Auswirkungen von beschränkter stochastischer Abhängigkeit auf statistische Verfahren

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund

Der Fakultät Statistik
der Technischen Universität Dortmund

vorgelegt von

Robert Kwiecien

Dortmund 2009

Promotionsausschuss

1. Gutachter:	Prof. Dr. Ursula Gather
2. Gutachter:	PD Dr. Rafael Weißbach
Vorsitzender der Prüfungskommission:	Prof. Dr. Roland Fried

Tag der mündlichen Prüfung

10. Juni 2009

Inhaltsverzeichnis

Kapitel 1

Einleitung

Statistische Verfahren dienen im Allgemeinen dazu, aus Daten wie z. B. wiederholten Messungen relevante Informationen zu extrahieren. In der Regel werden bei der Verwendung eines statistischen Verfahrens Annahmen getroffen, die sicherstellen, dass das statistische Verfahren die gewünschten Informationen aus dem zugrunde liegenden Datenmaterial zu extrahieren vermag. Eine der häufigsten Bedingungen ist die stochastische Unabhängigkeit von wiederholten Messungen. Auch wenn bei wiederholten Messungen die stochastische Unabhängigkeit plausibel erscheint, so ist es dennoch schwierig, diese sicherzustellen. Werden beispielsweise Blutdruckwerte von verschiedenen Versuchspersonen erhoben, dann erscheint es plausibel, dass die Messungen sich nicht gegenseitig beeinflussen, und diese deshalb als stochastisch unabhängig angesehen werden können. Werden allerdings die Blutdruckwerte am gleichen Ort zu nacheinander folgenden Zeitpunkten erhoben, so können die Messergebnisse voneinander stochastisch abhängig sein, wenn der Blutdruck beispielsweise über eventuelle Wetterfühligkeit der Versuchspersonen vom Wetter beeinflusst wird. Zeitnahe Messungen könnten über diesen Effekt stochastisch abhängig sein.

Bei Belastungstests ist es aus Kostengründen wünschenswert, an einem Testobjekt mehrere Belastungstests durchzuführen, um beispielsweise elastische Verformungen zu messen, anstatt bei jedem Experiment ein neues Testobjekt zu verwenden. Auch hier ist nicht ohne weiteres davon auszugehen, dass die Testobjekte gedächtnislos sind. Das Ausmaß der experimentell herbeigeführten Verformungen könnte die darauf folgenden Messergebnisse beeinflussen, so dass die stochastische Unabhängigkeit der Messergebnisse nicht mehr gegeben ist.

Die Daten aus einer jährlichen Erhebung einer Tierpopulation in einem beschränkten

Gebiet (um beispielsweise Aussagen über die mittlere Tierpopulation zu treffen) werden in der Regel ebenfalls nicht stochastisch unabhängig sein. Auch bei sich nicht ändernden Umwelteinflüssen wird eine gemessene Populationsgröße Einfluss auf die darauf folgende Messung haben.

Eine Modellierung der Abhängigkeitsstrukturen könnte sich in solchen Fällen als schwierig erweisen. Trotz dieser Betrachtung wird in vielen Fällen der Standpunkt vertreten, dass die stochastische Abhängigkeit nicht allzu groß und damit vernachlässigbar ist. Dabei stellt sich aber die Frage, ob nicht auch schwach ausgeprägte stochastische Abhängigkeit erheblich verzerrenden Einfluss auf Resultate, die mit statistischen Methoden gewonnen werden, haben kann. Insbesondere stellt sich hier auch die Frage, was „wenig“ bzw. „geringe“ stochastische Abhängigkeit bedeutet.
Ein Aspekt robuster Statistik ist die Analyse des Verhaltens statistischer Verfahren unter Störung gewisser Voraussetzungen. Beispiele zur Modellierung solcher Störungen sind Ausreißermodelle, oder auch semiparametrische Modelle.
In Kapitel 2 werden beispielhaft statistische Standardverfahren genannt, bei deren Anwendung in der Regel die stochastische Unabhängigkeit der eingehenden Daten vorausgesetzt wird, und welche im weiteren Verlauf auch detaillierter behandelt werden.
Diese Arbeit untersucht die Effekte leichter Verletzungen stochastischer Unabhängigkeit auf statistische Methoden. Dazu werden in Kapitel 3 zunächst Maße für Abhängigkeit eingeführt, um stochastische Abhängigkeit zu quantifizieren.
In Kapitel 4 werden Voraussetzungen charakterisiert, unter welchen die Auswirkungen auf statistische Methoden beliebig klein sind, wenn nur die Verletzung stochastischer Abhängigkeit im quantitativen Sinne hinreichend klein ist. Die nächsten Kapitel behandeln das asymptotische Verhalten statistischer Methoden, wenn die eingehenden Daten stochastisch abhängig sind.
In Kapitel 5 werden hinreichende Bedingungen für die Konsistenz von bestimmten Schätzerfolgen bei Abhängigkeit hergeleitet. Insbesondere wird das Verhalten der empirischen Verteilungsfunktion in Bezug auf Konsistenz untersucht, da sich viele Schätzer als Funktionale der empirischen Verteilungsfunktion darstellen lassen.
In Kapitel 6 werden Mittel geliefert, um asymptotische Verteilungen für Schätzerfolgen bei Abhängigkeit zu ermitteln. Dazu wird zunächst ein Zentraler Grenzwertsatz bewiesen, der beispielsweise in Kombination mit der Delta-Methode zur Bestimmung asymptotische Verteilungen vieler Schätzerfolgen verwendet werden kann. Darüber hinaus

wird auch das asymptotische Verhalten der empirischen Verteilungsfunktion in Bezug auf schwache Konvergenz untersucht. Dafür wird ein abstrakter funktionaler Zentraler Grenzwertsatz bei Abhängigkeit hergeleitet, der die schwache Konvergenz der empirischen Verteilungsfunktion als Spezialfall mitbehandelt.
In Kapitel 7 werden einige Anwendungsmöglichkeiten der Resultate aus Kapitel 6 demonstriert. Dazu werden die asymptotischen Verteilungen einiger Statistiken bei stochastischer Abhängigkeit bestimmt.
Die Arbeit schließt mit Kapitel 8 ab, das sowohl eine Zusammenfassung der gewonnenen Resultate beinhaltet, als auch Ansätze zur weiterführenden Untersuchung auf diesem Gebiet aufzeigt.

Kapitel 2

Motivation

In diesem Abschnitt soll anhand von Beispielen die Bedeutung stochastischer Unabhängigkeit in der Statistik kurz aufgezeigt werden, und die Problematik, die sich durch die Verletzung stochastischer Unabhängigkeit ergibt. Diese Beispiele sollen auch im Verlauf dieser Arbeit behandelt werden. Ferner soll die in dieser Arbeit aufgegriffene Idee zur mathematischen Beschreibung stochastischer Abhängigkeit erläutert werden.

Bei vielen statistischen Verfahren werden Bedingungen vorausgesetzt, unter denen ein günstiges bzw. wünschenswertes Verhalten des statistischen Verfahrens gezeigt werden kann. Ist beispielsweise $X_1, X_2, ...$ eine Folge reellwertiger Zufallsvariablen mit der gleichen Verteilungsfunktion F, dann bildet die Statistik

$$T_1[X_1, ..., X_n] := \overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

ein Verfahren zur Schätzung einer Lokation der Verteilungsfunktion F. Wird zusätzlich vorausgesetzt, dass $E[|X_i|] < \infty \; \forall i \in \mathbb{N}$ und dass die Zufallsvariablen $X_1, X_2, ...$ stochastisch unabhängig sind (Notation u.i.v.), dann stellt das starke Gesetz der großen Zahl die fast sichere Konvergenz der Schätzerfolge $T_1[X_1, ..., X_n]$ gegen den Erwartungswert $E[X_i]$ der Zufallsvariablen X_i für $n \longrightarrow \infty$ sicher. Die Statistik

$$T_2[X_1, ..., X_n] := \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

bildet ein Verfahren zur Schätzung eines Streuungsmaßes der Verteilung F. Unter der stärkeren Voraussetzung, dass $E[(X_i)^2] < \infty$ und die Zufallsvariablen $X_1, X_2, ...$ stochastisch unabhängig sind, folgt ebenfalls mit dem starken Gesetz der großen

Zahl die fast sichere Konvergenz der Schätzerfolge $T_2[X_1, ..., X_n]$ gegen die Varianz $Var[X_i] := E[(X_i - E[X_i])^2]$ der Zufallsvariable X_i. Unter den selben Voraussetzungen folgt

$$n^{1/2} \frac{\overline{X} - E[X_i]}{(Var[X_i])^{\frac{1}{2}}} \rightsquigarrow N(0,1),$$

falls $Var[X_i] > 0$, wobei „$\rightsquigarrow$“ die Verteilungskonvergenz symbolisiert, und mit $N(0,1)$ die Standardnormalverteilung bezeichnet wird. Das Lemma von Slutzky (vgl. van der Vaart (1998), Lemma 2.8) liefert

$$n^{1/2} \frac{\overline{X} - E[X_i]}{S[X_1, ..., X_n]} \rightsquigarrow N(0,1), \tag{2.1}$$

wobei

$$S[X_1, ..., X_n] := \left(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right)^{\frac{1}{2}}. \tag{2.2}$$

Wird zusätzlich vorausgesetzt, dass die Verteilungsfunktion F eine Normalverteilung mit Erwartungswert μ und Varianz σ^2 ist (Notation $X_i \sim N(\mu, \sigma^2)$), dann ist eine exakte Bestimmung der Verteilungen der Zufallsvariablen in (2.1) möglich:

$$n^{1/2} \frac{\overline{X} - E[X_i]}{S[X_1, ..., X_n]} \sim t_{n-1}, \tag{2.3}$$

wobei t_{n-1} die t-Verteilung mit $n-1$ Freiheitsgraden bezeichnet. Die Kenntnis der Verteilung der Statistik in (2.3) erlaubt die Herleitung des klassischen Einstichproben t-Tests und äquivalent dazu die Bestimmung von exakten $(1-\alpha)$-Konfidenzintervallen für das Funktional $\mu := E[X_i]$. Ist die Voraussetzung $X_i \sim N(\mu, \sigma^2)$ nicht gegeben, sondern lediglich

$$X_1, X_2, ... \overset{u.i.v.}{\sim} F \text{ mit } E[(X_i)^2] < \infty,$$

dann kann die beim t-Test verwendete Teststatistik wegen (2.1) weiterhin für einen asymptotischen Test verwendet werden. Die empirische Verteilungsfunktion soll im Folgenden mit F_n, gegeben durch

$$F_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq t\},$$

bezeichnet werden, wobei $\mathbb{1}$ für die Indikatorfunktion steht. Erfüllen die Zufallsvariablen $X_1, X_2, ... \overset{u.i.v.}{\sim} F$, so folgt aus dem Satz von Glivenko-Cantelli, dass für $n \longrightarrow \infty$

$$\|F_n(\cdot) - F(\cdot)\|_\infty \longrightarrow 0 \quad \text{fast sicher.}$$

Ist ein Funktional T hinreichend glatt an der Stelle F, dann folgt damit auch

$$T[F_n] \longrightarrow T[F] \quad \text{fast sicher.}$$

Viele Statistiken und Schätzer lassen sich als Funktionale der empirischen Verteilungsfunktion schreiben. So gilt beispielsweise

$$T_1[X_1, ..., X_n] = E[F_n], \quad T_2[X_1, ..., X_n] = Var[F_n],$$

wobei $E[F_n]$ und $Var[F_n]$ den Erwartungswert bzw. die Varianz einer Zufallsvariablen mit der Verteilungsfunktion F_n bezeichnen. Sei $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ mit der Metrik $d(s,t) := |\arctan(s) - \arctan(t)|$ versehen. Erfüllen die Zufallsvariablen $X_1, X_2, ... \overset{u.i.v.}{\sim} F$, dann folgt gemäß dem Satz von Donsker (vgl. van der Vaart (1998), Theorem 19.3), dass

$$n^{1/2}(F_n(\cdot) - F(\cdot)) \rightsquigarrow B_F(\cdot).$$

Dabei bezeichnet B_F eine F-Brownsche Brücke, und die schwache Konvergenz bezieht sich auf die Supremumsnorm auf dem Raum $D[-\infty, \infty]$ der auf $\overline{\mathbb{R}}$ definierten, reellwertigen, rechtsstetigen Funktionen mit existierenden linksseitigen Grenzwerten. Die Supremumsnorm als Abbildung von $D[-\infty, \infty]$ nach $\overline{\mathbb{R}}$ ist stetig, und mit dem Satz von der stetigen Abbildung folgt

$$n^{1/2}\|F_n(\cdot) - F(\cdot)\|_\infty \rightsquigarrow \|B_F(\cdot)\|_\infty.$$

Ist die Verteilungsfunktion F stetig, dann hängt die Verteilung von $\|B_F(t)\|_\infty$ nicht von F ab. Für eine Verteilungsfunktion F_0 kann dann $n^{1/2}\|F_n(\cdot) - F(\cdot)\|_\infty$ als Teststatistik für das Testproblem

$$H_0 : F = F_0 \quad \text{vs.} \quad H_1 : F \neq F_0$$

verwendet werden. Diese Überlegung führt schließlich zum klassischen Kolmogorov-Smirnov-Test. Durch Analyse der vom Stichprobenumfang n abhängigen Verteilung

der Teststatistik $n^{1/2}\|F_n(\cdot) - F(\cdot)\|_\infty$ ist hier ein exakter statistischer Test möglich.
Bei all diesen Verfahren stellt sich die Frage, wie sich diese verhalten, wenn die Voraussetzung der stochastischen Unabhängigkeit verletzt ist. Oftmals werden, abhängig von Sachverhalt, für erhobene Messreihen Modelle aufgestellt, die eventuelle Abhängigkeitsstrukturen möglichst gut erfassen. Die Möglichkeiten sind vielfältig. In einer Situation erscheint vielleicht ein ARMA-Modell vielversprechend. Wenn aus dem Sachzusammenhang plausibel hervorgeht, dass sich bei einer Zeitreihe (oder auch bei einem stochastischen Prozess) der Einfluss aus der Vergangenheit maßgeblich durch den zuletzt beobachteten Zustand erklären lässt, dann erscheint möglicherweise die Modellierung durch eine Markovkette oder durch ein Martingal sinnvoll. Weitere Möglichkeiten bietet hierbei auch die Ergodentheorie. Sachzusammenhänge und daraus resultierende Abhängigkeitsstrukturen bei stochastischen Prozessen können oftmals auch differenzierter über stochastische Differentialgleichungen modelliert werden.
In dieser Arbeit soll aber nicht die Modellierung von Abhängigkeitsstrukturen im Mittelpunkt stehen, sondern möglichst strukturlose Beschränkung stochastischer Abhängigkeit. An Stelle von Einschränkungen durch ein Modell soll hier stets ein quantitativ beschreibbares und beschränktes Ausmaß der stochastischen Abhängigkeit verwendet werden. Es sollen also „Umgebungen“ um den Fall stochastischer Unabhängigkeit betrachtet werden. Dazu ist es notwendig, einen Abstand zum Fall der stochastischen Unabhängigkeit quantitativ zu beschreiben. Das nächste Kapitel behandelt daher Maße für Abhängigkeit.

Kapitel 3

Abhängigkeitsmaße

Zu den wichtigsten Konzepten in der Wahrscheinlichkeitstheorie gehört das Konzept der stochastischen Unabhängigkeit. Anschaulich gesprochen sind zwei Zufallsvariablen stochastisch unabhängig, wenn durch Beobachtung der einen Zufallsvariablen keine Information über den Ausgang der anderen Zufallsvariable gewonnen werden kann. Die folgende Definition formalisiert diese Überlegung.

Definition 3.1 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, T eine Indexmenge, und $(\mathcal{A}_t)_{t\in T}$ eine Familie von Unter-σ-Algebren von $\mathcal{A}$. Die σ-Algebren $(\mathcal{A}_t)_{t\in T}$ heißen stochastisch unabhängig, wenn für je endlich viele, verschiedene Indizes $t_1, ..., t_u \in T$, die zugehörigen σ-Algebren $\mathcal{A}_{t_1}, ..., \mathcal{A}_{t_u}$ stochastisch unabhängig sind, d. h. wenn für beliebige Mengen $A_{t_1} \in \mathcal{A}_{t_1}, ..., A_{t_u} \in \mathcal{A}_{t_u}$*

$$P(A_{t_1} \cap ... \cap A_{t_u}) = P(A_{t_1}) \cdots P(A_{t_u}) \tag{3.1}$$

gilt. Sei $(X_t)_{t\in T}$ eine Familie von Zufallsvariablen. Die Zufallsvariablen $(X_t)_{t\in T}$ heißen stochastisch unabhängig, wenn die durch die Zufallsvariablen induzierten σ-Algebren $(\sigma(X_t))_{t\in T}$ stochastisch unabhängig sind (vgl. Kallenberg (2002), Chapter 3).

Während stochastische Unabhängigkeit ein genau definierter Begriff ist, liefert dieser kein kanonisches Konzept, um Abhängigkeit quantitativ zu messen. Viele statistische Methoden sind auf Daten aus wiederholten Messungen zugeschnitten, für welche die stochastische Unabhängigkeit vorausgesetzt wird. Auch wenn die stochastische Unabhängigkeit von wiederholten Messungen aufgrund eines sorgsam erstellten Versuchsplans zur Erhebung der Daten sehr plausibel erscheint, so ist sie damit nicht sichergestellt. Allerdings führen geringfügige Abweichungen von der Unabhängigkeitsannahme bei vielen statistischen Methoden möglicherweise zu nur geringfügig abweichenden Resultaten. Um Abweichungen von der stochastischen Unabhängigkeit zu quantifizieren, werden zunächst Konzepte und Maße für Abweichungen von der stochastischen Unabhängigkeit, also Abhängigkeitsmaße, benötigt. Als das Gegenstück zur stochastischen Unabhängigkeit zweier Zufallsvariablen X, Y könnte ein bestehender funktionaler Zusammenhang zwischen zwei Zufallsvariablen von der Form

$$Y = f(X) \text{ fast sicher, } f \text{ messbar,} \tag{3.2}$$

betrachtet werden. Sind zwei Zufallsvariablen stochastisch unabhängig, dann gibt es, bis auf triviale Fälle, zwischen diesen keinen solchen Zusammenhang. Angenommen zwei Zufallsvariablen X, Y mit Werten in $\mathbb{R}^u$ bzw. $\mathbb{R}^v$ wären stochastisch unabhängig, und es gäbe eine messbare Funktion f, so dass (3.2) erfüllt ist. Dann sind auch Y und $f(X)$ stochastisch unabhängig, und für jede Borelmenge $A \in \mathcal{B}(\mathbb{R}^v)$ gilt

$$\begin{aligned} P(Y \in A) &= P(Y \in A, f(X) \in A) \\ &= P(Y \in A)P(f(X) \in A) \\ &= (P(Y \in A))^2. \end{aligned}$$

Für jede Borelmenge A folgt daraus, dass entweder $P(Y \in A) = 1$ oder $P(Y \in A) = 0$, also ist Y fast sicher eine Konstante. Es erscheint daher naheliegend, stochastische Abhängigkeit durch einen Anteil eines funktionalen Zusammenhangs zu quantifizieren. Allerdings erscheint dieser Ansatz zu einer Quantifizierung stochastischer Abhängigkeit nicht erfolgsversprechend, wie das folgende Beispiel veranschaulicht: Ist X eine auf dem Intervall $[0, 1]$ gleichverteilte Zufallsvariable, dann ist $Y_m := f_m(X) := 10^m X - [10^m X]$, für $m \in \mathbb{N}$ ebenfalls auf $[0, 1]$ gleichverteilt, wobei $[t]$ die kleinste ganze Zahl bezeichnet, die nicht kleiner als t ist. Überdies sind die Zufallsvariablen $(X - 10^{-m} Y_m)$ und Y_m

stochastisch unabhängig. Sind I_1, I_2 beliebige Intervalle in $[0,1]$, dann folgt deswegen

$$\begin{aligned}
P(X \in I_1)\ &P(Y_m \in I_2) - 10^{-m} - 10^{-m} P(Y_m \in I_2) \\
&= (P(X \in I_1) - 10^{-m})\ P(Y_m \in I_2) - 10^{-m} \\
&\le P((X - 10^{-m} Y_m) \in I_1)\ P(Y_m \in I_2) - 10^{-m} \\
&= P((X - 10^{-m} Y_m) \in I_1,\ Y_m \in I_2) - 10^{-m} \\
&\le P((X - 10^{-m} Y_m) + 10^{-m} Y_m \in I_1,\ Y_m \in I_2) \\
&= P(X \in I_1,\ Y_m \in I_2).
\end{aligned}$$

Mit der gleichen Argumentation folgt auch

$$P(X \in I_1,\ Y_m \in I_2) \le P(X \in I_1)\ P(Y_m \in I_2) + 10^{-m} + 10^{-m} P(Y_m \in I_2).$$

Da natürlich $P(Y_m \in I_2) \in [0,1]$, kann die Differenz zwischen $P(X \in I_1,\ Y_m \in I_2)$ und $P(X \in I_1)\ P(Y_m \in I_2)$ demnach eingeschachtelt werden durch

$$-\frac{2}{10^m} \le P(X \in I_1,\ Y_m \in I_2) - P(X \in I_1)\ P(Y_m \in I_2) \le \frac{2}{10^m},$$

und damit gilt

$$|P(X \in I_1,\ Y_m \in I_2) - P(X \in I_1)\ P(Y_m \in I_2)| \le \frac{2}{10^m}$$

für alle Intervalle I_1, I_2. Obwohl zwischen X und Y_m ein funktionaler Zusammenhang der Form (3.2) besteht, verhalten sich diese Zufallsvariablen für große Werte m fast so, als seien sie stochastisch unabhängig. Dieses Beispiel ist nicht auf gleichverteilte Zufallsvariablen eingeschränkt. Wenn beispielsweise F, G stetige, streng monoton wachsende Verteilungsfunktionen auf $\mathbb{R}$ sind, dann können statt X und Y_m auch $F(X)$ und $G(Y_m)$ in gleicher Weise wie oben betrachtet werden, und $F(X), G(Y_m)$ haben die Verteilungsfunktionen F bzw. G. Daher soll zur quantitativen Bewertung stochastischer Abhängigkeit eine andere Strategie verfolgt werden.

3.1 Abhängigkeitsmaße zwischen zwei Zufallsvariablen

Zunächst wird ein klassisches Konzept betrachtet. Eines der bekanntesten Maße für die Abhängigkeit zweier σ-Algebren eines Wahrscheinlichkeitsraums ist der strong-mixing- oder auch α-Koeffizient, welcher auf die Arbeit Rosenblatt (1956) zurückgeht.

Definition 3.2 *Sei* $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum, und* $\mathcal{A}_1$ *und* $\mathcal{A}_2$ *seinen Unter-*σ*-Algebren von* $\mathcal{A}$*. Dann ist durch*

$$\alpha(\mathcal{A}_1, \mathcal{A}_2) := \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)|,\ A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\} \tag{3.3}$$

ein Maß für die Abhängigkeit zwischen den σ*-Algebren* $\mathcal{A}_1$ *und* $\mathcal{A}_2$ *gegeben.*

Mit diesem Maß kann auch die Abhängigkeit zwischen zwei Zufallsvariablen $X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^u, \mathcal{B}(\mathbb{R}^u))$ und $Y : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^v, \mathcal{B}(\mathbb{R}^v))$ quantitativ bewertet werden, indem die von den Zufallsvariablen induzierten σ-Algebren betrachtet werden. Die Zufallsvariablen X, Y sind genau dann unabhängig, wenn

$$\alpha(X, Y) := \alpha(\sigma(X), \sigma(Y)) = 0, \tag{3.4}$$

wobei $\sigma(X)$ und $\sigma(Y)$ die von den Zufallsvariablen X bzw. Y induzierten σ-Algebren bezeichnen. Sind zwei reellwertige Zufallsvariablen $X, Y : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ quadratisch integrierbar, so sind die Kovarianz und die Korrelation zwischen X und Y definiert durch

$$\begin{aligned} Cov[X, Y] &:= E[(X - E[X])(Y - E[Y])], \\ Corr[X, Y] &:= \frac{Cov[X, Y]}{(Var[X]Var[Y])^{1/2}}. \end{aligned} \tag{3.5}$$

Die Kovarianz und die Korrelation zwischen zwei Zufallsvariablen ist 0, wenn diese stochastisch unabhängig sind, der Umkehrschluß ist i. A. falsch. Ein aus der Korrelation abgeleitetes Maß für Abhängigkeit zwischen zwei σ-Algebren ist die maximale Korrelation.

Definition 3.3 *Es gelten die Voraussetzungen von Definition 3.2, und*

$$\Delta_2(\mathcal{A}_1) := \{X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}, \mathbb{R}),\ X \textit{ messbar bzgl. } \mathcal{A}_1,\ X \in L_2\},$$

$$\Delta_2(\mathcal{A}_2) := \{X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}, \mathbb{R}),\ X \textit{ messbar bzgl. } \mathcal{A}_2,\ X \in L_2\}$$

bezeichne die reellwertigen, quadratisch integrierbaren Zufallsvariablen, welche bzgl. $\mathcal{A}_1$ *bzw.* $\mathcal{A}_2$ *messbar sind. Durch*

$$\rho(\mathcal{A}_1, \mathcal{A}_2) := \sup\{|Corr[X, Y]|,\ X \in \Delta_2(\mathcal{A}_1),\ Y \in \Delta_2(\mathcal{A}_2)\} \tag{3.6}$$

wird ein Abhängigkeitsmaß zwischen zwei σ*-Algebren definiert, bezeichnet als maximale Korrelation zwischen* $\mathcal{A}_1$ *und* $\mathcal{A}_2$*.*

Wie in (3.4) kann auch hier $\rho(X,Y) := \rho(\sigma(X), \sigma(Y))$ gesetzt werden. Durch Betrachtung von Indikatorfunktionen kann gezeigt werden, dass

$$4\alpha(\mathcal{A}_1, \mathcal{A}_2) \leq \rho(\mathcal{A}_1, \mathcal{A}_2)$$

(vgl. Hall und Heyde (1980), Chapter 5.5). Wenn $\rho(\mathcal{A}_1, \mathcal{A}_2) = 0$, dann ist also auch $\alpha(\mathcal{A}_1, \mathcal{A}_2) = 0$. Somit sind auch die σ-Algebren $\mathcal{A}_1$ und $\mathcal{A}_2$ voneinander unabhängig. Die folgende Definition stellt noch weitere Abhängigkeitsmaße vor, die nach ähnlichen Prinzipien konstruiert sind.

Definition 3.4 *Es gelten die Voraussetzungen aus Definition 3.2, und $\mathcal{A}_1, \mathcal{A}_2$ seien zwei Unter-σ-Algebren von $\mathcal{A}$. Durch folgende Größen sind Maße für Abhängigkeit zwischen $\mathcal{A}_1$ und $\mathcal{A}_2$ gegeben:*

$$\phi(\mathcal{A}_1, \mathcal{A}_2) := \sup\left\{\left|\frac{P(A_1 \cap A_2)}{P(A_2)} - P(A_1)\right|,\ A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, P(A_2) > 0\right\},$$
$$\psi(\mathcal{A}_1, \mathcal{A}_2) := \sup\left\{\left|\frac{P(A_1 \cap A_2)}{P(A_1)P(A_2)} - 1\right|,\ A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, P(A_1) > 0, P(A_2) > 0\right\},$$
$$\beta(\mathcal{A}_1, \mathcal{A}_2) := \sup\left\{\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J} |P(A_{1,i} \cap A_{2,j}) - P(A_{1,i})P(A_{2,j})|\right\},$$

wobei für $\beta(\mathcal{A}_1, \mathcal{A}_2)$ das Supremum über alle endlichen Partitionen $\{A_{1,1}, ..., A_{1,I}\}$ und $\{A_{2,1}, ..., A_{2,J}\}$ von Ω mit $A_{1,i} \in \mathcal{A}_1$ bzw. $A_{2,j} \in \mathcal{A}_2$ gebildet wird.

Die Abhängigkeitsmaße aus Definition 3.2 - 3.4 werden auch in Bradley (2005) vorgestellt. Eine ausführlichere Untersuchung dieser Abhängigkeitsmaße ist in Bradley (2007) enthalten. Für die vorliegende Arbeit sind der α-Koeffizient $\alpha(\mathcal{A}_1, \mathcal{A}_2)$, die maximale Korrelation $\rho(\mathcal{A}_1, \mathcal{A}_2)$, sowie eine Abwandlung des Abhängigkeitsmaßes $\psi(\mathcal{A}_1, \mathcal{A}_2)$ von Bedeutung. Weitere Möglichkeiten ergeben sich durch die Betrachtung bedingter Erwartungswerte. Für eine L_p-integrierbare Zufallsvariable X und eine σ-Algebra $\mathcal{M}$ kann die Größe

$$\|E[X|\mathcal{M}] - E[X]\|_p := (E[|E[X|\mathcal{M}] - E[X]|^p])^{\frac{1}{p}}$$

Aspekte der Abhängigkeit zwischen der Zufallsvariablen X und der σ-Algebra $\mathcal{M}$ beleuchten. Offensichtlich nimmt diese Größe den Wert 0 an, wenn X und $\mathcal{M}$ voneinander stochastisch unabhängig sind. Anstatt die Zufallsvariable X direkt zu verwenden, kann auch die Zufallsvariable $g(X)$ für eine entsprechend messbare Abbildung g verwendet werden. Diese Überlegung führt zu Größen der Form $\|E[g(X)|\mathcal{M}] - E[g(X)]\|_p$. Dies kann beispielsweise hilfreich sein, wenn die Zufallsvariable X selbst nicht L_p-integrierbar ist. Häufig wird zur Definition eines Abhängigkeitsmaßes auch das Supremum solcher Größen über eine Funktionenklasse $\mathcal{G}$ herbeigezogen. Sei d eine Metrik auf $\mathbb{R}^v$ und

$$Lip(g) := \inf\{L \in \mathbb{R}_{\geq 0},\ |g(x) - g(y)| \leq L\, d(x, y)\ \forall x, y \in \mathbb{R}^v\}$$

die Lipschitzkonstante von g, wobei hier $\inf\{\emptyset\} := \infty$. Mit

$$\Lambda^{(l)}(d) := \{g : \mathbb{R}^v \longrightarrow \mathbb{R},\ Lip(g) \leq l\}$$

wird die Klasse der Lipschitz-stetigen Funktionen g mit Lipschitzkonstanten nicht größer als l bezeichnet. Wenn X eine L_p-integrierbare Zufallsvariable ist, d die euklidische Metrik und $g \in \Lambda^{(1)}(d)$, dann ist auch $g(X)$ eine L_p-integrierbare Zufallsvariable. Mittels dieser Funktionenklasse wird ein Maß für Abhängigkeit definiert (vgl. auch Dedecker und Prieur (2005), Section 1).

Definition 3.5 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{M}$ eine Unter-σ-Algebra von $\mathcal{A}$, d die euklidische Metrik auf $\mathbb{R}^v$, und $X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^u, \mathcal{B}(\mathbb{R}^u))$ sei L_p-integrierbar für ein $p \geq 1$. Dann stellt*

$$\theta_p(\mathcal{M}, X) := \sup\{\|E[g(X)|\mathcal{M}] - E[g(X)]\|_p\ ,\ g \in \Lambda^{(1)}(d)\} \tag{3.7}$$

ein Maß für die Abhängigkeit zwischen der σ-Algebra $\mathcal{M}$ und der Zufallsvariablen (bzw. des Zufallsvektors) X dar.

Dieses Abhängigkeitsmaß zeigt ein vielfältiges Konzept auf, um weitere Maße für Abhängigkeit zwischen einer Zufallsvariablen und einer σ-Algebra zu definieren. Dazu kann zum einen die Funktionenklasse variiert werden, zum anderen kann auch die Bildung der L_p-Norm mit der Bildung des Supremums vertauscht werden. Diese Überlegung liefert ein weiteres Maß für Abhängigkeit (vgl. auch Dedecker und Prieur (2005), Section 7).

Definition 3.6 *Unter den Voraussetzungen von Definition 3.5 sei* $\tau_p(\mathcal{M}, X)_p$ *definiert durch*

$$\begin{aligned}\tau_p(\mathcal{M}, X) := \Bigg\| \sup \Bigg\{ &\int_{\mathbb{R}^k} g(t) E[\mathbb{1}\{X \le t\}|\mathcal{M}] dt \\ &- \int_{\mathbb{R}^k} g(t) E[\mathbb{1}\{X \le t\}] dt \ , \ g \in \Lambda^{(1)}(d) \Bigg\} \Bigg\|_p .\end{aligned} \tag{3.8}$$

Dieses Abhängigkeitsmaß erlaubt es, eine reellwertige Zufallsvariable X, welche von der σ-Algebra $\mathcal{M}$ stochastisch abhängig ist, durch eine reellwertige Zufallsvariable X^* in L_1-Norm quantitativ zu approximieren. Unter leichten Regulatitätsbedingungen gibt es immer eine von der σ-Algebra $\mathcal{M}$ unabhängige Zufallsvariable X^*, so dass X^*, X die gleiche Verteilung besitzen und $\|X - X^*\|_1 = \tau_1(\mathcal{M}, X)$ (vgl. Dedecker und Prieur (2005), Lemma 2). Zusammenhänge dieser Art werden oftmals dazu genutzt, um Fälle von Abhängigkeit auf Fälle von Unabhängigkeit zurückzuführen. Sei nun d die euklidische Metrik auf $\mathbb{R}$. Ein weiteres Maß für Abhängigkeit liefert die Wahl einer Teilklasse von $\Lambda^{(1)}(d)$. Seien

$$\begin{aligned}\Lambda_{\cos} &:= \{g : \mathbb{R} \longrightarrow \mathbb{R},\ g(x) = (|t| \vee 1)^{-1} \cos(tx),\ t \in \mathbb{R}\},\\ \Lambda_{\sin} &:= \{g : \mathbb{R} \longrightarrow \mathbb{R},\ g(x) = (|t| \vee 1)^{-1} \sin(tx),\ t \in \mathbb{R}\} \text{ und}\\ \Lambda_{tri} &:= \Lambda_{\cos} \cup \Lambda_{\sin},\end{aligned}$$

wobei $a \vee b,\ a \wedge b$ das Maximum bzw. Minimum zweier Zahlen symbolisiert. Jede Funktion $g \in \Lambda_{tri}$ ist Lipschitz-stetig mit Lipschitzkonstante $Lip(g) \le 1$, da für jede Funktion $g \in \Lambda_{tri}$ die Ableitung $\frac{d}{dx} g$ betragsmäßig nicht größer als 1 ist. Zusätzlich sind die Funktionen $g \in \Lambda_{tri}$ selbst betragsmäßig nicht größer als 1. Analog zur Definition 3.5 kann mittels dieser Funktionenklasse ein weiteres Abhängigkeitsmaß definiert werden.

Definition 3.7 *Sei, wie in Definition 3.5,* $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum,* $\mathcal{M}$ *eine Unter-*σ*-Algebra von* $\mathcal{A}$ *und* $X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *eine Zufallsvariable. Dann sei* $\kappa_p(\mathcal{M}, X)$ *gegeben durch*

$$\kappa_p(\mathcal{M}, X) := \sup\{\|E[g(X)|\mathcal{M}] - E[g(X)]\|_p \ , \ g \in \Lambda_{tri}\}. \tag{3.9}$$

Bereits in Definition 3.6 tauchten die Verteilungsfunktion eines Zufallsvektors X und die auf eine σ-Algebra bedingte Verteilungsfunktion auf. Es ist naheliegend, stochastische Abhängigkeit durch die Unterschiede oder Differenzen zwischen einer Verteilungsfunktion und einer bedingten Verteilungsfunktion zu quantifizieren. Hierfür sind einige

Definitionen notwendig. Zunächst wird eine Ordnungsrelation auf $\mathbb{R}^u$ und der Begriff des multivariaten Intervalls eingeführt.

Definition 3.8 *Sei $s, t \in \mathbb{R}^u$. Dann sei $s \leq t :\Leftrightarrow s_i \leq t_i \; \forall i = 1, ..., u$ und,*

$$\begin{aligned}
[s,t] &:= [s_1, t_1] \times ... \times [s_u, t_u], \\
(s,t] &:= (s_1, t_1] \times ... \times (s_u, t_u], \\
(s,t] &:= [s,t] := \emptyset, \text{ falls } s_i > t_i \text{ für ein } i \in \{1, ..., u\}.
\end{aligned} \tag{3.10}$$

In natürlicher Weise können damit bedingte Verteilungsfunktionen über bedingte Erwartungswerte definiert werden.

Definition 3.9 *Sei $\omega \in \Omega$ und $\mathcal{M}$ eine Unter-σ-Algebra von $\mathcal{A}$. Die durch $\mathcal{M}$ bedingte Verteilungsfunktion $F_{X|\mathcal{M}}$ des Zufallsvektors $X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^u, \mathcal{B}(\mathbb{R}^u))$ und die Verteilungsfunktion F_X von X sind definiert bzw. gegeben durch*

$$\begin{aligned}
F_{X|\mathcal{M}}(t)(\omega) &:= E[\mathbb{1}\{X \leq t\}|\mathcal{M}](\omega), \\
F_X(t) &= E[\mathbb{1}\{X \leq t\}],
\end{aligned}$$

wobei die bedingte Verteilungsfunktion $F_{X|\mathcal{M}}$ bis auf eine Nullmenge eindeutig ist.

Da $\mathbb{R}^u$ ein polnischer Raum ist, gibt es eine Version von $F_{X|\mathcal{M}}(\cdot)(\omega)$, die fast sicher eine Verteilungsfunktion ist. Es gilt sogar, dass es ein $M \in \mathcal{M}$ mit $P(M) = 1$ gibt, so dass $F_{X|\mathcal{M}}(\cdot)(\omega)$ eine Verteilungsfunktion für jedes $\omega \in M$ ist (vgl. Dudley (2002), Theorem 10.2.2). Im Folgenden ist stets eine derartige Version gemeint. Für das nächste Abhängigkeitsmaß wird ein Abstandsmaß zwischen Funktionen benötigt, das die folgende Definition liefert.

Definition 3.10 *Sei $T \subset \mathbb{R}^u$ und $f : T \longrightarrow \mathbb{R}$, und λ^u bezeichne das Lebesguemaß auf $\mathbb{R}^u$. Dann definiert*

$$\operatorname*{esssup}_{t \in T} f(t) := \inf\{u \in \mathbb{R} \cup \{\infty\}, \; f(t) \leq u \; \forall t \in T \setminus N, \; \lambda^u(N) = 0\}$$

das essentielle Supremum von f auf T und

$$\|f\|_{\infty,T} := \operatorname*{esssup}_{t \in T} |f(t)| \tag{3.11}$$

die Supremumsnorm auf T. Für $T = \mathbb{R}^u$ heißt

$$\|f\|_\infty := \|f\|_{\infty,\mathbb{R}^u}$$

Supremumsnorm von f .

Es ist naheliegend, mittels der Supremumsnorm ein Abhängigkeitsmaß einzuführen, das auf der Differenz zwischen einer Verteilungsfunktion und einer bedingten Verteilungsfunktion basiert (vgl. auch Dedecker und Prieur (2005)).

Definition 3.11 *Sei $X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^u, \mathcal{B}(\mathbb{R}^u))$. Für eine Unter-$\sigma$-Algebra $\mathcal{M} \subset \mathcal{A}$ sei*

$$\xi(\mathcal{M}, X) := E\left[\|F_{X|\mathcal{M}} - F_X\|_\infty\right]. \tag{3.12}$$

D. h. $\xi(\mathcal{M}, X)$ ist der erwartete Supremumsabstand zwischen der Verteilungsfunktion von X und der auf $\mathcal{M}$ bedingten Verteilungsfunktion von X.

3.2 Abhängigkeitsmaße bei Folgen von Zufallsvariablen

Mit Hilfe von Abhängigkeitsmaßen zwischen σ-Algebren und Zufallsvariablen, oder alternativ zwischen zwei σ-Algebren, können auch Abhängigkeitsstrukturen von Folgen von Zufallsvariablen $(X_i)_{i\in\mathbb{Z}}$ charakterisiert werden. Liegt eine Folge $(X_i)_{i\in\mathbb{Z}}$ von Zufallsvariablen vor, dann ist in vielen Fällen aus der Praxis die Annahme plausibel, dass die Zufallsvariablen $(X_i)_{i\le n}$ keinen großen Einfluss auf die Zufallsvariablen $(X_i)_{i\ge n+k}$ haben, wenn $k \in \mathbb{N}$ groß ist. Wird die Folge $(X_i)_{i\in\mathbb{Z}}$ als Zeitreihe aufgefasst, so bedeutet dies anschaulich gesprochen, dass die Vergangenheit keinen großen Einfluss auf eine zeitlich weit entfernte Zukunft hat. Sei

$$\begin{aligned} \mathcal{A}_n &:= \sigma(X_i,\ i \le n), \\ \mathcal{A}^n &:= \sigma(X_i,\ i \ge n), \end{aligned} \tag{3.13}$$

wobei $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^u, \mathcal{B}(\mathbb{R}^u)),\ i \in \mathbb{Z}$, eine Folge von Zufallsvektoren mit Werten in $\mathbb{R}^u$ ist. Die folgenden Definitionen zeigen Konzepte zur Beschreibung von Beschränkungen von Abhängigkeitsstrukturen bei Folgen von Zufallsvariablen auf, die größtenteils mittels obiger Abhängigkeitsmaße formuliert werden können.

Definition 3.12 *Eine Folge von Zufallsvektoren* $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^u, \mathcal{B}(\mathbb{R}^u)),\ i \in \mathbb{Z}$ *(bzw.* $i \in \mathbb{N}$*), heißt m-dependent, falls (für den Wert* $m \in \mathbb{N}$*) die* σ*-Algebren* $\mathcal{A}_n$ *und* $\mathcal{A}^{n+l}$ *für jeden Wert* $n \in \mathbb{Z}$ *(bzw.* $n \in \mathbb{N}$*) und für jeden Wert* $l \geq m$ *stochastisch unabhängig sind.*

In Bezug auf das in Definition 3.2 gegebene Abhängigkeitsmaß führt die gleiche Überlegung zur folgenden Definition.

Definition 3.13 *Für* $l \geq 0$ *sei*

$$\begin{aligned} \alpha(l) &:= \sup\{\alpha(\mathcal{A}_n, \mathcal{A}^{n+l}),\ n \in \mathbb{Z}\} \text{ bzw.} \\ \alpha(l) &:= \sup\{\alpha(\mathcal{A}_n, \mathcal{A}^{n+l}),\ n \in \mathbb{N}\}, \end{aligned} \tag{3.14}$$

wobei $(X_i)_{i\in\mathbb{Z}}$ *oder* $(X_i)_{i\in\mathbb{N}}$ *eine Folge von Zufallsvektoren mit Werten in* $\mathbb{R}^u$ *ist, und* $\mathcal{A}_n, \mathcal{A}^{n+l}$ *wie in (3.13) definiert sind. Die Folge heißt* α*-mixing, falls* $\alpha(n) \longrightarrow 0$ *für* $n \longrightarrow \infty$*.*

Analog dazu kann anstelle des Abhängigkeitsmaßes α das Abhängigkeitsmaß ρ gesetzt werden.

Definition 3.14 *Unter den gleichen Voraussetzungen wie in Definition 3.13 sei für* $l \geq 0$

$$\begin{aligned} \rho(l) &:= \sup\{\rho(\mathcal{A}_n, \mathcal{A}^{n+l}),\ n \in \mathbb{Z}\} \text{ bzw.} \\ \rho(l) &:= \sup\{\rho(\mathcal{A}_n, \mathcal{A}^{n+l}),\ n \in \mathbb{N}\}. \end{aligned} \tag{3.15}$$

Die Folge wird ρ*-mixing genannt, falls* $\rho(n) \longrightarrow 0$ *für* $n \longrightarrow \infty$*.*

Ein im gewissen Sinne weniger einschränkendes Maß für Abhängigkeit als $\rho(l)$ liefert die folgende Definition.

Definition 3.15 *Für* $l \geq 0$ *sei*

$$\begin{aligned} \kappa_p(l) &:= \sup_{n\in\mathbb{Z}, m\in\mathbb{N}} \kappa_p\left(\mathcal{A}_n, \sum_{i=n+l+1}^{m+n+l} X_i\right) \text{ bzw.} \\ \kappa_p(l) &:= \sup_{n\in\mathbb{N}, m\in\mathbb{N}} \kappa_p\left(\mathcal{A}_n, \sum_{i=n+l+1}^{m+n+l} X_i\right), \end{aligned} \tag{3.16}$$

wobei $(X_i)_{i\in\mathbb{Z}}$ *bzw.* $(X_i)_{i\in\mathbb{N}}$ *eine Folge reellwertiger Zufallsvariablen ist.*

Für $\kappa_1(n)$ kann mit Hilfe der Größe $\rho(n)$ eine Abschätzung formuliert werden. Dafür wird zunächst folgendes Lemma benötigt.

Lemma 3.16 *Sei* $X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *mit* $E[X] = 0$ *und* $|X| \leq 1$ *fast sicher,* $\mathcal{A}^*$ *und* $\mathcal{A}_*$ *Unter-*σ*-Algebren von* $\mathcal{A}$*, und* X *sei* $\mathcal{A}^*$*-messbar. Dann gibt es Zufallsvariablen* Y, Z*, so dass* Z *messbar ist bezüglich* $\mathcal{A}^*$*,* Y *messbar bezüglich* $\mathcal{A}_*$*,* $E[Y] = E[Z] = 0$*,* $Var[Y] = Var[Z] = 1$ *und*

$$E[|E[X|\mathcal{A}_*]|] \leq |E[ZY]| = |Corr[Z, Y]|. \tag{3.17}$$

Beweis. Sei $E[X|\mathcal{A}_*]$ eine Version des bedingten Erwartungswerts. Die Mengen

$$\begin{aligned} A^+ &:= \{\omega \in \Omega,\ E[X|\mathcal{A}_*] \geq 0\}, \\ A^- &:= \{\omega \in \Omega,\ E[X|\mathcal{A}_*] < 0\} \end{aligned}$$

sind in $\mathcal{A}_*$ enthalten. Nach Definition bedingter Erwartungswerte gilt

$$\begin{aligned} E[|E[X|\mathcal{A}_*]|] &= E[|\mathbb{1}\{A^+\}E[X|\mathcal{A}_*] + \mathbb{1}\{A^-\}E[X|\mathcal{A}_*]|] \\ &\leq E[\mathbb{1}\{A^+\}E[X|\mathcal{A}_*] - \mathbb{1}\{A^-\}E[X|\mathcal{A}_*]] \\ &= E[E[X|\mathcal{A}_*](\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\})] \\ &= E[X(\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\})], \end{aligned}$$

da $\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\}$ messbar ist bezüglich $\mathcal{A}_*$. Seien

$$\begin{aligned} V &:= Var[\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\}], \\ E &:= E\left[\frac{\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\}}{V^{1/2}}\right], \\ Y &:= \frac{\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\}}{V^{1/2}} - E, \\ Z &:= \frac{X}{(Var[X])^{1/2}}. \end{aligned}$$

Da $|X|, |\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\}| \leq 1$, folgt $Var[X] \leq 1$ und $V \leq 1$. Angenommen $V^{1/2} > 0$. Weil $E[X] = 0$ folgt

$$\begin{aligned} E[X(\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\})] &\leq |E[X(\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\})]| \\ &\leq |E[X(Y + E)]| = |E[XY]| \\ &\leq |E[ZY]|, \end{aligned}$$

wobei Z messbar ist bezüglich $\mathcal{A}^*$ und Y bezüglich $\mathcal{A}_*$, womit die Behauptung für den Fall $V^{1/2} > 0$ bewiesen ist.
Falls $Var[\mathbb{1}\{A^+\} - \mathbb{1}\{A^-\}] = V = 0$, folgt unmittelbar $E[X|\mathcal{A}_*] = C$ fast sicher, $C \in \mathbb{R}$, und wegen $E[X] = 0$ gilt $E[X|\mathcal{A}_*] = 0$ fast sicher, und (3.17) gilt in diesem Fall auch wenn beispielsweise $Y = Z = 0$ gesetzt wird. □

Mittels Lemma 3.16 können der maximale Korrelationskoeffizient $\rho(n)$ und der Abhängigkeitskoeffizient $\kappa_1(n)$ in Relation gesetzt werden.

Lemma 3.17 *Für eine reellwertige Folge von Zufallsvariablen $(X_i)_{i\in\mathbb{Z}}$ bzw. $(X_i)_{i\in\mathbb{N}}$ folgt*

$$\kappa_1(n) \leq 2\rho(n). \tag{3.18}$$

Beweis. Seien $n \in \mathbb{Z},\ l, m \in \mathbb{N}$ und

$$X := (|t| \vee 1)^{-1} \cos\left(t \sum_{i=l+n+1}^{l+n+m} X_i\right) - E\left[(|t| \vee 1)^{-1} \cos\left(t \sum_{i=l+n+1}^{l+n+m} X_i\right)\right].$$

Die Betrachtung von X zeigt $E[X/2] = 0$ und $|X/2| \leq 1$. Mit Lemma 3.16 und der Definition des maximalen Korrelationskoeffizienten $\rho(l)$ folgt

$$E[|E[X|\mathcal{A}_n]|] \leq 2\rho(l).$$

Die gleiche Argumentation mit der Funktion $\sin(t)$ ergibt

$$(|t| \vee 1)^{-1} E\left|E\left[\sin\left(t \sum_{i=l+n+1}^{l+n+m} X_i\right) |\mathcal{B}_n\right] - E\left[\sin\left(t \sum_{i=l+n+1}^{l+n+m} X_i\right)\right]\right| \leq \ 2\rho(l).$$

Die Definition des Abhängigkeitskoeffizienten $\kappa_1(l)$ liefert schließlich

$$\kappa_1(l) \leq 2\rho(l). \tag{3.19}$$

□

Damit wäre die Konvergenz $\rho(n) \longrightarrow 0$ für $n \longrightarrow \infty$ eine stärkere Forderung als $\kappa_1(n) \longrightarrow 0$.

3.3 Abhängigkeitsmaße zwischen den Komponenten von Zufallsvektoren

Auf ähnliche Weise können Abhängigkeiten innerhalb eines Zufallsvektors $X = (X_1, ..., X_n)$ mit Werten in $\mathbb{R}^n$ (wobei n hier als fester Wert betrachtet wird) quantitativ beurteilt werden. Sei dazu $X' := (X'_1, ..., X'_n)$ ein Zufallsvektor mit unabhängigen Komponenten X'_i, so dass X_i und X'_i für jedes i die gleiche Verteilung haben. Eine erste einfache Möglichkeit bietet beispielsweise der direkte Vergleich zwischen der Verteilung von X und X' anhand der totalen Variation.

Definition 3.18 *Sei X ein Zufallsvektor mit Werten in $\mathbb{R}^n$. Maße für die Abhängigkeit innerhalb der Zufallsvektors X sind durch*

$$\begin{aligned} \alpha(X) &:= \sup\{|P^X(B) - P^{X'}(B)|,\ B \in \mathcal{B}(\mathbb{R}^n)\}, \\ \psi(X) &:= \sup\left\{\left|\frac{P^X(B)}{P^{X'}(B)} - 1\right|,\ B \in \mathcal{B}(\mathbb{R}^n),\ P^{X'}(B) > 0\right\} \end{aligned} \tag{3.20}$$

gegeben, wobei P^X bzw. $P^{X'}$ das Bildmaß von X bzw. X' bezeichnet.

Wenn $\alpha(X) = 0$, dann sind offensichtlich auch $X_1, ..., X_n$ stochastisch unabhängig. Mit dem Abhängigkeitsmaß ψ aus (3.20) verhält es sich genauso. Angenommen $\psi(X) = 0$. Dann folgt $P^X(B) = P^{X'}(B)$ falls $P^{X'}(B) \neq 0$. Dies gilt insbesondere für alle Mengen B von der Form $(-\infty, t] = (-\infty, t_1] \times ... \times (-\infty, t_n] \subset \mathbb{R}^n$. Sei nun $(-\infty, t]$, so dass $P^{X'}((-\infty, t]) = 0$. Wegen der Unabhängigkeit von $X'_1, ..., X'_n$ ist dies gleichbedeutend mit

$$P(X'_1 \in (-\infty, t_1])P(X'_2 \in (-\infty, t_2]) \cdots P(X'_n \in (-\infty, t_n]) = 0.$$

Da X_i und X'_i für jedes $i \in \{1, ..., n\}$ die gleichen Verteilungen besitzen, folgt

$$\begin{aligned} &P(X_i \in (-\infty, t_i]) = 0 \text{ für ein } i \in \{1, ..., n\} \text{ bzw.} \\ &\ P(X \in (-\infty, t]) = 0. \end{aligned}$$

Zusammenfassend ergibt sich

$$P^X((-\infty, t]) = P^{X'}((-\infty, t]) \quad \forall t \in \mathbb{R}^n.$$

Demnach sind die Verteilungen P^X und $P^{X'}$ gleich, und damit sind auch $X_1, ..., X_n$ stochastisch unabhängig.

Über die Definition 3.13 kann für eine Folge $(X_i)_{i\in\mathbb{Z}}$ von Zufallsvariablen mit

$$\alpha((X_i)_{i\in\mathbb{Z}}) := \sup\{\alpha(l),\ l \in \mathbb{N}\} \tag{3.21}$$

ebenfalls ein Maß für Abhängigkeit innerhalb der Folge $(X_i)_{i\in\mathbb{Z}}$ definiert werden. Mit den Größen $\rho(l), \kappa_p(l)$ aus Definition 3.14 bzw. Definition 3.15 kann gleichermaßen verfahren werden. Für einen Zufallsvektor $X = (X_i, ..., X_n)$ sei analog zu (3.13)

$$\mathcal{A}_m := \sigma(X_1, ..., X_m), \quad \mathcal{A}^m := \sigma(X_m, ..., X_n).$$

Definition 3.19 *Für einen Zufallsvektor $X = (X_1, ..., X_n)$ sei*

$$\alpha_X(l) := \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)|,\ A_1 \in \mathcal{A}_m,\ A_2 \in \mathcal{A}^{m+l},\ 1 \le m \le n-l\},$$
$$\xi_X(l) := \sup\{\xi(\mathcal{A}_m, X_{m+l}),\ 1 \le m \le n-l\}.$$

Der Unterschied von $\alpha(l)$ aus Definition 3.13 und $\alpha_X(l)$ aus Definition 3.19 besteht nur darin, dass im letzteren Fall Zufallsvektoren und nicht Folgen wie z. B. $(X_i)_{i\in\mathbb{Z}}$ betrachtet werden.

Bemerkung 3.20 *Ist $\hat{X}$ ein Zufallsvektor wie X, so dass einige (oder gar alle) Komponenten X_i aus X durch u.i.v.-Kopien $\hat{X}_i$ in $\hat{X}$ ersetzt werden, dann gilt offensichtlich*

$$\alpha_{\hat{X}}(l) \le \alpha_X(l).$$

Die folgenden Abhängigkeitsmaße sind von ähnlicher Gestalt wie in (3.21) und entsprechen einfach deren Übertragung auf Zufallsvektoren.

Definition 3.21 *Für einen Zufallsvektor X mit Werten in $\mathbb{R}^n$ sei*

$$\begin{aligned} \alpha'(X) &:= \sup\{\alpha_X(l),\ 1 \le l \le n-1\} \\ \xi(X) &:= \sup\{\xi_X(l),\ 1 \le l \le n-1\}. \end{aligned} \tag{3.22}$$

Wie bei den Abhängigkeitsmaßen aus (3.20) folgt aus $\alpha'(X) = 0$ ebenfalls die stochastische Unabhängigkeit von $X_1, ..., X_n$. Die stochastische Unabhängigkeit von X_1 und X_2 geht unmittelbar aus $\alpha'(X) = 0$ hervor. Darüber hinaus ist (X_1, X_2) von X_3, (X_1, X_2, X_3) von X_4, usw. stochastisch unabhängig. Nach Korollar 7.3 in Bauer (2002)

folgt daraus, dass die Zufallsvariablen $X_1, ..., X_n$ stochastisch unabhängig sind. Im nächsten Kapitel werden Voraussetzungen genannt, unter denen aus $\xi(X) = 0$ ebenfalls die stochastische Unabhängigkeit der Komponenten $X_1, ..., X_n$ folgt.
Weitere Möglichkeiten zur Definition von Abhängigkeitsmaßen ließen sich durch die Betrachtung von Copulae gewinnen. Sei X ein Zufallsvektor mit Werten in $\mathbb{R}^u$ und Verteilungsfunktion F. Hinsichtlich Sklar's Theorem (vgl. Nelsen (2006), Theorem 2.10.9) kann, anschaulich gesprochen, die Verteilungsfunktion zerlegt werden in ihre Randverteilungen und eine Copulafunktion, welche die Abhängigkeitsstrukturen der Komponenten von X untereinander widerspiegelt. Die analytische Handhabung von Copulae scheint allerdings schwierig zu sein. Diese Möglichkeit wird deshalb nicht weiter verfolgt.

Kapitel 4

Qualitative Robustheit bei Abhängigkeit

In diesem Abschnitt soll untersucht werden, wie sich geringe Abweichungen einer Zufallsstichprobe vom Unabhängigkeitsfall auf die Verteilung von Schätzern auswirkt. Ein quantitatives Maß für die Abweichung von Verteilungen ist durch die Prohorovmetrik gegeben. Zu jeder Teilmenge $B \subset \mathbb{R}^u$ und jedem $\epsilon > 0$ sei

$$B^\epsilon := \left\{ x \in \mathbb{R}^u,\ \inf_{y \in B} d(x,y) < \epsilon \right\},$$

wobei d die euklidische Metrik bezeichnet.

Definition 4.1 *Seien F, G Verteilungen auf $\mathbb{R}^u$, dann ist durch*

$$d_P(F,G) := \inf\{\epsilon > 0,\ F(B) \le G(B^\epsilon) + \epsilon,\ B \in \mathcal{B}(\mathbb{R}^u)\} \tag{4.1}$$

eine Metrik auf der Klasse aller Verteilungen auf $\mathbb{R}^u$ gegeben.

Diese Metrik induziert die schwache Konvergenz auf der Klasse der Verteilungen auf $\mathbb{R}^u$ (vgl. Huber (2004), Theorem 3.8). Sei $(X_k)_{k\in\mathbb{N}}$ eine Folge von Zufallsvektoren $X_1 = (X_{11}, ..., X_{1n}),\ X_2 = (X_{21}, ..., X_{2n}), X_3 = ...$ Sei $a = a(X_k)$ ein Abhängigkeitsmaß für die Komponenten der Zufallsvektoren X_k, und $(X_k)_{k\in\mathbb{N}}$ erfülle $a(X_k) \longrightarrow 0$ für $k \longrightarrow \infty$. Ferner sei $T : \mathcal{P} \longrightarrow \mathbb{R}^v$ ein Funktional, und F_n^k sei die empirische Verteilungsfunktion der Komponenten des Zufallsvektors X_k, wobei mit $\mathcal{P}$ eine Klasse von Verteilungen bzw. Verteilungsfunktionen bezeichnet wird. Es stellt sich die Frage, ob sich eine Statistik $T(F_n^k)$ für große Werte k im Verhalten kaum noch von der gleichen Statistik im Unabhängigkeitsfall unterscheidet, genauer ob

$$T(F_n^k) \rightsquigarrow T(F_n') \text{ für } k \longrightarrow \infty \tag{4.2}$$

gilt. Dabei sei F_n' die empirische Verteilungsfunktion eines Zufallsvektors $X' = (X_1', ..., X_n')$ mit u.i.v Komponenten X_i', so dass X_i' und X_{ki} die gleiche Verteilung F_i für jedes $k \in \mathbb{N}$ haben. F_n sei entsprechend die empirische Verteilungsfunktion eines Zufallsvektors X mit Dimension n. D. h. wenn das Funktional T als Abbildung

$$T : \mathcal{P} \longrightarrow \mathcal{P}', \; P^{(X_1,...,X_n)} \mapsto P^{T(F_n)}$$

betrachtet wird, so ist es in Bezug auf das Abhängigkeitsmaß a im Sinne von (4.2) gewissermaßen stetig an Stellen $P^{(X_1',...,X_n')}$. Dabei ist $\mathcal{P}$ der Raum der Verteilungen auf $\mathbb{R}^n$ und $\mathcal{P}'$ der Raum der Verteilungen auf $\mathbb{R}^v$. Anschaulich bedeutet dies, dass kleine Abweichungen von der Unabhängigkeitsforderung nur kleine Auswirkungen auf die Verteilung des Schätzers T haben können.
Wenn das Funktional T als Abbildung $T : \mathcal{X}^n \longrightarrow \mathbb{R}^v$, wobei $\mathcal{X}^n$ der Stichprobenraum ist, fast überall stetig ist (fast überall bzgl. der Verteilung von $(X_1', ..., X_n')$), dann kann der Satz von der stetigen Abbildung verwendet werden (vgl. van der Vaart (1998), Lemma 2.3), und (4.2) folgt aus

$$(X_{k1}, ..., X_{kn}) \rightsquigarrow (X_1', ..., X_n') \text{ für } k \longrightarrow \infty. \tag{4.3}$$

Beispiel 4.2 *Der Einstichproben-t-Test ist bezüglich einer Verletzung der Unabhängigkeitsforderung gemessen durch das Abhängigkeitsmaß a robust, weil das Funktional T der Teststatistik $T[X_1, ..., X_n] := n^{1/2}\frac{\overline{X}-\mu}{S}$, mit $S := (\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2)^{1/2}$, als Funktion von $\mathbb{R}^n$ nach $\mathbb{R}$ (bis auf einer Nullmenge) stetig ist, insofern die Verteilungen der Zufallsvariablen X_i nicht alle die gesamte Masse auf den gleichen Wert legen.*

Es soll daher das Ziel sein, unter gewissen Voraussetzungen die Gültigkeit von (4.3) zu zeigen, bzw. zu zeigen, dass der Prohorovabstand zwischen der gemeinsamen Verteilung von Zufallsvariablen $X_1, ..., X_n$ und ihrer Produktverteilung beliebig klein wird, wenn ein bestimmtes Abhängigkeitsmaß hinreichend klein wird. Zunächst wird ein einfacher Fall betrachtet. Wenn $\alpha(X_k) \longrightarrow 0$ für $k \longrightarrow \infty$, dann folgt aus der Definition für das Abhängigkeitsmaß α in (3.20) insbesondere, dass

$$\limsup_{k \longrightarrow \infty} P^{X_k}(B) = P^{X'}(B)$$

für jede abgeschlossene Menge B. Nach dem Portmanteau-Lemma (vgl. van der Vaart (1998), Lemma 2.2) ist dies äquivalent zu (4.3). Nun soll der etwas kompliziertere Fall des Abhängigkeitsmaßes α' aus der Definition 3.21 betrachtet werden. Wie oben sei X_k eine Folge von Zufallsvektoren, so dass die Komponenten X_{ki} eine von k unabhängige Verteilungsfunktion F_i haben.

Theorem 4.3 *Wenn $\alpha'(X_k) \longrightarrow 0$ für $k \longrightarrow \infty$, dann folgt*

$$(X_{k1}, ..., X_{kn}) \rightsquigarrow (X_1^{'}, ..., X_n^{'}) \text{ für } k \longrightarrow \infty. \tag{4.4}$$

Beweis. Die Schwierigkeit besteht – im Gegensatz zum vorherigen Fall – nun darin, dass die Definition von α' eine Abschätzung bzgl. eines Mengensystems liefert, welches kleiner ist als die ganze Borel-σ-Algebra. Insbesondere liefert die Voraussetzung von Theorem 4.3 für beliebiges $B \in \mathcal{B}(\mathbb{R}^n)$ nicht mehr zwangsläufig

$$P^{(X_{k1},...,X_{ki}),(X_{kj},...,X_{kn})}(B) - P^{(X_{k1},...,X_{ki})} \otimes P^{(X_{kj},...,X_{kn})}(B) \longrightarrow 0$$

wobei mit $P^1 \otimes P^2$ die Produktverteilung von P^1 und P^2 bezeichnet wird. Stattdessen werden nun halbabgeschlossene Intervalle $B_i = (a_i, b_i]$, wobei auch $B_i = \emptyset$ möglich sein soll, und halbabgeschlossene multivariate Intervalle $B_1 \times ... \times B_n$ betrachtet. Die Forderung $\alpha'(X_k) \longrightarrow 0$ impliziert, dass für $k \longrightarrow \infty$

$$\begin{aligned}
Q_1 &:= |P^{X_{k1},...,X_{kn}}(B_1 \times ... \times B_n) - P^{X_{k1}}(B_1)P^{X_{k2},...,X_{kn}}(B_2 \times ... \times B_n)| \longrightarrow 0,\\
Q_2 &:= |P^{X_{k2},...,X_{kn}}(B_2 \times ... \times B_n) - P^{X_{k2}}(B_2)P^{X_{k3},...,X_{kn}}(B_3 \times ... \times B_n)| \longrightarrow 0,\\
Q_3 &:= |P^{X_{k3},...,X_{kn}}(B_3 \times ... \times B_n) - P^{X_{k3}}(B_3)P^{X_{k4},...,X_{kn}}(B_4 \times ... \times B_n)| \longrightarrow 0,\\
&\vdots
\end{aligned}$$

Für den ersten Term folgt die Aussage unmittelbar aus den Voraussetzungen. Wird beispielsweise

$$A_1 := \{X_{k1} \in B_1\} \text{ und } A_2 := \{X_{k2} \in B_2, ..., X_{kn} \in B_n\}$$

gesetzt, dann folgt

$$|P(A_1 \cap A_2) - P(A_1)P(A_2)| \leq \alpha'(X_k).$$

Auch für die restlichen Terme gilt die Konvergenz gegen 0, da das Abhängigkeitsmaß $\alpha'(X)$ nur kleiner werden kann, wenn anstelle eines Zufallsvektors X derjenige Zufallsvektor eingesetzt wird, welcher durch Streichung der ersten Komponente in X entsteht.

(Die Konvergenz gegen 0 gilt darüber hinaus noch für weitere Mengen, aber hier genügt die Betrachtung der multivariaten, halbabgeschlossenen Intervalle.) Es wird noch folgende elementare Ungleichung benötigt. Seien $p_1, p_2, p_{1,n}, p_{2,n}, p_{3,n} \in [0,1]$. Dann gilt

$$\begin{aligned}|p_{1,n} - p_1(p_2 p_{3,n})| &= |p_{1,n} - p_1 p_{2,n} + p_1 p_{2,n} - p_1(p_2 p_{3,n})| \\ &\leq |p_{1,n} - p_1 p_{2,n}| + |p_1||p_{2,n} - p_2 p_{3,n}| \\ &\leq |p_{1,n} - p_1 p_{2,n}| + |p_{2,n} - p_2 p_{3,n}|.\end{aligned}$$

Seien nun $p_1, p_2, p_3, p_{1,n}, p_{2,n}, p_{3,n}, p_{4,n} \in [0,1]$. Eine zweifache Anwendung obiger Ungleichung ergibt

$$\begin{aligned}|p_{1,n} - p_1 p_2 p_3 p_{4,n})| &\leq |p_{1,n} - p_1 p_2 p_{3,n}| + |p_{3,n} - p_3 p_{4,n}| \\ &\leq |p_{1,n} - p_1 p_{2,n}| + |p_{2,n} - p_2 p_{3,n}| + |p_{3,n} - p_3 p_{4,n}|.\end{aligned}$$

Entsprechend gilt

$$\begin{aligned}&|P^{X_{k1},\ldots,X_{kn}}(B_1 \times \ldots \times B_n) - P^{X_{k1}}(B_1)P^{X_{k2}}(B_2)P^{X_{k3},\ldots,X_{kn}}(B_3 \times \ldots \times B_n)| \\ &\quad \leq |P^{X_{k1},\ldots,X_{kn}}(B_1 \times \ldots \times B_n) - P^{X_{k1}}(B_1)P^{X_{k2},\ldots,X_{kn}}(B_2 \times \ldots \times B_n)| \\ &\qquad + |P^{X_{k2},\ldots,X_{kn}}(B_2 \times \ldots \times B_n) - P^{X_{k2}}(B_2)P^{X_{k3},\ldots,X_{kn}}(B_3 \times \ldots \times B_n)| \\ &\quad = Q_1 + Q_1,\end{aligned}$$

bzw.

$$\begin{aligned}&|P^{X_{k1},\ldots,X_{kn}}(B_1 \times \ldots \times B_n) - P^{X_{k1}}(B_1)P^{X_{k2}}(B_2)P^{X_{k3}}(B_3)P^{X_{k4},\ldots,X_{kn}}(B_3 \times \ldots \times B_n)| \\ &\quad \leq Q_1 + Q_2 + Q_3.\end{aligned}$$

Eine iterative Fortsetzung dieses Arguments liefert die Abschätzung

$$|P^{X_{k1},\ldots,X_{kn}}(B_1 \times \ldots \times B_n) - P^{X_{k1}}(B_1)P^{X_{k2}}(B_2)\ldots P^{X_{kn}}(B_n)| \leq Q_1 + \ldots + Q_n,$$

und $Q_1 + \ldots + Q_n$ konvergiert für $k \longrightarrow \infty$ gegen 0. Das System aller multivariaten, halbabgeschlossenen Intervalle $B_1 \times \ldots \times B_n$ ist schnittabgeschlossen. Ein Mengensystem, das unter Schnitten abgeschlossen ist, wird auch als π-System bezeichnet. Zudem kann jede offene Menge in $\mathbb{R}^n$ durch eine abzählbare Vereinigung solcher Mengen gebildet werden. Damit folgt nach Theorem 2.2 in Billingsley (1999), dass

$$P^{X_{k1},\ldots,X_{kn}} \rightsquigarrow \bigotimes_{i=1}^{n} P^{X_{ki}},$$

bzw. es folgt (4.4). □

Damit ist z. B. wiederum der Einstichproben-t-Test in Bezug auf kleine Verletzung der Unabhängigkeitsforderung – klein im Sinne des Abhängigkeitsmaßes α' – robust.

Etwas aufwändiger ist die Betrachtung des Abhängigkeitsmaßes $\xi(X)$ aus Definition 3.21. Zunächst soll gezeigt werden, dass zu jeder Menge $T \subset \mathbb{R}^u$, deren Komplement T^c eine Nullmenge ist, eine dichte, abzählbare Teilmenge von T existiert, die gegenüber „rationalen Translationen“ invariant ist.

Definition 4.4 *Mit e_i wird der i-te Einheitsvektor bezeichnet, und*

$$W := \left\{ w : \mathbb{R}^u \longrightarrow \mathbb{R}^u,\ w(t) = t + \sum_{i=1}^{u} q_i e_i,\ q_i \in \mathbb{Q} \right\} \tag{4.5}$$

bildet eine Untergruppe der Gruppe der Translationen. Zu jeder Menge $T \subset \mathbb{R}^u$ wird eine Menge $\tilde{T} \subset T$ durch

$$\tilde{T} := \bigcup_{w \in W} w^{-1} \left(\bigcap_{w \in W} w(T) \right) \tag{4.6}$$

definiert.

Offensichtlich gilt, dass $\bigcap_{w \in W} w(T) \subset T$. Das folgende Lemma zeigt, dass Verschiebungen von $\bigcap_{w \in W} w(T)$ durch $w \in W$ ebenfalls Teilmengen von T sind.

Lemma 4.5 *Sei $T \subset \mathbb{R}^u$ und $\tilde{T}$ wie in Definition 4.4 definiert, dann folgt $\tilde{T} \subset T$.*

Beweis. Es gilt:

1) Zu jedem $t_1 \in \bigcap_{w \in W} w(T)$ und jedem $w \in W$ gibt es ein $t \in T$ mit $w(t) = t_1$.

2) Nach Definition von $\tilde{T}$ gibt es zu jedem $\tilde{t} \in \tilde{T}$ ein $w_0 \in W$, so dass $w_0(\tilde{t}) \in \bigcap_{w \in W} w(T)$.

Sei $\tilde{t} \in \tilde{T}$ beliebig, w_0 wie in 2), und $t_1 = w_0(\tilde{t})$. Nach 1) gibt es zu jedem $w \in W$ ein $t \in T$ mit $w(t) = t_1 = w_0(\tilde{t})$. Wird nun $w = w_0$ gesetzt, dann folgt $w_0(t) = w_0(\tilde{t})$. Dies impliziert $t = \tilde{t}$. Da also zu jedem $\tilde{t} \in \tilde{T}$ ein $t \in T$ gefunden werden kann, so dass

$t = \tilde{t}$, folgt $\tilde{T} \subset T$. □

Genau dann, wenn $T^c \subset \mathbb{R}^u$ eine Nullmenge ist, erfüllt T für jedes multivariate Intervall $[a, b]$

$$\lambda^u(T \cap [a, b]) = \lambda^u([a, b]), \tag{4.7}$$

wobei λ^u das Lebesguemaß bezeichnet. Das folgende Lemma zeigt, dass diese Eigenschaft beim Übergang von T zu $\tilde{T}$ erhalten bleibt.

Lemma 4.6 *Seien T und $\tilde{T}$ wie in Definition 4.4. Wenn T^c eine Nullmenge ist, dann ist $\tilde{T}^c$ auch eine Nullmenge.*

Beweis. Dazu soll $\left(\bigcap_{w \in W} w(T)\right)^c$ betrachtet werden. Wenn T (4.7) erfüllt, dann auch $w(T)$. Nach dem de Morganschen Gesetz gilt

$$\left(\bigcap_{w \in W} w(T)\right)^c = \bigcup_{w \in W} (w(T))^c. \tag{4.8}$$

Für jedes $w \in W$ ist $(w(T))^c$ eine Nullmenge, und jede abzählbare Vereinigung von Nullmengen ist auch eine Nullmenge. Da die Gruppe W nach Konstruktion eine abzählbare Menge ist, ist auch $\bigcup_{w \in W}(w(T))^c$ eine Nullmenge. Wegen (4.8) wird also (4.7) von $\bigcap_{w \in W} w(T)$ erfüllt. Da $\tilde{T}$ per Definition aus der Vereinigung von Mengen besteht, welche (4.7) erfüllen, muss $\tilde{T}$ selbst diese Eigenschaft auch erfüllen. □

Mittels obiger Konstruktion von $\tilde{T}$ kann die Existenz einer weiteren Teilmenge T' von T gezeigt werden, die einige hilfreiche Eigenschaften besitzt. Im weiteren Verlauf kann diese Menge verwendet werden, um unter gewissen Voraussetzungen Verteilungskonvergenz zu beweisen.

Lemma 4.7 *Sei $T \subset \mathbb{R}^u$ und erfülle (4.7). Dann gibt es eine Menge $T' \subset T$ mit folgenden Eigenschaften:*

i) T' ist abzählbar.

ii) T' liegt dicht in $\mathbb{R}^u$.

iii) Zu je zwei Punkten $s, t \in T'$ mit $s \leq t$ liegen auch die Eckpunkte des multivariaten Intervalls $(s, t]$ in T'.

Beweis. Nach Lemma 4.6 erfüllt $\tilde{T}$ (4.7). Damit liegt $\tilde{T}$ auch dicht in $\mathbb{R}^u$. Sei $\tilde{t} \in \tilde{T}$ beliebig, und sei

$$T' := \bigcup_{w \in W} w(\tilde{t}). \tag{4.9}$$

Nach Definition von $\tilde{T}$ gilt $T' \subset \tilde{T}$, und T' liegt offensichtlich dicht in $\mathbb{R}^u$ und ist abzählbar, weil W abzählbar ist. Zu je zwei Punkten $s, t \in T'$ gibt es nach Konstruktion von T' eine Abbildung $w \in W$ mit $w(t) = s$. Nach Definition von W kann diese Abbildung zerlegt werden in $w(t) = w_1 \circ ... \circ w_u(t)$ mit $w_i(t) = t + q_i e_i$. Wenn $s \leq t$ bzw. wenn $(s, t] \neq \emptyset$, kann offensichtlich jeder Eckpunkt v_j von $(s, t]$ dargestellt werden durch $v_j = w_j(t) = w_{i_1} \circ ... \circ w_{i_k}(t)$ mit $\{i_1, ..., i_k\} \subset \{1, ..., u\}$. Da $w_j \in W$ und $t \in T'$, folgt nach Definition von T', dass auch $w_j(t) \in T'$. Also erfüllt T' iii). □

Bemerkung 4.8 *Die Menge der halbabgeschlossenen multivariaten Intervalle, die durch T' erzeugt werden, sei durch*

$$MI(T') := \{(s, t],\ s, t \in T'\}$$

definiert. Nach der Eigenschaft iii) in Lemma 4.7 ist $MI(T')$ stabil unter Schnitten, d. h. wenn $A, B \in MI(T')$, dann ist auch $A \cap B \in MI(T')$. Demnach ist $MI(T')$ ein π-System. Darüber hinaus folgt aus i) und ii) in Lemma 4.7, dass jede offene Menge aus $\mathbb{R}^d$ durch abzählbare Vereinigung von Mengen aus $MI(T')$ gebildet werden kann.

Wie bereits beim Beweis von Theorem 4.3 ist die Konstruktion eines solchen π-Systems hilfreich, wenn Verteilungskonvergenz bewiesen werden soll. Das folgende Lemma soll helfen, aus fast überall punktweiser Konvergenz einer Folge von Verteilungsfunktionen die schwache Konvergenz dieser Verteilungen zu folgern.

Lemma 4.9 *Sei $T \subset \mathbb{R}^u$ und erfülle (4.7). Weiterhin sei P_k eine Folge von Verteilungen und P eine Verteilung, so dass*

$$P_k((-\infty, t]) \longrightarrow P((-\infty, t]) \ \text{ für } k \longrightarrow \infty\ \forall t \in T. \tag{4.10}$$

Dann existiert $T' \subset T$, so dass T' die Eigenschaften aus Lemma 4.7 erfüllt und

$$P_k((s,t]) \longrightarrow P((s,t]) \;\; \text{für } k \longrightarrow \infty \; \forall s \leq t \;,\; s,t \in T'.$$

Beweis. Sei nun T' wie in Lemma 4.7 konstruiert, $s,t \in T'$ und $s \leq t$. Die Eckpunkte des multivariaten Intervalls $(s,t]$ (bis auf den Eckpunkt s) werden mit $v_1, ..., v_{2^u-1}$ bezeichnet. Es gilt

$$(-\infty, t] = (s,t] \,\dot\cup\, \bigcup_{i=1}^{2^u-1} (-\infty, v_i] \text{ und}$$
$$(-\infty, t] \supset \bigcup_{i=1}^{2^u-1} (-\infty, v_i].$$

Mittels der Siebformel von Poincaré und Sylvester (vgl. Billingsley (1995), Chapter 1.2) folgt daraus

$$\begin{aligned} P_k((s,t]) &= P_k((-\infty,t]) - P_k\Big(\bigcup_{i=1}^{2^u-1} (-\infty, v_i]\Big) \\ &= P_k((-\infty,t]) - \Big[\sum_{i_1} P_k((-\infty, v_{i_1}]) - \sum_{i_1<i_2} P_k((-\infty, v_{i_1}] \cap (-\infty, v_{i_2}]) \\ &\quad + ... + \sum_{i_1<...<i_{2^u-1}} (-1)^{2^u-2} P_k((-\infty, v_{i_1}] \cap ... \cap (-\infty, v_{i_{2^u-1}}])\Big]. \end{aligned}$$

In der obigen Gleichung kann P_k auch durch P ersetzt werden. Für die Schnittmengen in den Termen der rechten Seite gilt:

$$(-\infty, v_{j_1}] \cap ... \cap (-\infty, v_{j_k}] = (-\infty, v'] \text{ für ein } v' \in T'.$$

Da $P_k((-\infty,t]) \longrightarrow P((-\infty,t])$ für $k \longrightarrow \infty$ für jedes $t \in T'$, gilt dies auch für jeden Summanden auf der rechten Seite der oberen Gleichung. Da es nur endlich viele Summanden sind, gilt die gewünschte Konvergenz auch für die Summe, also

$$P_k((s,t]) \longrightarrow P((s,t]) \;\; \text{für } k \longrightarrow \infty.$$

□

Bemerkung 4.10 *Wenn die Verteilungen P_k in Lemma 4.9 zufällig sind, und die Konvergenz in (4.10) nur in Wahrscheinlichkeit gilt, dann folgt unmittelbar*

$$P_k((s,t]) \longrightarrow_{st} P((s,t]) \;\; \text{für } k \longrightarrow \infty.$$

Dabei steht „$\longrightarrow_{st}$“ für stochastische Konvergenz bzw. für Konvergenz in Wahrscheinlichkeit.

Lemma 4.11 *Seien X_k und Y_k Folgen von Zufallsvektoren und T' eine Menge mit den Eigenschaften aus Lemma 4.7. Die Verteilungen $P^{X_k} =: P^X$ seien von k unabhängig. Die Verteilungen P^{Y_k} seien von k unabhängig oder konvergieren schwach gegen eine Lebesgue-stetige Verteilung P^Y. Wenn für jede Menge $A = (s,t]$ mit $s,t \in T'$, also $A \in MI(T')$, gilt*

$$E[\mathbb{1}\{X_k \in A\}|\sigma(Y_k)] - E[\mathbb{1}\{X_k \in A\}] = o_P(1) \text{ für } k \longrightarrow \infty,$$

dann gilt auch

$$P^{X_k,Y_k} \rightsquigarrow P^X \otimes P^Y.$$

Beweis. Sei $A_x = A$ und A_y eine Borelmenge, und $\mathring{A}_y$ bezeichne das topologische Innere von A_y. Sei P^Y Lebesgue-stetig und $P^{Y_k} \rightsquigarrow P^Y$. Dann ist $P^Y\left(\overline{A_y \setminus \mathring{A}_y}\right) = 0$, wobei $\overline{A_y \setminus \mathring{A}_y}$ den topologischen Abschluss von $A_y \setminus \mathring{A}_y$ bezeichnet. Nach dem Portmanteau-Lemma (van der Vaart (1998), Lemma 2.2) gilt

$$\lim_{k\longrightarrow\infty} P^{Y_k}(A_y) = P^Y(A_y). \tag{4.11}$$

Falls P^{Y_k} nicht von k abhängt, dann gilt (4.11) trivialerweise auch wenn P^Y nicht Lebesgue-stetig ist. Nach der Definition bedingter Erwartungswerte und nach Voraussetzung gilt:

$$\begin{aligned}\int_{\{Y_k\in A_y\}} \mathbb{1}\{X_k \in A_x\}dP &= \int_{\{Y_k\in A_y\}} E[\mathbb{1}\{X_k \in A_x\}|\sigma(Y_k)]dP \\ &= \int_\Omega \mathbb{1}\{Y_k \in A_y\}(E[\mathbb{1}\{X_k \in A_x\}] + o_p(1))dP \\ &\longrightarrow E[\mathbb{1}\{X_k \in A_x\}]P^Y(A_y).\end{aligned}$$

Dabei folgt die Konvergenz, weil $E[\mathbb{1}\{X_k \in A_x\}|\sigma(Y_k)]$ nicht negativ und beschränkt ist und in Wahrscheinlichkeit gegen eine Konstante $E[\mathbb{1}\{X_k \in A_x\}]$ (konstant, weil P^{X_k} von k unabhängig) konvergiert. Demnach gilt

$$\begin{aligned}\int_\Omega \mathbb{1}\{X_k \in A_x\}\mathbb{1}\{Y_k \in A_y\}dP &= P^{X_k,Y_k}(A_x \times A_y) \\ &\longrightarrow P^X \otimes P^Y(A_x \times A_y).\end{aligned}$$

Alle Mengen der Form $A_x \times A_y$ bilden nach Bemerkung 4.8 ein π-System für den Wertebereich $\mathbb{R}^u$ der Zufallsvariablen (X_k, Y_k). Zudem kann jede offene Menge aus $\mathbb{R}^u$ durch eine abzählbare Vereinigung solcher Mengen gebildet werden. Nach Theorem 2.2 aus Billingsley (1999) folgt damit $P^{X_k,Y_k} \rightsquigarrow P^X \otimes P^Y$. □

Das nächste Theorem zeigt, dass die gemeinsame Verteilung zweier Folgen von Zufallsvariablen schwach gegen die Produktverteilung konvergiert, wenn das Abhängigkeitsmaß aus Definition 3.11 für diese Zufallsvariablen gegen 0 konvergiert. An dieser Stelle wird zunächst die Landau-Notation eingeführt. Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von Zufallsvariablen (oder Zufallsvektoren).

$$\begin{aligned} &X_n = o_P(R_n) \text{ falls } X_n = R_n Y_n \text{ und } \|Y_n\| \longrightarrow_{st} 0, \\ &X_n = O_P(R_n) \text{ falls } X_n = R_n Y_n \text{ und für jedes } \epsilon > 0\ \exists\ M > 0 \text{ mit} \\ &\qquad \lim_{n\longrightarrow\infty} P(\|Y_n\| > M) \leq \epsilon \end{aligned} \tag{4.12}$$

Die erste Definition besagt anschaulich, dass Y_n asymptotisch vernachlässigbar ist, die zweite Definition besagt, dass Y_n stochastisch beschränkt ist. Sei $(x_i)_{i\in\mathbb{N}}$ nun eine (deterministische) Zahlenfolge.

$$\begin{aligned} &x_n = o(r_n) \text{ falls } x_n = r_n y_n \text{ und } \|Y_n\| \longrightarrow 0, \\ &x_n = O(r_n) \text{ falls } x_n = r_n y_n \text{ und } \|Y_n\| \text{ beschränkt.} \end{aligned} \tag{4.13}$$

Theorem 4.12 *Seien X_k, Y_k Folgen von Zufallsvariablen oder Zufallsvektoren wie in Lemma 4.11. Wenn $\xi(\sigma(Y_k), X_k) \longrightarrow 0$ für $k \longrightarrow \infty$, dann folgt $P^{X_k,Y_k} \rightsquigarrow P^X \otimes P^Y$.*

Beweis. Sei $H_k(t)(\omega) := |F_{X_k|\sigma(Y_k)}(t)(\omega) - F_{X_k}(t)|$. Nach Voraussetzung gilt nun $\lim_{k\longrightarrow\infty} E[\|H_k\|_\infty] = 0$ und offensichtlich $\|H_k(\omega)\|_\infty \geq 0\ \forall\omega \in \Omega$. Unter Verwendung der Markov-Ungleichung folgt

$$\|H_k\|_\infty = o_p(1), \text{ für } k \longrightarrow \infty. \tag{4.14}$$

Sei $(\epsilon_l)_{l\in\mathbb{N}}$ eine streng monoton fallende Nullfolge und sei

$$\begin{aligned} T_l &:= \{t \in \mathbb{R}^u,\ P(|H_k(t)| > \epsilon_l) \not\longrightarrow 0\} \text{ für l=1,2,...} \\ T_\infty &:= \{t \in \mathbb{R}^u,\ \exists\epsilon > 0,\ P(|H_k(t)| > \epsilon) \not\longrightarrow 0\}. \end{aligned}$$

Es soll nun gezeigt werden, dass T_∞ eine Lebesgue-Nullmenge ist, also $\lambda^u(T_\infty) = 0$. Sei nun $l \in \mathbb{N}$. Angenommen $\lambda^u(T_l) \neq 0$. Nach Definition 3.10, der Definition von T_l und mit der Annahme $\lambda^u(T_l) \neq 0$ gilt

$$P(\|H_k\|_{\infty,T_l} > \epsilon_l) \not\longrightarrow 0.$$

Wegen $\|H_k(\omega)\|_\infty \geq \|H_k(\omega)\|_{\infty,T_l}$ würde also auch

$$P(\|H_k\|_\infty > \epsilon_l) \not\longrightarrow 0$$

gelten, was ein Widerspruch zu (4.14) ist. Also ist $\lambda^u(T_l) = 0$ für jedes $l \in \mathbb{N}$. Weil $T_\infty = \bigcup_{l \in \mathbb{N}} T_l$ aus einer abzählbaren Vereinigung von Nullmengen besteht, ist T_∞ auch eine Nullmenge. Sei $T := (T_\infty)^c$. Nach Definition von T_∞ gilt demnach

$$|H_k(t)| = o_P(1) \ \forall t \in T,$$

und T erfüllt (4.7). Eine Umformulierung ergibt für $k \longrightarrow \infty$

$$E[\mathbb{1}\{X_k \leq t\}|\sigma(Y_k)] - E[\mathbb{1}\{X_k \leq t\}] = o_P(1) \ \forall t \in T.$$

Nach Lemma 4.9 und Bemerkung 4.10 gibt es eine Menge $T' \subset T$ mit den Eigenschaften aus Lemma 4.7, so dass für $k \longrightarrow \infty$

$$E[\mathbb{1}\{s \leq X_k \leq t\}|\sigma(Y_k)] - E[\mathbb{1}\{s \leq X_k \leq t\}] = o_P(1) \ \forall s, t \in T'.$$

Lemma 4.11 liefert nun $P^{X_k,Y_k} \rightsquigarrow P^X \otimes P^Y$. □

Wie zuvor wäre auch hier Verteilungskonvergenz gegen die Produktverteilung der Einzelkomponenten eines Zufallsvektors wünschenswert. Das nächste Theorem liefert im Zusammenhang mit dem Abhängigkeitsmaß $\xi(X)$ aus Definition 3.21 Bedingungen dafür.

Theorem 4.13 *Sei $X_k = (X_{k1}, ..., X_{ku})$ eine Folge von Zufallsvektoren mit Werten in $\mathbb{R}^p$. Weiterhin seien die Verteilungen $P^{X_{ki}}$ der Komponenten der Zufallsvektoren X_k von k und i unabhängig und Lebesgue-stetig. Ferner gelte*

$$\xi(X_k) \longrightarrow 0 \text{ für } k \longrightarrow \infty. \tag{4.15}$$

Wenn die Folge X_k in Verteilung gegen einen Zufallsvektor $Y = (Y_1, .., Y_u)$ konvergiert, dann gilt

$$P^{X_{k1},...,X_{ku}} \rightsquigarrow P^{X_{k1}} \otimes ... \otimes P^{X_{ku}} = P^{Y_1} \otimes ... \otimes P^{Y_u} = P^{Y_1,...,Y_u}.$$

D. h. Y hat zwangsläufig stochastisch unabhängige Komponenten Y_i.

Beweis. Wegen (4.15) und Theorem 4.12 folgt unmittelbar

$$P^{X_{k1},X_{k2}} \rightsquigarrow \quad P^{X_{k1}} \otimes P^{X_{k2}} = P^{Y_1} \otimes P^{Y_2} = P^{Y_1,Y_2}, \tag{4.16}$$

wobei die Verteilung von Y_1 Lebesgue-stetig und die Verteilung von (Y_1, Y_2) durch $P^{Y_1} \otimes P^{Y_2}$ gegeben ist. Demnach sind Y_1, Y_2 stochastisch unabhängig. Nochmals wegen (4.15), (4.16) und Theorem 4.12 gilt

$$P^{X_{k1},X_{k2},X_{k3}} \rightsquigarrow \quad P^{Y_1,Y_2} \otimes P^{X_{k3}} = P^{Y_1,Y_2} \otimes P^{Y_3} = P^{Y_1,Y_2,Y_3},$$

wobei die Verteilung von (Y_1, Y_2) wieder Lebesgue-stetig und (Y_1, Y_2) von Y_3 stochastisch unabhängig ist. Die iterative Anwendung dieses Arguments ergibt

$$\begin{aligned}
P^{X_{k1},X_{k2}} &\rightsquigarrow P^{X_{k1}} \otimes P^{X_{k2}} = P^{Y_1} \otimes P^{Y_2} = P^{Y_1,Y_2}, \\
P^{X_{k1},X_{k2},X_{k3}} &\rightsquigarrow P^{Y_1,Y_2} \otimes P^{X_{k3}} = P^{Y_1,Y_2} \otimes P^{Y_3} = P^{Y_1,Y_2,Y_3}, \\
&\vdots \\
P^{X_{k1},...,X_{ku}} &\rightsquigarrow P^{Y_1,...,Y_{u-1}} \otimes P^{X_{ku}} = P^{Y_1,...,Y_{u-1}} \otimes P^{Y_u} = P^{Y_1,...,Y_u}.
\end{aligned}$$

Nach Korollar 7.3 in Bauer (2002) ist dies hinreichend dafür, dass

$$P^{Y_1,...,Y_u} = \quad P^{Y_1} \otimes ... \otimes P^{Y_u}.$$

□

Sei $X = (X_1, ..., X_u)$ ein Zufallsvektor mit Werten in $\mathbb{R}^u$ und $\xi(X) = 0$. Wird $(X_{k1}, ..., X_{ku}) := (X_1, ..., X_u)$ gesetzt, so folgt wegen Theorem 4.12 und der gleichen Argumentation wie im Beweis von Theorem 4.13

$$P^X = P^{X_1} \otimes ... \otimes P^{X_u}.$$

Der Zufallsvektor X hat also stochastisch unabhängige Komponenten.

Kapitel 5

Konsistenz bei Abhängigkeit

Ziel dieses Abschnitts ist es, das Konvergenzverhalten von Punktschätzern bei Verletzung von u.i.v.−Forderungen zu untersuchen. Diese Punktschätzer sollen als Funktionale $T(F_n)$ von empirischen Verteilungen betrachtet werden. Die Funktionale T sind von der Form $T : \mathcal{P} \longrightarrow \mathbb{R}^v$, und $\mathcal{P}$ bezeichnet eine Menge von Verteilungen, die die diskreten Verteilungen enthält. Die Konsistenz solcher Schätzer resultiert häufig aus Stetigkeits- oder besser Glattheitseigenschaften des Funktionals T und der Konsistenz der empirischen Verteilungsfunktion F_n für die „wahre“ Verteilungsfunktion F. Unter der Bedingung, dass $X'_1, X'_2, ...$ eine Folge von unabhängigen, reellwertigen Zufallsvariablen mit Verteilung F ist, gilt nach dem Satz von Glivenko-Cantelli

$$\|F'_n - F\|_\infty \longrightarrow 0 \text{ für } n \longrightarrow \infty \text{ fast sicher.} \tag{5.1}$$

Dabei ist F'_n die empirische Verteilungsfunktion, resultierend aus Messungen $x'_1, ..., x'_n$ der Zufallsvariablen $X'_1, ..., X'_n$. Dieses Resultat geht auf die Arbeiten von Glivenko (1933) und Cantelli (1933) zurück. Für große Stichprobenumfänge n ist F'_n dann „nahe“ an F, und unter hinreichenden Glattheitseigenschaften des Funktionals T ist $T(F'_n)$ damit „nahe“ an $T(F)$.

5.1 Glivenko-Cantelli-Sätze

Um die Robustheit solcher Schätzer gegenüber Verletzung der stochastischen Unabhängigkeit in der u.i.v.-Forderung der Messungen zu untersuchen, scheint es unausweichlich, Sätze vom Typ Glivenko-Cantelli unter gewissen Verletzungen der u.i.v.-Forderungen zu beweisen oder zu widerlegen. Dafür werden in dieser Arbeit stets Wahrscheinlichkeiten für Folgen von Ereignissen abgeschätzt und auf diese Abschätzungen das 0-1 Gesetz von Borel-Cantelli angewendet. Im weiteren Verlauf dieses Abschnitts haben alle Zufallsvariablen in Folgen $X'_1, ..., X'_n, ...$ und $X_1, ..., X_n, ...$ die gleiche Verteilung, in der Folge $X_1, ..., X_n, ...$ sind diese aber nicht mehr zwangsläufig unabhängig. Mit F'_n bzw. F_n wird die empirische Verteilungsfunktion der Zufallsvariablen $(X'_1, ..., X'_n)$ bzw. $(X_1, ..., X_n)$ bezeichnet. Der allen Zufallsgrößen zugrunde liegende Wahrscheinlichkeitsraum sei $(\Omega, \mathcal{A}, P)$. Für die erste Konsistenzaussage wird eine Abschätzung für Wahrscheinlichkeiten der Form $P(\|F_n - F\|_\infty \geq \epsilon)$ benötigt, wie sie das folgende Lemma liefert.

Lemma 5.1 *Sei $X = (X_1, ..., X_n)$ ein Zufallsvektor mit abhängigen Komponenten, so dass alle Zufallsvariablen X_i die gleiche Verteilungsfunktion F haben, wobei F eine Verteilungsfunktion auf $\mathbb{R}$ ist. Es gelte $\psi(X) < \infty$ (vgl. Definition 3.18). Dann folgt*

$$P(\|F_n - F\|_\infty \geq \epsilon) \leq (1 + \psi(X))P(\|F'_n - F\|_\infty \geq \epsilon) \quad \forall n \in \mathbb{N} \text{ und } \forall \epsilon > 0. \tag{5.2}$$

Beweis. Sei $X' = (X'_1, ..., X'_n)$ ein Zufallsvektor mit stochastisch unabhängigen Komponenten die ebenfalls die Verteilungsfunktion F besitzen. Mit $P^X, P^{X'}$ werden die Verteilungen von X bzw. X' bezeichnet. Sei $n \in \mathbb{N}$ und $B \in \mathcal{B}(\mathbb{R}^n)$ beliebig mit $P^{X'}(B) \neq 0$. Angenommen $P^X(B) \geq P^{X'}(B)$, dann folgt aus (3.20)

$$P^X(B)/P^{X'}(B) - 1 \leq \psi(X) \tag{5.3}$$

und damit

$$P^X(B) \leq (1 + \psi(X))P^{X'}(B). \tag{5.4}$$

Wird umgekehrt angenommen, dass $P^X(B) < P^{X'}(B)$, dann gilt trivialerweise wegen $\psi(X) \geq 0$ ebenfalls (5.4). Nun werden folgende Borelmengen aus $\mathcal{B}(\mathbb{R}^n)$ betrachtet:

$$\begin{aligned} B_X &:= \{x \in \mathbb{R}^n,\ \|F_n - F\|_\infty \geq \epsilon\} \text{ und} \\ B_{X'} &:= \{x \in \mathbb{R}^n,\ \|F'_n - F\|_\infty \geq \epsilon\}, \end{aligned} \tag{5.5}$$

wobei die empirischen Verteilungsfunktionen, gegeben anhand der Vektoren x bzw. x', mit F_n, F'_n bezeichnet werden. Es gilt

$$P^X(B_X) = P(\|F_n - F\|_\infty \geq \epsilon) \text{ und } P^{X'}(B_{X'}) = P(\|F'_n - F\|_\infty \geq \epsilon). \tag{5.6}$$

Tatsache ist, dass weder B_X noch $B_{X'}$ von den Verteilungen $P^X(B)$ bzw. $P^{X'}(B)$ abhängen, sondern lediglich von ϵ und der Verteilungsfunktion F, d. h. $B_X = B'_X$. Mit (5.4) folgt daher

$$P(\|F_n - F\|_\infty \geq \epsilon) \leq (1 + \psi(X))P(\|F'_n - F\|_\infty \geq \epsilon). \tag{5.7}$$

Aus Gründen der Vollständigkeit muss noch gezeigt werden, dass gilt:

$$P(\|F'_n - F\|_\infty \geq \epsilon) = 0 \quad \Longrightarrow \quad P(\|F_n - F\|_\infty \geq \epsilon) = 0. \tag{5.8}$$

Angenommen, die Verteilung der Zufallsvariablen X'_i hat einen Träger mit Lebesguemaß größer als 0, und angenommen, $P^{X'}(B_1) = 0$, aber $P^X(B_1) = \delta_1 > 0$ für ein $B_1 \in \mathcal{B}(\mathbb{R}^n)$. Dann gibt es einen hinreichend kleinen Wert $\delta_2 > 0$, so dass $\frac{\delta_1}{\delta_2} > \psi(X)+1$, und es gibt ein $B_2 \in \mathcal{B}(\mathbb{R}^n)$ mit $B_2 \cap B_1 = \emptyset$ und $P^{X'}(B_1 \cup B_2) = \delta_2$. Daraus folgt auch, dass $P^X(B_1 \cup B_2) \geq \delta_1$ bzw.

$$\frac{P^X(B_1 \cup B_2)}{P^{X'}(B_1 \cup B_2)} = \frac{\delta_1}{\delta_2} > \psi(X) + 1. \tag{5.9}$$

Damit steht (5.9) im Widerspruch zur Definition in (3.20), woraus in diesem Fall (5.8) folgt. Angenommen, die Verteilung von X'_i hat einen Träger mit Lebesguemaß 0, und $P^X(c) = \delta > 0$ für ein $c = (c_1, ..., c_n) \in \mathbb{R}^n$. Dann folgt

$$\begin{aligned} P^{X'}(c) = P(X'_1 = c_1, ..., X'_n = c_n) &= P(X'_1 = c_1) \cdots P(X'_n = c_n) \\ &\geq P^X(c) \cdots P^X(c) \\ &= \delta^n > 0, \end{aligned}$$

womit (5.8) bewiesen ist. Also gilt (5.7) auch unter der Voraussetzung, dass $P(\|F'_n - F\|_\infty \geq \epsilon) = 0$, woraus (5.2) folgt. □

Mit Lemma 5.1 kann ein Glivenko-Cantelli-Satz für abhängige Zufallsvariablen gezeigt werden.

Theorem 5.2 *Sei $X = (X_1, ..., X_n)$ wie in Lemma 5.1, und es gelte*

$$\psi(X) \leq p(n) \; \forall n \in \mathbb{N}, \tag{5.10}$$

wobei $p(n)$ ein beliebiges Polynom mit $p(n) \geq 0 \; \forall n \in \mathbb{N}$ ist. Dann folgt

$$\|F_n - F\|_\infty \longrightarrow 0 \text{ für } n \longrightarrow \infty \text{ fast sicher.} \tag{5.11}$$

Beweis. Die Beweisidee unterscheidet sich im Grunde nicht von der des klassischen Glivenko-Cantelli-Satzes. Dafür wird die Dvoretzky-Kiefer-Wolfowitz-Ungleichung (DKW) aus Dvoretzky et al. (1956) verwendet. Die Terme $P(\|F_n - F\|_\infty > \epsilon)$ fallen mit steigendem n sehr schnell , d. h. wegen Lemma 5.1 und der DKW-Ungleichung gilt $\forall n \in \mathbb{N}, \; \forall \epsilon > 0$

$$\begin{aligned} P(\|F_n - F\|_\infty \geq \epsilon) &\leq (1 + p(n)) P(\|F_n' - F\|_\infty \geq \epsilon) \\ &\leq C(1 + p(n)) \exp(-2\epsilon^2 n) \text{ für ein } C \in \mathbb{R}_{\geq 0}. \end{aligned} \tag{5.12}$$

Die Folge $C(1+p(n)) \exp(-2\epsilon^2 n)$ konvergiert mit steigendem n gegen 0, woraus die stochastische Konvergenz von $\|F_n - F\|_\infty$ gegen 0 folgt. Die Terme $C(1+p(n)) \exp(-2\epsilon^2 n)$ summieren sich überdies für jeden Wert $\epsilon > 0$ über $n \in \mathbb{N}$ zu einem endlichen Wert, die Terme $P(\|F_n - F\|_\infty \geq \epsilon)$ demnach ebenfalls. Nach dem 0-1-Gesetz von Borel-Cantelli folgt daraus

$$\|F_n - F\|_\infty \longrightarrow 0 \text{ für } n \longrightarrow \infty \text{ fast sicher.} \tag{5.13}$$

□

Bemerkung 5.3

a) *Theorem 5.2 lässt sich problemlos auf eine Vielzahl weiterer Normen und Pseudonormen verallgemeinern. Wird für Verteilungen F, G auf $\mathbb{R}$ und für eine VC-Klasse $\mathcal{C} \subset \mathcal{B}$ (vgl. Definition 7.1)*

$$\|F - G\|_\mathcal{C} := \sup_{C \in \mathcal{C}} |F(C) - G(C)| \tag{5.14}$$

gesetzt, dann liefert eine kombinatorische Betrachtung im u.i.v.-Fall statt der DKW-Ungleichung die Ungleichung

$$P(\|F_n' - F\|_\mathcal{C} \geq \epsilon) \leq p'(n) \exp(-2\epsilon^2 n) \tag{5.15}$$

für ein Polynom p' (vgl. Pollard (1984), Chapter II). Wird diese Ungleichung wie im Beweis von Theorem 5.2 mit Lemma 5.1 (entsprechend für $\|\cdot\|_{\mathcal{C}}$ abgeändert) kombiniert, so resultiert dies wieder in fast sicherer Konvergenz (unter den Bedingungen von Lemma 5.1), also

$$\|F_n - F\|_{\mathcal{C}} \longrightarrow 0 \text{ für } n \longrightarrow \infty \text{ fast sicher.}$$

Im weiterem Verlauf soll aber weiterhin die Supremumsnorm betrachtet werden.

b) Aus der Ungleichung (5.12) kann direkt

$$\|F_n - F\|_\infty = o_P(n^{-\alpha}) \text{ für jedes } \alpha < 1/2 \tag{5.16}$$

abgeleitet werden, nicht aber

$$\|F_n - F\|_\infty = O_P(n^{-1/2}). \tag{5.17}$$

Die Relation (5.17) gilt, wenn das Polynom p in Theorem 5.2 durch eine nicht negative, beschränkte Funktion ersetzt wird.

c) Wird das Polynom p aus Theorem 5.2 durch

$$a_1 \exp(a_2 n^\alpha) \text{ mit } a_1, a_2 \in \mathbb{R}_{\geq 0},\ \alpha < 1 \tag{5.18}$$

ersetzt, dann gilt weiterhin

$$\|F_n - F\|_\infty \longrightarrow \ 0 \text{ für } n \longrightarrow \infty \text{ fast sicher.} \tag{5.19}$$

Eine Rechnung wie oben liefert die Konvergenzrate

$$\|F_n - F\|_\infty = o_P\left(n^{\frac{1-\beta}{2}}\right) \text{ für jedes } \beta < \alpha. \tag{5.20}$$

d) Der Wert α in c) kann i. A. nicht mehr erhöht werden: Wird $\alpha = 1$ erlaubt, dann erfüllt die Folge von (stark) abhängigen Zufallsvariablen $X_1 = X_2 = ... = X_n$ mit $X_1 \sim Bin(1, 1/2)$ die Ungleichung

$$\psi(X) \leq \frac{\exp(ln(2)\ n)}{2}. \tag{5.21}$$

Es ist aber in diesem Fall offensichtlich, dass $\|F_n - F\|_\infty \not\longrightarrow 0$ für $n \longrightarrow \infty$.

In vielen Situationen erscheint es plausibel, dass die Abhängigkeit von Zufallsgrößen mit zeitlichem Abstand abnimmt. Dazu soll der m-dependence Fall aus Definition 3.12 betrachtet werden. Hierfür wird folgendes Lemma benötigt:

Lemma 5.4 *Seien $x_1, ..., x_n$ reelle Zahlen, $N := \{1, ..., n\}$ und $N_1, ..., N_m$ eine disjunkte Zerlegung von N, d. h.*

$$N = N_1 \,\dot{\cup}\, N_2 \,\dot{\cup}\, ... \,\dot{\cup}\, N_m. \tag{5.22}$$

F_{N_j} bezeichne die empirische Verteilungsfunktion der Punkte $(x_i)_{i \in N_j}$ und F eine Verteilungsfunktion auf $\mathbb{R}$. Dann gilt:

$$\|F_N - F\| \leq \frac{|N_1|}{|N|}\|F_{N_1} - F\| + ... + \frac{|N_m|}{|N|}\|F_{N_m} - F\| \tag{5.23}$$

für jede Norm $\|\cdot\|$ auf dem Raum der Verteilungen auf $\mathbb{R}$, wobei mit $|N|$ die Kardinalität der Menge N bezeichnet wird.

Beweis. Der Beweis wird hier nur exemplarisch für eine disjunkte Zerlegung $N = N_1 \,\dot{\cup}\, N_2$ geführt:

$$\begin{aligned}
\|F_N - F\| &= \left\| \frac{1}{|N_1| + |N_2|}(|N_1|F_{N_1} + |N_2|F_{N_2} - (|N_1| + |N_2|)F) \right\| \\
&= \frac{1}{|N_1| + |N_2|}\||N_1|F_{N_1} - |N_1|F + |N_2|F_{N_2} - |N_2|F\| \\
&\leq \frac{|N_1|}{|N|}\|F_{N_1} - F\| + \frac{|N_2|}{|N|}\|F_{N_2} - F\|.
\end{aligned}$$

$\square$

Theorem 5.5 *Die Folge $(X_i)_{i\in\mathbb{N}}$ sei m-dependent (vgl. Definition 3.12), so dass die Folgeglieder X_i die gleiche Verteilungsfunktion F auf $\mathbb{R}$ besitzen. Dann gilt*

$$\|F_n - F\|_\infty \longrightarrow 0 \text{ für } n \longrightarrow \infty \text{ fast sicher, und} \tag{5.24}$$

$$\|F_n - F\|_\infty = O_P(n^{-1/2}). \tag{5.25}$$

Beweis. Zunächst wird folgende disjunkte Zerlegung von $N := \{1, ..., n\}$ gebildet:

$$\begin{aligned}
&N_1 := \{1, 1+m, 1+2m, ..., 1+k_1 m\} \text{ mit } k_1 := \max\{k,\ 1+km \in N\},\\
&N_2 := \{2, 2+m, 2+2m, ..., 2+k_2 m\} \text{ mit } k_2 := \max\{k,\ 2+km \in N\},\\
&\vdots\\
&N_m := \{m, m+m, m+2m, ..., m+k_m m\} \text{ mit } k_m := \max\{k,\ m+km \in N\}.
\end{aligned}$$

Nach Voraussetzung sind $\{X_i\}_{i\in N_j}$ für jedes $j \in \{1, ..., m\}$ unabhängig identisch verteilte Zufallsvariablen. Ferner wird noch

$$m' := (\max_{j\in\{1,...,m\}} \{|N_j|/|N|\})^{-1} \tag{5.26}$$

definiert. Da $||N_i| - |N_j|| \leq 1\ \forall i, j \in \{1, ..., m\}$, gilt auch $|m - m'| \leq 1$. Eine Anwendung von Lemma 5.4 und der DKW-Ungleichung ergibt

$$\begin{aligned}
P(\|F_N - F\|_\infty \geq \epsilon) &\leq P\left(\frac{|N_1|}{|N|}\|F_{N_1} - F\|_\infty + ... + \frac{|N_m|}{|N|}\|F_{N_m} - F\|_\infty \geq \epsilon\right)\\
&\leq P\left(\frac{1}{m'}\|F_{N_1} - F\|_\infty + ... + \frac{1}{m'}\|F_{N_m} - F\|_\infty \geq \epsilon\right)\\
&\leq P\left(\frac{m}{m'} \max_{j\in\{1,...,m\}} \{\|F_{N_j} - F\|_\infty\} \geq \epsilon\right)\\
&\leq P\left(\frac{m}{m'}\|F_{N_1} - F\|_\infty \geq \epsilon \text{ oder...oder } \frac{m}{m'}\|F_{N_m} - F\|_\infty \geq \epsilon\right)\\
&\leq m \max_{1\leq j\leq m} P\left(\|F_{N_j} - F\|_\infty \geq \frac{m'\epsilon}{m}\right)\\
&\leq m\, C \exp\left(-2\frac{m'^2\epsilon^2}{m^2}[n/m]\right) \text{ für ein } C \in \mathbb{R}_{\geq 0},
\end{aligned}$$

da $|N_j| = [n/m] + 1$ oder $|N_j| = [n/m]$ ist. Dabei wurde die Notation $[t] := \min\{n \in \mathbb{Z},\ n \geq t\}$ verwendet. Diese Ungleichung impliziert (5.25) (dazu kann ϵ durch $\epsilon/n^{1/2}$ ersetzt werden) und mit dem 0-1-Gesetz von Borel-Cantelli auch (5.24). □

Theorem 5.5 lässt sich mit der selben Beweisidee verallgemeinern. Statt m-dependence wird nun α-mixing betrachtet. Um Theorem 5.2 und Theorem 5.5 zu beweisen, wurde der Abhängigkeitsfall im Wesentlichen auf den bekannten Fall der Unabhängigkeit zurückgeführt. Um eine Verallgemeinerung von Theorem 5.5 zu erhalten, wird für diese Strategie folgendes Lemma benötigt.

Lemma 5.6 *Sei $X = (X_1, ..., X_n)$ wobei X_i reellwertig sind und die gleiche Verteilungsfunktion F haben. Seien $n, m \in \mathbb{N}$ mit $m < n$, und N_j für $j = 1, ..., m$ wie im Beweis von Theorem 5.5. Ferner sei $X^0 := (X_i)_{i \in N_j}$ für ein j, und $X^{0'}$ eine u.i.v.-Kopie der Zufallsvektors X^0, und $F_{X^{0'}}, F_{X^0}, ...$ bezeichnen die empirischen Verteilungsfunktion der Zufallsvektoren $X^{0'}, X^0, ...$ Dann gilt*

$$P(\|F_{X^0} - F\| \geq \epsilon) \leq P(\|F_{X^{0'}} - F\| \geq \epsilon) + [n/m]\ \alpha_X(m) \tag{5.27}$$

falls X^0 die Länge $[n/m] + 1$ hat, und

$$P(\|F_{X^0} - F\| \geq \epsilon) \leq P(\|F_{X^{0'}} - F\| \geq \epsilon) + ([n/m] - 1)\ \alpha_X(m) \tag{5.28}$$

falls X^0 die Länge $[n/m]$ hat.

Beweis. Nach der Definition der Partition der Indexmenge $N = \{1, ..., n\}$ in die Teilmengen $N_j,\ j = 1, ..., m$ haben die Zufallsvektoren $(X_i)_{i \in N_j}$ entweder die Länge $[n/m] + 1$ oder $[n/m]$. Sei $(X_i)_{i \in N_j}$ von der Länge $[n/m] + 1$. O.B.d.A sei

$$(X_i)_{i \in N_j} := (X_1, X_{1+m}, X_{1+2m}, ..., X_{1+[\frac{n}{m}]m}).$$

Zunächst werden folgende Zufallsvektoren definiert:

$$\begin{aligned}
X^0 &:= (X_1, X_{1+m}, X_{1+2m}, ..., X_{1+[\frac{n}{m}]m}),\\
X^1 &:= (X'_1, X_{1+m}, X_{1+2m}, ..., X_{1+[\frac{n}{m}]m}),\\
X^2 &:= (X'_1, X'_{1+m}, X_{1+2m}, ..., X_{1+[\frac{n}{m}]m}),\\
&\vdots\\
X^{[\frac{n}{m}]+1} &:= (X'_1, X'_{1+m}, X'_{1+2m}, ..., X'_{1+[\frac{n}{m}]m}),
\end{aligned}$$

d. h. die Komponenten X_i der Zufallsvektoren werden nach und nach durch u.i.v.-Kopien X'_i ersetzt, bis $X^{[\frac{n}{m}]+1}$ nur noch stochastisch unabhängige Komponenten enthält, also die gleiche Verteilung besitzt wie $X^{0'}$. Diese Bedingung wird allerdings bereits von $X^{[\frac{n}{m}]}$ erfüllt. Nun gilt

$$\begin{aligned}P(\|F_{X^0}-F\|_\infty \geq \epsilon) &\leq P(\|F_{X^1}-F\|_\infty \geq \epsilon)+\alpha_X(m)\\ &\leq P(\|F_{X^2}-F\|_\infty \geq \epsilon)+2\alpha_X(m)\\ &\;\vdots\\ &\leq P(\|F_{X^{[\frac{n}{m}]}}-F\|_\infty \geq \epsilon)+[n/m]\alpha_X(m),\end{aligned}$$

wobei die erste Ungleichung direkt aus Definition 3.19 für $\alpha_X(m)$ und der gleichen Überlegung wie beim Beweis von Lemma 5.1 folgt. Dazu muss lediglich $B := \{x \in \mathbb{R}^{[n/m]+1},\ \|F_{[n/m]+1}-F\|_\infty > \epsilon\}$ betrachtet werden. Es gilt

$$|P^{X_1,X_{1+m},\ldots,X_{1+[\frac{n}{m}]m}}(B)-P^{X_1',X_{1+m},\ldots,X_{1+[\frac{n}{m}]m}}(B)| \leq \alpha_X(m).$$

Die restlichen Ungleichungen folgen aus der Definition von $\alpha_X(m)$ und Bemerkung 3.20. Falls nun $(X_i)_{i\in N_j}$ die Länge $[n/m]$ besitzt, dann führt die gleiche Argumentation zu

$$P(\|F_{X^0}-F\|_\infty \geq \epsilon) \leq P(\|F_{X^{[\frac{n}{m}]-1}}-F\|_\infty \geq \epsilon)+([n/m]-1)\alpha_X(m).$$

□

Das nächste Theorem überträgt den Fall von m-dependence, wie er in in Theorem 5.5 betrachtet wurde, auf den Fall von α-mixing.

Theorem 5.7 *Sei X wie in Theorem 5.5 aber nicht zwangsläufig m-dependent, und es gelte*

$$\alpha_X(m) = O(m^{-\beta}) \text{ für ein } \beta > 2. \tag{5.29}$$

Dann folgt (wie in Theorem 5.5)

$$\|F_n - F\|_\infty \longrightarrow 0 \text{ für } n \longrightarrow \infty \text{ fast sicher.} \tag{5.30}$$

Beweis. Mit $F_{X'}, F_X, \ldots$ werden die empirischen Verteilungsfunktionen zu den Zufallsvektoren $(X_1',\ldots,X_n'),(X_1,\ldots,X_n),\ldots$ bezeichnet. Sei $m = m(n)$ eine Folge mit den Eigenschaften

$$\begin{aligned}&m(n) = O(n^\gamma) \text{ mit } \gamma<1, \quad m(n)<n \quad \forall n,\\ &m(n)\in\mathbb{N}\ \forall n, \quad m(n) \text{ monoton steigend.}\end{aligned} \tag{5.31}$$

Die Zufallsvariablen $X_1, ..., X_n$ werden wie im Beweis von Theorem 5.5 aufgeteilt. Dies liefert wie im Beweis von Theorem 5.5 die Ungleichung

$$P(\|F_X - F\|_\infty \geq \epsilon) \leq m \max_{1\leq j\leq m} P\left(\|F_{N_j} - F\|_\infty \geq \frac{m'\epsilon}{m}\right),$$

wobei $m' = m'(n)$ wie in (5.26) definiert ist, und $m'/m \longrightarrow 1$ für $n \longrightarrow \infty$ erfüllt. Lemma 5.6 und die DKW-Ungleichung implizieren nun

$$\begin{aligned} m \max_{1\leq j\leq m} P\left(\|F_{N_j} - F\|_\infty \geq \frac{m'\epsilon}{m}\right) &\leq m\, C \exp\left(-2\frac{m'^2\epsilon^2}{m^2}[n/m]\right) \\ &\quad + m\,\left[\frac{n}{m}\right]\alpha(m, X) \\ &= C \exp\left(-2\frac{m'^2\epsilon^2}{m^2}[n/m] + \ln(m)\right) \\ &\quad + m\,\left[\frac{n}{m}\right]\alpha(m, X). \end{aligned} \tag{5.32}$$

Der erste Summand im letzten Term von (5.32) erfüllt für jedes ϵ die Summierbarkeitsbedingung, wenn $m(n)$ die Eigenschaften aus (5.31) einhält. Für den zweiten Summanden gilt

$$m\,\left[\frac{n}{m}\right]\alpha(m, X) = O(n^{1-\beta\gamma}).$$

Da $\beta > 2$ und $\gamma < 1$ beliebig nahe an 1 gewählt werden kann, wird γ so gewählt, dass $1 - \beta\gamma < -1$. Damit erfüllt der zweite Summand ebenfalls die Summierbarkeitsbedingung, und nach dem 0-1-Gesetz von Borel-Cantelli folgt wieder (5.30). $\square$

Es gibt noch viele weitere Möglichkeiten, das Glivenko-Cantelli Theorem zu verallgemeinern. So wird z. B. in Tucker (1959) ein Ansatz aus der Ergodentheorie verwendet und damit auch ein Glivenko-Cantelli Theorem für abhängige Zufallsvariablen bewiesen. Auch für unabhängige Zufallsvariablen gibt es viele Möglichkeiten der Verallgemeinerung, indem beispielsweise größere Indexmengen verwendet werden, wie auch bereits in Bemerkung 5.3 a) angedeutet wurde.

5.2 Konsistenz von Schätzerfolgen

Ausgehend von der Gültigkeit der fast sicheren, gleichmäßigen Konvergenz der empirischen Verteilungsfunktion F_n gegen F und unter gewissen Verletzungen der u.i.v.-

Voraussetzung, d. h.

$$\|F_n - F\|_\infty \longrightarrow 0 \text{ für } n \longrightarrow \infty \text{ fast sicher,} \tag{5.33}$$

kann das asymptotische Verhalten statistischer Funktionale $T = T[F_n]$ bei Abhängigkeit analysiert werden.

Theorem 5.8 *Sei $T : \mathcal{P} \longrightarrow \mathbb{R}^v$ ein Funktional, wobei $\mathcal{P}$ eine Teilmenge der Verteilungen auf $\mathbb{R}$ bezeichnet. $X = (X_1, ..., X_n)$ erfülle die Voraussetzungen aus Theorem 5.2, $F_X, F_{X'}$ seien wie im Beweis von Theorem 5.7, $X_i \sim F$, und T erfülle die (Lipschitz-)Stetigkeitsbedingung*

$$\|T(G) - T(F)\| \leq L\|G - F\|_\infty \textit{ für ein } L > 0,\ \forall\, G \in \mathcal{P} \tag{5.34}$$

bzgl. einer beliebigen Norm $\|\cdot\|$ auf $\mathbb{R}^v$. Dann gilt

$$P(\|T(F_X) - T(F_{X'}))\| \geq \epsilon) \leq (2 + p(n)) \exp\left(-\frac{\epsilon^2 n}{2L^2}\right), \tag{5.35}$$

$$\textit{falls } \exp\left(-\frac{\epsilon^2 n}{2L^2}\right) \leq 1/2 \text{ , } \textit{und}$$

$$P(\|T(F_X) - T(F))\| \geq \epsilon) \leq (1 + p(n)) \exp\left(-\frac{2\epsilon^2 n}{L^2}\right), \tag{5.36}$$

$$\textit{falls } \exp\left(-\frac{2\epsilon^2 n}{L^2}\right) \leq 1/2.$$

Beweis. Zum Beweis von (5.35) wird folgende Ungleichungskette betrachtet, die sich an einer Stelle auf eine genauere Spezifikation der Konstanten C in der DKW-Ungleichung stützt. Massart (1990) beweist, dass die Konstante $C = 2$ generell gewählt werden kann. Wenn n so groß ist, dass $\exp(-2\epsilon^2 n) \leq 1/2$, kann sogar $C = 1$ gewählt werden. Diese

Wahl der Konstanten kann im Wesentlichen nicht mehr verbessert werden.

$$
\begin{aligned}
&P(\|T(F_X) - T(F_{X'}))\| \geq \epsilon) \\
&\quad = P(\|(T(F_X) - T(F)) - (T(F_{X'}) - T(F))\| \geq \epsilon) \\
&\quad \leq P(\|T(F_X) - T(F)\| + \|T(F_{X'}) - T(F)\| \geq \epsilon) \\
&\quad \leq P(2\max\{\|T(F_X) - T(F)\|, \|T(F_{X'}) - T(F)\|\} \geq \epsilon) \\
&\quad = P(\|T(F_X) - T(F)\| \geq \epsilon/2 \text{ oder } \|T(F_{X'}) - T(F)\| \geq \epsilon/2) \\
&\quad \leq (1 + p(n))P(L\|F_{X'} - F\|_\infty \geq \epsilon/2) + P(L\|F_{X'} - F\|_\infty \geq \epsilon/2) \\
&\quad = (2 + p(n))P(L\|F_{X'} - F\|_\infty \geq \epsilon/2) \\
&\quad \leq (2 + p(n)) \exp\left(-\frac{\epsilon^2 n}{2L^2}\right), \text{ falls } \exp\left(-\frac{\epsilon^2 n}{2L^2}\right) \leq 1/2,
\end{aligned}
$$

wobei die vorletzte Ungleichung aus Lemma 5.1 und (5.34), und die letzte Ungleichung aus Massart (1990), Theorem 1 folgt. Dies impliziert (5.35). In gleicher Weise folgt auch (5.36) wegen

$$
\begin{aligned}
&P(\|T(F_X) - T(F))\| \geq \epsilon) \\
&\quad \leq (1 + p(n))P(L\|F_{X'} - F\|_\infty \geq \epsilon) \\
&\quad \leq (1 + p(n)) \exp\left(-\frac{2\epsilon^2 n}{L^2}\right), \text{ falls } \exp\left(-\frac{2\epsilon^2 n}{L^2}\right) \leq 1/2,
\end{aligned} \tag{5.37}
$$

womit der Beweis abgeschlossen ist. □

Natürlich impliziert Theorem 5.8 wiederum die fast sichere Konvergenz von $T(F_X)$ gegen $T(F)$ (bzgl. einer Norm $\|\cdot\|$ auf $\mathbb{R}^v$), und es können wiederum Konvergenzraten für $\|(T(F_X) - T(F)\|$ ermittelt werden. Zudem liefert (5.37) (nichtasymptotische) Abschätzungen für Wahrscheinlichkeiten vom Typ $P(\|T(F_X) - T(F))\| \geq \epsilon)$, welche also auch für endliche Stichproben X gelten. Leider ist (5.34) eine sehr starke Forderung. So existiert beispielsweise kein Lokationsfunktional, welches diese erfüllt. Allerdings kann ein Satz von der Form Glivenko-Cantelli mittels analytischer Betrachtungen auch auf eine Reihe von Schätzern bzw. Funktionalen aus der robusten Statistik, wie z. B. den Median, angewendet werden. Die Konsistenz des Medians unter der Voraussetzung (5.33) kann direkt gezeigt werden. Diese ergibt sich aber auch aus der Betrachtung der Klasse der M-Schätzer.

Ein weiteres Beispiel liefert der folgende Lokationsschätzer. Für das α-getrimmte Mittel kann unter einigen Voraussetzungen ebenfalls Konsistenz gezeigt werden. Für einen

Wert $0 < \alpha < 1/2$ und eine Verteilungsfunktion F auf $\mathbb{R}$ ist das α-getrimmte Mittel gegeben durch

$$T[F] := \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} F^{-1}(t)dt$$

(vgl. Huber (2004), Example 3.3). Angenommen die Einschränkung

$$F : [F^{-1}(\alpha) - \epsilon, F^{-1}(1-\alpha) + \epsilon] \longrightarrow [\alpha - \delta_1, 1 - \alpha + \delta_2] \tag{5.38}$$

ist stetig und bijektiv bzw. streng monoton steigend für beliebig kleine Werte $\epsilon, \delta_1, \delta_2 > 0$. Für $T[F]$ gilt die Identität

$$T[F] = \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} t \, dF(t).$$

Die Stetigkeit und die Monotonie der eingeschränkten Abbildung in (5.38) erzwingt zusammen mit (5.33), dass für $n \longrightarrow \infty$ gilt:

$$\begin{aligned} &F_n^{-1}(\alpha) \longrightarrow F^{-1}(\alpha) \text{ fast sicher und} \\ &F_n^{-1}(1-\alpha) \longrightarrow F^{-1}(1-\alpha) \text{ fast sicher.} \end{aligned}$$

Wegen (5.33) konvergiert natürlich auch die Einschränkung von F_n auf dem kompakten Intervall $[F^{-1}(\alpha) - \epsilon, F^{-1}(1-\alpha) + \epsilon]$ fast sicher gleichmäßig gegen die eingeschränkte Abbildung in (5.38). Zusammen mit der Beschränktheit der Funktionen F_n, F folgt

$$\begin{aligned} T[F_n] &= \frac{1}{1-2\alpha} \int_{F_n^{-1}(\alpha)}^{F_n^{-1}(1-\alpha)} t \, dF_n(t) \\ &= \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} t \, dF_n(t) + o(1) \text{ fast sicher} \\ &= \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} t \, dF(t) + o(1) \text{ fast sicher} \\ &= T[F] + o(1) \text{ fast sicher,} \end{aligned}$$

und damit die Konsistenz der Schätzerfolge.
Das nächste Beispiel behandelt eine ganze Schätzerklasse aus der robusten Statistik. Eine Klasse von konsistenten Lokationsschätzern liefert die Klasse der M-Schätzer. Sei F eine Verteilungsfunktion auf $\mathbb{R}$ und $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ eine Funktion. Ein Lokations-M-Funktional $T[F]$ ist eine Lösung der Gleichung

$$\int_{\mathbb{R}} \psi(t - T[F])dF(t) = 0. \tag{5.39}$$

Diese ist nicht immer eindeutig. Falls aber die Lösung der Gleichung (5.39) für die Verteilungsfunktion F eindeutig und die Funktion ψ zudem monoton steigend und beschränkt ist, so folgt die schwache Stetigkeit des Funktionals T (vgl. Huber (2004), Theorem 2.6). Unter diesen Voraussetzungen konvergiert also jede Folge von Lösungen $T[F_n]$ gegen $T[F]$, falls $F_n \rightsquigarrow F$. Die Bedingung (5.33) impliziert die schwache Konvergenz $F_n \rightsquigarrow F$ fast sicher, also impliziert (5.33) auch

$$T[F_n] \longrightarrow T[F] \text{ für } n \longrightarrow \infty \text{ fast sicher,}$$

wenn $T[F_n]$ für jedes $n \in \mathbb{N}$ eine Lösung gemäß der Gleichung (5.39) ist. Der Median $Med[F]$ kann beispielsweise als ein Lokations-M-Funktional wie in (5.39) geschrieben werden, wenn für ψ die Funktion $t \mapsto sign(t)$ gesetzt wird, wobei mit $sign(t)$ die Vorzeichenfunktion bezeichnet wird. Da die Vorzeichenfunktion monoton steigend und beschränkt ist, folgt aus (5.33) auch

$$Med[F_n] \longrightarrow Med[F] \text{ für } n \longrightarrow \infty \text{ fast sicher,}$$

falls $Med[F]$ eindeutig ist.
Hinreichende Voraussetzungen für schwache Stetigkeit bei R-Schätzern (vgl. Chapter 3.4 in Huber (2004)) liefert auch Theorem 4.1 in Huber (2004), so dass auch für R-Schätzer Konsistenz unter der Voraussetzung (5.33) gezeigt werden kann.
Das gleiche Argument schlägt fehl, wenn nicht schwach stetige Funktionale wie z. B. der Erwartungswert betrachtet werden. Dennoch kann unter gewissen Voraussetzungen auch asymptotische Konsistenz für das arithmetische Mittel bei stochastische Abhängigkeit gezeigt werden. Das nächste Kapitel liefert u.a. Bedingungen dafür.

Kapitel 6

Asymptotische Verteilung bei Abhängigkeit

In vielen Situationen stellt sich die Frage nach der asymptotischen Verteilung einer Schätzerfolge. Ein Schätzer bzw. eine Schätzung kann beispielsweise durch die Angabe eines asymptotischen Konfidenzbereichs quantitativ bewertet werden. Der klassische Zentrale Grenzwertsatz ermöglicht es, das asymptotische Verteilungsverhalten von Mittelwerten zu bestimmen. Mit der Delta-Methode kann das asymptotische Verhalten vieler weiterer Schätzer ermittelt werden. Andererseits treten viele Schätzer als Funktionale der empirischen Verteilungsfunktion auf. Es ist naheliegend anzunehmen, dass das asymptotische Verhalten solcher Schätzer maßgeblich durch die asymptotische Verteilung der empirischen Verteilungsfunktion mitbestimmt wird. Die asymptotische Verteilung der empirischen Verteilungsfunktion kann oftmals mittels eines Funktionalen Zentralen Grenzwertsatzes durch einen Gaußprozess beschrieben werden. Sowohl der klassische Zentrale Grenzwertsatz wie auch der klassische Funktionale Grenzwertsatz für empirische Verteilungsfunktionen nach Donsker setzen stochastische Unabhängigkeit der Eingangsdaten voraus. Dieses Kapitel beschäftigt sich mit den Auswirkungen stochastischer Abhängigkeit auf Mittelwerte und empirische Prozesse in Bezug auf asymptotisches Verteilungsverhalten.

6.1 Ein Zentraler Grenzwertsatz bei Abhängigkeit

Es sind eine Reihe von Zentralen Grenzwertsätzen bei Folgen von abhängigen Zufallsvariablen $X_1, X_2, ...$ bekannt. Viele solcher Sätze setzen neben Regularitätsbedingungen auch Bedingungen an Summen von Kovarianzen zwischen Zufallsvariablen X_k und Zufallsvariablen $E[X_n|\sigma(X_1)]$ voraus, die Bedingungen an die Abhängigkeitsstruktur der Folge $X_1, X_2, ...$ darstellen (vgl. Hall und Heyde (1980), Theorem 5.2). Das folgende Theorem ist ein häufig verwendetes Mittel, um Verteilungskonvergenz gegen eine Normalverteilung nachzuweisen.

Theorem 6.1 *(Satz von Lindeberg-Feller) Sei $X_{n,i},\ 1 \leq i \leq k(n),\ n \in \mathbb{N},\ k(n) \longrightarrow \infty$ ein Dreiecksschema von zeilenweise stochastisch unabhängigen und reellwertigen Zufallsvariablen, d. h. für jedes $n \in \mathbb{N}$ sind $X_{n,1}, X_{n,2}, ..., X_{n,k(n)}$ stochastisch unabhängig. Ferner sei $E[X_{n,i}] = 0\ \forall n, i$. Wenn für jeden Wert $\epsilon > 0$ und für $n \longrightarrow \infty$*

$$\sum_{i=1}^{k(n)} E[\mathbb{1}\{|X_{n,i}| \geq \epsilon\} X_{n,i}^2] \longrightarrow 0 \text{ und}$$
$$\sum_{i=1}^{k(n)} E[\mathbb{1}\{|X_{n,i}| < \epsilon\} X_{n,i}^2] \longrightarrow 1,$$

dann folgt für $n \longrightarrow \infty$

$$(X_{n,1} + X_{n,2} + ... + X_{n,k(n)}) \rightsquigarrow N(0,1).$$

Beweis. Siehe Gnedenko und Kolmogorov (1954), Chapter 4, Theorem 3. □

Zentrale Grenzwertsätze können auch unter Beschränkung der Abhängigkeiten mittels Abhängigkeitskoeffizienten bzw. Abhängigkeitsmaßen gezeigt werden. Beispiele finden sich in Hall und Heyde (1980), Corollary 5.1, in welchem Abhängigkeit mittels des α-mixing-Koeffizienten beschränkt wird, oder im folgenden Theorem, bei dem die Abhängigkeit durch den maximalen Korrelationskoeffizienten aus Definition 3.15 beschränkt wird.

Theorem 6.2 *Sei $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})),\ i \in \mathbb{Z}$, eine stationäre und ergodische Folge von reellwertigen Zufallsvariablen (vgl. Kallenberg (2002), Chapter 10) mit $E[X_i] = 0\ \forall i \in \mathbb{Z}$. Sei $S_n := \sum_{i=1}^{n} X_i$, und es gelte $\sigma_n^2 := Var[S_n] \longrightarrow \infty$. Für einen Wert $\delta > 0$ sei $E[|X_i|^{2+\delta}] < \infty\ \forall i \in \mathbb{Z}$, und der maximale Korrelationskoeffizient $\rho(n)$ konvergiere für $n \longrightarrow \infty$ gegen 0. Dann folgt*

$$\frac{S_n}{\sigma_n} \rightsquigarrow N(0,1).$$

Beweis. Siehe Hall und Heyde (1980), Theorem 5.6. □

Es wäre wünschenswert, die Beschränkung der Abhängigkeit, wie sie in Theorem 6.2 durch die Bedingung $\rho(n) \longrightarrow 0$ gegeben ist, weiter abzuschwächen, und dennoch asymptotische Normalität der Summen der Zufallsvariablen $X_1, X_2, ...$ zu erhalten. Die Konvergenz $\rho(n) \longrightarrow 0$, wie in Theorem 6.2 gefordert, ist nach Lemma 3.17 eine stärkere Voraussetzung als $\kappa_1(n) \longrightarrow 0$. Das nächste Theorem liefert einen Zentralen Grenzwertsatz unter der schwächeren Voraussetzung, dass $\kappa_1(n) \longrightarrow 0$. Dafür wird allerdings ein Preis in Form einer Bedingung an die Folge $(\sigma_n^{2+\delta})^{-1}E[|S_n|^{2+\delta}]$ für einen Wert $\delta > 0$ gezahlt.

Theorem 6.3 *Es seien folgende Voraussetzungen erfüllt:*

i) $(X_i)_{i\in\mathbb{Z}}$ *mit* $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *ist eine reellwertige, stationäre Folge mit* $E[X_i] = 0$,

ii) $\sigma_n^2/n := E[S_n^2]/n \longrightarrow \sigma^2 \in \mathbb{R}_{>0}$ *mit* $S_n := \sum_{i=1}^n X_i$,

iii) $\kappa_1(n) \longrightarrow 0$,

iv) $\exists \delta, s, t > 0,\ s, t < 1$ *und* $\exists A_n$, *so dass* $E[|S_n|^{2+\delta}] \leq A_n \sigma_n^{2+\delta}$, *und* A_n *erfülle die Bedingungen* $A_n < \infty\ \forall n \in \mathbb{N}$, $A_n^{2/\delta} = o(\kappa_1([n^{1-s}])^{-t})$ *und* $A_n^{2/\delta} = o(n^{s/2})$.

Dann folgt

$$\frac{S_n}{n^{1/2}} \rightsquigarrow N(0, \sigma^2).$$

Darüber hinaus ist die Bedingung $\sum_{k\in\mathbb{Z}} |E[X_1 X_k]| < \infty$ *hinreichend dafür, dass* $E[S_n^2]/n$ *für* $n \longrightarrow \infty$ *konvergiert und der Grenzwert* σ^2 *durch*

$$\sigma^2 = \sum_{k\in\mathbb{Z}} E[X_1 X_k]$$

gegeben ist.

Beweis. Der Beweis ist stark angelehnt an den Beweis des Theorems 5.6 in Hall and Heyde und baut auf der Blocktechnik von Bernstein auf. Diese Beweistechnik ist auf die Arbeit von Bernstein (1927) zurückzuführen. Zunächst sollen für $n \in \mathbb{N}$ einige Zahlenfolgen definiert werden. Sei $w(n) > 0$ eine monotone Nullfolge, die so schnell gegen 0 konvergiert, dass

$$A_n^{2/\delta} w(n) = o(1), \; n^{s/2} w(n) \longrightarrow \infty. \tag{6.1}$$

Zur Existenz einer solchen Folge $w(n)$ wird

$$w'(n) := \left(\frac{1}{n^{s/2} \max\{A_n^{2/\delta}, C\}} \right)^{\frac{1}{2}}$$

für eine beliebige Konstante $C > 0$ betrachtet. Es folgt

$$\begin{aligned} w'(n) &\leq \left(\frac{1}{n^{s/2} C} \right)^{\frac{1}{2}} = o(1) \text{ und} \\ n^{s/2} w'(n) &= \left(\frac{n^{s/2}}{\max\{A_n^{2/\delta}, C\}} \right)^{\frac{1}{2}} \\ &\geq \min \left\{ \left(\frac{n^{s/2}}{C} \right)^{\frac{1}{2}}, \left(\frac{n^{s/2}}{A_n^{2/\delta}} \right)^{\frac{1}{2}} \right\} \longrightarrow \infty. \end{aligned}$$

Darüber hinaus gilt

$$\begin{aligned} A_n^{2/\delta} w'(n) &\leq \frac{n^{s/2} \max(A_n^{2/\delta}, C)}{n^{s/2}} w'(n) \\ &= \frac{\left(n^{s/2} \max(A_n^{2/\delta}, C) \right)^{\frac{1}{2}}}{n^{s/2}} \\ &= \left(\max \left\{ \frac{A_n^{2/\delta}}{n^{s/2}}, \frac{C}{n^{s/2}} \right\} \right)^{\frac{1}{2}} = o(1). \end{aligned}$$

Die Folge $w(n) := \sup_{k \geq n} w'(k)$ erbt die asymptotischen Eigenschaften der Folge $w'(n)$, erfüllt also (6.1), und ist monoton fallend. Seien

$$\begin{aligned} r(n) &:= n^{s/2} \kappa_1([n^{1-s}])^t + n^{s/2} w(n), \\ \lambda(n) &:= \max \left\{ \kappa_1([n^{1-s}])^{1-t}, \frac{1}{r(n)} \right\}, \\ p(n) &:= \max \left\{ \left[\frac{n \kappa_1([n^{1-s}])}{\lambda(n)} \right], \left[\frac{n^{1-\frac{s}{2}}}{\lambda(n)} \right] \right\}, \\ q(n) &:= [n^{1-s}], \\ k(n) &:= \max\{k \in \mathbb{N}, \; k(p(n) + q(n)) \leq n\}. \end{aligned} \tag{6.2}$$

Die Summen $S_n = \sum_{i=1}^{n} X_i$ werden in Blöcke zerlegt, bestehend aus der Summe $U_1 := X_1 + ... + X_{p(n)}$ der ersten $p(n)$ Summanden, der Summe $V_1 := X_{p(n)+1} + ... + X_{p(n)+q(n)}$ der nächsten $q(n)$ Summanden, der Summe der nächsten $p(n)$ Summanden usw. Dies führt zu folgender Darstellung von S_n:

$$\begin{aligned} S_n &= \sum_{i=1}^{p(n)} X_i + \sum_{i=p(n)+1}^{p(n)+q(n)} X_i + \sum_{i=p(n)+q(n)+1}^{2p(n)+q(n)} X_i \\ &\quad + ... + \sum_{i=k(n)p(n)+(k(n)-1)q(n))+1}^{k(n)(p(n)+q(n))} X_i + R_n \\ &=: U_1 + V_1 + U_2 + V_2 + ... + U_k + V_k + R_n, \end{aligned}$$

wobei R_n aus der Summe der restlichen Summanden besteht. Die Beweisstrategie bei der Blocktechnik beruht darauf, dass unter günstigen Voraussetzungen aus der Summe S_n ein Teil der Summe (hier $\sum_{i=i}^{k(n)} V_i + R_n$) weggelassen werden kann, ohne dass sich das asymptotische Verhalten von $\frac{S_n}{n^{1/2}}$ in Bezug auf die Verteilung ändert. Aufgrund der Abhängigkeitsstruktur, welche erzwingt, dass Summen $X_{n+1} + ... + X_{n+m}$ und $X_{n+l+1} + ... + X_{n+l+m}$ mit großem zeitlichen Abstand l näherungsweise unabhängig sind, werden im zweiten Schritt die Summanden U_i durch unabhängige Summanden U_i' approximiert und die Situation damit auf einen klassischen Fall von stochastischer Unabhängigkeit zurückgeführt.

Zunächst wird etwas Vorarbeit geleistet und das asymptotische Verhalten obiger Zahlenfolgen untersucht (vgl. auch Beweistechnik in Herrndorf (1984)). Zwischen zwei Zahlenfolgen a_n, b_n gelte die Relation $a_n \sim b_n :\Leftrightarrow a_n/b_n \longrightarrow 1$ für $n \longrightarrow \infty$. Es wird gezeigt, dass

1) $\frac{q(n)}{p(n)} = o(1)$,

2) $k(n) \sim \frac{n}{p(n)}$,

3) $k(n)\kappa_1(q(n)) = o(1)$,

4) $\frac{p(n)}{n} = o(1)$,

5) $\left(\frac{n}{p(n)}\right)^2 \frac{q(n)}{n} = o(1)$.

Es ist klar, dass $r(n) \longrightarrow \infty$ und damit auch $\lambda(n) \longrightarrow 0$, wobei $\lambda(n) > 0 \ \forall n \in \mathbb{N}$ wegen der Anforderungen an $w(n)$. Offensichtlich gilt $q(n) = [n^{1-s}] \longrightarrow \infty$. Allerdings wächst $p(n)$ mit n wesentlich schneller als $q(n)$. Da $\frac{a}{b \vee c} \leq \frac{a}{c}$ für $a, b, c > 0$, gilt

$$\begin{aligned}\frac{q(n)}{p(n)} &= \frac{\left[n^{1-s}\right]}{\max\left\{\left[\frac{n\kappa_1([n^{1-s}])}{\lambda(n)}\right], \left[\frac{n^{1-\frac{s}{2}}}{\lambda(n))}\right]\right\}} \\ &\leq \frac{\left[n^{1-s}\right]}{\left[\frac{n^{1-\frac{s}{2}}}{\lambda(n))}\right]} \sim n^{-s/2}\lambda(n) = o(1),\end{aligned}$$

womit 1) bewiesen ist. Wegen 1) gilt auch

$$\begin{aligned}k(n) &\sim \frac{n}{p(n)+q(n)} \\ &= \frac{n}{p(n)(1+o(1))} \\ &\sim \frac{n}{p(n)},\end{aligned}$$

und damit folgt 2). Dies hilft, 3) zu zeigen:

$$\begin{aligned}k(n)\kappa_1(q(n)) &\sim \frac{n}{p(n)}\kappa_1(q(n)) \\ &\sim \frac{n}{n\max\left\{\left[\frac{\kappa_1(q(n))}{\lambda(n)}\right], \left[\frac{1}{n^{s/2}\lambda(n))}\right]\right\}}\kappa_1(q(n)) \\ &\leq \frac{1}{\kappa_1(q(n))\lambda(n)^{-1}}\kappa_1(q(n)) = \lambda(n) = o(1).\end{aligned}$$

Eine eventuelle Nulldivision bereitet hier keine Probleme, da $k(n)\kappa_1(q(n)) = 0$, falls $\kappa_1(q(n)) = 0$. Als Nächstes wird 4) behandelt:

$$\begin{aligned}\frac{p(n)}{n} &\sim \max\left\{\left[\frac{\kappa_1(q(n))}{\lambda(n)}\right], \left[\frac{1}{n^{s/2}\lambda(n)}\right]\right\} \\ &\sim \max\left\{\frac{\kappa_1(q(n))}{\max\left\{\kappa_1([n^{1-s}])^{1-t}, \frac{1}{r(n)}\right\}}, \frac{1}{n^{s/2}\max\left\{\kappa_1([n^{1-s}])^{1-t}, \frac{1}{r(n)}\right\}}\right\} \\ &\leq \max\left\{\frac{\kappa_1(q(n))}{\kappa_1([n^{1-s}])^{1-t}}, \frac{1}{n^{s/2}\frac{1}{r(n)}}\right\} \\ &= \max\left\{\kappa_1(q(n))^t, \frac{r(n)}{n^{s/2}}\right\} = \kappa_1(q(n))^t + w(n) = o(1).\end{aligned} \tag{6.3}$$

Auch hier ist eine eventuelle Nulldivision unproblematisch. Ist $\kappa_1(q(n)) = 0$, so ist $\left[\frac{\kappa_1(q(n))}{\lambda(n)}\right] = 0$, und im zweiten Term wird das Maximum von $\left[\frac{1}{n^{s/2}\lambda(n)}\right]$ angenommen.

Aus der Definition von $q(n)$ folgt offensichtlich $\frac{q(n)}{n} \sim n^{-s}$, und damit

$$\begin{aligned}\left(\frac{n}{p(n)}\right)^2 \frac{q(n)}{n} \sim & \left(\frac{n}{\max\left\{\frac{n\kappa_1([n^{1-s}])}{\lambda(n)}, \frac{n^{1-\frac{s}{2}}}{\lambda(n)}\right\}}\right)^2 n^{-s} \\ \leq & \left(\frac{n}{\frac{n^{1-\frac{s}{2}}}{\lambda(n)}}\right)^2 n^{-s} = \lambda(n)^2 = o(1),\end{aligned}$$

woraus nun 5) folgt.
Die asymptotische Vernachlässigbarkeit des additiven Anteils $\frac{1}{n^{1/2}\sigma}\left(\sum_{i=1}^{k(n)} V_i + R_n\right)$ von $\frac{S_n}{n^{1/2}\sigma}$ erfolgt aus einer Betrachtung der zweiten Momente

$$M_n^2 := E\left[\left(\frac{\sum_{i=1}^{k(n)} V_i + R_n}{n^{1/2}\sigma}\right)^2\right].$$

Es soll nun gezeigt werden, dass $M_n^2 = o(1)$. Die Stationarität der Folge $(X_i)_{i\in\mathbb{Z}}$ und die Cauchy-Schwarz-Ungleichung liefern

$$\begin{aligned}M_n^2 &= \frac{1}{n\sigma^2} E\left[\left(\sum_{i=1}^{k(n)} V_i + R_n\right)^2\right] \\ &= \frac{1}{n\sigma^2}\left(\sum_{i=1}^{k(n)} E[V_i^2] + E[R_n^2] + \sum_{i\neq j}^{k(n)} E[V_iV_j] + 2\sum_{i=1}^{k(n)} V_iR_n\right) \\ &\leq \frac{1}{n\sigma^2}\Big(k(n)E[V_1^2] + E[R_n^2] + (k(n)^2 - k(n))(E[V_1^2])^{\frac{1}{2}}(E[V_1^2])^{\frac{1}{2}} + \\ &\quad 2k(n)(E[V_i^2])^{\frac{1}{2}}(E[R_n^2])^{\frac{1}{2}}\Big) \\ &\leq \frac{1}{n\sigma^2}\left(k(n)\sigma_{q(n)}^2 + \sigma_{q'(n)}^2 + k(n)^2\sigma_{q(n)}^2 + 2k(n)\sigma_{q(n)}\sigma_{q'(n)}\right),\end{aligned}$$

wobei $q'(n) := n - k(n)(p(n) + q(n))$ den Summenanteil der Terme R_n repräsentiert. Für den ersten Summanden gilt

$$\begin{aligned}\frac{1}{n\sigma^2}k(n)\sigma_{q(n)}^2 &\sim \frac{1}{n\sigma^2}\frac{n}{p(n)}\sigma_{q(n)}^2 = \frac{1}{\sigma^2}\frac{\sigma_{q(n)}^2}{q(n)}\frac{q(n)}{p(n)} \\ &\sim \frac{1}{n\sigma^2}\sigma^2\ o(1) = o(1),\end{aligned}$$

wegen 2) und 1), Voraussetzung ii) und $q(n) \longrightarrow \infty$. Nach Konstruktion der Folge $k(n)$ gilt $q'(n) \leq p(n) + q(n)$. Wegen 1) gilt $p(n) + q(n) \leq 2p(n)$ für hinreichend große Werte

n. Da $\frac{\sigma_n^2}{n} \longrightarrow \sigma^2$ folgt

$$\sup_{l \le 2p(n)} \frac{\sigma_l^2}{l} = O\left(\frac{\sigma_{2p(n)}^2}{2p(n)}\right)$$

und damit auch

$$\frac{\sigma_{q'(n)}^2}{2p(n)} = O\left(\frac{\sigma_{2p(n)}^2}{2p(n)}\right) = \frac{\sigma_{2p(n)}^2}{2p(n)}\, O\left(1\right).$$

Für den zweiten Summanden in der Abschätzung für M_n^2 gilt wegen 4) und Voraussetzung ii)

$$\begin{aligned}
\frac{1}{n\sigma^2}\sigma_{q'(n)}^2 &= \frac{1}{n\sigma^2}\; 2p(n)\; \frac{\sigma_{q'(n)}^2}{2p(n)} \\
&= \frac{2p(n)}{n\sigma^2}\frac{\sigma_{2p(n)}^2}{2p(n)}\; O\left(1\right) \\
&= o(1)O(1)O(1) = o(1).
\end{aligned}$$

Für den dritten Summanden liefern 2), 5) und die Voraussetzung ii)

$$\begin{aligned}
\frac{1}{n\sigma^2}k(n)^2\sigma_{q(n)}^2 &\sim \frac{1}{n\sigma^2}\left(\frac{n}{p(n)}\right)^2 \sigma_{q(n)}^2 \\
&= \frac{n}{\sigma^2}\; \frac{1}{p(n)^2}\; q(n)\; \frac{\sigma_{q(n)}^2}{q(n)} \\
&\sim \frac{nq(n)}{\sigma^2 p(n)^2}\sigma^2 = \frac{nq(n)}{p(n)^2} \\
&= \frac{n^2}{p(n)^2}\; \frac{q(n)}{n} = o(1).
\end{aligned}$$

Mit der gleichen Argumentation wie eben gilt $\frac{\sigma_{q'(n)}}{(q(n)+p(n))^{1/2}} = \frac{\sigma_{q(n)+p(n)}}{(q(n)+p(n))^{1/2}}\; O(1)$. Für den letzten Summanden liefern 1) und 2)

$$\begin{aligned}
\frac{1}{n\sigma^2}2k(n)\sigma_{q(n)}\sigma_{q'(n)} &= \frac{2k(n)}{n\sigma^2}q(n)^{1/2}(q(n)+p(n))^{1/2}\frac{\sigma_{q(n)}}{(q(n))^{1/2}}\frac{\sigma_{q'(n)}}{(q(n)+p(n))^{1/2}} \\
&\sim \frac{2}{\sigma^2 p(n)}(q(n)(q(n)+p(n)))^{1/2}\; O(1) \\
&= \frac{2}{\sigma^2}\left(\frac{q(n)}{p(n)}\frac{q(n)+p(n)}{p(n)}\right)^{1/2}\; O(1) \\
&= \frac{2}{\sigma^2}\left(o(1)\left(\frac{q(n)}{p(n)}+1\right)\right)^{1/2}\; O(1) \\
&= \frac{2}{\sigma^2}\left(o(1)O(1)\right)^{1/2}\; O(1) = o(1).
\end{aligned}$$

Zusammengefasst ergibt sich $M_n^2 = o(1)$, und mit der Markov-Ungleichung folgt

$$\frac{\sum_{i=1}^{k(n)} V_i + R_n}{n^{1/2}\sigma} = o_P(1).$$

Also ist dieser Summenanteil von $n^{-1/2}\sigma^{-1}S_n$ in Bezug auf die asymptotische Verteilung vernachlässigbar, und es genügt die asymptotische Verteilung von $n^{-1/2}\sigma^{-1}\sum_{i=1}^{k(n)} U_i$ zu ermitteln. Da die Zufallsvariablen $U_i,\ 1 \le i \le k(n)$, einen mit n wachsenden zeitlichen Abstand von $q(n)$ haben, ist in Hinblick auf die Voraussetzung iii) zu hoffen, dass die Zufallsvariablen U_i sich zu stochastisch unabhängigen Zufallsgrößen ähnlich verhalten und durch solche in angemessener Weise approximiert werden können. Diese Überlegung soll über die Betrachtung der charakteristischen Funktion von $n^{-1/2}\sigma^{-1}\sum_{i=1}^{k(n)} U_i$ präzisiert werden. Seien $U_i',\ 1 \le i \le k(n)$, stochastisch unabhängige Zufallsvariablen, so dass U_i und U_i' für $i = 1, ..., k(n)$ die gleichen Verteilungen besitzen. Sei $t \in \mathbb{R}$ und

$$g_t : \mathbb{R} \longrightarrow \mathbb{C}, \quad g_t(x) = \exp(itx) = \cos(tx) + i\sin(tx),$$

wobei $i \in \mathbb{C}$ hier die imaginäre Einheit mit $i^2 = -1$ bezeichnet. Für eine reellwertige Zufallsvariable Z ist die charakteristische Funktion von Z definiert durch

$$\varphi(Z, t) := E[g_t(Z)].$$

Als Nächstes soll gezeigt werden, dass sich die charakteristischen Funktionen von

$$\frac{\sum_{i=1}^{k(n)} U_i}{n^{1/2}\sigma} \quad \text{und} \quad \frac{\sum_{i=1}^{k(n)} U_i'}{n^{1/2}\sigma}$$

asymptotisch nicht unterscheiden. Dazu werden die charakteristischen Funktionen von Summen von abhängigen Zufallsvariablen und unabhängigen Zufallsvariablen über das Abhängigkeitsmaß κ_1 in Relation zueinander gebracht. Wie die folgende Rechnung zeigt, gilt für reellwertige Zufallsvariablen X, Y

$$\zeta(t) := |E[g_t(X)g_t(Y)] - E[g_t(X)]E[g_t(Y)]| \le 2(|t| \vee 1)\kappa_1(\sigma(Y), X). \tag{6.4}$$

Da die Betragsfunktion $|\cdot|$ konvex ist, liefern die Jensen-Ungleichung und die Eigenschaften bedingter Erwartungswerte die Abschätzung

$$\begin{aligned}\zeta(t) &= \left|E\left[E[g_t(X)g_t(Y)|\sigma(Y)]\right] - E[g_t(X)]E[g_t(Y)]\right| \\ &= \left|E\left[g_t(Y)E[g_t(X)|\sigma(Y)]\right] - E[g_t(X)]E[g_t(Y)]\right| \\ &= \left|E\left[g_t(Y)(E[g_t(X)|\sigma(Y)] - E[g_t(X)])\right.\right. \\ &\quad \left.\left. + g_t(Y)E[g_t(X)] - E[g_t(X)]E[g_t(Y)]\right]\right| \\ &= \left|E\left[g_t(Y)(E[g_t(X)|\sigma(Y)] - E[g_t(X)])\right]\right| \\ &\leq E\left[\left|g_t(Y)(E[g_t(X)|\sigma(Y)] - E[g_t(X)])\right|\right].\end{aligned}$$

Da $|g_t(z)| = 1 \;\forall z \in \mathbb{R}$, hat der Term $g_t(Y)$ auf den Erwartungswert in der letzten Zeile keinen Einfluss und kann weggelassen werden, so dass

$$E\left[\left|g_t(Y)(E[g_t(X)|\sigma(Y)] - E[g_t(X)])\right|\right] = E\left[\left|(E[g_t(X)|\sigma(Y)] - E[g_t(X)])\right|\right].$$

Für $z = z_1 + iz_2 \in \mathbb{C}$ gilt $|z| \leq |z_1| + |z_2|$ und damit

$$\begin{aligned}E\left[\left|(E[g_t(X)|\sigma(Y)] - E[g_t(X)])\right|\right] \leq\ & E\left[\left|(E[\cos(tX)|\sigma(Y)] - E[\cos(tX)])\right|\right] \\ &+ E\left[\left|(E[\sin(tX)|\sigma(Y)] - E[\sin(tX)])\right|\right].\end{aligned}$$

Die Summenterme auf der rechten Seite können mittels des Abhängigkeitsmaßes $\kappa_1(\sigma(Y), X)$ abgeschätzt werden durch

$$\begin{aligned}(|t| \vee 1)E\left[\left|(|t| \vee 1)^{-1}(E[\cos(tX)|\sigma(Y)] - E[\cos(tX)])\right|\right] &\leq (|t| \vee 1)\kappa_1(\sigma(Y), X) \\ (|t| \vee 1)E\left[\left|(|t| \vee 1)^{-1}(E[\sin(tX)|\sigma(Y)] - E[\sin(tX)])\right|\right] &\leq (|t| \vee 1)\kappa_1(\sigma(Y), X),\end{aligned}$$

und es folgt

$$|E[g_t(X)g_t(Y)] - E[g_t(X)]E[g_t(Y)]| \leq 2(|t| \vee 1)\kappa_1(\sigma(Y), X).$$

Die gleiche Argumentation kann auch mittels jeder anderen σ-Algebra $\mathcal{M}$ geführt werden (statt $\sigma(Y)$), insofern die Zufallsvariable Y bezüglich $\mathcal{M}$ messbar ist. Sei nun $Y := (n\sigma^2)^{-1/2} \sum_{j=1}^{k(n)-1} U_j$, $Y' := (n\sigma^2)^{-1/2} \sum_{j=1}^{k(n)-1} U_j'$ und $X := (n\sigma^2)^{-1/2} U_{k(n)}$. Die charakteristische Funktion von $(n\sigma^2)^{-1/2} \sum_{j=1}^{k(n)} U_j$ ist gegeben durch

$$\varphi(t, Y + X) = E[g_t(Y)g_t(X)],$$

die charakteristische Funktion von $(n\sigma^2)^{-1/2} \sum_{j=1}^{k(n)} U_j'$ ist wegen der Unabhängigkeit der Zufallsvariablen U_j' hingegen durch

$$\varphi(t, Y' + (n\sigma^2)^{-1/2} U_{k(n)}') = E[g_t(Y')]E[g_t(X)]$$

gegeben. Die Dreiecksungleichung liefert

$$\begin{aligned}|E[g_t(Y)g_t(X)] - E[g_t(Y')]E[g_t(X)]| &\leq |E[g_t(Y)g_t(X)] - E[g_t(Y)]E[g_t(X)]| \\ &\quad + |E[g_t(Y)]E[g_t(X)] - E[g_t(Y')]E[g_t(X)]|.\end{aligned}$$

Für den ersten Summanden liefert die Abschätzung (6.4), die Definition von $\kappa_1(n)$ mit $\tilde{t} := (n\sigma^2)^{-1/2}t$, $\tilde{X} := (n\sigma^2)^{1/2}X$ und $\tilde{Y} := (n\sigma^2)^{1/2}Y$, dass

$$\begin{aligned}|E[g_t(Y)g_t(X)] - E[g_t(Y)]E[g_t(X)]| &= |E[g_{\tilde{t}}(\tilde{Y})g_{\tilde{t}}(\tilde{X})] - E[g_{\tilde{t}}(\tilde{Y})]E[g_{\tilde{t}}(\tilde{X})]| \\ &\leq 2(|\tilde{t}| \vee 1)\kappa_1(q(n)) \\ &= 2(((n\sigma^2)^{-1/2}|t|) \vee 1)\kappa_1(q(n)).\end{aligned}$$

Da $|E[g_t(X)]| \leq 1$, kann der zweite Summand nach oben abgeschätzt werden durch

$$|E[g_t(Y)]E[g_t(X)] - E[g_t(Y')]E[g_t(X)]| \leq |E[g_t(Y)] - E[g_t(Y')]|. \tag{6.5}$$

Eine genaue Betrachtung des rechten Terms in (6.5) zeigt, dass hier die Situation auf die ursprüngliche Situation zurückgeführt wurde, genauer

$$|E[g_t(Y)] - E[g_t(Y')]| = |E[g_t(T)g_t(S)] - E[g_t(T')]E[g_t(S)]|$$

mit $T := (n\sigma^2)^{-1/2}\sum_{j=1}^{k(n)-2} U_j$, $T' := (n\sigma^2)^{-1/2}\sum_{j=1}^{k(n)-2} U'_j$ und $S := U_{k(n)-1}$. Die gleichen Überlegungen wie oben liefern also

$$\begin{aligned}|E[g_t(Y)] - E[g_t(Y')]| &\leq 2(((n\sigma^2)^{-1/2}|t|) \vee 1)\kappa_1(q(n)) + |E[g_t(T)] - E[g_t(T')]| \\ &= 2C_n\kappa_1(q(n)) + |E[g_t(T)] - E[g_t(T')]|,\end{aligned}$$

wobei $C_n := ((n\sigma^2)^{-1/2}|t|) \vee 1$. Aus der iterativen Anwendung dieser Argumentation folgt

$$\begin{aligned}\left|\varphi\left(t, \frac{\sum_{j=1}^{k(n)} U_j}{(n\sigma^2)^{1/2}}\right) - \varphi\left(t, \frac{\sum_{j=1}^{k(n)} U'_j}{(n\sigma^2)^{1/2}}\right)\right| &\leq 2C_n\kappa_1(q(n)) \\ &\quad + ... + 2C_n\kappa_1(q(n)) \\ &= 2C_n\ (k(n)-1)\ \kappa_1(q(n)).\end{aligned} \tag{6.6}$$

Wegen 3) folgt, dass diese Terme für jeden festen Wert $t \in \mathbb{R}$ gegen 0 konvergieren. Nach dem Stetigkeitssatz von Lévy-Cramér (vgl. Billingsley (1995), Theorem 26.3) konvergieren demnach die Folgen $(n\sigma^2)^{-1/2}\sum_{j=1}^{k(n)} U_j$ und $(n\sigma^2)^{-1/2}\sum_{j=1}^{k(n)} U'_j$ von Zufallsvariablen in Verteilung gegen die gleiche Verteilung, vorausgesetzt dass eine dieser

Folgen in Verteilung konvergiert. Mit Hilfe von Theorem 6.1 soll die schwache Konvergenz der Folge $(n\sigma^2)^{-1/2}\sum_{j=1}^{k(n)} U'_j$ gegen eine Standardnormalverteilung gezeigt werden. Nach Theorem 6.1 sind

$$\begin{aligned}
&\sum_{i=1}^{k(n)} E[\mathbb{1}\{|(n\sigma^2)^{-1/2}U'_i| \geq \epsilon\}(n\sigma^2)^{-1}U'^2_i] \longrightarrow 0 \text{ und} \\
&\sum_{i=1}^{k(n)} E[\mathbb{1}\{|(n\sigma^2)^{-1/2}U'_i| < \epsilon\}(n\sigma^2)^{-1}U'^2_i] \longrightarrow 1
\end{aligned}$$

für $n \longrightarrow \infty$ und jeden Wert $\epsilon > 0$ dafür hinreichend. Aufgrund der vorausgesetzten Stationarität der Folge $(X_i)_{i\in\mathbb{Z}}$ und weil U_1 und U'_1 die gleiche Verteilung besitzen ist dies gleichbedeutend mit

$$k(n)E[\mathbb{1}\{|(n\sigma^2)^{-1/2}U_1| \geq \epsilon\}(n\sigma^2)^{-1}U_1^2] \longrightarrow 0 \text{ und} \tag{6.7}$$

$$k(n)E[\mathbb{1}\{|(n\sigma^2)^{-1/2}U_1| < \epsilon\}(n\sigma^2)^{-1}U_1^2] \longrightarrow 1. \tag{6.8}$$

Zunächst soll (6.7) gezeigt werden. Für beliebige Werte $\epsilon, \delta > 0$ gilt

$$\begin{aligned}
E[\mathbb{1}\{|(n\sigma^2)^{-1/2}U_1| \geq \epsilon\}U_1^2] &= E[\mathbb{1}\{|U_1|^\delta \geq (n^{1/2}\sigma\epsilon)^\delta\}U_1^2] \\
&= \frac{1}{(n^{1/2}\sigma\epsilon)^\delta}E[\mathbb{1}\{|U_1|^\delta \geq (n^{1/2}\sigma\epsilon)^\delta\}(n^{1/2}\sigma\epsilon)^\delta U_1^2] \\
&\leq \frac{1}{(n^{1/2}\sigma\epsilon)^\delta}E[\mathbb{1}\{|U_1|^\delta \geq (n^{1/2}\sigma\epsilon)^\delta\}U_1^{2+\delta}] \\
&\leq \frac{1}{(n^{1/2}\sigma\epsilon)^\delta}E[U_1^{2+\delta}].
\end{aligned}$$

Wegen der Voraussetzung iv) und weil nach Definition $U_1 = S_{p(n)}$, gibt es ein $\delta > 0$ mit

$$\begin{aligned}
k(n)E[\mathbb{1}\{|(n\sigma^2)^{-1/2}U_1| \geq \epsilon\}(n\sigma^2)^{-1}U_1^2] &\leq \frac{k(n)}{n\sigma^2}\frac{1}{(n^{1/2}\sigma\epsilon)^\delta}E[S_p^{2+\delta}] \\
&\leq \frac{k(n)}{\epsilon^\delta(n^{1/2}\sigma)^{2+\delta}}A_{p(n)}\sigma_{p(n)}^{2+\delta} \\
&\sim \frac{n/p(n)}{\epsilon^\delta}A_{p(n)}\left(\frac{\sigma_{p(n)}}{\sigma_n}\right)^{2+\delta} \\
&\sim \frac{n/p(n)}{\epsilon^\delta}A_{p(n)}\left(\frac{p(n)}{n}\right)^{1+\delta/2} \\
&= \frac{A_{p(n)}}{\epsilon^\delta}\left(\frac{p(n)}{n}\right)^{\delta/2}
\end{aligned} \tag{6.9}$$

für jeden Wert $\epsilon > 0$. Mittels einer Abänderung der Argumentation bei (6.3) kann $\frac{p(n)}{n}$ abgeschätzt werden durch

$$\frac{p(n)}{n} \leq \kappa_1(q(n))^t + w(n).$$

Sei $A'_n := \max_{k \leq n} A_n$ die kleinste, monoton steigende Folge mit $A'_n \geq A_n \ \forall n \in \mathbb{N}$. Diese Folge erfüllt ebenfalls

$$A_n'^{2/\delta} = o(\kappa_1([n^{1-s}])^{-t}) \text{ und } A_n'^{2/\delta} w(n) = o(1)$$

und liefert

$$\begin{aligned} A_{p(n)}^{2/\delta}\kappa_1(p(n))^t \leq A_{p(n)}'^{2/\delta}\kappa_1(q(n))^t \leq A_n'^{2/\delta}\kappa_1([n^{1-s}])^t = o(1) \text{ und} \\ A_{p(n)}^{2/\delta} w(n) \leq A_{p(n)}'^{2/\delta} w(n) \leq A_n'^{2/\delta} w(n) = o(1). \end{aligned}$$

Demnach folgt

$$\frac{A_{p(n)}}{\epsilon^\delta}\left(\frac{p(n)}{n}\right)^{\frac{\delta}{2}} = o(1),$$

und wegen (6.9) ist damit (6.7) bewiesen. Mittels (6.7) kann nun (6.8) gezeigt werden:

$$\begin{aligned} k(n)E[\mathbb{1}\{|(n\sigma^2)^{-1/2}U_1| < \epsilon\}(n\sigma^2)^{-1}U_1^2] &= k(n)E[(n\sigma^2)^{-1}U_1^2] \\ &\quad - k(n)E[\mathbb{1}\{|(n\sigma^2)^{-1/2}U_1| \geq \epsilon\}(n\sigma^2)^{-1}U_1^2] \\ &= k(n)E[(n\sigma^2)^{-1}U_1^2] + o(1). \end{aligned}$$

Der erste Summand konvergiert gegen 1:

$$\begin{aligned} k(n)E[(n\sigma^2)^{-1}U_1^2] &= \frac{k(n)}{n\sigma^2}E[S_{p(n)}^2] = \frac{k(n)}{n\sigma^2}\sigma_{p(n)}^2 \\ &\sim \frac{n/p(n)}{n\sigma^2}\; p(n)\; \sigma^2 \longrightarrow 1 \end{aligned}$$

für $n \longrightarrow \infty$. Damit ist auch (6.8) bewiesen. Nach Theorem 6.1 folgt $(n\sigma^2)^{-1/2}\sum_{j=1}^{k(n)} U'_j \rightsquigarrow N(0,1)$ und damit auch

$$\frac{S_n}{n^{1/2}} \rightsquigarrow N(0, \sigma^2).$$

Unter der Voraussetzung, dass $\sum_{k\in\mathbb{Z}}|E[X_1X_k]| < \infty$, kann $\sigma^2 = \lim_{n\longrightarrow\infty} n^{-1}E[S_n^2]$ mit dem Satz von der dominierten Konvergenz ermittelt werden. Dazu wird für $E[S_n^2]$

eine andere Darstellung benötigt (vgl. auch Bradley (2007), Proposition 8.3). Wegen der vorausgesetzten Stationarität gilt $E[X_iX_j] = E[X_1X_{j-i+1}]$ und deshalb

$$\begin{aligned}
E[S_n^2] &= \sum_{i,j=1}^{n} E[X_iX_j] = nE[X_1X_1] + 2E[X_1X_2] + ... + 2(n-(n-1))E[X_1X_n] \\
&= nE[X_1X_1] + 2\sum_{k=2}^{n}(n-(k-1))E[X_1X_k] \\
&= nE[X_1X_1] + 2\sum_{k=2}^{\infty} \mathbb{1}\{k \leq n\}(n-(k-1))E[X_1X_k].
\end{aligned}$$

Wegen $\sum_{k\in\mathbb{Z}} |E[X_1X_k]| < \infty$ folgt insbesondere $\sum_{k\geq 2} |E[X_1X_k]| < \infty$ und

$$\begin{aligned}
n^{-1}E[S_n^2] &= E[X_1X_1] + 2\sum_{k=2}^{\infty} \mathbb{1}\{k \leq n\} \left(1 - \frac{k-1}{n}\right) E[X_1X_k] \\
&\leq E[X_1X_1] + 2\sum_{k=2}^{\infty} |E[X_1X_k]| < \infty.
\end{aligned}$$

Nach dem Satz von der dominierten Konvergenz können nun Grenzwertbildung und Integration (bzgl. des Zählmaßes auf $\{2,3,4,...\}$) vertauscht werden, und damit

$$\begin{aligned}
\lim_{n\longrightarrow\infty} n^{-1}E[S_n^2] &= \lim_{n\longrightarrow\infty} \left(E[X_1X_1] + 2\sum_{k=2}^{\infty} \mathbb{1}\{k \leq n\} \left(1 - \frac{k-1}{n}\right) E[X_1X_k] \right) \\
&= E[X_1X_1] + 2\sum_{k=2}^{\infty} \lim_{n\longrightarrow\infty} \mathbb{1}\{k \leq n\} \left(1 - \frac{k-1}{n}\right) E[X_1X_k] \\
&= E[X_1X_1] + 2\sum_{k=2}^{\infty} E[X_1X_k] = \sum_{k\in\mathbb{Z}} E[X_1X_k],
\end{aligned}$$

da die vorausgesetzte Stationarität $E[X_1X_2] = E[X_1X_0]$, $E[X_1X_3] = E[X_1X_{-1}], ..., E[X_1X_{1+l}] = E[X_1X_{1-l}], ...$ impliziert. Also folgt $\sigma^2 = \sum_{k\in\mathbb{Z}} E[X_1X_k]$, was noch zu zeigen war. $\square$

Bei Betrachtung von Bedingung iv) in Theorem 6.3 wird deutlich, dass eine konstante Folge $A_n = A \in \mathbb{R}_{>0}$ mit $E[|S_n|^{2+\delta}] \leq A\sigma_n^{2+\delta}$ die Voraussetzung iv) bereits erfüllt. Eine genauere Analyse des Beweises von Theorem 6.3 zeigt aber auch, dass die Voraussetzung iv) noch ein wenig abgeschwächt werden kann. In Hinblick auf Relation (6.9) und der unmittelbar folgenden Argumentation muss lediglich $A_{p(n)}$, mit $p(n)$ wie in (6.2) definiert, die Voraussetzung iv) erfüllen.

6.2 Ein Funktionaler Zentraler Grenzwertsatz bei Abhängigkeit

In Donsker (1952) wurde erstmals schwache Konvergenz auf gewissen Funktionenräumen definiert. Dabei wurde das asymptotische Verhalten von Partialsummenprozessen (vgl. auch Billingsley (1999), Chapter 2) und das asymptotische Verhalten empirischer Prozesse, die auf der empirischen Verteilungsfunktion basieren, untersucht. Ziel dieses Abschnitts ist es, das Konvergenzverhalten allgemeinerer, abstrakter empirischer Prozesse in Bezug auf schwache Konvergenz zu untersuchen, wenn die Eingangsdaten nicht stochastisch unabhängig sind. Dazu soll eine auf der Bernstein-Ungleichung basierende Beweisidee aus Ossiander (1987) aufgegriffen und auf bestimmte Fälle von Abhängigkeit übertragen werden. In Andrews und Pollard (1994) wird, aufbauend auf einer Maximalungleichung, beispielhaft aufgezeigt, wie die Problematik stochastischer Abhängigkeit in Form von α-mixing behandelt werden kann, wobei sich diese Arbeit auf die Problematik der asymptotischen Straffheit beschränkt. In Doukhan et al. (1995) wird stochastische Abhängigkeit über eine direkte Approximation an stochastisch unabhängige Zufallsvariablen behandelt und anschließend die Strategie aus Ossiander (1987) weiterverfolgt.
Die folgende Definition beschreibt empirische Prozesse von der Form wie sie in diesem Abschnitt untersucht werden sollen.

Definition 6.4 *Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von reellwertigen Zufallsvariablen, und sei $\mathcal{F}$ eine Klasse von reellwertigen Funktionen, so dass*

$$E[|f(X_i)|] \leq \infty \; \forall i \in \mathbb{N} \; \forall f \in \mathcal{F}.$$

Dann heißt

$$G_n : \mathcal{F} \longrightarrow \mathbb{R}, \quad G_n(f) := n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i) - E[f(X_i)]\right), \tag{6.10}$$

empirischer Prozess auf $\mathcal{F}$.

Der empirische Prozess aus (6.10) kann gelegentlich über die Eigenschaft der Linearität in natürlicher Weise auf größeren Klassen als $\mathcal{F}$ definiert werden. Seien $a, b \in \mathbb{R}$ und $f, g \in \mathcal{F}$. Für Funktionen $h = af + bg$ gilt

$$G_n(h) = G_n(af + bg) = aG_n(f) + bG_n(g). \tag{6.11}$$

Diese empirischen Prozesse sind zufällige Funktionale, die einer Funktion f einen reellwertigen Wert zuordnen. Die nächste Definition beschreibt eine wichtige Klasse von Funktionalen.

Definition 6.5 *Sei $\mathcal{F}$ eine Indexmenge, wie z. B. eine Klasse von Funktionen. Mit $l^\infty(\mathcal{F})$ wird die Menge der beschränkten Abbildungen $H : \mathcal{F} \longrightarrow \mathbb{R}$ bezeichnet. Durch*

$$\|H\|_{\mathcal{F}} := \sup\{|H(f)|,\ f \in \mathcal{F}\}$$

wird eine Norm auf $l^\infty(\mathcal{F})$ definiert. Dadurch ist auch eine Borel-σ-Algebra $\mathcal{B}$ auf $l^\infty(\mathcal{F})$ gegeben. Sei H eine bzgl. $\mathcal{B}$ messbare Zufallsgröße auf $l^\infty(\mathcal{F})$. H heißt straff, wenn es zu jedem Wert $\epsilon > 0$ eine kompakte Menge $K \subset l^\infty(\mathcal{F})$ gibt, so dass $P(H \notin K) < \epsilon$ gilt.

Das nächste Theorem gehört zu den wichtigsten Hilfsmitteln, um schwache Konvergenz von Folgen von stochastischen Prozessen zu zeigen.

Theorem 6.6 *Sei $(H_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsgrößen $H_n : (\Omega_n, \mathcal{A}_n, P_n) \longrightarrow (l^\infty(\mathcal{F}), \mathcal{B})$, wobei $\mathcal{F}$ eine Indexmenge und $\mathcal{B}$ die durch die Supremumsnorm $\|H\|_{\mathcal{F}}$ auf $l^\infty(\mathcal{F})$ erzeugte Borel-σ-Algebra ist. $H_n \rightsquigarrow H$ und H straff genau dann, wenn*

i) $(H_n(f_1), ..., H_n(f_u)) \rightsquigarrow (H(f_1), ..., H(f_u))$ für je endlich viele Funktionen $f_1, ..., f_u \in \mathcal{F}$, wobei $(H(f_1), ..., H(f_u))$ eine Zufallsvariable in $\mathbb{R}^u$ ist.

ii) Zu jedem $\epsilon, \eta > 0$ gibt es eine endliche Partition $\mathcal{F}^1, ..., \mathcal{F}^k$ von $\mathcal{F}$, so dass

$$\limsup_{n\longrightarrow\infty} P\left(\sup_i \sup_{f,g\in\mathcal{F}^i} |H_n(f) - H_n(g)| \geq \epsilon\right) \leq \eta.$$

Beweis. Siehe van der Vaart (1998), Theorem 18.14. □

Eine Reihe von Funktionalen Zentralen Grenzwertsätzen werden mittels der Bernstein-Ungleichung bewiesen. Diese gilt für stochastisch unabhängige Zufallsvariablen. Das folgende Theorem liefert eine Ungleichung vom Typ der Bernstein-Ungleichung bei abhängigen Zufallsvariablen.

Theorem 6.7 *Sei* $(Y_i)_{i\in\mathbb{Z}}$ *eine Folge von reellwertigen Zufallsvariablen* $Y_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *mit* $E(Y_i) = 0\ \forall i \in \mathbb{Z}$. *Sei* $\varphi : \mathbb{N}^2 \longrightarrow \mathbb{R}_{>0}$ *eine Funktion der Form* $\varphi(u,v) = 2v$ *oder* $\varphi(u,v) = \alpha(u+v) + (1-\alpha)(uv)$ *für ein* $\alpha \in [0,1]$. *Angenommen, es gibt Konstanten* $K, M, L_1, L_2, \mu, \nu \geq 0$ *und eine nichtnegative, monoton fallende Folge* $(\rho(n))_{n\in\mathbb{N}}$, *so dass* $\forall\ 1 \leq s_1 \leq ... \leq s_u \leq t_1 \leq ... \leq t_v$ *gilt:*

$$\begin{aligned} &|Cov[Y_{s_1}\cdots Y_{s_u}, Y_{t_1}\cdots Y_{t_v}]| \leq K^2 M^{u+v-2}((u+v)!)^\nu \varphi(u,v)\varrho(t_1 - s_1), \\ &\quad \sum_{s\geq 0}(s+1)^k \varrho(s) \leq L_1 L_2^k (k!)^\mu\ \forall k \geq 0, \\ &\quad E|Y_i| \leq (k!)^\nu M^k\ \forall k \geq 0. \end{aligned} \tag{6.12}$$

Dann gilt $\forall t \geq 0$*:*

$$P(S_n \geq t) \leq \exp\left(-\frac{t^2/2}{B_n + A_n^{1/\mu+\nu+2}\ t^{(2\mu+2\nu+3)/(\mu+\nu+2)}}\right),$$

wobei $S_n := Y_1 + \cdots + Y_n$, B_n *eine beliebige Folge mit* $B_n \geq Var(S_n)$ *und*

$$A_n = 2(K \vee M)L_2\left(\frac{2^{4+\mu+\nu}K^2 L_1}{B_n/n} \vee 1\right).$$

Beweis. Siehe Doukhan und Neumann (2007), Theorem 1. □

Die Folge $\varrho(l)$ in (6.12) kann wiederum als quantitatives Maß für Abhängigkeit zwischen der Vergangenheit und der um l Zeitpunkte verschobenen Zukunft betrachtet werden. Die Terme, in die K, M einfließen, sind zur Normierung wegen der Produkte $Y_{s_1}Y_{s_2}\cdots Y_{s_u}$ und $Y_{t_1}\cdots Y_{t_v}$ in den Kovarianzen unvermeidlich.

Recht ähnliche Bedingungen wie in Theorem 6.7 liefern auch einen Zentralen Grenzwertsatz.

Theorem 6.8 *Seien* $(Y_i)_{i\in\mathbb{Z}}$ *wie im obigen Theorem mit* $|Y_i| \leq W\ \forall i \in \mathbb{Z}$ *für eine Konstante* $W > 0$, *so dass*

$$\begin{aligned} &|Cov[Y_{s_1}\cdots Y_{s_u}, Y_{t_1}\cdots Y_{t_v}]| \leq K^2 M^{u+v-2}\varphi(u,v)\varrho(t_1 - s_1) \\ &\qquad \textit{und} \lim_{n\longrightarrow\infty}\frac{\varrho(n)}{n^p} = 0\ \forall p > 0. \end{aligned}$$

Dann gilt

$$\lim_{n\longrightarrow\infty}\frac{\sigma_n^2}{n}=\sum_{k\in\mathbb{Z}}E[Y_1Y_k]=:\sigma^2<\infty \text{ und } \frac{S_n}{n^{1/2}}\rightsquigarrow N(0,\sigma).$$

Dabei sind $S_n := Y_1 + ... + Y_n$ *und* $\sigma_n^2 := Var[S_n]$.

Beweis. Siehe Doukhan und Neumann (2007), Theorem 3 und Remark 5. □

Im weiteren Verlauf wird folgende Notation verwendet.

Definition 6.9 *Sei* $(X_i)_{i\in\mathbb{N}}$ *eine Folge reellwertiger Zufallsvariablen und* $f:\mathbb{R}\longrightarrow\mathbb{R}$ *eine messbare Funktion. Sei*

$$\sigma_n^2(f) := \frac{1}{n}Var[S_n(f)] := \frac{1}{n}Var[f(X_1)+...+f(X_n)],$$
$$\sigma^2(f) := \lim_{n\longrightarrow\infty}\sigma_n^2(f),$$

falls die Varianzen und der Limes existieren.

Die folgende Definition wird es erlauben, die Größe von Teilmengen von Funktionenklassen $\mathcal{F}$ unter Berücksichtigung stochastischer Abhängigkeit quantitativ zu beurteilen.

Definition 6.10 *Sei* $c > 0$ *und* $(X_i)_{i\in\mathbb{N}}$ *eine Folge von Zufallsvariablen* $X_i : (\Omega,\mathcal{A},P)\longrightarrow(\mathbb{R},\mathcal{B}(\mathbb{R}))$. *Durch*

$$p(f) := \max\Big((E[(f(X_1))^2])^{1/2},\ (\sigma^2(f)+c(E[f(X_1)])^2)^{1/2}\Big)$$

wird ein Funktional definiert, das für $c = 1$ *bei stochastischer Unabhängigkeit der Zufallsvariablen* $(X_i)_{i\in\mathbb{N}}$ *mit der Norm* $(E[(f(X_1))^2])^{1/2}$ *übereinstimmt. Seien* $\mathcal{F}$ *und* $\mathcal{F}_0$ *Klassen messbarer Funktionen* $f:\mathbb{R}\longrightarrow\mathbb{R}$ *und* $l\le u\in\mathcal{F}_0$ *mit* $p(u-l)<\epsilon$. *Dann wird die Menge*

$$[l,u] := \{f\in\mathcal{F},\ l\le f\le u\}$$

im Folgenden als ϵ-bracket bezeichnet. Dieses ist offenbar vom Funktional p abhängig. Ist $d : \mathcal{F}_0 \times \mathcal{F}_0 \longrightarrow \mathbb{R}_{\geq 0}$ eine Metrik, wobei $\mathcal{F}_0$ die Funktionenklasse $\mathcal{F}$ enthält, so wird auch die Menge

$$[l, u] := \{f \in \mathcal{F},\ l \leq f \leq u\}$$

als ϵ-bracket bezeichnet, wenn $d(u, l) < \epsilon$. Die Zahlen

$$\begin{aligned} N_{[]}(\epsilon, \mathcal{F}, p) &:= \inf \left\{ N,\ \exists l_1, u_1, ..., l_N, u_N,\ \mathcal{F} \subset \bigcup_{i=1}^{N} [l_i, u_i],\ p(l_i - u_i) < \epsilon \right\}, \\ N_{[]}(\epsilon, \mathcal{F}, d) &:= \inf \left\{ N,\ \exists l_1, u_1, ..., l_N, u_N,\ \mathcal{F} \subset \bigcup_{i=1}^{N} [l_i, u_i],\ d(l_i, u_i) < \epsilon \right\} \end{aligned} \tag{6.13}$$

heißen Überdeckungszahlen mit bracketing (engl. bracketing number), und werden im Folgenden kurz als Überdeckungszahlen bezeichnet. Für $f_0 \in \mathcal{F}_0$ sei

$$B_{f_0,\epsilon} := \{f \in \mathcal{F},\ d(f, f_0) < \epsilon\}. \tag{6.14}$$

Die Zahlen

$$N(\epsilon, \mathcal{F}, d) := \inf \left\{ N,\ \exists f_1, ..., f_N \in \mathcal{F}_0,\ \mathcal{F} \subset \bigcup_{i=1}^{N} B_{f_i,\epsilon} \right\} \tag{6.15}$$

werden im Folgenden als metrische Überdeckungszahlen bezeichnet.

Oftmals wird in Beweisen von Funktionalen Zentralen Grenzwertsätzen die Dreiecksungleichung für Normen verwendet. An einigen Stellen wird anstelle einer Norm ein Funktional von der in Definition 6.10 beschriebenen Form treten. Das folgende Lemma liefert für das Funktional p einen Ersatz für eine Dreiecksungleichung.

Lemma 6.11 *Sei p wie in Definition 6.10 und f, g so, dass $p(f), p(g)$ existieren. Dann gilt*

$$p(f + g) \leq (1 + 2^{1/2})(p(f) + p(g)). \tag{6.16}$$

Beweis. Für f, g gilt wegen der Dreiecksungleichung

$$\begin{aligned}\sigma(f+g) &= \lim_{n\longrightarrow\infty} n^{-1/2}(Var[S_n(f) - S_n(-g)])^{\frac{1}{2}}\\ &\leq \lim_{n\longrightarrow\infty} n^{-1/2}(Var[S_n(f)])^{\frac{1}{2}} + \lim_{n\longrightarrow\infty} n^{-1/2}(Var[S_n(g)])^{\frac{1}{2}}\\ &= \sigma(f) + \sigma(g).\end{aligned}$$

Daraus folgt

$$\sigma^2(f+g) \leq \sigma^2(f) + \sigma^2(g) + 2\sigma(f)\sigma(g).$$

Die Dreiecksungleichung bei der L_2-Norm und die Definition von p liefern

$$(E[(f(X_1) - g(X_1))^2])^{\frac{1}{2}} \leq (E[(f(X_1))^2])^{\frac{1}{2}} + (E[(g(X_1))^2])^{\frac{1}{2}} \leq p(f) + p(g).$$

Für $a, b \geq 0$ gilt außerdem $(a+b)^{1/2} \leq a^{1/2}+b^{1/2}$ und $a \vee b \leq a+b$. Dies kann kombiniert werden zu

$$\begin{aligned}p(f+g) &= \max\{(E[(f(X_1) + g(X_1))^2])^{\frac{1}{2}}, (\sigma^2(f+g) + c(E[(f+g)(X_1)])^2)^{1/2}\}\\ &\leq \max\{p(f) + p(g), (\sigma^2(f) + c(E[f(X_1)])^2 + \sigma^2(g) + c(E[g(X_1)])^2\\ &+ \quad 2\sigma(f)\sigma(g) + 2cE[f(X_1)]E[g(X_1)])^{1/2}\}\\ &\leq \max\{p(f) + p(g), ((p(f))^2 + (p(f))^2\\ &\quad + 2\sigma(f)\sigma(g) + 2cE[f(X_1)]E[g(X_1)])^{1/2}\}\\ &\leq \max\{p(f) + p(g), ((p(f))^2 + (p(f))^2\\ &\quad + 2((\sigma(f))^2 \vee (\sigma(g))^2) + 2c((E[f(X_1)])^2 \vee (E[g(X_1)])^2))^{1/2}\}\\ &\leq \max\{p(f) + p(g), p(f) + p(f)\\ &\quad + (2(\sigma(f))^2 + 2(\sigma(g))^2 + 2c(E[f(X_1)])^2 + 2c(E[g(X_1)])^2)^{1/2}\}\\ &\leq p(f) + p(g) + 2^{1/2}(((p(f))^2 + (p(g))^2)^{1/2}\\ &\leq (1 + 2^{1/2})(p(f) + p(g)).\end{aligned}$$

□

Das folgende Theorem zeigt Bedingungen auf, unter denen die in Definition 6.4 beschrieben empirischen Prozesse G_n schwach gegen einen Gauß-Prozess konvergieren. Die in den Prozess G_n eingehenden Zufallsvariablen $X_1, X_2, ...$ können dabei stochastisch abhängig sein.

Theorem 6.12 *Sei $(X_i)_{i\in\mathbb{Z}}$ eine streng stationäre Folge von reellwertigen Zufallsvariablen $X_i := (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Sei $\varphi : \mathbb{N}^2 \longrightarrow \mathbb{R}_{>0}$ eine Funktion von der Form $\varphi(u,v) = 2v$ oder $\varphi(u,v) = \alpha(u+v) + (1-\alpha)uv$ für ein $\alpha \in [0,1]$. Sei $\mathcal{F}_0$ eine Klasse von messbaren Funktionen $f : \mathbb{R} \longrightarrow \mathbb{R}$. Seien $J, L_1, L_2, \mu > 0$ und $(\varrho(n))_{n\geq 0}$ eine Folge, so dass $\forall s_1 \leq ... \leq s_u \leq t_1 \leq ... \leq t_v$ und für jede Funktion $f \in \mathcal{F}_0$ gilt*

$$\begin{aligned} &|Cov[f(X_{s_1})\cdots f(X_{s_u}), f(X_{t_1})\cdots f(X_{t_v})]| \\ &\qquad \leq \sigma^2(f)\|f\|_\infty^{u+v-2} J^{u+v}\varphi(u,v)\varrho(t_1 - s_u) \text{ und} \\ &\qquad \sum_{s\geq 0}(s+1)^k \varrho(s) \leq L_1 L_2^k (k!)^\mu \ \forall k \geq 0. \end{aligned} \tag{6.17}$$

Überdies erfülle die Funktionenklasse $\mathcal{F}_0$:

i) $\exists\, W > 0$, *so dass* $|f| \leq W \ \forall f \in \mathcal{F}_0$.

ii) *Für* $f_1, ..., f_u \in \mathcal{F}_0$ *und* $t = (t_1, ..., t_u) \in \mathbb{R}^u$ $\exists c \neq 0$ *mit* $c(t_1 f_1 +,,, + t_u f_u) \in \mathcal{F}_0$.

iii) $\exists C > 0$ *mit* $C(f-g) \in \mathcal{F}_0$, *falls* $f, g \in \mathcal{F}_0$.

iv) $f \in \mathcal{F}_0 \Longrightarrow f\mathbb{1}\{A\} \in \mathcal{F}_0$ *für jede Lebesgue-messbare Menge* $A \subset \mathbb{R}$.

v) *Es gelte*

$$\sup_{f\in\mathcal{F}_0} \frac{\sigma^2(f)}{\sigma_n^2(f)} - 1 = o(1) \quad \text{und} \quad \sup_{f\in\mathcal{F}_0} \frac{\sigma(f)}{\|f\|_\infty} < \infty, \ \text{wobei } \frac{0}{0} := 0.$$

vi) *Es existiere eine Konstante $c > 0$, so dass für jedes $f : \mathbb{R} \longrightarrow \mathbb{R}$ mit $f \geq 0$, $\sigma^2(f) \leq \infty$ und jede Indikatorfunktion $\mathbb{1}\{A\}$ gilt:*

$$\sigma^2(f\mathbb{1}\{A\}) + c(E[f(X_1)\mathbb{1}\{A\}(X_1)])^2 \leq \sigma^2(f) + c(E[f(X_1)])^2.$$

Sei $s := \frac{1}{\mu+2}$ und $\mathcal{F} \subset \mathcal{F}_0$, $\mathcal{F} \neq \emptyset$, so dass

$$\Xi(\mathcal{F}, p) := \sum_{q\geq q_0} 2^{-q}\left(\sum_{k=q_0}^{q} 1 \vee \log(N_{[]}(2^{-k}, \mathcal{F}, p))\right)^{\frac{1}{2s}} < \infty \text{ für ein } q_0 \in \mathbb{N}, \tag{6.18}$$

wobei p wie in Definition 6.10 mittels $c > 0$ aus Bedingung vii) definiert sei.
Dann folgt:

a) $G_n \rightsquigarrow G$, *wobei* G *zufälliges Element (mit stetigen Pfaden) in* $l^\infty(\mathcal{F})$ *ist.*

Dabei sei die Metrik d *auf* $\mathcal{F}$ *gegeben durch*

$$d(f_1, f_2) := (Var[G(f_1) - G(f_2)])^{1/2}.$$

b) G *ist straffer, zentrierter Gauß-Prozess auf* $\mathcal{F}$.

c) Die Autokovarianzstruktur von G *ist gegeben durch*

$$Cov[G(f), G(g)] = \frac{1}{2}\left(Var[G(f+g)] - Var[G(f)] - Var[G(g)]\right),$$

wobei für $h \in \mathcal{F}_0$ *gilt:*

$$Var[G(h)] = \sum_{k\in\mathbb{Z}} E[(h(X_1) - E[h(X_1)])(h(X_k) - E[h(X_k)])]$$

bzw.

$$Cov[G(f), G(g)] = \frac{1}{2}\sum_{k\in\mathbb{Z}} Cov[f(X_1), g(X_k)] + Cov[f(X_k), g(X_1)].$$

Beweis. Der Beweis des Theorems 6.12 ist stark angelehnt an die Beweise von Lemma 19.33, Lemma 19.34 und Theorem 19.5 in van der Vaart (1998) und stellt eine Übertragung des in van der Vaart (1998) behandelten Falls bei Unabhängigkeit auf den Abhängigkeitsfall dar. Nach Theorem 6.6 ist zum einen die Verteilungskonvergenz aller Randverteilungen von G_n, also die Verteilungskonvergenz der Zufallsvariablen $G_n(f_1), ..., G_n(f_u)$ mit $f_i \in \mathcal{F}$ und zum anderen die Eigenschaft ii) in Theorem 6.6 zu zeigen. Die Stetigkeit der Pfade von G geht dann unmittelbar aus Lemma 18.15 in van der Vaart (1998) hervor. Als erstes wird die Verteilungskonvergenz der Randverteilungen der Prozesse G_n behandelt. Aus Remark 3 in Doukhan und Neumann (2007) geht die Verteilungskonvergenz der eindimensionalen Randverteilungen hervor, also $G_n(f) \rightsquigarrow N(0, \sigma^2(f))\ \forall f \in \mathcal{F}_0$. Seien nun $f_1, ..., f_u \in \mathcal{F}_0$, und $t_1, ..., t_u \in \mathbb{R}$ beliebig. Dann gibt es nach Voraussetzung ii) ein $c \neq 0$ mit $c(t_1 f_1 + ... + t_u f_u) \in \mathcal{F}_0$, und damit folgt nach Remark 3 aus Doukhan und Neumann (2007) und wegen (6.11)

$$\begin{aligned} cG_n(t_1 f_1 + ... + t_u f_u) &= G_n(c(t_1 f_1 + ... + t_u f_u)) \\ &\rightsquigarrow N(0, \sigma^2(c(t_1 f_1 + ... + t_u f_u))). \end{aligned}$$

Offensichtlich existiert mit $\sigma^2(c(t_1f_1 + ... + t_uf_u))$ auch $\sigma^2((t_1f_1 + ... + t_uf_u)) = c^{-2}\sigma^2(c(t_1f_1 + ... + t_uf_u))$, und daraus folgt wieder mit (6.11)

$$\begin{aligned} G_n(t_1f_1 + ... + t_uf_u) = &\quad t_1G_n(f_1) + ... + t_uG_n(f_u)) \\ \rightsquigarrow &\quad N(0, \sigma^2(t_1f_1 + ... + t_uf_u)) \end{aligned}$$

für beliebige $t_1, ...t_u \in \mathbb{R}$. Nach dem Satz von Cramér und Wold (vgl. Billingsley (1995), Theorem 29.4) folgt, dass $(G_n(f_1), ..., G_n(f_u))$ asymptotisch multivariat normalverteilt ist, also in Verteilung gegen eine normalverteilte Zufallsvariable $(G(f_1), ..., G(f_u))$ konvergiert. Offensichtlich ist der Erwartungswert von $(G(f_1), ..., G(f_u))$ der Nullvektor $\vec{0}$ bzw. 0. Um die Verteilung von $(G(f_1), ..., G(f_u))$ zu identifizieren, muss noch die Kovarianzstruktur ermittelt werden. Seien $f, g \in \mathcal{F}_0$. Die Folgen $G_n(f+g)$ und $G_n(f)+G_n(g)$ besitzen trivialerweise die gleiche asymptotische Verteilung. Demnach besitzen $G(f+g)$ und $G(f) + G(g)$ die gleiche Verteilung. Die Betrachtung von $Var[G(f + g)]$ liefert

$$\begin{aligned} Cov[G(f), G(g)] &= \frac{1}{2}\left(Var[G(f + g)] - Var[G(f)] - Var[G(g)]\right) \\ &= \frac{1}{2}\left(\sigma^2(f + g) - \sigma^2(f) - \sigma^2(g)\right). \end{aligned}$$

Nach Doukhan und Neumann (2007), Remark 3, können für $h \in \mathcal{F}_0$ die rechten Terme beschrieben werden mittels

$$\begin{aligned} Var[G(h)] = \sigma^2(h) &= \sum_{k\in\mathbb{Z}} E[(h(X_1) - E[h(X_1)])(h(X_k) - E[h(X_k)])] \\ &= \sum_{k\in\mathbb{Z}} Cov[h(X_1), h(X_k)], \end{aligned}$$

und damit ist

$$
\begin{aligned}
Cov[G(f), G(g)] &= \frac{1}{2}\Big(Var[G(f+g)] - Var[G(f)] - Var[G(g)]\Big) \\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}} Cov[(f+g)(X_1), (f+g)(X_k)] \\
&\quad - \frac{1}{2}\sum_{k\in\mathbb{Z}} \Big(Cov[f(X_1), f(X_k)] + Cov[g(X_1), g(X_k)]\Big) \\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}} \Big(Cov[f(X_1), g(X_k)] + Cov[g(X_1), f(X_k)]\Big) \\
&\quad + \frac{1}{2}\sum_{k\in\mathbb{Z}} \Big(Cov[f(X_1), f(X_k)] + Cov[g(X_1), g(X_k)]\Big) \\
&\quad - \frac{1}{2}\sum_{k\in\mathbb{Z}} \Big(Cov[f(X_1), f(X_k)] + Cov[g(X_1), g(X_k)]\Big) \\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}} \Big(Cov[f(X_1), g(X_k)] + Cov[g(X_1), f(X_k)]\Big).
\end{aligned}
$$

Es steht noch aus, die Eigenschaft ii) aus Theorem 6.6 nachzuweisen. Dazu soll die exponentielle Ungleichung, die das Theorem 6.7 liefert, auf die Folgen G_n übertragen werden. Wird anstelle einer Funktion f die Funktion $\frac{f}{\|f\|_\infty}$ mit $f \in \mathcal{F}_0$ eingesetzt, dann gilt:

$$
\begin{aligned}
&\left| Cov\left[\frac{f(X_{s_1})}{\|f\|_\infty}\cdots\frac{f(X_{s_u})}{\|f\|_\infty}, \frac{f(X_{t_1})}{\|f\|_\infty}\cdots\frac{f(X_{t_v})}{\|f\|_\infty}\right]\right| \\
&\qquad = \frac{1}{\|f\|_\infty^{u+v}}\left|Cov\left[f(X_{s_1})\cdots f(X_{s_u}), f(X_{t_1})\cdots f(X_{t_v})\right]\right| \qquad (6.19)\\
&\qquad \le \left(\frac{J\sigma(f)}{\|f\|_\infty}\right)^2 J^{u+v-2}\varphi(u,v)\varrho(t_1 - s_1).
\end{aligned}
$$

Mit

$$
\begin{aligned}
D_1 &:= \sup_{f\in\mathcal{F}_0} 2L_2J\left(\frac{\sigma(f)}{\|f\|_\infty}\vee 1\right) \quad \text{und} \\
D_2 &:= 2^{4+\mu}L_1J^2
\end{aligned}
$$

sei nun

$$
\begin{aligned}
A_n &:= 2L_2J\left(\frac{\sigma(f)}{\|f\|_\infty}\vee 1\right)\left(\frac{2^{4+\mu}L_1J^2\frac{\sigma^2(f)}{\|f\|_\infty^2}}{\frac{\sigma_n^2(f)}{\|f\|_\infty^2}}\vee 1\right) \\
&\le D_1\left(D_2\frac{\sigma^2(f)}{\sigma_n^2(f)}\vee 1\right).
\end{aligned}
$$

Wegen der Voraussetzung v) gibt es Zahlen $N_1 \in \mathbb{N}$ und $A < \infty$, so dass $A_n \leq A$ für $n > N_1$ und jedes $f \in \mathcal{F}_0$. Des Weiteren folgt aus v), dass es Zahlen $B > 0, N_2 \in \mathbb{N}$ gibt, so dass

$$\sigma_n^2(f) \leq B\sigma^2(f) \text{ für jedes } f \in \mathcal{F}_0 \text{ und } \forall n > N_2.$$

Dabei sind $\tilde{b}, N_2$ von der Wahl der Funktion f unabhängig. Im Folgenden sei $N := N_1 \vee N_2$. Mittels Theorem 1 aus Doukhan und Neumann (2007) und wegen $\sigma_n^2(cf) = c^2\sigma_n^2(f)$ folgt für jedes $n > N$

$$\begin{aligned}
&P\left(\frac{S_n(f) - E[S_n(f)]}{n^{1/2}} \geq x\right) \\
&\quad = P\left(S_n\left(\frac{f}{\|f\|_\infty}\right) - E\left[S_n\left(\frac{f}{\|f\|_\infty}\right)\right] \geq \frac{x}{\|f\|_\infty n^{-1/2}}\right) \\
&\quad \leq \exp\left(-\frac{\frac{x^2}{2\|f\|_\infty^2 n^{-1}}}{Var\left[S_n\left(\frac{f}{\|f\|_\infty}\right)\right] + A_n^s\left(\frac{x}{\|f\|_\infty n^{-1/2}}\right)^{2-s}}\right) \\
&\quad \leq \exp\left(-\frac{x^2/2}{\frac{Var[S_n(f)]\|f\|_\infty^2 n^{-1}}{\|f\|_\infty^2} + A^s\frac{x^{2-s}\|f\|_\infty^2 n^{-1}}{\|f\|_\infty^{2-s} n^{-(2-s)/2}}}\right) \\
&\quad = \exp\left(-\frac{x^2/2}{\sigma_n^2(f) + A^s\|f\|_\infty^s n^{-s/2} x^{2-s}}\right).
\end{aligned}$$

Insbesondere kann das in den Voraussetzungen zu Theorem 1 aus Doukhan und Neumann (2007) auftretende ν gleich 0 gesetzt werden, weil die Funktionen $f \in \mathcal{F}_0$ wegen Voraussetzung i) betragsmäßig durch eine Konstante W nach oben beschränkt sind.
Da die Voraussetzungen in Theorem 1 aus Doukhan und Neumann (2007) für die Zufallsvariablen $-Y_i$ gleichermaßen gegeben sind, kann die gleiche Argumentation auch für die Funktion $-f$ angewendet werden. Diese Überlegung führt zu

$$\begin{aligned}
P\left(G_n(-f) \geq x\right) &= P\left(-\frac{S_n(f) - E[S_n(f)]}{n^{1/2}} \geq x\right) \\
&\leq \exp\left(-\frac{x^2/2}{\sigma_n^2(f) + A^s\|f\|_\infty^s n^{-s/2} x^{2-s}}\right)
\end{aligned}$$

für $n > N$. Diese Ungleichungen können miteinander kombiniert werden. Die Bonferroni-Ungleichung und die Voraussetzung v) liefern für $n > N$

$$\begin{aligned}
P\left(|G_n(f)| \geq x\right) &\leq P\left(\frac{S_n(f) - E[S_n(f)]}{n^{1/2}} \geq x\right) \\
&\quad + P\left(-\frac{S_n(f) - E[S_n(f)]}{n^{1/2}} \geq x\right) \\
&\leq 2\exp\left(-\frac{x^2/2}{\sigma_n^2(f) + A^s\|f\|_\infty^s n^{-s/2} x^{2-s}}\right) \qquad (6.20)\\
&\leq 2\exp\left(-\frac{x^2/2}{B\sigma^2(f) + A^s\|f\|_\infty^s n^{-s/2} x^{2-s}}\right) \\
&= 2\exp\left(-\frac{x^2}{b_f + a_f n^{-s/2} x^{2-s}}\right),
\end{aligned}$$

wobei $b_f := 2B\sigma^2(f)$ und $a_f := 2A^s\|f\|_\infty^s$. Der rechte Ausdruck enthält noch beide Terme b_f, a_f, was die Handhabung der Ungleichung (6.20) erschwert. Daher wird die Zufallsgröße $G_n(f)$ additiv zerlegt. Eine Zerlegung mittels Indikatorfunktionen

$$\begin{aligned}
A_f &:= G_n(f)\mathbb{1}\left\{|G_n(f)| > \left(\frac{2B\sigma^2(f)}{2A^s\|f\|_\infty^s n^{-s/2}}\right)^{\frac{1}{2-s}}\right\} \\
&= G_n(f)\mathbb{1}\left\{|G_n(f)| > \left(\frac{b_f}{a_f n^{-s/2}}\right)^{\frac{1}{2-s}}\right\}, \\
B_f &:= G_n(f)\mathbb{1}\left\{|G_n(f)| \leq \left(\frac{2B\sigma^2(f)}{2A^s\|f\|_\infty^s n^{-s/2}}\right)^{\frac{1}{2-s}}\right\} \qquad (6.21)\\
&= G_n(f)\mathbb{1}\left\{|G_n(f)| \leq \left(\frac{b_f}{a_f n^{-s/2}}\right)^{\frac{1}{2-s}}\right\}
\end{aligned}$$

liefert $G_n(f) = A_f + B_f$. Für $x^{2-s} > \frac{b_f}{a_f n^{-s/2}}$ folgt mittels (6.20)

$$\begin{aligned}
P(|G_n(f)| > x) &\leq 2\exp\left(-x^s \frac{x^{2-s}}{b_f + a_f n^{-s/2} x^{2-s}}\right) \\
&\leq 2\exp\left(-x^s \frac{\frac{b_f}{a_f n^{-s/2}}}{b_f + a_f n^{-s/2}\frac{b_f}{a_f n^{-s/2}}}\right) \\
&= 2\exp\left(-x^s \frac{1}{a_f n^{-s/2}} \frac{b_f}{b_f + b_f}\right) \\
&= 2\exp\left(-x^s \frac{1}{2}\frac{1}{a_f n^{-s/2}}\right).
\end{aligned}$$

Falls umgekehrt $0 < x^{2-s} \leq \frac{b_f}{a_f n^{-s/2}}$, so folgt

$$\begin{aligned} P(|G_n(f)| > x) &\leq 2\exp\left(-\frac{x^2}{b_f + a_f n^{-s/2} x^{2-s}}\right) \\ &\leq 2\exp\left(-\frac{x^2}{b_f + a_f n^{-s/2}\frac{b_f}{a_f n^{-s/2}}}\right) \\ &= 2\exp\left(-\frac{x^2}{2b_f}\right). \end{aligned}$$

Dies ist gleichbedeutend mit

$$\begin{aligned} &P(|A_f| > x) \leq 2\exp\left(-x^s \frac{1}{2}\frac{1}{a_f n^{-s/2}}\right) \quad \text{und} \\ &P(|B_f| > x) \leq 2\exp\left(\frac{x^2}{2b_f}\right). \end{aligned} \tag{6.22}$$

Ziel ist es nun, mittels obiger Resultate für Erwartungswerte der Form

$$E[\max_{f\in\tilde{\mathcal{F}}} |A_f|] \quad \text{und} \quad E[\max_{f\in\tilde{\mathcal{F}}} |B_f|]$$

für endliche Funktionenklassen $\tilde{\mathcal{F}}$ eine Abschätzung zu finden, die mit der Kardinalität $|\tilde{\mathcal{F}}|$ der Funktionenklasse nicht zu schnell wächst. Für $0 < s < 1$ sei

$$\psi_s : \mathbb{R}_{>0} \longrightarrow \mathbb{R}, \quad \psi_s(x) := \exp(x^s) - 1.$$

Des Weiteren sei $\tilde{\psi}_s$ die kleinste, konvexe Funktion, so dass $\tilde{\psi}_s(x) \geq \psi_s(x)\ \forall x > 0$. Eine analytische Betrachtung zeigt (unter Verwendung von $0 < s < 1$), dass es eine lineare Funktion L und eine nur von s abhängige Konstante $D(s) > 0$ gibt, so dass

$$\tilde{\psi}_s(x) = \begin{cases} L(x) & \text{für} \quad 0 < x \leq D(s) \\ \psi(x) & \text{für} \quad x > D(s) \end{cases} \leq \begin{cases} C(s) & \text{für} \quad 0 < x \leq D(s) \\ \psi(x) & \text{für} \quad x > D(s) \end{cases}$$

für eine nur von s abhängige Konstante $C(s)$. Sei $(a')^s := 4a_f n^{-s/2} > 0$. Wegen $\frac{d}{dx}\psi_s(x) = s\exp(x^s)x^{s-1}$ und mittels der Abschätzung für $P(|A_f| > x)$ in (6.22) und wegen des Satzes von Fubini gibt es eine von der Funktion f unabhängige Konstante

$C_A > 0$ mit

$$
\begin{aligned}
E\left[\tilde{\psi}_s\left(|A_f|/a'\right)\right] &\leq C(s) + E\left[\mathbb{1}\ \{|A_f|/a' > D(s)\}\,\psi_s\left(|A_f|/a'\right)\right] \\
&\leq C(s) + E\left[\psi_s\left(|A_f|/a'\right)\right] \\
&= C(s) + E\left[\int_0^{|A_f|/a'} s\exp(t^s)t^{s-1}dt\right] \\
&= C(s) + sE\left[\int_0^{\infty} \mathbb{1}\{t < |A_f|/a'\}\exp(t^s)t^{s-1}dt\right] \\
&= C(s) + s\int_0^{\infty} P(|A_f| > ta')\exp(t^s)t^{s-1}dt \\
&\leq C(s) + s\int_0^{\infty} 2\exp\left(-\frac{(ta')^s}{2a_f n^{-s/2}}\right)\exp(t^s)t^{s-1}dt \\
&\leq C(s) + 2s\int_0^{\infty} \exp\left(-2t^s\right)\exp(t^s)t^{s-1}dt \\
&\leq C_A < \infty,
\end{aligned}
\tag{6.23}
$$

wobei das letzte Integral wegen $s > 0$ endlich ist. Ist $\tilde{\mathcal{F}} \subset \mathcal{F}_0$ eine endliche Teilklasse, dann folgt aus (6.23) mittels der Jensen-Ungleichung

$$
\begin{aligned}
\psi_s\left(E\left[\max_{f\in\tilde{\mathcal{F}}}|A_f|/a'\right]\right) &\leq \tilde{\psi}_s\left(E\left[\max_{f\in\tilde{\mathcal{F}}}|A_f|/a'\right]\right) \\
&\leq E\left[\tilde{\psi}_s\left(\max_{f\in\tilde{\mathcal{F}}}|A_f|/a'\right)\right] \\
&\leq E\left[\sum_{f\in\tilde{\mathcal{F}}}\tilde{\psi}_s\left(|A_f|/a'\right)\right] \\
&\leq C_A|\tilde{\mathcal{F}}|.
\end{aligned}
\tag{6.24}
$$

Die inverse Funktion von ψ_s ist streng monoton steigend und gegeben durch $\psi_s^{-1}(t) = (\log(1+t))^{1/s}$. Die Anwendung der Funktion ψ^{-1} auf Ungleichung (6.24) und die Definitionen von a' und a_f liefern

$$
\begin{aligned}
E\left[\max_{f\in\tilde{\mathcal{F}}}|A_f|\right] &\leq \left(\frac{4}{n^{s/2}}\right)^{\frac{1}{s}}\left(\max_{f\in\tilde{\mathcal{F}}}a_f\right)^{\frac{1}{s}}\left(\log(1+C_A|\tilde{\mathcal{F}}|)\right)^{\frac{1}{s}} \\
&\leq \frac{4^{1/s}}{n^{1/2}}\left(2A^s\max_{f\in\tilde{\mathcal{F}}}\|f\|_\infty^s\right)^{\frac{1}{s}}\left(\log(1+C_A|\tilde{\mathcal{F}}|)\right)^{\frac{1}{s}} \\
&= \frac{8^{1/s}A}{n^{1/2}}\max_{f\in\tilde{\mathcal{F}}}\|f\|_\infty\left(\log(1+C_A|\tilde{\mathcal{F}}|)\right)^{\frac{1}{s}}.
\end{aligned}
\tag{6.25}
$$

Eine analoge Abschätzung ist auch für $E\left[\max_{f\in\tilde{\mathcal{F}}}|B_f|\right]$ möglich. Dazu wird die konvexe Funktion

$$\psi_2:\mathbb{R}_{>0}\longrightarrow\mathbb{R},\quad \psi_2(x):=\exp(x^2)-1,$$

betrachtet, deren Ableitung durch $\frac{d}{dx}\psi_2(x)=2x\exp(x^2)$ gegeben ist. Sei $b':=2(b_f)^{1/2}$. Nochmals mittels der Abschätzung für $P(|B_f|>x)$ und wegen des Satzes von Fubini gibt es eine von der Funktion f unabhängige Konstante $C_B>0$ mit

$$\begin{aligned}
E\left[\psi_2\left(|B_f|/b'\right)\right] &= E\left[\int_0^{|B_f|/a'} 2t\exp(t^2)dt\right]\\
&= 2E\left[\int_0^{\infty} \mathbb{1}\{t<|B_f|/b'\}\exp(t^2)\ t\ dt\right]\\
&= 2\int_0^{\infty} P(|B_f|>tb')\exp(t^2)\ t\ dt\\
&\le 2\int_0^{\infty} 2\exp\left(-\frac{(tb')^2}{2b_f}\right)\exp(t^2)\ t\ dt\\
&\le 4\int_0^{\infty} \exp\left(-2t^2\right)\exp(t^2)\ t\ dt\\
&\le C_B<\infty.
\end{aligned}\tag{6.26}$$

Wiederum folgt aus (6.26) für endliche Teilklassen $\tilde{\mathcal{F}}\subset\mathcal{F}_0$ mittels der Jensen-Ungleichung

$$\begin{aligned}
\psi_2\left(E\left[\max_{f\in\tilde{\mathcal{F}}}|B_f|/b'\right]\right) &\le E\left[\psi_2\left(\max_{f\in\tilde{\mathcal{F}}}|B_f|/b'\right)\right]\\
&\le E\left[\sum_{f\in\tilde{\mathcal{F}}}\psi_2\left(|B_f|/b'\right)\right]\\
&\le C_B|\tilde{\mathcal{F}}|.
\end{aligned}\tag{6.27}$$

Wird die monotone inverse Funktion von ψ_2, gegeben durch $\psi_2^{-1}(t)=(\log(1+t))^{1/2}$, auf Ungleichung (6.27) angewendet so ergibt sich

$$E\left[\max_{f\in\tilde{\mathcal{F}}}|B_f|\right]\le\ 4B\max_{f\in\tilde{\mathcal{F}}}\sigma(f)\left(\log(1+C_B|\tilde{\mathcal{F}}|)\right)^{\frac{1}{2}}.\tag{6.28}$$

Die Ergebnisse (6.25) und (6.28) können mittels der Dreiecksungleichung zusammengefasst werden zu

$$\begin{aligned}
E\left[\max_{f\in\tilde{\mathcal{F}}}|G_n(f)|\right] &\leq E\left[\max_{f\in\tilde{\mathcal{F}}}|A_f|\right]+E\left[\max_{f\in\tilde{\mathcal{F}}}|B_f|\right]\\
&\leq \frac{8^{1/s}A}{n^{1/2}}\max_{f\in\tilde{\mathcal{F}}}\|f\|_\infty\left(\log(1+C_A|\tilde{\mathcal{F}}|)\right)^{\frac{1}{s}}\\
&\quad+4B\max_{f\in\tilde{\mathcal{F}}}\sigma(f)\left(\log(1+C_B|\tilde{\mathcal{F}}|)\right)^{\frac{1}{2}}\\
&\leq C_3\left(\frac{1}{n^{1/2}}\max_{f\in\tilde{\mathcal{F}}}\|f\|_\infty\left(\log(1+C_A|\tilde{\mathcal{F}}|)\right)^{\frac{1}{s}}\right. \qquad (6.29)\\
&\quad\left.+\max_{f\in\tilde{\mathcal{F}}}\sigma(f)\left(\log(1+C_B|\tilde{\mathcal{F}}|)\right)^{\frac{1}{2}}\right)\\
&\leq C_4\left(\frac{1}{n^{1/2}}\max_{f\in\tilde{\mathcal{F}}}\|f\|_\infty\left(1\vee\log(C_A|\tilde{\mathcal{F}}|)\right)^{\frac{1}{s}}\right.\\
&\quad\left.+\max_{f\in\tilde{\mathcal{F}}}\sigma(f)\left(1\vee\log(C_B|\tilde{\mathcal{F}}|)\right)^{\frac{1}{2}}\right),
\end{aligned}$$

wobei $C_3, C_4 > 0$ von $\tilde{\mathcal{F}}$ unabhängige Konstanten sind. Mit (6.29) steht eine Maximalungleichung für endliche Funktionenklassen zur Verfügung.

Sei $\delta > 0$, $\mathcal{F}_1 \subset \mathcal{F}_0$ mit $p(f) < \delta\ \forall f \in \mathcal{F}_1$ und

$$\Xi(\mathcal{F}_1,p)=\sum_{q\geq q_0}2^{-q}\left(\sum_{k=q_0}^{q}1\vee\log N_{[]}(2^{-k},\mathcal{F}_1,p)\right)^{\frac{1}{2s}}<\infty$$

für ein $q_0 \in \mathbb{N}$. Ziel des nächsten Abschnitts ist es, mittels Ungleichung (6.29) eine Abschätzung für $E[\sup_{f\in\mathcal{F}_1}|G_n(f)|]$ zu erhalten, wobei $\mathcal{F}_1$, im Gegensatz zu $\tilde{\mathcal{F}}$ in Ungleichung (6.29), unendlich viele Elemente enthalten kann. Für Indikatorfunktionen $\mathbb{1}\{A\}$ und Funktionenklassen $\mathcal{F}_i$ wird im weiteren Verlauf

$$\begin{aligned}
E\|G_n(f)\|_{\mathcal{F}_i} &:= E[\sup_{f\in\mathcal{F}_i}|G_n(f)|],\\
E\|G_n(f\mathbb{1}\{A\})\|_{\mathcal{F}_i} &:= E[\sup_{f\in\mathcal{F}_i}|G_n(f\mathbb{1}\{A\})|]
\end{aligned}$$

gesetzt. Weiterhin sei

$$a(\delta) := \delta\left(1\vee\log(N_{[]}(\delta,\mathcal{F}_1,p))\right)^{-\frac{1}{2s}}. \qquad (6.30)$$

Offensichtlich ist $a(\delta)$ für $\delta \searrow 0$ streng monoton fallend, da die Überdeckungszahl $N_{[]}(\delta, \mathcal{F}_1, p)$ für kleinere Werte δ monoton steigt. Eine additive Zerlegung von $G_n(f)$ und die Dreiecksungleichung liefern mit (6.11)

$$\begin{aligned} E\|G_n(f)\|_{\mathcal{F}_1} &= E\|G_n(f\mathbb{1}\{W > n^{1/2}a(\delta)\} + f\mathbb{1}\{W \le n^{1/2}a(\delta)\})\|_{\mathcal{F}_1} \\ &\le E\|G_n(f\mathbb{1}\{W > n^{1/2}a(\delta)\})\|_{\mathcal{F}_1} \\ &\quad + E\|G_n(f\mathbb{1}\{W \le n^{1/2}a(\delta)\})\|_{\mathcal{F}_1}, \end{aligned} \tag{6.31}$$

wobei W die Schranke für $\mathcal{F}_0$ aus Voraussetzung i) ist. Die Betrachtung der Ungleichung

$$\begin{aligned} |G_n(f)| &= n^{1/2}\frac{1}{n}\left|\sum_{i=1}^n f(X_i) - E[f(X_i)]\right| \\ &\le n^{1/2}\left(\frac{1}{n}\sum_{i=1}^n |f(X_i)| + \frac{1}{n}\sum_{i=1}^n |E[f(X_i)]|\right) \\ &\le n^{1/2}\left(\frac{1}{n}\sum_{i=1}^n W + W\right) = 2n^{1/2}W \end{aligned}$$

erlaubt folgende Abschätzung für den ersten Summanden in Ungleichung (6.31):

$$\begin{aligned} E\|G_n(f\mathbb{1}\{W > n^{1/2}a(\delta)\})\|_{\mathcal{F}_1} &\le E[2n^{1/2}W\mathbb{1}\{W > n^{1/2}a(\delta)\}] \\ &= 2n^{1/2}W\mathbb{1}\{W > n^{1/2}a(\delta)\}. \end{aligned} \tag{6.32}$$

Für jedes feste $\delta > 0$ konvergiert der Term $E\|G_n(f\mathbb{1}\{W > n^{1/2}a(\delta)\})\|_{\mathcal{F}_1}$ also gegen 0. Es stellt sich heraus, dass der erste Summand in (6.31) deswegen in asymptotischer Betrachtung vernachlässigt werden kann. Somit genügt es, den zweiten Summanden zu betrachten. Der zweite Summand in Ungleichung (6.31) enthält nur noch Funktionen der Form $f\mathbb{1}\{W \le n^{1/2}a(\delta)\}$. Mit der Definition

$$\mathcal{F}_2 := \{f\mathbb{1}\{W \le n^{1/2}a(\delta)\},\ f \in \mathcal{F}_1\} \tag{6.33}$$

kann der zweite Summand in (6.31) auch geschrieben werden als $E\|G_n(f)\|_{\mathcal{F}_2}$. Die Funktionenklasse $\mathcal{F}_2$ erbt auch Eigenschaften der Funktionenklasse $\mathcal{F}_1$. Wegen der Voraussetzung iv) ist auch $\mathcal{F}_2$ eine Teilmenge von $\mathcal{F}_0$. Wenn Mengen der Form

$$[l, u] = \{f \in \mathcal{F}_1,\ l \le f \le u\}$$

die Klasse $\mathcal{F}_1$ überdecken, so bilden die Mengen der Form

$$[l', u'] = \{f \in \mathcal{F}_2,\ l' \le f \le u'\}$$

mit $l' = l\mathbb{1}\{W \leq n^{1/2}a(\delta)\}$ und $u' = u\mathbb{1}\{W \leq n^{1/2}a(\delta)\}$ eine Überdeckung von $\mathcal{F}_2$. Da die Monotonievoraussetzung aus vi) zusätzlich

$$p((u-l)\mathbb{1}\{W \leq n^{1/2}a(\delta)\}) \leq p(u-l)$$

impliziert, folgt

$$N_{[]}(\epsilon, \mathcal{F}_2, p) \leq N_{[]}(\epsilon, \mathcal{F}_1, p) \quad \forall \epsilon > 0.$$

Die Überdeckungszahlen von $\mathcal{F}_2$ sind also nicht größer als die von $\mathcal{F}_1$. Damit ist

$$\Xi(\mathcal{F}_2, p) < \infty \text{ für ein } q_0 \in \mathbb{N}$$

auch für die Funktionenklasse $\mathcal{F}_2$ erfüllt, vgl. (6.18). Zusätzlich gilt natürlich $|f| \leq n^{1/2}a(\delta)$, falls $f \in \mathcal{F}_2$. Sei $q_0 \in \mathbb{N}$ die kleinste Zahl, so dass

$$\begin{aligned} 2a(\delta) &\leq 2^{-q_0}\left(1 \vee \log(N_{[]}(2^{-(q_0+1)}), \mathcal{F}_2)\right)^{-\frac{1}{2s}} \\ &= 2a(2^{-(q_0+1)}). \end{aligned} \tag{6.34}$$

Wegen der Definition in (6.30) folgt daraus $\delta/2 \leq 2^{-q_0} \leq \delta$. Nach Voraussetzung kann für jede natürliche Zahl $q \geq q_0$ für $\mathcal{F}_2$ eine Überdeckung

$$[l_{q1}, u_{q1}], ..., [l_{qN'_q}, u_{qN'_q}]$$

gewählt werden, wobei $N'_q := N_{[]}(2^{-q}, \mathcal{F}_2)$, so dass $p(u_{q1} - l_{q1}), ..., p(u_{qN'_q} - l_{qN'_q}) \leq 2^{-q}$. Aus dieser von der Zahl q abhängigen Überdeckung kann eine von q abhängige Partition für $\mathcal{F}_2$

$$\mathcal{F}'_{q,1} \,\dot{\cup}\, ... \,\dot{\cup}\, \mathcal{F}'_{q,N'_q} = \mathcal{F}_2$$

konstruiert werden, also eine Folge (in q) von Partitionen von $\mathcal{F}_2$. Im weiteren Verlauf wird eine Folge von Partitionen erforderlich sein, so dass jedes Element aus der q-ten Partition aus einer (disjunkten) Vereinigung von Elementen aus der $(q+1)$-ten Partition besteht. Die obige Folge von Partitionen erfüllt diese Bedingung noch nicht zwangsläufig, aber eine solche Folge kann konstruiert werden. Für $q = q_0$ sei

$$\mathcal{F}_{q,1}, ..., \mathcal{F}_{q,N'_q} = \mathcal{F}'_{q,1}, ..., \mathcal{F}'_{q,N'_q}.$$

Für $q > q_0$ kann eine Folge von Partitionen rekursiv definiert werden durch

$$\begin{aligned}\mathcal{F}_{q,1}, ..., \mathcal{F}_{q,N_q} = \mathcal{F}_{q-1,1} &\cap \mathcal{F}'_{q,1}, ..., \mathcal{F}_{q-1,N_{q-1}} \cap \mathcal{F}'_{q,1},\\ &\mathcal{F}_{q-1,1} \cap \mathcal{F}'_{q,2}, ..., \mathcal{F}_{q-1,N_{q-1}} \cap \mathcal{F}'_{q,2},\\ &\vdots\\ &\mathcal{F}_{q-1,1} \cap \mathcal{F}'_{q,N'_q}, ..., \mathcal{F}_{q-1,N_{q-1}} \cap \mathcal{F}'_{q,N'_q},\end{aligned}$$

wobei $N_q \leq N'_{q_0} N'_{q_0+1} \cdots N'_q = N(2^{-q_0}, \mathcal{F}_2) \cdots N(2^{-q}, \mathcal{F}_2)$. Diese Folge erfüllt nun auch obige Forderung, d. h. für jedes $i \leq N_{q+1}$ $\exists j \leq N_q$, so dass

$$\mathcal{F}_{q+1,i} \subset \mathcal{F}_{q,j}. \tag{6.35}$$

Wenn $f, g \in [l, u]$, dann $l \leq f, g \leq u$, und damit $f - g \leq f - l \leq u - l$. Offensichtlich gilt also

$$|f - g| \leq \Delta_{q,i} := u_{qi} - l_{qi}, \text{ falls } f, g \in \mathcal{F}_{q,i}.$$

Zu jeder Menge $\mathcal{F}_{q,i}$ sei ein beliebiges Element $f_{q,i} \in \mathcal{F}_{q,i}$ ausgewählt. Durch die Abbildungen

$$\pi_q f := f_{q,i} \;\text{ und }\; \Delta_q f := \Delta_{q,i}, \;\text{falls } f \in \mathcal{F}_{q,i},$$

wird einerseits jeder Funktion f eine Approximation $\pi_q f$ zugeordnet. Andererseits wird jeder Funktion f eine Funktion $\Delta_q f$ zugeordnet, die die Differenzen von f und den Funktionen aus der Partition $\mathcal{F}_{q,i}$, zu der auch f gehört, betragsmäßig nach oben beschränkt. Im nächsten Schritt werden Funktionen $f \in \mathcal{F}_2$ in Teleskopsummen so zerlegt, dass die Summanden der Teleskopsummen aus Differenzen von Funktionen, die in gleichen Partitionen enthalten sind, bestehen. Dazu sei

$$\begin{aligned}a_q &:= 2^{-q} \left(\sum_{k=q_0}^{q+1} 1 \vee \log(N'_k) \right)^{-\frac{1}{2s}},\\ A_{q-1} f &:= \mathbb{1}\{\Delta_{q_0} f \leq n^{1/2} a_{q_0}, ..., \Delta_{q-1} f \leq n^{1/2} a_{q-1}\},\\ B_q f &:= \mathbb{1}\{\Delta_{q_0} f \leq n^{1/2} a_{q_0}, ..., \Delta_{q-1} f \leq n^{1/2} a_{q-1}, \Delta_q f > n^{1/2} a_q\},\\ B_{q_0} f &:= \mathbb{1}\{\Delta_{q_0} f > n^{1/2} a_{q_0}\}.\end{aligned} \tag{6.36}$$

Aus der Definition geht hervor, dass (für festes x) entweder $A_q f(x) = 1$ $\forall q \geq q_0$, und damit $B_q f(x) = 0$ $\forall q \geq q_0$ gilt, oder es gibt genau ein $q_1 \geq q_0$, so dass $B_{q_1} f(x) = 1$,

$A_q f(x) = 1\ \forall q < q_1$ und $A_q f(x) = 0\ \forall q \geq q_1$. Für ein beliebiges $q_0 < q_1 \in \mathbb{N}$ kann die Funktion $f - \pi_{q_0} f$ geschrieben werden als

$$f - \pi_{q_0} f = f - \pi_{q_1} f + (\pi_{q_1} f - \pi_{q_1 - 1} f) + ... + (\pi_{q_0+1} f - \pi_{q_0} f).$$

Dies liefert für $q_1 \longrightarrow \infty$

$$f - \pi_{q_0} f = \sum_{q=q_0}^{\infty} (f - \pi_q f) B_q f + \sum_{q=q_0+1}^{\infty} (\pi_q f - \pi_{q-1} f) A_{q-1} f,$$

da $|\pi_{q+k} f - \pi_{q-1} f| \leq \Delta_{q-1} f \leq n^{1/2} a_{q-1} \longrightarrow 0$ für $q \longrightarrow \infty\ \forall k \geq 0$ und aufgrund der Summierbarkeit der Folge a_q. Mit $|f| \leq g$ liefern die Dreiecksungleichung und die Stationarität der Folge $(X_i)_{i \in \mathbb{Z}}$

$$\begin{aligned}
|G_n(f)| &\leq n^{-1/2} \sum_{i=1}^{n} |f(X_i)| + |E[f(X_i)]| \\
&\leq n^{-1/2} \sum_{i=1}^{n} |f(X_i)| + E[|f(X_i)|] \\
&\leq n^{-1/2} \sum_{i=1}^{n} g(X_i) - E[g(X_i)] + E[g(X_i)] + E[g(X_i)] \\
&= G_n(g) + 2n^{1/2} E[g(X_i)].
\end{aligned}$$

Da $|(f - \pi_q f) B_q f| \leq \Delta_q f B_q f$, kann unter Zuhilfenahme der Dreiecksungleichung und mit (6.11) damit eine Abschätzung für $E\|G_n(f)\|_{\mathcal{F}_2}$ zerlegt werden in

$$\begin{aligned}
E\|G_n(f)\|_{\mathcal{F}_2} \leq E\|G_n(\pi_{q_0} f)\|_{\mathcal{F}_2} &+ E\| \sum_{q=q_0}^{\infty} G_n((f - \pi_q f) B_q f)\|_{\mathcal{F}_2} \\
&+ E\| \sum_{q=q_0+1}^{q_1} G_n((\pi_q f - \pi_{q-1} f) A_q f)\|_{\mathcal{F}_2} \\
\leq\ & E\|G_n(\pi_{q_0} f)\|_{\mathcal{F}_2} + \sum_{q=q_0}^{\infty} E\|G_n(\Delta_q f B_q f)\|_{\mathcal{F}_2} \\
&+ \sum_{q=q_0}^{\infty} 2n^{1/2} \sup_{f \in \mathcal{F}_2} E[\Delta_q f B_q f(X_1)] \\
&+ \sum_{q=q_0+1}^{\infty} E\|G_n(\Delta_{q-1} f A_q f)\|_{\mathcal{F}_2} \\
=:\ & E_1 + E_2 + E_3 + E_4.
\end{aligned} \tag{6.37}$$

Aufgrund der Voraussetzung iii) sind $\Delta_q f B_q f$ bzw. $\Delta_q f A_q f$ in $\mathcal{F}_0$ enthalten (eventuell bis auf einen konstanten Vorfaktor C, der von der Funktion f nicht abhängt). O.B.d.A gelte $C = 1$. Es gibt höchstens N_q viele Funktionen $\Delta_q f B_q f$ bzw. $\Delta_q f A_q f$, wenn f ganz $\mathcal{F}_2$ durchläuft, also

$$|\{\Delta_q f B_q f,\ f \in \mathcal{F}_2\}|,\ |\{\Delta_q f A_q f,\ f \in \mathcal{F}_2\}| \leq N_q. \tag{6.38}$$

Damit ist der Fall einer (eventuell) unendlichen Klasse $\mathcal{F}_2$ auf Fälle mit endlichen Funktionenklassen zurückgeführt. Ungleichung (6.29) hilft nun, die Terme E_2 und E_4 nach oben hin abzuschätzen. Wegen (6.35) und der Definition der Indikatorfunktion $B_q f$ gilt $\Delta_q f B_q f \leq \Delta_{q-1} f B_q f \leq n^{1/2} a_{q-1}$, und nach der Definition von Δ_{qi}, Voraussetzung vi) und der Definition von p gilt $\sigma(\Delta_q f B_q f) \leq p(\Delta_q f B_q f) \leq p(\Delta_q f) \leq 2^{-q}$. Ungleichung (6.29) auf den Term E_2 angewendet liefert

$$\begin{aligned}
E_2 &\leq C_4 \left(\sum_{q=q_0}^{\infty} \frac{1}{n^{1/2}} \max_{f \in \mathcal{F}_2} \|\Delta_q f B_q f\|_\infty \left(1 \vee \log(C_A N_q)\right)^{\frac{1}{s}} \right. \\
&\quad \left. + \max_{f \in \mathcal{F}_2} \sigma(\Delta_q f B_q f) \left(1 \vee \log(C_B N_q)\right)^{\frac{1}{2}} \right) \\
&\leq C_4 \left(\sum_{q=q_0}^{\infty} a_{q-1} \left(1 \vee \log(C_A N_q)\right)^{\frac{1}{s}} + 2^{-q} \left(1 \vee \log(C_B N_q)\right)^{\frac{1}{2}} \right) \\
&\leq C_4 \sum_{q=q_0}^{\infty} 2^{-(q-1)} \left(\sum_{k=q_0}^{q} 1 \vee \log(N'_k) \right)^{-\frac{1}{2s}} \left(1 \vee \log(C_A^2) \vee 2\log(N_q)\right)^{\frac{1}{s}} \\
&\quad + C_4 \sum_{q=q_0}^{\infty} 2^{-q} \left(1 \vee \log(C_A^2) \vee 2\log(N_q)\right)^{\frac{1}{2}} .
\end{aligned} \tag{6.39}$$

Da $N'_{q_0} \geq 1$, gilt

$$\begin{aligned}
1 \vee \log(C_A^2) \vee 2\log(N_q) &\leq 1 \vee \log(C_A^2) \vee 2 \sum_{k=q_0}^{q} 1 \vee \log(N'_k) \\
&\leq \left(\frac{\log(C_A^2)}{2\log(2)} \vee \frac{1}{2\log(2)} \vee 2 \right) \sum_{k=q_0}^{q} 1 \vee \log(N'_k).
\end{aligned}$$

Da $0 < s < 1$ ist, gilt $a^{\frac{1}{2}} \leq a^{\frac{1}{s}}$, falls $a \geq 1$. Somit folgt insgesamt

$$E_2 \leq \ C_5 \sum_{q=q_0}^{\infty} 2^{-q} \left(\sum_{k=q_0}^{q} 1 \vee \log(N'_k) \right)^{\frac{1}{2s}} < \infty$$

für eine Konstante $C_5 > 0$. Damit ist nach Voraussetzung (6.18) der Term $E_2 < \infty$ für q_0 hinreichend groß. Beim Term E_4 kann ähnlich vorgegangen werden. Genau wie eben gilt $\Delta_{q-1} f A_q f \leq n^{1/2} a_{q-1}$ und $\sigma(\Delta_{q-1} f A_q f) \leq p(\Delta_{q-1} f A_q f) \leq p(\Delta_{q-1} f) \leq 2^{-(q-1)}$. Die obige Argumentation führt zu

$$\begin{aligned}
E_4 &\leq C_4 \left(\sum_{q=q_0}^{\infty} \frac{1}{n^{1/2}} \max_{f \in \mathcal{F}_2} \|\Delta_{q-1} f A_q f\|_\infty \left(1 \vee \log(C_A N_q)\right)^{\frac{1}{s}} \right. \\
&\quad \left. + \max_{f \in \mathcal{F}_2} \sigma(\Delta_{q-1} f A_q f) \left(1 \vee \log(C_B N_q)\right)^{\frac{1}{2}} \right) \\
&\leq C_6 \sum_{q=q_0}^{\infty} 2^{-q} \left(\sum_{k=q_0}^{q} 1 \vee \log(N'_k) \right)^{\frac{1}{2s}} < \infty,
\end{aligned}$$

wobei $C_6 > 0$. Nun wird der Term E_1 behandelt. Für a_{q_0} liefert die Definition von $a(\delta)$ in (6.30), die Definition von a_{q_0} in (6.36), und die Wahl von q_0 durch (6.34)

$$\begin{aligned}
a_{q_0} &= 2^{-q_0} \left(\sum_{k=q_0}^{q_0+1} 1 \vee \log(N'_k) \right)^{-\frac{1}{2s}} \\
&\geq 2^{-q_0} \left(2(1 \vee \log(N'_{q_0+1})) \right)^{-\frac{1}{2s}} \\
&\geq \frac{2^{-\frac{1}{2s}}}{2} \delta \left((1 \vee \log(N'_{q_0+1})) \right)^{-\frac{1}{2s}} \\
&\geq \frac{2^{-\frac{1}{2s}}}{2} a(\delta).
\end{aligned}$$

Nach Konstruktion von $\mathcal{F}_2$ (siehe (6.33)) gilt $|\pi_{q_0} f| \leq a(\delta) n^{1/2} \leq a_{q_0} 2^{\frac{1}{2s}+1} n^{1/2}$ sowie $p(\pi_{q_0} f) < \delta$. Die Anwendung der Ungleichung (6.29) auf $E_1 = E\|G_n(\pi_{q_0} f)\|_{\mathcal{F}_2}$ ergibt wegen $\delta/2 \leq 2^{-q_0}$ (siehe (6.34)) damit

$$\begin{aligned}
E_1 &\leq C_4 \left(\frac{1}{n^{1/2}} \|\pi_{q_0} f\|_\infty \left(1 \vee \log(C_A N_q)\right)^{\frac{1}{s}} \right. \\
&\quad \left. + \sigma(\pi_{q_0} f) \left(1 \vee \log(C_B N_q)\right)^{\frac{1}{2}} \right) \\
&\leq C_4 \left(2^{\frac{1}{2s}+1} a_{q_0} \left(1 \vee \log(C_A N_q)\right)^{\frac{1}{s}} + \delta \left(1 \vee \log(C_B N_q)\right)^{\frac{1}{2}} \right) \\
&\leq C_7 2^{-q_0} \left(1 \vee \log(N'_{q_0})\right)^{\frac{1}{2s}} \\
&\leq C_7 \sum_{q=q_0}^{\infty} 2^{-q} \left(\sum_{k=q_0}^{q} 1 \vee \log(N'_k) \right)^{\frac{1}{2s}} < \infty
\end{aligned}$$

für eine Konstante $C_7 > 0$. Übrig bleibt noch der Term $E_3 = \sum_{q=q_0}^{\infty} 2n^{1/2} \sup_{f \in \mathcal{F}_2} E[\Delta_q f B_q f(X_1)]$. Aus der Definition von $B_q f$ geht die Ungleichung

$B_q f \leq \mathbb{1}\{\Delta_q f > n^{1/2} a_q\}$ hervor, und damit gilt

$$\begin{aligned}
n^{1/2} a_q E[\Delta_q f B_q f(X_1)] &\leq E[\Delta_q f \mathbb{1}\{\Delta_q f > n^{1/2} a_q\}(X_1)] \\
&\leq n^{1/2} a_q \int_0^{n^{1/2} a_q} P(\Delta_q f \mathbb{1}\{\Delta_q f > n^{1/2} a_q\}(X_1) > t) dt \\
&\quad + n^{1/2} a_q \int_{n^{1/2} a_q}^{\infty} P(\Delta_q f \mathbb{1}\{\Delta_q f > n^{1/2} a_q\}(X_1) > t) dt \\
&=: T_1 + T_2,
\end{aligned}$$

da $E[X] = \int_0^\infty P(X > t)dt$, falls $X \geq 0$ fast sicher und X integrierbar ist. Der erste Summand kann mit Hilfe der Markov-Ungleichung abgeschätzt werden durch

$$\begin{aligned}
T_1 &\leq n^{1/2} a_q \int_0^{n^{1/2} a_q} P(\Delta_q f(X_1) > n^{1/2} a_q) dt \\
&\leq (n^{1/2} a_q)^2 P(\Delta_q f(X_1) > n^{1/2} a_q) \\
&= (n^{1/2} a_q)^2 P((\Delta_q f(X_1))^2 > (n^{1/2} a_q)^2) \\
&\leq (n^{1/2} a_q)^2 \frac{1}{(n^{1/2} a_q)^2} E[(\Delta_q f(X_1))^2] \\
&\leq p((\Delta_q f)^2) \\
&\leq (2^{-q})^2.
\end{aligned}$$

Für den zweiten Summanden liefert die Markov-Ungleichung

$$\begin{aligned}
T_2 &= n^{1/2} a_q \int_{n^{1/2} a_q}^{\infty} t^{-2} t^2 P(\Delta_q f \mathbb{1}\{\Delta_q f > n^{1/2} a_q\}(X_1) > n^{1/2} a_q) dt \\
&\leq n^{1/2} a_q \int_{n^{1/2} a_q}^{\infty} t^{-2} t^2 P(\Delta_q f(X_1) > t) dt \\
&\leq n^{1/2} a_q \int_{n^{1/2} a_q}^{\infty} t^{-2} \left(\sup_{z>0} z^2 P((\Delta_q f(X_1))^2 > z^2) \right) dt \\
&\leq n^{1/2} a_q \int_{n^{1/2} a_q}^{\infty} t^{-2} E[(\Delta_q f(X_1))^2] dt \\
&= n^{1/2} a_q E[(\Delta_q f(X_1))^2] \frac{1}{n^{1/2} a_q} \\
&\leq (2^{-q})^2,
\end{aligned}$$

also $n^{1/2}\sup_{f\in\mathcal{F}_2} E[\Delta_q f B_q f(X_1)] \leq 2\frac{1}{a_q}(2^{-q})^2$. Werden diese Resultate für eine Abschätzung des Terms E_3 zusammengefasst, so ergibt sich

$$\begin{aligned}
E_3 &= \sum_{q=q_0}^{\infty} 2n^{1/2} \sup_{f\in\mathcal{F}_2} E[\Delta_q f B_q f(X_1)] \leq \sum_{q=q_0}^{\infty} 4\frac{1}{a_q}(2^{-q})^2 \\
&\leq \sum_{q=q_0}^{\infty} 4\frac{\left(\sum_{k=q_0}^{q+1} 1\vee\log(N_k')\right)^{\frac{1}{2s}}}{2^{-q}}(2^{-q})^2 = 4\sum_{q=q_0}^{\infty} 2^{-q}\left(\sum_{k=q_0}^{q+1} 1\vee\log(N_k')\right)^{\frac{1}{2s}} \\
&= 8\sum_{q=q_0}^{\infty} 2^{-(q+1)}\left(\sum_{k=q_0}^{q+1} 1\vee\log(N_k')\right)^{\frac{1}{2s}} = 8\sum_{q=q_0+1}^{\infty} 2^{-q}\left(\sum_{k=q_0}^{q} 1\vee\log(N_k')\right)^{\frac{1}{2s}} \\
&\leq C_8\sum_{q=q_0}^{\infty} 2^{-q}\left(\sum_{k=q_0}^{q} 1\vee\log(N_k')\right)^{\frac{1}{2s}} < \infty
\end{aligned}$$

für eine Konstante $C_8 > 0$. Aus den Abschätzungen für die Terme E_1, E_2, E_3, E_4 folgt

$$E\|G_n(f)\|_{\mathcal{F}_2} \leq (C_5 + C_6 + C_7 + C_8)\sum_{q=q_0}^{\infty} 2^{-q}\left(\sum_{k=q_0}^{q} 1\vee\log(N_k')\right)^{\frac{1}{2s}},$$

und die Terme $2^{-q}\left(\sum_{k=q_0}^{q} 1\vee\log(N_k')\right)^{\frac{1}{2s}}$ sind absolut summierbar. Sei $\eta > 0$, und sei δ so klein bzw. $q' > q_0$ so groß gewählt, dass

$$\begin{aligned}
\frac{E\|G_n(f)\|_{\mathcal{F}_2}}{(C_5 + C_6 + C_7 + C_8)} &\leq \sum_{q=q'}^{\infty} 2^{-q}\left(\sum_{k=q'}^{q} 1\vee\log(N_k')\right)^{\frac{1}{2s}} \\
&\leq \sum_{q=q'}^{\infty} 2^{-q}\left(\sum_{k=q_0}^{q} 1\vee\log(N_k')\right)^{\frac{1}{2s}} \\
&\leq \frac{\eta}{(C_5 + C_6 + C_7 + C_8)}.
\end{aligned}$$

Damit folgt, dass $E\|G_n(f)\|_{\mathcal{F}_2} \leq \eta$ für hinreichend kleines $\delta > 0$ bzw. zusammen mit (6.11), (6.32) und (6.33)

$$E\|G_n(f)\|_{\mathcal{F}_1} \leq \eta + 2n^{1/2}W\mathbb{1}\{W > n^{1/2}a(\delta)\}. \tag{6.40}$$

Es gelte weiterhin ohne Einschränkung, dass in Voraussetzung iii) $C = 1$. Bei Betrachtung von ii) in Theorem 6.6 fällt wegen $G_n(f) - G_n(g) = G_n(f-g)$ auf, dass nicht

Funktionen f aus $\mathcal{F}$ auftreten, sondern genauer Differenzen $f - g$ von Funktionen f, g aus $\mathcal{F}$. Sei deswegen

$$\mathcal{F}_1 := \{f = g - h, \text{ so dass } g, h \in \mathcal{F}, \ p(f) < \delta\}.$$

Als Nächstes ist zu überprüfen, dass $\mathcal{F}_1$ eine Relation wie in (6.18) erfüllt, diese Eigenschaft also von der Funktionenklasse $\mathcal{F}$ erbt. Wegen Lemma 6.11 gilt $p(f+g) \leq (1 + 2^{1/2})(p(f) + p(g))$. Damit können die Überdeckungszahlen $N_{[]}(\eta, \mathcal{F}, p)$ von $\mathcal{F}$ in Relation zu den Überdeckungszahlen $N_{[]}(\eta, \mathcal{F}_1, p)$ von $\mathcal{F}_1$ gebracht werden. Sei $\eta > 0$ und

$$[l_1, u_1], ..., [l_{N_\eta}, u_{N_\eta}] \text{ mit } p(u_i - l_i) < \eta$$

eine Überdeckung von $\mathcal{F}$ vom Umfang N_η. Für beliebige $f, g \in \mathcal{F}$ gibt es i, j mit

$$l_i \leq f \leq u_i \quad \text{und} \quad l_j \leq g \leq u_j.$$

Damit gilt auch

$$\begin{aligned} l_i - u_j \leq f - g \ &\leq \ u_i - l_j, \text{ also} \\ f - g \in [l_i - u_j, u_i - l_j]. \end{aligned}$$

Von den Mengen $[l_i - u_j, u_i - l_j]$ gibt es nicht mehr als N_η^2 viele, und wegen

$$\begin{aligned} p((u_i - l_j) - (l_i - u_j)) &= p((u_i - l_i) + (u_j - l_j)) \\ &< (1 + 2^{1/2})(\eta + \eta) \\ &= 2(1 + 2^{1/2})\eta \end{aligned}$$

bilden diese Mengen eine Überdeckung von $\mathcal{F}_1$ von einem Umfang nicht größer als N_η^2. Daraus folgt

$$N_{[]}(2(1 + 2^{1/2})\eta, \mathcal{F}_1, p) \leq \left(N_{[]}(\eta, \mathcal{F}, p)\right)^2$$

bzw. nach Umskalierung von η

$$N_{[]}(\eta, \mathcal{F}_1, p) \leq \left(N_{[]}(C_9\eta, \mathcal{F}, p)\right)^2,$$

wobei $C_9 := (2(1 + 2^{1/2}))^{-1}$. Sei $k_0 \in \mathbb{N}$ die kleinste Zahl, so dass $2^{-k_0} < C_9$, und q_0 so groß, dass $N_{[]}(2^{-q_0}, \mathcal{F}, p) \geq 1$. Daraus folgt $N_{[]}(C_9 2^{-k}, \mathcal{F}, p) \leq N_{[]}(2^{-(k+k_0)}, \mathcal{F}, p)$, und mit $C_{10} := 2^{\frac{1}{2s}} 2^{k_0}$ sowie $q_0' := q_0 + k_0$ ergibt sich

$$\begin{aligned}
&\sum_{q=q_0}^{\infty} 2^{-q} \left(\sum_{k=q_0}^{q} 1 \vee \log(N_{[]}(2^{-k}, \mathcal{F}_1, p)) \right)^{\frac{1}{2s}} \\
&\quad \leq \sum_{q=q_0}^{\infty} 2^{-q} \left(\sum_{k=q_0}^{q} 1 \vee \log((N_{[]}(2^{-(k+k_0)}, \mathcal{F}, p))^2) \right)^{\frac{1}{2s}} \\
&\quad \leq \sum_{q=q_0}^{\infty} 2^{-q} \left(\sum_{k=q_0+k_0}^{q+k_0} 2(1 \vee \log(N_{[]}(2^{-k}, \mathcal{F}, p))) \right)^{\frac{1}{2s}} \\
&\quad = 2^{\frac{1}{2s}} 2^{k_0} \sum_{q=q_0}^{\infty} 2^{-(q+k_0)} \left(\sum_{k=q_0+k_0}^{q+k_0} 1 \vee \log(N_{[]}(2^{-k}, \mathcal{F}, p)) \right)^{\frac{1}{2s}} \\
&\quad = C_{10} \sum_{q=q_0'}^{\infty} 2^{-q} \left(\sum_{k=q_0'}^{q} 1 \vee \log(N_{[]}(2^{-k}, \mathcal{F}, p)) \right)^{\frac{1}{2s}} \\
&\quad < \infty.
\end{aligned}$$

Dies gilt aufgrund der Eigenschaften der Logarithmusfunktion und nach Voraussetzung an die Funktionenklasse $\mathcal{F}$. Der Term wird, wie oben gezeigt, durch die Wahl einer hinreichend kleinen Zahl $\delta > 0$ bzw. einer hinreichend großen Zahl $q_0 \in \mathbb{N}$ (bzw. $q_0' \in \mathbb{N}$) beliebig klein. Damit ist auch (6.18) für $\mathcal{F}_1$ gezeigt. Nach Voraussetzung kann für jedes $\delta > 0$ (bzw. $q_0 \in \mathbb{N}$) die Funktionenklasse $\mathcal{F}$ überdeckt werden mit disjunkten Mengen

$$\mathcal{F}^1 \,\dot{\cup}\, ... \,\dot{\cup}\, \mathcal{F}^{N'_{q_0}} \subset \mathcal{F},$$

so dass $p(f - g) < \delta$, falls $f, g \in \mathcal{F}^i$. Mit der Markov-Ungleichung und mit (6.11) folgt

$$\begin{aligned}
P\left(\sup_i \sup_{f,g \in \mathcal{F}^i} |G_n(f) - G_n(g)| \geq \epsilon \right) &\leq \frac{1}{\epsilon} E\left[\sup_i \sup_{f,g \in \mathcal{F}^i} |G_n(f) - G_n(g)| \right] \\
&= \frac{1}{\epsilon} E\left[\sup_i \sup_{f,g \in \mathcal{F}^i} |G_n(f - g)| \right].
\end{aligned}$$

Aus $f, g \in \mathcal{F}^i$ folgt $f - g \in \mathcal{F}_1$ und wegen (6.40) auch

$$\begin{aligned}\frac{1}{\epsilon} E\left[\sup_i \sup_{f,g\in\mathcal{F}^i} |G_n(f-g)|\right] &\leq \frac{1}{\epsilon} E\left[\sup_{f\in\mathcal{F}_1} |G_n(f)|\right] \\ &= \frac{1}{\epsilon} E\|G_n(f)\|_{\mathcal{F}_1}. \\ &\leq \frac{1}{\epsilon}\left(\eta + 2n^{1/2} W \mathbb{1}\{W > a(\delta) n^{1/2}\}\right).\end{aligned}$$

Die Zusammenfassung dieser Resultate liefert

$$\limsup_{n\longrightarrow\infty} P\left(\sup_i \sup_{f,g\in\mathcal{F}^i} |G_n(f) - G_n(g)| \geq \epsilon\right) \leq \frac{\eta}{\epsilon},$$

wobei $\frac{\eta}{\epsilon}$ für jeden festen Wert $\epsilon > 0$ beliebig klein gewählt werden kann, wenn nur $\delta > 0$ hinreichend klein gewählt wird. Daraus folgt ii) aus Theorem 6.6, und der zweite Teil des Beweises ist abgeschlossen. □

Kapitel 7

Anwendungen

Dieses Kapitel behandelt beispielhaft einige Anwendungsmöglichkeiten der in Kapitel 6 gewonnenen Resultate. Die schwache Konvergenz empirischer Prozesse hängt in der Regel von der Beschaffenheit der Indexmenge ab, auf der die empirischen Prozesse definiert sind. Dabei spielt die „Größe“ der Indexmenge eine entscheidende Rolle. Über kombinatorische Betrachtungen liefert die Theorie der Vapnik-Červonenenkis-Klassen oftmals einfach nachzuweisende Aussagen über Überdeckungszahlen. Deshalb wird ein Schwerpunkt dieses Kapitels auf Methoden zum Nachweis der Voraussetzung (6.18) in Theorem 6.12 gelegt.

7.1 Anwendung des Zentralen Grenzwertsatzes

Zentrale Grenzwertsätze wie in Theorem 6.3 erlauben es, das asymptotische Verhalten des Einstichproben-t-Tests bei Verletzung stochastischer Unabhängigkeit der Eingangsdaten zu untersuchen. Sei $(Z_i)_{i\in\mathbb{Z}}$ eine stationäre Folge von reellwertigen und L_2-integrierbaren Zufallsvariablen mit $\mu_0 := E[Z_i]$, und die Folge gegeben durch $X_i := Z_i - \mu_0$ erfülle die Voraussetzungen aus Theorem 6.3. Sei wie in Theorem 6.3

$$\sigma := \left(\lim_{n\longrightarrow\infty} \frac{1}{n} Var[Z_1 + ... + Z_n]\right)^{\frac{1}{2}}$$

und $S'[Z_1, ..., Z_n]$ eine Schätzerfolge mit

$$S'[Z_1, ..., Z_n] \longrightarrow \sigma \text{ in Wahrscheinlichkeit.} \tag{7.1}$$

Mit Theorem 6.3 folgt dann

$$n^{1/2} \frac{\overline{Z} - \mu_0}{S'[Z_1, ..., Z_n]} \rightsquigarrow N(0, 1). \tag{7.2}$$

Die Statistik in (7.2) könnte also für einen asymptotisch unverzerrten Test für das Testproblem

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0$$

verwendet werden, falls eine für σ konsistente Schätzerfolge wie in (7.1) ermittelt werden kann. Leider scheint es schwierig, eine derartige Schätzerfolge zu finden, die zudem eine akzeptable Konvergenzrate aufweist.

Es stellt sich die Frage nach dem asymptotischen Verhalten des klassischen Einstichproben-t-Tests. Angenommen, dass nicht nur die Folge $(X_i)_{i\in\mathbb{Z}}$ die Voraussetzungen aus Theorem 6.3 erfüllt, sondern auch die Folge $(Y_i)_{i\in\mathbb{Z}}$, definiert durch $Y_i := X_i^2 - E[X_i^2]$. Zunächst soll das asymptotische Verhalten des Streuungsmaßes aus (2.2), wie es beim t-Test vorkommt, untersucht werden:

$$\begin{aligned}
S[Z_1, ..., Z_n] &= \left(\frac{n}{n-1} \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right)^{\frac{1}{2}} \\
&= \left(\frac{n}{n-1} \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - (\overline{X})^2 \right)^{\frac{1}{2}} \\
&= \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 \right)^{\frac{1}{2}} + o_p(1) \\
&= \left(\frac{1}{n} \sum_{i=1}^{n} Y_i \right)^{\frac{1}{2}} + (Var[Z_i])^{\frac{1}{2}} + o_p(1) \\
&= (Var[Z_i])^{\frac{1}{2}} + o_p(1),
\end{aligned}$$

wobei die dritte und die letzte Gleichung mit Theorem 6.3 zu begründen ist. Aus Theorem 6.3 folgt zudem

$$n^{1/2} \frac{\overline{Z} - \mu_0}{S[Z_1, ..., Z_n]} = \frac{\sigma}{(Var[Z_i])^{\frac{1}{2}}} \, n^{1/2} \frac{\overline{Z} - \mu_0}{\sigma} + o_p(1) \rightsquigarrow N(0, \sigma^2 / Var[Z_i]).$$

Der klassische Einstichproben-t-Test ist demnach asymptotisch unverzerrt, wenn $\sigma^2 = Var[Z_i]$.

7.2 VC-Klassen und Überdeckungszahlen

In diesem Abschnitt soll Theorem 6.12 weiter vertieft werden. Die Hauptschwierigkeit bei der Anwendung von Theorem 6.12 auf ein konkretes Modell scheint im Nachweis der Voraussetzungen an die Kovarianzen

$$Cov[f(X_{s_1})\cdots f(X_{s_u}), f(X_{t_1})\cdots f(X_{t_v})]$$

und im Nachweis der Voraussetzungen an die Überdeckungszahlen $N_{[]}(\epsilon, \mathcal{F}, p)$ zu liegen. Ansätze zum Nachweis der ersten Bedingung finden sich in Doukhan und Neumann (2007) bzw. in Doukhan und Louhichi (1999). Die darin beschriebenen Funktionenklassen erfüllen jedoch leider nicht die Abgeschlossenheit gegenüber Multiplikation mit Indikatorfunktionen, wie in der Voraussetzung iv) der Theorems 6.12 gefordert, und berücksichtigen zudem nicht die Terme $\sigma^2(f)\|f\|_\infty^{u+v-2}$ in (6.17). Der Nachweis der in (6.17) geforderten Bedingung scheint schwierig zu sein.
Zur analytischen Betrachtung der Überdeckungszahlen $N_{[]}(\epsilon, \mathcal{F}, p)$ kann der Begriff der VC-Klasse hilfreich sein (vgl. van der Vaart (1998), Chapter 19).

Definition 7.1 *Sei E eine Menge und $\mathcal{D} \subset 2^E$ ein Mengensystem, wobei 2^E die Potenzmenge von E bezeichnet. Wenn es ein Polynom $p(n)$ in $n \in \mathbb{N}$ gibt, so dass für jede n-elementige Teilmenge $N \subset E$ die Relation*

$$|\{N \cap D,\ D \in \mathcal{D}\}| \leq p(n)$$

gilt, dann heißt das Mengensystem $\mathcal{D}$ eine Vapnik-Červonenkis Klasse in E bzw. VC-Klasse.

VC-Klassen wurden erstmals in Vapnik und Červonenenkis (1971) betrachtet. Es kann gezeigt werden, dass es entweder ein solches Polynom $p(n)$ wie in Definition 7.1 gibt, oder dass alle endlichen Mengen $N \subset E$ in all ihre Teilmengen durch Bildung der Schnitte mit Mengen $D \in \mathcal{D}$ zerlegt werden können:

$$|\{N \cap D,\ D \in \mathcal{D}\}| = 2^n \quad \forall N \subset E,\ |N| = n.$$

Mit anderen Worten: Jede endliche Menge N wird im zweiten Fall von $\mathcal{D}$ zerschlagen („shattered“). Dies ist auch als Sauer’s Lemma bekannt (vgl. van der Vaart und Wellner

(1996), Corollary 2.6.3). Dieser Sachverhalt ermöglicht es, den Begriff von VC-Klassen über die VC-Dimension zu definieren

Definition 7.2 *Seinen $E, \mathcal{D}$ wie in Definition 7.1. Die kleinste Zahl $n \in \mathbb{N}$, so dass für jede Teilmenge $N \subset E$ mit $|N| = n$ die Relation*

$$|\{N \cap D,\ D \in \mathcal{D}\}| \leq 2^n - 1$$

gilt, heißt VC-Dimension von $\mathcal{D}$ in E und wird mit $V(\mathcal{D})$ bezeichnet. Gibt es keine solche Zahl $n \in \mathbb{N}$, so wird $V(\mathcal{D}) := \infty$ gesetzt.

Nach Sauers's Lemma ist ein Mengensystem $\mathcal{D}$ genau dann eine VC-Klasse, wenn die VC-Dimension von $\mathcal{D}$ endlich ist. Die Definition von VC-Klassen wird durch den nächsten Begriff verallgemeinert.

Definition 7.3 *Sei $\mathcal{X}$ eine Menge und $\mathcal{F}$ eine Klasse von Funktionen $f : \mathcal{X} \longrightarrow \mathbb{R}$. Für $f \in \mathcal{F}$ heißt die Menge*

$$\{(x,t),\ f(x) < t\} =: D_f \in 2^{\mathcal{X} \times \mathbb{R}}$$

Subgraf von f. Falls das Mengensystem von Subgrafen der Funktionen $f \in \mathcal{F}$

$$\{D_f,\ f \in \mathcal{F}\} =: \mathcal{D}_{\mathcal{F}} \subset 2^{\mathcal{X} \times \mathbb{R}}$$

eine VC-Klasse in der Menge $\mathcal{X} \times \mathbb{R}$ bildet, so heißt $\mathcal{F}$ eine VC-Klasse von Funktionen. Diese werden im Folgenden kurz als VC-Klassen bezeichnet. Die VC-Dimension des Mengensystems $\mathcal{D}_{\mathcal{F}}$ in $\mathcal{X} \times \mathbb{R}$ wird als die VC-Dimension von $\mathcal{F}$ bzw. $V(\mathcal{F})$, bezeichnet.

Es kann gezeigt werden, dass ein Mengensystem $\mathcal{D}$ genau dann eine VC-Klasse in E ist, wenn die Klasse von Funktionen

$$\mathcal{F} := \{\mathbb{1}\{D\} : E \longrightarrow \mathbb{R},\ D \in \mathcal{D}\}$$

eine VC-Klasse von Funktionen ist. In diesem Fall gilt $V(\mathcal{D}) = V(\mathcal{F})$. Demnach kann jede VC-Klasse als VC-Klasse von Funktionen (bestehend aus Indikatorfunktionen) definiert werden. Einige Zusammenhänge zwischen VC-Klassen und Überdeckungszahlen können in Bezug auf Theorem 6.12 hilfreich sein. Eine Schwierigkeit bzgl. der Anforderung von Theorem 6.12 an die Überdeckungszahlen $N_{[]}(\epsilon, \mathcal{F}, p)$ liegt in der Handhabung des Funktionals p. Ist jedoch aufgrund der Abhängigkeitsstruktur der Folge $(X_i)_{i\in\mathbb{Z}}$ für ein $r \geq 1$ und $C > 0$

$$p(f) \leq C\|f\|_r = C(E[|f(X_1)|^r])^{1/r} \quad \forall f \text{ mit } \|f\|_r < \infty \tag{7.3}$$

gegeben, so kann eine Argumentation über VC-Klassen helfen, $N_{[]}(\epsilon, \mathcal{F}, p)$ nach oben hin abzuschätzen. Aus (7.3) folgt unmittelbar

$$N_{[]}(\epsilon, \mathcal{F}, p) \leq N_{[]}(\epsilon, \mathcal{F}, C\|\cdot\|_r) = N_{[]}(\epsilon/C, \mathcal{F}, \|\cdot\|_r). \tag{7.4}$$

Wenn $\mathcal{F}$ eine VC-Klasse von Funktionen ist, lassen sich die Überdeckungszahlen $N_{[]}(\epsilon/C, \mathcal{F}, \|\cdot\|_r)$ in vielen Fällen nach oben hin abschätzen (vgl. van der Vaart und Wellner (1996), Chapter 2.7). Jedoch reichen diese Abschätzungen in der Regel noch nicht aus, um den Anforderungen in Theorem 6.12 an die Überdeckungszahlen zu genügen. Eine Relation zu metrischen Überdeckungszahlen $N(\epsilon/C, \mathcal{F}, \|\cdot\|_r)$ wäre hierbei hilfreich. Es gilt zwar

$$N(\epsilon, \mathcal{F}, \|\cdot\|_r) \leq N_{[]}(2\epsilon, \mathcal{F}, \|\cdot\|_r)$$

(vgl. van der Vaart und Wellner (1996), Definition 2.16), doch wird eine Abschätzung von $N_{[]}(2\epsilon, \mathcal{F}, \|\cdot\|_r)$ nach oben benötigt. Dazu sollen hier spezielle Funktionenklassen betrachtet werden. Sei $k > 0$,

$$f_A : \mathbb{R} \longrightarrow \mathbb{R}, \quad f_A(x) := \begin{cases} -k & \text{für} \quad x \leq -k \\ x & \text{für} \quad -k \leq x \leq k \\ k & \text{für} \quad k \leq x \end{cases} \tag{7.5}$$

und

$$\begin{aligned} \mathcal{F}_A &:= \{f_A(\cdot - t),\ t \in \mathbb{R}\}, \\ \mathcal{F}_0 &:= \{f\mathbb{1}\{B\},\ |f| \leq k,\ Lip(f) \leq 1,\ B \in \mathcal{B}(\mathbb{R})\}. \end{aligned} \tag{7.6}$$

Die Klasse $\mathcal{F}_0$ erfüllt die Voraussetzungen i) - iv) aus Theorem 6.12 und $\mathcal{F}_A \subset \mathcal{F}_0$. Auf der Klasse $\mathcal{F}_A$ ist in natürlicher Weise eine Ordnungsrelation gegeben. Zwei Funktionen aus $\mathcal{F}_A$ sind entweder gleich, oder eine Funktion ist punktweise kleiner oder gleich als die andere Funktion. Für $r \geq 1$, $\epsilon > 0$ und eine messbare Funktion $h : \mathbb{R} \longrightarrow \mathbb{R}$ sei

$$\begin{aligned} B_{h,\epsilon} &:= \{f \in \mathcal{F}_A,\ \|f - h\|_r < \epsilon\} \text{ und} \\ \overline{B_{h,\epsilon}} &:= \{f \in \mathcal{F}_A,\ \|f - h\|_r \leq \epsilon\}, \end{aligned}$$

wobei $\|f-h\|_r := (E[|f(X_1)-h(X_1)|^r])^{1/r}$. Ist die Verteilung von X_1 so, dass die Menge $\overline{B_{h,\epsilon}}$ für einen hinreichend kleinen Wert $\epsilon > 0$ stets eine kleinste und größte Funktion $u, l \in \mathcal{F}$ gegeben durch

$$\begin{aligned} l_A(x) &:= \inf\{f(x),\ f \in \overline{B_{h,\epsilon}}\} \\ u_A(x) &:= \sup\{f(x),\ f \in \overline{B_{h,\epsilon}}\} \end{aligned}$$

enthält, dann gilt $B_{h,\epsilon} \subset [l_A, u_A]$ und $\|u_A - h\|_r, \|h - l_A\|_r \leq \epsilon < \epsilon + \frac{\epsilon}{2}$. Durch Hinzunahme der Dreiecksungleichung folgt

$$\|u_A - l_A\|_r \leq \|u_A - h\|_r + \|h - l_A\|_r < 3\epsilon.$$

In Hinblick auf die parametrische Struktur der Klasse $\mathcal{F}_A$ bedeutet dies

$$\overline{B_{h,\epsilon}} \subset \{f_A(\cdot - t),\ t \in [a,b]\}, \tag{7.7}$$

wobei $[a,b]$ ein abgeschlossenes Intervall ist und $f_A(\cdot - a), f_A(\cdot - b) \in \overline{B_{h,\epsilon}}$. Also kann aus einer Überdeckung durch Mengen der Form $B_{h,\epsilon}$ eine Überdeckung aus Mengen der Form $[l_A, u_A]$ konstruiert werden mit $\|u_A - l_A\|_r < 3\epsilon$. Somit gilt

$$\begin{aligned} N_{[]}(3\epsilon, \mathcal{F}_A, \|\cdot\|_r) &\leq N(\epsilon, \mathcal{F}_A, \|\cdot\|_r) \text{ bzw.} \\ N_{[]}(\epsilon, \mathcal{F}_A, \|\cdot\|_r) &\leq N(\epsilon/3, \mathcal{F}_A, \|\cdot\|_r). \end{aligned} \tag{7.8}$$

Ist $\mathcal{F}_A$ eine VC-Klasse mit VC-Dimension $V(\mathcal{F}_A)$, dann können die rechten Terme in (7.8) durch Polynome in $1/\epsilon$ nach oben hin abgeschätzt werden. Nach Theorem 2.6.7 aus van der Vaart und Wellner (1996) gilt

$$N(\epsilon/3, \mathcal{F}_A, \|\cdot\|_r) \leq K_{V(\mathcal{F}_A)} \left(\frac{1}{\epsilon/3}\right)^{r(V(\mathcal{F}_A)-1)}, \tag{7.9}$$

wobei $K_{V(\mathcal{F}_A)}$ eine Konstante ist, welche nur von r, der VC-Dimension $V(\mathcal{F}_A)$ und der Schranke k abhängt. Die Schranke k fungiert hier als die einhüllende Funktion in Theorem 2.6.7 aus van der Vaart und Wellner (1996). Es kann leicht gezeigt werden, dass $\mathcal{F}_A$ aus (7.6) eine VC-Klasse ist mit VC-Dimension $V(\mathcal{F}_A) = 2$. Zusammenfassend liefern obige Resultate

$$N_{[]}(\epsilon, \mathcal{F}_A, p) \leq N(\epsilon/3C, \mathcal{F}_A, \|\cdot\|_r) \leq K_{V(\mathcal{F}_A)} \left(\frac{3C}{\epsilon}\right)^R, \tag{7.10}$$

wobei $R := r(V(\mathcal{F}_A) - 1)$. Eine polynomielle Abschätzung der Überdeckungszahlen $N_{[]}(\epsilon, \mathcal{F}_A, p)$ reicht aus, um den Anforderungen aus Theorem 6.12 in (6.18) an die Überdeckungszahlen zu genügen, ist also hinreichend dafür, dass

$$\Xi(\mathcal{F}_A, p) = \sum_{q \geq q_0} 2^{-q} \left(\sum_{k=q_0}^{q} 1 \vee \log(N_{[]}(2^{-k}, \mathcal{F}_A, p)) \right)^{\frac{1}{2s}} < \infty$$

für ein $q_0 \in \mathbb{N}$, wie folgende Rechnung zeigt. Wegen (7.10) gilt

$$\begin{aligned} \Xi(\mathcal{F}_A, p) &\leq K_{V(\mathcal{F}_A)} \sum_{q \geq q_0} 2^{-q} \left(\sum_{k=q_0}^{q} \log((3C2^k)^R) \right)^{\frac{1}{2s}} \\ &\leq K_{V(\mathcal{F}_A)} \sum_{q \geq q_0} 2^{-q} \left(\sum_{k=q_0}^{q} R \log(3C) + Rk \log(2) \right)^{\frac{1}{2s}} \\ &\leq K_{V(\mathcal{F}_A)} \sum_{q \geq q_0} 2^{-q} \left(R \log(3C) q + R \log(2) q^2 \right)^{\left[\frac{1}{2s}\right]+1} \\ &= K_{V(\mathcal{F}_A)} \sum_{q \geq q_0} 2^{-q} p(q) < \infty, \end{aligned} \tag{7.11}$$

da $p(q)$ ein Polynom in q ist. Mittels einer leichten Abänderung der Funktionenklassen in (7.6) kann eine weitere Funktionenklasse analysiert werden. Sei dazu

$$f_B : \mathbb{R} \longrightarrow \mathbb{R}, \quad f_B(x) := \begin{cases} |x| & \text{für} \quad -k \leq x \leq k \\ k & \text{sonst} \end{cases} \tag{7.12}$$

und die Funktionenklassen wie in (7.6) definiert durch

$$\begin{aligned} \mathcal{F}_B &:= \{f_B(\cdot - t),\ t \in \mathbb{R}\}, \\ \mathcal{F}_0 &:= \{f \mathbb{1}\{B\},\ |f| \leq k,\ Lip(f) \leq 1,\ B \in \mathcal{B}(\mathbb{R})\}. \end{aligned} \tag{7.13}$$

Zwischen den Klassen $\mathcal{F}_B$ und $\mathcal{F}_A$ besteht der einfache Zusammenhang

$$\mathcal{F}_B = \{|f|,\ f \in \mathcal{F}_A\}.$$

Unter den gleichen Voraussetzungen an die Verteilung von X_1 erfüllt ein ϵ-Ball $\overline{B_{h,\epsilon}} \subset \mathcal{F}_B$ eine Relation von der Form (7.7). Zu jeder Menge

$$[l_A, u_A] = [f_A(\cdot - a), f_A(\cdot - b)] \subset \mathcal{F}_A$$

mit $\|u_A - l_A\|_r < \epsilon$ können Funktionen

$$l_B(x) := \inf\{f(x) = f_B(x-t),\ t \in [a,b]\},$$
$$u_B(x) := \sup\{f(x) = f_B(x-t),\ t \in [a,b]\}$$

gewählt werden. l_B, u_B sind zwar nicht mehr in $\mathcal{F}_B$ enthalten, aber eine analytische Betrachtung zeigt, dass

$$|u_B(x) - l_B(x)| \leq |u_A(x) - l_A(x)|.$$

Aufgrund der Monotonie der Norm $\|\cdot\|_r$ folgt damit

$$\|u_B - l_B\|_r \leq \|u_A - l_A\|_r < \epsilon.$$

Aus einer minimalen Überdeckung der Klasse $\mathcal{F}_A$ mit Mengen $[l_A, u_A]$, so dass

$$\|u_A - l_A\|_r < \epsilon,$$

kann also eine Überdeckung der Klasse $\mathcal{F}_B$ mit Mengen $[l_B, u_B]$, so dass $\|u_B - l_B\|_r < \epsilon$, konstruiert werden. Demnach kann die Überdeckungszahl $N_{[]}(\epsilon, \mathcal{F}_B, \|\cdot\|_r)$ abgeschätzt werden durch

$$N_{[]}(\epsilon, \mathcal{F}_B, \|\cdot\|_r) \leq N_{[]}(\epsilon, \mathcal{F}_A, \|\cdot\|_r).$$

Mittels der polynomiellen Abschätzung der Überdeckungszahlen $N_{[]}(\epsilon, \mathcal{F}_A, \|\cdot\|_r)$ in (7.10) und der Rechnung in (7.11) folgt

$$\Xi(\mathcal{F}_B, p) < \infty.$$

Damit erfüllt auch $\mathcal{F}_B$ die Voraussetzung (6.18) aus Theorem 6.12. Ein weiteres Beispiel liefern folgende Funktionenklassen, welche in enger Beziehung zur Teststatistik der Kolmogorov-Smirnov-Tests stehen. Sei

$$\Lambda^1 := \left\{ f : \mathbb{R} \longrightarrow \mathbb{R},\ f(x) = \sum_{i=1}^{n} c_i \mathbb{1}\{I_i\}(x) \right\},$$

wobei, und $I_i \subset \mathbb{R}$ Intervalle und $c_i \in \mathbb{R}$ reelle Zahlen bezeichnen. Die Klasse Λ^1 ist in der Klasse der Treppenfunktionen enthalten. Analog zum ersten Beispiel sei

$$f_C : \mathbb{R} \longrightarrow \mathbb{R},\ f_C(x) := \mathbb{1}\{[0,\infty)\}(x)$$

und

$$\begin{aligned} \mathcal{F}_C &:= \{f_C(\cdot - t),\ t \in \mathbb{R}\}, \\ \mathcal{F}_0 &:= \{f\mathbb{1}\{B\},\ |f| \le 1,\ f \in \Lambda^1,\ B \in \mathcal{B}(\mathbb{R})\}. \end{aligned} \tag{7.14}$$

Auch die Klasse $\mathcal{F}_0$ erfüllt die Voraussetzungen i) - iv) aus Theorem 6.12 und $\mathcal{F}_C \subset \mathcal{F}_0$. Ist für die Funktionenklasse $\mathcal{F}_C$ aus (7.14) die Relation (7.3) gegeben, so kann wie für die Funktionenklasse aus (7.6) argumentiert werden. Auch hier ist auf der Klasse $\mathcal{F}_C$ in natürlicher Weise eine Ordnungsrelation gegeben, aufgrund dieser (unter milden Voraussetzungen) jeder Ball $B_{f_0,\epsilon}$ durch eine Menge $[l, u]$ überdeckt werden kann, so dass $\|u - l\|_r < 3\epsilon$. Darüber hinaus ist $\mathcal{F}_C$ aus (7.14) ebenfalls eine VC-Klasse mit $V(\mathcal{F}_C) = 2$. Mit obiger Argumentation kann damit gezeigt werden, dass in diesem Fall gleichermaßen die Bedingung (6.18) aus Theorem 6.12 erfüllt ist.

7.3 Anwendung des Funktionalen Zentralen Grenzwertsatzes

In diesem Abschnitt soll demonstriert werden, wie mittels Theorem 6.12 asymptotische Resultate gewonnen werden können. Die beispielhaft betrachteten Funktionenklassen $\mathcal{F}_B, \mathcal{F}_C$ in (7.13) bzw. (7.14) können dazu dienen, die asymptotische Verteilung von Statistiken zu ermitteln. Angenommen, eine stationäre, reellwertige Folge $(X_i)_{i\in\mathbb{Z}}$ mit der Randverteilungsfunktion F erfülle zusammen mit den Funktionenklassen aus (7.14) die Voraussetzungen aus Theorem 6.12. Wegen $E[\mathbb{1}\{X_i \le t\}] = P(X_i \le t) = F(t)$ liefert die Definition der Funktion f_C

$$n^{1/2}(F_n(t) - F(t)) = G_n(f_C(\cdot - t)), \tag{7.15}$$

wobei F_n die empirische Verteilungsfunktion der Zufallsvariablen $X_1, ..., X_n$ bezeichnet. Nach Theorem 6.12 konvergieren die empirischen Prozesse $t \mapsto n^{1/2}(F_n(t) - F(t))$

schwach gegen einen zentrierten Gaußprozess G' auf $\mathbb{R}$ mit der Kovarianzstruktur

$$\begin{aligned}
Cov[G'(s), G'(t)] &:= Cov[G(f_C(\cdot - t)), G(f_C(\cdot - s))] \\
&= \frac{1}{2} \sum_{k \in \mathbb{Z}} Cov[f_C(X_1 - t), f_C(X_k - s)] \\
&\quad + Cov[f_C(X_k - t), f_C(X_1 - s)] \\
&= \frac{1}{2} \sum_{k \in \mathbb{Z}} Cov[\mathbb{1}\{X_1 \leq t\}, \mathbb{1}\{X_k \leq s\}] \\
&\quad + Cov[\mathbb{1}\{X_k \leq t\}, \mathbb{1}\{X_1 \leq s\}] \\
&= \frac{1}{2} \sum_{k \in \mathbb{Z}} P(X_1 \leq t, X_k \leq s) - P(X_1 \leq t) P(X_k \leq s) \\
&\quad + P(X_k \leq t, X_1 \leq s) - P(X_k \leq t) P(X_1 \leq s).
\end{aligned} \tag{7.16}$$

Die Supremumsnorm $\| \cdot \|_{\mathcal{F}_C}$ als Abbildung von $l^\infty(\mathcal{F}_C)$ nach $\mathbb{R}$ ist stetig, da diese die Topologie auf $l^\infty(\mathcal{F}_C)$ definitionsgemäß induziert. Die Relation

$$n^{1/2} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| = \sup_{t \in \mathbb{R}} |G_n(f_C(\cdot - t))| = \|G_n(f)\|_{\mathcal{F}_C}$$

impliziert zusammen mit dem Satz von der stetigen Abbildung (vgl. van der Vaart (1998), Theorem 18.11), dass auch die Zufallsvariable $n^{1/2} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$ schwach konvergiert:

$$n^{1/2} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \rightsquigarrow \|G'\|_\infty, \tag{7.17}$$

wobei die Kovarianzstruktur des zentrierten Gaußprozesses G' durch (7.16) gegeben ist. Durch (7.17) ist damit die asymptotische Verteilung der Teststatistik $n^{1/2} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$ des Kolmogorov-Smirnov-Tests unter Abhängigkeitsbedingungen wie in Theorem 6.12 bestimmt. Viele Schätzer lassen sich darstellen als Funktionale T, ausgewertet an der empirischen Verteilungsfunktion F_n. Die Kenntnis des asymptotischen Verhaltens der Folge $n^{1/2}(F_n - F)$ kann die Möglichkeit bieten, das asymptotische Verhalten vieler Schätzer der Form $T(F_n)$ mittels einer funktionalen Entsprechung der Delta-Methode zu ermitteln (vgl. van der Vaart (1998), Theorem 20.8). Ist das Funktional T in gewissem Sinne glatt (Hadamard-differenzierbar), dann ist unter einigen Regularitätsbedingungen die asymptotische Verteilung der Folge $n^{1/2}(T(F_n) - T(F))$ durch die Verteilung der Zufallsvariable $T'_F(G')$ gegeben, wobei T'_F die Hadamard-Ableitung des Funktionals T an der Stelle F (die dann mit der Gâteaux-Ableitung übereinstimmt) bezeichnet (vgl. van der Vaart (1998), Chapter 20.2). Diese Methode ist nicht eingeschränkt auf empirische Verteilungsfunktionen. Sei H_n die Abbildung, die einer Funktion $f \in \mathcal{F}$ die Zufallsvariable $n^{-1}\sum_{i=}^{n} f(X_i)$ zuordnet, und entsprechend $H(f) := E[f(X_1)]$. Sei T ein Funktional, welches die Abbildungen H_n bzw. H beispielsweise in $\mathbb{R}$ abbildet, so dass durch $T(H_n)$ eine Folge reellwertiger Zufallsvariablen gegeben ist. Falls $G_n = n^{1/2}(H_n - H) \rightsquigarrow G$, liefert die funktionale Delta-Methode unter einigen Regularitätsbedingungen durch

$$n^{1/2}(T(H_n) - T(H)) \rightsquigarrow T'_H(G)$$

die asymptotische Verteilung der Folge $T(H_n)$. Mit der funktionalen Delta-Methode kann beispielsweise die asymptotische Verteilung der Teststatistik des Mann-Whitney-U-Tests ermittelt werden. Dazu kann die Argumentation aus Example 20.11 aus van der Vaart (1998) Punkt für Punkt übertragen werden. Vorausgesetzt $X_1, X_2, ...$ und $Y_1, Y_2, ...$ sind voneinander unabhängige Folgen reellwertiger Zufallsvariablen, welche in Bezug auf die Funktionenklassen aus (7.14) die Voraussetzungen aus Theorem 6.12 erfüllen. Mit F^X bzw. F^Y seien die Verteilungsfunktionen von X_i bzw. Y_i und mit F_n^X bzw. F_n^Y die empirischen Verteilungsfunktionen von $X_1, ..., X_n$ bzw. $Y_1, ..., Y_n$ bezeichnet. Aus Gründen der Einfachheit werden gleich große Stichprobenumfänge n betrachtet. Wird $T(F, G) := \int_{\mathbb{R}} F(t)dG(t)$ gesetzt (wobei mit F, G Verteilungfunktionen auf $\mathbb{R}$ bezeichnet werden), so besitzt die Teststatistik des Mann-Whitney-U-Tests die Darstellung

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{1}\{X_i \leq Y_j\} = T(F_n^X, F_n^Y).$$

Das Funktional T ist Hadamard-differenzierbar, und die asymptotische Verteilung von

$$(2n)^{1/2}\left(T(F_n^X, F_n^Y) - T(F^X, F^Y)\right)$$

kann mit der funktionalen Delta-Methode bestimmt werden. In diesem Fall ist sie eine Normalverteilung.

Ein weiteres Beispiel liefert die Betrachtung der Funktionenklassen $\mathcal{F}_B, \mathcal{F}_0$ aus (7.13). Diese kann hilfreich sein, um die asymptotische Verteilung von Folgen von der Form

$$M_n = \frac{1}{n}\sum_{i=1}^{n} |X_i - \overline{X}| \tag{7.18}$$

zu untersuchen. Im Falle der Unabhängigkeit der Zufallsvariablen $X_1, X_2, \ldots$ kann die asymptotische Verteilung der Folge M_n durch eine Argumentation über Donsker-Klassen ermittelt werden (vgl. van der Vaart (1998), Example 19.25). Diese Argumentation kann mittels Theorem 6.12 nicht Punkt für Punkt auf Fälle von Abhängigkeit übernommen werden. Das liegt im Wesentlichen an der geforderten Beschränkung der Funktionen der Klasse $\mathcal{F}_0$ (Voraussetzung i)). Deshalb soll die Statistik in (7.18) ein wenig abgeändert werden.

Sei $(X_i)_{i\in\mathbb{Z}}$ eine stationäre, reellwertige Folge von Zufallsvariablen, die zusammen mit den Funktionenklassen aus (7.13) die Voraussetzungen aus Theorem 6.12 erfüllt. Die Verteilungsfunktion von X_i wird mit F bezeichnet. Zusätzlich gelte

$$\begin{aligned} &\theta \mapsto E[|f_A(X_1 - f_A(\theta))|] \quad \text{differenzierbar bei} \quad \theta = 0, \\ &\overline{X} = o_P(1) \quad \text{und} \quad n^{1/2}\left(f_B\left(\overline{X}\right) - \frac{1}{n}\sum_{i=1}^{n} f_B\left(X_i\right)\right) = o_P(1). \end{aligned} \tag{7.19}$$

Die erste Bedingung an $\overline{X}$ kann bei Bedarf noch leicht abgeschwächt werden durch $f_A(\overline{X}) = o_P(1)$. Die letzte Bedingung in (7.19) ist beispielsweise erfüllt, wenn $X_i \leq k \ \forall i$ fast sicher, wobei k den Wert aus der Definition der Funktion f_B in (7.12) bezeichnet. Statt nun die Folge M_n zu betrachten, soll die Folge

$$M_n' = \frac{1}{n}\sum_{i=1}^{n} f_A(|X_i - f_A(\overline{X})|) = \frac{1}{n}\sum_{i=1}^{n} f_B(X_i - f_A(\overline{X}))$$

untersucht werden. Die Funktion f_A aus (7.5) bildet betragsmäßig kleine Werte auf sich selbst ab, verändert diese also nicht, und ersetzt nur betragsmäßig große Werte durch k bzw. $-k$. Wenn $X_i \leq k \ \forall i$ fast sicher, so folgt also $M_n = M_n'$ fast sicher. Zunächst soll gezeigt werden, dass

$$G_n(f_A(|\cdot - \overline{X}|)) = G_n(f_A(|\cdot|)) + o_P(1), \tag{7.20}$$

wobei $\overline{X}$ im linken Term von (7.20) als Konstante zu verstehen ist. Wegen der Dreiecksungleichung und der Monotonie der Funktion f_A gilt

$$\begin{aligned} f_A(|x-c|) - f_A(|x|) &\leq f_A(|x| + |c|) - f_A(|x|) \\ &\leq f_A(|x|) + f_A(|c|) - f_A(|x|) \\ &= f_A(|c|), \end{aligned} \tag{7.21}$$

wobei die letzte Ungleichung aus einer analytischen Betrachtung der Funktion f_A folgt. Sei

$$f_n(x) := f_A(|x - \overline{X}|), \quad f_0(x) := f_A(|x|),$$

wobei f_n als zufällige Funktion zu verstehen ist. Der zufällige Erwartungswert $E[(f_n(X_1) - f_0(X_1))^2]$ konvergiert stochastisch gegen 0. Wegen Ungleichung (7.21) und der Forderung (7.19) folgt

$$E[(f_n(X_1) - f_0(X_1))^2] \leq E[(f_A(|\overline{X}|))^2] = (f_A(|\overline{X}|))^2 = o_P(1). \tag{7.22}$$

Nach Theorem 6.12 konvergiert G_n schwach gegen einen Gaußprozess G auf $\mathcal{F}_B$. Eine genaue Analyse des Beweises von Lemma 19.24 in van der Vaart (1998) für den Unabhängigkeitsfall zeigt, dass die Forderung, die Funktionenklasse sei eine Donsker-Klasse, ersetzt werden kann durch die Voraussetzungen in Theorem 6.12. (Hier ist zu beachten, dass $f_A(|x|) = f_B(x)$ und dass an dieser Stelle mit der Funktionenklasse $\mathcal{F}_B$ gearbeitet wird.) Verwendet wird lediglich die schwache Konvergenz von G_n gegen einen (stetigen) Gaußprozess G, wie sie durch Theorem 6.12 zusammen mit Lemma 18.15 in van der Vaart (1998) gewährleistet ist. Die Relation (7.20) folgt nun mittels obiger Abänderung von Lemma 19.24 aus van der Vaart (1998) und aus (7.22). Die klassische Delta-Methode (vgl. Theorem 3.1, van der Vaart (1998)) und (7.22) implizieren mit $H_n := f_A(\overline{X})$, $T(\theta) := E[|f_A(X_1 - f_A(\theta))|]$ und $\theta_0 := 0$, dass

$$\begin{aligned}
&n^{1/2}\left(T(H_n) - T(\theta_0)\right)\\
&\quad= n^{1/2}\left(E[f_A(|X_1 - f_A(\overline{X})|)] - E[f_A(|X_1|)]\right)\\
&\quad= n^{1/2}\left(E[|f_A(X_1 - f_A(\overline{X}))|] - E[|f_A(X_1)|]\right)\\
&\quad= \frac{\partial}{\partial\theta} E[|f_A(X_1 - f_A(\theta))|]\Big|_{\theta=0}\; n^{1/2} f_A(\overline{X}) + o_P(1)\\
&\quad= \frac{\partial}{\partial\theta} \int_{\mathbb{R}} |f_A(t-\theta)| dF(t)\Big|_{\theta=0}\; n^{1/2} f_A(\overline{X}) + o_P(1)\\
&\quad= \frac{\partial}{\partial\theta} \int_{\mathbb{R}} f_A(|t-\theta|) dF(t)\Big|_{\theta=0}\; n^{-1/2} \sum_{i=1}^{n} f_A(X_i)) + o_P(1)\\
&\quad= \frac{\partial}{\partial\theta} \int_{\mathbb{R}} f_A(|t-\theta|) dF(t)\Big|_{\theta=0}\; G_n(f_A(\cdot)) + o_P(1),
\end{aligned} \tag{7.23}$$

wobei $\overline{X}$ in den Erwartungswerten von (7.23) wiederum als Konstante zu verstehen ist. Nun kann die asymptotische Verteilung der Folge $V_n := n^{1/2}(M_n' - E[f_A(|X_1|)])$

ermittelt werden:

$$
\begin{aligned}
V_n &= n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} f_A(|X_i - f_A(\overline{X})| - E[f_A(|X_1|)])\right) \\
&= n^{1/2}\big(E[f_A(|X_i - f_A(\overline{X})|)] - E[f_A(|X_1|)]\big) + G_n(f_A(|\cdot - \overline{X}|)) \\
&= \frac{\partial}{\partial\theta}\int_{\mathbb{R}} f_A(|t-\theta|)dF(t)\bigg|_{\theta=0} G_n(f_A(\cdot)) + G_n(f_A(|\cdot|)) + o_P(1) \\
&= G_n(h) + o_P(1) \rightsquigarrow G(h)
\end{aligned} \tag{7.24}
$$

wegen (6.11), wobei die Funktion h definiert ist durch

$$
h(x) := \frac{\partial}{\partial\theta}\int_{\mathbb{R}} f_A(|t-\theta|)dF(t)\bigg|_{\theta=0} f_A(x) + f_A(|x|).
$$

Sowohl die Funktion $f_A(\cdot)$ als auch die Funktion $f_B(\cdot) = f_A(|\cdot|)$ sind in $\mathcal{F}_0$ aus (7.13) enthalten und somit auch die Funktion h (eventuell bis auf einen konstanten Vorfaktor). Nach Theorem 6.8 folgt damit

$$
n^{1/2}(M'_n - E[f_A(|X_1|)]) \rightsquigarrow N\left(0, \sum_{k\in\mathbb{Z}} Cov[h(X_1), h(X_k)]\right), \tag{7.25}
$$

weil die asymptotische Normalverteilung von $G_n(f)$ für jedes feste $f \in \mathcal{F}_0$ gewährleistet ist.

Kapitel 8

Zusammenfassung und Ausblick

In dieser Arbeit wurde das Verhalten statistischer Methoden bei beschränkter Verletzung stochastischer Unabhängigkeit untersucht. Um stochastische Abhängigkeit zu beschränken wurden in Kapitel 3 zunächst Maße für Abhängigkeit eingeführt, um stochastische Abhängigkeit zu quantifizieren. Betrachtet wurden sowohl klassische Abhängigkeitsmaße, wie sie in der Zeitreihenanalyse unter dem Begriff „strong mixing“ auftreten, als auch neuere Abhängigkeitsmaße, die auf bedingten Verteilungsfunktionen basieren. Bekannte Konzepte aus der Literatur wurden variiert, um weitere Abhängigkeitsmaße zu konstruieren.
In Kapitel 4 wurden Auswirkungen infinitesimaler Abweichungen vom Unabhängigkeitsfall bei endlichen Stichproben untersucht. Es konnte für einige Abhängigkeitsmaße gezeigt werden, dass hinreichend kleine Verletzungen stochastischer Unabhängigkeit beliebig kleine Auswirkungen auf eine Vielzahl statistischer Verfahren haben.
Die nachfolgenden Kapitel behandeln das asymptotische Verhalten von Schätzerfolgen bei stochastischer Abhängigkeit. In Kapitel 5 wurden Schätzerfolgen auf Konsistenz hin untersucht. Ein wesentlicher Schwerpunkt lag dabei in der Betrachtung der empirischen Verteilungsfunktion. So konnten Bedingungen für stochastische Abhängigkeit gefunden werden, so dass die fast sichere, gleichmäßige Konvergenz der empirischen Verteilungsfunktion im Sinne des Glivenko-Cantelli-Satzes wie im Unabhängigkeitsfall sichergestellt ist. Darauf aufbauend konnte die Konsistenz für Klassen von Schätzerfolgen, die bereits aus der robusten Statistik bekannt sind, bei Abhängigkeit gezeigt werden.
In Kapitel 6 wurden die asymptotischen Verteilungen von Schätzerfolgen bei Abhängigkeit betrachtet. Es konnte ein Zentraler Grenzwertsatz gezeigt werden, bei dem die Ab-

hängigkeitsbedingung eines bereits bekannten Zentralen Grenzwertsatzes um den Preis anderer Bedingungen abgeschwächt wurde. Aufbauend auf einer Ungleichung vom Typ der Bernstein-Ungleichung und einer bekannten Beweistechnik mittels Überdeckungszahlen wurde auch ein Funktionaler Zentraler Grenzwertsatz für empirische Prozesse bei Verletzung stochastischer Unabhängigkeit bewiesen.
In Kapitel 7 wurde an Beispielen demonstriert, wie die Resultate aus Kapitel 6 verwendet werden können, um asymptotisches Verhalten von statistischen Verfahren zu untersuchen. Mit einem Zentralen Grenzwertsatz kann beispielsweise der Einstichproben-t-Test genauer beleuchtet werden, da damit die asymptotische Verteilung der t-Teststatistik bei stochastischer Abhängigkeit hergeleitet werden kann. Ein Funktionaler Zentraler Grenzwertsatz wie in Kapitel 6 erweist sich als hilfreich, um die asymptotische Verteilung einer Reihe von Schätzern zu ermitteln. Das liegt insbesondere daran, dass damit unter einigen Bedingungen auch die empirische Verteilungsfunktion auf schwache Konvergenz untersucht werden kann.

Die Resultate dieser Arbeit eröffnen weitere Themenbereiche. Es stellt sich die Frage, ob der Beweis des Zentralen Grenzwertsatzes unter leicht abgeänderten Bedingungen auf Folgen von Zufallsvektoren übertragbar ist. Der Nachweis der Bedingungen des Zentralen- und des Funktionalen Zentralen Grenzwertsatzes stellt eine weitere Schwierigkeit dar. Methoden zum Nachweis dieser Bedingungen könnten die Verwendung dieser Sätze vereinfachen. Hilfreich wären hierbei insbesondere Relationen zwischen Abhängigkeitsmaßen bzw. Schranken für Abhängigkeit, wie sie in obigen Grenzwertsätzen auftreten, und stochastischen Modellen.
Ein immer wieder auftretendes Problem bei asymptotischen Aussagen ist die Ermittlung von Konvergenzraten. Für praktische Anwendungen wäre Methoden zur Bestimmung von Stichprobenumfängen hilfreich, bei denen die asymptotischen Betrachtungen eine hinreichend gute Approximation darstellen. Dies gilt insbesondere für die Resultate aus Kapitel 6.
Wie in Kapitel 5 dargelegt, erweisen sich einige Schätzer in Bezug auf Konsistenz in gewisser Weise als unempfindlich gegenüber stochastischer Abhängigkeit. Wünschenswert wären weitere Methoden zur Konstruktion von Schätzern und statistischen Verfahren, die gegenüber Verletzungen stochastischer Unabhängigkeit im Sinne von Abhängigkeitsmaßen unempfindlich sind.
Dazu sollte bei fest vorgegebener oberer aber kleiner Schranke für Abhängigkeit bei ei-

ner endlichen Stichprobe die „schlimmste“ Auswirkung bzw. die „größte Abweichung“ vom Verhalten bei Unabhängigkeit ermittelt werden. Aus dieser Betrachtung könnte ein Konzept, das mit dem Konzept der Influenzfunktion aus der robusten Statistik vergleichbar wäre, entwickelt werden.
Sowohl die Verletzung stochastischer Unabhängigkeit als auch Ausreißer können die Resultate statistischer Methoden verzerren. Beide Formen von Störungen können möglicherweise auch kombiniert auftreten. Deswegen wären weiterführende Untersuchungen von Auswirkungen auf statistische Verfahren interessant, die aus Kombination stochastischer Abhängigkeit und dem Auftreten von Ausreißern herrühren.

Symbolverzeichnis

u.i.v.	unabhängig identisch verteilt	5
$\rightsquigarrow$	konvergiert schwach	6
$\longrightarrow_{st}$	Konvergenz in Wahrscheinlichkeit bzw. stochastische Konvergenz	33
$o_P(R_n), O_P(R_n)$	Landau-Symbole für Zufallsfolgen	34
$o(r_n), O(r_n)$	Landau-Symbole für Zahlenfolgen	34
$X \sim F$	X ist verteilt nach F	6
P^X	Bildmaß einer Zufallsvariablen bzw. eines Zufallsvektors X	21
$P^1 \otimes P^2$	Produktverteilung der Verteilungen P^1 und P^2	27
$a_n \sim b_n$	asymptotisch äquivalente Zahlenfolgen a_n, b_n	55
$\mathcal{B}(\mathbb{R}^v)$	Borel-σ-Algebra auf $\mathbb{R}^v$	10
$\sigma(X), (\sigma((X_t)_{t\in T}))$	durch eine oder mehrere Zufallsvariablen induzierte σ-Algebra	9
$E[X]$	Erwartungswert einer Zufallsvariablen bzw. eines Zufallsvektors $X \in L_1$	5
$Var[X]$	Varianz einer reellwertigen Zufallsvariablen $X \in L_2$	6
$Cov[X,Y]$	Kovarianz zwischen zwei reellwertigen Zufallsvariablen $X, Y \in L_2$	12
$Corr[X,Y]$	Korrelation zwischen zwei reellwertigen Zufallsvariablen $X, Y \in L_2$	12
$\mathbb{1}A,\ \mathbb{1}\{A\}$	Indikatorfunktion von A	7
$Lip(g)$	Lipschitzkonstante einer Funktion g	14
$[t]$	kleinste ganze Zahl, nicht kleiner als t	10
$a \vee b,\ a \wedge b$	Maximum bzw. Minimum von a und b	15
T^c	Komplement einer Menge $T \subset \mathbb{T}$	29

$[s,t], (s,t]$	Intervalle bzw. multivariate Intervalle zu Zahlen bzw. Vektoren s, t	16
$[u,l]$	bracket für Funktionen u, l bzgl. einer Funktionenklasse $\mathcal{F}$	69
$\|N\|$	Kardinalität einer Menge N	42
$N_{[]}(\epsilon,\mathcal{F},p),\ N(\epsilon,\mathcal{F},d)$	Überdeckungszahlen	69
$V(\mathcal{D})$	VC-Dimension eines Mengensystems $\mathcal{D}$	96
$V(\mathcal{F})$	VC-Dimension einer Klasse von Funktionen $\mathcal{F}$	96
L_p	Menge der L_p-integrierbaren Zufallsvariablen	13
$l^\infty(\mathcal{F})$	Raum der beschränkten Abbildungen von $\mathcal{F}$ nach $\mathbb{R}$	66
$\\|\cdot\\|_p$	L_p-Norm	13
$\\|\cdot\\|_\infty$	Supremumsnorm	17
$\\|\cdot\\|_{\mathcal{F}}$	Supremumsnorm auf $l^\infty(\mathcal{F})$	66

Literaturverzeichnis

Andrews, D. W. K., Pollard, D. (1994). An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes. *Int. Stat. Rev.*, 62,1:119–132.

Bauer, H. (2002). *Wahrscheinlichkeitstheorie.* Walter de Bruyter, Berlin, New York.

Bernstein, S. N. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.*, 97:1–59.

Billingsley, P. (1999). *Convergence of Probability Measures, 3rd edition.* Wiley, New York.

Billingsley, P. (1995). *Probability and Measure.* Wiley, New York.

Bradley, R. C. (2005). Basic Proberties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probab. Surv.*, 2:107–144.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions, Vol 1 - 3.* Kendrick Press, Heber City.

Cantelli, F. P. (1933). Sulla determinazione empirica delle leggi di probablità *Giornale dell'Instituto Italiano degli Attuari*, 4:421–424.

Dedecker, J., Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields*, 132:203–236.

Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* 23:277–281.

Doukhan, P., Louhichi, S. (1999). A new weak dependence condition and application to moment inequalities. *Stochastic Process. Appl.* 84:313–342.

Doukhan, P., Massart, P., Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincar´e Probab. Statist.* 31:393–427.

Doukhan, P., Neumann, M.H. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Process. Appl.* 117:878–903.

Dudley, R. M. (2002). *Real Analysis and Probability.* Cambridge University Press.

Dvoretzky, A., Kiefer, J., Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and the classical multinomial estimator. *Ann. Math. Statist.*, 33:624–669.

Glivenko, V. (1933). Sulla determinazione empirica delle leggi di probablità. *Giornale dell'Instituto Italiano degli Attuari*, 4:92–99.

Gnedenko, B. V., Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variavles.* Addison-Wesley, Reading, Massachusetts.

Hall, P., Heyde, C.C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.* 12:141–153.

Huber, P., J. (2004). *Robust Statistics.* Wiley, Hoboken.

Kallenberg, O. (2002). *Foundations of Modern Probability, 2nd edition.* Springer, New York.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *Ann. Probab.*, 18:1269–1283.

Nelsen, R., B. (2006). *An Introduction to Copulas, 2nd edition.* Springer, New York.

Ossiander, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.*, 15:897–919.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci.*, 42:43–47.

Tucker, H. G. (1959). A Generalization of the Glivenko-Cantelli Theorem. *Ann. Math. Statist.*, 30:828–830.

van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press.

van der Vaart, A.W., Wellner, J.A. (1996). *Weak convergence and empirical Processes.* Springer, Berlin.

Vapnik, V. N., Červonenenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, 16:264–280.