachheit halber nur Zufallsvariablen: Für beliebige $v \in \mathbb{R}^d$ existiere der Grenzwert $\ell(v) := \lim_{n\to\infty} \mathbb{E}\cos(v^\top Z_n)$, wobei ℓ im Nullpunkt stetig ist. Dann ist $(Z_n)_n$ asymptotisch straff. Denn für beliebige Vektoren $v \in \mathbb{R}^d$ ist

$$C(v) := \int_0^1 \ell(rv)\, dr \ = \ \lim_{n\to\infty} \int_0^1 \mathbb{E}\cos(rv^\top Z_n)\, dr \ = \ \lim_{n\to\infty} \mathbb{E}\, \frac{\sin(v^\top Z_n)}{v^\top Z_n},$$

wobei $\sin(0)/0 := 1$. Doch $\sin(x)/x \le 1\{|x| < 2\} + 1\{|x| \ge 2\}/2 = 1 - 1\{|x| \ge 2\}/2$, weshalb

$$\limsup_{n\to\infty} \mathbb{P}\big(|v^\top Z_n| \ge 2\big) \ \le \ 2\big(1 - C(v)\big).$$

Mit einer Orthonormalbasis $u_1, u_2, \ldots, u_d$ des $\mathbb{R}^d$ ergibt sich also für $R > 0$, dass

$$\limsup_{n\to\infty} \mathbb{P}\Big(\max_{j=1,\ldots,d} |u_j^\top Z_n| \ge 2R\Big) \ \le \ 2\sum_{i=1}^d \big(1 - C(R^{-1}u_i)\big) \ \to \ 0 \quad (R \to 0),$$

da nach Voraussetzung $\lim_{v\to 0} C(v) = \ell(0) = 1$. Aus diesen Betrachtungen und Satz 7.22 ergibt sich folgender Stetigkeitssatz für charakteristische Funktionen:

Korollar 7.24 (Lévy-Cramér). *Sei $(Z_n)_n$ eine Folge von $\mathbb{R}^d$-wertigen Zufallsvariablen derart, dass für beliebige $v \in \mathbb{R}^d$ der Grenzwert*

$$\vec{\ell}(v) \ := \ \lim_{n\to\infty} \mathbb{E}\exp\bigl(\sqrt{-1}\, v^\top Z_n\bigr) \ \in \ \mathbb{C}$$

existiert. Ist $\vec{\ell}$ im Nullpunkt stetig, dann ist $\vec{\ell}(v) = \mathbb{E}\exp\bigl(\sqrt{-1}\, v^\top Z\bigr)$ für eine $\mathbb{R}^d$-wertige Zufallsvariable, und $Z_n \to_{\mathcal{L}} Z$.

Beispiel 7.25 (Laplace-Transformierte). Sei $\mathbb{M} = [0,\infty)$ und $\mathcal{F}$ die Menge aller Funktionen

$$f(x) \ = \ \exp(-\lambda x)$$

mit $\lambda \ge 0$. Diese Familie erfüllt die Voraussetzungen von Satz 7.22. Die Verteilung einer $[0,\infty)$-wertigen Zufallsvariable Z wird also durch ihre *Laplace-Transformierte*,

$$[0,\infty) \ni \lambda \ \mapsto \ \mathbb{E}\exp(-\lambda Z),$$

eindeutig festgelegt.

Angenommen, für beliebige $\lambda \ge 0$ existiert der Grenzwert $\ell(\lambda) := \lim_{n\to\infty} \mathbb{E}_* \exp(-\lambda Z_n)$, und ℓ ist im Nullpunkt stetig. Dann ist $(Z_n)_n$ asymptotisch straff. Denn für $\lambda > 0$ und $x \ge 0$ ist $\exp(-\lambda x) \le 1\{x < \lambda^{-1}\} + e^{-1}1\{x \ge \lambda\} = 1 - (1-e^{-1})1\{x \ge \lambda^{-1}\}$, so dass

$$\mathbb{E}_* \exp(-\lambda Z_n) \ \le \ 1 - (1-e^{-1})\, \mathbb{P}^*(Z_n \ge \lambda^{-1}),$$

also

$$\limsup_{n\to\infty} \mathbb{P}^*(Z_n \ge \lambda^{-1}) \ \le \ \frac{1-\ell(\lambda)}{1-e^{-1}}.$$

Dies liefert einen Stetigkeitssatz für die Laplace-Transformierte von nichtnegativen Zufallsvariablen:

Korollar 7.26. *Sei $(Z_n)_n$ eine Folge von $[0,\infty)$-wertigen Zufallselementen derart, dass für beliebige $\lambda \ge 0$ der Grenzwert*

$$\ell(\lambda) \ := \ \lim_{n\to\infty} \mathbb{E}^* \exp(-\lambda Z_n) \ = \ \lim_{n\to\infty} \mathbb{E}_* \exp(-\lambda Z_n)$$

existiert. Ist ℓ im Nullpunkt stetig, dann ist $\ell(\lambda) = \mathbb{E}\exp(-\lambda Z)$ für eine $[0,\infty)$-wertige Zufallsvariable Z, und $Z_n \to_{\mathcal{L}} Z$.

Beispiel 7.27. Sei $\mathbb{M} = \ell_\infty(\mathcal{T})$, versehen mit der Supremumsnorm $\|\cdot\|_{\mathcal{T}}$. Dann erfüllt die Menge $\mathcal{F}$ aller Funktionen

$$f(x) \ := \ f_o\bigl((x(t))_{t\in\mathcal{T}_o}\bigr)$$

mit endlichen Mengen $\mathcal{T}_o \subset \mathcal{T}$ und beschränkten stetigen Funktionen $f_o : \mathcal{T}_o \to \mathbb{R}$ die Voraussetzungen von Satz 7.22. Hier ist asymptotische Straffheit schwieriger nachzuweisen; siehe Abschnitt 7.7.

7.7 Funktionale Grenzwertsätze

Hier betrachten wir die Menge $\ell_\infty(\mathcal{T})$ aller beschränkten Funktionen $x : \mathcal{T} \to \mathbb{R}$ versehen mit der Supremumsnorm $\|\cdot\| = \|\cdot\|_{\mathcal{T}}$. Das erste Resultat charakterisiert die separablen Teilmengen von $\ell_\infty(\mathcal{T})$.

Lemma 7.28. *Sei $\mathbb{M}_o$ eine separable Teilmenge von $\ell_\infty(\mathcal{T})$. Dann existiert eine Pseudometrik ρ auf $\mathcal{T}$, so dass gilt: $(\mathcal{T}, \rho)$ ist präkompakt, und alle Funktionen in $\mathbb{M}_o$ sind gleichmäßig stetig bezüglich ρ.*

Ein pseudometrischer Raum $(\mathcal{T}, \rho)$ heißt *präkompakt* (oder *totalbeschränkt*), wenn man für beliebige $\delta > 0$ die Menge $\mathcal{T}$ mit endlich vielen Kugeln mit Radius δ bezüglich ρ überdecken kann. Mit anderen Worten, für beliebige $\delta > 0$ ist die Überdeckungszahl $N(\delta, \mathcal{T}, \rho)$ endlich; siehe Abschnitt 6.1.

Beweis von Lemma 7.28. Sei $\{x_k : k \in \mathbb{N}\}$ eine dichte Teilmenge von $\mathbb{M}_o$. Dann definieren wir

$$\rho(s,t) \ = \ \sum_{k\in\mathbb{N}} \min\bigl(|x_k(s) - x_k(t)|, 2^{-k}\bigr).$$

Man kann zeigen, dass ρ eine Pseudometrik auf $\mathcal{T}$ ist; siehe unter anderem Aufgabe 7.9. Ferner sind alle Funktionen x_k gleichmäßig stetig bezüglich ρ, denn aus $\rho(s,t) \le \delta < 2^{-k}$ folgt notwendig, dass $|x_k(s) - x_k(t)| \le \delta$. Dann sind aber auch alle Funktionen in $\mathbb{M}_o$ gleichmäßig stetig bezüglich ρ; siehe Aufgabe 7.10.

Zu zeigen ist nun, dass $(\mathcal{T}, \rho)$ präkompakt ist. Für $\epsilon > 0$ wählen wir $N \in \mathbb{N}$ mit $2^{-N} \le \epsilon/2$. Dann ist $\rho(s,t) \le \sum_{k\le N} |x_k(s) - x(t)| + \epsilon/2$. Die Menge

$$\bigl\{(x_k(t))_{k\le N} : t \in \mathcal{T}\bigr\}$$

ist eine beschränkte Teilmenge des $\mathbb{R}^N$, da alle Funktionen x_k beschränkt sind. Folglich existiert eine endliche Teilmenge $\mathcal{T}_o$ von $\mathcal{T}$, so dass

$$\min_{s\in\mathcal{T}_o} \sum_{k\le N} |x_k(s) - x_k(t)| \ \le \ \epsilon/2 \quad \text{für alle } t \in \mathcal{T}.$$

Somit ist $\rho(t, \mathcal{T}_o) \le \epsilon$ für beliebige $t \in \mathcal{T}$. □

Im folgenden sei ρ eine feste Pseudometrik auf $\mathcal{T}$, so dass $(\mathcal{T}, \rho)$ präkompakt ist, und $\mathcal{T}_*$ sei eine dichte Teilmenge von $\mathcal{T}$ bezüglich ρ. Ferner sei $\mathcal{C}_{\mathrm{u}}(\mathcal{T}, \rho)$ die Menge aller bezüglich ρ gleichmäßig stetigen Funktionen auf $\mathcal{T}$. Mit dem Stetigkeitsmodul

$$\omega(x,\delta) = \omega(x,\delta \,|\, \rho) \ := \sup_{s,t\in\mathcal{T}:\rho(s,t)<\delta} |x(s) - x(t)|$$

einer Funktion $x \in \ell_\infty(\mathcal{T})$ kann man schreiben:

$$\mathcal{C}_{\mathrm{u}}(\mathcal{T}, \rho) \ = \ \Bigl\{x \in \ell_\infty(\mathcal{T}) : \lim_{\delta\downarrow 0} \omega(x,\delta) = 0\Bigr\}.$$

Lemma 7.29. *Der Raum $\mathcal{C}_{\mathrm{u}}(\mathcal{T}, \rho)$ ist vollständig und separabel bezüglich $\|\cdot\|$.*

Beweis von Lemma 7.29. Wir konstruieren zunächst eine spezielle Folge von Abbildungen, die wir mehrfach verwenden werden: Für $k \in \mathbb{N}$ sei $\mathcal{T}_k$ eine endliche Teilmenge von $\mathcal{T}_*$ derart, dass $\sup_{t\in\mathcal{T}} \rho(t, \mathcal{T}_k) \le (2k)^{-1}$, und es seien $\mathcal{T}_1 \subset \mathcal{T}_2 \subset \mathcal{T}_3 \subset \cdots$. Nun definieren wir

$$\lambda_k(t,u) \ := \ \bigl(1 - k\rho(t,u)\bigr)^+ \Big/ \sum_{v\in\mathcal{T}_k} \bigl(1 - k\rho(t,v)\bigr)^+$$

für $t \in \mathcal{T}$ und $u \in \mathcal{T}_k$. Man beachte, dass die Summe im Nenner größer oder gleich $1/2$ ist. Offensichtlich ist

$$0 \ \le \ \lambda_k(\cdot,u) \ \le \ 1\{\rho(\cdot,u) < k^{-1}\} \quad \text{und} \quad \sum_{u\in\mathcal{T}_k} \lambda_k(\cdot,u) \ \equiv \ 1.$$

Ferner wird in Aufgabe 7.11 gezeigt, dass die Funktionen $\lambda_k(\cdot,u)$, $u \in \mathcal{T}_k$ lipschitzstetig bezüglich ρ sind. Für $x \in \ell_\infty(\mathcal{T})$ oder $x \in \ell_\infty(\mathcal{T}_k) = \mathbb{R}^{\mathcal{T}_k}$ sei nun

$$\Pi_k x \ := \ \sum_{u\in\mathcal{T}_k} x(u)\lambda_k(\cdot,u) \ \in \ \mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho).$$

Dann ist Π_k eine lineare Abbildung von $\ell_\infty(\mathcal{T})$ beziehungsweise $\ell_\infty(\mathcal{T}_k)$ nach $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$, so dass gilt:

$$\begin{aligned} \|\Pi_k x\| &\le \|x\|_{\mathcal{T}_k} && \text{für alle } x \in \ell_\infty(\mathcal{T}) \cup \ell_\infty(\mathcal{T}_k), \\ \|x - \Pi_k x\| &\le \omega(x, k^{-1}) && \text{für alle } x \in \ell_\infty(\mathcal{T}). \end{aligned} \tag{7.17}$$

Letztere Ungleichung folgt aus

$$\begin{aligned} |x(t) - \Pi_k x(t)| \ &= \ \Bigl| \sum_{u\in\mathcal{T}_k} (x(t) - x(u))\lambda_k(t,u) \Bigr| \\ &\le \ \sum_{u\in\mathcal{T}_k} |x(t) - x(u)|\lambda_k(t,u) \\ &\le \ \sum_{u\in\mathcal{T}_k} \omega(x,k^{-1})\lambda_k(t,u) \\ &= \ \omega(x,k^{-1}). \end{aligned}$$

Nun zum eigentlichen Beweis von Lemma 7.29: Die Vollständigkeit von $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$ ist Gegenstand von Aufgabe 7.10. Aus (7.17) folgt, dass

$$\bigcup_{k\in\mathbb{N}} \bigl\{ \Pi_k y : y \in \mathbb{Q}^{\mathcal{T}_k} \bigr\}$$

eine abzählbare, dichte Teilmenge von $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$ darstellt. Denn für beliebige $x \in \mathcal{C}_{\mathrm{u}}(\mathcal{T})$ und $k \in \mathbb{N}$ ist

$$\inf_{y\in\mathbb{Q}^{\mathcal{T}_k}} \|x - \Pi_k y\| \ \le \ \omega(x,k^{-1}) + \inf_{y\in\mathbb{Q}^{\mathcal{T}_k}} \|x - y\|_{\mathcal{T}_k} \ = \ \omega(x,k^{-1}) \ \to \ 0 \quad (k \to \infty).$$

□

Lemma 7.30. *Ein $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$-Zufallselement Z ist genau dann messbar, wenn $Z(t)$ für beliebige $t \in \mathcal{T}_*$ messbar ist. Die Verteilung einer $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$-wertigen Zufallsvariable Z wird eindeutig bestimmt durch ihre endlichdimensionalen Randverteilungen,*

$$\mathcal{L}\big((Z(t))_{t\in\mathcal{T}_o}\big), \quad \mathcal{T}_o \subset \mathcal{T} \text{ endlich.}$$

Beweis. Sei $\mathcal{D}$ die Menge aller Durchschnitte von endlich vielen abgeschlossenen Kugeln in $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$. Dann ist $\mathcal{D}$ ein erzeugendes System der Borelmengen von $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$. Denn aufgrund der Separabilität von $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$ kann man jede offene Teilmenge hiervon als abzählbare Vereinigung abgeschlossener Kugeln darstellen. Doch jede Menge $D \in \mathcal{D}$ hat folgende Form: Für ein $m \in \mathbb{N}$, Funktionen $x_1,\ldots,x_m \in \mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$ und Zahlen $\eta_1,\ldots,\eta_m \geq 0$ ist

$$\begin{aligned} D &= \bigcap_{i=1}^{m} \big\{y \in \mathcal{C}_{\mathrm{u}}(\mathcal{T}) : \|y - x_i\| \leq \eta_i\big\} \\ &= \bigcap_{i=1}^{m}\bigcap_{k=1}^{\infty} \big\{y \in \mathcal{C}_{\mathrm{u}}(\mathcal{T}) : \|y - x_i\|_{\mathcal{T}_k} \leq \eta_i\big\} \\ &= \bigcap_{k\in\mathbb{N}} D_k \end{aligned}$$

mit $D_k := \bigcap_{i=1}^m \big\{y \in \mathcal{C}_{\mathrm{u}}(\mathcal{T}) : \|y - x_i\|_{\mathcal{T}_k} \leq \eta_i\big\}$ und den Mengen $\mathcal{T}_k \subset \mathcal{T}_*$ aus dem Beweis von Lemma 7.29. Die Indikatorfunktion $1\{Z \in D_k\}$ ist eine messbare Funktion von $(Z(t))_{t\in\mathcal{T}_k}$. Daher ist Z genau dann messbar, wenn $Z(t)$ für alle $t \in \mathcal{T}_*$ messbar ist. In diesem Falle ist $\mathbb{P}(Z \in D) = \lim_{k\to\infty} \mathbb{P}(Z \in D_k)$, da $D_1 \supset D_2 \supset D_3 \supset \cdots$. Also wird $\mathbb{P}(Z \in D)$ durch die Randverteilungen $\mathcal{L}\big((Z(t))_{t\in\mathcal{T}_k}\big)$, $k \in \mathbb{N}$, bestimmt. Da $\mathcal{D}$ durchschnittstabil ist, bestimmen diese Verteilungen sogar $\mathcal{L}(Z)$. □

Im Hinblick auf Abschnitt 7.6 ist eine naheliegende Frage, wie präkompakte Teilmengen von $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$ aussehen. Dies wird durch den Satz von Arzelà-Ascoli aus der Funktionalanalysis beantwortet. Es sei nur darauf hingewiesen, dass folgende Eigenschaft von $(Z_n)_n$ in engem Zusammenhang mit asymptotischer Straffheit steht.

Definition 7.31. Eine Folge von Folge $(Z_n)_{n\in\mathbb{N}}$ von $\ell_\infty(\mathcal{T})$-Zufallselementen heißt *asymptotisch stochastisch gleichstetig (bezüglich ρ)*, wenn

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathbb{P}^*\big(\omega(Z_n,\delta) > \epsilon\big) \;=\; 0 \quad \text{für beliebige } \epsilon > 0.$$

Nun kommen wir zum Hauptsatz dieses Abschnitts.

Satz 7.32. *Sei $(Z_n)_n$ eine Folge von $\ell_\infty(\mathcal{T})$-Zufallselementen. Dann sind folgende Aussagen äquivalent:*

(a) *$(Z_n)_n$ ist stochastisch gleichstetig, und für beliebige endliche Mengen $\mathcal{T}_o \subset \mathcal{T}_*$ konvergiert $(Z_n(t))_{t\in\mathcal{T}_o}$ in Verteilung gegen eine $\ell_\infty(\mathcal{T}_o)$-Zufallsvariable $Y_{(\mathcal{T}_o)}$.*

(b) *$(Z_n)_n$ konvergiert in Verteilung gegen eine $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$-wertige Zufallsvariable Z.*

Unter diesen Bedingungen ist $\mathcal{L}\big((Z(t))_{t\in\mathcal{T}_o}\big) = \mathcal{L}(Y_{(\mathcal{T}_o)})$ *für alle endlichen Teilmengen* $\mathcal{T}_o$ *von* $\mathcal{T}_*$.

Beweis. Daß Bedingung (b) Bedingung (a) impliziert, folgt aus dem Satz von der stetigen Abbildung und dem Portmanteau-Theorem. Denn $x \mapsto (x(t))_{t\in\mathcal{T}_o}$ und $x \mapsto \omega(x,\delta)$ sind stetige Abbildungen auf $\ell_\infty(\mathcal{T})$, wobei

$$\lim_{\delta\downarrow 0} \mathbb{P}\big(\omega(Z,\delta) \ge \epsilon\big) \;=\; 0 \quad \text{für beliebige } \epsilon > 0.$$

Umgekehrt sei Bedingung (a) erfüllt. Zu beliebigem $\epsilon > 0$ existiert ein $k \in \mathbb{N}$ mit

$$\limsup_{n\to\infty} \mathbb{P}^*\big(\omega(Z_n, k^{-1}) \ge \epsilon\big) \;\le\; \epsilon/2.$$

Mit $\mathcal{T}_k$ und Π_k aus dem Beweis von Lemma 7.29 ist also

$$\limsup_{n\to\infty} \mathbb{P}^*\big(\|Z_n - \Pi_k Z_n\| \ge \epsilon\big) \;\le\; \epsilon/2;$$

siehe (7.17). Doch $(Z_n(t))_{t\in\mathcal{T}_k}$ konvergiert in Verteilung gegen eine $\ell_\infty(\mathcal{T}_k)$-Zufallsvariable Y_k. Insbesondere existiert eine reelle Zahl $R > 0$ mit

$$\mathbb{P}\big(\|Y_k\|_{\mathcal{T}_k} \ge R\big) \;\le\; \epsilon/2.$$

Doch dann ist $K := \big\{\Pi_k y : \|y\|_{\mathcal{T}_k} \le R\big\}$ eine kompakte Menge in $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$ aufgrund der Stetigkeit von Π_k, und

$$\begin{aligned}
\mathbb{P}^*\big(Z_n \notin U(K,\epsilon)\big) \;&\le\; \mathbb{P}^*\big(\|Z_n - \Pi_k Z_n\| \ge \epsilon\big) + \mathbb{P}^*\big(\Pi_k Z_n \notin K\big) \\
&\le\; \mathbb{P}^*\big(\|Z_n - \Pi_k Z_n\| \ge \epsilon\big) + \mathbb{P}^*\big(\|Z_n\|_{\mathcal{T}_k} \ge R\big) \\
&\le\; \epsilon + o(1),
\end{aligned}$$

gemäß dem Portmanteau-Theorem. Nach Lemma 7.29 und Teil (II) von Satz 7.21 ist also $(Z_n)_n$ asymptotisch straff auf $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$. Doch nun folgt Aussage (b) aus Satz 7.22, angewandt auf die Menge $\mathcal{F}$ aller Funktionen der Form

$$f(x) \;=\; g\big((x(t))_{t\in\mathcal{T}_o}\big)$$

mit endlichen Teilmengen $\mathcal{T}_o$ von $\mathcal{T}_*$ und beschränkten stetigen Funktionen g auf $\mathbb{R}^{\mathcal{T}_o}$. □

Skizze eines anderen Beweises von “(a) $\Longrightarrow$ *(b)”.* Man kann die Resultate aus Abschnitt 7.6 wie folgt umgehen. Die Limites $Y_k \in \ell_\infty(\mathcal{T}_k)$ von $(Z_n(t))_{t\in\mathcal{T}_k}$, $k \in \mathbb{N}$, erfüllen folgende Bedingung:

$$\mathcal{L}(Y_k) \;=\; \mathcal{L}\big((Y_\ell(t))_{t\in\mathcal{T}_k}\big) \quad \text{für } 1 \le k < \ell.$$

Nach dem Existenzsatz von Kolmogorov existiert ein stochastischer Prozess Z auf $\mathcal{T}_{**} := \bigcup_k \mathcal{T}_k$ mit $\mathcal{L}\big((Z(t))_{t\in\mathcal{T}_k}\big) = \mathcal{L}(Y_k)$ für alle $k \in \mathbb{N}$. Mit den Stetigkeitsmoduli $\omega_k(\cdot,\cdot)$ und $\omega_{**}(\cdot,\cdot)$ von

Funktionen auf $\mathcal{T}_k$ bzw. $\mathcal{T}_{**}$ gilt: $\omega_k(\cdot,\cdot) \uparrow \omega_{**}(\cdot,\cdot)$ für $k \to \infty$, weshalb

$$\begin{aligned}
\mathbb{P}\big(\omega_{**}(Z,\delta) > \epsilon\big) &= \lim_{k\to\infty} \mathbb{P}\big(\omega_k(Z,\delta) > \epsilon\big) \\
&\le \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\big(\omega_k(Z_n,\delta) \ge \epsilon\big) \\
&\le \limsup_{n\to\infty} \mathbb{P}^*\big(\omega(Z_n,\delta) \ge \epsilon\big) \\
&\to 0 \quad (\delta \downarrow 0).
\end{aligned}$$

Also hat Z fast sicher gleichmäßig stetige Pfade auf $\mathcal{T}_{**}$ bezüglich ρ. Zusammen mit Lemma 7.30 kann man also folgern, dass eine $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$-Zufallsvariable Z existiert, so dass $\mathcal{L}\big((Z(t))_{t\in\mathcal{T}_k}\big) = \mathcal{L}(Y_k)$ für alle k. Für $f \in \mathcal{C}_{\mathrm{Lip}}(\ell_\infty(\mathcal{T}),[0,1])$ mit Lipschitzkonstante L gilt demnach:

$$\begin{aligned}
\big|\mathbb{E}^* f(Z_n) - \mathbb{E} f(Z)\big| &\le \mathbb{E}^*\big|f(Z_n) - f(\Pi_k Z_n)\big| + \mathbb{E}\big|f(Z) - f(\Pi_k Z)\big| \\
&\quad + \big|\mathbb{E}^* f(\Pi_k Z_n) - \mathbb{E} f(\Pi_k Z)\big| \\
&= \mathbb{E}^* |f(Z_n) - f(\Pi_k Z_n)| + \mathbb{E}\,|f(Z) - f(\Pi_k Z)| + o(1) \\
&\le 2L\epsilon + \mathbb{P}^*\big(\omega(Z_n,k^{-1}) \ge \epsilon\big) + \mathbb{P}\big(\omega(Z,k^{-1}) \ge \epsilon\big) + o(1),
\end{aligned}$$

und der limes superior der rechten Seite wird beliebig klein, für geeignete $k \in \mathbb{N}$ und $\epsilon > 0$. $\square$

7.8 Exkurs: Der Satz von Stone–Weierstrass

In diesem Abschnitt behandeln wir ein klassisches Resultat über die Approximation beliebiger durch spezielle stetige Funktionen. Zunächst erinnern wir an ein einfaches Resultat über die Approximation stetiger Funktionen durch *Bernstein-Polynome*:

Lemma 7.33. *Sei $f : [a,b] \to \mathbb{R}$ lipschitzstetig mit Konstante L. Für $n \in \mathbb{N}$ sei*

$$B_n f(x) := \sum_{j=0}^n \binom{n}{j} \Big(\frac{x-a}{b-a}\Big)^j \Big(\frac{b-x}{b-a}\Big)^{n-j} f\Big(a + (b-a)\frac{j}{n}\Big).$$

Dann ist

$$\big\|B_n f - f\big\|_{[a,b]} \le \frac{L(b-a)}{2\sqrt{n}}.$$

Beweis von Lemma 7.33. Für $p \in [0,1]$ sei $Y_p \sim \mathrm{Bin}(n,p)$. Dann ist

$$B_n f\big(a + (b-a)p\big) = \mathbb{E} f\big(a + (b-a)Y_p/n\big).$$

Folglich ist

$$\begin{aligned}
\Big|B_n f\big(a+(b-a)p\big) - f\big(a+(b-a)p\big)\Big| &\le \mathbb{E}\Big|f\big(a+(b-a)Y_p/n\big) - f\big(a+(b-a)p\big)\Big| \\
&\le \mathbb{E}\big(L(b-a)|Y_p/n - p|\big) \\
&\le L(b-a)\sqrt{\mathrm{Var}(Y_p/n)} \\
&\le L(b-a)/\sqrt{4n}.
\end{aligned}$$

$\square$

Nun formulieren wir das allgemeine Resultat für stetige Funktionen auf einem metrischen Raum $(\mathbb{M}, d)$. (Resultat und Beweis sind sogar für beliebige topologische Räume gültig.)

Satz 7.34 (Stone–Weierstrass). *Sei K eine kompakte Teilmenge von $\mathbb{M}$. Ferner sei $\mathcal{A}$ eine Teilmenge von $\mathcal{C}_{\mathrm{b}}(\mathbb{M})$ mit folgenden Eigenschaften:*

(i) *Für beliebige $\lambda \in \mathbb{R}$ und $g, h \in \mathcal{A}$ gehören auch λg und $g + h$ zu $\mathcal{A}$.*

(ii) *Für beliebige $g, h \in \mathcal{A}$ gehört auch gh zu $\mathcal{A}$.*

(iii) *Die konstante Funktion 1 gehört zu $\mathcal{A}$.*

(iv) *Für beliebige $x, y \in K$ mit $x \neq y$ existiert ein $g \in \mathcal{A}$ derart, dass $g(x) \neq g(y)$.*

Bezeichnen wir mit $\overline{\mathcal{A}}$ den Abschluss von $\mathcal{A}$ bezüglich $\|\cdot\|_{\mathbb{M}}$, dann existiert zu jedem $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{M})$ ein $g \in \overline{\mathcal{A}}$ derart, dass

$$g \;\equiv\; f \;\text{ auf } K \quad \text{ und } \quad \|g\|_{\mathbb{M}} \;\le\; \|f\|_K .$$

Beweis von Satz 7.34. Man kann sich leicht davon überzeugen, dass die Eigenschaften (i - iv) erhalten bleiben, wenn man $\mathcal{A}$ durch $\overline{\mathcal{A}}$ ersetzt. Wir können also ohne Einschränkung annehmen, dass $\mathcal{A} = \overline{\mathcal{A}}$.

Schritt 1: Aus den Bedingungen (i - iv) ergibt sich eine zusätzliche Eigenschaft von $\mathcal{A}$: Für beliebige $g, h \in \mathcal{A}$ gehören auch $\min(g, h)$ und $\max(g, h)$ zu $\mathcal{A}$.

Denn $\min(g, h)$ und $\max(g, h)$ lassen sich schreiben als $(g + h + |g - h|)/2$ bzw. $(g + h - |g - h|)/2$. Folglich genügt es zu zeigen, dass $|g| \in \mathcal{A}$ für beliebige $g \in \mathcal{A}$. Doch mit $[a, b] := \big[-\|g\|_{\mathbb{M}}, \|g\|_{\mathbb{M}}\big]$ und $f(x) := |x|$ in Lemma 7.33 folgt, dass $|g|$ der uniforme Grenzwert der Funktionen $B_n f \circ g$ für $n \to \infty$ ist. Wegen der Eigenschaften (i - iii) gehört mit g auch $B_n f \circ g$ zu $\mathcal{A}$.

Schritt 2: Für beliebige kompakte Mengen $A, B \subset K$ mit $A \cap B = \emptyset$ existiert eine Funktion $h \in \mathcal{A}$ mit Werten in $[0, 1]$, so dass $h \equiv 0$ auf A und $h \equiv 1$ auf B.

Um dies zu zeigen, wählen wir für beliebige $x, y \in K$ mit $x \neq y$ eine Funktion $h^o_{x,y} \in \mathcal{A}$ derart, dass $h^o_{x,y} < 0$ und $h^o_{x,y} > 1$. Dies ist möglich wegen (i, iii, iv). Dann ist aber

$$h_{x,y} \;:=\; \max\bigl(\min(h^o_{x,y}, 1), 0\bigr)$$

eine Funktion in $\mathcal{A}$ mit Werten in $[0, 1]$, und es existieren offene Umgebungen $U_{x,y}$ von x sowie $V_{x,y}$ von y, so dass $h_{x,y} \equiv 0$ auf $U_{x,y}$ und $h_{x,y} \equiv 1$ auf $V_{x,y}$. Für festes $x \in A$ bilden die Mengen $V_{x,y}$, $y \in B$, eine offene Überdeckung von B. Folglich existiert eine endliche Teilmenge $B_o(x)$ von B derart, dass $B \subset \bigcup_{y \in B_o(x)} V_{x,y}$. Insbesondere ist

$$h_x \;:=\; \max_{y \in B_o(x)} h_{x,y}$$

eine Funktion in $\mathcal{A}$ mit Werten in $[0, 1]$, so dass $h_x \equiv 0$ auf der offenen Umgebung $U_x := \bigcap_{y \in B_o(x)} U_{x,y}$ von x, und $h_x \equiv 1$ auf B. Nun bilden die Mengen U_x, $x \in A$, eine offene Überdeckung von A. Es existiert also eine endliche Teilmenge A_o von A derart, dass $A \subset \bigcup_{x \in A_o} U_x$.

Doch dann ist

$$h \ := \ \min_{x \in A_o} h_x$$

eine Funktion mit den gewünschten Eigenschaften.

Schritt 3: Nun sei f eine beliebige Funktion in $\mathcal{C}_{\mathrm{b}}(\mathbb{M})$. Dann konstruieren wir eine Funktion $h_1 \in \mathcal{A}$ wie folgt: Wir wählen Funktionen $h_1^{(\pm)}$ mit Werten in $\big[0, 2^{-1}\|f\|_K\big]$ derart, dass $h_1^{(\pm)} \equiv 0$ auf $\{\pm f \le 0\} \cap K$ und $h_1^{(\pm)} \equiv 2^{-1}\|f\|_K$ auf $\big\{\pm f \ge 2^{-1}\|f\|_K\big\} \cap K$. Mit $h_1 := h_1^{(+)} - h_1^{(-)}$ kann man leicht zeigen, dass $h_1 \in \mathcal{A}$ und

$$\|h_1\|_{\mathbb{M}} \ \le \ 2^{-1}\|f\|_K, \quad \|h_1 - f\|_K \ \le \ 2^{-1}\|f\|_K.$$

Nun wiederholen wir diesen Prozess induktiv: Seien $h_1, \ldots, h_\ell$ bereits gewählt, so dass

$$\|h_j\|_{\mathbb{M}} \ \le \ 2^{-j}\|f\|_K \quad \text{für } j = 1, \ldots, \ell \quad \text{und} \quad \Big\|\sum_{j=1}^{\ell} h_j - f\Big\|_K \ \le \ 2^{-\ell}\|f\|_K.$$

Dann wiederhole die obige Konstruktion von h_1 mit $f - \sum_{j=1}^{\ell} h_j$ an Stelle von f um $h_{\ell+1}$ zu erhalten. Letztlich ist dann der Grenzwert $g := \sum_{j=1}^{\infty} h_j$ eine Funktion in $\mathcal{A}$ mit den gewünschten Eigenschaften. □

7.9 Aufgaben

Aufgabe 7.1. Seien $(\mathbb{M}, \|\cdot\|)$ und $(\tilde{\mathbb{M}}, \|\cdot\|)$ normierte Räume, $T : \mathbb{M} \to \tilde{\mathbb{M}}$ eine Abbildung und x_o ein fester Punkt in $\mathbb{M}$. Zeigen Sie, dass die folgenden Aussagen äquivalent sind:

(a) Es existiert eine stetige Abbildung $H : \mathbb{M} \to \tilde{\mathbb{M}}$ derart, dass

$$\lim_{t \downarrow 0} \sup_{v \in K} \big\|t^{-1}[T(x_o + tv) - T(x_o)] - H(v)\big\| \ = \ 0$$

für jede kompakte Teilmenge K von $\mathbb{M}$.

(b) Es existiert eine Abbildung $H : \mathbb{M} \to \tilde{\mathbb{M}}$ derart, dass

$$\lim_{\lambda \to \infty,\, w \to v} \lambda\big(T(x_o + \lambda^{-1}w) - T(x_o)\big) \ = \ H(v)$$

für alle $v \in \mathbb{M}$. (Insbesondere ist H dann stetig.)

Aufgabe 7.2. Sei $(\mathcal{T}, \rho)$ ein metrischer Raum und $\mathcal{S}$ eine beliebige Teilmenge von $\mathcal{T}$. Zeigen Sie, dass die Menge aller $x \in \ell_\infty(\mathcal{T})$, welche in jedem Punkt $s \in \mathcal{S}$ stetig sind, ein abgeschlossener Untervektorraum von $\ell_\infty(\mathcal{T})$ ist.

Aufgabe 7.3. Zeigen Sie, dass die Menge $\mathbb{M}$ aller $x \in \ell_\infty(\mathbb{R})$ mit $\lim_{|t| \to \infty} x(t) = 0$ ein abgeschlossener linearer Teilraum von $\ell_\infty(\mathbb{R})$ ist.

Aufgabe 7.4. Beweisen Sie die kompakte Differenzierbarkeit des Funktionals T in Beispiel 7.13, i.e. Aussage (7.8).

Aufgabe 7.5. Sei P ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^d$ mit der Eigenschaft, dass $P\big(\{x \in \mathbb{R}^d : u^\top x = r\}\big) = 0$ für beliebige $u \in \mathbb{R}^d$ und $r \in \mathbb{R}$. Zeigen Sie, dass

$$\lim_{\epsilon \downarrow 0} \sup_{u \in \mathbb{R}^d, r \in \mathbb{R}} P\big(\{x \in \mathbb{R}^d : r \le u^\top x \le r + \epsilon\}\big) \ = \ 0.$$

Aufgabe 7.6 (Momentenmethode). Sei $\mathbb{M} = [0,1]$ und $\mathcal{F} = \{f_1, f_2, f_3, \ldots\}$ mit

$$f_k(x) \ := \ x^k.$$

Formulieren und begründen Sie nun mit Hilfe von Satz 7.22 Aussagen über (a) die Charakterisierung von Verteilungen auf $[0,1]$ und (b) Konvergenz in Verteilung von $[0,1]$-wertigen Zufallselementen.

Aufgabe 7.7. Sei $\mathbb{M} = \mathbb{M}_1 \times \mathbb{M}_2$ mit metrischen Räumen $(\mathbb{M}_1, d_1)$ und $(\mathbb{M}_2, d_2)$, und sei

$$d((x_1, x_2), (y_1, y_2)) \ := \ \max\big(d_1(x_1, x_2), d_2(x_2, y_2)\big).$$

(a) Angenommen, Z ist straff. Zeigen Sie, dass $(Z_n)_n$ genau dann in Verteilung gegen Z konvergiert, wenn $\lim_{n\to\infty} \mathbb{E}^* f(Z_n) = \mathbb{E}\, f(Z)$ für alle stetigen Funktionen f der Form

$$f(x_1, x_2) \ = \ f_1(x_1) f_2(x_2)$$

mit $f_i \in \mathcal{C}(\mathbb{M}_i, [0,1])$.

(b) Zeigen Sie, dass $(Z_n)_n$ genau dann asymptotisch straff und asymptotisch messbar ist, wenn mit $Z_n = (Z_{n1}, Z_{n2})$ beide Folgen $(Z_{n1})_n$ und $(Z_{n2})_n$ asymptotisch straff und asymptotisch messbar sind.

Aufgabe 7.8. Zeigen Sie, dass $\mathcal{N}_d(\mu_n, \Sigma_n) \to_w \mathcal{N}_d(\mu, \Sigma)$ genau dann, wenn $\mu_n \to \mu$ und $\Sigma_n \to \Sigma$.

Aufgabe 7.9. Sei $\rho(\cdot,\cdot)$ eine (Pseudo-) Metrik auf einer Menge $\mathcal{T}$. Ferner sei $h : [0,\infty) \to [0,\infty)$ monoton wachsend und stetig mit $h(0) = 0$, Grenzwert $h(\infty) > 0$, und

$$h(r+s) \ \le \ h(r) + h(s) \quad \text{für alle } r, s \ge 0.$$

Zeigen Sie, dass $h \circ \rho$ ebenfalls eine (Pseudo-) Metrik auf $\mathcal{T}$ darstellt und die gleiche Topologie wie ρ definiert.

Aufgabe 7.10. Sei $(\mathcal{T}, \rho)$ ein pseudometrischer Raum, und für $n \in \mathbb{N}$ sei $x_n : \mathcal{T} \to \mathbb{R}$ eine bezüglich ρ gleichmäßig stetige Funktion. Angenommen, $(x_n)_n$ konvergiert gleichmäßig gegen eine Funktion x. Zeigen Sie, dass auch x gleichmäßig stetig bezüglich ρ ist.

Aufgabe 7.11. Sei $\mathcal{T}_o$ eine endliche Teilmenge eines pseudometrischen Raumes $(\mathcal{T}, \rho)$, und sei

$$\sup_{t \in \mathcal{T}} \rho(t, \mathcal{T}_o) \ < \ \delta$$

für eine Konstante $\delta > 0$. Für $t \in \mathcal{T}$ und $u \in \mathcal{T}_o$ sei

$$\lambda(t,u) \;:=\; \bigl(1-\delta^{-1}\rho(t,u)\bigr)^+ \Big/ \sum_{v \in \mathcal{T}_o} \bigl(1-\delta^{-1}\rho(t,v)\bigr)^+.$$

Zeigen Sie, dass die Funktionen $\lambda(\cdot,u)$, $u \in \mathcal{T}_o$, lipschitzstetig bezüglich ρ sind.

Aufgabe 7.12. Sei $\hat{S}_n$ der Partialsummenprozess aus Abschnitt 1.2, und $\sup_{f\in\mathcal{F}} |f(x_{ni})|$ sei für alle $i \le n$ endlich. Zeigen Sie, dass $\hat{S}_n$ eine $\ell_\infty(\mathcal{F})$-Zufallsvariable ist.

Aufgabe 7.13. Zeigen Sie, dass die Abbildung

$$\mathbb{R} \ni y \;\mapsto\; 1\{y \le \cdot\} \in \ell_\infty(\mathbb{R})$$

nicht Borel-messbar ist.

Hinweis: Zeigen Sie, dass für beliebige Mengen $B \subset \mathbb{R}$ die Menge der Indikatorfunktionen $1\{y \le \cdot\}$, $y \in B$, abgeschlossen ist in $\ell_\infty(\mathbb{R})$.

Verallgemeinerung: Für $n \in \mathbb{N}$ ist die Abbildung $\mathbb{R}^n \ni y \mapsto \sum_{i=1}^n 1\{y_i \le \cdot\}$ nicht Borel-messbar.

Kapitel 8

Brownsche Bewegung und Brücke

In diesem Kapitel betrachten wir die Funktionenräume $\ell_\infty([0,1])$ und $\ell_\infty(\mathbb{R})$, wobei $\mathbb{R}$ mit der üblichen Metrik $\rho(s,t) := |t-s|$ versehen wird.

8.1 Stochastische Gleichstetigkeit auf $[0,1]$

Lemma 8.1. *Sei Z ein $\ell_\infty([0,1])$-Zufallselement. Für beliebige $0 < \delta \le 1$ und $\eta > 0$ ist*

$$\mathbb{P}^*\big(\omega(Z,\delta/2) > 2\eta\big) \ \le \ 2\delta^{-1} \sup_{t\in[0,1-\delta]} \mathbb{P}^*\Big(\sup_{u\in[t,t+\delta]} |Z(u)-Z(t)| > \eta\Big).$$

Beweis. Für $0 \le j \le \lfloor 2/\delta \rfloor$ sei $I_j := [j\delta/2, (j+1)\delta/2] \cap [0,1]$. Zu $u, v \in [0,1]$ mit $|u-v| < \delta/2$ existiert stets ein $j < \lfloor 2/\delta \rfloor$ mit $u, v \in I_j \cup I_{j+1}$, und $I_j \cup I_{j+1} \subset [t_j, t_j + \delta]$ mit $t_j := \min(j\delta/2, 1-\delta)$. Folglich ist

$$\begin{aligned} \mathbb{P}^*\big(\omega(Z,\delta/2) > 2\eta\big) &\le \sum_{j=0}^{\lfloor 2/\delta \rfloor - 1} \mathbb{P}^*\Big(\sup_{u,v\in I_j\cup I_{j+1}} |Z(u)-Z(v)| > 2\eta\Big) \\ &\le \sum_{j=0}^{\lfloor 2/\delta \rfloor - 1} \mathbb{P}^*\Big(\sup_{u\in I_j\cup I_{j+1}} |Z(u)-Z(t_j)| > \eta\Big) \\ &\le 2\delta^{-1} \sup_{t\in[0,1-\delta]} \mathbb{P}^*\Big(\sup_{u\in[t,t+\delta]} |Z(u)-Z(t)| > \eta\Big). \end{aligned}$$ □

Die Frage ist nun, wie man die rechte Seite der Ungleichung in Lemma 8.1 abschätzen kann.

Lemma 8.2. (Lévy). *Sei Z ein stochastischer Prozess auf $[a,b]$ mit rechtsseitig stetigen Pfaden und unabhängigen Zuwächsen; das heißt, für $a \le t < u \le b$ sind $(Z(s))_{s\in[a,t]}$ und $Z(u)-Z(t)$ stochastisch unabhängig. Für beliebige $\eta > 0$ ist*

$$\mathbb{P}\big(\|Z\|_{[a,b]} > \eta\big) \ \le \ \mathbb{P}\big(|Z(b)| > \eta/2\big) + \sup_{t\in[a,b]} \mathbb{P}\big(|Z(b)-Z(t)| > \eta/2\big).$$

Falls zusätzlich $0 \in \mathrm{Median}(Z(b)-Z(t))$ für $a \le t < b$, ist sogar

$$\mathbb{P}\big(\|Z\|_{[a,b]} > \eta\big) \ \le \ 2\,\mathbb{P}\big(|Z(b)| > \eta\big).$$

Beweis. Wegen der rechtsseitigen Stetigkeit der Pfade von Z genügt es, die Behauptung mit $\|Z\|_T$ an Stelle von $\|Z\|_{[a,b]}$ zu beweisen, wobei $T \supset \{b\}$ eine beliebige endliche Teilmenge von $[a, b]$ ist. Sei

$$\tau \ := \ \min\Bigl(\bigl\{t \in T : |Z(t)| > \eta\bigr\} \cup \{\infty\}\Bigr).$$

Dann ist

$$\{\|Z\|_T > \eta\} \ = \ \{\tau \in T\} \ \subset \ \bigl\{\tau \in T, |Z(b)| \le \eta/2\bigr\} \cup \bigl\{|Z(b)| > \eta/2\bigr\},$$

und $\bigl\{\tau \in T, |Z(b)| \le \eta/2\bigr\}$ ist gleich

$$\bigcup_{t\in T} \bigl\{\tau = t, |Z(b)| \le \eta/2\bigr\} \ \subset \ \bigcup_{t\in T} \{\tau = t\} \cap \bigl\{|Z(b) - Z(t)| > \eta/2\bigr\},$$

denn $|Z(t)| > \eta$ auf $\{\tau = t\}$. Folglich ist

$$\begin{aligned} \mathbb{P}(\|Z\|_T > \eta) \ &\le \ \mathbb{P}\bigl(|Z(b)| > \eta/2\bigr) + \mathbb{P}\bigl(\tau \in T, |Z(b)| \le \eta/2\bigr) \\ &\le \ \mathbb{P}\bigl(|Z(b)| > \eta/2\bigr) + \sum_{t\in T} \mathbb{P}(\tau = t)\, \mathbb{P}\bigl(|Z(b) - Z(t)| > \eta/2\bigr) \\ &\le \ \mathbb{P}\bigl(|Z(b)| > \eta/2\bigr) + \sup_{t\in[a,b]} \mathbb{P}\bigl(|Z(b) - Z(t)| > \eta/2\bigr). \end{aligned}$$

Wenn zusätzlich $\mathbb{P}\bigl(Z(b) - Z(t) \ge 0\bigr) \ge 1/2$ und $\mathbb{P}\bigl(Z(b) - Z(t) \le 0\bigr) \ge 1/2$ für $a \le t \le b$, dann ist

$$\begin{aligned} &\mathbb{P}(\tau = t) \\ &= \ \mathbb{P}\bigl(\tau = t, Z(t) > \eta\bigr) + \mathbb{P}\bigl(\tau = t, Z(t) < -\eta\bigr) \\ &\le \ 2\,\mathbb{P}\bigl(\tau = t, Z(t) > \eta, Z(b) - Z(t) \ge 0\bigr) + 2\,\mathbb{P}\bigl(\tau = t, Z(t) < -\eta, Z(b) - Z(t) \le 0\bigr) \\ &\le \ 2\,\mathbb{P}\bigl(\tau = t, Z(b) > \eta\bigr) + 2\,\mathbb{P}\bigl(\tau = t, Z(b) < -\eta\bigr) \\ &= \ 2\,\mathbb{P}\bigl(\tau = t, |Z(b)| > \eta\bigr), \end{aligned}$$

und die gewünschte Ungleichung folgt durch Aufsummieren über alle $t \in T$. □

8.2 Donskers Invarianzprinzipien

Definition 8.3 (Gaußprozess). Ein Gaußprozess auf einer beliebigen Indexmenge $\mathcal{T}$ ist ein stochastischer Prozess Z auf $\mathcal{T}$, so dass für beliebige $m \in \mathbb{N}$ und $t_1, t_2, \ldots, t_m \in \mathcal{T}$ der Vektor $(Z(t_j))_{1\le j\le m} \in \mathbb{R}^m$ Gauß-verteilt ist. Ist $\mathbb{E}(Z) \equiv 0$, so nennt man den Prozess Z *zentriert*.

Definition 8.4 (Brownsche Bewegung). Eine Brownsche Bewegung auf $\mathcal{T} = [0, 1]$ oder $\mathcal{T} = [0, \infty)$ ist ein zentrierter Gaußprozess W auf $\mathcal{T}$ mit fast sicher stetigen Pfaden und Kovarianzen

$$\mathbb{E}(W(t)W(u)) \ = \ \min(t, u) \quad \text{für } t, u \in \mathcal{T}.$$

Eine äquivalente Definition: Eine Brownsche Bewegung auf $\mathcal{T}$ ist ein zentrierter Gaußprozess W auf $\mathcal{T}$ mit fast sicher stetigen Pfaden und unabhängigen Zuwächsen, wobei $W(0) = 0$ fast sicher, und $W(u) - W(t) \sim \mathcal{N}(0, u - t)$ für $0 \le t < u \in \mathcal{T}$.

Definition 8.5 (Brownsche Brücke). Eine Brownsche Brücke ist ein zentrierter Gaußprozess B auf $[0, 1]$ mit stetigen Pfaden und Kovarianzen

$$\mathbb{E}(B(t)B(u)) \;=\; \min(t, u) - tu.$$

Anmerkung 8.6 (Eigenschaften der Brownschen Bewegung und Brücke). Die Existenz einer Brownschen Bewegung bzw. Brücke ist nicht offensichtlich sondern folgt implizit aus den Sätzen 8.7 und 8.8. Wir weisen bereits auf folgende Eigenschaften und Zusammenhänge hin:

Sei B eine Brownsche Brücke, und sei Y standardnormalverteilt und unabhängig von B. Dann definiert

$$\begin{aligned} W(t) &:= B(t) + tY \quad \text{eine Brownsche Bewegung auf } [0,1], \\ W(t) &:= (1+t)B\Big(\frac{t}{1+t}\Big) \quad \text{eine Brownsche Bewegung auf } [0,\infty). \end{aligned}$$

Sei W eine Brownsche Bewegung auf $[0, 1]$. Dann definiert

$$B(t) \;:=\; W(t) - tW(1)$$

eine Brownsche Brücke, weshalb man die Brownsche Brücke auch als *tied-down Brownian motion* bezeichnet.

Sei W eine Brownsche Bewegung auf $[0, \infty)$, und seien $\sigma > 0$, $v > u \geq 0$. Dann definiert

$$\begin{aligned} \tilde{W}(t) &:= \sigma^{-1}W(\sigma^2 t), \\ \tilde{W}(t) &:= \begin{cases} tW(1/t) & \text{falls } t > 0, \\ 0 & \text{falls } t = 0, \end{cases} \\ \tilde{W}(t) &:= W(v+t) - W(v) \end{aligned}$$

jeweils eine Brownsche Bewegung auf $[0, \infty)$. Ferner definiert

$$B(t) \;:=\; \frac{W((1-t)u + tv) - (1-t)W(u) - tW(v)}{\sqrt{v-u}}$$

eine Brownsche Brücke.

Nun kommen wir zu den Sätzen von Donsker. Im ersten Fall betrachten wir Partialsummenprozesse

$$[0,1] \ni t \;\mapsto\; W_n(t) \;:=\; n^{-1/2}\sum_{i=1}^{n} 1\{i \leq nt\}Y_{ni}.$$

Satz 8.7 (Donsker). *Angenommen, für jedes n sind $Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ unabhängig und identisch verteilt mit*

$$\mathbb{E}(Y_{n1}) \;=\; 0 \quad \textit{und} \quad \mathrm{Var}(Y_{n1}) \;=\; 1.$$

Ferner sei

$$\lim_{n\to\infty} \mathbb{E}\big(1\{Y_{n1}^2 \geq \epsilon n\}Y_{n1}^2\big) \;=\; 0 \quad \textit{für beliebige } \epsilon > 0.$$

Dann konvergiert $(W_n)_n$ in Verteilung gegen eine Brownsche Bewegung auf $[0, 1]$.

Im zweiten Fall betrachten wir den uniformen empirischen Prozess $\hat{G}_n$ aus Abschnitt 3.1, genauer gesagt, die Standardisierung

$$[0,1] \ni t \;\mapsto\; B_n(t) \;:=\; n^{1/2}(\hat{G}_n(t) - t) \;=\; n^{-1/2} \sum_{i=1}^{n} \bigl(1\{U_i \le t\} - t\bigr).$$

Satz 8.8 (Donsker). *Die Folge $(B_n)_n$ konvergiert in Verteilung gegen eine Brownsche Brücke.*

Beweis von Satz 8.7. Aus dem Lindebergschen Zentralen Grenzwertsatz kann man folgern, dass für beliebige $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{R})$ gilt:

$$\lim_{n\to\infty} \sup_{0\le u<v\le 1} \Bigl| \mathbb{E}\, f\bigl(W_n(u) - W_n(t)\bigr) - \mathbb{E}\, f\bigl(\sqrt{u-t}\, Y\bigr) \Bigr| \;=\; 0, \tag{8.1}$$

wobei Y standardnormalverteilt ist. Ferner hat der Prozess W_n unabhängige Zuwächse, weshalb

$$\mathcal{L}\Bigl(\bigl(W_n(t_i) - W_n(t_{i-1})\bigr)_{1\le i\le m} \Bigr) \;\to_w\; \mathcal{N}_m\Bigl(0, \mathrm{diag}\bigl((t_i - t_{i-1})_{i=1}^m\bigr)\Bigr)$$

für beliebige $m \in \mathbb{N}$ und $0 = t_0 < t_1 < t_2 < \cdots < t_m \le 1$. Zusammen mit $W_n(0) \equiv 0$ folgt hieraus die schwache Konvergenz der endlichdimensionalen Randverteilungen von W_n gegen zentrierte Normalverteilungen mit der gewünschten Kovarianzstruktur.

Zu zeigen bleibt die stochastische Gleichstetigkeit von $(W_n)_n$. Nach Lemma 8.1 und 8.2 genügt hierzu der Nachweis, dass

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sup_{0<u-t\le\delta} \delta^{-1}\, \mathbb{P}\bigl(|W_n(u) - W_n(t)| \ge \epsilon\bigr) \;=\; 0 \quad \text{für alle } \epsilon > 0.$$

Doch aus (8.1) folgt, dass

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{0<u-t\le\delta} \delta^{-1}\, \mathbb{P}\bigl(|W_n(u) - W_n(t)| \ge \epsilon\bigr) \\
&\le\; \sup_{0<u-t\le\delta} \delta^{-1}\, \mathbb{P}\bigl(\sqrt{u-t}\,|Y| \ge \epsilon\bigr) \;=\; \delta^{-1}\, \mathbb{P}\bigl(|Y| \ge \delta^{-1/2}\epsilon\bigr) \;\le\; \delta^{-1} \exp\bigl(-\delta^{-1}\epsilon^2/2\bigr),
\end{aligned}$$

und für $\delta \downarrow 0$ konvergiert die rechte Seite gegen Null. $\square$

Beweis von Satz 8.8. Für beliebiges festes n ist der Erwartungswert von B_n gleich Null, und $\mathbb{E}\bigl(B_n(t)B_n(u)\bigr) = \min(t,u) - tu$. Aus dem multivariaten Zentralen Grenzwertsatz folgt, dass

$$\mathcal{L}\bigl((B_n(t_i))_{1\le i\le m}\bigr) \;\to_w\; \mathcal{N}_m\Bigl(0, \bigl(\min(t_i,t_j) - t_i t_j\bigr)_{i,j=1}^m\Bigr)$$

für beliebige $m \in \mathbb{N}$ und $t_1, t_2, \ldots, t_m \in [0,1]$. Zu zeigen bleibt also die stochastische Gleichstetigkeit von $(B_n)_n$. Diese folgt direkt aus Lemma 8.9. $\square$

Lemma 8.9. *Für $0 < \delta \le 1$ und $\eta \ge \sqrt{\delta}$ ist stets*

$$\mathbb{P}\bigl(\omega(B_n, \delta/2) > 5\eta\bigr) \;\le\; \frac{32}{\delta}\Bigl(\exp\Bigl(-\frac{\eta^2}{4\delta}\Bigr) + \exp\Bigl(-\frac{3}{8}n\delta\Bigr)\Bigr).$$

Wählt man in Lemma 8.9 $\eta = \sqrt{\kappa\delta\log(e/\delta)}$ mit $\kappa > 4$, dann konvergiert die obere Schranke gegen Null für $\delta \to 0$ und $n\delta/\log n \to \infty$. Bis auf die große Konstante κ beschreibt dies das exakte Verhalten von $\omega(B_n, \cdot)$, denn W. Stute (1982) zeigte, dass hier $\omega(B_n, \delta) = \sqrt{2\delta\log(1/\delta)}(1 + o_p(1))$.

Beweis von Lemma 8.9. Wir betrachten zunächst B_n als stochastischen Prozess auf (einer abzählbaren dichten Teilmenge) der Menge $\mathcal{T}$ aller Paare $(s,t) \in [0,1]^2$ mit $0 < t - s < \delta/2$ vermöge $B_n(s,t) := B_n(t) - B_n(s)$. Dann ist nämlich $\omega(B_n, \delta/2) = \|B_n\|_{\mathcal{T}}$, und für alle $(s,t) \in \mathcal{T}$ gilt: $\mathrm{Var}[B_n(s,t)] \le \delta/2$, also $\mathbb{P}\big(|B_n(s,t)| \le \eta\big) \ge 1/2$. Mit einer unabhängigen Kopie B_n' von B_n und dem symmetrisierten Prozess $B_n^o(t) := n^{-1/2}\sum_{i=1}^n \xi_i 1\{U_i \le t\}$ gilt also:

$$\mathbb{P}\big(\omega(B_n, \delta/2) > 5\eta\big) \ \le\ 2\,\mathbb{P}\big(\omega(B_n - B_n', \delta/2) > 4\eta\big) \ \le\ 4\,\mathbb{P}\big(\omega(B_n^o, \delta/2) > 2\eta\big)$$

nach Symmetrisierungslemma 3.1 (a) und Symmetrisierungslemma 3.3 (a). Diese obere Schranke ist nach Lemma 8.1 und Lemma 8.2 nicht größer als

$$\begin{aligned}
\frac{8}{\delta}\sup_{t\in[0,1-\delta]}\mathbb{P}\Big(\sup_{u\in[t,t+\delta]}|B_n^o(u) - B_n^o(t)| > \eta\Big) \quad &= \quad \frac{8}{\delta}\,\mathbb{P}\Big(\sup_{u\in[0,\delta]}|B_n^o(u)| > \eta\Big)\\
&= \quad \frac{8}{\delta}\,\mathbb{E}\,\mathbb{P}\Big(\sup_{u\in[0,\delta]}|B_n^o(u)| > \eta\,\Big|\,\boldsymbol{U}\Big)\\
&\le \quad \frac{16}{\delta}\,\mathbb{E}\,\mathbb{P}\big(|B_n^o(\delta)| > \eta\,\big|\,\boldsymbol{U}\big)
\end{aligned}$$

mit $\boldsymbol{U} := (U_i)_{i=1}^n$. Nun verwenden wir Hoeffdings und Bennetts Ungleichung und erhalten

$$\begin{aligned}
\mathbb{P}\big(\omega(B_n, \delta/2) > 5\eta\big) \quad &\le \quad \frac{32}{\delta}\,\mathbb{E}\exp\Big(-\frac{\eta^2}{2\hat{G}_n(\delta)}\Big)\\
&\le \quad \frac{32}{\delta}\Big(\exp\Big(-\frac{\eta^2}{4\delta}\Big) + \mathbb{P}\big(\hat{G}_n(\delta) > 2\delta\big)\Big)\\
&\le \quad \frac{32}{\delta}\Big(\exp\Big(-\frac{\eta^2}{4\delta}\Big) + \exp\Big(-\frac{B(1)n\delta}{2}\Big)\Big).
\end{aligned}$$

Die Behauptung folgt jetzt aus der Ungleichung $B(1) \ge 3/4$; siehe Anmerkung 4.6. $\square$

8.3 Anwendung auf getrimmte Mittelwerte

Seien P und P_n Wahrscheinlichkeitsverteilungen auf $\mathbb{R}$ mit Verteilungsfunktionen F bzw. F_n. Ferner sei $\hat{P}_n$ die empirische Verteilung von unabhängigen Zufallsvariablen $X_{n1}, X_{n2}, \ldots, X_{nn}$ mit Verteilung P_n. Die Verteilungsfunktion von $\hat{P}_n$ sei $\hat{F}_n$.

Eine Standardmethode, um den empirischen Mittelwert zu "robustifizieren", besteht darin, die Variablen X_{ni} der Größe nach zu ordnen und die $k(n)$ größten sowie die $k(n)$ kleinsten Zahlen aus der Stichprobe zu entfernen, wobei $k(n) := \lfloor n\alpha \rfloor$ mit $0 < \alpha < 1/2$. Der Mittelwert dieser reduzierten Stichprobe ist ein *getrimmter Mittelwert*. Im Falle von paarweise verschiedenen Beobachtungen X_{ni} ist dieser getrimmte Mittelwert bis auf einen Faktor $1 + O(1/n)$ gleich $T(\hat{P}_n)$, wobei allgemein

$$T(P) \ := \ (1-2\alpha)^{-1} \int 1\{\alpha < F(x) \leq \beta\}\, x\, P(dx)$$

mit $\beta := 1 - \alpha$. Es ist $T(P) = \mu$, falls P stetig und um μ symmetrisch ist.

Satz 8.10. *Angenommen $(P_n)_n$ konvergiert schwach gegen P, wobei F stetig und auf dem Intervall $\{0 < F < 1\}$ streng isoton ist. Ferner sei*

$$\omega(F_n, 0+) \Big(= \max_{x \in \mathbb{R}} P\{x\}\Big) \ = \ o(n^{-1/2}).$$

Dann ist

$$T(\hat{P}_n) - T(P_n) \ = \ (1-2\alpha)^{-1} \int H\, d(\hat{P}_n - P_n) + o_p(n^{-1/2}),$$

wobei

$$H(x) \ := \ 1\{F(x) \leq \alpha\} F^{-1}(\alpha) + 1\{\alpha < F(x) \leq \beta\}\, x + 1\{\beta < F(x)\} F^{-1}(\beta).$$

Die Funktion H ist stetig und beschränkt. Aus dem Lindebergschen Zentralen Grenzwertsatz folgt daher, dass

$$L_n(P_n) := \mathcal{L}\big(n^{1/2}[T(\hat{P}_n) - T(P_n)]\big) \ \to_w \ \mathcal{N}\big(0, (1-2\alpha)^{-2} \mathrm{Var}_{X \sim P}[H(X)]\big).$$

Andererseits folgt aus den Voraussetzungen des Satzes, dass auch

$$\|\hat{F}_n - F\|_{\mathbb{R}} \ = \ o_p(1) \quad \text{und} \quad \omega(\hat{F}_n, 0+) \ = \ o_p(n^{-1/2}) \quad \text{(Aufgabe 8.2)}.$$

Folglich konvergiert auch $L_n(\hat{P}_n)$ schwach gegen $\mathcal{N}\big(0, (1-2\alpha)^{-2} \mathrm{Var}_{X \sim P}[H(X)]\big)$ in Wahrscheinlichkeit, ist also ein konsistenter Schätzer für $L_n(P_n)$. Die Verteilung $L_n(\hat{P}_n)$ ist ein sogenannter *Bootstrap-Schätzer* für $L_n(P_n)$ und kann auch wie folgt beschrieben werden: Sei $\hat{P}_n^*$ die empirische Verteilung von Zufallsvariablen $X_{n1}^*, X_{n2}^*, \ldots, X_{nn}^*$ mit

$$\mathcal{L}\big(X_{n1}^*, X_{n2}^*, \ldots, X_{nn}^* \,\big|\, X_{n1}, X_{n2}, \ldots, X_{nn}\big) \ = \ \hat{P}_n \otimes \hat{P}_n \otimes \cdots \otimes \hat{P}_n.$$

Dann ist

$$L_n(\hat{P}_n) \ = \ \mathcal{L}\big(n^{1/2}[T(\hat{P}_n^*) - T(\hat{P}_n)] \,\big|\, X_{n1}, X_{n2}, \ldots, X_{nn}\big).$$

Der Beweis von Satz 8.10 beruht im wesentlichen auf folgender Ungleichung:

Lemma 8.11. *Sei Q ein weiteres Wahrscheinlichkeitsmaß auf $\mathbb{R}$ mit Verteilungsfunktion G. Sei*

$$t(P) \ := \ \int 1\{\alpha < F(x) \leq \beta\}\, x\, P(dx).$$

Falls $\|G - F\| < \alpha$, so ist

$$t(Q) - t(P) \ = \ \int h\, d(Q - P) + R,$$

wobei

$$h(x) \;:=\; 1\{F(x) \le \alpha\}F^{-1}(\alpha+) + 1\{\alpha < F(x) \le \beta\}\, x + 1\{\beta < F(x)\}F^{-1}(\beta+).$$

und

$$\begin{aligned}
|R| &\le 10(\Delta + \delta)\Big(\omega(G - F, \delta+) + \omega(F, 0+)\Big) + 4\delta\|G - F\|, \\
\Delta &:= \max_{u \in \{\alpha,\beta\}} \Big|F^{-1}(u+)\Big|, \\
\delta &:= \max_{u \in \{\alpha,\beta\}} \Big(F^{-1}\Big((u + \|G - F\|)+\Big) - F^{-1}\Big((u - \|G - F\|)+\Big)\Big).
\end{aligned}$$

Beweis von Satz 8.10. Aus den Voraussetzungen an F und F_n folgt, dass $\|F_n - F\|_{\mathbb{R}} = o(1)$ und

$$\delta_n \;:=\; \sup_{u \in \{\alpha,\beta\}, |r| \le n^{1/3}} \Big|F_n^{-1}((u + r)+) - F^{-1}(u)\Big| \;=\; o(1).$$

Mit dem uniformen empirischen Prozess $\hat{G}_n$ und $B_n := n^{1/2}(\hat{G}_n - \mathrm{id})$ ist ferner $\hat{F}_n - F_n$ verteilt wie $n^{-1/2}B_n \circ F_n$, also

$$\begin{aligned}
\|\hat{F}_n - F_n\|_{\mathbb{R}} &= O_p(n^{-1/2}), \\
\omega(\hat{F}_n - F_n, \delta_n) &=_{\mathcal{L}} n^{-1/2}\omega(B_n \circ F_n, \delta_n) \\
&\le n^{-1/2}\omega\Big(B_n, \omega(F_n, \delta_n)+\Big) \\
&= o_p(n^{-1/2}).
\end{aligned}$$

Aus Lemma 8.11 folgt dann, dass

$$T(\hat{P}_n) - T(P_n) \;=\; (1 - 2\alpha)^{-1} \int H_n \, d(\hat{P}_n - P_n) + o_p(n^{-1/2}),$$

wobei

$$H_n(x) \;:=\; 1\{F_n(x) \le \alpha\}F_n^{-1}(\alpha+) + 1\{\alpha < F_n(x) \le \beta\}x + 1\{\beta < F_n(x)\}F_n^{-1}(\beta+).$$

Doch aus $\|F_n - F\|_{\mathbb{R}} = o(1)$ kann man leicht ableiten, dass $\|H_n - H\|_{\mathbb{R}} = o(1)$, weshalb $\int H_n \, d(\hat{P}_n - P_n) = \int H \, d(\hat{P}_n - P_n) + o_p(n^{-1/2})$. □

Beweis von Lemma 8.11. Wir teleskopieren

$$\begin{aligned}
t(Q) - t(P) &= \int \Big(1\{\alpha < G(x) \le \beta\} - 1\{\alpha < F(x) \le \beta\}\Big)\, x \, Q(dx) \\
&\quad + \int 1\{\alpha < F(x) \le \beta\}\, x \,(Q - P)(dx)
\end{aligned}$$

und analysieren nun das erste Integral auf der rechten Seite. Es ist

$$\begin{aligned}
1\{\alpha < G \le \beta\} - 1\{\alpha < F \le \beta\} &= 1\{\alpha < G\} - 1\{\beta < G\} - 1\{\alpha < F\} + 1\{\beta < F\} \\
&= 1_{J_1(\alpha)} - 1_{J_2(\alpha)} - 1_{J_1(\beta)} + 1_{J_2(\beta)},
\end{aligned}$$

wobei

$$J_1(u) \ := \ \{F \le u < G\} \quad \text{und} \quad J_2(u) \ := \ \{G \le u < F\}$$

für $u \in \{\alpha, \beta\}$.

Zum einen ist

$$\left.\begin{aligned} J_1(u) \ &\subset \ \{F \le u < F + \|G-F\|\} \ = \ \big\{F \in \big(u - \|G-F\|, u\big]\big\} \\ J_2(u) \ &\subset \ \{F - \|G-F\| \le u < F\} \ = \ \big\{F \in \big(u, u + \|G-F\|\big]\big\} \end{aligned}\right\}$$
$$\begin{aligned} &\subset \ \big\{F \in \big(u - \|G-F\|, u + \|G-F\|\big]\big\} \\ &\subset \ \Big[F^{-1}\big((u - \|G-F\|)+\big), F^{-1}\big((u + \|G-F\|)+\big)\Big]. \end{aligned}$$

Mit $x_u := F^{-1}(u+)$ gilt demnach für beliebige $\delta_o > \delta$:

$$\mathrm{Leb}(J_k(u)) \ < \ \delta_o \quad \text{und} \quad \sup_{x \in J_k(u)} |x - x_u| \ < \ \delta_o$$

für $k = 1, 2$. Insbesondere ist

$$\big|Q(J_k(u)) - P(J_k(u))\big| \ \le \ \omega_o := \omega(G - F, \delta_o).$$

Ferner ist

$$P(J_k(u)) \ \le \ \|G-F\| + 2\omega(F, 0+)$$

aufgrund der Tatsache, dass

(8.2) $$\big|P(F \in K) - \mathrm{Leb}(K)\big| \ \le \ 2\omega(F, 0+) \quad \text{für beliebige Intervalle } K \subset [0,1].$$

Alles in allem ist

$$\begin{aligned} &\Big|\int_{J_k(u)} x\, Q(dx) - x_u P(J_k(u))\Big| \\ &\le \ \int_{J_k(u)} |x - x_u|\, Q(dx) + \Delta\big|Q(J_k(u)) - P(J_k(u))\big| \\ &\le \ \delta_o Q(J_k(u)) + \Delta\big|Q(J_k(u)) - P(J_k(u))\big| \\ &\le \ (\Delta + \delta_o)\omega_o + \delta_o\|G-F\| + 2\delta_o\omega(F, 0+), \end{aligned}$$

und somit

$$\begin{aligned} &\int \big(1\{\alpha < G(x) \le \beta\} - 1\{\alpha < F(x) \le \beta\}\big)\, x\, Q(dx) \\ &= \ x_\alpha\big(P(J_1(\alpha)) - P(J_2(\alpha))\big) - x_\beta\big(P(J_1(\beta)) - P(J_2(\beta))\big) + R', \end{aligned}$$

wobei

$$|R'| \ \le \ 4(\Delta + \delta_o)\omega_o + 4\delta_o\|G-F\| + 8\delta_o\omega(F, 0+).$$

Nun betrachten wir die Differenz $P(J_1(u)) - P(J_2(u))$. Es ist

$$\begin{aligned}
J_1(u) &= \big\{x \in J_1(u) : F(x) \le u < F(x) + (G-F)(x)\big\} \\
&\subset \big\{F \le u < F + (G-F)(x_u) + \omega_o\big\} \\
&= \big\{F \in \big(u - (G-F)(x_u) - \omega_o, u\big]\big\}, \\
J_1(u) &= \big\{x : u - \|G-F\| < F(x) \le u < F(x) + (G-F)(x)\big\} \\
&\supset \big\{x : u - \|G-F\| \le F(x) \le u < F(x) + (G-F)(x_u) - \omega_o\big\} \\
&= \big\{x : F(x) \le u < F(x) + (G-F)(x_u) - \omega_o\big\} \\
&= \big\{F \in \big(u - (G-F)(x_u) + \omega_o, u\big]\big\}.
\end{aligned}$$

Analog zeigt man, dass

$$\big\{F \in \big(u, u - (G-F)(x_u) - \omega_o\big]\big\} \ \subset \ J_2(u) \ \subset \ \big\{F \in \big(u, u - (G-F)(x_u) + \omega_o\big]\big\}.$$

Zusammen mit (8.2) folgt also, dass

$$\big|P(J_1(u)) - P(J_2(u)) - (G-F)(x_u)\big| \ \le \ 2\omega_o + 4\omega(F, 0+).$$

Ferner ist

$$\Big|(G-F)(x_u) - \int 1\{F \le u\}\, d(Q-P)\Big| \ \le \ \max_{x \in \mathbb{R}} |Q\{x\} - P\{x\}| \ \le \ \omega_o,$$

und $\int 1\{F \le u\}\, d(Q-P) = -\int 1\{u < F\}\, d(Q-P)$. Folglich ist

$$\begin{aligned}
& x_\alpha\big(P(J_1(\alpha)) - P(J_2(\alpha))\big) - x_\beta\big(P(J_1(\beta)) - P(J_2(\beta))\big) \\
&= \int \big(1\{F < \alpha\} x_\alpha + 1\{\beta < F\} x_\beta\big)\, d(Q-P) + R'',
\end{aligned}$$

wobei

$$|R''| \ \le \ 6\Delta\omega_o + 8\Delta\omega(F, 0+).$$

Insgesamt erhalten wir die Entwicklung $t(Q) - t(P) = \int h\, d(Q-P) + R$ mit $|R| \le 10(\Delta + \delta_o)(\omega_o + \omega(F, 0+)) + 4\delta_o\|G-F\|$. Für $\delta_o \downarrow \delta$ ergibt sich die Behauptung. □

8.4 Verteilungen von Funktionalen von Gaussprozessen

8.4.1 Maximum und Supremumsnorm der Brownschen Brücke

Nachfolgend ermitteln wir die Verteilungen von $\max_{t \in [0,1]} B(t)$ und $\|B\|_{[0,1]}$ für eine Brownsche Brücke B. Als Hilfsmittel benötigen wir zwei nützliche Eigenschaften der Brownschen Bewegung.

Satz 8.12 (Starke Markov-Eigenschaft). *Sei M ein stochastischer Prozess auf $[0,\infty)$ mit folgenden Eigenschaften:*

$M(0) \equiv 0$, *und M hat rechtsseitig stetige Pfade;*

M hat unabhängige Zuwächse, wobei $\mathcal{L}(M(t+\delta) - M(t)) = \mathcal{L}(M(\delta))$ für $t, \delta \geq 0$.

Sei $\tau = \tau(M)$ eine Markovzeit. Das heißt, τ ist eine Zufallsvariable mit Werten in $[0,\infty]$, so dass

$$\{\tau \leq t\} \ \in \ \sigma(M(s) : s \leq t) \quad \text{für } 0 \leq t < \infty.$$

Ferner sei M' eine unabhängige Kopie von M. Dann definiert

$$\tilde{M}(t) \ := \ M(\min(t,\tau)) + M'((t-\tau)^+) \ = \ \begin{cases} M(t) & \text{für } t \leq \tau, \\ M(\tau) + M'(t-\tau) & \text{für } t \geq \tau, \end{cases}$$

einen stochastischen Prozess $\tilde{M}$ mit der gleichen Verteilung und den gleichen Pfadeigenschaften wie M.

Beispiele für solche Prozesse M sind der Standard-Poissonprozess und die Brownsche Bewegung auf $[0,\infty)$. Im letzteren Falle ist M symmetrisch, und man kann folgendes Resultat anwenden.

Korollar 8.13 (Spiegelungsprinzip). *Sei M ein stochastischer Prozess wie in Satz 8.12, so dass $\mathcal{L}(-M) = \mathcal{L}(M)$. Sei $\tau = \tau(M)$ eine Markovzeit. Dann definiert*

$$M_o(t) \ := \ M(\min(t,\tau)) - \big(M(t) - M(\min(t,\tau))\big) \ = \ \begin{cases} M(t) & \text{für } t \leq \tau, \\ 2M(\tau) - M(t) & \text{für } t \geq \tau, \end{cases}$$

einen stochastischen Prozess M_o mit der gleichen Verteilung und den gleichen Pfadeigenschaften wie M.

Man sagt, der Prozess M_o geht aus M durch "Spiegelung zur Markovzeit τ" hervor. Korollar 8.13 folgt aus Satz 8.12, indem man dort $\tilde{M}_o$ wie $\tilde{M}$ mit $-M'$ anstelle von M' definiert. Denn man kann leicht nachweisen, dass $\tau(M) = \tau(\tilde{M})$, und es ist

$$\tilde{M}_o(t) \ = \ \tilde{M}(\min(t,\tau)) - \big(\tilde{M}(t) - \tilde{M}(\min(t,\tau))\big).$$

Beweis von Satz 8.12. Man sieht leicht, dass sich die Pfadeigenschaften von M und M' auf $\tilde{M}$ übertragen. Zu zeigen bleibt, dass $\tilde{M}$ ein stochastischer Prozess mit der gleichen Verteilung wie M ist. Zu diesem Zweck betrachten wir für $n \in \mathbb{N}$ die Variablen $\tau_n := 2^{-n}\lceil 2^n\tau\rceil$ und $\tilde{\tau}_n := 2^{-n}\lfloor 2^n\tau\rfloor$. Auch τ_n ist eine Markovzeit, denn $\{\tau_n \leq t\} = \big\{\tau \leq 2^{-n}\lfloor 2^n t\rfloor\big\}$. Ferner konvergiert $(\tau_n)_n$ antiton und $(\tilde{\tau}_n)_n$ isoton gegen τ mit $\tau_n - \tilde{\tau}_n \leq 2^{-n}$, falls $\tau < \infty$. Wegen der rechtsseitigen Stetigkeit von M und M' ist

$$M(\min(t,\tau)) \ = \ \lim_{n\to\infty} M(\min(t,\tau_n)) \quad \text{und} \quad M'((t-\tau)^+) \ = \ \lim_{n\to\infty} M((t-\tilde{\tau}_n)^+),$$

also ist auch $\tilde{M}(t)$ messbar. Ferner ist

$$\begin{aligned}
&\mathbb{P}\Big(\big|M'((t-\tau_n)^+) - M'((t-\tilde{\tau}_n)^+)\big| > \epsilon\Big)\\
\leq\ & \sum_{k=1}^{\infty} \mathbb{P}\Big(\tau_n = 2^{-n}k, \sup_{t\in[2^{-n}(k-1),2^{-n}k]} |M'(t) - M'(2^{-n}k)| > \epsilon\Big)\\
=\ & \sum_{k=1}^{\infty} \mathbb{P}(\tau_n = 2^{-n}k)\,\mathbb{P}\Big(\sup_{t\in[2^{-n}(k-1),2^{-n}k]} |M'(t) - M'(2^{-n}k)| > \epsilon\Big)\\
\leq\ & \mathbb{P}\Big(\sup_{t\in[0,2^{-n}]} |M'(t) - M'(2^{-n})| > \epsilon\Big)\\
\to\ & 0 \quad (n\to\infty)
\end{aligned}$$

für beliebige $\epsilon > 0$. Definiert man also $\tilde{M}_n$ wie $\tilde{M}$ mit τ_n anstelle von τ, dann konvergiert $(\tilde{M}_n(t))_n$ stochastisch gegen $\tilde{M}(t)$, und es genügt zu zeigen, dass $\tilde{M}_n$ genauso verteilt ist wie M. Zu diesem Zwecke benutzen wir die Tatsache, dass für beliebige $k < \infty$ gilt:

$$\begin{aligned}
&\mathcal{L}\big((M(2^{-n}k+t))_{t\geq 0}\,\big|\,M(s) : s \leq 2^{-n}k\big)\\
=\ & \mathcal{L}\big((M(2^{-n}k) + M'(t))_{t\geq 0}\,\big|\,M(s) : s\leq 2^{-n}k\big).
\end{aligned} \tag{8.3}$$

Für $m \in \mathbb{N}$ und $0 \leq t_1 < t_2 < \cdots < t_m$ sowie Borelmengen $B_1, B_2, \ldots, B_m \subset \mathbb{R}$ ist folglich

$$\begin{aligned}
&\mathbb{P}\big(\tilde{M}_n(t_i) \in B_i \text{ für } 1 \leq i \leq m\big)\\
=\ & \sum_{0\leq k\leq\infty} \mathbb{P}\big(\tau_n = 2^{-n}k, \tilde{M}_n(t_i) \in B_i \text{ für } 1\leq i\leq m\big)\\
=\ & \sum_{0\leq k\leq\infty} \mathbb{P}\Big(\tau_n = 2^{-n}k, M(t_i)\in B_i \text{ falls } t_i \leq 2^{-n}k,\\
&\qquad M(2^{-n}k) + M'(t_i - 2^{-n}k) \in B_i \text{ falls } t_i > 2^{-n}k\Big)\\
=\ & \sum_{0\leq k\leq\infty} \mathbb{E}\,\mathbb{P}\Big(\tau_n = 2^{-n}k, M(t_i)\in B_i \text{ falls } t_i \leq 2^{-n}k,\\
&\qquad M(2^{-n}k) + M'(t_i - 2^{-n}k) \in B_i \text{ falls } t_i > 2^{-n}k\,\Big|\,M(s) : s\leq 2^{-n}k\Big)\\
=\ & \sum_{0\leq k\leq\infty} \mathbb{E}\,\mathbb{P}\Big(\tau_n = 2^{-n}k, M(t_i)\in B_i \text{ für } 1\leq i\leq m\,\Big|\,M(s) : s\leq 2^{-n}k\Big)\\
=\ & \sum_{0\leq k\leq\infty} \mathbb{P}\big(\tau_n = 2^{-n}k, M(t_i)\in B_i \text{ für } 1\leq i\leq m\big)\\
=\ & \mathbb{P}\big(M(t_i)\in B_i \text{ für } 1\leq i\leq m\big),
\end{aligned}$$

wobei die vierte (drittletzte) Gleichung eine Folgerung aus (8.3) ist. □

Satz 8.14. *Sei B eine Brownsche Brücke. Für beliebige $\eta > 0$ ist*

$$\mathbb{P}\Big(\max_{t\in[0,1]} B(t) \geq \eta\Big) = \exp(-2\eta^2) \qquad \textit{und} \tag{8.4}$$

$$\mathbb{P}\big(\|B\|_{[0,1]} \geq \eta\big) = 2\sum_{k=1}^{\infty} (-1)^{k-1}\exp(-2k^2\eta^2). \tag{8.5}$$

Anmerkung 8.15. Man kann zeigen, dass eine Brownsche Brücke B fast sicher eine eindeutige Maximalstelle $U = \arg\max_{t\in[0,1]} B(t)$ besitzt, wobei $U \sim \mathcal{U}[0,1]$; siehe Aufgabe 8.5.

Beweis von Satz 8.14. Sei W eine Brownsche Bewegung auf $[0,\infty)$ und $B(t) = W(t) - tW(1)$ für $t \in [0,1]$. Dann sind B und $W(1)$ stochastisch unabhängig, weshalb

$$\mathcal{L}\big((W(t))_{t\in[0,1]} \,\big|\, |W(1)| \le \epsilon\big) \;\to_w\; \mathcal{L}(B) \quad \text{für } \epsilon \downarrow 0.$$

Nach dem Portmanteau-Theorem ist also

$$\lim_{\epsilon\downarrow 0} \mathbb{P}\big(\phi(W) \ge \eta \,\big|\, |W(1)| \le \epsilon\big) \;=\; \mathbb{P}(\phi(B) \ge \eta), \tag{8.6}$$

falls $\phi(B) \ne \eta$ fast sicher. Hierbei ist $\phi(x) = \max_{[0,1]} x$ oder $\phi(x) = \|x\|_{[0,1]}$. Angenommen, wir weisen nach, dass der Grenzwert in (8.6) für alle $\eta > 0$ existiert und in η stetig ist. Aus der Monotonie von $\mathbb{P}\big(\phi(B) \ge \eta\big)$ in η folgt dann die Gültigkeit von (8.6) für alle $\eta > 0$.

Im Falle von $\phi(x) = \max_{[0,1]} x$ sei

$$\tau = \tau(W) \;:=\; \inf\big\{t \ge 0 : W(t) \ge \eta\big\}.$$

Spiegeln von W zur Stoppzeit τ liefert die Brownsche Bewegung $W_o(t) := W(\min(t,\tau)) - \big(W(t) - W(\min(t,\tau))\big)$, und für $0 < \epsilon \le \eta$ gilt:

$$\begin{aligned}
\mathbb{P}\Big(\max_{[0,1]} W \ge \eta, |W(1)| \le \epsilon\Big) &= \mathbb{P}\big(\tau \le 1, |W(1)| \le \epsilon\big) \\
&= \mathbb{P}\big(\tau \le 1, |W_o(1) - 2\eta| \le \epsilon\big) \\
&= \mathbb{P}\big(|W_o(1) - 2\eta| \le \epsilon\big),
\end{aligned}$$

denn aus $|W_o(1) - 2\eta| \le \epsilon \le \eta$ folgt bereits, dass $\tau(W) = \tau(W_o) \le 1$. Daher ist

$$\begin{aligned}
\mathbb{P}\Big(\max_{[0,1]} W \ge \eta \,\Big|\, |W(1)| \le \epsilon\Big) &= \frac{\mathbb{P}\big(|W_o(1) - 2\eta| \le \epsilon\big)}{\mathbb{P}\big(|W_o(1)| \le \epsilon\big)} \\
&= \frac{\mathcal{N}(0,1)\big([2\eta - \epsilon, 2\eta + \epsilon]\big)}{\mathcal{N}(0,1)([-\epsilon,\epsilon])} \\
&\to \exp(-2\eta^2).
\end{aligned}$$

Dies beweist Behauptung (8.4).

Im Falle von $\phi(x) = \|x\|_{[0,1]}$ ist $\phi(W) \ge \eta$ genau dann, wenn $W \in A_1 \cup B_1$ mit den Mengen

$$\begin{aligned}
A_1 &:= \big\{x \in \mathcal{C}([0,\infty)) : x(t) = \eta \text{ für ein } t \in [0,1]\big\}, \\
B_1 &:= \big\{x \in \mathcal{C}([0,\infty)) : x(t) = -\eta \text{ für ein } t \in [0,1]\big\}.
\end{aligned}$$

Allerdings sind diese Ereignisse nicht disjunkt. Vielmehr ist $A_1 \cap B_1 = A_2 \cup B_2$, wobei allgemein

$$\begin{aligned}
A_k &:= \big\{x \in \mathcal{C}([0,\infty)) : \text{ es existieren } 0 \le t_1 < t_2 < \cdots < t_k \le 1 \text{ derart, dass} \\
&\qquad\qquad x(t_i) = (-1)^{i-1}\eta \quad \text{für } 1 \le i \le k\big\}, \\
B_k &:= \big\{x \in \mathcal{C}([0,\infty)) : \text{ es existieren } 0 \le t_1 < t_2 < \cdots < t_k \le 1 \text{ derart, dass} \\
&\qquad\qquad x(t_i) = (-1)^{i}\eta \quad \text{für } 1 \le i \le k\big\}.
\end{aligned}$$

Im Falle von $\|x\|_{[0,1]} \geq \eta$ existiert ein $k(x) \in \mathbb{N}$ mit

$$x \begin{cases} \in \; A_k \cap B_k & \text{falls } 1 \leq k < k(x), \\ \in \; A_k \triangle B_k & \text{falls } k = k(x), \\ \notin \; A_k \cup B_k & \text{falls } k > k(x). \end{cases}$$

Insbesondere kann man nun zeigen, dass

$$1\big\{\|x\|_{[0,1]} \geq \eta\big\} \;=\; \sum_{k=1}^{\infty} (-1)^{k-1} \big(1\{x \in A_k\} + 1\{x \in B_k\}\big)$$

und

$$\sum_{k=1}^{N} (-1)^{k-1} \big(1\{x \in A_k\} + 1\{x \in B_k\}\big) \;\in\; \{0,1,2\} \quad \text{für alle } N \in \mathbb{N}.$$

Aus dem Satz von der majorisierten Konvergenz folgt also, dass

$$\begin{aligned} & \mathbb{P}\big(\|W\|_{[0,1]} \geq \eta, |W(1)| \leq \epsilon\big) \\ =\; & \sum_{k=1}^{\infty} (-1)^{k-1} \Big(\mathbb{P}\big(W \in A_k, |W(1)| \leq \epsilon\big) + \mathbb{P}\big(W \in B_k, |W(1)| \leq \epsilon\big)\Big) \\ =\; & 2 \sum_{k=1}^{\infty} (-1)^{k-1} \, \mathbb{P}\big(W \in A_k, |W(1)| \leq \epsilon\big). \end{aligned}$$

Um die Wahrscheinlichkeit von $\{W \in A_k, |W(1)| \leq \epsilon\}$ zu berechnen, betrachten wir für $a \geq 0$, $k \in \mathbb{N}$ und $0 < \epsilon \leq \eta$ folgende Menge:

$$C_{a,k} \;:=\; \big\{x \in \mathcal{C}([0,\infty)) : \begin{array}{l} \text{es existieren } 0 \leq t_1 < t_2 < \cdots < t_k \leq 1 \text{ derart, dass} \\ x(t_i) = a + (-1)^{i-1}\eta \;\text{ für } 1 \leq i \leq k \quad \text{und} \quad |x(1) - a| \leq \epsilon \big\}. \end{array}$$

Die Spiegelung von $x \in \mathcal{C}([0,\infty))$ zur Stoppzeit $\tau(x) := \inf\{t \geq 0 : x(t) \geq a + \eta\}$ ergibt eine bijektive Abbildung von $C_{a,k}$ nach $C_{a+2\eta,k-1}$, wobei $C_{b,0} := \{x : |x(1) - b| \leq \epsilon\}$. Aus dem Spiegelungsprinzip für W folgt also, dass $\mathbb{P}(W \in C_{a,k}) = \mathbb{P}(W \in C_{a+2\eta,k-1})$, und induktiv ergibt sich die Formel

$$\begin{aligned} \mathbb{P}(W \in A_k, |W(1)| \leq \epsilon) \;&=\; \mathbb{P}(W \in C_{0,k}) \;=\; \mathbb{P}(W \in C_{2\eta,k-1}) \;=\; \cdots \\ &=\; \mathbb{P}\big(|W(1) - 2k\eta| \leq \epsilon\big). \end{aligned}$$

Folglich ist

$$\begin{aligned} \mathbb{P}\big(\|W\|_{[0,1]} \geq \eta \,\big|\, |W(1)| \leq \epsilon\big) \;&=\; 2 \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\mathcal{N}(0,1)\big([2k\eta - \epsilon, 2k\eta + \epsilon]\big)}{\mathcal{N}(0,1)([-\epsilon, \epsilon])} \\ &\to\; 2 \sum_{k=1}^{\infty} (-1)^{k-1} \exp(-2k^2\eta^2) \quad \text{für } \epsilon \downarrow 0. \end{aligned}$$

Dabei ist zu beachten, dass stets

$$\frac{\mathcal{N}(0,1)\big([2k\eta - \epsilon, 2k\eta + \epsilon]\big)}{\mathcal{N}(0,1)([-\epsilon, \epsilon])} \;\leq\; \exp\big(\epsilon^2/2 - (2k\eta - \epsilon)^2/2\big) \;\leq\; \exp\big((1 - (2k\eta - 1)^2)/2\big).$$

Man kann also den Grenzübergang summandenweise durchführen. $\square$

8.4.2 Lineare und quadratische Funktionale von Gaußprozessen

Sei $(\mathcal{T}, \rho)$ ein präkompakter pseudometrischer Raum und X ein Gaußprozess auf $\mathcal{T}$ mit gleichmäßig stetigen Pfaden. Dies impliziert, dass auch $\mu := \mathbb{E}(X)$ und $K := \operatorname{Cov}(X)$ gleichmäßig stetige Funktionen auf $\mathcal{T}$ beziehungsweise $\mathcal{T} \times \mathcal{T}$ sind. (Dabei wird $\mathcal{T} \times \mathcal{T}$ vermöge $\tilde{\rho}((s,t),(s',t')) := \rho(s,s') + \rho(t,t')$ metrisiert.) Dies folgt im wesentlichen aus Aufgabe 7.8. Ferner sei Q ein Wahrscheinlichkeitsmaß auf den Borelmengen in $\mathcal{T}$.

Lemma 8.16. *Sei Y ein stochastischer Prozess auf $\mathcal{T}$ mit gleichmäßig stetigen Pfaden, definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann ist die Abbildung*

$$\mathcal{T} \times \Omega \ni (t,w) \ \mapsto \ Y(t)(w)$$

$\operatorname{Borel}(\mathcal{T}) \otimes \mathcal{A}$*-messbar. Insbesondere ist*

$$\mathbb{E}\, Q\big(\{t \in \mathcal{T} : Y(t) = f(t)\}\big) \ = \ \int \mathbb{P}(Y(t) = f(t))\, Q(dt)$$

für beliebige messbare Funktionen f auf $\mathcal{T}$, und

$$\mathbb{E} \int Y(t)\, Q(dt) \ = \ \int \mathbb{E}\, Y(t)\, Q(dt),$$

falls $Y \geq 0$ oder $\int \mathbb{E}\, |Y(t)|\, Q(dt) < \infty$.

Beweis. Für $n \in \mathbb{N}$ sei $(B_{n1}, B_{n2}, \ldots, B_{nk(n)})$ eine Partition von $\mathcal{T}$ in Borelmengen B_{ni} mit $\operatorname{diam}(B_{ni}) < n^{-1}$. Für feste Punkte $t_{ni} \in B_{ni}$ ist

$$Y_n(t) \ := \ \sum_{i=1}^{k(n)} 1\{t \in B_{ni}\} Y(t_{ni})$$

ein stochastischer Prozess auf $\mathcal{T}$ mit $|Y_n(t)(w) - Y(t)(w)| \leq \omega(Y, n^{-1}) \to 0$ für beliebige $(t,w) \in \mathcal{T} \times \Omega$. Außerdem ist Y_n messbar als Funktion auf $\mathcal{T} \times \Omega$, da

$$\Big\{(t,w) \in \mathcal{T} \times \Omega : Y_n(t)(w) \leq r\Big\} \ = \ \bigcup_{i=1}^{k(n)} B_{ni} \times \Big\{w \in \Omega : Y(t_{ni})(w) \leq r\Big\}.$$

Somit ist auch Y eine messbare Funktion auf $\mathcal{T} \times \Omega$. Die weiteren Aussagen folgen nun aus dem Satz von Fubini. □

Satz 8.17. *Für beliebige $g \in \mathcal{L}^1(Q)$ ist $\int X g\, dQ$ normalverteilt mit Mittelwert $\int \mu g\, dQ$ und Varianz $\int K(s,t) g(s) g(t)\, Q(ds) Q(dt)$.*

Korollar 8.18. *Für beliebige $d \in \mathbb{N}$ und $g_1, g_2, \ldots, g_d \in \mathcal{L}^1(Q)$ ist $\big(\int X g_i\, dQ\big)_{i=1}^d$ verteilt nach*

$$\mathcal{N}\bigg(\Big(\int \mu g_i\, dQ\Big)_{1 \leq i \leq d}, \Big(\int K(s,t) g_i(s) g_j(t)\, Q(ds) Q(dt)\Big)_{1 \leq i,j \leq d}\bigg).$$ □

Eine mögliche Charakterisierung von Normalverteilungen auf $\mathbb{R}^d$ lautet wie folgt: Ein Zufallsvektor $Y \in \mathbb{R}^d$ ist normalverteilt genau dann, wenn $\ell^\top Y \in \mathbb{R}$ normalverteilt ist für beliebige $\ell \in \mathbb{R}^d$. Die Verallgemeinerung auf Banachräume lautet dann: Sei Y eine Zufallsvariable mit Werten in einem Banachraum $(\mathbb{B}, \|\cdot\|)$. Man nennt Y normalverteilt, wenn LY normalverteilt ist für beliebige stetige Linearformen L auf $\mathbb{B}$. Speziell für $\big(\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho), \|\cdot\|_{\mathcal{T}}\big)$ besagt der Rieszsche Darstellungssatz, dass jede stetige Linearform L auf $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$ von der Form $Lx = \int x\, d\nu$ mit einem endlichen signierten Maß ν auf $\mathcal{T}$ ist. Aus Satz 8.17 und Lemma 7.30 folgt somit:

Korollar 8.19. *Die Menge der normalverteilten $\mathcal{C}_{\mathrm{u}}(\mathcal{T},\rho)$-wertigen Zufallsvariablen ist identisch mit der Menge aller Gaußprozesse mit gleichmäßig stetigen Pfaden.*

Satz 8.20. *Sei X zentriert. Es existiert ein vollständiges Orthonormalsystem $(\phi_k)_k$ von $L^2(Q)$ und eine monoton fallende Folge $(\lambda_k)_k$ in $[0,\infty)$ mit*

$$\int K(s,t)\phi_j(s)\phi_k(t)\, Q(ds)Q(dt) \;=\; 1\{j=k\}\lambda_k \quad \text{für alle } j,k.$$

Ferner ist $\sum_k \lambda_k = \int K(t,t)\, Q(dt)$ und

$$\int X^2\, dQ \;=_{\mathcal{L}}\; \sum_k \lambda_k Z_k^2$$

mit unabhängigen, standardnormalverteilten Zufallsvariablen $Z_1, Z_2, Z_3, \ldots$.

Im Falle einer Brownschen Brücke B und der Gleichverteilung auf $[0,1]$ definiert

$$\phi_k(t) \;:=\; \sqrt{2}\,\sin(k\pi t)$$

ein vollständiges Orthonormalsystem von $L^2[0,1]$ mit

$$\int_0^1\int_0^1 \phi_j(t)\phi_k(u)K(t,u)\,dt\,du \;=\; \frac{1\{j=k\}}{\pi^2k^2},$$

wobei hier $K(t,u) = \min(t,u) - tu$. Wie im Beweis von Satz 8.20 zeigt man, dass

$$\int_0^1 B(t)^2\, dt \;=_{\mathcal{L}}\; \sum_{k=1}^\infty \frac{Z_k^2}{\pi^2k^2}.$$

Beweis von Satz 8.17. Sei $X_n(t) := \sum_{i=1}^{k(n)} 1\{t \in B_{ni}\}X(t_{ni})$ mit B_{ni} und $t_{ni} \in B_{ni}$ wie im Beweis von Lemma 8.16. Dann konvergiert $\int X_n g\, dQ$ in Verteilung gegen $\int Xg\, dQ$. Doch

$$\begin{aligned}
\int X_n g\, dQ \;&=\; \sum_i X(t_{ni})\int_{B_{ni}} g\, dQ \\
&\sim\; \mathcal{N}\Big(\sum_i \mu(t_{ni})\int_{B_{ni}} g\, dQ, \sum_{i,j} K(t_{ni},t_{nj})\int_{B_{ni}\times B_{nj}} g(s)g(t)\, Q(ds)Q(dt)\Big),
\end{aligned}$$

und aus der gleichmäßigen Stetigkeit von μ und K folgt, dass Mittelwert und Varianz dieser Normalverteilung gegen $\int \mu g\, dQ$ beziehungsweise $\int K(s,t)g(s)g(t)\, Q(ds)Q(dt)$ konvergieren. $\square$

Beweis von Satz 8.20. Aus der Funktionalanalysis ist bekannt, dass

$$Sg(t) \ := \ \int K(t,u)g(u)\,Q(du) \quad (g \in L^2(Q))$$

einen kompakten linearen Operator von $L^2(Q)$ nach $L^2(Q)$ definiert. Da $K(s,t) = K(t,s)$ für alle $s,t \in \mathcal{T}$, ist dieser Operator S symmetrisch und positiv semidefinit, denn

$$\begin{aligned} \int (Sg)h\,dQ &= \int K(s,t)g(s)h(t)\,Q(ds)Q(dt) \\ &= \mathrm{Cov}\Big(\int Xg\,dQ, \int Xh\,dQ\Big) \begin{cases} = \int (Sh)g\,dQ, \\ \geq \ 0 & \text{falls } g = h. \end{cases} \end{aligned}$$

Aus dem Spektralsatz für kompakte, normale Operatoren auf Hilberträumen folgt die Existenz eines vollständigen Orthonormalsystems $(\phi_k)_k$ von $L^2(Q)$ mit $S\phi_k = \lambda_k\phi_k$, wobei $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq 0$. Nach Korollar 8.18 sind die Variablen $\int X\phi_k\,dQ$, $k \geq 0$, stochastisch unabhängig und normalverteilt mit Mittelwert Null und Varianz λ_k. Nach der Parsevalschen Gleichung ist somit

$$\int X^2\,dQ \ = \ \sum_k \Big(\int X\phi_k\,dQ\Big)^2 \ =_{\mathcal{L}} \ \sum_k \lambda_k Z_k^2.$$

Insbesondere ist

$$\sum_k \lambda_k \ = \ \mathbb{E}\int X^2\,dQ \ = \ \int K(t,t)\,Q(dt)$$

nach Lemma 8.16. $\square$

8.5 Der uniforme Quantilprozess

Mithilfe von Satz 8.8 kann man auch einen funktionalen Grenzwertsatz für den uniformen Quantilprozess $\hat{G}_n^{-1}$ beweisen. Für $u \in [0,1]$ sei $\mathrm{id}(u) := u$.

Satz 8.21 (Bahadur-Kiefer). *Die Folge $\big(n^{1/2}(\hat{G}_n^{-1} - \mathrm{id})\big)_n$ konvergiert im Raum $\ell_\infty([0,1])$ in Verteilung gegen eine Brownsche Brücke. Ferner ist*

$$\big\| n^{1/2}(\hat{G}_n^{-1} - \mathrm{id}) + n^{1/2}(\hat{G}_n - \mathrm{id})\big\| \ = \ o_p(1).$$

Für den Beweis dieses Satzes benötigen wir folgendes Resultat:

Lemma 8.22 (Vervaat). *Sei T eine Abbildung von $\ell_\infty([0,1])$ nach $\ell_\infty([0,1])$ mit $T_1 f \leq Tf \leq T_2 f$, wobei*

$$\begin{aligned} T_1 f(u) &:= \inf\big\{t \in [0,1] : f(t) \geq u\big\} \quad (\textit{mit } \inf(\emptyset) := 1), \\ T_2 f(u) &:= \sup\big\{t \in [0,1] : f(t) \leq u\big\} \quad (\textit{mit } \sup(\emptyset) := 0). \end{aligned}$$

Für beliebige $x \in \ell_\infty([0,1])$ ist dann

$$\big\|T(\mathrm{id} + x) - \mathrm{id}\big\| \ \leq \ \|x\| \quad \textit{und} \quad \big\|T(\mathrm{id} + x) - \mathrm{id} + x\big\| \ \leq \ \omega\big(x, \|x\|+\big).$$

Man beachte, dass $\hat{G}_n^{-1} = T_1\hat{G}_n$ und $\hat{G}_n = T_2\hat{G}_n^{-1}$. Daher folgt Satz 8.21 aus Satz 8.8 und Lemma 8.22, angewandt auf $x = \hat{G}_n - \mathrm{id}$.

Beweis von Lemma 8.22. Für beliebiges festes $u \in [0,1]$ und $t \in [-u, 1-u]$ ist

$$(\mathrm{id} + x)(u+t) \; = \; u + t + x(u+t) \begin{cases} > u & \text{falls } t > \|x\|, \\ < u & \text{falls } t < -\|x\|. \end{cases}$$

Folglich ist $T_2(\mathrm{id} + x)(u) \le u + \|x\|$ und $T_1(\mathrm{id} + x)(u) \ge u - \|x\|$, also insbesondere $\big\|T(\mathrm{id} + x) - \mathrm{id}\big\| \le \|x\|$. Mit $\omega_o := \omega\big(x, \|x\|+\big)$ ist ferner

$$\begin{aligned} u + t + x(u+t) \; &= \; u + t + x(u) + \big(x(u+t) - x(u)\big) \\ &\begin{cases} > u & \text{falls } |t| \le \|x\| \text{ und } t > -x(u) + \omega_o, \\ < u & \text{falls } |t| \le \|x\| \text{ und } t < -x(u) - \omega_o. \end{cases} \end{aligned}$$

Dies zeigt, dass $T_2(\mathrm{id} + x)(u) \le -x(u) + \omega_o$ und $T_1(\mathrm{id} + x)(u) \ge -x(u) - \omega_o$, also insbesondere $\big\|T(\mathrm{id} + x) - \mathrm{id} + x\big\| \le \omega_o$. $\square$

Anmerkung 8.23. Mithilfe von Satz 8.8 und Lemma 8.9 ergibt sich sogar, dass

$$\Big\| n^{1/2}(\hat{G}_n - \mathrm{id}) + n^{1/2}(\hat{G}_n^{-1} - \mathrm{id}) \Big\| \; = \; O_p\big(n^{-1/4}\log(n)^{1/2}\big).$$

Anmerkung 8.24. Mithilfe des Invarianzprinzips für Partialsummenprozesse, Satz 8.7, kann man auch direkt nachweisen, dass $\big(n^{1/2}(\hat{G}_n^{-1} - \mathrm{id})\big)_n$ in Verteilung gegen eine Brownsche Brücke konvergiert. Sei nämlich $W_n(t) := n^{-1/2}\sum_{i \le \lceil nt \rceil}(Y_i - 1)$ mit unabhängigen, exponentialverteilten Zufallsvariablen $Y_1, Y_2, Y_3, \ldots$. Dann ist

$$\hat{G}_n^{-1} \; =_{\mathcal{L}} \; \Big(\frac{\sum_{i \le \lceil nt \rceil} Y_i}{\sum_{i \le n+1} Y_i}\Big)_{t \in [0,1]} \; = \; \Big(\frac{n^{-1/2}W_n(t) + n^{-1}\lceil nt \rceil}{1 + n^{-1/2}W_n(1) + n^{-1}Y_{n+1}}\Big)_{t \in [0,1]};$$

siehe Abschnitt 3.4. Das heißt,

$$n^{1/2}(\hat{G}_n^{-1} - \mathrm{id}) \; =_{\mathcal{L}} \; \Big(\frac{W_n(t) - tW_n(1) + R_n(t)}{1 + S_n}\Big)_{t \in [0,1]}$$

mit $S_n := n^{-1/2}W_n(1) + n^{-1}Y_{n+1} = O_p(n^{-1/2})$ und

$$R_n(t) \; := \; n^{-1/2}(\lceil nt \rceil - nt) - tn^{-1/2}Y_{n+1}, \quad \|R_n\| \; = \; O_p(n^{-1/2}).$$

Demnach konvergiert $n^{1/2}(\hat{G}_n^{-1} - \mathrm{id})$ in Verteilung gegen $W - W(1)\,\mathrm{id}$, eine Brownsche Brücke. Aus Lemma 8.5, angewandt auf $x = \hat{G}_n^{-1} - \mathrm{id}$ und $T = T_2$, kann man nun Satz 8.21 sowie Satz 8.8 ableiten.

8.6 Anwendungen auf Rangstatistiken und -prozesse

Mit Hilfe von Donskers Theorem (Satz 8.8) und einfachen zusätzlichen Überlegungen kann man verschiedene Verfahren, die auf Rangstatistiken basieren, behandeln. Dazu eine einfache Vorüberlegung: Sei $\hat{F}_n$ die empirische Verteilungsfunktion von unabhängigen Zufallsvariablen X_{n1}, X_{n2},

$\ldots, X_{nn}$ mit stetiger Verteilungsfunktion F_n. Dann sind die X_{ni} fast sicher paarweise verschieden, und der Rang RX_{ni} von X_{ni} lässt sich schreiben als

$$RX_{ni} \ = \ n\hat{F}_n(X_{ni}).$$

Wenn man keine Stetigkeit von F_n voraussetzen möchte und die übliche Konvention bei Bindungen anwendet, ergibt sich

$$RX_{ni} \ = \ n\big(\hat{F}_n(X_{ni}-) + \hat{F}_n(X_{ni})\big)/2 + 1/2.$$

Aus Gründen der Bequemlichkeit betrachten wir aber nur stetige Verteilungsfunktionen.

Spearmans Rangkorrelationskoeffizient. Seien $(X_{n1}, Y_{n1}), (X_{n2}, Y_{n2}), \ldots, (X_{nn}, Y_{nn})$ stochastisch unabhängig und identisch verteilt nach einer Verteilung H_n auf $\mathbb{R} \times \mathbb{R}$. Ferner seien die marginalen Verteilungsfunktionen F_n und G_n von X_{n1} bzw. Y_{n1} stetig. Mit $\hat{F}_n$ bezeichnen wir die empirische Verteilungsfunktion (und Verteilung) der Beobachtungen X_{ni}, mit $\hat{G}_n$ diejenige der Beobachtungen Y_{ni}. Ferner sei $\hat{H}_n$ die empirische Verteilung aller Beobachtungen (X_{ni}, Y_{ni}). Mit den Rängen $RX_{ni} = n\hat{F}_n(X_{ni})$ und $RY_{ni} = n\hat{G}_n(Y_{ni})$ lässt sich Spearmans Rangkorrelationskoeffizient $\hat{\rho}_n$ wie folgt darstellen:

$$\sum_{i=1}^n \Big(RX_{ni} - \frac{n+1}{2}\Big)\Big(RY_{ni} - \frac{n+1}{2}\Big) \ = \ n^3\Big(\int (\hat{F}_n - 1/2) \otimes (\hat{G}_n - 1/2)\, d\hat{H}_n - \frac{1}{4n^2}\Big)$$

und

$$\sum_{i=1}^n \Big(RX_{ni} - \frac{n+1}{2}\Big)^2 \ = \ \sum_{i=1}^n \Big(RY_{ni} - \frac{n+1}{2}\Big)^2 \ = \ n^3\Big(\frac{1}{12} - \frac{1}{12n^2}\Big).$$

Dabei bezeichnet allgemein $h \otimes k$ die Funktion $(x, y) \mapsto h(x)k(y)$ auf $\mathbb{R} \times \mathbb{R}$ für Funktionen $h, k : \mathbb{R} \to \mathbb{R}$. Also ist

$$\begin{aligned} \hat{\rho}_n \ &= \ \frac{\sum_{i=1}^n \big(RX_{ni} - \frac{n+1}{2}\big)\big(RY_{ni} - \frac{n+1}{2}\big)}{\sqrt{\sum_{i=1}^n \big(RX_{ni} - \frac{n+1}{2}\big)^2 \sum_{i=1}^n \big(RY_{ni} - \frac{n+1}{2}\big)^2}} \\ &= \ 12 \int (\hat{F}_n - 1/2) \otimes (\hat{G}_n - 1/2)\, d\hat{H}_n + O_p(n^{-2}). \end{aligned}$$

Man kann $\hat{\rho}_n$ als Schätzer für die theoretische Größe

$$\begin{aligned} \rho_n \ &:= \ \mathrm{Corr}\big(F_n(X_{n1}), G_n(Y_{n1})\big) \\ &= \ 12 \int (F_n - 1/2) \otimes (G_n - 1/2)\, dH_n \end{aligned}$$

auffassen. Nun gilt folgende Entwicklung für $\hat{\rho}_n$:

Lemma 8.25.

$$\sqrt{n}\big(\hat{\rho}_n - \rho_n\big) \ = \ n^{-1/2} \sum_{i=1}^n \big(h_n(X_{ni}, Y_{ni}) - \mathbb{E}\, h_n(X_{n1}, Y_{n1})\big) + o_p(1)$$

mit

$$
\begin{aligned}
h_n(x,y) \;&:=\; 12(F_n(x)-1/2)(G_n(y)-1/2) \\
&\quad + 12\int \big(1_{[x,\infty)}\otimes(G_n-1/2)+(F_n-1/2)\otimes 1_{[y,\infty)}\big)\,dH_n.
\end{aligned}
$$

Anmerkung 8.26. Sind X_{n1} und Y_{n1} stochastisch unabhängig, dann ist $h_n = 12(F_n-1/2)\otimes(G_n-1/2)$, und $\mathrm{Var}(h_n(X_{n1},Y_{n1}))=1$.

Beweis von Lemma 8.25. Wir schreiben $\sqrt{n}(\hat F_n - F_n) = B_{n1}\circ F_n$ und $\sqrt{n}(\hat G_n - G_n) = B_{n2}\circ G_n$ mit stochastischen Prozessen B_{n1}, B_{n2} auf $[0,1]$, die in Verteilung gegen eine Brownsche Brücke konvergieren (und im allgemeinen abhängig sind). Dann ist

$$
\begin{aligned}
12^{-1}(\hat\rho_n-\rho_n) \;&=\; \int (F_n-1/2)\otimes(G_n-1/2)\,d(\hat H_n-H_n) \\
&\quad + \int (\hat F_n-F_n)\otimes(G_n-1/2)\,d\hat H_n + \int (F_n-1/2)\otimes(\hat G_n-G_n)\,d\hat H_n \\
&\quad + \int (\hat F_n-1/2)\otimes(\hat G_n-1/2)\,d\hat H_n + O_p(n^{-2}) \\
&=\; \int (F_n-1/2)\otimes(G_n-1/2)\,d(\hat H_n-H_n) \\
&\quad + \int (\hat F_n-F_n)\otimes(G_n-1/2)\,dH_n + \int (F_n-1/2)\otimes(\hat G_n-G_n)\,dH_n \\
&\quad + R_{n0}+R_{n1}+R_{n2}+O_p(n^{-2}),
\end{aligned}
$$

wobei

$$
|R_{n0}| \;=\; \Big|\int (\hat F_n-1/2)\otimes(\hat G_n-1/2)\,d\hat H_n\Big| \;\le\; n^{-1}\|B_{n1}\|_{[0,1]}\|B_{n2}\|_{[0,1]} \;=\; O_p(n^{-1})
$$

und

$$
\begin{aligned}
\sqrt{n}|R_{n1}| \;&=\; \Big|\int (B_{n1}\circ F_n)\otimes(G_n-1/2)\,d(\hat H_n-H_n)\Big|, \\
\sqrt{n}|R_{n2}| \;&=\; \Big|\int (F_n-1/2)\otimes(B_{n2}\circ G_n)\,d(\hat H_n-H_n)\Big|.
\end{aligned}
$$

Wir werden später zeigen, dass für beliebige $\delta\in(0,1]$ gilt:

$$
\sqrt{n}|R_{nj}| \;\le\; \omega(B_{nj},\delta+)\Big(1+\frac{\|\hat H_n-H_n\|_{\mathcal F}}{\delta}\Big) \tag{8.7}
$$

mit

$$
\mathcal F \;:=\; \big\{1_{[a,\infty)}\otimes(G_n-1/2), a\in\mathbb R\big\}\cup\big\{(F_n-1/2)\otimes 1_{[b,\infty)} : b\in\mathbb R\big\}.
$$

Man kann leicht zeigen, dass $\mathbb E\,\|\hat H_n - H_n\|_{\mathcal F}\to 0$. Wählt man nun $\delta=\delta_n$ von der gleichen Größenordnung wie $\mathbb E\,\|\hat H_n-H_n\|_{\mathcal F}$, dann ergibt sich, dass $R_{nj}=o_p(n^{-1/2})$ für $j=1,2$.

Mit den Techniken in Kapitel 9 kann man sogar zeigen, dass $\mathbb E\,\|\hat H_n-H_n\|_{\mathcal F} = O(n^{-1/2})$, und für $\delta_n = n^{-1/2}$ ergibt sich aus Lemma 8.9, dass $R_{nj} = O_p\big(\log(n)^{1/2}n^{-3/4}\big)$.

Ungleichung (8.7) ergibt sich aus folgender Konstruktion: Für eine beschränkte Funktion b auf $[0,1]$ mit $b(0) = 0$ setzen wir

$$b_\delta \ := \ \sum_{i=1}^{\lfloor \delta^{-1} \rfloor} \bigl(b(i\delta) - b((i-1)\delta)\bigr) \, 1_{[i\delta,1]}.$$

Dann ist $\|b - b_\delta\|_{[0,1]} \le \omega(b, \delta+)$, und b_δ ist eine Linearkombination von $\lfloor \delta^{-1} \rfloor$ Indikatorfunktionen der Form $1_{[v,1]}$, die mit Faktoren, welche absolut kleiner oder gleich $\omega(b, \delta+)$ sind, multipliziert werden. Ferner ist $1_{[v,1]} \circ F_n(x) = 1\{x \ge F_n^{-1}(v)\}$ und $1_{[v,1]} \circ G_n(y) = 1\{y \ge G_n^{-1}(v)\}$.

Wir wissen also, dass $\hat\rho_n - \rho_n$ bis auf einen Fehler der Ordnung $o_p(n^{-1/2})$ gleich

$$\begin{aligned} &12 \int (F_n - 1/2) \otimes (G_n - 1/2) \, d(\hat H_n - H_n) \\ &\quad + 12 \int (\hat F_n - F_n) \otimes (G_n - 1/2) \, dH_n + 12 \int (F_n - 1/2) \otimes (\hat G_n - G_n) \, dH_n \end{aligned}$$

ist, und man leicht nachrechnen, dass dies gleich $\int h_n \, d(\hat H_n - H_n)$ mit der angegebenen Funktion $h_n : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ ist. $\Box$

Mann-Whitney-Statistiken. Seien $X_{m1}, X_{m2}, \ldots, X_{mm}$ und $Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ stochastisch unabhängige Zufallsvariablen mit stetigen Verteilungsfunktionen $F_m(r) := \mathbb{P}(X_{mi} \le r)$ und $G_n(r) := \mathbb{P}(Y_{nj} \le r)$. Die Mann-Whitney-U-Statistik ist definiert als

$$U_{m,n} \ := \ \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n 1\{X_i \ge Y_j\}.$$

Sie ist eng verwandt mit der Wilcoxon-Rangsummenstatistik, lässt sich aber besser interpretieren als letztere. Mitunter betrachtet man noch einen Shift-Parameter $\theta \in \mathbb{R}$ und definiert

$$\hat u_{m,n}(\theta) \ := \ \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n 1\{X_i - \theta \ge Y_j\}.$$

Unter der Annahme, dass $F_m = G_n(\cdot - \theta_{m,n})$ für einen unbekannten Shift-Parameter $\theta_{m,n}$, sucht man dann $\hat\theta_{m,n}$ derart, dass $\hat u_{m,n}(\hat\theta_{m,n}) \approx 1/2$.

Mit den empirischen Verteilungsfunktionen $\hat F_m$ der X_{mi} und $\hat G_n$ der Y_{nj} kann man schreiben

$$\hat u_{m,n}(\theta) \ = \ \int \hat G_n(x - \theta) \, \hat F_m(dx) \ = \ 1 - \int \hat F_m((y + \theta)-) \, \hat G_n(dy).$$

Offensichtlich ist dies ein Schätzer für

$$u_{m,n}(\theta) \ := \ \int G_n(x - \theta) \, F_m(dx) \ = \ 1 - \int F_m(y + \theta) \, G_n(dy) \ = \ \mathbb{P}(X_{m1} - \theta \ge Y_{n1}).$$

Für $\hat u_{m,n} - u_{m,n}$ gilt folgende Entwicklung:

Lemma 8.27.

$$\hat{u}_{m,n}(\theta) - u_{m,n}(\theta) \;=\; \int G_n(x-\theta)\,(\hat{F}_m - F_m)(dx) - \int F_m(y+\theta)\,(\hat{G}_n - G_n)(dy) + R_{m,n}(\theta)$$

wobei

$$\sup_{\theta\in\mathbb{R}} |R_{m,n}(\theta)| \;=\; O_p\Big(\frac{\min(m,n)^{-1/4}(\log\min(m,n))^{1/2}}{\sqrt{\max(m,n)}}\Big) \quad \textit{für } \min(m,n)\to\infty.$$

Anmerkung 8.28. Falls $F_m \equiv G_n$, ist $u_{m,n}(0) = 1/2$, und $U_{m,n} - 1/2 = \hat{u}_{m,n}(0) - 1/2$ ist gleich

$$\frac{1}{m}\sum_{i=1}^m (F_m(X_{mi}) - 1/2) - \frac{1}{n}\sum_{j=1}^n (G_n(Y_{nj}) - 1/2) + o_p\big(\max(m,n)^{-1/2}\big).$$

Hieraus folgt, dass

$$\sqrt{\frac{12mn}{m+n}}\,(U_{m,n} - 1/2) \;\to_{\mathcal{L}}\; \mathcal{N}(0,1).$$

Anmerkung 8.29. Schreibt man $\sqrt{m}(\hat{F}_m - F_m) = B_{m1}\circ F_m$ und $\sqrt{n}(\hat{G}_n - G_n) = B_{n2}\circ G_n$ mit stochastisch unabhängigen uniformen empirischen Prozessen $B_{m1}, B_{n2} \to_{\mathcal{L}} B$, dann kann man auch schreiben:

$$\begin{aligned}\sqrt{\frac{mn}{m+n}}\big(\hat{u}_{m,n}(\theta) - u_{m,n}(\theta)\big) \;&=\; \sqrt{\frac{m}{m+n}}\int B_{n2}\circ G_n(x-\theta)\,F_m(dx) \\ &\quad - \sqrt{\frac{n}{m+n}}\int B_{m1}\circ F_m(y+\theta)\,G_n(dy) + \tilde{R}_{m,n}(\theta),\end{aligned}$$

wobei $\sup_{\theta\in\mathbb{R}} |\tilde{R}_{m,n}(\theta)| = o_p(1)$. Für große Werte von $\min(m,n)$ verhält sich dies in etwa wie

$$\sqrt{\frac{m}{m+n}}\int B_2\circ G_n(x-\theta)\,F_m(dx) - \sqrt{\frac{n}{m+n}}\int B_1\circ F_m(y+\theta)\,G_n(dy)$$

mit stochastisch unabhängigen Brownschen Brücken B_1, B_2.

Beweis von Lemma 8.27. Durch Teleskopieren und eine einfache Rechnung ergibt sich die Darstellung

$$\begin{aligned}&\hat{u}_{m,n}(\theta) - u_{m,n}(\theta) \\ &= \int G_n(x-\theta)\,(\hat{F}_m - F_m)(dx) + \int (\hat{G}_n - G_n)(x-\theta)\,F_m(dx) + R_{m,n}(\theta) \\ &= \int G_n(x-\theta)\,(\hat{F}_m - F_m)(dx) - \int F_m(y+\theta)\,(\hat{G}_n - G_n)(dy) + R_{m,n}(\theta)\end{aligned}$$

mit

$$\begin{aligned}R_{m,n}(\theta) \;&:=\; \int (\hat{G}_n - G_n)(x-\theta)\,(\hat{F}_m - F_m)(dx) \\ &=\; \int (\hat{F}_m - F_m)((y+\theta)-)\,(\hat{G}_n - G_n)(dy).\end{aligned}$$

Nun schreiben wir $\sqrt{m}(\hat{F}_m - F_m) = B_{m1} \circ F_m$ und $\sqrt{n}(\hat{G}_n - G_n) = B_{n2} \circ G_n$, wobei $B_{m1}, B_{n2} \to_{\mathcal{L}} B$. Durch Approximation von B_{m1} und B_{n2} durch Treppenfunktionen wie im Beweis von Lemma 8.25 kann man zeigen, dass

$$\begin{aligned}
|R_{m,n}(\theta)| &= n^{-1/2}\Big|\int (B_{n2} \circ G_n)(x-\theta)\,(\hat{F}_m - F_m)(dx)\Big| \\
&\le n^{-1/2}\omega(B_{n2}, \delta+)\Big(1 + \frac{\|\hat{F}_m - F_m\|_{\mathbb{R}}}{\delta}\Big) \qquad \text{bzw.} \\
|R_{m,n}(\theta)| &= m^{-1/2}\Big|\int (B_{m1} \circ F_m)((y+\theta)-)\,(\hat{G}_n - G_n)(dx)\Big| \\
&\le m^{-1/2}\omega(B_{m1}, \delta+)\Big(1 + \frac{\|\hat{G}_n - G_n\|_{\mathbb{R}}}{\delta}\Big)
\end{aligned}$$

für beliebige $\delta \in (0,1]$. Setzt man $\delta = \min(m,n)^{-1/2}$, dann ergibt dies zusammen mit Lemma 8.9 die behauptete Ungleichung für $\sup_{\theta\in\mathbb{R}} |R_{m,n}(\theta)|$. □

Wilcoxons signierte Ränge. Wie zu Beginn seinen $X_{n1}, X_{n2}, \ldots, X_{nn}$ stochastisch unabhängige Zufallsvariablen mit stetiger Verteilung(sfunktion) F_n. Die empirische Verteilung(sfunktion) der X_{ni} sei $\hat{F}_n$. Für die Nullhypothese, dass F_n um einen Punkt $\theta \in \mathbb{R}$ symmetrisch ist, betrachten wir Wilcoxons Signed-Rank-Statistik

$$\begin{aligned}
\hat{\tau}_n(\theta) &:= \frac{1}{n}\sum_{i=1}^n \operatorname{sign}(X_{ni}-\theta)\hat{F}_n\big([\theta - |X_{ni}-\theta|, \theta + |X_{ni}-\theta|]\big) \\
&= \int \operatorname{sign}(x-\theta)\big(\hat{F}_n(\theta + |x-\theta|) - \hat{F}_n(\theta - |x-\theta|)\big)\,\hat{F}_n(dx) + R_{n1}(\theta) \\
&= \int \big(\hat{F}_n(x) - \hat{F}_n(2\theta - x)\big)\,\hat{F}_n(dx) + R_{n1}(\theta) \\
&= 1/2 - \int \hat{F}_n(2\theta - x)\,\hat{F}_n(dx) + R_{n1}(\theta) + (2n)^{-1},
\end{aligned}$$

wobei $|R_{n1}(\theta)| \le n^{-1}$. Dies ist ein Schätzer für

$$\begin{aligned}
\tau_n(\theta) &:= \int \operatorname{sign}(x-\theta)F_n\big([\theta \pm |x-\theta|]\big)\,F_n(dx) \\
&= \int \big(F_n(x) - F_n(2\theta - x)\big)\,F_n(dx) \\
&= 1/2 - \int F_n(2\theta - x)\,F_n(dx).
\end{aligned}$$

Auch hier kann man durch Teleskopieren und Abschätzung eines Restterms die Differenz $\hat{\tau}_n - \tau_n$ approximieren:

Lemma 8.30.

$$\hat{\tau}_n(\theta) - \tau_n(\theta) = -2\int F_n(2\theta - x)\,(\hat{F}_n - F_n)(dx) + R_n(\theta)$$

wobei

$$\sup_{\theta\in\mathbb{R}} |R_n(\theta)| = O_p\big(n^{-3/4}(\log n)^{1/2}\big).$$

Anmerkung 8.31. Ist F_n symmetrisch um $\theta_n \in \mathbb{R}$, das heißt,

$$F_n(\theta_n + h) \; = \; 1 - F_n(\theta_n - h) \quad \text{für alle } h \in \mathbb{R},$$

dann ist $-F_n(2\theta_n - x) = F_n(x) - 1$, so dass $\tau_n(\theta_n) = 0$ und

$$\sqrt{n}\,\hat{\tau}_n(\theta_n) \; = \; n^{-1/2}\sum_{i=1}^{n}(1 - 2F_n(X_{ni})) + o_p(1) \; \to_{\mathcal{L}} \; \mathcal{N}(0, 1/3).$$

Anmerkung 8.32. Schreibt man $\sqrt{n}(\hat{F}_n - F_n) = B_n \circ F_n$ mit $B_n \to_{\mathcal{L}} B$, dann ist

$$\sqrt{n}\big(\hat{\tau}_n(\theta) - \tau_n(\theta)\big) \; = \; 2\int B_n \circ F_n(2\theta - x)\, F_n(dx) + \tilde{R}_n(\theta)$$

mit $\sup_{\theta\in\mathbb{R}} |\tilde{R}_n(\theta)| = o_p(1)$. Dies verhält sich asymptotisch wie

$$2\int B \circ F_n(2\theta - x)\, F_n(dx).$$

Beweis von Lemma 8.30. Schreibt man $\sqrt{n}(\hat{F}_n - F_n) = B_n \circ F_n$ mit $B_n \to_{\mathcal{L}} B$, dann gilt für beliebige $\theta \in \mathbb{R}$:

$$\begin{aligned}
&\int \hat{F}_n(2\theta - x)\, \hat{F}_n(dx) - \int F_n(2\theta - x)\, F_n(dx) \\
&= \; \int (\hat{F}_n - F_n)(2\theta - x)\, F_n(dx) + \int F_n(2\theta - x)\, (\hat{F}_n - F_n)(dx) + R_{n2}(\theta) \\
&= \; 2\int F_n(2\theta - x)\, (\hat{F}_n - F_n)(dx) + R_{n2}(\theta),
\end{aligned}$$

wobei

$$\begin{aligned}
|R_{n2}(\theta)| \; &= \; \Big|\int (\hat{F}_n - F_n)(2\theta - x)\, (\hat{F}_n - F_n)(dx)\Big| \\
&= \; n^{-1/2}\Big|\int B_n \circ F_n(2\theta - x)\, (\hat{F}_n - F_n)(dx)\Big| \\
&\le \; n^{-1/2}\omega\big(B_n, n^{-1/2}+\big)\big(1 + \sqrt{n}\|\hat{F}_n - F_n\|_{\mathbb{R}}\big) \\
&= \; O_p\big(n^{-3/4}(\log n)^{1/2}\big)
\end{aligned}$$

nach Lemma 8.9; siehe auch den Beweis von Lemma 8.25. □

8.7 Aufgaben

Aufgabe 8.1. Beweisen Sie die unter Anmerkung 8.6 aufgeführten Eigenschaften der Brownschen Bewegung und -Brücke. (Vorsicht bei $t \mapsto tW(1/t)$!)

Aufgabe 8.2. Seien $F, F_n, \hat{F}_n$ wie in Abschnitt 8.3. Zeigen Sie, dass aus $\omega(F_n, 0+) = o(n^{-1/2})$ folgt, dass auch $\omega(\hat{F}_n, 0+) = o_p(n^{-1/2})$.

Aufgabe 8.3. Formulieren und beweisen Sie einen zentralen Grenzwertsatz analog zu Satz 8.10 für folgendes Funktional:

$$T(P) := \int xw(F(x))\,P(dx) \Big/ \int w(F(x))\,P(dx),$$

wobei $w : [0,1] \to [0,\infty)$ stetig differenzierbar ist mit $\{w > 0\} = (\alpha, 1-\alpha)$ für eine Konstante $0 < \alpha < 1/2$.

Aufgabe 8.4. Sei B eine Brownsche Brücke. Zeigen Sie, dass

$$\mathbb{P}\big(B(t) \geq (a(1-t)+bt)\eta \text{ für ein } t \in [0,1]\big) \;=\; \exp(-2ab\eta^2)$$

für beliebige $\eta, a, b \geq 0$.

Hinweis: Verwenden Sie die Tatsachen, dass

$$W(t) \;:=\; (1+t)B(t/(1+t)) \quad \text{und} \quad \tilde{W}(t) \;:=\; \sigma^{-1}W(\sigma^2 t)$$

mit $\sigma > 0$ Brownsche Bewegungen auf $[0,\infty)$ definieren.

Aufgabe 8.5. Sei B eine Brownsche Brücke. Zeigen Sie, dass $U := \arg\max_{t\in[0,1]} B(t)$ fast sicher eindeutig definiert ist und folgende Eigenschaft hat: Für beliebige $r \in (0,1)$ ist

$$\mathbb{P}\big(U \leq r \,\big|\, B(r)\big) \;=\; r \quad \text{fast sicher.}$$

Hinweis: Zeigen Sie, dass man B mit zwei stochastisch unabhängigen Brownschen Brücken B_1, B_2 und einer davon unabhängigen Zufallsvariable $Y \sim \mathcal{N}(0, r(1-r))$ darstellen kann: Für $0 \leq \lambda \leq 1$ setze

$$\begin{aligned} B(\lambda r) &:= \sqrt{r}B_1(\lambda) + \lambda Y, \\ B(r + \lambda(1-r)) &:= \sqrt{1-r}B_2(\lambda) + (1-\lambda)Y. \end{aligned}$$

Verwenden Sie nun Aufgabe 8.4.

Aufgabe 8.6. Sei B eine Brownsche Brücke. Bestimmen Sie mit Hilfe des Spiegelungsprinzips für die Brownsche Bewegung die Wahrscheinlichkeit

$$\mathbb{P}\Big(\min_{t\in[0,1]} B(t) \leq -r \text{ und } \max_{t\in[0,1]} B(t) \geq s\Big)$$

für $r, s > 0$. (Hieraus kann man die Verteilung von $\max_{t,u\in[0,1]}\big(B(u) - B(t)\big)$ ableiten.)

Kapitel 9

Chaining

9.1 Maximalungleichungen

Sei Z ein stochastischer Prozess auf einem präkompakten pseudometrischen Raum $(\mathcal{T}, \rho)$. Wir nehmen stets an, dass entweder $\mathcal{T}$ abzählbar ist oder Z stetige Pfade bezüglich ρ hat.

Das *Chaining* ist eine Technik, mit der man $\|Z\|$ oder $\omega(Z, \delta)$ nach oben abschätzen kann. Ein entscheidender Schritt ist dabei, das Maximum endlich vieler Zufallsvariablen $Y_1, Y_2, \ldots, Y_m$ abzuschätzen. Eine naive Schranke ist $\mathbb{E} \max_{i \le m} |Y_i| \le \sum_{i \le m} \mathbb{E}\,|Y_i|$. Das folgende Lemma liefert eine Verfeinerung hiervon:

Lemma 9.1 (Pisier). *Sei $\psi : \mathbb{R} \to [0, \infty)$ eine gerade, konvexe Funktion mit $\psi(0) = 0$ und $\psi \not\equiv 0$. Für beliebige $m \in \mathbb{N}$ und Zufallsvariablen $Y_1, Y_2, \ldots, Y_m$ ist*

$$\mathbb{E} \max_{i \le m} |Y_i| \ \le \ \psi^{-1}\Big(\sum_{i=1}^m \mathbb{E}\, \psi(Y_i) \Big),$$

wobei $\psi^{-1}(v) := \max\big\{u \ge 0 : \psi(u) \le v\big\}$ für $v \ge 0$.

Mithilfe dieses Lemmas kann man folgenden Satz beweisen:

Satz 9.2 (Chaining, I). *Sei $\psi : \mathbb{R} \to [0, \infty)$ wie in Lemma 9.1. Für beliebige $s, t \in \mathcal{T}$ sei*

$$\begin{aligned} &\mathbb{E}\, \psi\Big(\frac{Z(s) - Z(t)}{\rho(s,t)} \Big) \le 1, \ \text{falls } \rho(s,t) > 0, \\ &Z(s) = Z(t) \text{ fast sicher, falls } \rho(s,t) = 0. \end{aligned}$$

***(a)** Für beliebige $t_o \in \mathcal{T}$ und $\delta(t_o) := \sup_{t \in \mathcal{T}} \rho(t, t_o)$ ist*

$$\begin{aligned} \mathbb{E}\, \|Z\| \ &\le \ \mathbb{E}\, |Z(t_o)| + 4 \int_0^{\delta(t_o)/2} \psi^{-1}\big(D(u, \mathcal{T}, \rho)\big)\, du \\ &\le \ \mathbb{E}\, |Z(t_o)| + 8 \int_0^{\delta(t_o)/4} \psi^{-1}\big(N(u, \mathcal{T}, \rho)\big)\, du. \end{aligned}$$

(b) *Für beliebige* $\delta > 0$ *ist*

$$\mathbb{E}\,\omega(Z,\delta) \ \le\ 16\int_0^{\delta/8} \psi^{-1}\big(N(u,\mathcal{T},\rho)^2\big)\,du.$$

Beweis von Lemma 9.1. Aus der Konvexität von ψ und der Jensenschen Ungleichung folgt, dass

$$\begin{aligned}\psi\Big(\mathbb{E}\max_{i\le m}|Y_i|\Big) &\le \mathbb{E}\,\psi\Big(\max_{i\le m}|Y_i|\Big)\\ &= \mathbb{E}\max_{i\le m}\psi(Y_i)\\ &\le \sum_{i=1}^m \mathbb{E}\,\psi(Y_i).\end{aligned}$$

Dabei ergibt sich der zweite Schritt aus der Tatsache, dass ψ auf $[0,\infty)$ monoton wachsend ist, und der dritte Schritt folgt aus der Tatsache, dass $\psi \ge 0$. Nun folgt er der Definition von ψ^{-1}, dass $\mathbb{E}\max_{i\le m}|Y_i|$ nicht größer ist als $\psi^{-1}\big(\sum_{i\le m}\mathbb{E}\,\psi(Y_i)\big)$. □

Beweis von Satz 9.2, Teil (a). Sei $\mathcal{T} = \{t_o, t_1, t_2, \ldots\}$, oder sei $\{t_o, t_1, t_2, \ldots\}$ eine dichte Teilmenge von $\mathcal{T}$, und Z habe stetige Pfade. Nach dem Satz von der monotonen Konvergenz ist in beiden Fällen $\mathbb{E}\,\|Z\| = \lim_{k\to\infty}\mathbb{E}\,\|Z\|_{\{t_o,t_1,\ldots,t_k\}}$. Deshalb kann man ohne Einschränkung annehmen, dass $\mathcal{T}$ endlich ist. Außerdem kann man annehmen, dass $\rho(s,t) > 0$ für verschiedene $s,t\in\mathcal{T}$.

Nun sei $\mathcal{T}_0 := \{t_o\}$ und $\delta_0 := \delta(t_o)$. Für $k = 1,2,3,\ldots$ wählen wir induktiv eine maximale Teilmenge $\mathcal{T}_k$ von $\mathcal{T}$, so dass gilt:

$$\mathcal{T}_{k-1} \subset \mathcal{T}_k \quad\text{und}\quad \rho(s,t) > \delta_k := 2^{-k}\delta_0 \quad \text{für verschiedene } s,t\in\mathcal{T}_k.$$

Die Maximalität von $\mathcal{T}_k$ für $k\ge 1$ und die Definition von $\delta(t_o) = \delta_0$ implizieren, dass

$$\#\mathcal{T}_k \ \le\ D(\delta_k,\mathcal{T},\rho) \quad\text{und}\quad \rho(t,\mathcal{T}_k) \ \le\ \delta_k \quad \text{für alle } t\in\mathcal{T} \text{ und } k\in\mathbb{N}_0.$$

Insbesondere ist $\mathcal{T}_0 \subset \mathcal{T}_1 \subset \mathcal{T}_2 \subset \cdots \subset \mathcal{T}_\ell = \mathcal{T}$ für ein $\ell\in\mathbb{N}$. Für jedes $k\in\mathbb{N}_0$ und $t\in\mathcal{T}$ wählen wir einen Punkt $\pi_k t\in\mathcal{T}_k$ derart, dass $\rho(t,\pi_k t)\le\delta_k$. Dann ist

$$\begin{aligned}\|Z\|_{\mathcal{T}_{k+1}} &\le \max_{t\in\mathcal{T}_{k+1}}\big(|Z(\pi_k t)| + \big|Z(t)-Z(\pi_k t)\big|\big)\\ &\le \|Z\|_{\mathcal{T}_k} + \max_{t\in\mathcal{T}_{k+1}}|Z(t)-Z(\pi_k t)|,\end{aligned} \tag{9.1}$$

und nach Lemma 9.1 ist

$$\begin{aligned}\mathbb{E}\max_{t\in\mathcal{T}_{k+1}}|Z(t)-Z(\pi_k t)| &= \delta_k\,\mathbb{E}\max_{t\in\mathcal{T}_{k+1}}\frac{|Z(t)-Z(\pi_k t)|}{\delta_k}\\ &\le \delta_k\psi^{-1}(\#\mathcal{T}_{k+1})\\ &\le \delta_k\psi^{-1}\big(D(\delta_{k+1},\mathcal{T},\rho)\big)\\ &= 4(\delta_{k+1}-\delta_{k+2})\psi^{-1}\big(D(\delta_{k+1},\mathcal{T},\rho)\big)\\ &\le 4\int_{\delta_{k+2}}^{\delta_{k+1}}\psi^{-1}\big(D(u,\mathcal{T},\rho)\big)\,du.\end{aligned}$$

Letztere Ungleichung folgt aus der Tatsache, dass $\psi^{-1}(D(u,\mathcal{T},\rho))$ monoton fallend ist in $u > 0$. Kombiniert man diese Ungleichungen für $k = 0,1,2,\ldots,\ell-1$, dann folgt, dass

$$\begin{aligned} \mathbb{E}\,\|Z\|_{\mathcal{T}} &\le \mathbb{E}\,\|Z\|_{\mathcal{T}_0} + 4\sum_{k=0}^{\infty}\int_{\delta_{k+2}}^{\delta_{k+1}} \psi^{-1}\big(D(u,\mathcal{T},\rho)\big)\,du \\ &= \mathbb{E}\,|Z(t_o)| + 4\int_0^{\delta(t_o)/2} \psi^{-1}\big(D(u,\mathcal{T},\rho)\big)\,du. \end{aligned}$$

Zusammen mit der Ungleichung $D(u,\mathcal{T},\rho) \le N(u/2,\mathcal{T},\rho)$ folgt hieraus auch die zweite behauptete Ungleichung. □

Beweis von Satz 9.2, Teil (b). Dieses Resultat folgt aus Teil (a), indem man das dortige Tupel $(\mathcal{T},\rho,Z,t_o)$ durch $(\tilde{\mathcal{T}},\tilde\rho,\tilde Z,\tilde t_o)$ ersetzt, wobei

$$\begin{aligned} \tilde{\mathcal{T}} = \tilde{\mathcal{T}}(\delta) &:= \big\{(s,t)\in\mathcal{T}\times\mathcal{T} : \rho(s,t) < \delta\big\}, \\ \tilde\rho[(s,t),(s',t')] &:= \min\big(\rho(s,s')+\rho(t,t'),\rho(s,t)+\rho(s',t')\big) \quad \text{für } s,t,s',t'\in\mathcal{T}, \\ \tilde Z(s,t) &:= Z(s)-Z(t), \\ \tilde t_o &:= (t_o,t_o) \end{aligned}$$

für ein beliebiges $t_o \in \mathcal{T}$. Dann ist nämlich $\omega(Z,\delta) = \|\tilde Z\|_{\tilde{\mathcal{T}}}$ und $\tilde Z(\tilde t_o) \equiv 0$.

Man kann leicht zeigen, dass $\tilde\rho$ eine Pseudometrik auf $\tilde{\mathcal{T}}$ ist, und dass

$$N(u,\tilde{\mathcal{T}},\tilde\rho) \le N(u,\mathcal{T}\times\mathcal{T},\tilde\rho) \le N(u/2,\mathcal{T},\rho)^2 \quad \text{für alle } u>0.$$

In der Tat gilt letztere Ungleichung für die “Manhattan-Distanz” bzw. “Mannheim-Distanz”

$$d_{\mathrm{M}}[(s,t),(s',t')] := \rho(s,s')+\rho(t,t')$$

anstelle von $\tilde\rho$; siehe Abbildung 9.1. Ferner ist

$$\tilde\rho[(s,t),(s',t')] = \min\big(d_{\mathrm{M}}[(s,t),(s',t')], d_{\mathrm{M}}[(s,t),\Delta] + d_{\mathrm{M}}[(s',t'),\Delta]\big)$$

mit der “Diagonalen” $\Delta := \{(t,t) : t\in\mathcal{T}\}$ von $\mathcal{T}$. Eine mögliche anschauliche Deutung: Der Abstand d_{M} misst Fußwege in Mannheim. Man kann zu Fuß von (s,t) nach (s',t') gelangen und legt dabei den Weg $d_{\mathrm{M}}[(s,t),(s',t')]$ zurück. Oder man läuft von (s,t) aus auf dem schnellsten Wege zur Diagonalen Δ, zum Beipiel nach (s,s) oder (t,t), fährt ein Stück mit einer virtuellen Luxus-U-Bahn (mit beliebigen Haltestellen!) entlang Δ bis zum Punkt (s',s') oder (t',t'), und läuft von dort aus nach (s',t'). Also beschreibt $\tilde\rho$ den kürzesten Fußweg bei möglicher U-Bahnbenutzung.

Schließlich kann man mithilfe der Konvexität von ψ zeigen, dass

$$\mathbb{E}\,\psi\Big(\frac{\tilde Z(\tilde s)-\tilde Z(\tilde t)}{\tilde\rho(\tilde s,\tilde t)}\Big) \le 1 \quad \text{für beliebige } \tilde s,\tilde t\in\tilde{\mathcal{T}},$$

wobei $\psi(0/0) := 0$ und $\psi(a/0) := \infty$ für $a>0$. Da $\sup_{\tilde t\in\tilde{\mathcal{T}}}\tilde\rho(\tilde t,\tilde t_o)\le\delta$, folgt aus Teil (a), dass

$$\mathbb{E}\,\|\tilde Z\|_{\tilde{\mathcal{T}}} \le 8\int_0^{\delta/4}\psi^{-1}\big(N(u/2,\mathcal{T},\rho)^2\big)\,du = 16\int_0^{\delta/8}\psi^{-1}\big(N(u,\mathcal{T},\rho)^2\big)\,du.$$ □

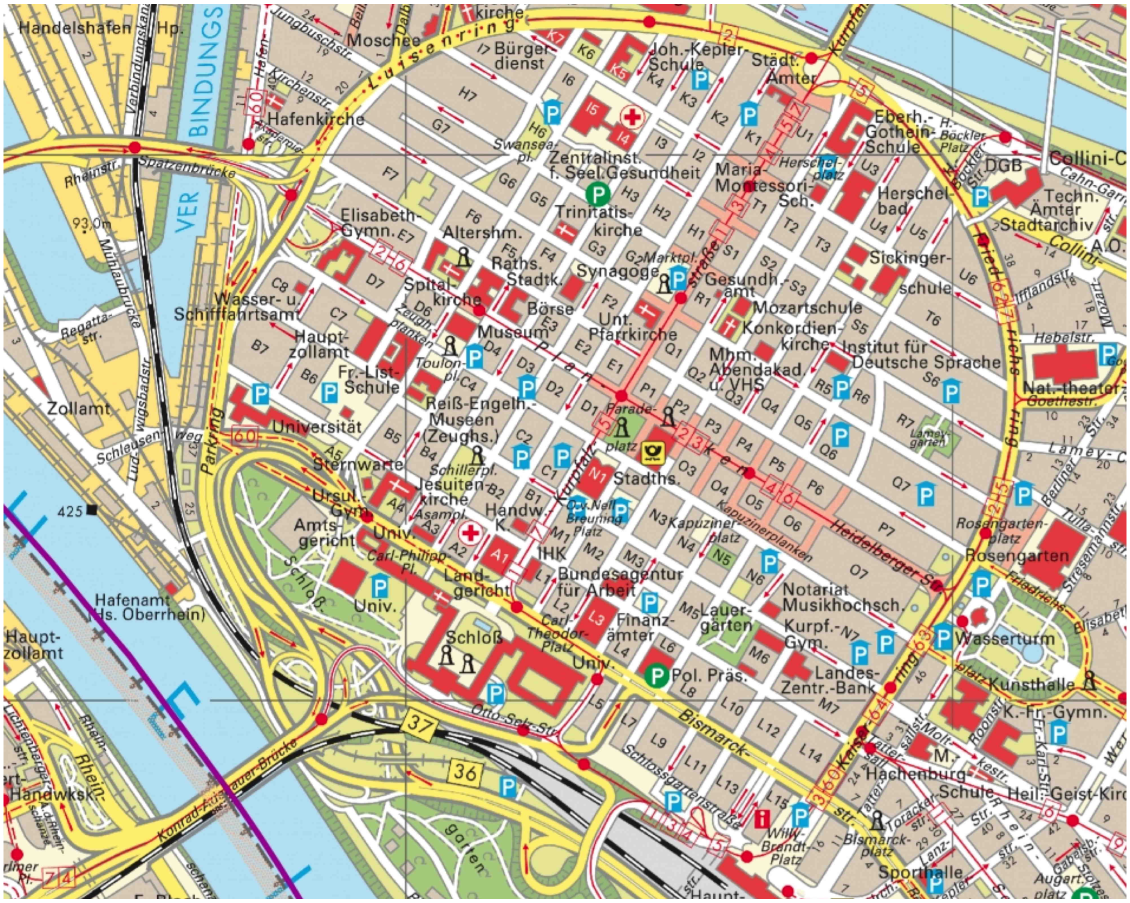

Abbildung 9.1: Die Innenstadt von Mannheim.

Die Maximalungleichung von Lemma 9.1 kann man auf vielfältige Weisen verfeinern, siehe zum Beispiel Aufgabe 9.8. Nun beschreiben wir eine Verfeinerung unter einer zusätzlichen Annahme an die Funktion ψ. Dazu definieren wir für eine Zufallsvariable Y ihre *Orlicz-Norm*

$$\|Y\|_\psi \ := \ \inf\bigl\{r > 0 : \mathbb{E}\,\psi(Y/r) \le 1\bigr\} \quad (\text{mit } \inf(\emptyset) := \infty).$$

Dies ist eine Seminorm auf dem Vektorraum $\mathcal{L}^\psi(\mathbb{P})$ aller Zufallsvariablen Y auf $(\Omega, \mathcal{A}, \mathbb{P})$, so dass $\mathbb{E}\,\psi(Y/r) < \infty$ für ein $r > 0$; siehe Aufgabe 9.2. Für den Spezialfall $\psi(x) := |x|^p$ erhält man den üblichen Raum $\mathcal{L}^p(\mathbb{P})$ mit der Seminorm $\|Y\|_p := (\mathbb{E}\,|Y|^p)^{1/p}$.

Lemma 9.3. *Sei $\psi : \mathbb{R} \to [0, \infty)$ wie in Lemma 9.1, und es sei*

$$\limsup_{x,y\to\infty} \frac{\psi(x)\psi(y)}{\psi(xy)} \ < \ \infty.$$

Dann existiert eine universelle Konstante $C(\psi)$, so dass für beliebige $m \in \mathbb{N}$ und Zufallsvariablen $Y_1, Y_2, \ldots, Y_m$ gilt:

$$\Bigl\|\max_{i\le m} |Y_i|\Bigr\|_\psi \ \le \ C(\psi)\psi^{-1}(m) \max_{i\le m} \|Y_i\|_\psi.$$

Dieses Lemma ergibt eine verbesserte Version von Satz 9.2:

Satz 9.4 (Chaining, II). *Sei $\psi : \mathbb{R} \to [0,\infty)$ wie in Lemma 9.3. Für beliebige $s,t \in \mathcal{T}$ sei*

$$\|Z(s) - Z(t)\|_\psi \ \leq\ \rho(s,t).$$

Dann gilt für eine universelle Konstante $C(\psi) \in (0,\infty)$:

***(a)** Für beliebige $t_o \in \mathcal{T}$ und $\delta(t_o) := \sup_{t\in\mathcal{T}} \rho(t,t_o)$ ist*

$$\big\|\|Z\|\big\|_\psi \ \leq\ \|Z(t_o)\|_\psi + C(\psi)\int_0^{\delta(t_o)/4} \psi^{-1}\big(N(u,\mathcal{T},\rho)\big)\,du.$$

***(b)** Für beliebige $\delta > 0$ ist*

$$\big\|\omega(Z,\delta)\big\|_\psi \ \leq\ 2C(\psi)\int_0^{\delta/8} \psi^{-1}\big(N(u,\mathcal{T},\rho)^2\big)\,du.$$

Subgaußsche Variablen. Ein wichtiger Spezialfall für ψ ist die Funktion

$$\psi_o(x) \ :=\ \exp(x^2) - 1.$$

Hier kann man zeigen, dass

$$\psi_o^{-1}(r) \ =\ \sqrt{\log(1+r)},$$

und für beliebige Zufallsvariablen $Y \in \mathbb{R}$ und Konstanten $L > 0$ gilt:

$$\mathbb{P}(|Y| \geq \eta) \ \leq\ 2\exp(-\eta^2/L) \ \text{ für alle } \eta \geq 0, \quad \text{falls} \quad \|Y\|_{\psi_o} \ \leq\ \sqrt{L}, \tag{9.2}$$

$$\|Y\|_{\psi_o} \ \leq\ \sqrt{3L} \quad \text{falls} \quad \mathbb{P}(|Y| > \eta) \ \leq\ 2\exp(-\eta^2/L) \ \text{ für alle } \eta \geq 0; \tag{9.3}$$

siehe Aufgabe 9.3 und Aufgabe 9.4. Außerdem ist

$$\begin{aligned} \psi_o^{-1}(r) \ &\leq\ \sqrt{\log(3)/\log(2)}\sqrt{\log r} \quad \text{und} \\ \psi_o^{-1}(r^2) \ &\leq\ \sqrt{2\log(5)/\log(4)}\sqrt{\log r} \quad \text{für alle } r \geq 2. \end{aligned}$$

Aus Satz 9.4 kann man daher folgendes Resultat ableiten:

Korollar 9.5. *Angenommen, Z hat subgaußsche Zuwächse bezüglich ρ. Das bedeutet, für beliebige $s,t \in \mathcal{T}$ und $\eta \geq 0$ ist $\mathbb{P}\big(|Z(s) - Z(t)| > \rho(s,t)\eta\big) \leq 2\exp(-\eta^2/2)$. Für eine universelle Konstante C gilt dann:*

***(a)** Für beliebige $t_o \in \mathcal{T}$ und $\delta(t_o) := \sup_{t\in\mathcal{T}} \rho(t,t_o)$ ist*

$$\big\|\|Z\|\big\|_{\psi_o} \ \leq\ \|Z(t_o)\|_{\psi_o} + C\int_0^{\delta(t_o)} \sqrt{\log N(u,\mathcal{T},\rho)}\,du.$$

***(b)** Für beliebige $\delta > 0$ ist*

$$\big\|\omega(Z,\delta)\big\|_{\psi_o} \ \leq\ C\int_0^{\delta} \sqrt{\log N(u,\mathcal{T},\rho)}\,du. \qquad \square$$

Beweis von Lemma 9.3. Ohne Einschränkung sei $\|Y_i\|_\psi \le 1$ für alle $i \le m$. Nach Voraussetzung existieren Konstanten $r_o, s_o \ge 1$ derart, dass $\psi(x)\psi(y) \le s_o \psi(xy)$ für alle $x, y \ge r_o$. Für $z \ge 0$ und $r \ge r_o$ ist also

$$\psi(z/r) \;\le\; 1\{z/r < r_o\}\psi(r_o) + 1\{z/r \ge r_o\}\psi(z/r) \;\le\; \psi(r_o) + s_o\psi(z)/\psi(r).$$

Folglich ist

$$\begin{aligned} \mathbb{E}\,\psi\Big(\max_{i\le m}|Y_i| \Big/ \max\big(\psi^{-1}(m), r_o\big)\Big) &\le \psi(r_o) + s_o\,\mathbb{E}\,\psi\Big(\max_{i\le m}|Y_i|\Big) \Big/ \max(m, \psi(r_o)) \\ &\le \psi(r_o) + s_o. \end{aligned}$$

Wegen $\psi(0) = 0$ und der Konvexität von ψ ist $\psi(\cdot/q) \le \psi/q$ für beliebige $q \ge 1$. Folglich ist

$$\begin{aligned} \Big\|\max_{i\le m}|Y_i|\Big\|_\psi &\le (\psi(r_o)+s_o)\max\big(\psi^{-1}(m), r_o\big) \\ &\le (\psi(r_o)+s_o)\max\big(r_o/\psi^{-1}(1), 1\big)\psi^{-1}(m). \end{aligned}$$

□

Beweis von Satz 9.4. Mit den gleichen Bezeichnungen wie im Beweis von Satz 9.2 folgt aus (9.1) und Lemma 9.3, dass

$$\begin{aligned} \big\|\|Z\|_{\mathcal{T}_{k+1}}\big\|_\psi &\le \big\|\|Z\|_{\mathcal{T}_k}\big\|_\psi + C(\psi)\delta_k\psi^{-1}\big(N(\delta_{k+2}, \mathcal{T}, \rho)\big) \\ &\le \big\|\|Z\|_{\mathcal{T}_k}\big\|_\psi + 8C(\psi)\int_{\delta_{k+3}}^{\delta_{k+2}} \psi^{-1}\big(N(u, \mathcal{T}, \rho)\big)\,du. \end{aligned}$$

Nun argumentiert man wie im Beweis von Satz 9.2. □

Beweis von Korollar 9.5.. Nach (9.3) erfüllt der Prozess $Z/\sqrt{6}$ die Voraussetzungen von Satz 9.4 mit $\psi = \psi_o$. Daher ist

$$\begin{aligned} \big\|\|Z\|\big\|_{\psi_o} &\le \|Z(t_o)\|_{\psi_o} + \sqrt{6}C(\psi_o)\int_0^{\delta(t_o)/4} \psi_o^{-1}\big(N(u, \mathcal{T}, \rho)\big)\,du, \\ \big\|\omega(Z, \delta)\big\|_{\psi_o} &\le 2\sqrt{6}C(\psi_o)\int_0^{\delta/8} \psi_o^{-1}\big(N(u, \mathcal{T}, \rho)^2\big)\,du. \end{aligned}$$

Doch $N(u, \mathcal{T}, \rho) \ge 2$ für $0 < u < \delta(t_o)$. Und aus $N(u_o, \mathcal{T}, \rho) = 1$ für ein $u_o \le \delta/8$ würde folgen, dass $\omega(Z, \delta) = \omega(Z, 2u_o)$. Wir können also ohne Einschränkung annehmen, dass die in den vorangehenden Integralen auftretenden Überdeckungszahlen $N(u, \mathcal{T}, \rho)$ größer oder gleich 2 sind. Doch dann ist $\psi_o^{-1}\big(N(u, \mathcal{T}, \rho)$ bzw. $\psi_o^{-1}\big(N(u, \mathcal{T}, \rho)^2\big)$ nicht größer als $C'\sqrt{\log N(u, \mathcal{T}, \rho)}$ für eine geeignete Konstante C'. □

9.2 Anwendungen

Sei $Z_n = \sum_{i=1}^n \phi_{ni}$ mit unabhängigen stochastischen Prozessen ϕ_{ni} auf einer abzählbaren Menge $\mathcal{T}$. Anstelle von $\hat\rho_n(s,t) = \sum_{i=1}^n |\phi_{ni}(s) - \phi_{ni}(t)|$ betrachten wir die zufällige Pseudometrik

$$\hat\rho_{n,2}(s,t) = \hat\rho_{n,2}(s,t\,|\,\boldsymbol{\phi}_n) \;:=\; \sqrt{\sum_{i=1}^n (\phi_{ni}(s) - \phi_{ni}(t))^2}$$

auf $\mathcal{T}$, wobei $\boldsymbol{\phi}_n := (\phi_{ni})_{i=1}^n$. Sei nämlich Z_n^o der symmetrisierte Prozess $\sum_{i=1}^n \xi_i \phi_{ni}$. Gegeben $\boldsymbol{\phi}_n$, hat der Prozess Z_n^o stetige Pfade und subgaußsche Zuwächse bezüglich $\hat{\rho}_{n,2}$. Diese Tatsache folgt aus Hoeffdings Ungleichung (Korollar 4.3) und wird nun in verschiedenen Situationen angewandt.

9.2.1 Eine Maximalungleichung für empirische Prozesse

Sei $\mathcal{F}$ eine punktweise separable Familie von messbaren Funktionen $f : \mathcal{X} \to [0,1]$ mit $0 \in \mathcal{F}$ und $\#\mathcal{F} \geq 2$. Ferner sei

$$J(\mathcal{F}) \ := \ \int_0^1 \sqrt{\log N(u^2, \mathcal{F})} \, du \ < \ \infty$$

mit den uniformen Überdeckungszahlen $N(\cdot, \mathcal{F})$ aus Abschnitt 6.2, wobei $F := 1$. Dann ist stets

$$\Big\| \sqrt{n} \big\| \hat{P}_n - P \big\|_{\mathcal{F}} \Big\|_{\psi_o} \ \leq \ CJ(\mathcal{F})$$

mit einer universellen Konstante C. Denn für $r > 0$ folgt aus den beiden Symmetrisierungen, dass

$$\Big\| \sqrt{n} \big\| \hat{P}_n - P \big\|_{\mathcal{F}} \Big\|_{\psi_o} \ \leq \ 2 \big\| \|Z_n^o\|_{\mathcal{F}} \big\|_{\psi_o}.$$

Hier ist $Z_n^o(f) := n^{-1/2} \sum_{i=1}^n \xi_i f(X_i)$ und

$$\hat{\rho}_{n,2}(f,g) \ = \ \sqrt{\hat{P}_n(|f-g|^2)} \ \leq \ \sqrt{\hat{P}_n(|f-g|)}.$$

Folglich ist $N(u, \mathcal{F}, \hat{\rho}_{n,2}) \leq N(u^2, \mathcal{F}, \rho_{\hat{P}_n}) \leq N(u^2, \mathcal{F})$, und Korollar 9.5 ergibt

$$\mathbb{E}\, \psi_o\Big(\frac{\|Z_n^o\|_{\mathcal{F}}}{CJ(\mathcal{F})} \Big) \ = \ \mathbb{E}\, \mathbb{E} \Big(\psi_o\Big(\frac{\|Z_n^o\|_{\mathcal{F}}}{CJ(\mathcal{F})} \Big) \, \Big| \, X_1, \ldots, X_n \Big) \ \leq \ 1$$

für die dortige Konstante C.

Es sei noch angemerkt, dass im Falle von $\#\mathcal{F} \geq 2$ gilt:

$$J(\mathcal{F}) \ \leq \ C' \sqrt{V[\mathrm{sgr}(\mathcal{F})]}$$

mit der VC-Dimension $V[\mathrm{sgr}(\mathcal{F})]$ von $\mathrm{sgr}(\mathcal{F})$ und einer universellen Konstanten C'. Diese Ungleichung folgt direkt aus Satz 6.3.

9.2.2 Prozesse mit Lipschitz-stetigen Pfaden

Angenommen, die Prozesse ϕ_{ni} haben Lipschitz-stetige Pfade, und $\mathbb{E}\, \phi_{ni}$ existiere in $\mathbb{R}^{\mathcal{T}}$. Für $x : \mathcal{T} \to \mathbb{R}$ definieren wir

$$\|x\|_{\mathrm{Lip}} \ := \ \sup_{s,t \in \mathcal{T}} \frac{|x(s) - x(t)|}{\rho(s,t)}$$

mit den Konventionen $0/0 := 0$ und $a/0 := \infty$ für $a > 0$. Dann gilt:

Lemma 9.6. *Es existiert eine universelle Konstante $C < \infty$, so dass*

$$\mathbb{E}\,\omega(Z_n - \mathbb{E}\,Z_n, \delta) \ \le\ C\sqrt{\mathbb{E}\sum_{i=1}^n \|\phi_{ni}\|^{2(1-\gamma)}\|\phi_{ni}\|_{\mathrm{Lip}}^{2\gamma}}\int_0^{\delta^\gamma}\sqrt{\log N(u^{1/\gamma}, \mathcal{T}, \rho)}\,du$$

für beliebige $\delta > 0$ und $\gamma \in (0, 1]$.

Beweis. Nach Symmetrisierungslemma 3.1 (b) und 3.3 (b) ist $\mathbb{E}\,\omega(Z_n - \mathbb{E}\,Z_n, \delta)$ nicht größer als $2\,\mathbb{E}\,\omega(Z_n^o, \delta)$. Wie schon erwähnt hat der symmetrisierte Prozess Z_n^o subgaußsche Zuwächse bezüglich $\hat\rho_{n,2}$, gegeben $\boldsymbol{\phi}_n$. Außerdem ist

$$\begin{aligned}(\phi_{ni}(s) - \phi_{ni}(t))^2 &= |\phi_{ni}(s) - \phi_{ni}(t)|^{2(1-\gamma)}|\phi_{ni}(s) - \phi_{ni}(t)|^{2\gamma}\\ &\le 4^{1-\gamma}\|\phi_{ni}\|^{2(1-\gamma)}\|\phi_{ni}\|_{\mathrm{Lip}}^{2\gamma}\,\rho(s,t)^{2\gamma},\end{aligned}$$

weshalb

$$\hat\rho_{n,2}(s,t) \ \le\ \check\rho_n(s,t) \ :=\ D_n\rho(s,t)^\gamma \quad \text{mit} \quad D_n \ :=\ 2^{1-\gamma}\sqrt{\sum_{i=1}^n \|\phi_{ni}\|^{2(1-\gamma)}\|\phi_{ni}\|_{\mathrm{Lip}}^{2\gamma}}.$$

Ferner ist $\omega(Z_n^o, \delta\,|\,\rho) = \omega(Z_n^o, D_n\delta^\gamma\,|\,\check\rho_n)$ und $N(u, \mathcal{T}, \check\rho_n) = N\big((u/D_n)^{1/\gamma}, \mathcal{T}, \rho\big)$. Daher folgt aus Korollar 9.5, dass

$$\begin{aligned}\mathbb{E}\big(\omega(Z_n^o, \delta)\,\big|\,\boldsymbol{\phi}_n\big) &\le C\int_0^{D_n\delta^\gamma}\sqrt{\log N(u, \mathcal{T}, \check\rho_n)}\,du\\ &= C\int_0^{D_n\delta^\gamma}\sqrt{\log N\big((u/D_n)^{1/\gamma}, \mathcal{T}, \rho\big)}\,du\\ &= CD_n\int_0^{\delta^\gamma}\sqrt{\log N(u^{1/\gamma}, \mathcal{T}, \rho)}\,du,\end{aligned}$$

und die Behauptung folgt durch Integration beider Seiten bezüglich der Verteilung von $\boldsymbol{\phi}_n$. □

Beispiel 9.7. Ein spezielles Anwendungsbeispiel sind *empirische charakteristische Funktionen*. Für $R > 0$ sei $\mathcal{T}_R = \{t \in \mathbb{R}^d : |t| \le R\}$, und sei $\rho(s,t) := |s - t|$. Für unabhängige, identisch verteilte Zufallsvariablen $X_1, X_2, X_3, \ldots \in \mathbb{R}^d$ sei

$$\begin{aligned}Z_n(t,1) &:= n^{-1/2}\sum_{i=1}^n \cos(t^\top X_i),\\ Z_n(t,2) &:= n^{-1/2}\sum_{i=1}^n \sin(t^\top X_i).\end{aligned}$$

Falls

$$\mathbb{E}\,|X_1|^\epsilon \ <\ \infty \quad \text{für ein } \epsilon > 0,$$

dann konvergiert $(Z_n - \mathbb{E}\,Z_n)_n$ im Raum $\ell_\infty(\mathcal{T}_R \times \{1,2\})$ in Verteilung gegen einen zentrierten Gaußprozess Z mit gleichmäßig stetigen Pfaden bezüglich der Metrik

$$\tilde\rho((s,j),(t,k)) \ :=\ |s - t| + |j - k|,$$

und

$$\mathbb{E}\big(Z(s,j)Z(t,k)\big) \;=\; \mathbb{E}\big(Z_n(s,j)Z_n(t,k)\big) \quad \text{für alle } s,t \in \mathbb{R}^d \text{ und } j,k \in \{1,2\}.$$

Denn aus dem multivariaten Zentralen Grenzwertsatz folgt, dass die endlichdimensionalen Randverteilungen von Z_n schwach gegen die gewünschten zentrierten Normalverteilungen konvergieren. Nach Satz 7.32 in Abschnitt 7.7 genügt es zu zeigen, dass die Folgen $(Z_n(\cdot,1))_n$ und $(Z_n(\cdot,2))_n$ asymptotisch stochastisch gleichstetig auf $\mathcal{T}_R$ bezüglich der euklidischen Metrik ρ sind. Dies folgt aber aus Lemma 9.6, angewandt auf $\phi_{ni}(t) := n^{-1/2}\cos(t^\top X_i)$ bzw. $\phi_{ni}(t) := n^{-1/2}\sin(t^\top X_i)$. Denn

$$\|\phi_{ni}\|_{\mathcal{T}_R} \;\le\; n^{-1/2}, \quad \|\phi_{ni}\|_{\mathrm{Lip}} \;\le\; n^{-1/2}|X_i| \quad \text{und} \quad N(u,\mathcal{T}_R,\rho) \;\le\; \Big(1+\frac{2R}{u}\Big)^d.$$

Letztere Ungleichung folgt aus Lemma 6.2. Folglich ist $\|\phi_{ni}\|^{2(1-\gamma)}\|\phi_{ni}\|_{\mathrm{Lip}}^{2\gamma} \le n^{-1}|X_i|^{2\gamma}$ und

$$\mathbb{E}\,\omega(Z_n(\cdot,j),\delta) \;\le\; C\sqrt{\mathbb{E}\,|X_1|^{2\gamma}}\int_0^{\delta^\gamma}\sqrt{d\,\log\Big(1+\frac{2R}{u^{1/\gamma}}\Big)}\,du.$$

9.2.3 Ein Funktionaler Zentraler Grenzwertsatz

Die Resultate in diesem Abschnitt lehnen sich stark an die Arbeit von K.S. Alexander (1987) an. Neben der zufälligen Pseudometrik $\hat\rho_{n,2}$ betrachten wir nun auch die deterministische Metrik

$$\rho_{n,2}(s,t) \;:=\; \sqrt{\mathbb{E}\,\hat\rho_{n,2}(s,t)^2}.$$

Mit den Prozessen

$$\tilde\phi_{ni}(s,t) \;:=\; \phi_{ni}(s)\phi_{ni}(t) \quad \text{und} \quad \tilde Z_n \;:=\; \sum_{i=1}^n \tilde\phi_{ni}$$

auf $\mathcal{T}\times\mathcal{T}$ kann man schreiben:

$$\hat\rho_{n,2}(s,t)^2 \;=\; \tilde Z_n(s,s) + \tilde Z_n(t,t) - 2\tilde Z_n(s,t).$$

Insbesondere ist $\|\hat\rho_{n,2}^2 - \rho_{n,2}^2\|_{\mathcal{T}\times\mathcal{T}} \le 4\|\tilde Z_n - \mathbb{E}\,\tilde Z_n\|_{\mathcal{T}\times\mathcal{T}}$.

Wir nehmen im folgenden an, dass die Prozesse ϕ_{ni} und Z_n “separabel” sind in folgendem Sinne: Es existiere eine abzählbare Teilmenge $\mathcal{T}_o$ von $\mathcal{T}$ derart, dass $\|\phi_{ni}\|$, $\|Z_n - \mathbb{E}\,Z_n\|$, $\|\tilde Z_n - \mathbb{E}\,\tilde Z_n\|_{\mathcal{T}\times\mathcal{T}}$ und $\omega(Z_n - \mathbb{E}\,Z_n, \delta \,|\, \rho_{n,2})$ unverändert bleiben, wenn man $\mathcal{T}$ durch $\mathcal{T}_o$ ersetzt.

Lemma 9.8. *Angenommen, die folgenden drei Bedingungen sind erfüllt:*

$$\mathbb{E}\sum_{i=1}^n \|\phi_{ni}\|^2 \quad = \quad O(1), \tag{9.4}$$

$$\mathbb{E}\sum_{i=1}^n 1\big\{\|\phi_{ni}\|^2 > u\big\}\|\phi_{ni}\|^2 \quad = \quad o(1) \quad \textit{für beliebige } u > 0, \tag{9.5}$$

$$\int_0^\delta \sqrt{\log N(u,\mathcal{T},\hat\rho_{n,2})}\,du \quad \to_p \quad 0 \quad \textit{für } n\to\infty, \delta\downarrow 0. \tag{9.6}$$

Daraus folgt, dass

$$(9.7)\qquad \mathbb{E}\,\|\tilde{Z}_n - \mathbb{E}\,\tilde{Z}_n\|_{\mathcal{T}\times\mathcal{T}} = o(1) \quad \textit{und} \quad N(u,\mathcal{T},\rho_{n,2}) = O(1) \textit{ für alle } u>0;$$

$$(9.8)\qquad \omega\big(Z_n - \mathbb{E}\,Z_n, \delta \,\big|\, \rho_{n,2}\big) \ \to_p\ 0 \quad \textit{für } n\to\infty, \delta\downarrow 0;$$

$$(9.9)\qquad \|Z_n - \mathbb{E}\,Z_n\| \ = \ O_p(1).$$

Anmerkung 9.9. Seien die ϕ_{ni} signierte Maße auf $\mathcal{X}$ und $\mathcal{F}$ eine punktweise separable Klasse von messbaren Funktionen auf $\mathcal{X}$ mit Einhüllender $F := \sup_{f\in\mathcal{F}}|f|$; siehe Abschnitt 6.2. Dann sind die Voraussetzungen (9.4-9.5) erfüllt, falls

$$\mathbb{E}\sum_{i=1}^n |\phi_{ni}|(F)^2 = O(1) \quad \text{und} \quad \mathbb{E}\sum_{i=1}^n 1\big\{|\phi_{ni}|(F)^2 > u\big\}\,|\phi_{ni}|(F)^2 = o(1) \ \text{ für alle } u>0.$$

Setzt man zusätzlich noch voraus, dass

$$J(\mathcal{F}) \ := \ \int_0^1 \sqrt{\log N(u^2,\mathcal{F})}\,du \ < \ \infty,$$

dann ist auch Bedingung (9.6) erfüllt. Denn $\hat{\rho}_{n,2}(f,g)^2 \ \le\ \rho_{\hat{M}_n}(f,g)$ mit dem Maß $\hat{M}_n := 2\sum_{i=1}^n |\phi_{ni}|(F)\,|\phi_{ni}|$, weshalb $N(u,\mathcal{F},\hat{\rho}_{n,2}) \le N\big(u^2/\hat{M}_n(F),\mathcal{F}\big)$ und

$$\begin{aligned}\int_0^\delta \sqrt{\log N(u,\mathcal{F},\hat{\rho}_{n,2})}\,du \ &\le\ \sqrt{\hat{M}_n(F)}\int_0^{\delta/\sqrt{\hat{M}_n(F)}} \sqrt{\log N(u^2,\mathcal{F})}\,du \\ &\le\ \sqrt{1+\hat{M}_n(F)}\int_0^\delta \sqrt{\log N(u^2,\mathcal{F})}\,du\end{aligned}$$

für $u,\delta>0$, wobei $\hat{M}_n(F) = 2\sum_{i=1}^n |\phi_{ni}|(F)^2 = O_p(1)$.

Beweis von Lemma 9.8. Der erste Teil von Behauptung (9.7) folgt aus Korollar 6.10, angewandt auf die Prozesse $\tilde{\phi}_{ni}$ anstelle von ϕ_{ni}. Zum einen ist $\|\tilde{\phi}\|_{\mathcal{T}\times\mathcal{T}} = \|\phi_{ni}\|^2$, was die ersten beiden Bedingungen von Korollar 6.10 impliziert. Zum anderen folgt aus (9.6), dass

$$(9.10)\qquad N(u,\mathcal{T},\hat{\rho}_{n,2}) \ = \ O_p(1) \quad \text{für alle } u>0,$$

denn für $0<\delta\le u$ ist stets

$$\sqrt{\log N(u,\mathcal{T},\hat{\rho}_{n,2})} \ \le\ \delta^{-1}\int_0^\delta \sqrt{\log N(v,\mathcal{T},\hat{\rho}_{n,2})}\,dv.$$

Ferner ist

$$\begin{aligned}\hat{\rho}_n\big((s,t),(s',t')\,\big|\,\tilde{\boldsymbol{\phi}}_n\big) \ &\le\ \sum_{i=1}^n \big(|\phi_{ni}(t)|\big|\phi_{ni}(s)-\phi_{ni}(s')\big| + |\phi_{ni}(s')|\big|\phi_{ni}(t)-\phi_{ni}(t')\big|\big) \\ &\le\ M_n\big(\hat{\rho}_{n,2}(s,s') + \hat{\rho}_{n,2}(t,t')\big)\end{aligned}$$

mit $M_n := \bigl(\sum_{i=1}^n \|\phi_{ni}\|^2\bigr)^{1/2} = O_p(1)$, woraus man ableiten kann, dass

$$N\bigl(u, \mathcal{T}\times\mathcal{T}, \hat\rho_n(\cdot,\cdot\,|\,\tilde{\boldsymbol{\phi}}_n)\bigr) \ \le\ N\Bigl(\frac{u}{2M_n}, \mathcal{T}, \hat\rho_{n,2}\Bigr) \ =\ O_p(1).$$

Also ist auch die letzte Bedingung von Korollar 6.10 erfüllt. Folglich ist

$$\mathbb{E}\,\|\hat\rho_{n,2}^2 - \rho_{n,2}^2\|_{\mathcal{T}\times\mathcal{T}} \ \le\ 4\,\mathbb{E}\,\|\tilde Z_n - \mathbb{E}\,\tilde Z_n\|_{\mathcal{T}\times\mathcal{T}} \ =\ o(1).$$

Zusammen mit (9.10) folgt hieraus der zweite Teil von (9.7).

Zur Behauptung (9.8): Da $\mathrm{Var}[Z_n(s) - Z_n(t)] \le \rho_{n,2}(s,t)^2$, kann man aus Symmetrisierungslemma 3.1 (a) und 3.3 (a) ableiten, dass für $\eta \ge \sqrt{2}\delta > 0$ gilt:

$$\begin{aligned}
&\mathbb{P}\bigl(\omega(Z_n - \mathbb{E}\,Z_n, \delta\,|\,\rho_{n,2}) \ge 3\eta\bigr)\\
&\quad\le\ 4\,\mathbb{P}\bigl(\omega(Z_n^o, \delta\,|\,\rho_{n,2}) \ge \eta\bigr)\\
&\quad\le\ 4\,\mathbb{P}\bigl(\|\hat\rho_{n,2}^2 - \rho_{n,2}^2\|_{\mathcal{T}\times\mathcal{T}} \ge 3\delta^2\bigr) + 4\,\mathbb{P}\bigl(\omega(Z_n^o, 2\delta\,|\,\hat\rho_{n,2}) \ge \eta\bigr)\\
&\quad=\ o(1) + 4\,\mathbb{P}\bigl(\omega(Z_n^o, 2\delta\,|\,\hat\rho_{n,2}) \ge \eta\bigr),
\end{aligned}$$

gemäß (9.7). Doch Z_n^o hat subgaußsche Zuwächse bezüglich $\hat\rho_{n,2}$ gegeben $\boldsymbol{\phi}_n$, und nach Korollar 9.5 ist

$$\mathbb{E}\bigl(\omega(Z_n^o, 2\delta\,|\,\hat\rho_{n,2})\,\big|\,\boldsymbol{\phi}_n\bigr) \ \le\ C\int_0^{2\delta} \sqrt{\log N(u, \mathcal{T}, \hat\rho_{n,2})}\,du \ \to_p\ 0 \quad \text{für } n\to\infty, \delta\downarrow 0.$$

Dies ergibt Behauptung (9.8).

Zu beweisen bleibt (9.9). Gemäß (9.7) und (9.8) gibt es zu beliebigem $\epsilon > 0$ eine Folge von Mengen $\mathcal{T}_n \subset \mathcal{T}$ mit

$$\#\mathcal{T}_n \ =\ O(1) \quad\text{und}\quad \limsup_{n\to\infty}\ \mathbb{P}\bigl(\|Z_n - \mathbb{E}\,Z_n\| \ge \|Z_n - \mathbb{E}\,Z_n\|_{\mathcal{T}_n} + 1\bigr) \ \le\ \epsilon.$$

Doch nach (9.4) ist

$$\begin{aligned}
\mathbb{E}\,\|Z_n - \mathbb{E}\,Z_n\|_{\mathcal{T}_n} \ &\le\ \sum_{t\in\mathcal{T}_n} \mathbb{E}\bigl|Z_n(t) - \mathbb{E}\,Z_n(t)\bigr|\\
&\le\ \#\mathcal{T}_n\sqrt{\sup_{t\in\mathcal{T}} \mathrm{Var}[Z_n(t)]}\\
&\le\ \#\mathcal{T}_n\sqrt{\mathbb{E}\sum_{i=1}^n \|\phi_{ni}\|^2}\\
&=\ O(1).
\end{aligned}$$

□

Satz 9.10. *Angenommen, die Voraussetzungen von Lemma 9.8 sind erfüllt. Desweiteren gebe es Funktionen H und K auf $\mathcal{T}\times\mathcal{T}$, so dass*

(9.11) $$\|\,\mathbb{E}\,\tilde Z_n - H\|_{\mathcal{T}\times\mathcal{T}} \ \to\ 0 \quad\textit{und}\quad \|\mathrm{Cov}(Z_n) - K\|_{\mathcal{T}\times\mathcal{T}} \ \to\ 0.$$

Dann ist $\mathcal{T}$ präkompakt bezüglich der Pseudometrik

$$\rho(s,t) \;:=\; \sqrt{H(s,s)+H(t,t)-2H(s,t)},$$

und $(Z_n - \mathbb{E}\, Z_n)_n$ konvergiert im Raum $\ell_\infty(\mathcal{T})$ in Verteilung gegen einen zentrierten Gaußprozess W auf $\mathcal{T}$ mit Kovarianzfunktion K und gleichmäßig stetigen Pfaden bezüglich ρ.

Anmerkung 9.11. Im Falle von zentrierten Prozessen ϕ_{ni}, also $\mathbb{E}\,\phi_{ni} \equiv 0$, ist $\mathbb{E}\,\tilde{Z}_n \equiv \mathrm{Cov}(Z_n)$. Tatsächlich bleiben die Bedingungen (9.4-9.5) gültig, wenn man die Prozesse ϕ_{ni} jeweils durch $\phi_{ni} - \mathbb{E}\,\phi_{ni}$ ersetzt (Aufgabe 9.5). Dagegen überträgt sich Bedingung (9.6) nicht automatisch auf die zentrierten Prozesse.

Beweis von Satz 9.10. Aus (9.4-9.5) und (9.11) folgt zusammen mit dem Lindebergschen Zentralen Grenzwertsatz, dass die endlichdimensionalen Randverteilungen von $Z_n - \mathbb{E}\,Z_n$ schwach gegen die entsprechenden Randverteilungen eines zentrierten Gaußprozesses mit Kovarianzfunktion K konvergieren.

Aus (9.11) folgt, dass $\|\rho_{n,2}^2 - \rho^2\|_{\mathcal{T}\times\mathcal{T}} = o(1)$. Zusammen mit (9.7) ergibt sich die Präkompaktheit von $(\mathcal{T},\rho)$, und (9.8) impliziert die stochastische Gleichstetigkeit der Folge $(Z_n - \mathbb{E}\,Z_n)_n$ bezüglich ρ. Die Konvergenz von $(Z_n - \mathbb{E}\,Z_n)_n$ in Verteilung ist also eine direkte Folge des Satzes 7.32. $\square$

9.2.4 Empirische Prozesse und P-Brücken

Wie in Kapitel 6.2 sei $\mathcal{F}$ eine punktweise separable Familie von messbaren Funktionen auf $\mathcal{X}$ mit Einhüllender $F := \sup_{f\in\mathcal{F}} |f|$. Falls

$$P(F^2) \;<\; \infty \quad \text{und} \quad J(\mathcal{F}) := \int_0^1 \sqrt{\log N(u^2,\mathcal{F})}\, du \;<\; \infty,$$

dann konvergiert der empirische Prozess $\big(\sqrt{n}(\hat{P}_n - P)(f)\big)_{f\in\mathcal{F}}$ im Raum $\ell_\infty(\mathcal{F})$ in Verteilung gegen einen zentrierten Gaußprozess B_P mit Kovarianzen

$$\mathbb{E}\big(B_P(f)B_P(g)\big) \;=\; P(fg) - P(f)P(g)$$

und gleichmäßig stetigen Pfaden bezüglich der Metrik

$$\rho(f,g) \;:=\; \sqrt{P((f-g)^2)}.$$

Dies folgt im wesentlichen aus Satz 9.10; siehe auch Anmerkung 9.9. Den Grenzprozess B_P nennt man auch eine P*-Brücke*.

Bootstrap. Bei festem P hängen sowohl $\rho_{n,2}$ als auch $\mathrm{Cov}[\sqrt{n}(\hat{P}_n - P)]$ nicht von n ab. Lemma 9.8 und Satz 9.10 beinhalten aber auch folgendes Resultat: Sei $\hat{P}_n$ die empirische Verteilung von unabhängigen Zufallsvariablen $X_{n1}, X_{n2}, \ldots, X_{nn}$ mit Verteilung P_n auf $\mathcal{X}$. Angenommen,

$$J(\mathcal{F}) \;<\; \infty;$$

$$P_n(F^2) = O(1) \quad \text{und} \quad P_n(1\{F^2 > nu\}F^2) = o(1) \ \text{ für beliebige } u > 0;$$

$$\sup_{f,g \in \mathcal{F}\cup\{1\}} \big|P_n(fg) - P(fg)\big| \ = \ o(1).$$

Dann konvergiert $\big(\sqrt{n}(\hat{P}_n - P_n)(f)\big)_{f\in\mathcal{F}}$ im Raum $\ell_\infty(\mathcal{F})$ in Verteilung gegen eine P-Brücke. Desweiteren gilt:

$$\mathbb{E}\,\hat{P}_n(F^2) = O(1) \quad \text{und} \quad \mathbb{E}\,\hat{P}_n(1\{F^2 > nu\}F^2) = o(1) \ \text{ für beliebige } u > 0;$$

$$\mathbb{E} \sup_{f,g \in \mathcal{F}\cup\{1\}} \big|\hat{P}_n(fg) - P(fg)\big| \ = \ o(1).$$

Dabei folgt letztere Tatsache aus (9.7). Man kann also die Verteilung von Statistiken, die stetig von $\big(\sqrt{n}(\hat{P}_n - P_n)(f)\big)_{f\in\mathcal{F}}$ abhängen, konsistent schätzen, indem man die unbekannte Verteilung P_n durch $\hat{P}_n$ ersetzt.

9.2.5 Partialsummenprozesse und P-Bewegungen

Sei $\mathcal{F}$ wie in Abschnitt 9.2.4, und sei

$$\hat{S}_n(f) \ := \ n^{-1/2} \sum_{i=1}^n Y_i f(x_{ni})$$

ein Partialsummenprozess wie in Abschnitt 1.2. Angenommen, die folgenden vier Bedingungen sind mit $P_n := n^{-1}\sum_{i=1}^n \delta_{x_{ni}}$ erfüllt:

$$J(\mathcal{F}) \ < \ \infty;$$

$$\mathbb{E}(Y_1) = 0 \quad \text{und} \quad \mathbb{E}(Y_1^2) = 1;$$

$$P_n(F^2) = O(1) \quad \text{und} \quad P_n(1\{F^2 > nu\}F^2) = o(1) \ \text{ für beliebige } u > 0;$$

$$\sup_{f,g \in \mathcal{F}} \big|P_n(fg) - P(fg)\big| \ = \ o(1)$$

für ein Wahrscheinlichkeitsmaß P auf $\mathcal{X}$. Dann konvergiert $\hat{S}_n$ im Raum $\ell_\infty(\mathcal{F})$ in Verteilung gegen einen zentrierten Gaußprozess W_P mit Kovarianzen

$$\mathbb{E}\big(W_P(f)W_P(g)\big) \ = \ P(fg)$$

und gleichmäßig stetigen Pfaden bezüglich der Metrik

$$\rho(f,g) \ := \ \sqrt{P((f-g)^2)} \ = \ \mathrm{Var}[W_P(f) - W_P(g)].$$

Diesen Grenzprozess W_P nennt man auch eine P-*Bewegung*. Ist $1 \in \mathcal{F}$, was man ohne Einschränkung annehmen kann, dann definiert $B_P(f) := W_P(f) - P(f)W_P(1)$ eine P-Brücke, und B_P, $W_P(1)$ sind stochastisch unabhängig.

9.2.6 Vorzeichentests für lineare Regression

Seien $Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ unabhängige Zufallsvariablen mit stetigen Verteilungsfunktionen

$$F_{ni}(r) \ := \ \mathbb{P}(Y_{ni} \le r).$$

Eine Variante des linearen Modells ist die Annahme, dass für gegebene feste Vektoren x_{n1}, x_{n2}, $\ldots$, x_{nn} im $\mathbb{R}^d$ und einen unbekannten Parameter $\theta_n \in \mathbb{R}^d$ gilt:

$$F_{ni}(\theta_n^\top x_{ni}) \ = \ 1/2 \quad \text{für } 1 \le i \le n \tag{9.12}$$

Mit

$$\vec{Z}_n(\gamma) \ := \ \sum_{i=1}^n \bigl(1\{Y_{ni} \le \gamma^\top x_{ni}\} - 1/2\bigr)\, x_{ni} \ \in \ \mathbb{R}^d$$

ist allgemein

$$\mathbb{E}\, \vec{Z}_n(\gamma) \ = \ \sum_{i=1}^n \bigl(F_{ni}(\gamma^\top x_{ni}) - 1/2\bigr)\, x_{ni}.$$

Unter der Modellannahme (9.12) ist $\mathbb{E}\, \vec{Z}_n(\gamma) = 0$ genau dann, wenn $F_{ni}(\gamma^\top x_{ni}) = 1/2$ für alle $i \le n$, und

$$2\vec{Z}_n(\theta_n) \ =_{\mathcal{L}} \ \sum_{i=1}^n \xi_i x_{ni}$$

mit einer Rademacher-Folge $(\xi_1, \xi_2, \ldots, \xi_n)$. Ist also $B_n(\alpha) \subset \mathbb{R}^d$ eine Menge mit

$$\mathbb{P}\Bigl(\sum_{i=1}^n \xi_i x_{ni} \in B_n(\alpha)\Bigr) \ \ge \ 1 - \alpha,$$

dann definiert $1\bigl\{2\vec{Z}_n(\gamma) \in B_n(\alpha)\bigr\}$ einen Test der Hypothese "$\theta_n = \gamma$" zum Niveau $1 - \alpha$, und

$$C_n(\alpha) \ := \ \bigl\{\gamma \in \mathbb{R}^d : 2\vec{Z}_n(\gamma) \in B_n(\alpha)\bigr\}$$

ist ein Konfidenzbereich für θ_n mit Konfidenzniveau $1 - \alpha$.

Nun betrachten wir das asymptotische Verhalten des vektorwertigen Prozesses $\vec{Z}_n : \mathbb{R}^d \to \mathbb{R}^d$. Wir setzen voraus, dass

$$\sum_{i=1}^n x_{ni} x_{ni}^\top \ = \ I; \tag{9.13}$$

$$\max_{i \le n} |x_{ni}| \ = \ o(1). \tag{9.14}$$

Dabei sei $|v|$ die Euklidische Norm $\sqrt{v^\top v}$. Durch lineare Transformation der Vektoren x_{ni} und Dimensionsreduktion, falls nötig, kann man stets erreichen, dass Bedingung (9.13) erfüllt ist. Insbesondere ist dann

$$\sum_{i=1}^n |x_{ni}|^2 \ = \ d.$$

Unter Bedingung (9.13) ist $|x_{ni}|^2$ die "Hebelkraft" (leverage) des Design-Punktes x_{ni}. Bedingung (9.14) verbietet also die Existenz von "Hebelarmpunkten".

Lemma 9.12. *Unter der Voraussetzung (9.13) ist*

$$\Big\| \|\vec{Z}_n - \mathbb{E}\,\vec{Z}_n\|_{\mathbb{R}^d} \Big\|_{\psi_o} \;\le\; Cd$$

für eine universelle Konstante C. Setzt man zusätzlich (9.14) voraus, dann gilt:

$$\omega\big(\vec{Z}_n - \mathbb{E}\,\vec{Z}_n, \delta \,\big|\, \rho_n\big) \;\to_p\; 0 \quad \text{für } n \to \infty, \delta \downarrow 0,$$

wobei

$$\rho_n(\gamma, \tilde{\gamma}) \;:=\; \sum_{i=1}^n |x_{ni}|^2 \,\big|F_{ni}(\gamma^\top x_{ni}) - F_{ni}(\tilde{\gamma}^\top x_{ni})\big|.$$

Unter den Annahmen (9.12), (9.13) und (9.14) folgt aus dem Lindebergschen Zentralen Grenzwertsatz, dass

$$\mathcal{L}\big(2\vec{Z}_n(\theta_n)\big) \;\to_w\; \mathcal{N}_d(0, I).$$

Daher ist

$$C_{n,r} \;:=\; \big\{\gamma \in \mathbb{R}^d : |\vec{Z}_n(\gamma)| \le r\big\}$$

ein spezieller Konfidenzbereich für θ_n mit asymptotischem Konfidenzniveau

$$\lim_{n\to\infty} \mathbb{P}(\theta_n \in C_{n,r}) \;=\; \chi^2_d([0, 4r^2]).$$

Die Berechnung von $C_{n,r}$ ist alles andere als trivial, obwohl wir gleich sehen werden, dass es sich unter gewissen Annahmen approximativ um eine Kugel mit Mittelpunkt $\hat{\theta}_n$ handelt. Letzterer ist ein Schätzer

$$\hat{\theta}_n \;\in\; \mathop{\arg\min}_{\gamma \in \mathbb{R}^d} H_n(\gamma) \quad \text{mit} \quad H_n(\gamma) \;:=\; \sum_{i=1}^n |Y_{ni} - \gamma^\top x_{ni}|.$$

Die Funktion H_n ist konvex, und man kann leicht zeigen, dass $H_n(\gamma) \to \infty$ für $|\gamma| \to \infty$, da die Vektoren x_{ni} unter Bedingung (9.13) den Raum $\mathbb{R}^d$ aufspannen. Also ist $\arg\min(H_n)$ eine nichtleere, kompakte und konvexe Teilmenge des $\mathbb{R}^d$. Ein Punkt $\hat{\theta}_n$ gehört zu $\arg\min(H_n)$ genau dann, wenn

$$DH_n(\hat{\theta}_n, \delta) := \lim_{t \downarrow 0} \frac{H_n(\hat{\theta}_n + t\delta) - H_n(\hat{\theta}_n)}{t} \;\ge\; 0 \quad \text{für alle } \delta \in \mathbb{R}^d.$$

Doch man kann leicht nachweisen, dass

$$\begin{aligned} DH_n(\hat{\theta}_n, \delta) \;&=\; \sum_{i=1}^n \operatorname{sign}(\hat{\theta}_n^\top x_{ni} - Y_{ni})\delta^\top x_{ni} + \sum_{i=1}^n 1\{Y_{ni} = \hat{\theta}_n^\top x_{ni}\}|\delta^\top x_{ni}| \\ &=\; 2\delta^\top \vec{Z}_n(\hat{\theta}_n) + \sum_{i=1}^n 1\{Y_{ni} = \hat{\theta}_n^\top x_{ni}\}(\delta^\top x_{ni})^-. \end{aligned}$$

Aus $DH_n(\hat{\theta}_n, \delta) \wedge DH_n(\hat{\theta}_n, -\delta) \ge 0$ folgt also, dass

$$(9.15) \qquad |\vec{Z}_n(\hat{\theta}_n)| \;\le\; \sum_{i=1}^n 1\{Y_{ni} = \hat{\theta}_n^\top x_{ni}\}|x_{ni}| \;\le\; d \max_{i \le n} |x_{ni}| \qquad \text{fast sicher;}$$

siehe Aufgabe 9.6. Folglich ist $\hat{\theta}_n \in C_{n,r}$, sofern $d \max_{i\le n} |x_{ni}| \le r$.

Lemma 9.13. *Angenommen, es gelten (9.12), (9.13) und (9.14). Ferner sei*

$$F_{ni}(\theta_n^\top x_{ni} + \cdot) \ = \ F(\cdot)$$

mit einer Verteilungsfunktion F, so dass

$$\lim_{y \to 0} \frac{F(y) - 1/2}{y} \ = \ c > 0.$$

Dann gilt:

(9.16) $$\inf_{v \in \mathbb{R}^d : |v| \ge b} \bigl| \vec{Z}_n(\theta_n + v) \bigr| \ \to_p \ \infty \quad \text{für } n \to \infty, b \to \infty;$$

(9.17) $$\sup_{v \in \mathbb{R}^d : |v| \le b} \bigl| \vec{Z}_n(\theta_n + v) - \vec{Z}_n(\theta_n) - cv \bigr| \ = \ o_p(1) \quad \text{für beliebige } b > 0.$$

Insbesondere ist

(9.18) $$\hat{\theta}_n \ = \ \theta_n - c^{-1} \vec{Z}_n(\theta_n) + o_p(1),$$

und für $C^o_{n,r} := \bigl\{ v \in \mathbb{R}^d : |v - \hat{\theta}_n| \le c^{-1} r \bigr\}$ gilt:

$$\mathbb{P}\bigl(C^o_{n,r-\epsilon} \subset C_{n,r} \subset C^o_{n,r+\epsilon} \bigr) \ \to \ 1 \quad \text{für beliebige } \epsilon > 0.$$

Beweis von Lemma 9.12. Als Hilfsmittel betrachten wir reellwertige Prozesse

$$\phi_{ni}(\gamma, \lambda) \ := \ 1\{Y_{ni} \le \gamma^\top x_{ni}\} \lambda^\top x_{ni}$$

und $Z_n = \sum_{i=1}^n \phi_{ni}$ auf $\mathcal{T} := \mathbb{R}^d \times \mathbb{S}^{d-1}$, wobei $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d : |v| = 1\}$. Dann ist $\lambda^\top (\vec{Z}_n - \mathbb{E}\, \vec{Z}_n)(\gamma) = (Z_n - \mathbb{E}\, Z_n)(\gamma, \lambda)$, also

$$\|\vec{Z}_n - \mathbb{E}\, \vec{Z}_n\|_{\mathbb{R}^d} \ = \ \|Z_n - \mathbb{E}\, Z_n\|_{\mathcal{T}}.$$

Nach den zwei Symmetrisierungen ist $\bigl\| \|Z_n - \mathbb{E}\, Z_n\| \bigr\|_{\psi_o}$ nicht größer als $2 \bigl\| \|Z_n^o\| \bigr\|_{\psi_o}$. Ferner ist

$$\begin{aligned}
&\hat{\rho}_{n,2}[(\gamma, \lambda), (\tilde{\gamma}, \tilde{\lambda})]^2 \\
&= \ \sum_{i=1}^n \Bigl(\bigl(1\{Y_{ni} \le \gamma^\top x_{ni}\} - 1\{Y_{ni} \le \tilde{\gamma}^\top x_{ni}\} \bigr) \lambda^\top x_{ni} \\
&\qquad\qquad + 1\{Y_{ni} \le \tilde{\gamma}^\top x_{ni}\} (\lambda^\top x_{ni} - \tilde{\lambda}^\top x_{ni}) \Bigr)^2 \\
&\le \ 2 \sum_{i=1}^n |x_{ni}|^2 \bigl| 1\{Y_{ni} \le \gamma^\top x_{ni}\} - 1\{Y_{ni} \le \tilde{\gamma}^\top x_{ni}\} \bigr| + 4 \sum_{i=1}^n |x_{ni}|^2 |\lambda - \tilde{\lambda}| \\
&= \ \hat{M}_n \bigl(\bigl| 1_{H(\gamma)} - 1_{H(\tilde{\gamma})} \bigr| \bigr) + 2 \hat{M}_n(1) \rho(\lambda, \tilde{\lambda}),
\end{aligned}$$

wobei

$$\begin{aligned}
\hat{M}_n \ &:= \ 2 \sum_{i=1}^n |x_{ni}|^2 \delta_{(x_{ni}^\top, Y_{ni})^\top}, \\
H(\gamma) \ &:= \ \bigl\{ v \in \mathbb{R}^{d+1} : (-\gamma^\top, 1) v \le 0 \bigr\}, \\
\rho(\lambda, \tilde{\lambda}) \ &:= \ |\lambda - \tilde{\lambda}|.
\end{aligned}$$

Man beachte, dass stets $\hat{M}_n(1) = 2d$. Die Menge $H(\gamma)$ gehört zu der Familie der abgeschlossenen Halbräume in $\mathbb{R}^d$. Letztere ist eine VC-Klasse mit VC-Index $d+1$; siehe Beispiel 5.14. Aus Lemma 6.2 sowie Satz 6.3 folgt, dass

$$\begin{aligned} N\big(u, \mathbb{S}^{d-1}, 2\hat{M}_n(1)\rho\big) &\leq \Big(\frac{9d}{u}\Big)^d \quad \text{und} \\ N\big(u, \{1_{H(\gamma)} : \gamma \in \mathbb{R}^d\}, \rho_{\hat{M}_n}\big) &\leq \Big(\frac{6d}{u}\Big)^{5(d+1)} \quad \text{für } 0 < u \leq d. \end{aligned}$$

Folglich ist

$$\begin{aligned} \log N(u, \mathcal{T}, \hat{\rho}_{n,2}) &\leq \log N\Big(\frac{3u^2}{5}, \mathbb{S}^{d-1}, 2\hat{M}_n(1)\rho\Big) + \log N\Big(\frac{2u^2}{5}, \{1_{H(\gamma)} : \gamma \in \mathbb{R}^d\}, \rho_{\hat{M}_n}\Big) \\ &\leq (6d+5)\log\Big(\frac{15d}{u^2}\Big) \quad \text{für } 0 < u \leq \sqrt{d}. \end{aligned}$$

Ferner ist $\phi_{ni}(\gamma_n, \cdot) \equiv 0$ für alle $i \leq n$ und ein geeignetes $\gamma_n = \gamma_n(Y_{n1}, \ldots, Y_{nn}) \in \mathbb{R}^d$, und $\hat{\rho}_{n,2}[(\gamma_n, \cdot), (\cdot, \cdot)]^2 \leq \sum_{i=1}^n \|\phi_{ni}\|^2 \leq d$. Aus Korollar 9.5, angewandt auf die bedingte Verteilung von Z_n^o gegeben $\boldsymbol{\phi}_n$, folgt also, dass

$$\Big\| \|Z_n^o\| \Big\|_{\psi_o} \leq C\sqrt{6d+5} \int_0^{\sqrt{d}} \sqrt{\log\Big(\frac{15d}{u^2}\Big)}\, du \leq 4Cd \int_0^1 \sqrt{\log\Big(\frac{15}{u^2}\Big)}\, du.$$

Mit obigen Überlegungen folgt aus (9.13) und (9.14), dass die Voraussetzungen von Lemma 9.8 erfüllt sind. Also gilt:

$$\omega\big(Z_n - \mathbb{E}\, Z_n, \delta \,\big|\, \rho_{n,2}\big) \to_p 0 \quad \text{für } n \to \infty, \delta \downarrow 0.$$

Doch

$$\rho_{n,2}\big((\gamma, \lambda), (\tilde{\gamma}, \tilde{\lambda})\big)^2 \leq 2\rho_n(\gamma, \tilde{\gamma}) + 4d|\lambda - \tilde{\lambda}|,$$

so dass

$$\begin{aligned} \omega\big(\vec{Z}_n - \mathbb{E}\, \vec{Z}_n, \delta \,\big|\, \rho_n\big) &= \sup_{\gamma, \tilde{\gamma} \in \mathbb{R}^d, \lambda \in \mathbb{S}^{d-1} : \rho_n(\gamma, \tilde{\gamma}) < \delta} \big|(Z_n - \mathbb{E}\, Z_n)(\gamma, \lambda) - (Z_n - \mathbb{E}\, Z_n)(\tilde{\gamma}, \lambda)\big| \\ &\leq \omega\big(Z_n - \mathbb{E}\, Z_n, \sqrt{2\delta} \,\big|\, \rho_{n,2}\big). \end{aligned}$$

$\square$

Beweis von Lemma 9.13. Sei

$$\vec{\Delta}_n(v) := \mathbb{E}\, \vec{Z}_n(\theta_n + v) = \sum_{i=1}^n \big(F(v^\top x_{ni}) - 1/2\big)\, x_{ni},$$

also $\vec{Z}_n(\theta_n + v) = (\vec{Z}_n - \mathbb{E}\, \vec{Z}_n)(\theta_n + v) + \vec{\Delta}_n(v)$. Dann ist

$$\inf_{|v| \geq b} |\vec{Z}_n(\theta_n + v)| \geq \inf_{|v| \geq b} |\vec{\Delta}_n(v)| - \|\vec{Z}_n - \mathbb{E}\, \vec{Z}_n\|_{\mathbb{R}^d} = \inf_{|v| \geq b} |\vec{\Delta}_n(v)| + O_p(1).$$

Für beliebige feste $b > 0$ folgt aus den Voraussetzungen an die Verteilungsfunktionen F_{ni} und F, dass

$$\begin{aligned}
\sup_{|v|\le b} \bigl|\vec{\Delta}_n(v) - cv\bigr| &= \sup_{|v|\le b} \Bigl|\sum_{i=1}^n \bigl(F(v^\top x_{ni}) - 1/2 - cv^\top x_{ni}\bigr)\, x_{ni}\Bigr| \\
&= o(1), \\
\inf_{|v|\ge b} \bigl|\vec{\Delta}_n(v)\bigr| &\ge \inf_{r\ge b, \lambda\in\mathbb{S}^{d-1}} \lambda^\top \vec{\Delta}_n(r\lambda) \\
&= \inf_{r\ge b, \lambda\in\mathbb{S}^{d-1}} \sum_{i=1}^n \bigl(F(r\lambda^\top x_{ni}) - 1/2\bigr)\, \lambda^\top x_{ni} \\
&= \inf_{\lambda\in\mathbb{S}^{d-1}} \sum_{i=1}^n \bigl(F(b\lambda^\top x_{ni}) - 1/2\bigr)\, \lambda^\top x_{ni} \\
&= cb + o(1).
\end{aligned}$$

Hieraus folgt (9.16). Doch

$$\delta_n(b) \ := \ \sup_{|v|\le b} \rho_n(\theta_n + v, \theta_n) \ = \ \sup_{|v|\le b} \sum_{i=1}^n |x_{ni}|^2 \bigl|F(v^\top x_{ni}) - 1/2\bigr| \ = \ o(1),$$

weshalb

$$\begin{aligned}
&\sup_{|v|\le b} \bigl|\vec{Z}_n(\theta_n + v) - \vec{Z}_n(\theta_n) - cv\bigr| \\
&\quad \le \ \sup_{|v|\le b} \bigl|\vec{\Delta}_n(v) - cv\bigr| + \omega\bigl(\vec{Z}_n - \mathbb{E}\, Z_n, \delta_n(b) \,\big|\, \rho_n\bigr) \ = \ o_p(1).
\end{aligned}$$

Aus diesen Entwicklungen für den Prozess $\vec{Z}_n(\theta + v)$ und der Ungleichung (9.15) folgt die Entwicklung (9.18) für $\hat{\theta}_n$. Ferner kann man leicht schließen, dass die Menge $C_{n,r}$ approximativ gleich der Kugel

$$C_{n,r}^1 \ := \ \theta_n + \Bigl\{v \in \mathbb{R}^d : |\vec{Z}_n(\theta_n) + cv| \le r\Bigr\}$$

ist. Diese Kugel hat Radius $c^{-1}r$ und Mittelpunkt $\theta_n - c^{-1}\vec{Z}_n(\theta_n) = \hat{\theta}_n + o_p(1)$. Also kann man $C_{n,r}^1$ wiederum durch $C_{n,r}^o$ approximieren. □

9.3 Aufgaben

Aufgabe 9.1. Sei $\psi : [0,\infty) \to [0,\infty)$ monoton wachsend und konvex, wobei $\psi(0) = 0$ aber $\psi \not\equiv 0$. Ferner sei $\psi^{-1}(t) := \max\{s \ge 0 : \psi(s) \le t\}$. Zeigen Sie, dass ψ^{-1} monoton wachsend, konkav und unbeschränkt ist.

Aufgabe 9.2. Sei m ein Maß auf einer σ-Algebra $\mathcal{A}$ auf Ω und $\psi : \mathbb{R} \to [0,\infty)$ eine gerade, konvexe Funktion, wobei $\psi(0) = 0$ aber $\psi \not\equiv 0$. Es bezeichne $\mathcal{L}^\psi(m)$ die Menge aller messbaren Funktionen f auf $(\Omega, \mathcal{A})$, so dass $\psi(f/r)$ für ein $r > 0$ integrierbar ist bezüglich m. Man zeige, dass $\mathcal{L}^\psi(m)$ ein Vektorraum und

$$\|f\|_\psi \ := \ \inf\Bigl\{r > 0 : \int \psi(f/r)\, dm \le 1\Bigr\}$$

eine Seminorm auf $\mathcal{L}^\psi(m)$ ist.

(Zusatz: Man zeige, dass $\mathcal{L}^\psi(m)$ bezüglich $\|\cdot\|_\psi$ vollständig ist.)

Aufgabe 9.3. Beweisen Sie die Ungleichungen (9.2-9.3).

Aufgabe 9.4. Sei $h : \mathbb{R} \to \mathbb{R}$ eine konvexe und gerade Funktion derart, dass $h(0) = 0$ aber $h \not\equiv 0$. Beweisen Sie folgende Ungleichungen für eine reellwertige Zufallsvariable Y:

(a) Ist $\mathbb{E}\exp(h(Y)) \le K$, dann ist $\mathbb{P}(|Y| \ge \eta) \le K\exp(-h(\eta))$ für alle $\eta \ge 0$.

(b) $\mathbb{P}(|Y| \ge \eta) \le K\exp(-h(\eta))$ für eine Konstante $K > 0$ und beliebige $\eta \ge 0$, dann ist $\mathbb{E}\exp(h(aY)) \le 1 + Ka/(1-a)$ für beliebige $0 < a < 1$.

Aufgabe 9.5. Zeigen Sie, dass

$$1\Big\{\sum_{i=1}^d y_i > u\Big\}\sum_{i=1}^d y_i \ \le\ d\sum_{i=1}^d 1\{y_i > u/d\}y_i \quad \text{für } y \in \mathbb{R}^d$$

und

$$1\{\mu \ge u\}\mu \ \le\ 2\,\mathbb{E}\,1\{Y \ge u/2\}Y \quad \text{für Zufallsvariablen } Y \text{ mit Erwartungswert } \mu.$$

Leiten Sie hieraus die Behauptung von Anmerkung 9.11 ab.

Aufgabe 9.6. Zeigen Sie, dass in Abschnitt 9.2.6 aus der Stetigkeit der Verteilungsfunktionen F_{ni} folgt, dass $\max_{\gamma\in\mathbb{R}^d}\sum_{i=1}^n 1\{Y_{ni} = \gamma^\top x_{ni}\} \le d$ fast sicher.

Hinweis: Für beliebige $S \subset \{1, 2, \ldots, n\}$ kann man zeigen, dass

$$\begin{aligned}&\#\big\{\gamma \in \mathbb{R}^d : \gamma^\top x_{ni} = Y_{ni} \text{ für alle } i \in S\big\}\\ &\qquad\begin{cases} = \ 1, & \text{falls die } x_{ni},\, i \in S, \text{ linear unabhängig sind,}\\ = \ 0 \text{ fast sicher,} & \text{falls die } x_{ni},\, i \in S, \text{ linear abhängig sind.}\end{cases}\end{aligned}$$

Aufgabe 9.7. Angenommen, in Abschnitt 9.2.6 ist $F_{ni}(\theta_n^\top x_{ni} + \cdot) = F(\cdot/\sigma_{ni})$ mit Skalenfaktoren $\sigma_{ni} > 0$ und einer Verteilungsfunktion F wie in Lemma 9.13. Formulieren und beweisen Sie eine Verallgemeinerung von Lemma 9.13.

Aufgabe 9.8. Hier ist noch eine Verfeinerung von Pisiers Lemma (Lemma 9.1) für subgaußsche Zufallsvariablen: Seien $Y_1, Y_2, \ldots, Y_m$ Zufallsvariablen derart, dass

$$\mathbb{E}\exp(tY_i) \ \le\ \exp(t^2/2) \quad \text{für } 1 \le i \le m \text{ und } t \in \mathbb{R}.$$

Dann ist

$$\mathbb{E}\Big(\max_{i\le m} Y_i^2\Big) \ \le\ 2\log(2m).$$

Anleitung: Für festes $\lambda > 0$ und $u \in \mathbb{R}$ sei

$$\psi_\lambda(u) \ :=\ \cosh\big(\lambda\sqrt{|u|}\big) - 1.$$

Zeigen Sie, dass ψ_λ eine gerade und konvexe Funktion ist. Zeigen Sie ferner, dass

$$\psi^{-1}(v) := \max\bigl\{u \geq 0 : \psi(u) \leq v\bigr\} \ \leq \ \bigl(\lambda^{-1} \log\bigl(2(1+v)\bigr)\bigr)^2$$

für $v \geq 0$. Zeigen Sie nun, dass

$$\mathbb{E}\Bigl(\max_{i \leq m} Y_i^2\Bigr) \ \leq \ \psi_\lambda^{-1}\Bigl(\sum_{i=1}^m \mathbb{E}\bigl(\cosh(\lambda Y_i) - 1\bigr)\Bigr),$$

und schätzen Sie die rechte Seite geeignet ab.

Kapitel 10

Kombinatorische Prozesse

Bisher wurden nur Prozesse $Z_n = \sum_{i=1}^n \phi_{ni}$ mit unabhängigen Summanden betrachtet. In diesem Abschnitt behandeln wir eine allgemeinere Klasse von Prozessen mit möglicherweise abhängigen Summanden, welche im Zusammenhang mit Permutationstests auftauchen. Sei

$$\Phi_n \;=\; (\phi_{nij})_{1\le i,j\le n}$$

eine Matrix bestehend aus n^2 unabhängigen stochastischen Prozessen ϕ_{nij} auf einer Menge $\mathcal{T}$. Ferner sei $\Pi = \Pi_n$ eine von Φ_n unabhängige und rein zufällig gewählte Permutation der Menge $\{1, 2, \ldots, n\}$. Nun betrachten wir den Prozess

$$S_n \;:=\; \sum_{i=1}^n \phi_{n\,i\,\Pi(i)}.$$

Aus technischen Gründen verlangen wir, dass entweder $\mathcal{T}$ abzählbar ist oder die ϕ_{nij} feste Funktionen sind.

Die bisher betrachteten Prozesse Z_n passen in diesen Rahmen. Sei nämlich $\mathcal{L}(\phi_{nij}) = \mathcal{L}(\phi_{ni})$ für alle i, j. Dann ist $\mathcal{L}(S_n) = \mathcal{L}(Z_n)$. Auch im allgemeinen Fall ist natürlich $\mathcal{L}(S_n \,|\, \Pi)$ vom gleichen Typ wie $\mathcal{L}(Z_n)$, aber wir interessieren uns für die *unbedingte* Verteilung von S_n.

Beispiel 10.1. Sei $a_n = (a_{ni})_{i=1}^n$ ein fester Vektor in $\mathbb{R}^n$ mit

$$\sum_{i=1}^n a_{ni} \;=\; 0 \quad \text{und} \quad \sum_{i=1}^n a_{ni}^2 \;=\; 1. \tag{10.1}$$

Für $t \in [0, 1]$ definieren wir

$$B_n(t) = B_n(t \,|\, a_n) \;:=\; \sum_{i\le nt} a_{n\,\Pi(i)}.$$

Dieser Prozess B_n ist in Verteilung gleich S_n, wenn

$$\phi_{nij}(t) \;=\; 1\{i \le nt\} a_{nj}.$$

Man kann leicht nachrechnen, dass

$$\mathbb{E}\, B_n(t) \;=\; 0 \quad \text{und} \quad \mathbb{E}\big(B_n(s)B_n(t)\big) \;=\; \frac{n}{n-1}\Big(\frac{\lfloor n\min(s,t)\rfloor}{n} - \frac{\lfloor ns\rfloor}{n}\frac{\lfloor nt\rfloor}{n}\Big)$$

für $s,t \in [0,1]$; siehe auch Lemma 10.3. Also hat B_n approximativ die gleiche Kovarianzfunktion wie die Brownsche Brücke. Man beachte, dass die Kovarianzfunktion von $B_n(\cdot \,|\, a_n)$ nicht von a_n abhängt, (10.1) vorausgesetzt. Realisationen dieses Prozesses werden in Abbildung 10.1 gezeigt. Hier gilt folgender Grenzwertsatz:

Satz 10.2 (vgl. Billingsley 1968). *Für $n \geq 2$ sei $a_n = (a_{ni})_{i=1}^n$ ein Vektor aus $\mathbb{R}^n$, welcher Bedingung (10.1) erfüllt, so dass*

$$\max_{i \leq n} |a_{ni}| \;=\; o(1). \tag{10.2}$$

Dann konvergiert $B_n(\cdot \,|\, a_n)$ in Verteilung gegen eine Brownsche Brücke. □

Im folgenden werden wir eine Verallgemeinerung dieses Satzes beweisen. Zunächst berechnen wir die ersten und zweiten Momente von S_n. Falls existent seien

$$\begin{array}{rclcrcl} M_n(t) & := & \big(\mu_{nij}(t)\big)_{i,j=1}^n & \text{mit} & \mu_{nij}(t) & := & \mathbb{E}\,\phi_{nij}(t), \\ \Gamma_n(t,u) & := & \big(\gamma_{nij}(t,u)\big)_{i,j=1}^n & \text{mit} & \gamma_{nij}(t,u) & := & \mathrm{Cov}[\phi_{nij}(t), \phi_{nij}(u)]. \end{array}$$

Für $A = (a_{ij})_{i,j}$ und $B = (b_{ij})_{i,j}$ aus $\mathbb{R}^{n\times n}$ sei

$$\langle A, B\rangle \;:=\; \mathrm{trace}(A^\top B) \;=\; \sum_{i,j} a_{ij}b_{ij} \quad \text{und} \quad |A| \;:=\; \sqrt{\langle A, A\rangle}.$$

Ferner sei $\mathbf{1} := (1)_{i,j} \in \mathbb{R}^{n\times n}$.

Lemma 10.3. *Angenommen, $M_n : \mathcal{T} \to \mathbb{R}^{n\times n}$ existiert. Dann ist*

$$\mathbb{E}\, S_n \;=\; \frac{1}{n}\langle M_n, \mathbf{1}\rangle.$$

Zusätzlich existiere $\Gamma_n : \mathcal{T} \times \mathcal{T} \to \mathbb{R}^{n\times n}$. Dann ist

$$\mathrm{Cov}[S_n(t), S_n(u)] \;=\; \frac{1}{n}\big\langle \Gamma_n(t,u), \mathbf{1}\big\rangle + \frac{1}{n-1}\big\langle GM_n(t), GM_n(u)\big\rangle$$

wobei

$$GA \;:=\; \Big(a_{ij} - \frac{a_{i+}}{n} - \frac{a_{+j}}{n} + \frac{a_{++}}{n}\Big)_{1\leq i,j\leq n}.$$

Den Beweis dieses Lemmas überlassen wir dem Leser als Übungsaufgabe. Man beachte, dass G die orthogonale Projektion von $\mathbb{R}^{n\times n}$ auf den Teilraum

$$\big\{(\alpha_i + \beta_j)_{i,j=1}^n : \alpha, \beta \in \mathbb{R}^n\big\}^\perp$$

ist.

Eine wesentliche Frage ist, ob und wie man den Prozess S_n symmetrisieren kann. Unser Ziel ist, den Erwartungswert von $\|S_n - \mathbb{E}\, S_n\|$ durch den Erwartungswert von $\|S_n^o\|$ nach oben abzuschätzen, wobei

$$S_n^o \ := \ \sum_{i=1}^n \xi_i \phi_{n\,i\,\Pi(i)}$$

mit einer von (Φ_n, Π) unabhängigen Rademacherfolge $(\xi_i)_{i\geq 1}$. Naive Anwendung der ersten Symmetrisierung führt zu dem Prozess

$$\sum_{i=1}^n \bigl(\phi_{n\,i\,\Pi(i)} - \phi_{n\,i\,\Pi'(i)}\bigr)$$

mit einer unabhängigen Kopie Π' von Π. Aber es ist nicht klar, wie man nun die Rademachervariablen ξ_i ins Spiel bringen soll. Die tatsächliche Methode ist etwas komplizierter und führt zu folgendem Resultat:

Lemma 10.4 (Symmetrisierung von S_n). *Angenommen, $M_n : \mathcal{T} \to \mathbb{R}^{n\times n}$ existiert. Für eine beliebige gerade, konvexe Funktion $\psi : \mathbb{R} \to [0,\infty)$ ist*

$$\mathbb{E}\, \psi\bigl(\|S_n - \mathbb{E}\, S_n\|\bigr) \ \leq \ \mathbb{E}\, \psi\bigl(8\|S_n^o\|\bigr).$$

Beweis von Lemma 10.4.

Erster Schritt: Sei $n \geq 2$ und $m := \lfloor n/2 \rfloor$. (Der Fall $n = 1$ ist uninteressant und wird direkt durch Lemma 3.1 (b) und 3.3 (b) abgedeckt.) Für $L \subset \{1, 2, \ldots, n\}$ sei $S_L := \sum_{i\in L} \phi_{n\,i\,\Pi(i)}$. Dann kann man schreiben:

$$S_n \ = \ \sum_{i=1}^n \phi_{n\,i\,\Pi(i)} \binom{n-1}{m-1}^{-1} \sum_{L:\#L=m} 1\{i \in L\} \ = \ \binom{n}{m}^{-1} \sum_{L:\#L=m} \frac{n}{m}\, S_L.$$

Folglich ist

$$\begin{aligned}
\mathbb{E}\, \psi\bigl(\|S_n - \mathbb{E}\, S_n\|\bigr) \ &= \ \mathbb{E}\, \psi\Bigl(\Bigl\| \binom{n}{m}^{-1} \sum_{L:\#L=m} \frac{n}{m}\, (S_L - \mathbb{E}\, S_L) \Bigr\|\Bigr) \\
&\leq \ \binom{n}{m}^{-1} \sum_{L:\#L=m} \mathbb{E}\, \psi\Bigl(\frac{n}{m}\, \|S_L - \mathbb{E}\, S_L\|\Bigr) \\
&\leq \ \max_{L:\#L=m} \mathbb{E}\, \psi\bigl(c_*\|S_L - \mathbb{E}\, S_L\|\bigr)
\end{aligned}$$

mit $c_* := n/m \leq 3$. Durch Umordnen der Zeilen von Φ_n kann man erreichen, dass

$$\mathbb{E}\, \psi\bigl(\|S_n - \mathbb{E}\, S_n\|\bigr) \ \leq \ \mathbb{E}\, \psi\bigl(c_*\|S_J - \mathbb{E}\, S_J\|\bigr), \tag{10.3}$$

wobei

$$J \ := \ \{1, \ldots, m\}.$$

Zweiter Schritt: Wir definieren

$$S_J^* \ := \ \sum_{i\in J} \phi_{n\,i\,\Pi(m+i)}.$$

Die Prozesse S_J und S_J^* sind identisch verteilt aber im allgemeinen abhängig. Gegeben $\Pi(J)$ sind sie im allgemeinen unterschiedlich verteilt aber unabhängig. Es ist

$$S_J - S_J^* \;=\; \sum_{i\in J}(\phi_{n\,i\,\Pi(i)} - \phi_{n\,i\,\Pi(m+i)}) \;=_{\mathcal{L}}\; \sum_{i\in J}\xi_i(\phi_{n\,i\,\Pi(i)} - \phi_{n\,i\,\Pi(m+i)}),$$

was man durch Bedingen auf die zweipunktigen Mengen $\{\Pi(i),\Pi(m+i)\}$, $1\le i\le m$, nachweisen kann. Mit $\mathbb{E}_o := \mathbb{E}(\cdot\,|\,\Phi_n,\Pi)$ gilt also für beliebige $c>0$:

$$\begin{aligned}
\mathbb{E}\,\psi\big(c\|S_J - S_J^*\|\big) \;&\le\; \mathbb{E}\,\psi\Big(c\Big\|\sum_{i\in J}\xi_i\phi_{n\,i\,\Pi(i)}\Big\| + c\Big\|\sum_{i\in J}\xi_i\phi_{n\,i\,\Pi(m+i)}\Big\|\Big)\\
&\le\; \frac{1}{2}\Big(\mathbb{E}\,\psi\Big(2c\Big\|\sum_{i\in J}\xi_i\phi_{n\,i\,\Pi(i)}\Big\|\Big) + \mathbb{E}\,\psi\Big(2c\Big\|\sum_{i\in J}\xi_i\phi_{n\,i\,\Pi(m+i)}\Big\|\Big)\Big)\\
&=\; \mathbb{E}\,\psi\Big(2c\Big\|\sum_{i\in J}\xi_i\phi_{n\,i\,\Pi(i)}\Big\|\Big)\\
&=\; \mathbb{E}\,\mathbb{E}_o\,\psi\Big(2c\Big\|\sum_{i\in J}\xi_i\phi_{n\,i\,\Pi(i)} + \mathbb{E}_o\sum_{i\notin J}\xi_i\phi_{n\,i\,\Pi(i)}\Big\|\Big)\\
&\le\; \mathbb{E}\,\mathbb{E}_o\,\psi\Big(2c\Big\|\sum_{i=1}^n\xi_i\phi_{n\,i\,\Pi(i)}\Big\|\Big) \quad \text{(nach Lemma 3.1 (b))}\\
&=\; \mathbb{E}\,\psi\big(2c\|S_n^o\|\big). \qquad (10.4)
\end{aligned}$$

Dritter und letzter Schritt: Nun vergleichen wir $\|S_J - \mathbb{E}\,S_J\|$ und $\|S_J - S_J^*\|$. Mit der Schreibweise $\mathbb{E}_o := \mathbb{E}(\cdot\,|\,\Pi(J))$ ist

$$\mathbb{E}_o\,S_J \;=\; \frac{1}{m}\sum_{i\in J}\sum_{j\in\Pi(J)}\mu_{nij}, \quad \mathbb{E}_o\,S_J^* \;=\; \frac{1}{n-m}\sum_{i\in J}\sum_{j\notin\Pi(J)}\mu_{nij}$$

und

$$\mathbb{E}\,S_J \;=\; \frac{1}{n}\sum_{i\in J}\sum_{j=1}^n\mu_{nij} \;=\; \frac{m}{n}\,\mathbb{E}_o\,S_J + \frac{n-m}{n}\,\mathbb{E}_o\,S_J^*.$$

Gegeben $\Pi(J)$ sind die Prozesse S_J und S_J^* stochastisch unabhängig. Folglich ist

$$\begin{aligned}
&\mathbb{E}_o\,\psi\big(c_*\|S_J - \mathbb{E}\,S_J\|\big)\\
&=\; \mathbb{E}_o\,\psi\Big(c_*\Big\|S_J - \frac{m}{n}\,\mathbb{E}_o\,S_J - \frac{m-n}{n}\,\mathbb{E}_o\,S_J^*\Big\|\Big)\\
&=\; \mathbb{E}_o\,\psi\Big(c_*\Big\|S_J - \mathbb{E}_o\,S_J^* - \frac{m}{n}\,\mathbb{E}_o(S_J - S_J^*)\Big\|\Big)\\
&\le\; \mathbb{E}_o\,\psi\big(c_*\|S_J - \mathbb{E}_o\,S_J^*\| + \|\,\mathbb{E}_o(S_J - S_J^*)\|\big)\\
&\le\; \frac{c_*}{1+c_*}\,\mathbb{E}_o\,\psi\big((1+c_*)\|S_J - \mathbb{E}_o\,S_J^*\|\big) + \frac{1}{1+c_*}\psi\big((1+c_*)\|\,\mathbb{E}_o(S_J - S_J^*)\|\big)\\
&\le\; \mathbb{E}_o\,\psi\big((1+c_*)\|S_J - S_J^*\|\big)
\end{aligned}$$

nach Symmetrisierungslemma 3.1 (b), angewandt auf das Paar $(Z,Z') = (S_J,S_J^*)$ bzw. $(Z,Z') = (0, S_J - S_J^*)$, gegeben $\Pi(J)$. Integration beider Seiten bezüglich der Verteilung von $\Pi(J)$ liefert die Ungleichung

$$\mathbb{E}\,\psi\big(c_*\|S_J - \mathbb{E}\,S_J\|\big) \;\le\; \mathbb{E}\,\psi\big((1+c_*)\|S_J - S_J^*\|\big). \qquad (10.5)$$

Kombiniert man nun (10.3), (10.4) mit $c = 1 + c_* \leq 4$ und (10.5), dann folgt die behauptete Ungleichung. $\square$

Nun kann man die Argumente des Beweises von Satz 6.8 auf die Prozesse S_n übertragen:

Satz 10.5. *Der Erwartungswert von $\|S_n - \mathbb{E}\, S_n\|$ konvergiert gegen Null, falls für eine Nullfolge $(\delta_n)_n$ folgende zwei Bedingungen erfüllt sind:*

$$\mathbb{E}\,\frac{1}{n}\sum_{i,j=1}^{n}\|\phi_{nij}\| = O(1) \quad \textit{und} \quad \mathbb{E}\,\frac{1}{n}\sum_{i,j=1}^{n} 1\bigl\{\|\phi_{nij}\| > \delta_n\bigr\}\|\phi_{nij}\| \;=\; o(1); \tag{10.6}$$

$$\log N(u, \mathcal{T}, \hat{\rho}_n) \;=\; o_p(\delta_n^{-1}) \quad \textit{für beliebige } u > 0, \tag{10.7}$$

wobei $\hat{\rho}_n(s,t) := \hat{\rho}_n\bigl(s,t \,\big|\, (\phi_{n\,i\,\Pi(i)})_{i=1}^n\bigr)$.

Auch der allgemeine funktionale Zentrale Grenzwertsatz, Lemma 9.8 und Satz 9.10, können auf $S_n - \mathbb{E}\, S_n$ übertragen werden. Allerdings kann man die endlichdimensionalen Randverteilungen von $S_n - \mathbb{E}\, S_n$ nicht mehr mit dem Lindebergschen Zentralen Grenzwertsatz behandeln. Der Einfachheit halber formulieren und beweisen wir nur einen einfachen Spezialfall. Allgemeinere Resultate findet man in Dümbgen (1994).

Satz 10.6. *Sei $\mathcal{F}$ eine Familie von Funktionen auf einem messbaren Raum $\mathcal{X}$ mit messbarer Einhüllender $F := \sup_{f\in\mathcal{F}}|f|$. Ferner sei $\phi_{nij}(f) = c_n f(x_{nij})$ mit festen Punkten $x_{nij} \in \mathcal{X}$ und einer Konstanten $c_n > 0$. Angenommen, folgende vier Bedingungen sind erfüllt:*

$$\frac{1}{n}\sum_{i,j=1}^{n} c_n^2 F(x_{nij})^2 \;=\; O(1), \tag{10.8}$$

$$\frac{1}{n}\sum_{i,j=1}^{n} 1\bigl\{c_n^2 F(x_{nij})^2 > u\bigr\} c_n^2 F(x_{nij})^2 \;=\; o(1) \;\; \textit{für beliebige } u > 0; \tag{10.9}$$

$$\int_0^1 \sqrt{\log N(u^2, \mathcal{F})}\, du \;<\; \infty; \tag{10.10}$$

$$\|\mathrm{Cov}(S_n) - K\|_{\mathcal{F}\times\mathcal{F}} \;\to\; 0 \quad \textit{für eine feste Funktion } K \textit{ auf } \mathcal{F}\times\mathcal{F}. \tag{10.11}$$

Dann ist $\mathcal{F}$ präkompakt bezüglich der Pseudometrik

$$\rho(f,g) \;:=\; \sqrt{K(f,f) + K(g,g) - 2K(f,g)},$$

und $S_n - \mathbb{E}\, S_n$ konvergiert in Verteilung im Raum $\ell_\infty(\mathcal{F})$ gegen einen zentrierten Gaußprozess W mit Kovarianzfunktion K und gleichmäßig stetigen Pfaden bezüglich ρ.

Satz 10.2 ist ein Spezialfall von Satz 10.6. Allgemein sei $\mathcal{D}$ eine VC-Klasse von messbaren Teilmengen eines messbaren Raumes $\mathcal{X}$, seien $z_{n1}, z_{n2}, \ldots, z_{nn}$ feste Punkte in $\mathcal{X}$, und sei $(a_n)_n$ eine Folge von Vektoren wie in Satz 10.2. Für die empirische Verteilung $P_n := n^{-1} \sum_{i=1}^n \delta_{z_{ni}}$ und ein Wahrscheinlichkeitsmaß P auf $\mathcal{X}$ gelte:

$$\sup_{D,E \in \mathcal{D}} |P_n(D \cap E) - P(D \cap E)| \;\to\; 0.$$

Definiert man $\phi_{nij}(D) := 1\{z_{ni} \in D\} a_{nj} = f_D(x_{nij})$ mit $x_{nij} := (z_{ni}, a_{nj})$ und $f_D(z, a) := 1\{z \in D\} a$, dann konvergiert $S_n - \mathbb{E}\, S_n$ in Verteilung im Raum $\ell_\infty(\mathcal{D})$ gegen eine P-Brücke auf $\mathcal{D}$.

Beweis von Satz 10.6. Dass die endlichdimensionalen Randverteilungen von $S_n - \mathbb{E}\, S_n$ gegen die entsprechenden Randverteilungen eines zentrierten Gaußprozesses mit Kovarianzfunktion K konvergieren, folgt aus (10.8-10.9), (10.11) und Satz 10.7 unten.

Die stochastische Gleichstetigkeit von $(S_n - \mathbb{E}\, S_n)_n$ bezüglich ρ kann man wie folgt nachweisen: Zunächst weisen wir darauf hin, dass

$$\sum_{i=1}^n (G\Phi_n)_{n\, i\, \Pi(i)} \;=\; S_n - \frac{1}{n} \big\langle \Phi_n, \mathbf{1} \big\rangle \;=\; S_n - \mathbb{E}\, S_n.$$

(Letztere Gleichung ist nur richtig im Falle von festen Funktionen ϕ_{nij}!) Daher ersetzen wir von nun an $\phi_{nij}(f) = c_n f(x_{nij})$ durch

$$\phi_{nij}(f) \;:=\; (G\Phi_n(f))_{ij} \;=\; \int h_{nij} f \, dm_{nij}$$

mit dem Maß

$$m_{nij} \;:=\; c_n \Big(\delta_{x_{nij}} + \frac{1}{n} \sum_{\ell=1}^n \delta_{x_{ni\ell}} + \frac{1}{n} \sum_{k=1}^n \delta_{x_{nkj}} + \frac{1}{n^2} \sum_{k,\ell=1}^n \delta_{x_{nk\ell}} \Big)$$

und einer Funktion $h_{nij} : \mathcal{X} \to [-1, 1]$. Es ist $\|\tilde{\phi}_{nij}\|^2 \le m_{nij}(F)^2$, und man kann zeigen, dass Bedingungen (10.8-10.9) erfüllt bleiben, wenn man $F(x_{nij})$ durch $m_{nij}(F)$ ersetzt; siehe Aufgabe 9.5. Der Vorteil dieser Substitution liegt darin, dass für die Pseudometriken $\hat{\rho}_{n,2}(f, g) := \hat{\rho}_{n,2}(f, g \,|\, (\phi_{n\, i\, \Pi(i)})_i)$ und $\rho_{n,2}(f, g) := \sqrt{\mathbb{E}\, \hat{\rho}_{n,2}(f, g)^2}$ gilt:

$$\rho_{n,2}(f, g) \;=\; \sqrt{\mathrm{Var}[S_n(f) - S_n(g)]},$$

und nach (10.11) ist

$$\|\rho_{n,2}^2 - \rho^2\|_{\mathcal{F} \times \mathcal{F}} \;=\; o(1). \tag{10.12}$$

Nach dem Symmetrisierungslemma 10.4 ist

$$\mathbb{E}\, \omega(S_n - \mathbb{E}\, S_n, \delta) \;\le\; 8\, \mathbb{E}\, \omega(S_n^o, \delta) \;\le\; 8\epsilon + 16 \sqrt{\mathbb{E}(\|S_n^o\|^2)\, \mathbb{P}\big(\omega(S_n^o, \delta) > \epsilon\big)}$$

für beliebige $\epsilon > 0$. Folglich genügt es zu zeigen, dass

$$\mathbb{E}(\|S_n^o\|^2) \;=\; O(1) \quad \text{und} \quad \omega(S_n^o, \delta) \;\to_p\; 0 \;\text{ für } n \to \infty, \delta \downarrow 0.$$

Gegeben Π hat der Prozess S_n^o subgaußsche Zuwächse bezüglich $\hat{\rho}_{n,2}$, und mit

$$\hat{M}_n \;:=\; 2\sum_{i=1}^n m_{n\,i\,\Pi(i)}(F)\, m_{n\,i\,\Pi(i)}$$

ist $\hat{\rho}_{n,2}(f,g)^2 \le \rho_{\hat{M}_n}(f,g)$. Also ist

$$\int_0^\delta \sqrt{\log N(u, \mathcal{F}, \hat{\rho}_{n,2})}\, du \;\le\; \sqrt{\hat{M}_n(F) + 1}\int_0^\delta \sqrt{\log N(u^2, \mathcal{F})}\, du$$

für $\delta > 0$; siehe auch Anmerkung 9.9. Wir nehmen ohne Einschränkung an, dass $0 \in \mathcal{F}$. Wegen $y^2 \le \exp(y^2) - 1 = \psi_o(y)$ folgt dann aus Korollar 9.5, dass

$$\begin{aligned}
\mathbb{E}\big(\|S_n^o\|^2 \,\big|\, \Pi\big) &\;\le\; C^2(\hat{M}_n(F) + 1)\Big(\int_0^1 \sqrt{\log N(u^2, \mathcal{F})}\, du\Big)^2, \\
\mathbb{E}\big(\omega(S_n^o, \delta \,|\, \hat{\rho}_{n,2})^2 \,\big|\, \Pi\big) &\;\le\; C^2(\hat{M}_n(F) + 1)\Big(\int_0^\delta \sqrt{\log N(u^2, \mathcal{F})}\, du\Big)^2.
\end{aligned}$$

Also ist $\mathbb{E}(\|S_n^o\|^2) = O(1 + \mathbb{E}\,\hat{M}_n(F)) = O(1)$, und $\mathbb{E}\,\omega(S_n^o, \delta \,|\, \hat{\rho}_{n,2})^2 \to 0$ für $n \to \infty$ und $\delta \downarrow 0$. Doch ähnlich wie im Beweis von Lemma 9.8 kann man aus Satz 10.5 ableiten, dass

$$\mathbb{E}\,\|\hat{\rho}_{n,2}^2 - \rho_{n,2}^2\|_{\mathcal{F}\times\mathcal{F}} \;\le\; 4\,\mathbb{E}\,\|\tilde{S}_n - \mathbb{E}\,\tilde{S}_n\|_{\mathcal{F}\times\mathcal{F}} \;\to\; 0,$$

wobei $\tilde{\phi}_{nij}(f,g) := \phi_{nij}(f)\phi_{nij}(g)$ und $\tilde{S}_n := \sum_{i=1}^n \tilde{\phi}_{n\,i\,\Pi(i)}$. Denn

$$\|\tilde{\phi}_{nij}\|_{\mathcal{F}\times\mathcal{F}} \;=\; \|\phi_{nij}\|^2 \;\le\; m_{nij}(F)^2,$$

und die zufällige Pseudometrik $\hat{\rho}_n\big(\cdot,\cdot \,\big|\, (\tilde{\phi}_{n\,i\,\Pi(i)})_i\big)$ erfüllt die Ungleichung

$$\hat{\rho}_n\big((f,g),(\tilde{f},\tilde{g}) \,\big|\, (\tilde{\phi}_{n\,i\,\Pi(i)})_i\big) \;\le\; \big(\rho_{\hat{M}_n}(f,\tilde{f}) + \rho_{\hat{M}_n}(g,\tilde{g})\big)/2,$$

also

$$N\big(u, \mathcal{F}\times\mathcal{F}, \hat{\rho}_n(\cdot,\cdot \,|\, (\tilde{\phi}_{n\,i\,\Pi(i)})_i)\big) \;\le\; N\Big(\frac{u}{\hat{M}_n(F)}, \mathcal{F}\Big)^2 \;=\; O_p(1).$$

□

Satz 10.7 (Hoeffding 1951). *Für $n \ge 2$ sei $A_n = (a_{nij})_{i,j} \in \mathbb{R}^{n\times n}$, so dass gilt:*

$$\frac{1}{n}|A_n|^2 \;=\; O(1) \quad \textit{und} \quad \frac{1}{n}\sum_{i,j=1}^n 1\{a_{nij}^2 > u\}a_{nij}^2 \;=\; o(1) \;\; \textit{für beliebige } u > 0;$$

$$\frac{1}{n}|GA_n|^2 \;\to\; \sigma^2.$$

Dann konvergiert $\mathcal{L}\big(\sum_{i=1}^n a_{n\,i\,\Pi(i)}\big)$ schwach gegen $\mathcal{N}(0,\sigma^2)$. □

$$a_{ni} = \frac{1\{i \leq n/2\} - 1\{i > n/2\}}{\sqrt{n}} \qquad a_{ni} = \frac{1\{i \leq n/10\} - 1\{i > n/10\}/9}{\sqrt{n/9}}$$

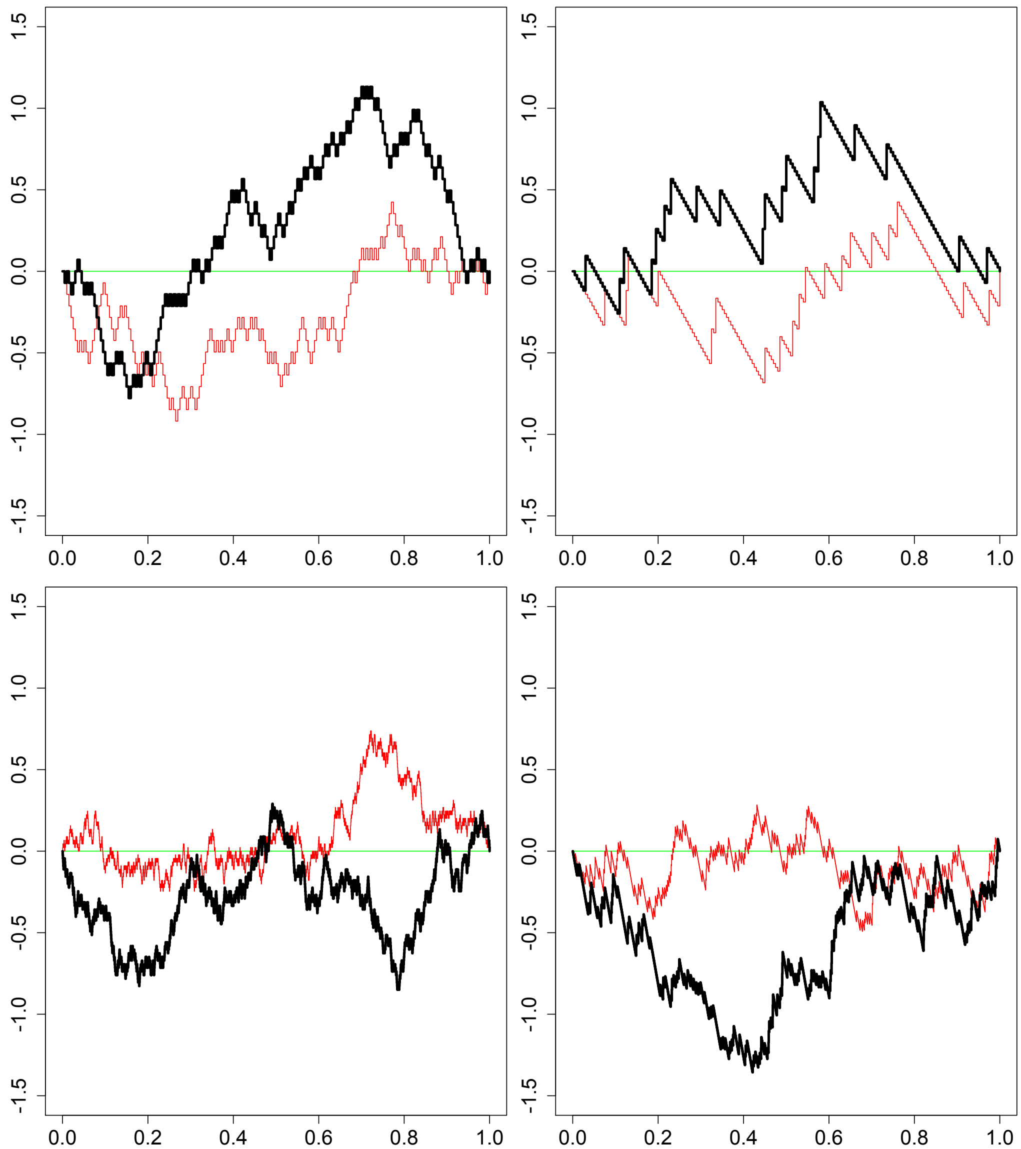

Abbildung 10.1: Realisationen von $B_n(\cdot \mid a_n)$ für $n = 200$ (oben) und $n = 2000$ (unten).

10.1 Aufgaben

Aufgabe 10.1. Beweisen Sie Lemma 10.3.

Aufgabe 10.2. Sei $(a_n)_n$ eine Folge wie in Satz 10.2, und für $t \in [0,1]$ sei

$$\phi_{nij}(t) \;:=\; \sqrt{\frac{n}{k_n}}\, 1\{i \le k_n t\} a_{nj}$$

mit einer Folge von Zahlen $k_n > 0$, so dass

$$k_n \;\to\; \infty \quad \text{aber} \quad \frac{k_n}{n} \;\to\; 0.$$

Zeigen Sie, dass $(S_n - \mathbb{E}\, S_n)_n$ in Verteilung konvergiert.

Kapitel 11

Weitere statistische Anwendungen

11.1 Maximum-Likelihood-Schätzer

In diesem Abschnitt beschreiben wir Anwendungen, welche auf S.A. van de Geer (1993, 2000) zurückgehen. Seien $X_1, X_2, \ldots, X_n$ stochastisch unabhängige Zufallsvariablen mit Verteilung P auf $\mathcal{X}$. Wir nehmen an, dass $P(dx) = f_*(x)\mu(dx)$ für ein σ-endliches Maß μ auf $\mathcal{X}$, und f_* sei eine unbekannte Wahrscheinlichkeitsdichte bezüglich μ, die zu einer gegebenen Famile $\mathcal{F}$ solcher Funktionen gehört.

Eine Dichtefunktion $\hat{f} \in \mathcal{F}$ heißt Maximum-Likelihood-Schätzer von f_*, falls

$$\begin{aligned}\hat{f} &\in \operatorname*{arg\,max}_{f\in\mathcal{F}} \sum_{i=1}^{n} \log f(X_i) \\ &= \operatorname*{arg\,max}_{f\in\mathcal{F}} \int \log f \, d\hat{P}_n.\end{aligned}$$

Um zu beweisen, dass $\hat{f}$ konsistent ist, benötigen wir einen geeigneten Abstand zwischen $\hat{f}$ und f_*. Hier sind zwei von vielen Möglichkeiten:

- **Test-Distanz (Totalvariation).**

$$\begin{aligned}T(f,g) &:= 2^{-1}\int |f-g|\,d\mu \\ &= 1 - \inf_{\text{Tests }\phi}\Big(\int \phi f\,d\mu + \int (1-\phi) g\,d\mu\Big).\end{aligned}$$

Anmerkung: Für einen Test $\phi : \mathcal{X} \to [0,1]$ von $\{f\}$ versus $\{g\}$ sind $\int \phi f\,d\mu$ und $\int (1-\phi) g\,d\mu$ die Wahrscheinlichkeiten für einen Fehler der ersten bzw. zweiten Art.

- **Hellinger-Abstand.**

$$\begin{aligned}H(f,g) &:= \sqrt{2^{-1}\int \left(f^{1/2} - g^{1/2}\right)^2 \, d\mu} \\ &= \sqrt{1 - \int (fg)^{1/2}\,d\mu}.\end{aligned}$$

Zwischen der Test-Distanz und dem Hellinger-Abstand besteht ein enger Zusammenhang, wie das folgende Lemma zeigt.

Lemma 11.1 (L. LeCam).

$$H(f,g)^2 \;\le\; T(f,g) \;\le\; H(f,g)\sqrt{2 - H(f,g)^2}.$$

Beweis von Lemma 11.1. Zum einen ist

$$\begin{aligned} H(f,g)^2 &\le 2^{-1} \int \left| f^{1/2} - g^{1/2} \right| \left(f^{1/2} + g^{1/2} \right) d\mu \\ &= 2^{-1} \int |f - g| \, d\mu \\ &= T(f,g). \end{aligned}$$

Andererseits folgt aus der Cauchy-Schwarz-Ungleichung, dass

$$\begin{aligned} T(f,g) &= 2^{-1} \int \left| f^{1/2} - g^{1/2} \right| \left(f^{1/2} + g^{1/2} \right) d\mu \\ &\le \sqrt{2^{-1} \int \left(f^{1/2} - g^{1/2} \right)^2 d\mu} \sqrt{2^{-1} \int \left(f^{1/2} + g^{1/2} \right)^2 d\mu} \\ &= H(f,g) \sqrt{1 + \int (fg)^{1/2} \, d\mu} \\ &= H(f,g) \sqrt{2 - H(f,g)^2}. \end{aligned}$$

□

Oftmals kann man mit Hilfe der folgenden Ungleichungen und eines uniformen Gesetzes der großen Zahlen nachweisen, dass der Maximum-Likelihood-Schätzer $\hat{f}$ konsistent bezüglich des Hellinger-Abstandes ist.

Lemma 11.2 (S. van de Geer 1993). *Für beliebige Familien $\mathcal{F}$ ist*

$$H(\hat{f}, f_*)^2 \;\le\; \int \sqrt{\frac{\hat{f}}{f_*}} \, d(\hat{P}_n - P) \;\le\; \sup_{f \in \mathcal{F}} \int \sqrt{\frac{f}{f_*}} \, d(\hat{P}_n - P).$$

Im Falle einer konvexen Familie $\mathcal{F}$ ist

$$H(\hat{f}, f_*)^2 \;\le\; 2 \int \sqrt{\frac{2\hat{f}}{f_* + \hat{f}}} \, d(\hat{P}_n - P) \;\le\; 2 \sup_{f \in \mathcal{F}} \int \sqrt{\frac{2f}{f_* + f}} \, d(\hat{P}_n - P).$$

Beweis von Lemma 11.2. Nach Definition von $\hat{f}$ ist

$$0 \;\le\; 2^{-1} \left(\int \log \hat{f} \, d\hat{P}_n - \int \log f_* \, d\hat{P}_n \right) \;=\; \int \log \sqrt{\frac{\hat{f}}{f_*}} \, d\hat{P}_n.$$

Wegen $\log(x) \leq x - 1$ ist die rechte Seite kleiner oder gleich

$$\begin{aligned}
\int \Big(\sqrt{\frac{\hat f}{f_*}} - 1\Big)\, d\hat P_n &= \int \Big(\sqrt{\frac{\hat f}{f_*}} - 1\Big)\, d(\hat P_n - P) + \int \Big(\sqrt{\frac{\hat f}{f_*}} - 1\Big)\, dP \\
&= \int \sqrt{\frac{\hat f}{f_*}}\, d(\hat P_n - P) + \int \Big(\sqrt{\frac{\hat f}{f_*}} - 1\Big) f_*\, d\mu \\
&= \int \sqrt{\frac{\hat f}{f_*}}\, d(\hat P_n - P) - \Big(1 - \int (\hat f f_*)^{1/2}\, d\mu\Big) \\
&= \int \sqrt{\frac{\hat f}{f_*}}\, d(\hat P_n - P) - H(\hat f, f_*)^2.
\end{aligned}$$

Im Falle einer konvexen Familie $\mathcal{F}$ folgt aus der Konkavität von $f \mapsto \int \log f\, d\hat P_n$, dass

$$\begin{aligned}
0 &\leq 2^{-1} \left(\int \log \hat f\, d\hat P_n - \int \log\Big(\frac{f_* + \hat f}{2}\Big)\, d\hat P_n \right) \\
&\leq \int \sqrt{\frac{2\hat f}{f_* + \hat f}}\, d(\hat P_n - P) - \int \left(1 - \sqrt{\frac{2\hat f}{f_* + \hat f}}\right) f_*\, d\mu.
\end{aligned}$$

Doch

$$\begin{aligned}
&\int \Big(1 - \sqrt{\frac{2g}{f+g}}\Big) f\, d\mu \\
&= \int \frac{(f+g)^{1/2} - (2g)^{1/2}}{(f+g)^{1/2}}\, f\, d\mu \\
&= \int \frac{f - g}{(f+g)^{1/2}\Big((f+g)^{1/2} + (2g)^{1/2}\Big)}\, f\, d\mu \\
&= \int \frac{f^{1/2}(f^{1/2} + g^{1/2})}{(f+g)^{1/2}\big((f+g)^{1/2} + (2g)^{1/2}\big)}\, f^{1/2}(f^{1/2} - g^{1/2})\, d\mu.
\end{aligned}$$

und man kann leicht nachweisen, dass

$$\frac{f^{1/2}(f^{1/2} + g^{1/2})}{(f+g)^{1/2}\Big((f+g)^{1/2} + (2g)^{1/2}\Big)} \begin{cases} \leq 1/2 & \text{falls } f \leq g, \\ \geq 1/2 & \text{falls } f \geq g. \end{cases}$$

Folglich ist

$$\int \Big(1 - \sqrt{\frac{2\hat f}{f_* + \hat f}}\Big) f_*\, d\mu \;\geq\; 2^{-1} \int f_*^{1/2}(f_*^{1/2} - \hat f^{1/2})\, d\mu \;=\; 2^{-1} H(\hat f, f_*)^2. \qquad \square$$

11.2 Zensierte Daten

In vielen Anwendungsbereichen wie Medizin oder Qualitätskontrolle muss man "zensierte" Daten analysieren. Man interessiert sich für die Verteilung von "Überlebenszeiten". Genauer gesagt, seien $X_{n1}, X_{n2}, \ldots, X_{nn}$ unabhängige Zufallsvariablen mit unbekannter Verteilung F_n auf $[0, \infty]$.

Mit F_n bezeichnen wir auch die Verteilungsfunktion

$$F_n(t) \;:=\; \mathbb{P}(X_{n1} \geq t);$$

aus dem Zusammenhang wird klar, welche Interpretation von F_n gemeint ist. (In der Überlebenszeitanalyse ist es üblich, "$\geq$" oder "$>$" anstelle von "$\leq$" zu verwenden.) Nun betrachten wir feste Zahlen $c_{n1}, c_{n2}, \ldots, c_{nn}$ aus $[0, \infty]$, sogenannte Zensierungszeitpunkte. Anstelle von X_{ni} beobachtet man nur

$$Y_{ni} \;:=\; X_{ni} \wedge c_{ni} \quad \text{und} \quad \Delta_{ni} \;:=\; 1\{X_{ni} \leq c_{ni}\}.$$

Beispiel 11.3. In medizinischen Anwendungen beschreibt X_{ni} oftmals die Überlebenszeit des i-ten Patienten nach einer bestimmten Operation, oder es ist die Zeit bis zum Auftreten von Abstoßungsreaktionen nach einer bestimmten Transplantation. Für Zensierung gibt es unterschiedliche Gründe. Beispielsweise könnte es sein, dass der i-te Patient nach einer Zeit c_{ni} tödlich verunglückt, oder dass er in eine andere Stadt umzieht und vor Ort nicht weiter untersucht wird. Wenn das Ereignis, welches durch X_{ni} beschrieben wird bis dahin noch nicht eintrat, dann weiß man eben nur, dass $X_{ni} > c_{ni}$.

Beispiel 11.4. In der Qualitätskontrolle beschreibe X_{ni} die Lebensdauer eines bestimmten Gerätes. Im einfachsten Falle ist c_{ni} die Zeit, vor welcher das i-te Gerät in Betrieb genommen wurde, und somit bekannt. Falls aber das Gerät schon früher aus dem Betrieb genommen wurde, obwohl es noch funktionierte, muß man mit einer anderen Zensierungszeit arbeiten. Sobald unvorhersehbare Zensierungsmechanismen auftreten können, muss man die c_{ni} als im allgemeinen unbekannt behandeln!

Anmerkung 11.5. In vielen Arbeiten werden die Zensierungszeiten c_{ni} als Zufallsvariablen modelliert, wobei $(X_{ni})_{i \leq n}$ und $(c_{ni})_{i \leq n}$ stochastisch unabhängig sind. Durch Bedingen auf $(c_{ni})_{i \leq n}$ kann man aber dieses Modell auf den hier betrachteten Fall fester Zensierungszeiten zurückführen.

Die Frage ist nun, inwieweit man F_n schätzen kann. Dazu betrachten wir die empirischen Verteilung(sfunktion)en

$$\begin{aligned}
\hat{H}_n \;&:=\; \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_{ni}} \quad \text{und} \quad \hat{H}_n(t) \;:=\; \frac{1}{n} \sum_{i=1}^{n} 1\{Y_{ni} \geq t\}, \\
\hat{H}_n^{\Delta} \;&:=\; \frac{1}{n} \sum_{i=1}^{n} \Delta_{ni} \delta_{Y_{ni}}, \\
C_n \;&:=\; \frac{1}{n} \sum_{i=1}^{n} \delta_{c_{ni}} \quad \text{und} \quad C_n(t) \;:=\; \frac{1}{n} \sum_{i=1}^{n} 1\{c_{ni} \geq t\}.
\end{aligned}$$

Die beiden ersteren sind tatsächlich Funktionen der vorhandenen Daten, wohingegen C_n im allgemeinen unbekannt ist. Mit $H_n := \mathbb{E}\,\hat{H}_n$ und $H_n^\Delta := \mathbb{E}\,\hat{H}_n^\Delta$ ist

$$(11.1) \qquad \begin{aligned} H_n(t) &= \frac{1}{n}\sum_{i=1}^n 1\{c_{ni} \ge t\}\,\mathbb{P}(X_{ni} \ge t) \;=\; C_n(t)F_n(t), \\ H_n^\Delta(B) &= \frac{1}{n}\sum_{i=1}^n \mathbb{P}(X_{ni} \le c_{ni}, X_{ni} \in B) \;=\; \int_B C_n(x)\,F_n(dx). \end{aligned}$$

Betrachtet man diese Formeln genauer, dann erkennt man, dass

$$\int_B \frac{1}{H_n(x)}\,H_n^\Delta(dx) \;=\; \int_B \frac{1}{F_n(x)}\,F_n(dx) \quad \text{falls } B \subset \bigl\{x \in [0,\infty] : C_n(x) > 0\bigr\}.$$

Tatsächlich handelt es sich bei dem Maß Λ_n mit $\Lambda_n(dx) := F_n(x)^{-1}F_n(dx)$ um das sogenannte Hazard-Maß der Verteilung F_n. Noch geläufiger ist der Ausdruck "kumulative Hazard-Funktion" für die entsprechende Verteilungsfunktion

$$\Lambda_n(t) \;:=\; \int \frac{1\{x \le t\}}{F_n(x)}\,F_n(dx).$$

Jede Verteilung auf $[0,\infty]$ wird durch ihre kumulative Hazardfunktion eindeutig festgelegt. Wir demonstrieren dies nur im Falle einer stetigen Verteilung F_n. Dann ist nämlich

$$\Lambda_n(t) \;=\; \int_{F_n(t)}^1 \frac{1}{u}\,du \;=\; -\log F_n(t).$$

Hat F_n eine Lebesgue-Dichte f_n, so nennt man die Lebesgue-Dichte $\lambda_n := f_n/F_n$ von Λ_n die Hazard-Rate oder Hazard-Funktion von F_n. Im Falle einer stetigen Dichte f_n ist

$$\lambda_n(t) \;=\; \lim_{\epsilon \downarrow 0} \epsilon^{-1}\,\mathbb{P}\bigl(X_{ni} \le t+\epsilon \,\big|\, X_{ni} \ge t\bigr).$$

Ein naheliegender Schätzer für $\Lambda_n(t)$ ist der *Nelson-Schätzer*

$$\hat{\Lambda}_n(t) \;:=\; \int \frac{1\{x \le t\}}{\hat{H}_n(x)}\,\hat{H}_n^\Delta(dx) \;=\; \frac{1}{n}\sum_{i=1}^n \frac{\Delta_{ni}1\{Y_{ni} \le t\}}{\hat{H}_n(Y_{ni})}.$$

Der folgende Satz gibt Aufschluss über seine asymptotischen Eigenschaften.

Satz 11.6 (Bresley und Crowley 1974). *Angenommen, $(F_n)_n$ und $(C_n)_n$ konvergieren schwach gegen Verteilung(sfunktion)en F beziehungsweise C. Ferner sei F stetig, und für eine feste Zahl $T > 0$ sei $F(T) := F[T,\infty] > 0$, $C(T+) := C]T,\infty[> 0$. Dann konvergiert $\bigl(\sqrt{n}(\hat{\Lambda}_n - \Lambda_n)(t)\bigr)_{t\in[0,T]}$ in Verteilung im Raum $\ell_\infty[0,T]$ gegen einen Prozess $W \circ \Gamma$. Dabei ist W eine Brownsche Bewegung auf $[0,\infty[$, und*

$$\Gamma(t) \;:=\; \int \frac{1\{x \le t\}}{C(x)F(x)^2}\,F(dx).$$

Beweis von Satz 11.6. Der Beweis ist ähnlich zu dem von Breslow und Crowley (1974). Ein anderer Zugang über Punktprozesse und Martingale wird von Andersen *et al.* (1993) dargestellt.

Zunächst schreiben wir

$$\begin{aligned}\hat{\Lambda}_n(t) &= \int \frac{1\{x \leq t\}}{H_n(x)} \hat{H}_n^{\Delta}(dx) - \int \frac{1\{x \leq t\}(\hat{H}_n - H_n)(x)}{\hat{H}_n(x) H_n(x)} \hat{H}_n^{\Delta}(dx) \\ &= \int \frac{1\{x \leq t\}}{H_n(x)} \hat{H}_n^{\Delta}(dx) - \int \frac{1\{x \leq t\}(\hat{H}_n - H_n)(x)}{H_n(x)^2} \hat{H}_n^{\Delta}(dx) + R_{n1}(t),\end{aligned}$$

wobei

$$R_{n1}(t) := \int \frac{1\{x \leq t\}(\hat{H}_n - H_n)(x)^2}{\hat{H}_n(x) H_n(x)^2} \hat{H}_n^{\Delta}(dx).$$

1. Schritt: Nun untersuchen wir den Prozess $Z_n := \sqrt{n}(\hat{H}_n - H_n)$. Ohne Einschränkung der Allgemeinheit nehmen wir an, dass $c_{n1} \geq c_{n2} \geq \cdots \geq c_{nn}$. Dann ist

$$\begin{aligned}Z_n(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \bigl(1\{c_{ni} \geq t\} 1\{X_{ni} \geq t\} - C_n(t) F_n(t)\bigr) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n 1\{c_{ni} \geq t\} \bigl(1\{X_{ni} \geq t\} - F_n(t)\bigr) \\ &= \frac{1}{\sqrt{n}} \sum_{i \leq n C_n(t)} \bigl(1\{X_{ni} \geq t\} - F_n(t)\bigr) \\ &= \tilde{Z}_n(t, C_n(t))\end{aligned}$$

mit dem *Kiefer-Müller-Prozess*

$$\tilde{Z}_n(t, v) := \frac{1}{\sqrt{n}} \sum_{i \leq nv} \bigl(1\{X_{ni} \geq t\} - F_n(t)\bigr), \quad (t, v) \in [0, \infty] \times [0, 1].$$

Mithilfe von Satz 9.10 kann man zeigen, dass der Prozess $\tilde{Z}_n$ in Verteilung im Raum $\ell_\infty([0,\infty] \times [0,1])$ gegen einen zentrierten Gaußprozess $\tilde{Z}$ mit stetigen Pfaden konvergiert. Folglich ist

$$\|Z_n\|_{[0,\infty]} = O_p(1).$$

Hieraus folgt, zusammen mit $\liminf_{n\to\infty} H_n(T) > 0$, dass

$$\|R_{n1}\|_{[0,T]} \leq \frac{\|Z_n\|_{[0,\infty]}^2 / n}{H_n(T)^2 \bigl(H_n(T) - \|Z_n\|_{[0,\infty]} / \sqrt{n}\bigr)} = O_p\Bigl(\frac{1}{n}\Bigr).$$

2. Schritt: Nun werden wir zeigen, dass

$$\int \frac{1\{x \leq t\}(\hat{H}_n - H_n)(x)}{H_n(x)^2} \hat{H}_n^{\Delta}(dx) = \int \frac{1\{x \leq t\}(\hat{H}_n - H_n)(x)}{H_n(x)^2} H_n^{\Delta}(dx) + R_{n2}(t)$$

mit $\|R_{n2}\|_{[0,T]} = o_p(n^{-1/2})$. Zunächst kann man mithilfe von Satz 6.8 leicht nachweisen, dass

$$\sup_{t \in [0,T]} \Bigl| \int \frac{1\{x \leq t\} g_n(x)}{H_n(x)^2} (\hat{H}_n^{\Delta} - H_n^{\Delta})(dx) \Bigr| = o_p(1)$$

für beliebige messbare Funktionen g_n auf $[0,\infty]$, falls $\|g_n\|_{[0,\infty]} = O(1)$. Dies wollen wir eigentlich auf Z_n anwenden, aber diese Funktion ist zufällig. Doch wegen der Separabilität von $(\mathcal{C}(\mathcal{T}), \|\cdot\|_{\mathcal{T}})$, wobei $\mathcal{T} := [0,\infty] \times [0,1]$, gibt es zu jedem $\epsilon > 0$ eine endliche Teilmenge $\mathcal{G}(\epsilon)$ von $\mathcal{C}(\mathcal{T})$, so dass gilt:

$$\mathbb{P}\Big(\min_{g \in \mathcal{G}(\epsilon)} \|\tilde{Z} - g\|_{\mathcal{T}} \ge \epsilon \Big) \ \le \ \epsilon.$$

Mit

$$\mathcal{G}_n(\epsilon) \ := \ \big\{ [t \mapsto g(t, C_n(t))] : g \in \mathcal{G}(\epsilon) \big\}$$

ist dann $\#\mathcal{G}_n(\epsilon) \le \#\mathcal{G}(\epsilon)$, und aus dem Portmanteau-Theorem folgt, dass

$$\limsup_{n\to\infty} \mathbb{P}\Big(\min_{g \in \mathcal{G}_n(\epsilon)} \|Z_n - g\|_{[0,\infty]} \ge \epsilon \Big) \ \le \ \epsilon.$$

Also können wir schreiben:

$$\begin{aligned}
&\sup_{t\in[0,T]} \Big| \int \frac{1\{x \le t\} Z_n(x)}{H_n(x)^2} \, (\hat{H}_n^\Delta - H_n^\Delta)(dx) \Big| \\
\le \quad & \frac{2 \min_{g\in\mathcal{G}_n(\epsilon)} \|Z_n - g\|_{[0,\infty]}}{H_n(T)^2} + \sup_{t\in[0,T], g\in\mathcal{G}_n(\epsilon)} \Big| \int \frac{1\{x \le t\} g(x)}{H_n(x)^2} \, (\hat{H}_n^\Delta - H_n^\Delta)(dx) \Big| \\
= \quad & \frac{2 \min_{g\in\mathcal{G}_n(\epsilon)} \|Z_n - g\|_{[0,\infty]}}{H_n(T)^2} + o_p(1) \\
\le \quad & \frac{2\epsilon}{H_n(T)^2} + o_p(1)
\end{aligned}$$

mit asymptotischer Wahrscheinlichkeit mindestens $1 - \epsilon$. Da $\epsilon > 0$ beliebig klein sein kann, folgt hieraus die Behauptung des zweiten Schrittes.

Letzter Schritt: Wir wissen nun, dass

$$\begin{aligned}
&(\hat{\Lambda}_n - \Lambda_n)(t) \\
= \quad & \int \frac{1\{x \le t\}}{H_n(x)} \, \hat{H}_n^\Delta(dx) - \int \frac{1\{x \le t\}(\hat{H}_n - H_n)(x)}{H_n(x)^2} \, H_n^\Delta(dx) - \Lambda_n(t) + R_{n3}(t) \\
= \quad & \int \frac{1\{x \le t\}}{H_n(x)} \, \hat{H}_n^\Delta(dx) - \int \frac{1\{x \le t\}\hat{H}_n(x)}{H_n(x)^2} \, H_n^\Delta(dx) + R_{n3}(t) \\
= \quad & \frac{1}{n} \sum_{i=1}^n \Big(\frac{\Delta_{ni} 1\{Y_{ni} \le t\}}{H_n(Y_{ni})} - \int \frac{1\{x \le t\} 1\{x \le Y_{ni}\}}{H_n(x)^2} \, H_n^\Delta(dx) \Big) + R_{n3}(t),
\end{aligned}$$

wobei $\|R_{n3}\|_{[0,T]} = o_p(n^{-1/2})$. Mit anderen Worten,

$$\sqrt{n}(\hat{\Lambda}_n - \Lambda_n)(t) \ = \ \sum_{i=1}^n \phi_{ni}[0,t] + \sqrt{n} R_{n3}(t)$$

mit den zufälligen signierten Maßen

$$\phi_{ni}(B) \ := \ \frac{1}{\sqrt{n}} \Big(\frac{\Delta_{ni}}{H_n(Y_{ni})} 1\{Y_{ni} \in B\} - \int_B \frac{1\{x \le Y_{ni}\}}{H_n(x)^2} \, H_n^\Delta(dx) \Big)$$

auf $[0, T]$. Eine längere, aber elementare Rechnung zeigt, dass

$$\begin{aligned}
\mathbb{E}\,\phi_{ni}(B) &= 0, \\
\mathbb{E}\,\phi_{ni}(B)\phi_{ni}(\tilde{B}) &= \frac{1}{n}\int_{B\cap\tilde{B}} \frac{1\{x \le c_{ni}\}}{H_n(x)^2}\,F_n(dx) - \frac{1}{n}\sum_{x\in B\cap\tilde{B}} \frac{1\{x \le c_{ni}\}F_n(\{x\})^2}{H_n(x)^2 F_n(x)}.
\end{aligned}$$

Zusammen mit (11.1) folgt hieraus, dass

$$\begin{aligned}
\operatorname{Cov}\Big(\sum_{i=1}^n \phi_{ni}[0,s], \sum_{i=1}^n \phi_{ni}[0,t]\Big) &= \sum_{i=1}^n \mathbb{E}\,\phi_{ni}[0,s]\phi_{ni}[0,t] \\
&= \int_{[0,s\wedge t]} \frac{1}{H_n(x)F_n(x)}\,F_n(dx) - \sum_{x\in]0,s\wedge t]} \frac{F_n(\{x\})^2}{H_n(x)F_n(x)^2} \\
&\to \Gamma(s,t) := \int_{[0,\min(s,t)]} \frac{1}{C(x)F(x)^2}\,F(dx)
\end{aligned}$$

gleichmäßig in $s, t \in [0, T]$. Ferner gilt für das Totalvariationsmaß $|\phi_{ni}|$ von ϕ_{ni}:

$$\max_{i\le n} |\phi_{ni}|([0,T])^2 \le \frac{4}{nH_n(T)^4} = O\Big(\frac{1}{n}\Big).$$

Aus Satz 9.10 folgt somit, dass $\big(\sum_{i=1}^n \phi_{ni}[0,t]\big)_{t\in[0,T]}$ in Verteilung im Raum $\ell_\infty[0,T]$ gegen einen zentrierten Gaußprozess $\tilde{W}$ mit Kovarianzfunktion Γ und gleichmäßig stetigen Pfaden bezüglich $\rho(s,t) := \sqrt{\Gamma(s,s) + \Gamma(t,t) - 2\Gamma(s,t)}$ konvergiert. $\square$

11.3 Aufgaben

Aufgabe 11.1. Man nehme an, dass alle Zensierungszeiten c_{ni} bekannt sind. Ein naheliegender Schätzer für $F_n(t)$ ist dann

$$\check{F}_n(t) := \frac{\hat{H}_n(t)}{C_n(t)} \quad \text{für } t \in T_n := \Big[0, \max_{i\le n} c_{ni}\Big].$$

Berechnen Sie die Kovarianzfunktion dieses Schätzers. Formulieren und beweisen Sie einen Zentralen Grenzwertsatz für $(\check{F}_n(t))_{t\in[0,T]}$ unter geeigneten Annahmen an die Folgen $(C_n)_n$ und $(F_n)_n$.

Aufgabe 11.2. Unter der Annahme, dass F_n stetig ist, beschreibe man einen Bootstrap-Schätzer für die Verteilung von $\big(\sqrt{n}(\hat{\Lambda}_n - \Lambda_n)(t)\big)_{t\in[0,T]}$.

Hinweis: Man muss sich üeberlegen, wie man F_n und C_n konsistent (auf $[0, T]$) schätzen kann.

Anhang A

Anhang

A.1 Literatur

K.S. ALEXANDER (1987). Central limit theorems for stochastic processes under random entropy conditions. *Probability Theory and Related Fields* ***75***, 351-378.

P.K. ANDERSEN, O. BORGAN, R.D. GILL and N. KEIDING (1993). *Statistical Models Based on Counting Processes.* Springer, New York

P. BILLINGSLEY (1968). *Convergence of Probability Measures.* Wiley

P. BILLINGSLEY and F. TOPSØE (1967). Uniformity in weak convergence. *Zeitschrift für Wahrscheinlichkeitsrechnung und verwandte Gebiete* ***7***, 1-16.

L. BREIMAN (1968). *Probability.* Addison-Wesley.

N. BRESLOW and J. CROWLEY (1974). A large sample study and of the life table and product limits estimates under random censorship. *Annals of Statistics* ***2***, 437-453.

L. DÜMBGEN (1994). Combinatorial stochastic processes. *Stochastic Process. Appl.* ***52***, 75-92.

L. DÜMBGEN (1998). Symmetrization and decoupling of combinatorial random elements. *Statistics and Probability Letters* ***39***, 355-361.

S.A. VAN DE GEER (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Annals of Statistics* ***21***, 14-44.

S.A. VAN DE GEER (2000). *Empirical Processes in M-Estimation.* Cambridge University Press.

R. GRÜBEL (1988). The length of the shorth. *Annals of Statistics* ***16***, 619-628.

W. HOEFFDING (1951). A combinatorial Central Limit Theorem. *Annals of Mathematical Statistics* ***22***, 558-566.

P. MASSART (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Annals of Probability **18***, 1269-1283.

A. PAJOR (1985). *Sous-Espaces l_1^n des Espaces de Banach.* Travaux en Cours [Works in Progress] **16**, Hermann, Paris.

D. POLLARD (1984). *Convergence of Stochastic Processes.* Springer.

D. POLLARD (1990). *Empirical Processes: Theory and Applications.* NSF-CBMS Regionial Conference Series in Probability and Statistics **2**, Institute of Mathematical Statistics.

R.R. RAO (1962). Relations between weak and uniform convergence of measures with applications. *Annals of Mathematical Statistics **33***, 659-680.

G.R. SHORACK and J.A. WELLNER (1986). *Empirical Processes with Applications to Statistics.* Wiley.

A.W. VAN DER VAART and J.A. WELLNER (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer.

V.N. VAPNIK and A.JA. ČERVONENKIS (1971). The uniform convergence of frequencies of the appearance of events to their probabilities. *Theory of Probability and its Applications **16***, 264-279.

A.2 Notation

Kapitel 1:

$\hat{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ mit unabhängigen Zufallsvariablen $X_i \in \mathcal{X}$ mit Verteilung P

$\hat{S}_n(f) := c_n \sum_{i=1}^n f(x_{ni}) Y_i$ mit festen Punkten $x_{ni} \in \mathcal{X}$ und unabhängigen, identisch verteilten Zufallsvariablen Y_i

$Z_n := \sum_{i=1}^n \phi_{ni}$ mit unabhängigen Prozessen ϕ_{ni} auf $\mathcal{T}$

$\|x\| = \|x\|_{\mathcal{T}} := \sup_{t \in \mathcal{T}} |x(t)|$

Kapitel 3:

$\hat{F}_n(t) := n^{-1} \sum_{i=1}^n 1\{X_i \le t\}$

$\hat{G}_n(v) := n^{-1} \sum_{i=1}^n 1\{U_i \le v\}$ für unabhängige, uniform auf $[0,1]$ verteilte Zufallsvariablen U_i

$Z_n^o := \sum_{i=1}^n \xi_i \phi_{ni}$ mit einer von $(\phi_{ni})_i$ unabhängigen Rademacherfolge $(\xi_i)_i$.

Kapitel 6:

$N(\epsilon, \mathcal{T}, \rho) := \inf\{\#\mathcal{T}_o : \mathcal{T}_o \subset \mathcal{T}$ und $\rho(t, \mathcal{T}_o) \le \epsilon$ für alle $t \in \mathcal{T}\}$

$D(\epsilon, \mathcal{T}, \rho) := \sup\{\#\mathcal{T}_o : \mathcal{T}_o \subset \mathcal{T}$ und $\rho(s, t) > \epsilon$ für verschiedene $s, t \in \mathcal{T}\}$

$\hat{\rho}_n(s, t) := \sum_{i=1}^n |\phi_{ni}(s) - \phi_{ni}(t)|$

$\rho_M(f, g) := \int |f - g| \, dM$

$N(\epsilon, \mathcal{F}) := \max\{N(\epsilon \, M(F), \mathcal{F}, \rho_M) : M$ ein Maß auf $\mathcal{X}\}$ mit $F := \sup_{f \in \mathcal{F}} |f|$

Kapitel 7:

$\omega(x, \delta) = \omega(x, \delta \,|\, \rho) := \sup_{s,t \in \mathcal{T} : \rho(s,t) < \delta} |x(s) - x(t)|$

$\mathcal{C}_{\mathrm{u}}(\mathcal{T}) = \mathcal{C}_{\mathrm{u}}(\mathcal{T} \,|\, \rho) := \{x \in \mathbb{R}^{\mathcal{T}} : \lim_{\delta \downarrow 0} \omega(x, \delta) = 0\}.$

$\hat{\rho}_{n,2}(s, t) := \left(\sum_{i=1}^n |\phi_{ni}(s) - \phi_{ni}(t)|^2\right)^{1/2}$

$\rho_{n,2}(s, t) := \left(\mathbb{E}\left(\hat{\rho}_{n,2}(s, t)^2\right)\right)^{1/2}$