Eine Analyse bedingter Tests mit bedingten Zentralen Grenzwertsätzen für Resampling-Statistiken

Inaugural-Dissertation

zur

Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Markus Pauly

aus Düsseldorf

Januar 2009

Aus dem Mathematischen Institut der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. A. Janssen

Koreferenten: Apl. Prof. Dr. H. Finner

Prof. Dr. E. Häusler

Tag der mündlichen Prüfung: 12. Dezember 2008

Zusammenfassung

Diese Arbeit befasst sich mit der Qualität von bedingten Resampling-Tests φ_n^* , wie Bootstrap- oder Randomisationstests, die für nichtparametrische Nullhypothesen konstruiert werden. Solche Tests benötigen keine Verteilungsannahme, können aber im Vergleich zum parametrischen Test an Güte verlieren.

Verglichen werden diese Tests mit unbedingten Tests φ_n , die für eine eingeschränkte, z.B. parametrische, Nullhypothese \mathcal{P}_0 asymptotisch das Fehlerniveau einhalten und gute asymptotische Güteeigenschaften aufweisen. Die unbedingten Tests φ_n besitzen jedoch oft den Nachteil, dass sie für einen festen Stichprobenumfang n den Fehler 1. Art nicht allgemein (nichtparametrisch) kontrollieren können. Deshalb wird vom bedingten Resampling-Test φ_n^* nun gefordert, dass dieser wenigstens asymptotisch effektiv im Bezug auf die Meßlatte φ_n ist, d.h. für $n \to \infty$ liegt Konvergenz

$$E_P(|\varphi_n - \varphi_n^*|) \to 0 \text{ für alle } P \in \mathcal{P}_0$$
 (1)

vor. Zum Nachweis dieser Eigenschaft benötigt man dabei sogenannte bedingte Zentrale Grenzwertsätze für die Resampling-Version der Teststatistik gegeben den Daten. Diese werden aufbauend auf den Arbeiten von Janssen et. Al., vgl. [33] und [35], mit Hilfe eines gewichteten Bootstrap-Ansatzes, der die simultane Behandlung von verschiedenen Resampling-Verfahren erlaubt, in den Kapiteln 3 und 4 hergeleitet. Diese Sätze werden zuerst für (insbesondere multivariate) lineare und dann asymptotisch lineare Statistiken sowie für Quadratische Formen in jeweils allgemeinen Dreiecksschemata von Zufallsvariablen untersucht. Als Anwendung hiervon werden anschließend studentisierte Resampling-Tests φ_n^* für konkrete nichtparametrische Ein- und Zweistichprobenprobleme konstruiert, wobei neben dem Nachweis der Eigenschaft (1) auch die Gütefunktion dieser Tests untersucht wird. So erhält man beispielsweise, dass sich der (parametrische) F-Test zum Vergleich zweier Varianzen im nichtparametrischen Fall ohne asymptotischen Qualitätsverlust als studentisierter Permutationstest ausführen lässt. Der letzte Teil dieser Arbeit stellt verschiedene Resampling-Verfahren dar, die sich in den gewichteten Bootstrap-Ansatz integrieren lassen. Zu den bekannteren Methoden zählen dabei das Permutations-, das m(n)- und m(n)-doublesowie das Wild-Bootstrap-Verfahren.

Im abschließenden Anhang werden die im Rahmen dieser Arbeit verwendeten wesentlichen Hilfsmittel bereitgestellt.

Abstract

This thesis deals with the quality of conditional resampling tests φ_n^* , such as bootstrap and randomization tests that are established for nonparametrical null-hypothesis. Such tests are distribution-free but can lose power in comparison with parametrical tests.

These tests are compared with unconditional tests φ_n , which are asymptotical α -level tests for a restricted, possibly parametrical, null-hypothesis \mathcal{P}_0 and posses good asymptotical power qualities. However, the unconditional tests often have the disadvantage that the true alpha error probability can in general not be controlled (nonparametrically) for fixed sample sizes n.

Therefore the conditional resampling test φ_n^* requires at least to be asymptotically effective with respect to the benchmark φ_n , viz. the convergence

$$E_P(|\varphi_n - \varphi_n^*|) \to 0 \text{ for all } P \in \mathcal{P}_0$$
 (1)

holds for $n \to \infty$. To proof this property so-called conditional central limit theorems for the resampling version of the test statistic given the data are required. Based on the articles of Janssen et al., see [33] and [35], these are derived in chapters 3 and 4 by means of a weighted bootstrap approach that allows the simultaneous treatment of different resampling procedures. Under fairly general assumptions about the underlying triangular arrays of random variables these theorems are first established for (especially multivariate) linear statistics. In a second step asymptotically linear statistics as well as quadratic forms are studied.

As application studentized resampling tests for specific nonparametric one- and two-sample problems are presented. In addition to the verification of property (1) the power functions of these tests are discussed. In this way it is for instance shown that the (parametric) F-Test for comparing two variances can be carried out in a nonparametrical setting without asymptotical loss of power as a studentized permutation test.

The last part of this thesis demonstrates different resampling methods which are covered by the weighted bootstrap approach. The permutation procedure, the m(n) and m(n)-double as well as the wild bootstrap are thereby ranked among the more popular methods.

In an concluding appendix the essential tools that are used within this work are provided.

Inhaltsverzeichnis

	EIII	leitung	1		
2	${f Res}$	Resampling-Verfahren - Die Grundlagen			
	2.1	Motivation	6		
	2.2	Das Permutationsverfahren	7		
	2.3	Das Bootstrap-Verfahren	8		
		2.3.1 Gegenbeispiele und Erweiterungen	9		
	2.4	Allgemeines Resampling durch Gewichtsfunktionen	12		
	2.5	Resampling-Tests	14		
I liı		dingte Zentrale Grenzwertsätze für lineare und nichte Resampling-Statistiken und ihre Anwendungen	17		
3	\mathbf{Bec}	ingte Zentrale Grenzwertsätze für lineare Resampling-			
3		ingte Zentrale Grenzwertsätze für lineare Resampling- istiken	19		
3			19 21		
3	Sta	zistiken			
3	Sta	cistiken Der eindimensionale Fall			
3	Sta	Der eindimensionale Fall	21		
3	Sta 3.1	Der eindimensionale Fall	2129		
3	Sta 3.1	Der eindimensionale Fall	212936		
	Star 3.1 3.2	Der eindimensionale Fall	21 29 36 36		
$oldsymbol{4}$	Sta 3.1 3.2 Bed	Der eindimensionale Fall	21 29 36 36		
	Sta 3.1 3.2 Bed	Der eindimensionale Fall	21 29 36 36 39		

In	HALT	SVERZEICHNIS	II	
	4.2	Resampling Quadratischer Formen	. 51	
5	Explizite Anwendungen in der Testtheorie			
	5.1	Resampling-Tests für Einstichprobenprobleme	. 63	
	5.2	Resampling-Tests für Zweistichprobenprobleme $\dots \dots$. 73	
II	E	xkurs: Implementierbare Resampling-Verfahren	101	
6	Res	ampling-Verfahren für das weighted Resampling	103	
\mathbf{A}	Apj	pendix	124	
	A.1	Schwache Konvergenz und bedingte Verteilungen	. 124	
	A.2	Stationäre und ergodische Zufallsvariablen	. 130	
Sv	mhe	ol- und Ahkiirzungsverzeichnis	132	

134

 ${\bf Literaturverzeichnis}$

Kapitel 1

Einleitung

Die klassischen statistischen Schätz- und Testverfahren beruhen auf einer parametrischen Verteilungsannahme, wie z.B. einer Normalverteilungsannahme. Diese Annahme ist jedoch in vielen praktischen Situationen nicht gerechtfertigt. In allgemeineren sogenannten nichtparametrischen Modellen stellt sich dann aber unmittelbar die Frage, in welcher Weise valide statistische Tests und Konfidenzintervalle konstruiert werden können. Insbesondere geht es darum, Tests für eine (nichtparametrische) zusammengesetzte Nullhypothese zu konstruieren, die auch tatsächlich das vorgegebene Fehler-Niveau einhalten und dann als valide bezeichnet werden.

Eine Lösung dieses Problems besteht nun in der Anwendung von sogenannten Resampling-Verfahren. Zu den bekanntesten Verfahren zählen hierbei Bootstrapund Permutationsmethoden. Diese Methoden zählen seit der richtungsweisenden Arbeit von Efron [14] zum Standardwerkzeug der Statistik. Dabei geht die Entwicklung in der mathematischen Statistik Hand in Hand mit dem Fortschritt der Computertechnologie, da diese Techniken in der Regel mit rechenintensiven Monte Carlo Prozeduren umgesetzt werden.

Diese Methoden haben inzwischen in allen Wissenschaftsbereichen, in denen Statistik verwendet wird, Verwendung gefunden. Als Referenz kann dabei beispielsweise auf die Lehrbücher von Davison und Hinkley [10], Efron und Tibshirani [15], Manly [47] und die darin zitierten Arbeiten sowie auf den Artikel von Horowitz [28] verwiesen werden. Diese enthalten eine Vielzahl von Anwendungen der Bootstrap- und Permutationsmethoden in der Bioinformatik, Biologie, Medizin und Ökonometrie. Neben den oben genannten Büchern findet

Einleitung ______ 2

man in der Literatur mehrere tausend wissenschaftliche Arbeiten, die die Qualität solcher Verfahren untersuchen.

Um nun derartige praktische Situationen mathematisch zu beschreiben, fasst man einen gegebenen Datensatz (x_1, \ldots, x_n) als Realisierungen einer Zufallsvariablen $X = (X_1, \ldots, X_n)$ auf, wobei X auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) operiert. Ziel ist es nun, die unbekannte Verteilung P^{T_n} einer datenabhängigen Statistik $T_n = T_n(X)$ durch geeignete Methoden zu approximieren. Dies erfolgt bei Resampling-Verfahren dadurch, dass man aus den bekannten Daten X mittels eines speziellen Algorithmus eine neue Stichprobe X^* erzeugt, aus der man nun Informationen über die unbekannte Verteilung gewinnen kann. Beim klassischen Bootstrap-Verfahren von Efron [14] erhält man den neuen Datensatz $x^* = (x_1^*, \dots, x_n^*)$ durch n-maliges Ziehen mit Zurücklegen aus den gegebenen Daten x. Dabei wird jede Beobachtung x_i mit gleicher Wahrscheinlichkeit $\frac{1}{n}$ aus der Ausgangsstichprobe gezogen. Der Vorteil besteht nun darin, dass man die bedingte Verteilung der Bootstrap-Stichprobe X^* gegeben den Daten X kennt, so dass dort Aussagen über die gesuchte Verteilung getroffen werden, die dann im zweiten Schritt wieder auf das gegebene Experiment übertragen werden können.

Im Gegensatz zum obigen Verfahren, werden beim Permutationsverfahren, die neuen Daten durch Ziehen ohne Zurücklegen gewonnen, indem man für eine auf der symmetrischen Gruppe S_n gleichverteilten Permutation π die neuen Daten $(x_{\pi(1)}, \ldots, x_{\pi(n)})$ bestimmt. Die zu diesem Verfahren gehörenden Permutationstests haben sich insbesondere für Zweistichprobenprobleme als sehr nützlich erwiesen.

Die Anwendung dieser Resampling-Techniken kann nun durch den Vergleich der asymptotischen Grenzverteilung von T_n und der Resampling-Statistik $T_n^* := T_n(X^*)$ erfolgen. Stimmen diese überein, so halten beispielsweise die hiermit konstruierten Resampling-Tests ein vorgegebenes Fehler-Niveau asymptotisch ein und sind valide. Für die analytische Behandlung dieser Problemstellung benötigt man sogenannte bedingte Zentrale Grenzwertsätze (ZGS) für zufällige Maße.

In der Situation von unabhängig und identisch verteilten Zufallsvariablen X_1, \ldots, X_n findet man hierzu in der Literatur ausführliche Untersuchungen. Beispiels-

Einleitung ______ 3

weise konnte Mammen [46] zeigen, dass die Bootstrap-Version T_n^* der Mittelwertsstatistik aus dem klassischen ZGS genau dann einen bedingten ZGS erfüllt, wenn dies auch schon für die unbedingte Statistik T_n gilt. Ähnliche Aussagen wurden mittlerweile auch im Rahmen der empirischen Prozesstheorie gezeigt, vgl. z.B. Kapitel 2 (insbesondere Theorem 2.2.) bei GINÉ [19] oder auch Kapitel 3.6. bei VAN DER VAART UND WELLNER [62], so dass das Verhalten im i.i.d. Fall für den klassischen Bootstrap genauestens erforscht ist. Der Ansatz dieser Arbeit ist jedoch ein anderer. Hierbei wird

- das Verhalten von Resampling-Statistiken in allgemeinen Dreiecksschemata $(X_{n,i})_{i\leq n}$ untersucht,
- die unter Umständen auch voneinander abhängig sein können.
- Dabei wird ein *gewichteter* Ansatz verwendet, durch den mehrere Resampling-Verfahren *simultan* behandelt werden können.

Es wird sich zeigen, dass diese allgemeine Betrachtungsweise ein viel flexibleres Hilfsmittel zur Behandlung der verschiedensten testtheoretischen Fragestellungen darstellt.

Die vorliegende Arbeit gliedert sich in sechs Kapitel.

Nach einer kurzen Einleitung wird im zweiten Kapitel zunächst die Verwendung von Resampling-Verfahren im Rahmen der Testtheorie motiviert. Dabei liegt das Hauptaugenmerk erst einmal auf den oben angesprochenen Bootstrap- und Permutationsverfahren. Ausgehend hiervon wird die Notwendigkeit der Betrachtung weiterer Resampling-Verfahren mit Hilfe von einigen Beispielen aufgezeigt und so das gewichtete Resampling eingeführt. Zum Abschluss des Kapitels werden die hierzugehörenden Resampling-Tests genauer untersucht und dabei das oben angesprochene Bedürfnis zur Behandlung von bedingten ZGS auch mathematisch begründet.

Der sich anschließende mathematische Hauptteil leitet zuerst allgemeine bedingte ZGS für lineare und nicht-lineare Resampling-Statistiken her, um diese im Anschluss zur Konstruktion von validen Resampling-Tests für spezielle Testprobleme zu verwenden.

Dabei liegt in Kapitel 3 das Hauptaugenmerk auf dem asymptotischen Verhalten

Einleitung ______ 4

der linearen, gewichteten Resampling-Statistik

$$T_n^* := \sqrt{k(n)} \sum_{i=1}^{k(n)} W_{n,i} (X_{n,i} - \overline{X}_n). \tag{1.1}$$

Dort ist $(X_{n,i})_{i\leq k(n)}$ ein beliebiges Dreiecksschema von Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und $(W_{n,i})_{i\leq k(n)}$ ein weiteres, das Resampling-Verfahren beschreibendes Dreiecksschema von reellen Gewichtsfunktionen auf einem anderen Wahrscheinlichkeitsraum $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$.

Für reellwertige $(X_{n,i})_{i \leq k(n)}$ hat JANSSEN [35] bereits Bedingungen entwickelt, unter denen (1.1) einem bedingten ZGS genügt. Aufbauend hierauf werden zum einen Voraussetzungen geschaffen, unter denen für T_n^* auch die stärkere Konvergenz in der Mallows-Metrik vorliegt und zum anderen gezeigt, dass die obigen Bedingungen beispielsweise für spezielle Martingaldifferenzenschemata und stationäre Zufallsvariablen erfüllt sind. Im weiteren Verlauf wird dann auch das Verhalten von T_n^* für höhere Dimensionen untersucht.

Das sich anschließende Kapitel 4 befasst sich mit dem Resampling nicht-linearer Statistiken und gliedert sich in zwei Abschnitte. Hierbei wird im ersten Teil u.a. gezeigt, dass sich das unbedingte Konvergenzverhalten von asymptotisch linearen Statistiken auch auf das bedingte Verhalten unter suffizienten σ -Algebren übertragen lässt. Als Anwendung erhält man Bedingungen für die asymptotische Validität von allgemeinen bedingten Tests, zu denen auch die Permutationstests zählen.

Im zweiten Abschnitt des vierten Kapitels werden bedingte ZGS für Resampling-Versionen von quadratischen Statistiken entwickelt. Als Konsequenz hiervon erhält man insbesondere verteilungsfreie Anpassungstests für diskrete Zweistichprobenprobleme.

In Kapitel 5 werden dann die Ergebnisse aus den vorangegangenen Kapiteln zur Entwicklung von validen und verteilungsfreien Resampling-Tests für konkrete Ein- und Zweistichprobenprobleme angewendet. Da Permutationstests bei Einstichprobenproblemen nicht anwendbar sind, rücken hierfür zuerst die verschiedenen Bootstrap-Verfahren in den Mittelpunkt. Erst im weiteren Verlauf des Kapitels werden Zweistichprobentests untersucht, die als studentisierte Permutationstests ausgeführt werden. Diese studentisierten Tests sind bei austauschbaren Zufallsgrößen immer noch exakt, besitzen aber den Vorteil, dass sie unter größeren Nullhypothesen - anders als die nichtstudentisierten Versionen - auch

Einleitung _____ 5

im unbalancierten Fall asymptotisch das Niveau einhalten. So erhält man beispielsweise eine neue, studentisierte Permutationsversion des F-Tests zum Vergleich zweier Varianzen.

Für die verschiedenen Resampling-Tests wird zusätzlich noch das asymptotische Verhalten der Gütefunktion für feste, sowie - in Spezialfällen - auch für lokale Alternativen untersucht.

Das sechste Kapitel zeigt u.a., dass sich die in Kapitel 2 angesprochenen Verfahren in den gewichteten Ansatz einordnen lassen und untersucht darüber hinaus viele weitere Beispiele von Resampling-Gewichten, die den oben angesprochenen Voraussetzungen genügen.

Der abschließende Appendix fasst einige wichtige Aussagen über die schwache Konvergenz von bedingten Verteilungen sowie über die Theorie der stark stationären und ergodischen Zufallsvariablen zusammen, die für diese Arbeit benötigt werden.

Abschließend möchte ich meinem Lehrer Herrn Prof. Dr. Arnold Janssen sowohl für die vielen anregenden Gespräche als auch für die großartige Betreuung und Hilfsbereitschaft bei der Erstellung dieser Arbeit danken. Ein herzlicher Dank gebührt auch Herrn Apl. Prof. Dr. Helmut Finner und Herrn Prof. Dr. Erich Häusler für die Übernahme und Erstellung der weiteren Gutachten.

Bedanken möchte ich mich auch bei den Professoren und Mitarbeitern des Lehrstuhls für Mathematische Statistik und Wahrscheinlichkeitstheorie am Mathematischen Institut der HHU Düsseldorf. Durch ihre stete Diskussions- und Hilfsbereitschaft füllen sie die Stochastik mit Leben.

Darüber hinaus danke ich der Gründerstiftung zur Förderung von Forschung und wissenschaftlichem Nachwuchs an der Heinrich-Heine-Universität Düsseldorf (Düsseldorfer Entrepreneurs Foundation) für deren Unterstützung in Form eines eineinhalbjährigen Stipendiums. Hierdurch konnte ich die vorliegende Arbeit sehr voranbringen.

Mein tiefer Dank gilt meiner Frau Daniela Pauly sowie meinen Eltern Gabriele und Werner Pauly für ihre Liebe und Unterstützung. Ohne sie wäre ich nicht der Mensch, der ich heute bin.

Kapitel 2

Resampling-Verfahren - Die Grundlagen

2.1 Motivation

Beim Übergang von einem parametrischen zu einem nichtparametrischen Testproblem besteht das Hauptproblem in der Festlegung geeigneter kritischer Werte. Angenommen man hat für das Testproblem eine geeignete Teststatistik T_n gefunden, so kann man in der Regel die zugehörige Verteilung nicht oder nur in Spezialfällen bestimmen. Eine Möglichkeit zur Bestimmung des kritischen Wertes besteht dann darin, die asymptotische Verteilung von T_n unter der Nullhypothese zu bestimmen und deren Quantile als (approximative) kritische Werte zu verwenden. Hierbei ist man jedoch mit mehreren Problemen konfrontiert, die im Folgenden kurz aufgelistet werden:

- Es kann vorkommen, dass man die Quantile der Grenzverteilung erst gar nicht berechnen kann, da diese zu kompliziert ist.
- In vielen Fällen hängt die Grenzverteilung von unbekannten Parametern ab, die man nur unzureichend bzw. fehlerbehaftet schätzen kann.
- Ist die Konvergenz gegen die Grenzverteilung recht langsam, so lässt sich der Fehler, der bei der Approximation entsteht insbesondere für kleine Stichprobenumfänge schlecht kontrollieren.

Speziell der letzte Punkt kann dazu führen, dass der so konstruierte Test das gewünschte Fehlerniveau 1.Art zu stark überschreitet.

Eine mögliche Lösung dieser Probleme besteht nun - wie schon in der Einleitung bemerkt - in der Verwendung von Resampling-Verfahren. Die für diese Arbeit wichtigsten, das Permutations- und das Bootstrap-Verfahren, werden in den folgenden Abschnitten vorgestellt.

2.2 Das Permutationsverfahren

Möchte man zu gegebenen Zufallsvariablen $X_1,\ldots,X_n:(\Omega,\mathcal{A},P)\to\mathbb{R}$ die Nullhypothese

$$H_0: X_1, \ldots, X_n$$
 sind i.i.d.

mittels einer adäquaten Teststatistik $T_n = T(X_1, \ldots, X_n) : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ zu einem gegebenen Fehlerniveau $\alpha \in (0,1)$ überprüfen, so bietet das Permutationsverfahren eine geeignete, verteilungsfreie Möglichkeit zur Bestimmung kritischer Werte an.

Hierbei hält man zunächst für $\omega \in \Omega$ die Beobachtungen $X_1(\omega), \ldots, X_n(\omega)$ fest und berechnet dann den datenabhängigen kritischen Wert $c_n(\alpha) = c_n(\alpha, \omega)$ für eine von den Daten unabhängige, auf der symmetrischen Gruppe \mathcal{S}_n gleichverteilte Permutation $\tau : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}) \to \mathcal{S}_n$, als das $(1 - \alpha)$ -Quantil der Permutationsverteilung von T_n , d.h. als Lösung von

$$\alpha = \tilde{P}(T_n((X_{\tau(i)}(\omega))_{i \le n}) > c_n(\alpha)) + \gamma_n \tilde{P}(T_n((X_{\tau(i)}(\omega))_{i \le n}) = c_n(\alpha)) \quad (2.1)$$

für $\gamma_n=\gamma_n(\omega)\in[0,1]$ geeignet. Der zugehörige einseitige obere Permutationstest ist dann gegeben durch

$$\varphi_n^* = \begin{cases} 1 & > \\ \gamma_n & \text{, falls } T_n = c_n(\alpha). \\ 0 & < \end{cases}$$
 (2.2)

Da die gemeinsame Verteilung der Zufallsvariablen (X_1, \ldots, X_n) unter H_0 invariant unter Koordinatenpermutationen ist, besitzt der Test den Vorteil, dass er das Niveau unter der Nullhypothese exakt einhält. Dies gilt sogar unter der erweiterten Nullhypothese

$$\tilde{H}_0: X_1, \ldots, X_n$$
 sind austauschbar,

vgl. z.B. Kapitel 8 bei Janssen [30].

Bemerkung 2.1

1. Ist die Stichprobenanzahl zu groß, so kostet die Bestimmung der genauen Permutationsverteilung und somit der kritischen Werte mittels (2.1) zu viel Rechenzeit. Aus diesem Grunde wird diese in vielen praktischen Situationen mittels Monte-Carlo-Methoden approximiert, vgl. hierzu auch den Bootstrap-Algorithmus im nächsten Abschnitt.

2. Eine alternative Variante zur Einführung von Permutationstests kann im Rahmen von allgemeinen bedingten Tests erfolgen, vgl. hierzu Kapitel 4.1. □

2.3 Das Bootstrap-Verfahren

Eine weitere Möglichkeit zur Bestimmung der kritischen Werte in der obigen Situation liefert das 1979 von EFRON vorgestellte Bootstrap-Verfahren. Hierzu definiert man sich zu der gegebenen Stichprobe $X_1(\omega), \ldots, X_n(\omega)$ neue, unabhängig und identisch verteilte Zufallsvariablen X_1^*, \ldots, X_n^* mit gleicher bedingter Verteilung

$$P^{X_1^*|(X_1,...,X_n)=(X_1(\omega),...,X_n(\omega))} = \frac{1}{n} \sum_{i=1}^n \varepsilon_{X_i(\omega)}$$

und berechnet den zum Bootstrap-Test gehörigen datenabhängigen kritischen Wert $c_n^*(\alpha) = c_n^*(\alpha, \omega)$ als das $(1 - \alpha)$ -Quantil der bedingten Verteilung von $T_n(X_1^*, \ldots, X_n^*)$ gegeben den Daten $(X_1, \ldots, X_n) = (X_1(\omega), \ldots, X_n(\omega))$.

Nur in wenigen Fällen lässt sich diese bedingte Bootstrap-Verteilung geschlossen angeben und damit $c_n^*(\alpha)$ bestimmen. Ist dies nicht der Fall, so werden approximative kritische Werte mit Hilfe von Monte-Carlo-Verfahren berechnet. Das gesamte Verfahren wird im folgenden Algorithmus nochmal zusammengefasst:

- 1. Beobachte Realisierungen der Zufallsvariablen X_1, \ldots, X_n .
- 2. Erzeuge eine Bootstrap-Stichprobe $X_1^{*(1)}, \ldots, X_n^{*(1)}$ durch Ziehen mit Zurücklegen aus X_1, \ldots, X_n .
- 3. Berechne die Bootstrap-Statistik $T_n^{*(1)} = T_n(X_1^{*(1)}, \dots, X_n^{*(1)})$.
- 4. Wiederhole die Schritte 2.-3. B-mal, $B \in \mathbb{N}$, und erhalte so $T_n^{*\,(b)}$ für $1 \leq b \leq B$.
- 5. Approximiere $c_n^*(\alpha)$ durch das $(1-\alpha)$ -Quantil von $\frac{1}{B} \sum_{b=1}^B \mathbb{1}_{(-\infty,x]}(T_n^{*(b)})$.

Das im vorherigen Abschnitt beschriebene Permutationsverfahren erhält man aus obigem Algorithmus, indem man ohne Zurücklegen zieht. Aus diesem Grund wird hierfür in der Literatur auch häufig der Begriff Bootstrap ohne Zurücklegen verwendet.

Um die Anwendung des Bootstraps zu rechtfertigen, benötigt man natürlich noch Aussagen darüber, wie genau die bedingte Bootstrap-Verteilung die wahre Verteilung von T_n approximiert. Diese sollten wenigstens im Limes übereinstimmen.

Eine der ersten Begründungen - und somit auch Anwendungen - findet sich im Bootstrap der klassischen Mittelwertsstatistik aus dem Zentralen Grenzwertsatz. Mit Hilfe des Satzes von Berry-Esseen lässt sich nämlich u.a. für den Fall von i.i.d. Zufallsvariablen X_1, \ldots, X_n mit Erwartungswert $\mu \in \mathbb{R}$ und endlicher Varianz $\sigma^2 > 0$ leicht zeigen, dass für $n \to \infty$

$$\sup_{x \in \mathbb{R}} \left| P(\sqrt{n}(\overline{X}_n - \mu) \le x) - P(\sqrt{n}(\overline{X}_n^* - \overline{X}_n) \le x | X_1, \dots, X_n) \right| \xrightarrow{f.s.} 0 \quad (2.3)$$

gilt, vgl. beispielsweise Pauly [50, Satz 3.2]. Da die Konvergenzgeschwindigkeit hierbei in vielen Fällen schneller ist als beim Zentralen Grenzwertsatz, ist die Verwendung des Bootstrap-Verfahrens angebracht.

Eine genaue Untersuchung der Konvergenzgeschwindigkeiten in Abhängigkeit von verschiedenen Momentenbedingungen findet man für den klassischen Bootstrap insbesondere in Theorem 1 bei SINGH [59] und für eine Form des gewichteten Bootstraps, auf den weiter unten noch genauer eingegangen wird, in den Kapiteln 4 und 5 bei HÄUSLER ET. AL. [21]

2.3.1 Gegenbeispiele und Erweiterungen

Dieses so einfach zu implementierende Verfahren ist mittlerweile ein Standardwerkzeug für viele Problemstellungen in der nichtparametrischen Statistik. Aus diesem Grunde hat es auch zu Recht den Einzug in viele statistische Anwendungsgebiete und Programmpakete gefunden und die dortigen Verfahren erweitert und verbessert. Es ist jedoch kein Allheilmittel. Neben vielen Anwendungsmöglichkeiten findet man nämlich auch Situationen, wo das oben beschriebene, klassische Verfahren keine adäquate Lösung für die gegebene Fragestellung liefert. Um diese Fälle trotzdem behandeln zu können, kann man versuchen den Bootstrap-Ansatz von Efron geeignet zu modifizieren. Hierfür finden sich in

der Literatur zahlreiche Beispiele, von denen im Folgenden einige exemplarisch vorgestellt werden.

Beispiel 2.2

Einem der ersten Gegenbeispiele begegnet man in der frühen Arbeit von BICKEL UND FREEDMAN (vgl. [5, S.1210]). Dort wird u.a. bewiesen, dass sich die Verteilung der zentrierten Maximumsstatistik

$$M_n := n(1 - \max_{i \le n} X_i)$$

von unabhängigen und auf (0,1) gleichverteilten Zufallsvariablen X_1,\ldots,X_n nicht mittels der nichtparametrischen Bootstrap-Version

$$M_n^* := n(\max_{i \le n} X_i - \max_{i \le n} X_i^*)$$

approximieren lässt. Während nämlich M_n in Verteilung gegen eine Standardexponentialverteilung konvergiert, existiert für die bedingte Verteilung $\mathcal{L}(M_n^*|\vec{X})$ mit Wahrscheinlichkeit 1 keine Grenzverteilung.

Eine Lösung dieses Problems kann z.B. durch eine bestimmte Variation des Stichprobenumfangs der Bootstrap-Stichprobe erreicht werden¹. Zieht man dabei insgesamt m(n)-mal, $m(n) \in \mathbb{N}$, mit Zurücklegen aus der Grundgesamtheit, so spricht man vom sogenannten $\mathbf{m(n)}$ -Bootstrap, siehe auch Beispiel 2.5 weiter unten.

Ein weiteres Gegenbeispiel findet man im Zusammenhang mit abhängigien Daten.

Beispiel 2.3

SINGH [59, Remark 2.1.] zeigt u.a., dass für eine Folge X_1, X_2, \ldots von stationären, m-dependent² Zufallsvariablen mit Erwartungswert μ und endlicher Varianz $0 < \sigma^2 < \infty$, die Konvergenz in (2.3) nicht eintritt.

Die bedingte Verteilung der Bootstrap-Version konvergiert zwar für P-fast alle ω schwach gegen eine normalverteilte Zufallsvariable, d.h. für $n \to \infty$ gilt

$$\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|X_1, \dots, X_n) \xrightarrow{w} N(0, \sigma^2)$$
 P-fast sicher,

¹Zur genauen Wahl von m(n) in der obige Situation vgl. SWANEPOEL [61, S.3195f, Theorem 2.1.]. Das daraus resultierende Resampling-Verfahren gehört wegen $m(n)/n \to 0$ zu den sog. Low-Resampling-Verfahren, auf die in den nächsten Kapiteln noch näher eingegangen wird.

²Dabei heißt eine Folge $(X_n)_n$ von Zufallsvariablen m-dependent, für $n \in \mathbb{N}_0$, falls (X_1, \ldots, X_k) unabhängig von (X_{k+m+1}, \ldots) ist.

jedoch besitzt die ursprüngliche Statistik eine andere Grenzverteilung

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{D}} Z \sim N(0, \sigma^2 + 2\sum_{i=1}^{m-1} \text{Cov}(X_1, X_{i+1})),$$

so dass keine Approximation über das klassische Bootstrap-Verfahren möglich ist. Wie man an der Grenzvarianz erkennen kann, liegt der Grund hierfür gerade in der Tatsache, dass die klassische Bootstrap-Methode die Abhängigkeitsstruktur des Datensatzes nicht berücksichtigt.

Dieser Umstand kann z.B. behoben werden, indem man die Daten in verschiedene Blöcke aufteilt und dann hieraus die neue Stichprobe durch Ziehen mit Zurücklegen ermittelt. In diesem Fall spricht man vom sogenannten Block-Bootstrap. Die Beschreibung der genauen Vorgehensweise bei dieser Methode sowie viele weitere, verfeinerte Resampling-Techniken findet man in Kapitel 6.

Neben den oben genannten Gegenbeispielen beschreiben BABU [1] und DATTA [9] in ihren Arbeiten einen weiteren Fall, in dem der klassische Ansatz nicht zu funktionieren scheint. Dabei beschäftigen sich beide mit komplizierteren, nichtlinearen Statistiken, deren asymptotische Grenzverteilung keine Normalverteilung ist. Genauer behandeln beide die Statistik $T_n := H(\overline{Z}_n)$ für eine ausreichend glatte Funktion $H: \mathbb{R}^k \to \mathbb{R}$ und i.i.d. Zufallsvektoren $Z_1, \ldots, Z_n : \Omega \to \mathbb{R}^k$ mit Erwartungswert μ und existierender, nicht-singulärer Kovarianzmatrix Σ . Dabei wird für den Fall $k = 1, \mu = 0$ und $H(x) = x^2$ beispielhaft festgestellt, dass die Statistik $nT_n = n(\overline{Z}_n)^2$ und die Bootstrap-Statistik $n((\overline{Z}_n^*)^2 - (\overline{Z}_n)^2)$ i.a. nicht dieselbe asymptotische Grenzverteilung besitzen.

Dieses Problem kann man jedoch ganz leicht umgehen, indem man die Daten **zuerst zentriert** und dann die Bootstrap-Stichprobe aus den zentrierten Daten erzeugt³, d.h. man vergleicht $n(\overline{Z}_n - \mu)^2$ mit $n(\overline{Z}_n^* - \overline{Z}_n)^2$. Diese Statistiken besitzen beide nach dem Satz über stetige Abbildungen und (2.3) die χ_1^2 -Verteilung als asymptotische Grenzverteilung.

Im oberen Fall macht es also auch allgemein Sinn

$$H(\sqrt{n}(\overline{Z}_n - \mu))$$

mit Hilfe von

$$H(\sqrt{n}(\overline{Z}_n^* - \overline{Z}_n)) = H(\frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \overline{Z}_n)^*)$$

³Dabei spielt diese Zentrierungsreihenfolge bei linearen Statistiken keine Rolle.

zu approximieren. Dies folgt auch später in Spezialfällen aus Satz 3.20.

2.4 Allgemeines Resampling durch Gewichtsfunktionen

In diesem Abschnitt wird ein Verfahren vorgestellt, mit dem man die oben vorgestellten Methoden simultan untersuchen kann. Dieses sogenannte gewichtete Resampling-Verfahren wurde zunächst von Mason und Newton, vgl. [48], vorgeschlagen. Dabei zeigen sie u.a., dass sich die gleichmässige Konsistenz des Bootstrap-Verfahrens für die Mittelwertstatistik (2.3) unter denselben Voraussetzungen auch auf weitere Bootstrap-Verfahren⁴ übertragen lässt. Dieser Ansatz wurde von weiteren Autoren aufgegriffen und erweitert, vgl. u.a. HÄUSLER ET. AL. [21], JANSSEN ET.AL. [33],[35],[12] sowie Praestgaard und Wellner [53]. Dabei wurden zum einen weitere Resampling-Techniken wie die Permutationsmethode, der Wild-Bootstrap (vgl. Kapitel 6 sowie Mammen [46, Theorem 3 und 4) und der iterative Double-Bootstrap (siehe Beispiel 2.5 unten sowie [53]) mit aufgenommen und zum anderen die Bedingung der Unabhängigkeit abgeschwächt, vgl. insbesondere [33],[35] und [12]. Die hierzu gehörigen mathematischen Voraussetzungen werden im dritten Kapitel genauer vorgestellt. Die Idee für den gewichteten Ansatz lässt sich allerdings schon an dieser Stelle sehr schön motivieren. Sei dazu X_1, \ldots, X_n eine Folge von Zufallsvariablen mit unbekannter Verteilung und davon unabhängig $U_1, \ldots, U_{m(n)}$ unabhängige, auf der Menge $\{1,\ldots,n\}$ gleichverteilte Zufallsvariable. Hiermit lassen sich die m(n)-Bootstrap-Variablen wie folgt verteilungsgleich ausdrücken

$$X_{i}^{*} = X_{U_{i}}$$
.

Definiert man nun zusätzlich Zufallsvariablen

$$M_{n,j} := \frac{1}{m(n)} \sum_{i=1}^{m(n)} \mathbb{1}_j(U_i), \quad 1 \le j \le n, \tag{2.4}$$

so kann man die Bootstrap-Version der Mittelwertsstatistik durch diese Gewichte und die Zufallsvariablen X_i wie folgt ausdrücken

$$m(n)^{1/2} \left(\overline{X}_{m(n)}^* - \overline{X}_n \right) = \sum_{i=1}^n m(n)^{1/2} \left(M_{n,i} - \frac{1}{n} \right) \left(X_i - \overline{X}_n \right). \tag{2.5}$$

⁴Hierbei handelte es sich um den gewichteten i.i.d. Bootstrap und um Rubins Bayesian Bootstrap, vgl. Kapitel 6

Dabei gilt $\overline{X}_{m(n)}^*:=\frac{1}{m(n)}\sum_{i=1}^{m(n)}X_i^*$. Diese Motivation gibt Anlass allgemeinere Statistiken der Form

$$T_n^* = \sum_{i=1}^{k(n)} W_{n,i} (X_i - \overline{X}_n)$$
 (2.6)

für verschiedene, das Resampling-Verfahren beschreibende Gewichte $W_{n,i}$ zu behandeln. Dies wird nun in den folgenden zentralen Beispielen aufgegriffen. Hierbei seien - wie auch in der gesamten Arbeit - k(n) und m(n) Teilfolgen der natürlichen Zahlen, die für $n \to \infty$ auch gegen ∞ konvergieren.

Beispiel 2.4

Wie man aus der obigen Motivation und Formel (2.5) ablesen kann, sind die Gewichte für den **m(n)-Bootstrap** definiert als

$$W_{n,i} = \sqrt{m(n)} \left(\frac{1}{m(n)} M_{n,i} - \frac{1}{k(n)} \right), \quad 1 \le i \le k(n).$$
 (2.7)

Dabei ist $M := (M_{n,1}, \ldots, M_{n,k(n)})$ eine multinomial verteilte Zufallsvariable zum Stichprobenumfang $m(n) = \sum_{i=1}^{k(n)} M_{n,i}$ mit jeweils gleicher Auswahlwahrscheinlichkeit 1/k(n).

Beispiel 2.5

Eine weitere Modifikation des m(n)-Bootstraps kann darin bestehen, die gezogene m(n)-Bootstrap-Stichprobe $\{X_1^*,\ldots,X_{m(n)}^*\}$ als Grundgesamtheit zu betrachten und hieraus eine zweite Bootstrap-Stichprobe $\{X_1^{**},\ldots,X_{m(n)}^{**}\}$ zu ziehen. Man spricht dann vom sogenannten m(n)-double Bootstrap. Die zugehörigen Gewichte sind in diesem Fall

$$W_{n,i} := \left(\frac{m(n)}{2}\right)^{1/2} \left(\frac{1}{m(n)} M'_{n,i} - \frac{1}{k(n)}\right), \quad 1 \le i \le k(n). \tag{2.8}$$

Dabei ist $(M'_{n,1},\ldots,M'_{n,k(n)})$ bedingt unter M (aus Beispiel 2.4) multinomialverteilt zum Stichprobenumfang m(n) mit Auswahlwahrscheinlichkeiten $(\frac{M_{n,i}}{m(n)})_{i\leq k(n)}$

Beispiel 2.6

Seien

$$c_{n,i} = \left(\frac{n_1 \cdot n_2}{k(n)}\right)^{1/2} \cdot \begin{cases} \frac{1}{n_1}, & \text{für } 1 \le i \le n_1, \\ -\frac{1}{n_2}, & \text{für } n_1 < i \le k(n) \end{cases}$$
 (2.9)

die Regressionskoeffizienten des Zweistichprobenproblems und τ wie zu Beginn des Kapitels eine auf der symmetrischen Gruppe $\mathcal{S}_{k(n)}$ gleichverteilte Permutation, die unabhängig von dem vorliegenden Datensatz $X_1, \ldots, X_{k(n)}$ ist. Dann

sind die Zweistichprobenpermutationsgewichte definiert durch

$$W_{n,i} := c_{n,\tau(i)}, i \le k(n) = n = n_1 + n_2. \tag{2.10}$$

. 14

Viele weitere Beispiele mit Bezug zu den noch näher zu erläuternden mathematischen Voraussetzungen unter denen die allgemeine Resampling-Statistik T_n^* einen bedingten Zentralen Grenzwertsatz der Form (2.3) erfüllt, finden sich in Kapitel 3 und 6.

2.5 Resampling-Tests

Resampling-Tests _

Nachdem im vorangegangenen Abschnitt das allgemeine Resampling-Verfahren vorgestellt wurde, werden an dieser Stelle die zugehörigen Resampling-Tests genauer untersucht. Seien dazu H_0 eine Nullhypothese von Interesse und T_n : $\Omega^{k(n)} \to \mathbb{R}$ eine geeignete Teststatistik. Das nachfolgende Lemma dient als Haupthilfsmittel zum Vergleich von Resampling-Tests mit unbedingten T_n -Tests und entspricht Lemma 1 von Janssen und Pauls [33]. Hierzu werden die folgenden Annahmen getroffen:

(A) Sei $P \in H_0$. Die Teststatistik T_n konvergiere unter P für $n \to \infty$ in Verteilung gegen eine weitere Zufallsvariable T

$$\mathcal{L}(T_n|P) \xrightarrow{w} \mathcal{L}(T|P),$$
 (2.11)

deren Verteilungsfunktion F_T stetig und streng monoton wachsend auf ihrem Träger ist.

(B) Für $\alpha \in (0,1)$ sei $\varphi_{n,\alpha}$ der durch den unbedingten kritischen Wert $c_n(\alpha)$ gegebene obere T_n -Test

$$\varphi_{n,\alpha} = 1 \text{ für } T_n > c_n(\alpha) \text{ und } \varphi_{n,\alpha} = 0 \text{ für } T_n < c_n(\alpha),$$
so dass $E_P(\varphi_{n,\alpha}) \to \alpha \text{ für } n \to \infty \text{ gelte.}$

(C) Sei T_n^* eine Teststatistik (in unserem Fall eine Resampling-Statistik) mit bedingter Verteilungsfunktion F_n^* von $\mathcal{L}(T_n^*|X_{n,1},\ldots,X_{n,k(n)})$. Bezeichne mit $c_n^*(\alpha) = c_n^*(\alpha|X_{n,1},\ldots,X_{n,k(n)}) := F_n^{*-1}(1-\alpha)$ das $(1-\alpha)$ -Quantil der bedingten Verteilung gegeben den Daten $X_{n,1},\ldots,X_{n,k(n)}$. Ein bedingter oberer (Resampling-) T_n -Test ist dann durch

$$\varphi_{n,\alpha}^* = 1 \text{ für } T_n > c_n^*(\alpha) \text{ und } \varphi_{n,\alpha}^* = 0 \text{ für } T_n < c_n^*(\alpha)$$
 (2.13)

gegeben.

Lemma und Definition 2.7

Unter den obigen Bedingungen (A)-(C) sind die beiden folgenden Aussagen äquivalent 5 :

(a)
$$d_1(\mathcal{L}(T_n^*|X_{n,1},\dots,X_{n,k(n)}),\mathcal{L}(T)) \xrightarrow{P} 0, \qquad (2.14)$$

(b) Für alle $\alpha \in (0,1)$ gilt

$$E_P(|\varphi_{n,\alpha} - \varphi_{n,\alpha}^*|) \to 0.$$
 (2.15)

Dabei heißen Tests $\varphi_{n,\alpha}^*$ und $\varphi_{n,\alpha}$, die der Bedingung (2.15) genügen **asymptotisch äquivalent** (unter P). Bezeichnet man mit $\mathcal{P}_0 \subset H_0$ die Klasse von Verteilungen, für die (B) gilt, so heißt $\varphi_{n,\alpha}^*$ asymptotisch effektiv im Bezug auf $\varphi_{n,\alpha}$, falls (2.15) für alle $P \in \mathcal{P}_0$ gilt.

Bemerkung 2.8

Unter der Voraussetzung (A) ist (B) erfüllt, falls die kritischen Werte $c_n(\alpha)$ gegen das $(1-\alpha)$ -Quantil $F_T^{-1}(1-\alpha)$ der Grenzverteilung konvergieren. Bezeichnet man mit $F_{T_n}^{-1}$ die inverse Verteilungsfunktion von T_n , so liegt dies insbesondere für die Wahl $c_n(\alpha) = F_{T_n}^{-1}(1-\alpha)$ vor, vgl. Lemma 11.3 bei Janssen [30].

Der Nachweis der asymptotischen Effektivität des zu $\varphi_{n,\alpha}$ gehörigen Resampling-Tests $\varphi_{n,\alpha}^*$ erfordert daher die Kenntnis von bedingten Grenzwertsätzen der Form (2.14). Diese werden in den kommenden Abschnitten behandelt.

Um den Zusammenhang mit der anfänglichen Problemstellung zu sehen, werden zum Abschlus dieses Kapitels noch einmal die Vorteile von Resampling-Tests zusammengefasst:

- Oft hängt die Verteilung der Grenzvariablen T aus (A) von unbekannten Parametern ab. Dieses Problem tritt bei Resampling-Verfahren nicht auf, da diese verteilungsfrei ausgeführt werden.
- ullet Selbst wenn man die Verteilung von T kennt, kann es schwierig sein, die zugehörigen Quantile genau zu bestimmen. In diesem Fall überzeugen die

⁵Dabei bezeichnet d_1 die Levy-Metrik, vgl. z.B. WITTING UND MÜLLER-FUNK [66, Anmerkung 5.42].

Resampling-Verfahren wieder dadurch, dass man ihre bedingte Verteilung leicht durch Monte-Carlo-Methoden approximieren kann.

- Auch wenn keiner der beiden obigen Fälle vorliegt, kann es immer noch sinnvoll sein, Resampling-Tests zu verwenden, da diese im Finiten ihren unbedingten Konkurrenztests i.d.R. überlegen sind, vgl. hierzu beispielsweise die Arbeiten von HOROWITZ [28] und JANSSEN [35].
- Permutationstests halten i.a. sogar unter der Bedingung der Austauschbarkeit das Niveau exakt ein und besitzen somit insbesondere für Zweistichprobenprobleme Vorteile. Bei Einstichprobenproblemen sind diese jedoch nicht anwendbar. Dies liegt daran, dass die gemeinsame Verteilung der beobachteten Zufallsvariablen auch unter bestimmten Alternativen austauschbar sein kann, so dass der Einstichprobenpermutationstest diese nicht von der Hypothese trennen kann.

Teil I

Bedingte Zentrale
Grenzwertsätze für lineare und
nicht-lineare
Resampling-Statistiken und ihre
Anwendungen

Kapitel 3

Bedingte Zentrale Grenzwertsätze für lineare ResamplingStatistiken

In diesem Kapitel werden einige bedingte Grenzwertsätze für gewichtete Resampling-Statistiken bewiesen. Dabei ist durch jede geeignete Wahl von Gewichten ein spezielles Resampling-Verfahren gegeben. Wie in Kapitel 2 bereits erwähnt, besteht der Vorteil dieser allgemeinen Betrachtungsweise darin, dass man mehrere Resampling-Verfahren simultan behandeln kann.

Als Grundlage für die Beweise dienen bedingte Zentrale Grenzwertsätze für reelle Zufallsvariablen von Janssen [35, Theorem 2.1] und Del Barrio et.al. [12, Theorem 2.2].

Aufbauend auf den Gleichungen (2.4) - (2.6) wird zunächst der Begriff der Resampling-Statistik definiert.

Definition 3.1 (Resampling-Statistik)

Sei $(X_{n,i})_{i \leq k(n)}$ ein Dreiecksschema von Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , welches Werte auf dem separablen \mathbb{R} -Vektorraum $(\mathbb{E}, \mathcal{E})$ mit Borelscher σ -Algebra \mathcal{E} annimmt. Hiervon unabhängig sei $(W_{n,i})_{i \leq k(n)}$ ein Dreiecksschema von reellen Gewichtsfunktionen auf einem Wahrscheinlichkeits $raum\ (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}),\ dann\ wird\ durch$

$$T_n^* = T_n^*((X_{n,i})_{i \le k(n)}, (W_{n,i})_{i \le k(n)}) := \sqrt{k(n)} \sum_{i=1}^{k(n)} W_{n,i}(X_{n,i} - \overline{X}_n)$$
(3.1)

eine allgemeine Resampling-Statistik definiert.

Dabei bedeutet Unabhängigkeit der beiden Dreiecksschemata, dass diese auf dem gemeinsamen Wahrscheinlichkeitsraum $(\Omega \times \tilde{\Omega}, \mathcal{A} \otimes \tilde{\mathcal{A}}, P \otimes \tilde{P})$ unabhängig sind, wenn man beide kanonisch mittels Projektionen auf diesem gemeinsamen Raum definiert¹.

Um nun Aussagen über das Konvergenzverhalten von T_n^* treffen zu können, muss erst noch genauer spezifiziert werden, welche Art von Gewichtsfunktionen man betrachten möchte. Als Auswahlkriterium werden - analog zu den Arbeiten von Janssen et al. [12] [35] [33] - die folgenden Generalvoraussetzungen zu Grunde gelegt.

Definition 3.2

Sei $\overline{W}_n := \frac{1}{k(n)} \sum_{i=1}^{k(n)} W_{n,i}$. Die Gewichte $W_{n,i}$ erfüllen die **Generalvoraus**setzungen, falls die folgenden Aussagen erfüllt sind:

$$(W_{n,1},\ldots,W_{n,k(n)})$$
 sind austauschbar, (3.2)

$$\max_{1 \le i \le k(n)} |W_{n,i} - \overline{W}_n| \to 0 \ \tilde{P}\text{-stochastisch}, \tag{3.3}$$

$$\sum_{i=1}^{k(n)} (W_{n,i} - \overline{W}_n)^2 \to 1 \ \tilde{P}\text{-stochastisch}$$
 (3.4)

und es liegt Verteilungskonvergenz

$$\sqrt{k(n)}(W_{n,1} - \overline{W}_n) \stackrel{\mathcal{D}}{\to} W_1 \tag{3.5}$$

gegen eine Zufallsvariable W_1 vor.

Unter bestimmten Voraussetzungen an m(n) und k(n) erfüllen die Gewichte aus den Beispielen 2.4 - 2.6 die obigen Bedingungen. Für die Bootstrap-Gewichte

 $^{^{1}}$ d.h. es gilt $X_{n,i}: \Omega \times \tilde{\Omega} \ni (\omega, \tilde{\omega}) \mapsto X_{n,i}(\omega) \in E$ sowie $W_{n,i}: \Omega \times \tilde{\Omega} \ni (\omega, \tilde{\omega}) \mapsto W_{n,i}(\tilde{\omega}) \in \mathbb{R}$

(2.7) liegt dabei der "normale" Fall $\operatorname{Var}(W_1) = 1$ vor, falls $m(n)/k(n) \to c \in (0, \infty]$ gilt und das sog. "Low-Resampling" mit $\operatorname{Var}(W_1) = 0$ ergibt sich für den anderen Fall $m(n)/k(n) \to 0$. Analog erhält man für die Zweistichprobenpermutationsgewichte aus Beispiel 2.6 u.a. unter der Bedingung $n_1/n \to p \in (0, 1)$ den normalen und für $n_1/n \to 0, n_1 \to \infty$, den Low-Resampling Fall. Die genauen Voraussetzungen für diese und weitere Resampling-Verfahren werden in Kapitel 6 untersucht.

Für den Fall $E = \mathbb{R}$ existieren für die reelle Resampling-Statistik T_n^* in den beiden Extremfällen $Var(W_1) = 1$ (vgl. [35]) und $Var(W_1) = 0$ (vgl. [12]), d.h. $W_1 = 0$ fast sicher, bereits bedingte Zentrale Grenzwertsätze. Da der Fokus dieser Arbeit hauptsächlich auf dem ersten Fall liegt, handelt es sich bei den auftretenden Gewichten - sofern nicht anders erwähnt - vorwiegend um solche mit Varianz der Grenzvariablen W_1 gleich 1.

3.1 Der eindimensionale Fall

Wir betrachten nun zuerst den Fall, dass ein Dreiecksschema $X_{n,i}$, $i \leq k(n)$, von reellen Zufallssvariablen vorliegt. Da Resampling-Verfahren bedingt nach den Daten ausgeführt werden, hängt die Verteilungskonvergenz von ω ab. Um Missverständnissen bei der Erklärung von P-stochastischer Konvergenz in Verteilung vorzubeugen, wird die schwache Konvergenz auf $\mathcal{M}_1(\mathbb{R}, \mathcal{B})$ mittels einer Metrik d_1 definiert, vgl. hierzu auch Seite 125 im Appendix.

Dabei sei nochmals bemerkt, dass alle auftretenden Zufallsvariablen kanonisch mittels Projektionen auf dem gemeinsamen Wahrscheinlichkeitsraum $(\Omega \times \tilde{\Omega}, \mathcal{A} \otimes \tilde{\mathcal{A}}, P \otimes \tilde{P})$ definiert werden können. Als unmittelbare Folgerungen aus der Arbeit von Janssen, vgl. [35, S.10, Theorem 2.1.], erhält man dann den folgenden Satz.

Satz 3.3

Sei $\overline{X}_n \stackrel{P}{\longrightarrow} 0$. Erfüllen die Gewichte die Generalvoraussetzungen (3.2)-(3.5) mit $\operatorname{Var}(W_1) = 1$ und gelten

$$\max_{i \le k(n)} |X_{n,i} - \overline{X}_n| \to 0 \ P\text{-stochastisch}, \tag{3.6}$$

sowie

$$\sum_{i=1}^{k(n)} (X_{n,i} - \overline{X}_n)^2 \stackrel{\mathcal{D}}{\to} V^2, \tag{3.7}$$

 $mit\ einer\ nichtnegativen\ reellen\ Zufallsvariable\ V^2,\ dann\ folgt$

$$d_1(\mathcal{L}(T_n^*|(X_{n,i})_{i\leq k(n)}), \mathcal{L}(Z)) \xrightarrow{P} 0$$
(3.8)

für eine Zufallsvariable Z mit² $Z(\omega,\cdot) \sim N(0,V^2(\omega))$.

Beweis. Zum Beweis der Gültigkeit von (3.8) reicht es aus, die Voraussetzungen von Theorem 2.1. von Janssen [35] nachzuweisen. Aufgrund von $\overline{X}_n \stackrel{P}{\longrightarrow} 0$ hat man dann nur noch die Konvergenz

$$((X_{i:k(n)})_{i\in\mathbb{N}}, (X_{k(n)+1-j:k(n)})_{j\in\mathbb{N}}) \stackrel{P}{\longrightarrow} 0$$

auf $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ zu zeigen, da hieraus durch Anwendung von Voraussetzung (3.7) zusammen mit dem Lemma von Slutzky die Behauptung folgt. Dabei seien hier - wie auch im weiteren Verlauf der Arbeit - die Räume $\mathbb{R}^{\mathbb{N}}$ und $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ jeweils mit der Produktopologie versehen.

Man setzt nun analog zu Janssen $X_{i:k(n)}=0$ für $i\notin\{1,\ldots,k(n)\}$. Für die Metrik $\rho_{\mathbb{N}}(x,y):=\sum_{i=1}^{\infty}\frac{1}{2^i}\frac{|x_i-y_i|}{1+|x_i-y_i|}$ auf dem Folgenraum $\mathbb{R}^{\mathbb{N}}$ erhält man so für alle $\epsilon>0$

$$P\left(\rho((X_{i:k(n)})_{i\in\mathbb{N}},0)\geq\epsilon\right) \leq P\left(\sum_{i=1}^{k(n)}\frac{\left|X_{i:k(n)}\right|}{2^{i}}\geq\epsilon\right)$$
(3.9)

$$\leq P\left(\max_{i\leq k(n)}|X_{n,i}|\geq\epsilon\right)\to 0.$$
 (3.10)

Analog ergibt sich die Konvergenz gegen 0 der absteigenden Folge von Orderstatistiken und damit auch die gemeinsame Konvergenz gegen 0.

Bemerkung 3.4

- 1. Die Bedingungen $\overline{X}_n \stackrel{P}{\longrightarrow} 0$ und (3.6) sind nach der Dreiecksungleichung zusammen äquivalent zur Konvergenz $\max_{i \leq k(n)} |X_{n,i}| \stackrel{P}{\longrightarrow} 0$.
- 2. Ist die Grenzvariable in (3.7) konstant und positiv, d.h. $V^2 = \sigma^2 > 0$, so konvergiert die bedingte Verteilungsfunktion $F_{T_n^*}(\omega, x) := P(T_n^* \le x | (X_{n,i})_{i \le k(n)} = 0)$

 $^{^2}$ Für $V^2=0$ fast sicher erhält man die entartete Grenzverteilung $N(0,0)=\varepsilon_0$.

 $(X_{n,i}(\omega))_{i\leq k(n)}$) nach dem Satz von Polya, vgl. WITTING UND MÜLLER-FUNK [66, Satz 5.75], sogar gleichmässig gegen $\Phi(x/\sigma)$. Genauer gilt

$$\sup_{x \in \mathbb{R}} \left| F_{T_n^*}(\cdot, x) - \Phi\left(\frac{x}{\sigma}\right) \right| \xrightarrow{P} 0, \tag{3.11}$$

wobei mit Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet wird.

3. Gilt $P(\sum_{i=1}^{k(n)} (X_{n,i} - \overline{X}_n)^2 > 0) \to 1$, so kann man die Zufallsvariablen $X_{n,i}$ durch die normierten Zufallsvariablen

$$Y_{n,i} := \frac{X_{n,i} - \overline{X}_n}{(\sum_{i=1}^{k(n)} (X_{n,i} - \overline{X}_n)^2)^{1/2}} \mathbb{1}_{\{\sum_{i=1}^{k(n)} (X_{n,i} - \overline{X}_n)^2 > 0\}}$$
(3.12)

ersetzen. Diese erfüllen offensichtlich die Bedingung (3.7) mit V=1.

Im Zusammenhang mit Resampling-Verfahren wird in der Litaratur auch häufig die stärkere Konvergenz in der Mallows-Metrik gezeigt, vgl. z.B. BICKEL UND FREEDMAN [5]. Der folgende Zusatz zeigt, dass die Kovergenz (3.8) unter stärkeren Voraussetzungen auch in der Mallows-Metrik³ d_{M2} vorliegt.

Zusatz 3.5

Gilt zusätzlich zu den Voraussetzungen von Satz 3.3 noch $V = \sigma \in \mathbb{R}_{\geq 0}$ sowie

$$\mathrm{E}_{\tilde{P}}\left(\sum_{i=1}^{k(n)} (W_{n,i} - \overline{W}_n)^2\right) \to 1,\tag{3.13}$$

so folgt

$$d_{M2}(\mathcal{L}(T_n^*|(X_{n,i})_{i < k(n)}), N(0, \sigma^2)) \xrightarrow{P} 0. \tag{3.14}$$

Beweis. Ohne Einschränkung gelte $\overline{X}_n = 0^4$. Nach Satz 3.3 findet man mit Hilfe des Teilfolgenkriteriums für stochastische Konvergenz eine gemeinsame Teilfolge m_n sowie eine Menge M mit P(M) = 1, so dass in (3.7) sowie (3.8) Konvergenz entlang dieser Teilfolge für jedes $\omega \in M$ vorliegt. Zur besseren Übersicht setzt man dabei im Folgenden $m_n = n$.

Sei $\tau: (\Omega_1, \mathcal{A}_1, P_1) \to \mathcal{S}_{k(n)}$ eine gleichverteilte Zufallsvariable, so dass τ , $(W_{n,i})_{i \leq k(n)}$ und $(X_{n,i})_{i \leq k(n)}$ gemeinsam unabhängig sind. Definiert man nun alle Zufallsvariablen auf dem gemeinsamen Wahrscheinlichkeitsraum $(\Omega \times \tilde{\Omega} \times \tilde{\Omega})$

³Vgl. hierzu Definition A.10 im Appendix.

⁴Sonst betrachtet man zentrierte Dreiecksschemata $X_{n,i} - \overline{X}_n$.

 $\Omega_1, \mathcal{A} \otimes \tilde{\mathcal{A}} \otimes \mathcal{A}_1, P \otimes \tilde{P} \otimes P_1)$, so liegt aufgrund der Austauschbarkeit der Gewichte Verteilungsgleichheit

$$(W_{n,i})_{i \le k(n)} \stackrel{\mathcal{D}}{=} (W_{n,\tau(i)})_{i \le k(n)}$$

vor. Aufgrund der Unabhängigkeit von $(W_{n,i})_i$ und $(X_{n,i})_i$ gilt daher auch

$$T_n^* = T_n^*((X_{n,i})_{i \le k(n)}, (W_{n,i})_{i \le k(n)}) \stackrel{\mathcal{D}}{=} T_n^*((X_{n,i})_{i \le k(n)}, (W_{n,\tau(i)})_{i \le k(n)}).$$

Des Weiteren erhält man für den bedingten Erwartungswert von T_n^*

$$E\left(\sum_{i=1}^{k(n)} W_{n,\tau(i)} X_{n,i} | (X_{n,i})_{i \le k(n)}, (W_{n,i})_{i \le k(n)}\right) = \sum_{i=1}^{k(n)} X_{n,i} \overline{W}_n = 0.$$
 (3.15)

Somit ergibt sich mit Hilfe der Varianzformel von Hájek und Šidak, vgl. z.B. HÁJEK ET. AL. [22, Theorem 3],

$$V_{n}^{2} := \mathbb{E}\left(\left(\sqrt{k(n)}\sum_{i=1}^{k(n)}W_{n,\tau(i)}X_{n,i}\right)^{2}|(X_{n,i})_{i\leq k(n)}, (W_{n,i})_{i\leq k(n)}\right)$$

$$= \mathbb{V}\operatorname{ar}\left(\sqrt{k(n)}\sum_{i=1}^{k(n)}W_{n,\tau(i)}X_{n,i}|(X_{n,i})_{i\leq k(n)}, (W_{n,i})_{i\leq k(n)}\right)$$

$$= \frac{k(n)}{k(n)-1}\sum_{i=1}^{k(n)}X_{n,j}^{2}\sum_{i=1}^{k(n)}(W_{n,i}-\overline{W}_{n})^{2}.$$

Wegen der Unabhängigkeit von $(W_{n,i})_i$ und $(X_{n,i})_i$ folgt hieraus für alle $\omega \in M$

$$E\left((T_n^*)^2|(X_{n,i})_{i\leq k(n)}\right)(\omega,\cdot)$$

$$= E\left(\left(\sqrt{k(n)}\sum_{i=1}^{k(n)}W_{n,\tau(i)}X_{n,i}\right)^2|(X_{n,i})_{i\leq k(n)}\right)(\omega,\cdot)$$

$$= E\left(V_n^2|(X_{n,i})_{i\leq k(n)}\right)(\omega,\cdot)$$

$$= \frac{k(n)}{k(n)-1}\sum_{j=1}^{k(n)}X_{n,j}^2(\omega) E\left(\sum_{i=1}^{k(n)}(W_{n,i}-\overline{W}_n)^2\right) \to \sigma^2.$$

Da in (3.8) für alle $\omega \in M$ punktweise Konvergenz vorliegt, folgt hieraus mit Hilfe von Lemma A.11 aus dem Appendix die fast sichere Konvergenz in (3.14) entlang der Teilfolge m_n . Eine erneute Anwendung des Teilfogenkriteriums für stochastische Konvergenz liefert so die Behauptung.

Bemerkung 3.6

Die Bedingung (3.13) ist z.B. für die modifizierten Gewichte

$$\tilde{W}_{n,i} := \frac{W_{n,i}}{(\sum_{i=1}^{k(n)} (W_{n,i} - \overline{W}_n)^2)^{1/2}}, \quad i \le k(n)$$

erfüllt. Hierunter fallen u.a. die Zweistichprobenpermutationsgewichte, vgl. Beispiel 2.6. Weitere Gewichte, die dieser Bedingung genügen sind die Bootstrap-, Double-Bootstrap- und Wild-Bootstrap-Gewichte, vgl. Beispiele 2.4 und 2.5 sowie Kapitel 6. Dies kann man direkt nachrechnenen oder den Beweisen von Janssen [35] und Janssen und Pauls [33] entnehmen.

Die folgenden Korollare liefern einen ersten Eindruck über die allgemeinen Anwendungsmöglichkeiten von Satz 3.3.

Sei dazu zunächst $(X_n)_{n\in\mathbb{N}}$ eine Folge von stark stationären Zufallsvariablen und \mathcal{H} die zugehörige σ -Algebra der invarianten Mengen, vgl. Abschnitt A.2 im Appendix.

Korollar 3.7

Ist $(X_n)_n$ stark stationär mit $E(X_1^2) < \infty$, so erfüllt das Dreiecksschema $X_{n,i} := (X_i - \overline{X}_n)/\sqrt{k(n)}, i \le k(n)$, die Voraussetzungen (3.6) sowie (3.7). Falls die Gewichte die Voraussetzungen (3.2)-(3.5) mit $Var(W_1) = 1$ erfüllen, erhält man somit einen bedingten Zentralen Grenzwertsatz

$$d_1\left(\mathcal{L}\left(\sum_{i=1}^{k(n)} W_{n,i}(X_i - \overline{X}_n)|(X_i)_{i \le k(n)}\right), \mathcal{L}(Z)\right) \xrightarrow{P} 0, \tag{3.16}$$

wobei Z eine Zufallsvariable mit $Z(\omega, \cdot) \sim N(0, \operatorname{Var}(X_1|\mathcal{H})(\omega))$ ist⁵.

Beweis. Es wird Satz 3.3 angewendet. Hierzu sei zuerst bemerkt, dass für beliebige Dreiecksschema $(X_{n,i})_{i \leq k(n)}$ die Bedingungen

$$\max_{i \le k(n)} |X_{n,i}| \xrightarrow{P} 0 \tag{3.17}$$

und

$$\sum_{i=1}^{k(n)} X_{n,i}^2 \mathbb{1}_{\{|X_{n,i}| > \epsilon\}} \xrightarrow{P} 0 \text{ für alle } \epsilon > 0,$$
(3.18)

äquivalent sind. Der Beweis ist elementar. Die Äquivalenz wird z.B. auch von HALL UND HEYDE [24, S.53] verwendet. Um dies nun hier anzuwenden zeigt man

⁵Dabei definiert man $Var(X_1|\mathcal{H}) := E(X_1^2|\mathcal{H}) - E(X_1|\mathcal{H})^2$.

zunächst mit Hilfe der Tschebyscheff-Ungleichung die stochastische Konvergenz von $X_1/\sqrt{k(n)}$ gegen 0

$$P(|X_1| > \epsilon \sqrt{k(n)}) \le \epsilon^{-2} \operatorname{Var}(X_1)/k(n) \to 0.$$

Somit erhält man aus dem Lemma von Pratt

$$E(\frac{1}{k(n)}\sum_{i=1}^{k(n)}X_i^2\mathbb{1}_{\{|X_i|>\epsilon\sqrt{k(n)}\}}) = E(X_1^2\mathbb{1}_{\{|X_1|>\epsilon\sqrt{k(n)}\}}) \to 0,$$

woraus sich insbesondere (3.18) ergibt. Dies impliziert schon (3.6). Die zweite Voraussetzung (3.7) von Satz 3.3 erhält man nun mit Hilfe von Bemerkung A.13 aus dem starken Gesetz der grossen Zahlen für stark stationäre Folgen von Zufallsvariablen, vgl. Satz A.14. Hiernach liegen nämlich die Konvergenzen

$$\frac{1}{k(n)} \sum_{i=1}^{k(n)} X_i \xrightarrow{f.s.} \mathrm{E}(X_1 | \mathcal{H}) \text{ und } \frac{1}{k(n)} \sum_{i=1}^{k(n)} X_i^2 \xrightarrow{f.s.} \mathrm{E}(X_1^2 | \mathcal{H})$$

vor, so dass die Behauptung aufgrund der fast sicheren Konvergenz

$$\sum_{i=1}^{k(n)} X_{n,i}^2 \xrightarrow{f.s.} \operatorname{Var}(X_1 | \mathcal{H})$$

folgt.

Bemerkung 3.8

1. Ist die Folge $(X_n)_n$ zusätzlich zu den obigen Bedingungen noch ergodisch, vgl. Definition A.12, so erhält man mit Hilfe von A.13 und Satz A.14

$$d_1\left(\mathcal{L}\left(\sum_{i=1}^{k(n)} W_{n,i}(X_i - \overline{X}_n)|(X_i)_{i \le k(n)}\right), N(0, \operatorname{Var}(X_1))\right) \xrightarrow{P} 0.$$
 (3.19)

Dies zeigt insbesondere die Konsistenz bei unabhängig und identisch verteilten Zufallsvariablen mit endlicher Varianz.

2. Für den Nachweis der Maximumsbedingung (3.6) kann man die Voraussetzungen des Korollars (starke Stationarität und endliche Varianz) abschwächen, indem man für beliebige Dreiecksschemata $(X_{n,i})_{i \leq k(n)}$ nur

$$k(n) \max_{i \le k(n)} E(X_{n,i}^2 \mathbb{1}_{\{|X_{n,i}| \ge \epsilon\}}) \to 0 \text{ für alle } \epsilon > 0$$
(3.20)

fordert. In diesem Fall kann Bemerkung 3.4 verwendet werden, um Grenzwertsätze für normierte Zufallsvariablen der Form (3.12) zu erhalten.

Dabei ist die Bedingung (3.20) beispielsweise für Dreiecksschemata $X_{n,i} = \frac{X_i}{\sqrt{k(n)}}$ in identisch verteilten Zufallsvariablen X_i mit endlicher Varianz $\operatorname{Var}(X_1) < \infty$ erfüllt.

Das nächste Korollar befasst sich mit dem Resampling von Martingaldifferenzenschemata. Dabei heißt ein integrierbares Dreiecksschema $(X_{n,i})_{i\leq k(n)}$ von reellen Zufallsvariablen **Martingaldifferenzenschema**, vgl. HALL UND HEYDE [24, S.51ff.], falls es für jedes $n \in \mathbb{N}$ an ein Dreiecksschema von Filtrationen $(\mathcal{F}_{n,i})_{i\leq k(n)}$ adaptiert ist und der Gleichung

$$E(X_{n,i}|\mathcal{F}_{n,i-1}) = 0 \text{ für alle } 1 \le i \le k(n) \text{ und } n \in \mathbb{N}$$
 (3.21)

genügt. Dabei wurden $X_{n,0} := 0$ und $\mathcal{F}_{n,0} := \{\emptyset, \Omega\}$ gesetzt.

Im Folgenden wird angenommen, dass die Gewichte die Voraussetzungen (3.2)-(3.5) mit $Var(W_1) = 1$ erfüllen. Unter dieser Voraussetzung werden nachfolgend hinreichende Bedingungen für das konsistente Resamplen der Statistik $\sum_{i=1}^{k(n)} X_{n,i}$ aufgezeigt.

Korollar 3.9

Sei $(X_{n,i}, \mathcal{F}_{n,i})_{i \leq k(n)}$ ein quadratintegrierbares Martingaldifferenzenschema, welches die nachfolgenden Annahmen erfüllt

$$\max_{i \le k(n)} |X_{n,i}| \xrightarrow{P} 0, \tag{3.22}$$

$$\sum_{i=1}^{k(n)} X_{n,i}^2 \xrightarrow{P} \sigma^2 \in \mathbb{R}_{\geq 0}, \tag{3.23}$$

$$E(\max_{i \le k(n)} X_{n,i}^2) \text{ ist beschränkt in } n.$$
(3.24)

Dann gelten

$$d_1\left(\mathcal{L}\left(\sum_{i=1}^{k(n)} X_{n,i}\right), N(0, \sigma^2)\right) \longrightarrow 0$$
(3.25)

und

$$d_1(\mathcal{L}(T_n^*|X_{n,1}\dots X_{n,k(n)}), N(0,\sigma^2)) \xrightarrow{P} 0. \tag{3.26}$$

Beweis. Die Voraussetzungen implizieren nach Theorem 3.2 und den anschließenden Bemerkungen von HALL UND HEYDE [24, S.58f.] gerade die unbedingte

Verteilungskonvergenz (3.25).

Das Grenzverhalten von T_n^* folgt durch Anwendung von Satz 3.3 auf das zentrierte Dreiecksschema $X_{n,i} - \overline{X}_n, i \leq k(n)$. Da die Voraussetzung (3.22) nach der Dreiecksungleichung schon die Bedingung (3.6) impliziert, muss hierzu nur noch die Konvergenz der Summe der Quadrate (3.7) nachgewiesen werden. Dies erhält man aber unmittelbar mit Hilfe des Lemmas von Slutzky aus den Konvergenzen (3.25) und (3.23)⁶

$$\sum_{i=1}^{k(n)} (X_{n,i} - \overline{X}_n)^2 = \sum_{i=1}^{k(n)} X_{n,i}^2 - \overline{X}_n (\sum_{i=1}^{k(n)} X_{n,i})$$
$$= \sum_{i=1}^{k(n)} X_{n,i}^2 + o_P(1) \xrightarrow{P} \sigma^2.$$

Bemerkung 3.10

Ist die Grenzvariable σ^2 in (3.23), anders als gefordert, nicht konstant sondern eine fast sicher endliche Zufallsvariable V^2 , so stimmen die Grenzverteilungen in (3.25) und (3.26) i.a. nicht mehr überein⁷, so dass die Verwendung von Resampling-Verfahren in diesem Falle zu nicht konsistenten Ergebnissen führen würde.

Die bisherigen Grenzwertsätze für T_n^* haben sich auf den Fall konzentriert, dass die Grenzvariable W_1 aus Generalvoraussetzung (3.5) Varianz 1 besitzt. Für den anderen Extremfall $Var(W_1) = 0$ finden sich in der Arbeit von Del Barrio ett. Al. [12, S.9, Theorem 2.1] unter etwas stärkeren Voraussetzungen ähnliche Sätze. Da in Kapitel 6 u.a. solche sogenannten Low-Resampling-Gewichte, wie z.B. Subsampling-Gewichte, behandelt werden und der nächste Teilabschnitt auch Aussagen für diesen Fall benötigt, wird der Vollständigkeit halber noch eine vereinfachte Version der Hauptaussage ihrer Arbeit angegeben.

⁶Dabei bezeichnet man mit dem Symbol $o_P(1)$ eine Zufallsvariable, die für $n \to \infty$ P-stochastisch gegen 0 konvergiert.

 $^{^7}$ Genauer zeigen Hall und Heyde [24, S.58f.], dass die Grenzverteilung in (3.25) unter geeigneten Meßbarkeitsbedingungen an V^2 nicht von ω abhängt. Die bedingte Resampling-Verteilung konvergiert jedoch nach Satz 3.3 gegen eine Grenzverteilung, deren Varianz von $V^2(\omega)$ bestimmt wird.

Satz 3.11

Sei $\tilde{X}_{n,i} := X_{n,i} - \overline{X}_n$. Erfüllen die Gewichte die Generalvoraussetzungen (3.2)-(3.5) mit $\operatorname{Var}(W_1) = 0$ und gelten zusätzlich zu (3.6)

$$\sum_{i=1}^{k(n)} \tilde{X}_{n,i}^2 \xrightarrow{P} \sigma^2 \in \mathbb{R}_{\geq 0}$$
 (3.27)

und

$$\lim_{d \to \infty} \limsup_{n \to \infty} \sum_{i=1}^{k(n)} \tilde{X}_{n,i}^2 \mathbb{1}_{[d,\infty)}(\sqrt{k(n)}|\tilde{X}_{n,i}|) = 0, \tag{3.28}$$

so folgt (3.8) mit $V^2 = \sigma^2$.

Beweis. Die Aussage folgt unmittelbar aus Theorem 2.1. von Del Barrio et. Al. [12, Remark 2.4.], da die obigen Bedingungen deren Voraussetzungen mit $\bar{\sigma}^2 = 1$ und $\xi_0^2 = \sigma^2$ implizieren.

Offensichtlich sind die Bedingungen für die ergodischen und stationären Dreiecksschemata aus dem ersten Teil von Bemerkung 3.8 erfüllt. Genauso lässt sich zeigen, dass diese Voraussetzungen in klassischen Zweistichprobenproblemen mit zwei voneinander unabhängigen i.i.d. Gruppen mit jeweils endlicher Varianz erfüllt sind, vgl. Del Barrio etc. Al. [12].

Das abschliessende Korollar zeigt, dass die bedingte Verteilungskonvergenz in den obigen Sätzen 3.3 und 3.11 schon die unbedingte impliziert.

Korollar 3.12

Liegt die Konvergenz (3.8) mit $V^2 = \sigma^2 \in \mathbb{R}_{\geq 0}$ vor, so konvergiert T_n^* auch unbedingt gegen dieselbe Grenzverteilung

$$T_n^* \xrightarrow{\mathcal{D}} X \sim N(0, \sigma^2).$$
 (3.29)

Beweis. Dies folgt mit Hilfe von Korollar A.4 aus dem Appendix.

3.1.1 Anwendung: Das schwache Gesetz der großen Zahlen für das Bootstrap-Mittel

Sei $(Y_i)_{i \leq k(n)}$ eine Folge von reellen Zufallsvariablen und $(Y_i^*)_{i \leq m(n)}$ die zugehörige m(n)-Bootstrap Stichprobe mit arithmetischem Mittel $\overline{Y}_{m(n)}^* := \frac{1}{m(n)} \sum_{i=1}^{m(n)} Y_i^*$. Wie auch bei den Zentralen Grenzwertsätzen existieren für das Gesetz der

großen Zahlen Gegenstücke in der Bootstrap-Welt. Hierbei unterscheidet man zwischen dem bedingten schwachen Gesetz der großen Zahlen für das Bootstrap-Mittel

$$d_1(\mathcal{L}(\overline{Y}_{m(n)}^* - \overline{Y}_n | Y_1, \dots, Y_{k(n)}), \varepsilon_0) \xrightarrow{P} 0$$
(3.30)

sowie dem unbedingten

$$\overline{Y}_{m(n)}^* - \overline{Y}_n = \frac{1}{m(n)} \sum_{i=1}^{m(n)} Y_i^* - \frac{1}{k(n)} \sum_{i=1}^{k(n)} Y_j \stackrel{P \otimes \tilde{P}}{\longrightarrow} 0.$$
 (3.31)

Dabei impliziert (3.30) nach Korollar 3.12 immer (3.31).

In der Literatur finden sich zahlreiche Arbeiten, die sich mit geeigneten Bedingungen für diese und stärkere Konvergenzen beschäftigen, vgl. beispielsweise den Übersichtsartikel von Csörgő und Rosalsky [7] und die dort enthaltenen Referenzen. In den meisten Fällen wird dabei jedoch gefordert, dass die Zufallsvariablen $(Y_i)_{i \leq k(n)}$ gewisse Unabhängigkeitsvoraussetzungen erfüllen. Eine Ausnahme ist hierbei der Artikel von Einmahl und Rosalsky [16] aus dem Jahre 2005. Diese zeigen in ihrer Arbeit, dass unter der Bedingung

$$\max_{i \le k(n)} \frac{|Y_i|}{\sqrt{m(n)}} \to 0 \text{ fast sicher}$$
 (3.32)

in (3.30) fast sichere Konvergenz gegen 0 vorliegt. Der zugehörige Beweis beruht dabei auf der interessanten Tatsache, dass die Konvergenz in (3.32) äquivalent zur Konvergenz

$$\frac{Y_{k(n)}}{\sqrt{m(n)}} \to 0 \text{ fast sicher} \tag{3.33}$$

ist. Eine vergleichbare Äquivalenzaussage für P-stochastische Konvergenz gilt aber nicht, da die zweite Implikation verletzt ist.

Um dies einzusehen, betrachtet man eine Folge von i.i.d. standard Fréchet verteilten Zufallsvariablen $(X_i^F)_{i\in\mathbb{N}}$ mit Verteilungsfunktion $P(X_1^F \leq x) = \exp(-1/x) \, \mathbbm{1}_{(0,\infty)}(x)$. Hierfür liegt die stochastische Konvergenz $n^{-1/2} \cdot X_n^F \stackrel{P}{\longrightarrow} 0$ vor. Dies lässt sich jedoch nicht auf die Maximumsstatistik übertragen, da für alle $\epsilon > 0$ gilt

$$P\left(\max_{i \le n} \left| \frac{X_i^F}{\sqrt{n}} \right| \ge \epsilon \right) = 1 - \exp\left(-\frac{n}{\sqrt{n}\epsilon}\right) \to 1.$$

Es wird nun gezeigt, dass schwache Gesetze der großen Zahlen der obigen Form (3.30) bzw. (3.31) in vielen Situationen schon unter schwächeren Voraussetzungen folgen.

Korollar 3.13

Sei $(Y_i)_{i \leq k(n)}$ eine Folge von reellen Zufallsvariablen für die eine der beiden Bedingungen erfüllt ist.

1. Es qelte

$$\max_{i \le k(n)} \frac{\left| Y_i - \overline{Y}_n \right|}{\sqrt{k(n)}} \xrightarrow{P} 0 \tag{3.34}$$

und $\lim_{n\to\infty} m(n)/k(n) = \kappa \in (0,\infty]$.

2. Es gelte

$$\max_{i \le k(n)} \frac{\left| Y_i - \overline{Y}_n \right|}{\sqrt{m(n)}} \xrightarrow{P} 0 \tag{3.35}$$

 $und \lim_{n\to\infty} m(n)/k(n) = 0.$

Dann gilt das schwache Gesetz der großen Zahlen für das Bootstrap-Mittel (3.30).

Beweis. Für den ersten Fall setzt man zunächst

$$X_{n,i}^{(1)} := \frac{Y_i}{k(n)}, \quad i \le k(n),$$

und $\overline{X}_n^{(1)}:=\frac{1}{k(n)}\sum_{i=1}^{k(n)}X_{n,i}^{(1)}$. Dann folgt aus (3.34) schon (3.6) für das zentrierte Dreiecksschema $X_{n,i}^{(1)}-\overline{X}_n^{(1)}$. Außerdem gilt

$$\sum_{i=1}^{k(n)} \left(X_{n,i}^{(1)} - \overline{X}_n^{(1)} \right)^2 \leq k(n) \cdot \max_{i \leq k(n)} \left(X_{n,i}^{(1)} - \overline{X}_n^{(1)} \right)^2$$

$$= \max_{i \leq k(n)} \left(\frac{Y_i - \overline{Y}_n}{\sqrt{k(n)}} \right)^2 \xrightarrow{P} 0$$
(3.36)

und somit (3.7) mit $V^2 = 0$. Für die speziellen m(n)-Bootstrap-Gewichte

$$W_{n,i} = \sqrt{m(n)} \left(\frac{1}{m(n)} M_{n,i} - \frac{1}{k(n)} \right)$$

aus Beispiel 2.4 ergibt sich somit

$$T_{n}^{*}((X_{n,i}^{(1)})_{i \leq k(n)}) = \sqrt{k(n)} \sum_{i=1}^{k(n)} W_{n,i}(X_{n,i}^{1} - \overline{X}_{n}^{1})$$

$$= \frac{1}{\sqrt{k(n)}} \sum_{i=1}^{k(n)} W_{n,i}(Y_{i} - \overline{Y}_{n})$$

$$= \left(\frac{m(n)}{k(n)}\right)^{1/2} \cdot \left(\frac{1}{m(n)} \sum_{i=1}^{k(n)} M_{n,i}Y_{i} - \frac{1}{k(n)} \sum_{i=1}^{k(n)} Y_{i}\right)$$

$$= \left(\frac{m(n)}{k(n)}\right)^{1/2} \cdot \left(\overline{Y}_{m(n)}^{*} - \overline{Y}_{n}\right). \tag{3.38}$$

Da die obigen Gewichte unter beiden Bedingungen die Generalvoraussetzungen erfüllen, vgl. Beispiel 6.2, folgt mit Hilfe von Satz 3.3 die Konvergenz

$$d_1(\mathcal{L}((m(n)/k(n))^{1/2}(\overline{Y}_{m(n)}^* - \overline{Y}_n)|Y_1, \dots, Y_{k(n)}), \varepsilon_0) \xrightarrow{P} 0.$$

Wegen $m(n)/k(n) \to \kappa \in (0, \infty]$ folgt hieraus die Behauptung. Der zweite Fall folgt aus dem folgenden Korollar 3.15.

Bemerkung 3.14

1. Wie man an Gleichung (3.37) erkennen kann, funktioniert der Beweis auch für viele weitere Resampling-Verfahren, für die die Generalvoraussetzungen erfüllt sind. Beispielsweise werden in der Literatur Gewichte der Form

$$W_{n,i} = \frac{1}{\sqrt{m(n)}}\widetilde{W}_{n,i} - \frac{\sqrt{m(n)}}{k(n)}, \quad i \le k(n)$$

für geeignete Dreiecksschemata $(\widetilde{W}_{n,i})_{i \leq k(n)}$ betrachtet. Hierzu zählt neben vielen weiteren Verfahren u.a. der Double-Bootstrap aus Beispiel 2.5.

2. Im Beweis wird nur benötigt, dass die Bedingungen (3.6) und (3.7) bzw. (3.27) mit $V = \sigma = 0$ erfüllt sind. Somit können auch allgemeine Dreiecksschemata $(Y_{n,i})_i$ betrachtet werden.

Die zweite Bemerkung wird im Spezialfall im folgenden Korollar nochmals aufgegriffen und schwächt Teile von Theorem 2 von Einmahl und Rosalsky ab, die für das schwache Gesetz der großen Zahlen nicht benötigt werden.

Korollar 3.15

Sei $(Y_i)_{i \leq k(n)}$ eine Folge von reellen Zufallsvariablen, für die die beiden Bedingungen

$$\max_{i \le k(n)} \frac{\left| Y_i - \overline{Y}_n \right|}{\sqrt{k(n)m(n)}} \xrightarrow{P} 0 \tag{3.39}$$

und

$$\frac{1}{k(n)m(n)} \sum_{i=1}^{k(n)} (Y_i - \overline{Y}_n)^2 \xrightarrow{P} 0 \tag{3.40}$$

gelten. In diesem Fall folgt das bedingte schwache Gesetz der großen Zahlen für das Bootstrap-Mittel (3.30).

Beweis. Wiederholt man den Beweis von Korollar 3.13 für den Fall $m(n)/k(n) \rightarrow \kappa > 0$ mit dem Dreiecksschema

$$X_{n,i} := \frac{Y_i}{\sqrt{k(n)m(n)}}, \quad i \le k(n),$$

so fällt der Vorfaktor in Zeile (3.38) weg und man erhält die bedingte stochastische Konvergenz von

$$T_n^*((X_{n,i})_{i \le k(n)}) = \overline{Y}_{m(n)}^* - \overline{Y}_n$$

gegen 0, d.h. (3.30). Wie im ersten Fall erhält man auch für $m(n)/k(n) \to 0$ die Bedingungen (3.6) und (3.27) mit $\sigma^2 = 0$, woraus unmittelbar auch (3.28) folgt. Durch Anwendung von Satz 3.11 ergibt sich so wieder die Behauptung.

Da die Bedingung (3.34) bzw. (3.35) schon (3.39) und (3.40) implizieren, ist hierin auch das erste Korollar 3.13 enthalten. Um dies einzusehen, muss man in Zeile (3.36) nur $X_{n,i}^{(1)}$ durch $X_{n,i}$ ersetzen und die zwei Fälle $m(n)/k(n) \to \kappa \in (0,\infty]$ sowie $m(n)/k(n) \to 0$ betrachten.

Als Spezialfall erhält man aus dem folgenden Korollar bedingte schwache Gesetze der grossen Zahlen für unkorellierte Zufallsvariablen.

Korollar 3.16

Erfüllt die Folge $(Y_i)_{i \leq k(n)}$ die Voraussetzungen

$$\frac{1}{m(n)k(n)} \sum_{i=1}^{k(n)} E(Y_i^2) \to 0$$
 (3.41)

und

$$\frac{1}{\sqrt{m(n)}} \frac{1}{k(n)} \sum_{i=1}^{k(n)} Y_i \xrightarrow{P} 0, \tag{3.42}$$

so folgt das bedingte schwache Gesetz der großen Zahlen für das Bootstrap-Mittel (3.30).

Beweis. Sei $X_{n,i}$ das Dreiecksschema aus dem Beweis von Korollar 3.15. Dann impliziert die Tschebyscheff Ungleichung die Bedingung (3.39)

$$P(\max_{i \le k(n)} |X_{n,i}| \ge \epsilon) \le \sum_{i=1}^{k(n)} P(|X_{n,i}| \ge \epsilon) \le \frac{1}{\epsilon^2} \frac{1}{m(n)k(n)} \sum_{i=1}^{k(n)} E(Y_i^2) \to 0.$$

Da Bedingung (3.40) eine unmittelbare Folgerung von (3.41) und (3.42) ist, erhält man die Aussage aus Korollar 3.15.

Bemerkung 3.17

Ist die Folge Y_i unkorelliert sowie zentriert, d.h. $E(Y_i) = 0$, und gilt $m(n)/k(n) \to \tilde{\kappa} \in [0, \infty)$, so impliziert (3.41) schon das klassische Gesetz der großen Zahlen $\overline{Y}_n \xrightarrow{P} 0$. Da hieraus insbesondere (3.42) folgt, erhält man auch für diesen Fall die bedingten Gesetze der grossen Zahlen (3.30) und (3.31).

Als weitere Anwendung erhält man Gesetze der großen Zahlen für stark stationäre Zufallsvariablen mit endlicher Varianz. Anders als bei EINMAHL UND ROSALSKY [16, Corollary 2] benötigt man dabei keine weiteren Voraussetzungen.

Korollar 3.18

Ist die Folge $(Y_i)_i$ stark stationär mit $E(Y_1^2) < \infty$, so erfüllt sie die schwachen Gesetze der großen Zahlen für das Bootstrap-Mittel (3.30) und (3.31).

Beweis. Dies folgt unmittelbar aus Korollar 3.7.

EINMAHL UND ROSALSKY [16, Corollary 3] haben als Konsequenz hiervon nun schwache Gesetze der großen Zahlen für den Bootstrap einer speziellen stark stationären Zeitreihen angegeben. Dies gilt nach obigem Korollar nun auch in größerer Allgemeinheit für jede stark stationäre Zeitreihe mit endlicher Varianz. Allerdings ist die praktische Bedeutung dieser Aussage für die Testtheorie und die damit verbundene Berechnung kritischer Werte von Teststatistiken erstmal

gering, da dort nicht entartete Grenzverteilungen von Bedeutung sind. Um diese zu erhalten, muss man jedoch die Statistik aus (3.30) und (3.31) noch mit einem geeigneten Normierungsfaktor versehen, vgl. z.B. Beispiel 2.3. Wie dort bereits gesehen, ist das klassische Bootstrap-Verfahren selbst für schwach abhängige Variablen nicht konsistent. Da man die obigen Beweise allerdings nach Bemerkung 3.14 auch für andere Resampling-Gewichte durchführen kann, erhält man für die stationären Zufallsvariablen aus obigem Korollar unter geeigneten Zusatzvoraussetzungen z.B. Gesetze der großen Zahlen für das Nonoverlapping Block-Bootstrap-Verfahren, auf das in Kapitel 6 noch etwas genauer eingegangen wird.

3.2 Höhere Dimensionen

Da man in der Praxis häufig mit mehrdimensionalen Datensätzen konfrontiert wird, benötigt man auch in diesem Fall Zentrale Grenzwertsätze, die als Entscheidungshilfe dienen können. Aus diesem Grund wird der bedingte Zentrale Grenzwertsatz aus Theorem 3.3 in den folgenden Abschnitten auf den Fall höherer Dimensionen übertragen.

3.2.1 Der multivariate Fall

Sei $k \in \mathbb{N}$. Beim klassischen Beweis des multivariaten Zentralen Grenzwertsatzes für i.i.d. Zufallsvariablen wird das mehrdimensionale Problem mittels des Cramér-Wold-Devices auf ein eindimensionales Problem reduziert. Für bedingte Grenzwertsätze der Form (3.8) lässt sich dieser Ansatz nicht direkt anwenden, da man die Verteilungskonvergenz jeder beliebigen Linearkombination unter Umständen nicht auf einer gemeinsamen 1-Menge nachweisen kann⁸. Aus diesem Grund wird der klassische Cramér-Wold-Device im nächsten Satz geeignet modifiziert. Als Spezialfall erhält man hierbei insbesondere Satz 8.2. von Pauls [49].

Dabei bezeichnet d_k eine Metrik der schwachen Konvergenz auf $\mathcal{M}_1(\mathbb{R}^k, \mathcal{B}^k)$.

Satz 3.19

Seien P_n und Q_n zwei Folgen von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^k, \mathcal{B}^k)$ und D eine dichte Teilmenge des \mathbb{R}^k , die alle Einheitsvektoren $e_j, 1 \leq j \leq k$, enthält. Für $\lambda \in \mathbb{R}^k$ sei durch P^{λ} das Bildmaß von $P \in \mathcal{M}_1(\mathbb{R}^k, \mathbb{B}^k)$ unter der Abbildung $\mathbb{R}^k \ni x \to x^t \lambda$ definiert. Dann gelten:

- (a) P_n ist genau dann straff, wenn $P_n^{e_j}$ straff ist für alle $1 \le j \le k$.
- (b) Ist Q_n straff und gilt

$$d_1(P_n^{\lambda}, Q_n^{\lambda}) \to 0 \text{ für alle } \lambda \in D,$$
 (3.43)

so folgt

$$d_k(P_n, Q_n) \to 0. \tag{3.44}$$

⁸Es treten überabzählbar viele Nullmengen auf.

⁹Zur Auswahl der Metrik vgl. z.B. WITTING UND MÜLLER-FUNK [66, Anmerkung 5.42 a)].

Beweis. (a) ist leicht zu zeigen und folgt analog zu PAULY [50, Lemma 2.7]. Wir zeigen nun (b). Nach (a) ist auch $Q_n^{e_j}$ für jedes $1 \leq j \leq k$ straff, so dass es nach dem Satz von Prohorov, vgl. BILLINGSLEY [6, S.35 f.], zu jeder Teilfolge eine weitere Teilfolge $\{m_1\}$ und ein Wahrscheinlichkeitmaß $Q_0 \in \mathcal{M}_1(\mathbb{R}, \mathcal{B})$ gibt mit

$$d_1(Q_{m_1}^{e_j}, Q_0) \to 0.$$

Mit Hilfe der Dreiecksungleichung folgt hieraus die Straffheit von $P_n^{e_j}$ für alle $1 \leq j \leq k$, was nach (a) wiederum äquivalent zur Straffheit von P_n ist. Nun existiert nach Definition eine Teilfoge $\{m\} \subset \mathbb{N}$ mit

$$d_k(P_m, Q_m) \to a := \limsup_{n \to \infty} d_k(P_n, Q_n).$$

Ausserdem existieren aufgrund der Straffheit von P_n und Q_n Wahrscheinlichkeitsmaße $P,Q \in \mathcal{M}_1(\mathbb{R}^k,\mathcal{B}^k)$ und eine Teilfolge $\{m_2\} \subset \{m\}$, so dass P_{m_2} schwach gegen P und Q_{m_2} schwach gegen Q konvergieren. Die Anwendung der Dreiecksungleichung ergibt somit die Konvergenz

$$d_k(P_{m_2}, Q_{m_2}) \to d_k(P, Q) = a.$$

Hieraus folgt durch Anwendung des Cramér-Wold-Devices für alle $\lambda \in \mathbb{R}^k$

$$d_1(P_{m_2}^{\lambda}, Q_{m_2}^{\lambda}) \to d_1(P^{\lambda}, Q^{\lambda}).$$

Da aber nach Voraussetzung $d_1(P^{\lambda}, Q^{\lambda}) = 0$ für alle $\lambda \in D$ gilt, folgt aus dem Eindeutigkeitssatz die Gleichheit der zugehörigen Fouriertransformierten

$$\widehat{P}^{\lambda} = \widehat{Q}^{\lambda}$$
 für alle $\lambda \in D$.

Aufgrund der Stetigkeit der Abbildung $\lambda \to \widehat{P^{\lambda}}(y)$ ergibt sich

$$\widehat{P^{\lambda}} = \widehat{Q^{\lambda}}$$
 für alle $\lambda \in \mathbb{R}^k$

und somit wiederum aus dem Eindeutigkeitssatz auch

$$P^{\lambda} = Q^{\lambda}$$
 für alle $\lambda \in \mathbb{R}^k$.

Der Cramér-Wold-Device liefert schließlich die Behauptung.

Da im Folgenden mehrdimensionale Normalverteilungen $N(\mu, \Sigma)$ mit Erwartungswert μ und Kovarianzmatrix Σ auftreten, wird die Menge der symmetrischen, positiv semidefiniten bzw. positiv definiten $(k \times k)$ -Matrizen über den

reellen Zahlen zur Vereinfachung mit $\mathbb{R}_{p.s.}^{k \times k}$ bzw. $\mathbb{R}_{p.d.}^{k \times k}$ bezeichnet (vgl. WITTING [65, S.96, Fussnote 1]).

Im Mittelpunkt der Betrachtungen stehen nun multivariate Resampling-Statistiken T_n^* der Form (3.1) in Dreiecksschemata von \mathbb{R}^k -wertigen Zufallsvariablen $(X_{n,i})_{i \leq k(n)}$ mit $X_{n,i} = (X_{n,i}^{(1)}, \dots, X_{n,i}^{(k)}), i \leq k(n)$, auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Das arithmetische Mittel der j-ten Koordinaten sei definiert durch $\overline{X}_n^{(j)} := \frac{1}{k(n)} \sum_{i=1}^{k(n)} X_{n,i}^{(j)}$.

Analog zu Bedingung (3.7) ist nun die Konvergenz der empirischen Kovarianzmatrix

$$\Gamma_n := \Gamma_n(X_{n,i}) := \left(\sum_{i=1}^{k(n)} (X_{n,i}^{(j)} - \overline{X}_n^{(j)}) (X_{n,i}^{(r)} - \overline{X}_n^{(r)}) \right)_{1 \le j,r \le k}$$
(3.45)

von Interesse, wie die folgende Verallgemeinerung von Satz 3.3 zeigt. Dabei seien wieder alle auftretenden Zufallsvariablen mittels Projektionen auf dem gemeinsamen Wahrscheinlichkeitsraum $(\Omega \times \tilde{\Omega}, \mathcal{A} \otimes \tilde{\mathcal{A}}, P \otimes \tilde{P})$ definiert.

Satz 3.20

Gelten die Voraussetzungen (3.2)-(3.5) mit $Var(W_1) = 1$ sowie

$$\max_{i \le k(n)} \left| X_{n,i}^{(j)} \right| \to 0 \ P\text{-}stochastisch$$
 (3.46)

für alle $1 \le j \le k$ und liegt Verteilungskonvergenz

$$\Gamma_n = \left(\sum_{i=1}^{k(n)} (X_{n,i}^{(j)} - \overline{X}_n^{(j)}) (X_{n,i}^{(r)} - \overline{X}_n^{(r)})\right)_{1 < j,r < k} \xrightarrow{\mathcal{D}} \Gamma$$
 (3.47)

gegen eine Zufallsmatrix Γ vor, dann folgt

$$d_k(\mathcal{L}(T_n^*|(X_{n,i})_{i< k(n)}), \mathcal{L}(Z)) \xrightarrow{P} 0$$
(3.48)

für eine Zufallsvariable Z mit $Z(\omega,\cdot) \sim N(0,\Gamma(\omega))$.

Beweis. Sei D := $\{\lambda_l : l \in \mathbb{N}\}$ eine abzählbar dichte Teilmenge des \mathbb{R}^k , die alle Einheitsvektoren e_i enthält. Sei $\langle \cdot, \cdot \rangle$ das euklidische Skalarprodukt auf \mathbb{R}^k . Im Folgenden wird Satz 3.3 auf Folgen $M_{n,i,1} := \langle \lambda_1, X_{n,i} \rangle$ für ein festes $\lambda_1 = (\lambda_1^{(j)})_{j \leq k} \in \mathbb{D}$ angewendet. Hierfür gilt

$$\max_{i \le k(n)} \left| M_{n,i,1} - \overline{M}_{n,1} \right| \le k \cdot \max_{j \le k} \left| \lambda_1^{(j)} \right| \cdot \max_{i \le k(n)} \left| X_{n,i}^{(j)} - \overline{X}_n^{(j)} \right| \stackrel{P}{\longrightarrow} 0.$$

Der Satz über stetige Abbildungen liefert nach Voraussetzung (3.47) die Verteilungskonvergenz

 $\lambda_1^t \Gamma_n \lambda_1 = \sum_{i=1}^{k(n)} \langle \lambda_1, X_{n,i} \rangle^2 \xrightarrow{\mathcal{D}} \lambda_1^t \Gamma \lambda_1.$

Durch Anwendung von Satz 3.3 auf eindimensionale Resampling-Statistiken der Form

$$T_{n,1}^* := \langle T_n^*, \lambda_1 \rangle = (k(n))^{\frac{1}{2}} \sum_{i=1}^{k(n)} W_{n,i}(\langle \lambda_1, X_{n,i} \rangle - \langle \lambda_1, \overline{X}_n \rangle)$$

erhält man entlang einer Teilfolge $\{m_1\}$ für alle $\omega \in N_1^c$ mit $P(N_1)=0$ punktweise Konvergenz

$$d_1(\mathcal{L}(\langle T_{m_1}^*(\omega), \lambda_1 \rangle | (X_{n,i})_{i \le k(n)})(\omega, \cdot), N(0, \lambda_1^t \Gamma(\omega) \lambda_1)) \to 0.$$
 (3.49)

Genauso findet man für $\lambda_2 \in D$ eine weitere Teilfolge $\{m_2\} \subset \{m_1\}$ und eine Nullmenge N_2 , so dass (3.49) auch für λ_2 und alle $\omega \in N_2^c$ entlang dieser Teilfolge richtig ist. So fortfahrend, erhält man eine gemeinsame Teilfolge $\{m\}$ und eine gemeinsame Nullmenge N, so dass für alle $\lambda_k \in D$ und $\omega \in N^c$ schwache Konvergenz (3.49) entlang $\{m\}$ vorliegt.

Da nach Voraussetzung alle Einheitsvektoren e_j , $1 \leq j \leq k$, in D enthalten sind, ist die Folge $\mathcal{L}(< T_m^*, e_j > |(X_{m,i})_{i \leq k(n)})(\omega, \cdot)$ straff. Nach Teil (a) von Satz 3.19 folgt hieraus die Straffheit von $\mathcal{L}(T_m^*|(X_{m,i})_{i \leq k(n)})(\omega, \cdot)$. Die punktweise Anwendung von Satz 3.19 für $\omega \in N^c$ ergibt damit

$$d_k(\mathcal{L}(T_m^*|(X_{n,i})_{i\leq k(n)})(\omega,\cdot), N(0,\Gamma(\omega)))\to 0.$$

Eine erneute Anwendung des Teilfolgenkriteriums für stochastische Konvergenz zeigt demnach die Behauptung.

Bemerkung 3.21

- 1. Man rechnet leicht nach, dass $\Gamma_n(\omega)$ für alle $\omega \in \Omega$ eine Kovarianzmatrix ist.
- 2. Aufgrund der Verteilungskonvergenz (3.47) ist auch $\Gamma(\omega)$ für fast alle $\omega \in \Omega$ nach dem Portmanteu-Theorem eine Kovarianzmatrix.

3.2.2 Diskrete Prozesse

Wir betrachten in diesem Teilabschnitt ein Dreiecksschema von diskreten Prozessen

$$(X_{n,i})_{i \le k(n)} : \Omega \to \mathbb{R}^{\mathbb{N}}$$
(3.50)

mit $X_{n,i} = (X_{n,i}^{(j)})_{j \in \mathbb{N}}$, wobei $\mathbb{R}^{\mathbb{N}}$ als metrischer Raum die Produkttopologie trägt, und untersuchen den diskreten Resampling-Prozess der Form (3.1)

$$T_n^* = \sqrt{k(n)} \cdot \sum_{i=1}^{k(n)} W_{n,i} (X_{n,i} - \overline{X}_n).$$

Um auch hierfür Grenzwertsätze angeben zu können, wird im Folgenden eine Aussage von Billingsley [6, S.29f] aus dem Abschnitt über stetige Abbildungen zitiert. Dabei sei für $k \in \mathbb{N}$ die kanonische Projektionen von $\mathbb{R}^{\mathbb{N}}$ auf \mathbb{R}^k gegeben durch $\pi_k : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^k, (x_i)_{i \in \mathbb{N}} \mapsto (x_1, \dots, x_k)$.

Lemma 3.22

Für Wahrscheinlichkeitsmaße $P_n, n \in \mathbb{N}_0$ auf $\mathcal{M}(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ sind äquivalent:

(a)
$$P_n \xrightarrow{w} P_0$$
 und

(b)
$$P_n^{\pi_k} \xrightarrow{w} P_0^{\pi_k}$$
 für alle $k \in \mathbb{N}$.

Hiermit lassen sich die obigen Grenzwertsätze auf diskrete Prozesse verallgemeinern. Dabei metrisiert $d_{\mathbb{N}}$ die schwache Konvergenz auf dem polnischen Raum $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$, vgl. Appendix S.125f..

Satz 3.23

Gelten die Voraussetzungen (3.2)-(3.5) mit $Var(W_1) = 1$ sowie

$$\max_{i \le k(n)} \left| X_{n,i}^{(j)} \right| \to 0 \ P\text{-stochastisch}$$
 (3.51)

für alle $j \in \mathbb{N}$ und liegt Verteilungskonvergenz

$$\Gamma_n := \left(\sum_{i=1}^{k(n)} (X_{n,i}^{(j)} - \overline{X}_n^{(j)}) (X_{n,i}^{(r)} - \overline{X}_n^{(r)})\right)_{j,r \in \mathbb{N}} \xrightarrow{\mathcal{D}} \Gamma \tag{3.52}$$

gegen eine Zufallsvariable Γ vor, so folgt

$$d_{\mathbb{N}}(\mathcal{L}(T_n^*|(X_{n,i})_{i < k(n)}), \mathcal{L}(Z)) \xrightarrow{P} 0 \tag{3.53}$$

für eine Zufallsvariable Z mit $Z(\omega, \cdot) \sim N(0, \Gamma(\omega))$.

Beweis. Dies folgt aus Satz 3.20 durch Anwendung von Lemma 3.22. ■

Eine Anwendung hiervon findet man später in dem Abschnitt über das Resampling Quadratischer Formen, vgl. Kapitel 4.2.

Kapitel 4

Bedingte Zentrale Grenzwertsätze für nicht-lineare Resampling-Statistiken

Ist $\tilde{T}_n:(\Omega,\mathcal{A},P)\to(\mathbb{R},\mathbb{B})$ anders als in Kapitel 3 eine komplizierte, nichtlineare Statistik von Interesse, so sind die dort entwickelten Methoden i.a. nicht mehr anwendbar. Aus diesem Grund werden in den beiden nächsten Abschnitten Voraussetzungen untersucht, unter denen Resampling-Versionen von \tilde{T}_n noch konsistent sind. Dabei liegt das Hauptaugenmerk zum einen auf quadratischen Statistiken und zum anderen auf Statistiken, die eine asymptotisch lineare Entwicklung besitzen.

4.1 Resampling asymptotisch linearer Statistiken

Sei $\tilde{T}_n:(\Omega,\mathcal{A},P)\to(\mathbb{R},\mathbb{B})$ wie zu Beginn eine nicht-lineare Statistik von Interesse. Dann existiert hierfür häufig unter geeigneten Regularitätsbedingungen eine Entwicklung in einen linearen und einen asymptotisch vernachlässigbaren Anteil, d.h. es gilt

$$\tilde{T}_n := \sum_{i=1}^{k(n)} c_{n,i} X_{n,i} + R_n =: T_n + R_n \tag{4.1}$$

für adäquate Regressionskoeffizienten $(c_{n,i})_{i \leq k(n)}$ und Dreiecksschema von Zufallsvariablen $X_{n,i}: (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathbb{B}), i \leq k(n)$, sowie $R_n = o_P(1)$, d.h.

 $R_n \stackrel{P}{\longrightarrow} 0$. Konvergieren nun die Verteilung des linearen Anteils T_n und die bedingte Verteilung der hierzu gehörigen linearen Resampling-Statistik T_n^* , wie beispielsweise in Kapitel 3, schwach gegen dieselbe Grenzverteilung, so stellt sich die Frage, ob sich diese Konsistenz auch auf \tilde{T}_n und \tilde{T}_n^* überträgt. In dieser Situation liegt nach dem bedingten Lemma von Slutzky¹ genau dann Konsistenz des gewählten Resampling-Verfahrens vor, wenn die Resampling-Version R_n^* von R_n ebenfalls bedingt gegen 0 konvergiert.

Dabei wird im Folgenden gezeigt, dass dies u.a. für Randomisationsverfahren - wie beispielsweise das Permuationsverfahren - erfüllt ist, falls die zugrunde liegende Verteilung invariant unter der Randomisationsgruppe ist.

Dass sich diese schöne Schlußweise, selbst in den leichtesten Fällen nicht auf das Bootstrap-Verfahren übertragen lässt, zeigt das nächste Beispiel.

Beispiel 4.1 (Bootstrap und Restterme)

Seien X_1, \ldots, X_n reelle i.i.d. Zufallsvariablen mit stetiger Verteilungsfunktion F. Zu der Menge

$$A_n := \{(x_1, \dots, x_n) \in \mathbb{R}^n : \text{ es existieren } i \neq j \text{ mit } x_i = x_j\}$$

und einer beliebigen Folge $(a_n)_n \subset \mathbb{R}$ mit $a_n \to +\infty$ definiert man die reelle Statistik

$$R_n = R_n(X_1, \dots, X_n) := a_n \cdot \mathbb{1}_{A_n}(X_1, \dots, X_n).$$
 (4.2)

Hierfür gilt offensichtlich $R_n = 0$ P-f.s., also insbesondere $R_n \to 0$ P-stochastisch. Die klassische Bootstrap-Version² von R_n

$$R_n^* := R_n(X_1^*, \dots, X_n^*)$$

erfüllt dies aber nicht. Man kann sogar zeigen, dass R_n^* bedingt nach den Daten stochastisch gegen $+\infty$ konvergiert. Sei dazu $\kappa > 0$ beliebig gewählt, so gilt für alle $n \geq n_0(\kappa)$

$$P(R_n^* \ge \kappa | X_1, \dots, X_n) = P(\exists i \ne j : X_i^* = X_j^* | X_1, \dots, X_n)$$
$$= 1 - 1(1 - \frac{1}{n})(1 - \frac{2}{n}) \cdot \dots \cdot \frac{1}{n} \to 1.$$

¹Vgl. insbesondere Lemma A.9 im Appendix.

²Dabei macht es in dieser Situation keinen Unterschied, ob man aus der Menge $\{X_i: i \leq n\}$ oder $\{X_i - \overline{X}_n: i \leq n\}$ mit Zurücklegen zieht.

Dieses Gegenbeispiel funktioniert somit für alle Resampling-Verfahren mit zugehöriger Resampling-Stichprobe X^* , für die

$$P(\exists i \neq j : X_i^* = X_i^* | X_1, \dots, X_n) \to 1$$

gilt. Außerdem kann man aus obigem Beispiel entnehmen, dass selbst stärkere Momentenbedingungen wie $E(|R_n|^p) \to 0$ für $p \ge 1$ nicht die bedingte stochastische Konvergenz von R_n^* gegen 0 nach sich ziehen.

4.1.1 Restterme unter Suffizienz

An dieser Stelle werden nun Bedingungen gesucht, unter denen man von der unbedingten Konvergenz auf die bedingte Konvergenz schließen kann. Um dabei möglichst allgemein zu bleiben, zielen die Untersuchungen auf das zugehörige asymptotische Verhalten bedingter Tests ab. Als Spezialfall sind hierin Randomisationstests wie beispielsweise Permutationstests enthalten.

In diesem Abschnitt sei vorausgesetzt, dass $(\Omega_n, \mathcal{A}_n, \mathcal{P}_n)$ für jedes $n \in \mathbb{N}$ ein statistisches Experiment sei, d.h. \mathcal{P}_n ist eine Familie von Wahrscheinlichkeitsmaßen auf dem Messraum $(\Omega_n, \mathcal{A}_n)$. Mit $\mathcal{P}_{0,n} \subset \mathcal{P}_n$ wird dann eine zusammengesetzte Nullhypothese von Interesse bezeichnet. Das Testproblem ist dann durch

$$\mathcal{P}_{0,n} \text{ gegen } \mathcal{P}_n \backslash \mathcal{P}_{0,n}$$
 (4.3)

gegeben. Darüber hinaus sei

$$S_n: (\Omega_n, \mathcal{A}_n) \to (\Omega_{S_n}, \mathcal{A}_{S_n})$$
 (4.4)

stets eine $\mathcal{P}_{0,n}$ -suffiziente Statistik mit Werten in dem Messraum $(\Omega_{S_n}, \mathcal{A}_{S_n})$. Als nächstes wird wiederholt, was man unter einem bedingten Test versteht.

Definition 4.2

Ein Test $\psi: \Omega \to [0,1]$ für das obige Testproblem (4.3) heißt bedingter Test oder Test mit Neyman Struktur, falls für die zugehörige bedingte Erwartung gilt

$$E.(\psi|S_n)$$
 ist konstant unter $\mathcal{P}_{0,n}$, (4.5)

vgl. z.B. Janssen und Völker [36].

Bedingte Tests, die durch suffiziente σ -Algebren $\mathcal{F}_n \subset \mathcal{A}_n$ gegeben sind, definiert man mittels der suffizienten Statistik

$$S_n = id : (\Omega_n, \mathcal{A}_n) \to (\Omega_n, \mathcal{F}_n).$$
 (4.6)

Sei nun

$$T_n: (\Omega_n, \mathcal{A}_n) \to \mathbb{R}$$
 (4.7)

eine beliebige Teststatistik.

Es wird nun gezeigt, wie der obere T_n -Test als (S_n) -bedingter Test $\varphi_{n,\alpha}^*$ ausgeführt werden kann. Dazu sei zunächst bemerkt, dass die bedingte Verteilung von T_n gegeben $S_n = s$ durch einen stochastischen Kern $K_{T_n}(\cdot, \cdot) : \Omega_{S_n} \times \mathcal{B} \to [0, 1]$ repräsentiert werden kann. Genauer gilt

$$K_{T_n}(s,\cdot) = P_n^{T_n|S_n=s}(\cdot) \text{ für alle } P_n \in \mathcal{P}_{0,n}, \tag{4.8}$$

vgl. z.B. Dudley [8, Kapitel 10.2]. Sei $\alpha \in (0,1)$. Dann basiert der bedingte α -Niveau Test

$$\varphi_{n,\alpha}^* = \mathbb{1}_{(c_n^*(S_n),\infty)}(T_n) + \gamma_n(S_n)\mathbb{1}_{\{c_n^*(S_n)\}}(T_n)$$
(4.9)

auf der \mathcal{A}_{S_n} messbaren Quantilfunktion c_n^* von (4.8) an der Stelle (1 – α) und auf der Randomisierung $0 \le \gamma(S_n) \le 1$. Diese erhält man als Lösungen von

$$\int \varphi_{n,\alpha}^* dK_{T_n}(s,\cdot) = \alpha \text{ für } s \in \Omega_{S_n}.$$
(4.10)

In diesem allgemeinen Ansatz sind z.B. Randomisationstests enthalten, vgl. beispielsweise Janssen und Völker [36, Example 2.3] für konstanten Stichprobenumfang n.

Beispiel 4.3 (Randomisationstests)

Sei $\mathcal{G}_n = \{g_n : (\Omega_n, \mathcal{A}_n) \to (\Omega_n, \mathcal{A}_n)\}$ eine endliche Gruppe von Transformationen und $\mathcal{P}_{0,n}$ die Nullhypothese der \mathcal{G}_n -invarianten Verteilungen, d.h.

$$\mathcal{P}_{0,n} = \{ P_n : P_n^{g_n} = P_n \text{ für alle } g_n \in \mathcal{G}_n \}. \tag{4.11}$$

Weiter bezeichne $\mathcal{F}_n := \{A \in \mathcal{A}_n : g_n(A) = A \text{ für alle } g_n \in \mathcal{G}_n\}$ die σ -Algebra der \mathcal{G}_n -invarianten Mengen und $S_n = id : (\Omega_n, \mathcal{A}_n) \to (\Omega_n, \mathcal{F}_n)$ die zugehörige $\mathcal{P}_{0,n}$ -suffiziente Statistik. Die bedingte Randomisationsverteilung $K_{T_n}(\omega, \cdot) = P_n^{T_n|S_n=\omega}$ der Statistik T_n ist dann durch

$$K_{T_n}(\omega, \cdot) = \frac{1}{|\mathcal{G}_n|} \sum_{g_n \in \mathcal{G}_n} \varepsilon_{T_n(g_n(\omega))}$$
(4.12)

gegeben. In diesem Fall heißt der durch (4.9) definierte bedingte Test $\varphi_{n,\alpha}^*$ Randomisationstest. Als Spezialfall erhält man

- (a) **Permutationtests:** In diesem Fall ist $\mathcal{P}_{0,n}$ die Familie der austauschbaren Verteilungen auf dem Produktraum $(\Omega_n, \mathcal{A}_n)$ und \mathcal{G}_n die Gruppe aller Koordinatenpermutationen von Ω^n .
- (b) **Symmetrietests:** Hier betrachtet man den Produktraum $(\Omega_n, \mathcal{A}_n) = (\mathbb{R}^n, \mathcal{B}^n)$ und die Gruppe \mathcal{G}_n der Spiegelungen $(x_1, \ldots, x_n) \mapsto (\epsilon_1 x_1, \ldots, \epsilon_n x_n)$, die durch die Vektoren $(\epsilon_1, \ldots, \epsilon_n) \in \{-1, 1\}^n$ gegeben ist. Als \mathcal{G}_n -invariante Verteilungen erhält man z.B. 0-symmetrische Produktmaße

$$\mathcal{P}_{0,n} = \{ P^n : P \text{ ist } 0\text{-symmetrisch auf } (\mathbb{R}, \mathcal{B}) \}. \tag{4.13}$$

Die Asymptotik dieser allgemeinen bedingten Tests wird nun im Folgenden untersucht. Hierzu wird für eine feste Folge $P_n \in \mathcal{P}_{0,n}$ die Bedingung (A) von Lemma 2.7 auf die aktuelle Situation übertragen. Dabei sei $T: (\Omega_0, \mathcal{A}_0, P_0) \to \mathbb{R}$ eine Statistik auf einem weiteren Wahrscheinlichkeitsraum $(\Omega_0, \mathcal{A}_0, P_0)$ mit Verteilung $\mathcal{L}(T) = \mu$.

(A) Die Verteilung der Statistik T_n unter der Folge $P_n \in \mathcal{P}_{0,n}$ konvergiere schwach

$$\mathcal{L}(T_n|P_n) \xrightarrow{w} \mathcal{L}(T) = \mu \tag{4.14}$$

gegen ein Wahrscheinlichkeitsmaß μ . Dabei wird angenommen, dass die Verteilungsfunktion von μ stetig und streng monoton wachsend auf ihrem Träger ist.

An dieser Stelle sei bemerkt, dass man in der vorliegenden Situation alle auftretenden Zufallsvariablen mittels Projektionen³ auf dem Produktraum

$$\left(\Omega := \prod_{n=0}^{\infty} \Omega_n, \mathcal{A} := \bigotimes_{n=0}^{\infty} \mathcal{A}_n, P := \bigotimes_{n=0}^{\infty} P_n\right)$$

definieren kann. Nach der definierenden Gleichung für bedingte Verteilungen gilt dann auch $P^{T_n|S_n} = K_{T_n}(S_n, \cdot)$ P-fast sicher. Diese Darstellung wird im weiteren Verlauf des Kapitels stets verwendet.

 $[\]overline{^{3}\text{D.h. z.B. }T_{n}:\Omega\ni\omega=(\omega_{i})_{i\in\mathbb{N}_{0}}}\mapsto T_{n}(\omega):=T_{n}(\omega_{n})\in\mathbb{R}.$

Der Vollständigkeit halber wird nun Lemma 2.7 mit Hilfe der vorigen Bemerkung auf die obige Situation angepasst. Sei dazu wieder $\varphi_{n,\alpha}$ der zu T_n gehörige unbedingte obere T_n -Test.

Lemma 4.4

Unter der Voraussetzung (A) sind die beiden folgenden Aussagen äquivalent:

$$d(K_{T_n}(S_n,\cdot),\mu) \xrightarrow{P_n} 0, \tag{4.15}$$

$$E_{P_n}(|\varphi_{n,\alpha} - \varphi_{n,\alpha}^*|) \to 0 \text{ für alle } \alpha \in (0,1).$$
 (4.16)

Nachdem alle Vorbereitungen getroffen sind, können nun Antworten auf die anfängliche Problemstellung gegeben werden. Sei dazu \tilde{T}_n eine Teststatistik, die eine Entwicklung wie in (4.1) in einen linearen Anteil T_n und einen asymptotisch vernachlässigbaren Anteil $R_n: (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}, \mathcal{B})$, mit $R_n \xrightarrow{P_n} 0$, besitzt. Das nächste Lemma gibt Auskunft über das Verhalten der bedingten Verteilung $K_{R_n}(S_n,\cdot)$ von R_n unter einer suffizienten σ -Algebra. Sei dazu wieder $P_n \in \mathcal{P}_{0,n}$ beliebig aber fest gewählt.

Lemma 4.5

Unter den obigen Voraussetzung und (A) ist die Aussage (4.15) äquivalent zu

$$d_1(K_{T_n+R_n}(S_n,\cdot),\mu) \xrightarrow{P_n} 0. \tag{4.17}$$

Beweis. Durch Anwendung des Lemmas von Slutzky erhält man aus (A) zunächst

$$\mathcal{L}(T_n + R_n) \xrightarrow{w} \mathcal{L}(T).$$
 (4.18)

Somit reicht es aus nur eine Beweisrichtung zu zeigen (betrachte sonst $\tilde{T}_n := T_n + R_n$ und $\tilde{T}_n + \tilde{R}_n$ mit $\tilde{R}_n := -R_n$). Sei also die Konvergenz (4.15) gegeben. Für $\alpha \in (0,1)$ bezeichnet man mit $c(\alpha)$ das $(1-\alpha)$ -Quantil von $\mathcal{L}(T)$ und definiert den Test $\psi_n := \mathbb{1}_{(c(\alpha),\infty)}(T_n + R_n)$. Dieser ist aufgrund von (4.18) asymptotisch α -ähnlich, d.h. es gilt $E_P(\psi_n) \to \alpha$. Nun existieren wie in (4.8) - (4.10) S_n -messbare Funktionen c_n, γ_n , so dass $\psi_n^* := \mathbb{1}_{(c_n(\alpha),\infty)}(T_n + R_n) + \gamma_n \mathbb{1}_{\{c_n(\alpha)\}}(T_n + R_n)$ ein S_n -bedingter Test mit der Eigenschaft

$$\int \psi_n^* dK_{T_n + R_n}(s, \cdot) = \alpha \text{ für alle } s \in \Omega_{S_n}$$
(4.19)

ist. Für den Nachweis von (4.17) reicht es nach (4.18) und Lemma 4.4 aus die asymptotische Äquivalenz dieser Tests

$$E_P(|\psi_n - \psi_n^*|) \to 0$$
 für alle $\alpha \in (0,1)$

nachzuweisen. Hierfür werden noch zwei weitere Tests benötigt. Dazu definiert man wieder mittels (4.8) - (4.10) die Tests

$$\phi_n^* := \mathbb{1}_{(b_n(\alpha),\infty)}(T_n) + \tau_n \mathbb{1}_{\{b_n(\alpha)\}}(T_n)$$

für messbare Abbildungen b_n, τ_n , so dass $\int \phi_n^* dK_{T_n}(s, \cdot) = \alpha$ für alle $s \in \Omega_{S_n}$ gilt, sowie

$$\tilde{\phi}_n^* := \mathbb{1}_{(b_n(\alpha),\infty)}(T_n + R_n) + \tau_n \mathbb{1}_{\{b_n(\alpha)\}}(T_n + R_n).$$

Bedingung (4.15) impliziert nun $b_n(\alpha) \xrightarrow{P} c(\alpha)$ für alle $P \in \mathcal{P}_0$. Eine erneute Anwendung des Lemmas von Slutzky zeigt

$$E_P(|\phi_n^* - \tilde{\phi}_n^*|) \to 0.$$

Hieraus folgt unmittelbar $E_P(|\mathbb{1}_{(c(\alpha),\infty)}(T_n) - \tilde{\phi}_n^*|) \to 0$. Genauso weist man die Gültigkeit von $E_P(|\mathbb{1}_{(c(\alpha),\infty)}(T_n) - \psi_n|) \to 0$ nach. Dies impliziert zusammen die Konvergenz

$$E_P(|\psi_n - \tilde{\phi}_n^*|) \to 0.$$

Somit reicht es aus $E_P(|\psi_n^* - \tilde{\phi}_n^*|) \to 0$ zu beweisen.

Aufgrund von $T_n - |R_n| \le T_n + R_n \le T_n + |R_n|$ wird zunächst der Fall betrachtet, dass $R_n \ge 0$ P-fast sicher gilt. Damit erhält man $b_n(\alpha) \le c_n(\alpha)$ und somit $\tilde{\phi}_n^* \ge \psi_n^*$. Zusammen mit Gleichung (4.19) folgt hieraus

$$E_{P}(|\psi_{n}^{*} - \tilde{\phi}_{n}^{*}|) = E_{P}(\tilde{\phi}_{n}^{*}) - E_{P}(\psi_{n}^{*})$$

$$= E_{P}(\tilde{\phi}_{n}^{*}) - \int E_{P}(\psi_{n}^{*}|S_{n}) dP$$

$$= E_{P}(\tilde{\phi}_{n}^{*}) - \int K_{T_{n}+R_{n}}(\cdot, (c_{n}(\alpha), \infty))$$

$$+ \gamma_{n} \cdot K_{T_{n}+R_{n}}(\cdot, \{c_{n}(\alpha)\}) dP$$

$$= E_{P}(\tilde{\phi}_{n}^{*}) - \int \alpha dP \to 0.$$

Den Fall $R_n \leq 0$ erhält man analog.

Definiert man jetzt Tests $\tilde{\phi}_n^+$ und $\tilde{\phi}_n^-$, indem man R_n in $\tilde{\phi}_n^*$ durch $|R_n|$ bzw. $-|R_n|$ ersetzt, und wiederholt für diese Tests die obigen Beweisschritte, so kann man wieder auf die Konvergenz $E_P(|\psi_n - \tilde{\phi}_n^+|) + E_P(|\psi_n - \tilde{\phi}_n^-|) \to 0$ schließen. Wegen $\tilde{\phi}_n^- \leq \psi_n^* \leq \tilde{\phi}_n^+$ impliziert dies

$$E_P(|\psi_n - \psi_n^*|) \le E_P(|\psi_n - \tilde{\phi}_n^+|) + E_P(|\psi_n - \tilde{\phi}_n^-|) \to 0$$

und somit die Behauptung (4.17).

Bemerkung 4.6

Der obige Beweis lässt sich für Randomisationsverfahren stark vereinfachen, vgl. hierzu Lemma A.7 im Appendix. \Box

Die nächsten beiden Aussagen befassen sich mit notwendigen Bedingungen für die Voraussetzung (4.15). Hierbei sei wieder auf die Bemerkung nach Voraussetzung (A) verwiesen.

Lemma 4.7

Sei $\mathcal{L}(T) = \mu \in \mathcal{M}_1(\mathbb{R})$ ein Wahrscheinlichkeitsmaß, für dass die Konvergenz in (4.15) erfüllt ist. Falls die ersten Momente von T_n und T existieren und für $n \to \infty$ gegeneinander konvergieren $E(|T_n|) \to E(|T|) < \infty$, so konvergieren auch schon die bedingten Erwartungen von T_n unter S_n stochastisch gegen den Erwartungswert von T

$$E(T_n|S_n) \xrightarrow{P} E(T).$$
 (4.20)

Beweis. Zunächst sei bemerkt, dass aufgrund der Suffizienz eine von $P_n \in \mathcal{P}_{0,n}$ unabhängige Version des bedingten Erwartungswertes $E(T_n|S_n) = E_P(T_n|S_n) = E_{P_n}(T_n|S_n)$ existiert, vgl. WITTING [65, Satz 3.13].

Es wird nun im ersten Fall angenommen, dass $T_n \ge 0$ für alle $n \in \mathbb{N}$ gilt. Dann folgt aus der Markoffschen Ungleichung für jedes $\delta > 0$

$$P(E(T_n|S_n) \ge \delta) \le \delta^{-1} \sup_{n} E(T_n).$$

Da man nun zu vorgegebenem $\epsilon > 0$ den Wert δ so groß wählen kann, dass $\delta^{-1}\sup_n E(T_n) \leq \epsilon$ gilt, folgt hieraus die Straffheit der Folge $(P^{E(T_n|S_n)})_n$. Nach dem Satz von Prohorov existiert somit eine Teilfolge $(n_k)_k$ und eine Zufallsvariable T_0 , so dass Verteilungskonvergenz $E(T_{n_k}|S_{n_k}) \stackrel{\mathcal{D}}{\longrightarrow} T_0$ vorliegt. Zeigt man nun, dass $T_0 = E(T)$ gilt, so folgt (4.20), da der Grenzwert konstant ist und somit nicht von der speziellen Teilfolge n_k abhängt. Hierzu wird nun Voraussetzung (4.15) verwendet. Danach existiert nämlich eine weitere Teilfolge $(n_m)_m \subset (n_k)_k$, so dass für jede Konstante j > 0 gilt

$$E(T_{n_m}|S_{n_m}) = \int_{[0,\infty)} t \, dP^{T_{n_m}|S_{n_m}}(t) \ge \int_{[0,\infty)} \min(t,j) \, K_{T_{n_m}}(S_{n_m},dt) =: Y_m^j$$

$$\to \int_{[0,\infty)} \min(t,j) \, dP^T(t) =: C^j$$

P-fast sicher für $m \to \infty$. Dabei konvergiert der Grenzwert C^j nach dem Satz von der dominierten Konvergenz für $j \to \infty$ gegen E(T). Somit gibt es für alle $\delta > 0$ ein $j_0(\delta)$, so dass für alle $j \geq j_0(\delta)$ gilt

$$P(E(T_{n_m}|S_{n_m}) \ge E(T) - 2\delta) \ge P(Y_m^j \ge E(T) - 2\delta)$$

$$\ge P(\{C^j \ge E(T) - \delta\} \cap \{|Y_m^j - C^j| < \delta\})$$

$$\to 1 \text{ für } m \to \infty,$$

woraus $P(T_0 \geq E(T) - 2\delta) = 1$ folgt. Weil dies für alle $\delta > 0$ stimmt, erhält man

$$P(T_0 \ge E(T)) = 1. \tag{4.21}$$

Die andere Richtung erhält man mit Hilfe des Lemmas von Fatou sowie dem Satz von Skohorod

$$E(T) = \lim_{m \to \infty} E(T_{n_m}) = \lim_{m \to \infty} \int E(T_{n_m}|S_{n_m}) dP$$

$$\geq \int \liminf_{m \to \infty} E(T_{n_m}|S_{n_m}) dP = E(T_0).$$

In Kombination mit (4.21) impliziert dies nun $P^{T_0} = \varepsilon_{E(T)}$ und somit schon (4.20).

Für den allgemeinen Fall splittet man die Statistik T_n in ihren Positivteil $T_n^+ := \max(T_n, 0)$ und ihren Negativteil $T_n^- := \max(-T_n, 0)$ auf, d.h. man schreibt $T_n = T_n^+ - T_n^-$. Nach dem Satz über stetige Abbildungen folgt wiederum

$$d(\mathcal{L}(T^+), \mathcal{L}(T_n^+|S_n)) \xrightarrow{P} 0.$$

Da auch

$$E(|T_n|) = E(T_n^+) + E(T_n^-) \to E(T^+) + E(T^-) = E(|T|)$$

und $\liminf_{n\to\infty} E(T_n^+) \geq E(T^+)$ sowie $\liminf_{n\to\infty} E(T_n^-) \geq E(T^-)$ gelten, ist auch die zweite Bedingung $E(T_n^+) \to E(T^+)$ erfüllt, so das man den ersten Fall verwenden kann, um

$$E(T_n^+|S_n) \xrightarrow{P} E(T^+)$$

zu folgern. Da man dasselbe auch in analoger Weise für T_n^- zeigen kann, folgt hieraus die Behauptung.

Lemma 4.8

Es gelte $E(T_n) = E(T) = 0$ sowie $Var(T_n) \to Var(T) < \infty$ mit $Var(T_n) < \infty$ für alle $n \in \mathbb{N}$. Dann ist die folgende Bedingung notwendig für (4.15)

$$Var(T_n - E(T_n|S_n)) \to Var(T).$$
 (4.22)

Darüber hinaus ist die Bedingung (4.22) äquivalent zu $\int E(T_n|S_n)^2 dP_n \to 0$.

Beweis. Sei $\|\cdot\|_2^2$ die $L_2(P)$ -Norm. Aufgrund von

$$||T_n||_2^2 = ||T_n - E(T_n|S_n)||_2^2 + ||E(T_n|S_n)||_2^2$$

reicht es aus zu zeigen, dass $||E(T_n|S_n)||_2^2 \to 0$ gilt.

Da die bedingte Konvergenz (4.15) nach Korollar A.4 die unbedingte Konvergenz $T_n \stackrel{\mathcal{D}}{\longrightarrow} T$ nach sich zieht, gilt (4.14). Demnach existieren nach dem Satz von Skohorod auf einem geeigneten Wahrscheinlichkeitsraum $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P})$ Zufallsvariablen \hat{T}_n, \hat{T} mit $\hat{P}^{\hat{T}_n} = P^{T_n}, \hat{P}^{\hat{T}} = P^T$ und $\hat{T}_n \to \hat{T}$ \hat{P} -fast sicher. Hierfür sind wegen

$$\infty > \int \hat{T}_n^2 d\hat{P} = \int T_n^2 dP \to \int T^2 dP = \int \hat{T}^2 d\hat{P} < \infty \tag{4.23}$$

und $0 \le |\hat{T}_n| \le 1 + \hat{T}_n^2$ die Voraussetzungen des Lemmas von Pratt erfüllt, so dass man die Konvergenz

$$\int |T_n|dP = \int |\hat{T}_n|d\hat{P} \to \int |\hat{T}|d\hat{P} = \int |T|dP$$

erhält. Nach Lemma 4.7 folgt hieraus nun die Konvergenz der bedingten Erwartungswerte

$$E(T_n|S_n) \xrightarrow{P} E(T) = 0.$$

Zum Abschluss des Beweises ist nach dem Satz von Vitali, vgl. WITTING [65, Satz 1.181], nur noch die gleichgradige Integrierbarkeit der Folge $(E(T_n|S_n)^2)_n$ zu beweisen.

Hierzu wird zunächst die gleichgradige Integrierbarkeit der Folge $(T_n^2)_n$ untersucht.

Da $(\hat{T}_n^2)_n$ aufgrund der Konvergenz in (4.23) nach dem Satz von Vitali schon in $L_1(\hat{P})$ gegen \hat{T}^2 konvergiert, folgt hieraus die gleichgradige Integrierbarkeit von \hat{T}_n^2 . Weil diese Eigenschaft nur von der Verteilung von \hat{T}_n^2 abhängt, folgt die

gleichgradige Integrierbarkeit der Folge $(T_n^2)_n$. Außerdem erhält man aus der Markoffschen Ungleichung, dass

$$\mathbb{1}_{\left\{E(T_n^2|S_n)\geq a_n\right\}} \stackrel{P}{\longrightarrow} 0$$

für jede Folge a_n gilt, die für $n \to \infty$ gegen ∞ konvergiert.

Die gleichgradige Integrierbarkeit von $(E(T_n|S_n)^2)_n$ ergibt sich somit mit Hilfe der Jensenschen Ungleichung

$$\int \mathbb{1}_{\{E(T_n|S_n)^2 \ge a_n\}} E(T_n|S_n)^2 dP \le \int \mathbb{1}_{\{E(T_n^2|S_n) \ge a_n\}} E(T_n^2|S_n) dP$$
$$= \int \mathbb{1}_{\{E(T_n^2|S_n) \ge a_n\}} T_n^2 dP \to 0.$$

4.2 Resampling Quadratischer Formen

Nachdem nun im vorigen Abschnitt das Verhalten von asymptotisch linearen Statistiken untersucht wurde, rückt an dieser Stelle das asymptotische Verhalten einer weiteren Form von nicht-linearen Statistiken - den quadratischen Formen - in den Fokus. Diese treten häufig bei Testproblemen in mehrdimensionalen Datensätzen oder auch bei Anpassungstests als Teststatistiken auf, vgl. u.a. Kapitel 5.4 und 7.5.3. - 7.5.4. bei WITTING UND MÜLLER-FUNK [66], Kapitel 5 bei SHORACK UND WELLNER [58] sowie den Abschnitt 3.5. bei SERFLING [57]. Als erste einfache Anwendung des Satzes über stetige Abbildungen wird an dieser Stelle eine Aussage über das Verhalten von endlichdimensionalen quadratischen Formen zitiert.

Satz 4.9

Konvergiert die Folge der k-dimensionalen Zufallsvariablen $Z_n: (\Omega, \mathcal{A}, P) \to (\mathbb{R}^k, \mathcal{B}^k)$ in Verteilung gegen eine $N(\mu, \Gamma)$ -verteilte Zufallsvariable mit Erwartungswert $\mu \in \mathbb{R}^k$ und positiv definiter Kovarianzmatrix $\Gamma \in \mathbb{R}^{k \times k}_{p.d.}$ und existiert eine Folge von Zufallsmatrizen Γ_n mit $\Gamma_n \xrightarrow{P} \Gamma$, so liegt schwache Konvergenz

$$\mathcal{L}(Z_n^t \Gamma_n^{-1} Z_n) \xrightarrow{w} \chi_k^2(\delta^2) \tag{4.24}$$

gegen eine nichtzentrale χ^2 -Verteilung mit k Freiheitsgraden und Nichtzentralitätsparameter $\delta^2=\mu^t\Gamma^{-1}\mu$ vor.

Beweis. Da aus den Voraussetzungen $\Gamma_n^{-1} \xrightarrow{P} \Gamma^{-1}$ folgt, erhält man die Behauptung mit Hilfe des Lemmas von Slutzky aus Satz 5.125. von WITTING UND MÜLLER-FUNK [66].

Dieser Hilfssatz findet u.a. in den noch folgenen Erweiterungen sowie in dem späteren Abschnitt über Anwendungen in speziellen Testproblemen Verwendung.

Als nächstes werden für eine Familie von reellen Folgen $(\lambda_{n,j})_{j,n}$ die komplizierteren unendlichdimensionalen quadratischen Statistiken der Form

$$\tilde{T}_n = \sum_{j=1}^{\infty} \lambda_{n,j} Z_{n,j}^2 \tag{4.25}$$

untersucht, die durch Dreieckschema von Zufallsvariablen $(Z_{n,j})_{j,n}$ gegeben sind. Hierbei werden Bedingungen angegeben, unter denen man, ähnlich wie im endlichdimensionalen Fall, von dem asymptotischen Verhalten der Folge $(Z_{n,j})_{j,n}$ auf das von \tilde{T}_n schließen kann. Für $j \in \mathbb{N}$ werden im Folgenden die kanonischen Projektionen mit $\pi_j : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}) \to (\mathbb{R}, \mathcal{B}), j \in \mathbb{N}$, bezeichnet. Da der Raum $\mathbb{R}^{\mathbb{N}}$ (wieder versehen mit der Produkttopologie) polnisch ist, ist auch der Raum der Maße $\mathcal{M}_1(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ polnisch bzgl. der Metrik der schwachen Konvergenz, vgl. BILLINGSLEY [6, Appendix III, S.236ff.]. Das nächste Lemma gibt nun einen geeigneten Teilraum an, auf dem sich die Konvergenz von (4.25) untersuchen lässt.

Lemma 4.10

Für eine feste Konstante $K \geq 0$ und $r \in \mathbb{N}$ ist die Menge

$$\mathcal{M}^r := \{ P \in \mathcal{M}_1(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}) : \int |\pi_j|^r dP \le K \text{ f\"{u}r alle } j \in \mathbb{N} \}$$
 (4.26)

ebenfalls polnisch bzgl. der schwachen Konvergenz.

Beweis. Da \mathcal{M}^r ein Unterraum des polnischen Raumes $\mathcal{M}_1(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ ist, reicht es aus zu zeigen, dass \mathcal{M}^r bzgl. der Metrik der schwachen Konvergenz abgeschlossen ist, vgl. BAUER [3, S.179].

Sei also $(P_n)_{n\in\mathbb{N}}\subset\mathcal{M}^r$ eine Folge von Wahrscheinlichkeitsmaßen mit

$$P_n \xrightarrow{w} P_0 \in \mathcal{M}_1(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}).$$

⁴Der Beweis kann an dieser Stelle nicht mit dem Stetigkeitssatz erfolgen, da unendlichdimensionale quadratischen Formen $\mathbb{R}^{\mathbb{N}} \ni (x_i)_i \mapsto \sum_{i=1}^{\infty} \lambda_i x_i^2$ i.d.R. nicht stetig sind.

Aufgrund der Stetigkeit der Projektionen folgt hieraus

$$P_n^{\pi_j} \xrightarrow{w} P_0^{\pi_j} \text{ für jedes } j \in \mathbb{N}.$$
 (4.27)

Nach dem Satz von Skohorod existieren somit auf einem geeigneten Wahrscheinlichkeitsraum $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P})$ Zufallsvariablen $\hat{X}_n, n \in \mathbb{N}_0$, mit $\mathcal{L}(\hat{X}_n) = P_n^{\pi_j}$ und $\mathcal{L}(\hat{X}_0) = P_0^{\pi_j}$ sowie $\hat{X}_n \to \hat{X}_0$ \hat{P} fast sicher. Die Behauptung erhält man demnach aus dem Lemma von Fatou

$$K \ge \liminf_{n \to \infty} \int |\pi_j|^r dP_n \ge \int \liminf_{n \to \infty} |\hat{X}_n|^r d\hat{P}$$
$$= \int |\pi_j|^r dP_0$$

Auf diesem Teilraum lässt sich nun auf die Konvergenz der quadratischen Statistik schließen. Aus Gründen der Übersichtlichkeit schreibt man dies mittels geeigenten Bildmaßen auf. Hierzu definiert man nun für $n \in \mathbb{N}_0$ die Abbildung $f_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ mit

$$(x_j)_{j \in \mathbb{N}} =: x \mapsto f_n(x) := \begin{cases} \sum_{j=1}^{\infty} \lambda_{n,j} x_j^2 &, \text{ falls } \sum_{j=1}^{\infty} |\lambda_{n,j}| x_j^2 < \infty, \\ 0 & \text{sonst,} \end{cases}$$

wobei $(\lambda_{n,j})_{j\in\mathbb{N}}$ für jedes n eine reelle Folge ist.

Satz 4.11

Konvergiert die Folge von Wahrscheinlichkeitsmaßen $(P_n)_{n\in\mathbb{N}}\subset\mathcal{M}^2$ schwach gegen P_0 , so konvergieren auch die zugehörigen Bildmaße $P_n^{f_n}$ schwach gegen $P_0^{f_0} = \mathcal{L}\left(\sum_{j=1}^{\infty} \lambda_{0,j} (X_0^{(j)})^2\right)$, falls die reelle Folge $(\lambda_{0,j})_{j\in\mathbb{N}}$ absolut summierbar ist und zusätzlich eine der beiden Bedingungen erfüllt ist:

- a) Es gilt $\lambda_{n,j} = \lambda_{0,j}$ für alle $n \in \mathbb{N}$.
- b) Für alle $j \in \mathbb{N}$ gilt $\lim_{n \to \infty} \lambda_{n,j} = \lambda_{0,j}$ und es liegt die folgende Konvergenz vor

$$\lim_{k \to \infty} \limsup_{n \to \infty} \sum_{j=k}^{\infty} |\lambda_{n,j}| = 0.$$
 (4.28)

Dabei ist $X_0 := (X_0^{(j)})_{j \in \mathbb{N}}$ eine $\mathbb{R}^{\mathbb{N}}$ -wertige Zufallsvariable mit Verteilung $\mathcal{L}(X_0) = P_0$.

Beweis. Es wird zunächst Teil b) bewiesen.

Für $n \in \mathbb{N}_0$ seien $X_n := (X_n^{(j)})_{j \in \mathbb{N}} \mathbb{R}^{\mathbb{N}}$ -wertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Verteilung $\mathcal{L}(X_n) = P_n$. Der Beweis erfolgt nun mit Hilfe der Abschneidetechnik von BILLINGSLEY (vgl. [6, Theorem 4.2.]). Dazu splittet man die Statistik $f_n(X_n)$ in zwei Teile auf

$$\sum_{j=1}^{k} \lambda_{n,j} (X_n^{(j)})^2 + \sum_{j=k+1}^{\infty} \lambda_{n,j} (X_n^{(j)})^2 =: T_n^k + Z_n^k.$$

Weil T_n^k eine endlichdimensionale quadratische Form ist, liegt hierfür analog zu Satz 4.9 nach dem Stetigkeitssatz für $n\to\infty$ Verteilungskonvergenz

$$T_n^k = (X_n^{(j)})_{j \le k}^t Diag(\lambda_{n,j} : j \le k) (X_n^{(j)})_{j \le k} \xrightarrow{\mathcal{D}} \sum_{i=1}^k \lambda_{0,j} (X_0^{(j)})^2 =: T_0^k \quad (4.29)$$

vor. Aufgrund der Summierbarkeit der Folge $(\lambda_{0,j})_{j\in\mathbb{N}}$ existiert für jedes $\epsilon>0$ ein $m_0\in\mathbb{N}$, so dass für alle $n,m\geq m_0$

$$\mathbb{E}\left(\left|\sum_{j=m}^{n} \lambda_{0,j} (X_0^{(j)})^2\right|\right) \le \sum_{j=m}^{n} |\lambda_{0,j}| \,\mathbb{E}((X_0^{(j)})^2) \le K \sum_{j=m}^{n} |\lambda_{0,j}| \le \epsilon \tag{4.30}$$

gilt. Somit ist die Grenzvariable T_0^k eine Cauchy-Folge in L_1 mit zugehörigem, fast sicher endlichem L_1 -Limes $\sum_{j=1}^{\infty} \lambda_{0,j}(X_0^{(j)})^2$. Genauso gilt nach dem Satz von Beppo-Levi⁵

$$E(|Z_n^k|) \le \sum_{j=k+1}^{\infty} |\lambda_{n,j}| E((X_n^{(j)})^2) \le K \sum_{j=k+1}^{\infty} |\lambda_{n,j}|,$$
 (4.31)

woraus mit Hilfe der Markoffschen Ungleichung auch

$$\lim_{k \to \infty} \limsup_{n \to \infty} P(|Z_n^k| \ge \epsilon) \le \frac{K}{\epsilon} \lim_{k \to \infty} \limsup_{n \to \infty} \sum_{j=k+1}^{\infty} |\lambda_{n,j}| = 0$$

folgt. Die Behauptung b) ergibt sich nun nach Theorem 4.2. von Billingsley. Teil a) folgt aufgrund von $\sum_{j=k+1}^{\infty} \lambda_{0,j} \to 0$ für $k \to \infty$ aus b).

Um nun Resampling-Versionen von diesen Quadratischen Formen zu behandeln, benötigt man nur Abschätzungen für die 2. Momente der einzelnen Einträge der hochdimensionalen Resampling-Statistik

$$T_n^* = (T_n^{*(j)})_{j \in \mathbb{N}} := (T_n^*(X_{n,i}^{(j)}))_{j \in \mathbb{N}} = \sqrt{k(n)} \sum_{i=1}^{k(n)} W_{n,i}(X_{n,i} - \overline{X}_n).$$

⁵Hieraus folgt insbesondere zusammen mit $E(T_n^k) \leq K \sum_{j=1}^k |\lambda_{n,j}|$ und (4.28), dass $T_n := f_n((X_n^{(j)})_{j \in \mathbb{N}})$ schließlich fast sicher existiert.

Diese lassen sich aber auf Varianzen von Rangstatistiken zurückführen. Für den folgenden Satz gelte $\lambda_{n,j} = \lambda_j$ und somit $f_n = f_1$ für alle $j, n \in \mathbb{N}$.

Satz 4.12

Seien $X_{n,i} = (X_{n,i}^{(j)})_{j \in \mathbb{N}}$ ein Dreiecksschema von $\mathbb{R}^{\mathbb{N}}$ -wertigen Zufallsvariablen, das den Bedingungen aus Satz 3.23 mit konstantem $\Gamma = (\gamma_{j,r})_{j,r \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ genüge und $Z := (Z_j)_{j \in \mathbb{N}}$ eine $N(0,\Gamma)$ -verteilte, $\mathbb{R}^{\mathbb{N}}$ -wertige Zufallsvariable. Weiter existiere eine Menge M mit P(M) = 1 sowie eine Abbildung $K : \Omega \to \mathbb{N}$, so dass für alle $\omega \in M$

$$\limsup_{n \to \infty} \sup_{j \in \mathbb{N}} \sum_{i=1}^{k(n)} \left(X_{n,i}^{(j)}(\omega) \right)^2 \le K(\omega)$$
(4.32)

gelte⁶. Erfüllen die Gewichte $(W_{n,i})_{i \leq k(n)}$ zuätzlich zu den Generalvoraussetzungen mit $Var(W_1) = 1$ noch die Bedingung

$$a := \limsup_{n \to \infty} k(n) E_{\tilde{P}} \left((W_{n,1} - \overline{W}_n)^2 \right) < \infty, \tag{4.33}$$

so $folgt^7$

$$d_1\left(\mathcal{L}(f_1((T_n^*(X_{n,i}^{(j)}))_{j\in\mathbb{N}})|(X_{n,i})_{i\leq k(n)}), \mathcal{L}(\sum_{i=1}^{\infty}\lambda_j Z_j^2)\right) \stackrel{P}{\longrightarrow} 0, \tag{4.34}$$

falls die Folge $(\lambda_j)_{j\in\mathbb{N}}$ absolut summierbar ist.

Beweis. Aufgrund von (4.33) existiert für jedes $\epsilon > 0$ ein $n_0 \in \mathbb{N}$, so dass

$$E_{\tilde{P}}\left(\sum_{i=1}^{k(n)} (W_{n,i} - \overline{W}_n)^2\right) \le a + \epsilon$$

für alle $n \ge n_0$ gilt. Weiter liegt nach Satz 3.23 bedingte Konvergenz

$$d_{\mathbb{N}}(\mathcal{L}(T_n^*|(X_{n,i})_{i < k(n)}), \mathcal{N}(0,\Gamma)) \xrightarrow{P} 0 \tag{4.35}$$

vor. Durch Anwendung des Teilfolgenkriteriums für stochastische Konvergenz findet man für jede Teilfolge eine weitere Teilfolge $\{m\}$ und eine Menge M_1

 $^{^6}$ Zusammen mit den Voraussetzungen von Satz 3.23 impliziert dies insbesondere die Beschränkheit der Diagonaleinträge sup $_i |\gamma_{j,j}| < \infty$.

⁷Alle auftretenden Zufallsvariablen werden wieder mittels Projektionen auf dem gemeinsamen Wahrscheinlichkeitsraum $(\Omega \times \tilde{\Omega}, \mathcal{A} \otimes \tilde{\mathcal{A}}, P \otimes \tilde{P})$ definiert.

mit $P(M_1) = 1$, so dass entlang dieser Teilfolge für alle $\omega \in M_1$ punktweise Konvergenz in (4.35) vorliegt und

$$\sup_{j\in\mathbb{N}}\sum_{i=1}^{k(m)}\left(X_{m,i}^{(j)}(\omega)\right)^2\leq K(\omega) \text{ für alle } m\geq m_0(\omega)\in\mathbb{N}$$

gelte. Für festes $\omega \in M_1$ erhält man so, aufgrund von $\mathrm{E}(T_n^{*(j)}|(X_{n,i})_{i \leq k(n)}) = 0$, wie im Beweis von Zusatz 3.5 die Abschätzung

$$\mathbb{E}\left((T_m^{*\,(j)})^2|(X_{m,i})_{i\leq k(m)}\right)(\omega)$$

$$= \frac{k(m)}{k(m)-1} \sum_{i=1}^{k(m)} (X_{m,i}^{(j)}(\omega))^2 \operatorname{E}\left(\sum_{i=1}^{k(m)} (W_{m,i} - \overline{W}_m)^2\right)$$
(4.36)

$$\leq K(\omega)(a+\epsilon) =: \tilde{K}(\omega) < \infty$$
 (4.37)

für alle $j \in \mathbb{N}$ und $m \ge \max(m_0(\omega), n_0)$. Durch Anwendung von Satz 4.11 erhält man hieraus für $m \to \infty^8$

$$\mathcal{L}(\sum_{j=1}^{\infty} \lambda_j(T_m^*(X_{m,i}^{(j)}))^2 | (X_{m,i})_{i \leq k(m)})(\omega) \xrightarrow{w} \mathcal{L}(\sum_{j=1}^{\infty} \lambda_j Z_j^2).$$

Zusammenfassend hat man also für jede Teilfolge eine weitere Teilfolge gefunden, so dass die Konvergenz in (4.34) entlang dieser Teilfolge fast sicher gilt. Dies zeigt das Gewünschte.

Bemerkung 4.13

Die Bedingung (4.33) ist insbesondere für die Gewichte aus Bemerkung 3.6 erfüllt. Weitere Gewichtsfunktionen findet man in Kapitel 6.

Als einfache Anwendung des obigen Satzes erhält man das folgende Korollar.

Korollar 4.14

Gegeben seien eine absolut summierbare Folge reeller Zahlen $(\lambda_j)_{j\in\mathbb{N}}$ sowie eine Funktionenfolge $f_j:[0,1]\to\mathbb{R},\ j\in\mathbb{N},\ mit\ den\ folgenden\ Eigenschaften$

$$\sup_{x} |f_j(x)| \le K \in \mathbb{R}, \ E(f_j(U_1)) = 0, \tag{4.38}$$

⁸Die fast sichere Existenz der linken Seite ab $m \ge \max(m_0(\omega), n_0)$ folgt wie im Beweis von Satz 4.11 mit Hilfe von (4.37). Wegen $\sup_j |\gamma_{j,j}| < \infty$ erhält man so auch die fast sichere Existenz der rechten Seite.

$$\int_0^1 f_j(x) f_r(x) \ d\lambda(x) = \delta_{jr},\tag{4.39}$$

für alle $j, r \in \mathbb{N}$. Definiert man für unabhängige auf [0, 1] gleichverteilte Zufallsvariable $U_1, \ldots, U_{k(n)}$ das Dreiecksschema

$$X_{n,i} = (X_{n,i}^{(j)})_{j \in \mathbb{N}} := \left(\frac{f_j(U_i)}{\sqrt{k(n)}}\right)_{j \in \mathbb{N}}, \ i \le k(n), \tag{4.40}$$

von $\mathbb{R}^{\mathbb{N}}$ -wertigen Zufallsvariablen, so folgt der bedingte Grenzwertsatz (4.34), sofern die Gewichte den Generalvoraussetzungen mit $Var(W_1) = 1$ sowie der Bedingung (4.33) genügen.

Beweis. Wegen $\max_{i \leq k(n)} |X_{n,i}^{(j)}| \leq K/k(n)^{1/2} \to 0$ und

$$\sum_{i=1}^{k(n)} (X_{n,i}^{(j)} - \overline{X}_{k(n)}^{(j)})(X_{n,i}^{(r)} - \overline{X}_{k(n)}^{(r)}) = \frac{1}{k(n)} \sum_{i=1}^{k(n)} f_j(U_i) f_r(U_i)
- \left(\frac{1}{k(n)} \sum_{i=1}^{k(n)} f_j(U_i)\right) \left(\frac{1}{k(n)} \sum_{i=1}^{k(n)} f_r(U_i)\right)
\xrightarrow{f.s.} \quad \operatorname{E} \left(f_j(U_1) f_r(U_1)\right) - \operatorname{E} \left(f_j(U_1)\right) \operatorname{E} \left(f_r(U_1)\right)
= \delta_{jr},$$

sowie $\sum_{i=1}^{k(n)} (X_{n,i}^{(j)} - \overline{X}_{k(n)}^{(j)})^2 \le K^2 < \infty$ für alle $j \in \mathbb{N}$ folgt dies aus dem vorigen Satz 4.12.

Beispiel 4.15 (Resampling der Cramér von Mises Statistik)

Gegeben sei eine i.i.d. Folge von reellen Zufallsvariablen X_1, \ldots, X_n mit stetiger Verteilungsfunktion F und empirischer Verteilungsfunktion F_n . Aufgrund der Quantiltransformation gilt $F^{-1}(U_i) \stackrel{\mathcal{D}}{=} X_i$ für alle $i \leq n$, wobei $(U_i)_i$ unabhängige auf [0,1] gleichverteilte Zufallsvariable sind. Mit Hilfe von Anpassungstests kann man nun überprüfen, ob die Zufallsvariablen eine spezielle Verteilungsfunktion F_0 besitzen, d.h. man testet die Hypothese $H: \{F=F_0\}$. Eine geläufige Teststatistik ist die Cramér von Mises Statistik

$$W_n^2 := n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 dF(x). \tag{4.41}$$

Hierfür ist der folgende Grenzwertsatz, vgl. Shorack und Wellner [58, S.145,

(11) und S.215, Theorem 1.], bekannt⁹

$$W_n^2 \xrightarrow{\mathcal{D}} W^2 := \sum_{j=1}^{\infty} \frac{Z_j^2}{(j\pi)^2}, \tag{4.42}$$

wobei Z_1, Z_2, \ldots eine Folge von unabhängigen und standardnormalverteilten Zufallsvariablen ist. Da die rechte Seite nicht von F_0 abhängt, kann der zugehörige Test verteilungsfrei ausgeführt werden. Weil die Verteilung von W^2 zusätzlich auch gut vertafelt ist, ist es aus praktischer Sicht eigentlich nicht nötig W_n^2 in diesem Einstichprobenfall zu resamplen. Aus mathematischer Sicht ist es aber dennoch interessant zu wissen, ob die Resampling-Verfahren auch in diesem Fall konsistent sind, d.h. ob sich auch hier ein bedingter Zentraler Grenzwertsatz beweisen lässt. Für den m(n)- und den Bayesian-Bootstrap wurde dies beispielsweise in den Arbeiten von Lo und Sazonov, vgl. [43] und [44], untersucht. Da sich dies auch hier als erste Anwendung von Satz 4.12 anbietet, wird die Vorgehensweise im Folgenden kurz skizziert.

Nach dem Satz von Durbin und Knott, vgl. [58, S.215f, Theorem 2.], liegt für W_n^2 eine Spektralzerlegung der Form

$$W_n^2 = \sum_{j=1}^{\infty} \lambda_j Z_{n,j}^2$$

mit Eigenwerten $\lambda_j = (j\pi)^{-2}$ und, für festes $n \in \mathbb{N}$, unkorellierten und identisch verteilten Zufallsvariablen $Z_{n,j} = \sqrt{\frac{2}{n}} \cdot \sum_{i=1}^n \cos(j\pi U_i), \ j \geq 1$ mit Erwartungswert 0 und Varianz 1 vor. Die Eigenwerte sind offensichtlich positiv und summierbar (mit $\sum_{j=1}^{\infty} (j\pi)^{-2} = 1/6$). Als Resampling-Version von W_n^2 kommt nun

$$W_n^{*2} := \sum_{i=1}^{\infty} \frac{1}{(j\pi)^2} (T_n^*(X_{n,i}^{(j)}))^2$$

in Frage, wobei $X_{n,i}^{(j)} := \sqrt{2}n^{-1/2}\cos(j\pi U_i), \ j \in \mathbb{N}$ gelte. Hierfür erhält man

$$d_1\left(\mathcal{L}(\sum_{j=1}^{\infty} \frac{1}{(j\pi)^2} (T_n^*(X_{n,i}^{(j)}))^2 | (X_{n,i})_{i \le n}), \mathcal{L}(\sum_{j=1}^{\infty} \frac{1}{(j\pi)^2} Z_j^2)\right) \stackrel{P}{\longrightarrow} 0, \tag{4.43}$$

falls die Gewichte den Voraussetzungen von Satz 4.12 genügen.

Dies folgt aus Korollar 4.14, weil aus Symmetriegründen

$$\int_0^1 \cos(j\pi x) \ dx = 0 \text{ und } \int_0^1 \cos(j\pi x)^2 \ dx = 1 - \int_0^1 \sin(j\pi x)^2 \ dx = \frac{1}{2},$$

⁹Dabei existiert die rechte Seite nach WITTING UND MÜLLER-FUNK [66, Hilfssatz 5.145].

sowie

$$\int_0^1 \cos(j\pi x)\cos(r\pi x) \ dx = 0$$

für $j \neq r$ nach [58, S.217 (g)] gilt.

Als weitere Anwendung von Satz 4.11 erhält man einen möglichen Anpassungstest für diskrete Verteilungen.

Beispiel 4.16 (goodness-of-fit Test für diskrete Verteilungen)

Gegeben seien zwei voneinander unabhängige i.i.d. Stichproben X_1, \ldots, X_{n_1} und X_{n_1+1}, \ldots, X_n mit Werten in den natürlichen Zahlen $\mathbb N$ und jeweils identischen Verteilungsfunktionen F bzw. G. Möchte man nun die Hypothese $H_0: \{F=G\}$, dass die beiden Gruppen dieselbe Verteilung besitzen, testen, so kommt als Teststatistik

$$_{d}W_{n}^{2} := \frac{n_{1}n_{2}}{n} \sum_{k=1}^{\infty} (F_{n_{1}}(k) - G_{n_{2}}(k))^{2} p_{n}(k)$$
(4.44)

in Frage¹⁰, wobei der linke Index kennzeichnen soll, dass es sich um eine Teststatistik für diskrete Verteilungen handelt.

Dabei seien $F_{n_1}(k) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbbm{1}_{(-\infty,k]}(X_i)$ und $G_{n_2}(k) = \frac{1}{n_2} \sum_{i=n_1+1}^{n} \mathbbm{1}_{(-\infty,k]}(X_i)$ die empirischen Verteilungsfunktionen der einzelnen Gruppen und $p_n(k)$ geeignete Gewichtsfunktionen, die für die folgenden Untersuchungen als $p_n(k) := \frac{1}{n} \sum_{i=1}^{n} \mathbbm{1}_{\{k\}}(X_i)$ spezifiziert seien. Es wird nun zunächst gezeigt, dass diese Teststatistik unter H_0 in Verteilung gegen eine gewichtete, unendliche Summe $\sum_{j=1}^{\infty} p_j Z_j^2 =: dW^2$ von χ^2 -verteilten Zufallsvariablen Z_j^2 konvergiert. Da die Gewichte $p_j, j \in \mathbb{N}$, dabei gerade die Zähldichten der unbekannten Verteilungen F = G unter H_0 sind, ist die Grenzverteilung nicht verteilungsfrei. Dieses Problem wird mit Hilfe der obigen Theorie gelöst. Dabei wird sich zeigen, dass die bedingte Permutationsverteilung von dW_n^2 ebenfalls schwach gegen dW^2 konvergiert, so dass sich der dW_n^2 -Test als verteilungsfreier Permutationstest ausführen lässt.

Für die folgenden Aussagen gelte stets H_0 .

Weil Verteilungskonvergenz auf $\mathbb{R}^{\mathbb{N}}$ äquivalent zur Konvergenz der endlichdi-

¹⁰Diese lässt sich z.B. durch die Einstichproben Cramér von Mises Statistik für diskrete Verteilungen, vgl. z.B. Henze [27], motivieren.

¹¹Die folgenden Rechnungen funktionieren i.a. auch für nicht zufällige Gewichte $p_n(k)$, die den Bedingungen (4.28) sowie $p_n(j) \to p_j \in \mathbb{R}$ für $n \to \infty$ genügen.

mensionalen Verteilungen ist, erhält man mit Hilfe des Cramér-Wold-Devices aus dem klassischen Zentralen Grenzwertsatz die Verteilungskonvergenz

$$\left(\sqrt{\frac{n_1 n_2}{n}} (F_{n_1}(j) - G_{n_2}(j))\right)_{j \in \mathbb{N}} \xrightarrow{\mathcal{D}} Z = (Z^{(j)})_{j \in \mathbb{N}} \sim N(0, \Gamma) \tag{4.45}$$

mit $\Gamma = (F(j \wedge r) - F(j)F(r))_{j,r \in \mathbb{N}}$. Außerdem liegt für $n \to \infty$ nach dem Gesetz der grossen Zahlen die fast sichere Konvergenz $p_n(k) \to F(k) - F(k-1) =: p_k$ für $k \in \mathbb{N}$ vor, so dass man für $k \in \mathbb{N}_{\geq 3}$ analog zu (4.29) mit Hilfe des Lemmas von Slutzky die Konvergenz

$${}_{d}W_{n}^{k} := \frac{n_{1}n_{2}}{n} \sum_{j=1}^{k} (F_{n_{1}}(j) - G_{n_{2}}(j))^{2} p_{n}(j) \xrightarrow{\mathcal{D}} \sum_{j=1}^{k} p_{j}(Z^{(j)})^{2} =: {}_{d}W^{k} \quad (4.46)$$

erhält. Aufgrund von $\operatorname{Var}(Z^{(j)}) = F(j) - F(j)^2 \leq 1$ konvergiert dW^k - analog zu (4.30) - in L_1 gegen $\sum_{j=1}^{\infty} p_j(Z^{(j)})^2$ für $k \to \infty$.

Schreibt man nun die mit $p_n(j)$ erweiterte linke Seite von (4.45) mit Hilfe der Zweistichprobenregressionskoeffizienten $c_{n,i}$ aus Beispiel 2.6 als

$$\sqrt{p_n(j)}\sqrt{\frac{n_1n_2}{n}}(F_{n_1}(j) - G_{n_2}(j)) = \sqrt{n}\sum_{i=1}^n c_{n,i}X_{n,i}^{(j)}$$
(4.47)

für $X_{n,i}^{(j)}:=n^{-1/2}p_n(j)^{1/2}\mathbbm{1}_{(-\infty,j]}(X_i)=:p_n(j)^{1/2}Y_{n,i}^{(j)},$ so folgt aufgrund von $p_n\leq 1$

$$E\left(\left(\sqrt{n}\sum_{i=1}^{n}c_{n,i}X_{n,i}^{(j)}\right)^{2}\right) \leq \operatorname{Var}(\sqrt{n}\sum_{i=1}^{n}c_{n,i}Y_{n,i}^{(j)})$$

$$= \sum_{i=1}^{n}c_{n,i}^{2}\operatorname{Var}(\mathbb{1}_{(-\infty,j]}(X_{1}))$$

$$= p_{j} - p_{j}^{2} \leq p_{j}.$$

Hieraus erhält man mit Hilfe des Lemmas von Fatou

$$\mathbb{E}\left(\frac{n_1 n_2}{n} \sum_{j=k+1}^{\infty} (F_{n_1}(j) - G_{n_2}(j))^2 p_n(j)\right) \le \sum_{j=k+1}^{\infty} p_j \to 0$$

für $k \to \infty$ und so nach Theorem 4.2. von Billingsley

$$_{d}W_{n}^{2} \xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} p_{j}(Z^{(j)})^{2}.$$

Somit ist die Grenzverteilung von ${}_{d}W_{n}^{2}$ unter H_{0} nicht verteilungsfrei.

Um dies zu umgehen, kann ein Permutationsverfahren verwendet werden. Sei dazu wieder τ eine auf der symmetrischen Gruppe \mathcal{S}_n gleichverteilte Permutation. Gesucht ist nun die Grenzverteilung von¹² (vgl. Gleichung (4.47))

$$\sum_{j=1}^{\infty} (\sqrt{n} \sum_{i=1}^{n} c_{n,i} X_{n,\tau(i)}^{(j)})^{2} = \sum_{j=1}^{\infty} p_{n}(j) (\sqrt{n} \sum_{i=1}^{n} c_{n,i} Y_{n,\tau(i)}^{(j)})^{2}$$

$$\stackrel{\mathcal{D}}{=} \sum_{j=1}^{\infty} p_{n}(j) (\sqrt{n} \sum_{i=1}^{n} c_{n,\tau(i)} Y_{n,i}^{(j)})^{2}.$$

Die Konvergenzen $\max_{i \leq n} \left| Y_{n,i}^{(j)} - \overline{Y}_n^{(j)} \right| \leq 1/\sqrt{n} \to 0$ sowie

$$\sum_{i=1}^{n} (Y_{n,i}^{(j)} - \overline{Y}_{n}^{(j)}) (Y_{n,i}^{(r)} - \overline{Y}_{n}^{(r)}) \xrightarrow{f.s.} F(j \wedge r) - F(j)F(r) =: \gamma_{j,r}$$

für alle $j,r\in\mathbb{N}$ implizieren nach Satz 3.23 die bedingte Konvergenz

$$d_{\mathbb{N}}(\mathcal{L}(\sqrt{n}\sum_{i=1}^{n}c_{n,\tau(i)}Y_{n,i}|(X_{i})_{i\leq n}), N(0,\Gamma)) \xrightarrow{P} 0$$

$$(4.48)$$

für $Y_{n,i} := (Y_{n,i}^{(j)})_{j \in \mathbb{N}}$. In Analogie zum unbedingten Fall (vgl. (4.46)) erhält man

$$d_1(\mathcal{L}(\sum_{i=1}^k p_n(j)(\sqrt{n}\sum_{i=1}^n c_{n,\tau(i)}Y_{n,i}^{(j)})^2|(X_i)_{i\leq n}), \mathcal{L}(\sum_{i=1}^k p_j(Z^{(j)})^2)) \xrightarrow{P} 0. \quad (4.49)$$

Nach der Varianzformel von HAJEK UND ŠIDAK gilt außerdem

$$E((\sqrt{n}\sum_{i=1}^{n}c_{n,\tau(i)}X_{n,i}^{(j)})^{2}|(X_{i})_{i\leq n}) = Var(\sqrt{n}\sum_{i=1}^{n}c_{n,\tau(i)}X_{n,i}^{(j)}|(X_{i})_{i\leq n})$$

$$= \frac{n}{n-1}\sum_{r=1}^{n}c_{n,r}^{2}\sum_{i=1}^{n}(X_{n,i}^{(j)}-\overline{X}_{n}^{(j)})^{2}$$

$$\leq 2p_{n}(j),$$

so dass das Lemma von Fatou die folgende Ungleichung impliziert

$$E(\sum_{j=k+1}^{\infty} (\sqrt{n} \sum_{i=1}^{n} c_{n,\tau(i)} X_{n,i}^{(j)})^{2} | (X_{i})_{i \le n}) \le 2 \sum_{j=k+1}^{\infty} p_{n}(j) = \frac{2}{n} \sum_{i=1}^{n} \mathbb{1}_{[k+1,\infty)}(X_{i}).$$

¹² Dies reicht wegen $\sum_{i=1}^{n} \mathbb{1}_{(-\infty,k]}(X_{\tau(i)}) = \sum_{i=1}^{n} \mathbb{1}_{(-\infty,k]}(X_i) = np_n(k)$.

Hieraus folgt nach dem Gesetz der grossen Zahlen

$$\limsup_{n \to \infty} \mathbb{E}\left(\sum_{j=k+1}^{\infty} (\sqrt{n} \sum_{i=1}^{n} c_{n,\tau(i)} X_{n,i}^{(j)})^{2} | (X_{i})_{i \le n}\right) \le \limsup_{n \to \infty} \frac{2}{n} \sum_{i=1}^{n} \mathbb{1}_{[k+1,\infty)}(X_{i})$$

$$= 2 \mathbb{E}(\mathbb{1}_{[k+1,\infty)}(X_{1})) \to 0$$

für $k \to \infty$. Insgesamt erhält man also nach Theorem 4.2. von BILLINGSLEY

$$d_1(\mathcal{L}(\sum_{j=1}^{\infty} p_n(j)(\sqrt{n}\sum_{i=1}^{n} c_{n,\tau(i)} Y_{n,i}^{(j)})^2 | (X_i)_{i \le n}), \mathcal{L}(\sum_{j=1}^{\infty} p_j(Z^{(j)})^2)) \xrightarrow{P} 0, \quad (4.50)$$

so dass $_dW_n^2$ -Tests, deren kritischen Werte mittels der Permutationsverteilung von $_dW_n^2$ bestimmt werden, verwendet werden können.

Auf die gleiche Art und Weise erhält man auch Zweistichproben-Bootstrap- oder Double-Bootstrap-Tests. Die zugehörige Bootstrap-Version der obigen Teststatistik ist dann

$$_{d}W_{n}^{*} := \frac{n_{1}n_{2}}{n} \sum_{k=1}^{\infty} (T_{n_{1}}^{*}(k) - T_{n_{2}}^{*}(k))^{2} p_{n}(k),$$

wobei $T_{n_1}^*(k) = \sum_{i=1}^{n_1} W_{n,i} \mathbb{1}_{(-\infty,k]}(X_i)$ und $T_{n_2}^*(k) = \sum_{i=1}^{n_2} \tilde{W}_{n,i} \mathbb{1}_{(-\infty,k]}(Y_i)$ für jeweils unabhängige Bootstrap- oder Double-Bootstrap-Gewichte $(W_{n,i})_{i \leq n_1}$ und $(\tilde{W}_{n,i})_{i \leq n_2}$ gelte. Aufgrund der bedingten Unabhängigkeit von $T_{n_1}^*$ und $T_{n_2}^*$ folgt mit $T_{n_2}^*$

$$\operatorname{Var}\left(\frac{\sqrt{n_1} T_{n_1}^*(k)\sqrt{p_n(k)}}{(\sum_{i=1}^{n_1} W_{n,i}^2)^{1/2}} \middle| \vec{X}\right) = \frac{n_1 p_n(k)}{n_1 - 1} \sum_{i=1}^{n_1} (\mathbb{1}_{(-\infty,k]}(X_i) - \sum_{j=1}^{n_1} \mathbb{1}_{(-\infty,k]}(X_j))^2 \\ \leq 2p_n(k)$$

und unter Berücksichtigung, dass diese Gewichte die Bedingung (3.13) aus Zusatz 3.5, also insbesondere (4.33) erfüllen, alles analog zu oben.

 $^{^{13}}$ Man setzt $Y_i := X_{n_1+i}$ für $i > n_1$.

 $^{^{14}}$ Für $\vec{X} := (X_1, \dots, X_n)$.

Kapitel 5

Explizite Anwendungen in der Testtheorie

In diesem Kapitel werden nun die Grenzwersätze aus den vorangegangenen Kapiteln verwendet, um die asymptotische Validität von Resampling-Tests für spezielle Testprobleme aufzuzeigen. Hierbei wird zum einen die asymptotische Konsistenz unter der Nullhypothese in Form von (2.15) gezeigt und zum anderen das Verhalten der Gütefunktion unter Alternativen untersucht.

Wie schon zu Beginn dieser Arbeit in Kapitel 2 bemerkt wurde, wird dabei in der Vorgehensweise zwischen Ein- und Zweistichprobenproblemen unterschieden.

5.1 Resampling-Tests für Einstichprobenprobleme

Da Permutationsverfahren bei Einstichprobenproblemen mit austauschbaren Daten die Nullhypothese nicht von der Alternative trennen können, liegen in diesem Abschnitt die verschiedenen Bootstrap-Verfahren im Mittelpunkt der Betrachtung.

In der ersten Anwendung wird ein Test zur Überprüfung der Kovarianz eines zweidimensionalen Vektors entwickelt.

Resampling-Test für Kovarianzen

Sei $\binom{X_1}{Y_1},\ldots,\binom{X_n}{Y_n}$ eine i.i.d. Folge \mathbb{R}^2 -wertiger Zufallsvariablen mit unbekann-

tem Erwartungswert $E((X_1, Y_1)^t) = (\mu_1, \mu_2)^t$ und unbekannter Kovarianzmatrix

$$Cov((X_1, Y_1)^t) = \begin{pmatrix} \sigma_1^2 & c \\ c & \sigma_2^2 \end{pmatrix}$$

mit $\sigma_i^2 > 0$ für i = 1, 2. Es wird zunächst das einseitige Testproblem

$$H: \{c \le 0\} \text{ gegen } K: \{c > 0\}$$
 (5.1)

untersucht. Zweiseitige Tests auf Korreliertheit der beiden Koordinaten (d.h. $H=\{c=0\}$ gegen $K=\{c\neq 0\}$) erhält man wie üblich als Summe zweier einseitiger Test. Sei nun $J=\{c=0\}$ der Rand der einseitigen bzw. zweiseitigen Nullhypothese.

Als mögliche Teststatistik kommt die empirische Kovarianz (oder im zweiseitigen Fall deren Betrag)

$$T_n := \sqrt{n} \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n) (Y_i - \overline{Y}_n)$$

in Frage. Da T_n nicht vom Erwartungswert abhängt, wird im Weiteren ohne Einschränkung $\mu_1 = \mu_2 = 0$ angenommen. Dann gibt der nächste Satz Auskunft über das asymptotische Verhalten der Teststatistik unter dem Rand der Nullhypothese.

Satz 5.1

Existieren vierte Momente $\mathrm{E}(X_1^4)+\mathrm{E}(Y_1^4)<\infty$, so konvergiert die Teststatistik unter J in Verteilung gegen eine normalverteilte Zufallsvariable X

$$T_n \xrightarrow{\mathcal{D}} X \sim N(0, \operatorname{Var}(X_1 Y_1)).$$
 (5.2)

Beweis. Nach Slutzky folgt die Behauptung direkt

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \overline{X}_n)(Y_i - \overline{Y}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i Y_i - \sqrt{n} \, \overline{X}_n \overline{Y}_n$$

$$\xrightarrow{\mathcal{D}} Z_1 \sim N(0, \operatorname{Var}(X_1 Y_1)).$$

An dieser Stelle tritt nun das Problem auf, dass die Grenzverteilung von der unbekannten Größe $\tilde{\sigma}^2 := \operatorname{Var}(X_1Y_1)$ abhängt, so dass der zugehörige asymptotische Test $\varphi_n = \mathbb{1}_{(\tilde{\sigma}u_{1-\alpha},\infty)}(T_n)$ nicht verteilungsfrei ausgeführt werden kann.

Um dies zu umgehen, kann man die Teststatistik entweder geeignet Studentisieren oder datenabhängige kritische Werte, die durch Resampling-Verfahren erzeugt werden, verwenden.

Es wird zunächst der zweite Fall untersucht. Seien hierfür wieder $(W_{n,i})_{i\leq n}$ die Resampling-Gewichte aus Kapitel 3 mit $\overline{W}_n = 0$. Da die ursprüngliche Stichprobe aus den Paaren $\binom{X_i}{Y_i}$, $i \leq n$, besteht, wird die Resampling-Stichprobe auch hieraus gezogen. Dies kann z.B. durch die folgende Resampling-Version von T_n beschrieben werden

$$T_n^* := \sum_{i=1}^n W_{n,i} X_i Y_i - \left(\frac{1}{\sqrt{n}} \sum_{j=1}^n W_{n,j} X_j\right) \left(\sum_{j=1}^n W_{n,j} Y_j\right). \tag{5.3}$$

Für die klassischen Bootstrap-Gewichte, vgl. Beispiel 2.4 mit m(n) = n, erhält man beispielsweise

$$T_n^* = \frac{1}{\sqrt{n}} (\sum_{i=1}^n X_i^* Y_i^* - \sum_{j=1}^n X_j Y_j) - \sqrt{n} (\overline{X}_n^* - \overline{X}_n) (\overline{Y}_n^* - \overline{Y}_n).$$

Satz 5.2

Erfüllen die Gewichte die Generalvoraussetzungen mit $Var(W_1) = 1$ sowie $\overline{W}_n = 0$ und gilt wie in Satz 5.1 $E(X_1^4) + E(Y_1^4) < \infty$, so liegt für die Resampling-Version der Teststatistik ein bedingter Zentraler Grenzwertsatz vor

$$d_1\Big(\mathcal{L}(T_n^*|X_1Y_1,\dots,X_nY_n),N\big(0,\operatorname{Var}(X_1Y_1)\big)\Big) \xrightarrow{P} 0.$$
 (5.4)

Beweis. Es wird zunächst gezeigt, dass der zweite Term aus (5.3) asymptotisch vernachlässigbar ist. Aufgrund der Voraussetzungen erfüllt das Dreiecksschema $Z_{n,i} := \frac{1}{\sqrt{n}} {X_i \choose Y_i}$ nach Bemerkung 3.8 die Maximumsbedingung (3.46). Weiter gilt

$$\frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 & \sum_{i=1}^{n} (X_i - \overline{X}_n) (Y_i - \overline{Y}_n) \\ \sum_{i=1}^{n} (X_i - \overline{X}_n) (Y_i - \overline{Y}_n) & \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \sigma_1^2 & c \\ c & \sigma_2^2 \end{pmatrix},$$

so dass aus dem bedingten multivariaten Zentralen Grenzwertsatz 3.20 die Konvergenz

$$d_2\left(\mathcal{L}(\sqrt{n}\sum_{i=1}^n W_{n,i}Z_{n,i}|Z_{n,1},\ldots,Z_{n,n}),N\left(0,\begin{pmatrix}\sigma_1^2&c\\c&\sigma_2^2\end{pmatrix}\right)\right)\stackrel{P}{\longrightarrow}0$$

folgt. Hieraus erhält man mit Hilfe des Satzes über stetige Abbildungen und des bedingten Lemmas von Slutzky, vgl. Lemma A.9,

$$d_1\Big(\mathcal{L}\Big(\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^n W_{n,j}X_j\Big)\Big(\sum_{j=1}^n W_{n,j}Y_j\Big)|Z_{n,1},\ldots,Z_{n,n}\Big),\varepsilon_0\Big)\stackrel{P}{\longrightarrow} 0.$$

Da das zeilenweise unabhängige und identisch verteilte Dreiecksschema

$$X_{n,i} := \frac{1}{\sqrt{n}} X_i Y_i, \quad i \le n,$$

nach Bemerkung 3.8 den Bedingungen von Satz 3.3 genügt, folgt für den ersten Term aus (5.3) die Konvergenz

$$d_1\left(\mathcal{L}\left(\sum_{i=1}^n W_{n,i}X_iY_i|X_{n,1},\ldots,X_{n,n}\right),N\left(0,\operatorname{Var}(X_1Y_1)\right)\right)\stackrel{P}{\longrightarrow} 0,$$

woraus sich nach dem bedingten Lemma von Slutzky A.9 und Satz A.3 die Behauptung ergibt.

Somit kann man zum Testen von (5.1) den Resampling-Test $\varphi_n^* = \mathbb{1}_{(c_n^*(\alpha),\infty)}(T_n)$ verwenden, wobei $c_n^*(\alpha)$ das datenabhängige $(1-\alpha)$ -Quantil der bedingten Verteilung $\mathcal{L}(T_n^*|X_{n,1}\ldots X_{n,k(n)})$ ist. Dieser lässt sich somit ohne Kenntnis von $\tilde{\sigma}$ ausführen und ist nach Lemma 2.7 auch asymptotisch äquivalent zu φ_n .

Da Hall und Wilson [25] u.a. bemerkt haben, dass es von Vorteil ist studentisierte Resampling-Statistiken zur Bestimmung des kritischen Wertes zu verwenden, werden diese als nächstes untersucht. Eine geeignete Studentisierung ist im unbedingten Fall durch

$$V_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i Y_i - \frac{1}{n} \sum_{j=1}^n X_j Y_j)^2$$
 (5.5)

gegeben, da hierfür $V_n^2 \xrightarrow{f.s.} \operatorname{Var}(X_1Y_1)$ gilt. Ist $\operatorname{Var}(X_1Y_1) > 0$, so gilt dies schliesslich auch für V_n^2 fast sicher und man kann den asymptotischen α -Niveau-Test $\psi_n = \mathbb{1}_{(u_{1-\alpha},\infty)}(T_n/V_n)$ verwenden.¹

Um nun eine Resampling-Version dieses Tests zu finden, muss auch der Nenner der Teststatistik mit resampelt werden. Dazu werden aus beweistechnischen Gründen wie bei Janssen [35] nur Resampling-Gewichte mit der Eigenschaft

$$W_{n,i} = W'_{n,i} - \frac{m(n)^{1/2}}{k(n)}, \ \overline{W}_n = 0,$$
 (5.6)

behandelt. Hierzu zählen u.a. die m(n)-Bootstrap- und m(n)-double-Bootstrap-Gewichte, sowie viele weitere Beispiele aus Kapitel 6.

¹Die genauere Wahl währe hier - wie auch bei den weiteren studentisierten Tests - eigentlich $\mathbbm{1}_{(u_{1-\alpha},\infty)}(\mathbbm{1}_{\{V_n^2>0\}}T_n/V_n)$. Da dies aber zu unübersichtlich wäre, wird dies im Folgenden unterdrückt.

Auf diese Weise kann man analog zu Janssen [35, Kapitel 3]

$$V_n^{*2} := \frac{1}{m(n)^{1/2}} \sum_{i=1}^n W'_{n,i} X_i^2 Y_i^2 - \frac{1}{m(n)} \left(\sum_{j=1}^n W'_{n,i} X_j Y_j \right)^2$$
 (5.7)

als Resampling-Version von V_n^2 verwenden und erhält so den folgenden Satz.

Satz 5.3

Erfüllen die Gewichte die Generalvoraussetzungen mit $Var(W_1) = 1$ sowie Gleichung (5.6) und existieren vierte Momente $E(X_1^4) + E(Y_1^4) < \infty$, so gilt unter der Zusatzvoraussetzung $\liminf_{n\to\infty} m(n)/k(n) > 0$ der folgende bedingte Zentrale Grenzwertsatz

$$d_1\left(\mathcal{L}\left(\frac{T_n^*}{V_n^*}|X_1Y_1,\dots,X_nY_n\right),N(0,1)\right) \xrightarrow{P} 0,\tag{5.8}$$

wobei T_n^* und V_n^* durch (5.3) bzw. (5.7) gegeben sind.

Beweis. Dies folgt nach Theorem 3.1. von Janssen [35] aufgrund der Zerlegung in (5.3) zunächst für die bedingte Verteilung des ersten Terms $\sum_{i=1}^{n} W_{n,i} X_i Y_i / V_n^*$ und dann mit Hilfe des Lemmas von Slutzky und des Beweises von Satz 5.2 auch für T_n^* / V_n^* .

Der zu ψ_n gehörige studentisierte asymptotisch effektive Resampling-Test ist $\psi_n^* = \mathbb{1}_{(c_{n,1}^*(\alpha),\infty)}(T_n/V_n)$, wobei $c_{n,1}^*(\alpha)$ diesmal das $(1-\alpha)$ -Quantil der bedingten Verteilung $\mathcal{L}(T_n^*/V_n^*|X_1Y_1...X_nY_n)$ ist. Hierfür erhält man nun Aussagen über das Verhalten der Gütefunktion.

Satz 5.4

Unter den Voraussetzungen von Satz 5.2 gelten

$$E_0(|\psi_n^* - \psi_n|) \to 0,$$
 (5.9)

$$E_c(\psi_n^*) \to 1 \text{ für festes } c > 0$$
 (5.10)

$$E_c(\psi_n^*) \to 0 \text{ für festes } c < 0.$$
 (5.11)

Für lokale Alternativen $c = c_n$ mit $\sqrt{n}c_n \to \tau \in \mathbb{R}$ erhält man zusätzlich

$$E_{c_n}(\psi_n^*) \to 1 - \Phi\left(u_{1-\alpha} - \frac{\tau}{\tilde{\sigma}}\right).$$
 (5.12)

Beweis. Der erste Teil (5.9) folgt aufgrund von Satz 5.3 aus Lemma 2.7. Für die anderen Teile bemerkt man, dass aufgrund des Beweises von Satz 5.1

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i Y_i + o_P(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i Y_i - c) + o_P(1) + \sqrt{n} \cdot c$$

gilt und somit für feste Alternativen $c \in \mathbb{R}$ nach Slutzky stochastische Konvergenz $T_n \xrightarrow{P} \operatorname{sign}(c) \cdot \infty$ und für lokale Alternativen c_n Verteilungskonvergenz $T_n \xrightarrow{\mathcal{D}} X \sim N(\tau, \tilde{\sigma}^2)$ folgt. Zusammen mit Satz 5.3 impliziert dies die restlichen Behauptungen.

Bemerkung 5.5

- 1. Für die oben angesprochenen zweiseitigen Tests erhält man an dieser Stelle analoge Resultate. 2
- 2. Man kann genauso auch auf den Korrelationskoeffizienten $\rho = 0$ mittels der empirischen Korrelation testen. Da unter den obigen Bedingungen die empirischen Varianzen der einzelnen Koordinatengruppen nämlich gegen die jeweiligen Varianzen konvergieren, erhält man alles auf analoge Weise. Allerdings muss man die Studentisierungen in (5.5) und (5.7) noch geeignet erweitern.

Janssen untersucht in seiner Arbeit von 2005, vgl. [35], gewichtete Resampling-Versionen der studentschen t-Statistik. Dies wird im Folgenden aufgegriffen und auf den mehrdimensionalen Fall erweitert.

Resampling-Tests in endlichdimensionalen Quadratischen Formen

Gegeben sei ein Dreiecksschema von k-dimensionalen Zufallsvektoren $(X_{n,i})_{i \leq k(n)}$ mit $X_{n,i} = (X_{n,i}^{(j)})_{j \leq k}$ für die man die Hypothese

$$H = \{ E(X_{n,i}) = 0 \text{ für alle } i \le k(n) \}$$

gegen geeignete Alternativen testen möchte. Um hierfür Satz 3.20 anzuwenden, wird im Folgenden angenommen, dass das Dreiecksschema der Bedingung (3.46) genügt, und dass die zugehörige empirische Kovarianzmatrix stochastisch gegen

²Mit dem Unterschied, dass in (5.11) auch Konvergenz gegen 1 vorliegt, da der Betrag der Teststatistik betrachtet wird.

eine fast sicher positiv definite Kovarianzmatrix $\Gamma \in \mathbb{R}_{p.d.}^{k \times k}$ konvergiert, d.h. es gilt

$$\Gamma_n = \left(\sum_{i=1}^{k(n)} X_{n,i}^{(j)} X_{n,i}^{(r)} - k(n) \overline{X}_n^{(j)} \overline{X}_n^{(r)}\right) \Big|_{1 \le j,r \le k} \xrightarrow{P} \Gamma.$$
 (5.13)

Da dann auch Γ_n schließlich fast sicher positiv definit ist, kann man zum Testen von H die Quadratische Form

$$Q_n := \left(\sum_{i=1}^{k(n)} X_{n,i}\right)^t \Gamma_n^{-1} \sum_{i=1}^{k(n)} X_{n,i} \, \mathbb{1}_{\{\Gamma_n^{-1} \text{ existient}\}}$$

verwenden. Diese konvergiert nach Satz 4.9 unter der Zusatzvoraussetzung

$$\sum_{i=1}^{k(n)} X_{n,i} \xrightarrow{\mathcal{D}} Z \sim N(0,\Gamma)$$
 (5.14)

unter H in Verteilung gegen eine χ_k^2 -verteilte Zufallsvariable, so dass $\psi_n = \mathbb{1}_{(\chi_{k;\alpha}^2,\infty)}(Q_n)$ ein asymptotischer α -Niveau Test ist. Dabei bezeichnet $\chi_{k;\alpha}^2$ das $(1-\alpha)$ -Quantil der χ_k^2 -Verteilung. Möchte man nun den zugehörigen Resampling-Test konstruieren, so ist es von Vorteil, auch die empirische Kovarianzmatrix Γ_n nach der selben Prozedur zu resamplen, vgl. Hall und Wilson [25] für den eindimensionalen Fall. Dazu werden aus beweistechnischen Gründen wieder nur Resampling-Gewichte der Form (5.6) betrachtet. Für diese Gewichte schreibt sich dann die Resampling-Version von Γ_n als

$$\Gamma_n^* = \frac{k(n)}{\sqrt{m(n)}} \Big(\sum_{i=1}^{k(n)} W_{n,i}' X_{n,i}^{(j)} X_{n,i}^{(r)} - \frac{1}{\sqrt{m(n)}} (\sum_{i=1}^{k(n)} W_{n,i}' X_{n,i}^{(j)}) (\sum_{i=1}^{k(n)} W_{n,i}' X_{n,i}^{(r)}) \Big)_{1 \leq j,r \leq k}.$$

Ist $T_n^* = T_n^*((X_{n,i})_{i \leq k(n)})$ die zugehörige multivariate Resampling-Statistik aus Abschnitt 3.2.1, so erhält man als Resampling-Version von Q_n die Statistik

$$Q_n^* = T_n^{*t} \Gamma_n^{*-1} T_n^* \mathbb{1}_{\{\Gamma_n^{*-1} \text{ existient}\}}.$$

Der nächste Satz zeigt nun, dass Q_n^* bedingt nach den Daten gegen eine χ_k^2 verteilte Zufallsvariable konvergiert.

Satz 5.6

Erfüllen die Gewichte die Generalvoraussetzungen mit $Var(W_1) = 1$ sowie die Bedingung (5.6) und gilt $\frac{m(n)}{k(n)} \to c > 0$, so implizieren die obigen Voraussetzungen (3.46) und (5.13) den folgenden bedingten Grenzwertsatz

$$d_1(\mathcal{L}(Q_n^*|X_{n,1},\dots,X_{n,k(n)}),\chi_k^2) \xrightarrow{P} 0.$$
 (5.15)

Beweis. Nach Voraussetzung (5.13) konvergiert die Wahrscheinlichkeit der Menge

$$I_n := \bigcup_{j \le k} \{ \sum_{s=1}^{k(n)} (X_{n,s}^{(j)} - \overline{X}_n^{(j)})^2 > 0 \}$$

gegen 1, so dass man sich für die folgenden Betrachtungen auf diese Menge beschränken kann.

Es wird zunächst gezeigt, dass man ohne Einschränkung annehmen kann, dass $\sum_{s=1}^{k(n)} (X_{n,s}^{(j)} - \overline{X}_n^{(j)})^2 = 1$ für alle $1 \leq j \leq k$ gilt. Hierzu rechnet man mit Hilfe der standardisierten Zufallsvariablen

$$\tilde{X}_{n,i}^{(j)} := \frac{X_{n,i}^{(j)}}{\left(\sum_{s=1}^{k(n)} (X_{n,s}^{(j)} - \overline{X}_n^{(j)})^2\right)^{1/2}}, \quad i \le k(n),$$

die Gleichheit $Q_n^*((X_{n,i})_{i\leq k(n)})=Q_n^*((\tilde{X}_{n,i})_{i\leq k(n)})$ nach. Diese Gleichheit wird durch die Gleichungen

$$T_n^*(\tilde{X}) = Diag\left(\left(\sum_{s=1}^{k(n)} (X_{n,s}^{(j)} - \overline{X}_n^{(j)})^2\right)^{-1/2}\right)_{j \le k} T_n^*(X) =: D_n T_n^*(X)$$

und

$$(D_n\Gamma_n^{*-1}(\tilde{X})D_n)^{-1} = D_n^{-1}\Gamma_n^*(\tilde{X})D_n^{-1} = \Gamma_n^*(X)$$

ersichtlich.

Als nächstes wird gezeigt, dass man zusätzlich ohne Einschränkung $\overline{X}_n^{(j)}=0$ für alle $1\leq j\leq k$ fordern kann. Analog zu Janssen [35, Theorem 3.1] folgt dies für $\hat{X}_{n,i}^{(j)}:=X_{n,i}^{(j)}-\overline{X}_n^{(j)}$ wegen $\sum_{i=1}^{k(n)}W_{n,i}'=m(n)^{1/2}$ direkt aus der Gleichheit

$$\sum_{i=1}^{k(n)} W'_{n,i} \hat{X}_{n,i}^{(j)} \hat{X}_{n,i}^{(r)} - \frac{1}{m(n)^{1/2}} \left(\sum_{i=1}^{k(n)} W'_{n,i} \hat{X}_{n,i}^{(j)} \right) \left(\sum_{i=1}^{k(n)} W'_{n,i} \hat{X}_{n,i}^{(r)} \right) \\
= \sum_{i=1}^{k(n)} W'_{n,i} X_{n,i}^{(j)} X_{n,i}^{(r)} - \frac{1}{m(n)^{1/2}} \left(\sum_{i=1}^{k(n)} W'_{n,i} X_{n,i}^{(j)} \right) \left(\sum_{i=1}^{k(n)} W'_{n,i} X_{n,i}^{(r)} \right) (5.16)$$

für alle $1 \leq j, r \leq k$. Multipliziert man den Term auf der rechten Seite von (5.16) noch mit dem Normierungsfaktor $k(n)/m(n)^{1/2}$, so konvergiert dieser aufgrund des bedingten Zentralen Grenzwertsatzes 3.3 und dem Lemma von

Slutzky stochastisch gegen 0

$$\frac{k(n)}{m(n)} \left(\sum_{i=1}^{k(n)} W'_{n,i} X_{n,i}^{(j)} \right) \left(\sum_{i=1}^{k(n)} W'_{n,i} X_{n,i}^{(r)} \right)
= \left(\frac{1}{m(n)^{1/2}} \left(k(n)^{1/2} \sum_{i=1}^{k(n)} W_{n,i} X_{n,i}^{(j)} \right) + k(n)^{1/2} \overline{X}_{n}^{(j)} \right) \left(\frac{k(n)}{m(n)} \right)^{1/2} \sum_{i=1}^{k(n)} W'_{n,i} X_{n,i}^{(r)}
\xrightarrow{P} 0.$$

Zur Behandlung des noch fehlenden Terms, wendet man Satz 3.3 auf das Dreiecksschema $Y_{n,i} := (\frac{k(n)}{m(n)})^{1/2} X_{n,i}^{(j)} X_{n,i}^{(r)}$ an. Da hierfür nach (3.46) die Maximumsbedingung erfüllt ist und

$$\sum_{i=1}^{k(n)} (Y_{n,i} - \overline{Y}_n)^2 \le \sum_{i=1}^{k(n)} Y_{n,i}^2 \le \frac{k(n)}{m(n)} \max_{i \le k(n)} (X_{n,i}^{(j)})^2 \max_{i \le k(n)} (X_{n,i}^{(r)})^2 \xrightarrow{P} 0$$

gilt, folgt somit aufgrund von $\overline{W}_n = 0$ die Konvergenz

$$k(n)^{1/2} \sum_{i=1}^{k(n)} W_{n,i} Y_{n,i} \stackrel{P \otimes \tilde{P}}{\longrightarrow} 0.$$

Setzt man nun die Gewichte ein, so liefert dies die Konvergenz

$$\frac{k(n)}{m(n)^{1/2}} \sum_{i=1}^{k(n)} W'_{n,i} X_{n,i}^{(j)} X_{n,i}^{(r)} - \sum_{i=1}^{k(n)} X_{n,i}^{(j)} X_{n,i}^{(r)} \stackrel{P \otimes \tilde{P}}{\longrightarrow} 0.$$

Da $\overline{X}_n = 0$ gilt, impliziert dies

$$\Gamma_n - \Gamma_n^* \stackrel{P \otimes \tilde{P}}{\longrightarrow} 0$$

und somit auch

$$\Gamma_n^* \stackrel{P \otimes \tilde{P}}{\longrightarrow} \Gamma.$$

Wendet man jetzt Satz 3.20 wie in den vorangegangenen Kapiteln punktweise an, so folgt die Behauptung aus Satz 4.9.

Bemerkung 5.7

1. Die Aussage bleibt auch dann gültig, wenn Γ_n in (5.13) stochastisch gegen eine Zufallsvariable Γ mit Werten in $\mathbb{R}_{p.d.}^{k\times k}$ konvergiert. Hierzu braucht man nur die Schritte des obigen Beweises mit Hilfe des Teilfolgenkriteriums für stochastische Konvergenz durchzugehen.

2. Die oben geforderten Voraussetzungen an die Gewichte werden u.a. von den m(n)-Bootstrap- sowie von bestimmten m(n)-double-Bootstrap-Gewichten erfüllt, vgl. Kapitel 6.

Als Anwendung erhält man den im Bezug auf ψ_n asymptotisch effektiven Resampling-Test $\psi_n^* = \mathbb{1}_{(q_\alpha^*,\infty)}(Q_n)$, wobei $q_\alpha^* = q_\alpha^*(\cdot)$ das $(1-\alpha)$ -Quantil der bedingten Verteilung $\mathcal{L}(Q_n^*|X_{n,1},\ldots,X_{n,k(n)})$ ist.

Das asymptotische Verhalten der Gütfunktion wird im Folgenden untersucht.

Satz 5.8

Sei $E(X_{n,i}) = \mu_{n,i}$ für $1 \le i \le k(n)$. Liegt zusätzlich zu den Voraussetzungen von Satz 5.6 unter H die Konvergenz (5.14) vor, so verhält sich die Gütefunktion von ψ_n^* wie folgt:

$$E(\psi_n^*) \to 1 \ , falls \ \| \sum_{i=1}^{k(n)} \mu_{n,i} \| \to \infty \ gilt \ und$$
 (5.17)

$$E(\psi_n^*) \to \chi_k^2(\delta^2) \left((\chi_{k;\alpha}^2, \infty) \right) , falls \sum_{i=1}^{k(n)} \mu_{n,i} \to \mu \in \mathbb{R}^k \ gilt.$$
 (5.18)

Dabei ist $\delta^2 = \mu^t \Gamma^{-1} \mu$.

Beweis. Da sich der kritische Wert nach Satz 5.6 unter den obigen Voraussetzungen nicht verändert und somit immer gegen das $(1 - \alpha)$ -Quantil der χ_k^2 -Verteilung konvergiert, muss nur noch das Verhalten der Teststatistik Q_n untersucht werden. Da Γ positiv definit ist, trifft dies auch auf Γ^{-1} zu, so dass auch Γ_n^{-1} schließlich fast sicher postiv definit ist. Hiermit folgt

$$Q_n = \left(\sum_{i=1}^{k(n)} X_{n,i} - E(X_{n,i}) + \sum_{j=1}^{k(n)} \mu_{n,j}\right)^t \Gamma_n^{-1} \left(\sum_{i=1}^{k(n)} X_{n,i} - E(X_{n,i}) + \sum_{j=1}^{k(n)} \mu_{n,j}\right).$$

Da im ersten Fall für den dominierenden Term die stochastische Konvergenz

$$\left(\sum_{i=1}^{k(n)} \mu_{n,i}\right)^t \Gamma_n^{-1} \sum_{i=1}^{k(n)} \mu_{n,i} \xrightarrow{P} \infty$$

vorliegt, erhält man mit Hilfe des Lemmas von Slutzky die stochastische Konvergenz von Q_n gegen ∞ und so (5.17).

Die Behauptung (5.18) erhält man aus Satz 4.9.

Bemerkung 5.9

- 1. Die obigen Voraussetzungen sind z.B. für Dreiecksschema $X_{n,i} = X_i / \sqrt{k(n)}$, $i \leq k(n)$, in unabhängig und identisch verteilten Zufallsvektoren $(X_i)_i$ mit positiv definiter Kovarianzmatrix $\Gamma \in \mathbb{R}_{p.d.}^{k \times k}$ erfüllt.
- 2. Analog zu Kapitel 4 bei WITTING [65] kann man das Testproblem für zeilenweise konstante Erwartungswerte $E(X_{n,i}) = \mu_n$ erweitern, indem man testet, ob μ_n in einem linearen Teilraum liegt oder nicht. Da hierfür keine neuen Beweisideen benötigt werden, wird dies nicht weiter ausgeführt, um den Umfang dieser Arbeit im Rahmen zu halten.

5.2 Resampling-Tests für Zweistichprobenprobleme

Anders als bei Einstichprobenproblemen besitzen Permutationstests für Zweistichprobenprobleme im Gegensatz zu den einzelnen Bootstrap-Varianten den Vorteil, dass sie unter den Teilmengen der Nullhypothese, für die die beiden Stichproben austauschbar sind, exakt das Niveau einhalten. Aus diesem Grund liegt in diesem Abschnitt der Fokus auf der Entwicklung solcher verteilungsfreier Permutationstests. Da die Beweise wieder mit Hilfe des gewichteten Resamplings geführt werden, gelten viele der folgenden Aussagen auch für Zweistichproben-Bootstrap-Tests³.

Für diesen Abschnitt sei stets vorausgesetzt, dass $k(n) = n = n_1 + n_2$ mit $\min(n_1, n_2) \to \infty$ und $n_1/n \to p \in (0, 1), q = 1 - p$ gelte. Der allgemeinere Fall $0 < \liminf \frac{n_1}{n} \le \limsup \frac{n_1}{n} < 1$ folgt dann hieraus - wie im Beweis zu Satz 3.19 - durch Betrachtung konvergenter Teilfolgen.

Als erstes Anwendungsbeispiel wird ein Permutationstest zum Vergleich zweier Varianzen konstruiert. Als Resultat ergibt sich dabei in den Sätzen 5.15 und 5.16, dass sich der F-Test im nichtparametrischen Fall ohne asymptotischen Qualitätsverlust als studentisierter Permutationstest ausführen lässt.

³Hierauf wird später in Bemerkung 5.26 noch etwas genauer eingegangen.

Permutations-F-Test

Nun betrachten wir auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) zwei voneinander unabhängige reelle Stichproben

$$X_1, \dots, X_{n_1}$$
 i.i.d. P^{X_1} , (5.19)
 Y_1, \dots, Y_{n_2} i.i.d. P^{Y_1}

mit unbekannten Erwartungswerten $E(X_1) = \mu_1$, $E(Y_1) = \mu_2$ und unbekannten, positiven Varianzen $\sigma_1^2 := Var(X_1)$, $\sigma_2^2 := Var(Y_1) > 0$ sowie unbekannten vierten Momenten $\rho_1^4 := E(X_1^4) < \infty$, $\rho_2^4 := (Y_1^4) < \infty$. Das erweiterte Behrens-Fisher-Modell, vgl. Janssen [29, Example 1.1],

$$X_{i} = \mu_{1} + \sigma_{1} Z_{i}, \ 1 \le i \le n_{1},$$

$$Y_{i} = \mu_{2} + \sigma_{2} Z_{n_{1} + i}, \ 1 \le i \le n_{2}$$

$$(5.20)$$

für eine Folge von i.i.d. Zufallsvariablen $(Z_i)_{i\in\mathbb{N}}$ mit $\mathrm{E}(Z_1)=0, \mathrm{Var}(Z_1)=1, \mathrm{E}(Z_1^4)<\infty$, ist somit als Teilmodell enthalten. Der Fokus liegt nun auf dem Testproblem

$$H: \{\sigma_1^2 \le \sigma_2^2\}$$
 gegen $K: \{\sigma_1^2 > \sigma_2^2\}$.

Sind die Zufallsvariablen normalverteilt, so ist durch den F-Test mit der Teststatistik

$$F_n = \frac{\frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \overline{X}_{n_1})^2}{\frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \overline{Y}_{n_2})^2} =: \frac{\hat{\sigma}_{n_1}^2(X)}{\hat{\sigma}_{n_2}^2(Y)}$$
(5.21)

ein optimaler (und exakter) α -Niveau Test gegeben, vgl. WITTING [65, S.201]. In einer nichtparametrischen Situation ist dies nicht mehr der Fall, vgl. z.B. VAN DER VAART [63, Example 3.3]. Aus diesem Grund wird nun in mehreren Schritten eine Permutationsversion des F-Tests für das obige Testproblem entwickelt. Dabei wird die Teststatistik (5.21) zunächst geeignet transformiert, um dann deren Permutationsverteilung mit Hilfe der Sätze aus den vorangegangenen Kapiteln zu studieren, vgl. Satz 5.12 unten. Hierbei wird sich herausstellen, dass der zugehörige, nicht-studentisierte Permutationstest nur für Teilmodelle von (5.19) konsistent ist. Um dies zu umgehen, verwendet man eine adäquate Studentisierung V_n , vgl. Gleichung (5.38) unten, und erhält so zum Abschluss

den studenisierten Permutationstest

$$\varphi_{n,PF}^* = \begin{cases} 1 & \text{, falls } \left(\frac{n_1 n_2}{n}\right)^{1/2} \cdot \frac{\log(F_n) \mathbb{1}_{\{V_n^2 > 0\}}}{V_n} > c_{n,PF}^*(\alpha). \end{cases}$$
(5.22)

Hierbei ist $c_{n,PF}(\alpha)^*$ das bedingte $(1-\alpha)$ -Quantil der Permutationsverteilung der Teststatistik.

Zur Abarbeitung der obigen Vorgehensweise wird nun für das Folgende ohne Einschränkung angenommen, dass $E(X_1) = E(Y_1) = 0$ gilt (da die Teststatistik nicht von μ_i abhängt).

So erhält man zunächst unter dem Rand der Hypothesen $H_0 = {\sigma_1^2 = \sigma_2^2 =: \sigma^2}$ für die transformierte Statistik $\log(F_n)$ aus der Taylor-Entwicklung des Logarithmus um σ^2

$$\log(x) = \log(\sigma^2) + \frac{1}{\sigma^2}(x - \sigma^2) + R_2(x) \text{ mit } \frac{R_2(x)}{x - \sigma^2} \xrightarrow[x \to \sigma^2]{} 0$$

die Darstellung

$$a_{n} \log(F_{n}) = a_{n} \left(\log(\hat{\sigma}_{n_{1}}^{2}(X)) - \log(\hat{\sigma}_{n_{2}}^{2}(Y)) \right)$$

$$= a_{n} \left(\log(\sigma^{2}) + \frac{\hat{\sigma}_{n_{1}}^{2}(X) - \sigma^{2}}{\sigma^{2}} + R_{2}(\hat{\sigma}_{n_{1}}^{2}(X)) \right)$$

$$- \left(\log(\sigma^{2}) + \frac{\hat{\sigma}_{n_{2}}^{2}(Y) - \sigma^{2}}{\sigma^{2}} + R_{2}(\hat{\sigma}_{n_{2}}^{2}(Y)) \right) \right)$$

$$= a_{n} \frac{\hat{\sigma}_{n_{1}}^{2}(X) - \hat{\sigma}_{n_{2}}^{2}(Y)}{\sigma^{2}}$$

$$+ a_{n} \left(R_{2}(\hat{\sigma}_{n_{1}}^{2}(X)) - R_{2}(\hat{\sigma}_{n_{2}}^{2}(Y)) \right),$$

$$(5.23)$$

wobei a_n eine noch zu bestimmende Normierungskonstante ist. Zur Behandlung der Restterme in (5.25) benötigen wir noch eine Aussage über die Konvergenzgeschwindigkeit der empirischen Varianz im i.i.d. Fall. Dies liefert der folgende Hilfssatz.

Hilfssatz 5.10

Für eine Folge von zentrierten i.i.d. Zufallsvariablen $Z_1, Z_2 \dots$ mit Varianz σ^2 und existierendem vierten Moment $\rho^4 := E(Z_1^4) < \infty$ gilt

$$\sqrt{n} \left(\frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n)^2 - \sigma^2 \right) \xrightarrow{\mathcal{D}} Z \sim N(0, \rho^4 - \sigma^4).$$
 (5.26)

Beweis. Vgl. VAN DER VAART [63, Example 3.2].

Wählt man nun $a_n := \left(\frac{n_1 \cdot n_2}{n}\right)^{1/2}$, so gilt für das Konvergenzverhalten der Teststatitik $a_n \log(F_n)$ der folgende Satz.

Satz 5.11

Unter den Voraussetzungen des Modells (5.19) konvergiert die Teststatistik unter H_0 in Verteilung gegen eine normalverteilte Zufallsvariable X

$$a_n \log(F_n) \xrightarrow{\mathcal{D}} X \sim N\left(0, \frac{q\rho_1^4 + p\rho_2^4 - \sigma^4}{\sigma^4}\right).$$
 (5.27)

Beweis. Für die Restterme in (5.25) gilt nach dem Lemma von Slutzky

$$a_n \cdot R_2(\hat{\sigma}_{n_1}^2(X)) = \sqrt{\frac{n_2}{n}} \cdot \sqrt{n_1} \left(\hat{\sigma}_{n_1}^2(X) - \sigma_1^2 \right) \cdot \frac{R_2(\hat{\sigma}_{n_1}^2(X))}{(\hat{\sigma}_{n_1}^2(X) - \sigma_1^2)} \xrightarrow{P} 0, \quad (5.28)$$

weil der letzte Faktor nach dem starken Gesetz der großen Zahlen fast sicher gegen 0 konvergiert. Da dies auch für den zweiten Restterm $a_n \cdot R_2(\hat{\sigma}_{n_2}^2(Y))$ gilt, erhält man

$$a_n \log(F_n) = a_n \frac{\hat{\sigma}_{n_1}^2(X) - \hat{\sigma}_{n_2}^2(Y)}{\sigma^2} + o_P(1).$$
 (5.29)

Aus der Unabhängigkeit der beiden Stichproben folgt hierfür aus (5.26) zusammen mit dem Lemma von Slutzky die Behauptung

$$\sigma^{2} \cdot a_{n} \log(F_{n}) = \sqrt{\frac{n_{2}}{n}} \cdot \sqrt{n_{1}} \left(\hat{\sigma}_{n_{1}}^{2}(X) - \sigma^{2} \right)$$
$$-\sqrt{\frac{n_{1}}{n}} \cdot \sqrt{n_{2}} \left(\hat{\sigma}_{n_{2}}^{2}(Y) - \sigma^{2} \right) + o_{P}(1)$$
$$\xrightarrow{\mathcal{D}} N \left(0, q \rho_{1}^{4} + p \rho_{2}^{4} - \sigma^{4} \right).$$

Als nächstes wird die permutierte Version von $a_n \log(F_n)$ untersucht. Sei hierfür τ , wie in Beispiel 2.6, eine auf der symmetrischen Gruppe \mathcal{S}_n gleichverteilte Permutation auf einem anderen Wahrscheinlichkeitsraum $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$, welche bzgl. des Produktmaßes $P \otimes \tilde{P}$ unabhängig⁴ von den Daten ist. Setzt man $X_{n_1+i} := Y_i$ für $1 \leq i \leq n_2$, so definiert man für feste Beobachtungen $(X_i(\omega))_{i \leq n}$ die Permutationsstatistik durch $(\tau(i))_{i \leq n} \mapsto a_n \log F_n((X_{\tau(i)}(\omega))_{i \leq n})$. Der nächste Satz gibt Aufschluss über das zugehörige Konvergenzverhalten.

⁴Hierzu definiert man die Zufallsvariablen τ, X_i, Y_i wie in Kapitel 3 mittels Projektionen kanonisch auf dem gemeinsamen Raum $\Omega \times \tilde{\Omega}$.

Satz 5.12

Unter den Voraussetzungen des Modells (5.19) gilt für die Permuationsversion der Teststatistik unter H_0 ein bedingter Zentraler Grenzwertsatz

$$d_1\left(\mathcal{L}(a_n \log F_n((X_{\tau(i)})_{i \le n}) | X_1^2, \dots, X_n^2), N\left(0, \frac{p\rho_1^4 + q\rho_2^4 - \sigma^4}{\sigma^4}\right)\right) \stackrel{P}{\longrightarrow} 0. \quad (5.30)$$

Für diesen Beweis benötigt man noch den folgenden kleinen Hilfssatz.

Hilfssatz 5.13

Sei τ wie in Beispiel 2.6 eine auf S_n gleichverteilte Permutation. Dann erfüllen die Gewichte

$$W_{n,i} := c_{n,\tau(i)}, i \le n, \text{ mit } c_{n,i} := \sqrt{\frac{n_1 n}{n_2}} \begin{cases} 1/n_1 & i \le n_1 \\ 0 & i > n_1 \end{cases}$$
 (5.31)

die Generalvoraussetzungen (3.2)-(3.5) mit $Var(W_1) = 1$, falls $n_1/n \to p \in (0,1)$ für $n \to \infty$ gilt.

Beweis. Wegen $\overline{W}_n = \overline{c}_n = \sqrt{\frac{n_1}{nn_2}}$ ist die zweite Bedingung erfüllt

$$\max_{i \le n} \left| W_{n,i} - \overline{W}_n \right| = \max \left\{ \overline{c}_n, \sqrt{\frac{n_1 n}{n_2}} \left(\frac{1}{n_1} - \frac{1}{n} \right) \right\} \to 0.$$

Bedingung drei ergibt sich aus

$$\sum_{i=1}^{n} (W_{n,i} - \overline{W}_n)^2 = n_2 \overline{c}_n^2 + n_1 \frac{n_2}{nn_1} \to p + q = 1.$$

Aufgrund von

$$\tilde{P}(\sqrt{n}W_{n,1} = 0) = \frac{n_2}{n} \to q \text{ und } \tilde{P}(\sqrt{n}W_{n,1} = \frac{n}{\sqrt{n_1 n_2}}) = \frac{n_1}{n} \to p$$

liegt nach dem Lemma von Slutzky Verteilungskonvergenz

$$\sqrt{n}(W_{n,1} - \overline{W}_n) \xrightarrow{\mathcal{D}} \frac{1}{\sqrt{pq}} \mathbb{1}_A - \sqrt{\frac{p}{q}}$$

vor, wobei $\tilde{P}(A)=p$ gilt. Hieraus erhält man die Behauptung.

Hiermit lässt sich nun der Beweis von Satz 5.12 führen.

Beweis. Analog zu (5.28) wird zunächst gezeigt, dass auch die Permutationsversionen der Restterme in (5.25) von niederer Ordnung sind. Hierzu betrachtet man die folgende Entwicklung

$$\sqrt{n_1} \left(\frac{n_1 - 1}{n_1} \hat{\sigma}_{n_1}^2 ((X_{\tau(i)})_{i \le n}) - \sigma^2 \right)
= \sqrt{n_1} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} (X_{\tau(i)} - \frac{1}{n_1} \sum_{j=1}^{n_1} X_{\tau(j)})^2 - \sigma^2 \right)
= \sqrt{n_1} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} X_{\tau(i)}^2 - \left(\frac{1}{n_1} \sum_{j=1}^{n_1} X_{\tau(j)} \right)^2 - \sigma^2 \right)
= \sqrt{n_1} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} X_{\tau(i)}^2 - \frac{1}{n} \sum_{j=1}^{n} X_j^2 \right) + \sqrt{n_1} \left(\frac{1}{n} \sum_{j=1}^{n} X_j^2 - \sigma^2 \right)
- \sqrt{n_1} \left(\frac{1}{n_1} \sum_{j=1}^{n_1} X_{\tau(j)} \right)^2
=: A_n + B_n - C_n.$$

Nach Theorem 3.3 und Hilfssatz 5.13 gilt mit den Gewichten $W_{n,i}$ aus (5.31)

$$d_1\left(\mathcal{L}\left(\sqrt{\frac{n}{n_2}}\left(\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}X_{\tau(i)}^2 - \sqrt{n_1}\frac{1}{n}\sum_{j=1}^nX_j^2\right)|X_1^2,\dots,X_n^2\right), N\left(0,\hat{\sigma}^2\right)\right)$$
(5.32)
= $d_1\left(\mathcal{L}\left(\sum_{i=1}^nW_{n,i}\left(X_i^2 - \frac{1}{n}\sum_{j=1}^nX_j^2\right)\right)|X_1^2,\dots,X_n^2\right), N\left(0,\hat{\sigma}^2\right)\right) \stackrel{P}{\longrightarrow} 0$

für $\hat{\sigma}^2 = p\rho_1^4 + q\rho_2^4 - \sigma^4$. Somit konvergiert die bedingte Verteilung von A_n gegeben den Daten gegen eine zentrierte Normalverteilung. Genauso konvergiert B_n nach dem Zentralen Grenzwertsatz von Lindeberg-Feller in Verteilung gegen eine normalverteilte Zufallsvariable. Mulitpliziert man nun die Statistik aus Zeile (5.32) mit $\sqrt{\frac{n_2}{nn_1}}$, so folgt wegen $\overline{X}_n \stackrel{P}{\longrightarrow} 0$ nach dem (bedingten) Lemma von Slutzky, vgl. Lemma A.9 im Appendix, auch

$$\frac{1}{n_1} \sum_{i=1}^{n_1} X_{\tau(i)} \stackrel{P \otimes \tilde{P}}{\longrightarrow} 0. \tag{5.33}$$

Hiermit erhält man die asymptotische Vernachlässigbarkeit von C_n

$$C_n = \left(\frac{1}{n_1} \sum_{j=1}^{n_1} X_{\tau(j)}\right) \left(\sqrt{n_1} \left(\frac{1}{n_1} \sum_{j=1}^{n_1} X_{\tau(j)} - \overline{X}_n\right) + \sqrt{\frac{n_1}{n}} \sqrt{nX_n}\right) \stackrel{P \otimes \tilde{P}}{\longrightarrow} 0. \quad (5.34)$$

Zusammengefasst folgt

$$a_{n}R_{2}(\hat{\sigma}_{n_{1}}^{2}((X_{\tau(i)})_{i\leq n_{1}})$$

$$= \sqrt{\frac{n_{2}}{n}}\sqrt{n_{1}}(\hat{\sigma}_{n_{1}}^{2}((X_{\tau(i)})_{i\leq n_{1}}) - \sigma^{2})\frac{R_{2}(\hat{\sigma}_{n_{1}}^{2}((X_{\tau(i)})_{i\leq n_{1}})}{(\hat{\sigma}_{n_{1}}^{2}((X_{\tau(i)})_{i\leq n_{1}}) - \sigma^{2})} \xrightarrow{P \otimes \tilde{P}} 0$$

und somit eine Darstellung wie in (5.29)

$$\begin{split} \sigma^2 \ a_n \log(F_n(X_{\tau(i)})_{i \leq n}) &= a_n \Big(\frac{n_1}{n_1 - 1} \Big(\frac{1}{n_1} \sum_{i=1}^{n_1} X_{\tau(i)}^2 - \Big(\frac{1}{n_1} \sum_{j=1}^{n_1} X_{\tau(j)}^2\Big)\Big) \\ &- \frac{n_2}{n_2 - 1} \Big(\frac{1}{n_2} \sum_{i=n_1+1}^n X_{\tau(i)}^2 - \Big(\frac{1}{n_2} \sum_{j=n_1+1}^n X_{\tau(j)}^2\Big)\Big) \\ &+ o_{P \otimes \tilde{P}}(1) \\ &= \frac{n_1}{n_1 - 1} \frac{a_n}{n_1} \sum_{i=1}^{n_1} X_{\tau(i)}^2 - \frac{n_2}{n_2 - 1} \frac{a_n}{n_2} \sum_{i=n_1+1}^n X_{\tau(i)}^2 \\ &- \frac{n_1 \sigma^2}{n_1 - 1} \frac{a_n}{n_1} \Big(\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{X_{\tau(i)}}{\sigma}\Big)^2 \\ &+ \frac{n_2 \sigma^2}{n_2 - 1} \frac{a_n}{n_2} \Big(\frac{1}{\sqrt{n_2}} \sum_{i=n_1+1}^n \frac{X_{\tau(i)}}{\sigma}\Big)^2 + o_{P \otimes \tilde{P}}(1) \\ &= \frac{n_1}{n_1 - 1} \frac{a_n}{n_1} \sum_{i=1}^{n_1} X_{\tau(i)}^2 - \frac{n_2}{n_2 - 1} \frac{a_n}{n_2} \sum_{i=n_1+1}^n X_{\tau(i)}^2 \\ &+ o_{P \otimes \tilde{P}}(1). \end{split}$$

Demnach lässt sich die Permutationsversion von $a_n \log(F_n)$ mit Hilfe von

$$c_{n,i} = a_n \cdot \begin{cases} \frac{1}{n_1}, & \text{für } 1 \le i \le n_1, \\ -\frac{1}{n_2}, & \text{für } n_1 < i \le n \end{cases} \text{ und } X_{n,i} = \frac{1}{\sqrt{n}} \begin{cases} \frac{n_1}{n_1 - 1} X_i^2, & \text{für } 1 \le i \le n_1, \\ \frac{n_2}{n_2 - 1} Y_{i - n_1}^2, & \text{für } n_1 < i \le n. \end{cases}$$

als asymptotisch lineare Statistik schreiben

$$\sigma^2 a_n \log(F_n)((X_{\tau(i)})_{i \le n}) = \sqrt{n} \sum_{i=1}^n c_{n,i} X_{n,\tau(i)} + o_{P \otimes \tilde{P}}(1).$$
 (5.35)

Definiert man nun Permutationsgewichte $W_{n,i} := c_{n,\tau(i)}$, so erfüllen diese nach Beispiel 2.6 die Generalvoraussetzungen mit $\text{Var}(W_1) = 1$, so dass für die bedingte Konvergenz der Permutationsstatistik $T_n^* := \sqrt{n} \sum_{i=1}^n W_{n,i} X_{n,i}$ aufgrund

von Theorem 3.3 nur noch die zwei Bedingungen (3.6) und (3.7) nachgewiesen werden müssen. Dabei ist die Maximumsbedingung (3.6) erfüllt, da $\rho_1^4, \rho_2^4 < \infty$ gilt, so dass man dies aus Bemerkung 3.8 erhält

$$\max_{i \le n} |X_{n,i}| \le \max_{i \le n} |X_{n,i}| + \max_{n_1 < i \le n} |X_{n,i}| \xrightarrow{P} 0.$$

Es fehlt also nur noch die Bestimmung der Varianz der Grenzverteilung. Diese ist durch die Konvergenz von

$$\sum_{i=1}^{n} (X_{n,i} - \overline{X}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i\right)^2$$

gegeben, wobei $\tilde{X}_i := \sqrt{n} X_{n,i}$ gilt. Aus dem starken Gesetz der großen Zahlen erhält man für die beiden Summen die Konvergenzen

$$\frac{1}{n} \sum_{i=1}^{n} \tilde{X}_{i} = \frac{n_{1}}{n_{1} - 1} \frac{n_{1}}{n} \left(\frac{1}{n_{1}} \sum_{i=1}^{n_{1}} X_{i}^{2} \right) + \frac{n_{2}}{n_{2} - 1} \frac{n_{2}}{n} \left(\frac{1}{n_{2}} \sum_{i=1}^{n_{2}} Y_{i}^{2} \right) \xrightarrow{f.s.} p\sigma^{2} + q\sigma^{2} = \sigma^{2}$$

sowie

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{i}^{2} = \frac{n_{1}}{n_{1}-1}\frac{n_{1}}{n}\left(\frac{1}{n_{1}}\sum_{i=1}^{n_{1}}X_{i}^{4}\right) + \frac{n_{2}}{n_{2}-1}\frac{n_{2}}{n}\left(\frac{1}{n_{2}}\sum_{i=1}^{n_{2}}Y_{i}^{4}\right) \xrightarrow{f.s.} p\rho_{1}^{4} + q\rho_{2}^{4}.$$

Der bedingte Zentrale Grenzwertsatz 3.3 liefert somit unter H_0

$$d_1\left(\mathcal{L}(T_n^*|X_{n,1},\dots,X_{n,n}),N\left(0,p\rho_1^4+q\rho_2^4-\sigma^4\right)\right) \xrightarrow{P} 0,$$
 (5.36)

so dass die Behauptung aufgrund der Darstellung (5.35) aus dem bedingten Lemma von Slutzky folgt.

Bemerkung 5.14

Im Teilmodell (5.20) kann man den obigen Beweis mit Hilfe von Lemma 4.5 stark verkürzen, da die beiden Stichproben dann austauschbar wären. □

Damit also die Grenzverteilung von $a_n \log(F_n)$, vgl. (5.27), und der zugehörigen Permutationsversion, vgl. (5.30), unter H_0 übereinstimmen, benötigt man entweder einen balancierten Stichprobenumfang, d.h. p = q = 1/2, oder identische vierte Momente $\rho_1^4 = \rho_2^4$. Die zweite Bedingung ist z.B. in der Situation des erweiterten Behrens-Fisher-Modells (5.20) erfüllt.

Da diese Einschränkungen aber nicht unbedingt wünschenswert sind, wird im Folgenden gezeigt, wie dies durch eine geeignete Studentisierung behoben werden kann. Dabei geht die Idee auf JANSSEN [29] zurück, der im Behrens-Fisher-Modell (5.20) mittels einem studentisiertem Permutationstest auf Unterschiede in den Erwartungswerten testet.

Für die weiteren Betrachtungen liegt jedoch wieder das allgemeinere Modell (5.19) zu Grunde.

Dabei definiert man für vierte Momentschätzer $\hat{\rho}_1^4 := \frac{1}{n_1} \sum_{i=1}^{n_1} X_i^4$ und $\hat{\rho}_2^4 := \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i^4$ die Statistik

$$\tilde{V}_n^2 := a_n^2 \left(\frac{1}{n_1} \hat{\rho}_1^4 + \frac{1}{n_2} \hat{\rho}_2^4 \right) - \left(a_n^2 \left(\frac{1}{n_1} \hat{\sigma}_{n_1}^2 + \frac{1}{n_2} \hat{\sigma}_{n_2}^2 \right) \right)^2. \tag{5.37}$$

Diese konvergiert (unter H_0) fast sicher gegen $q\rho_1^4 + p\rho_2^4 - \sigma^4$. Als geeignete Studentisierung kann demnach

$$V_n^2 := \frac{\tilde{V}_n^2}{\left(a_n^2 \left(\frac{1}{n_1} \hat{\sigma}_{n_1}^2 + \frac{1}{n_2} \hat{\sigma}_{n_2}^2\right)\right)^2}$$
 (5.38)

verwendet werden, da diese Statisitk unter H_0 gerade gegen die Varianz der Grenzverteilung von $a_n \log(F_n)$ konvergiert, vgl. (5.27). Somit liegt nach dem Lemma von Slutzky Verteilungskonvergenz

$$\frac{a_n \log(F_n)}{V_n} \xrightarrow{\mathcal{D}} X_0 \sim N(0, 1) \tag{5.39}$$

vor. Dies gilt unter denselben Voraussetzungen auch für die Permutationsversion von $a_n \log(F_n)/V_n$.

Satz 5.15

Unter den Voraussetzungen des Modells (5.19) gilt unter H_0 ein bedingter Zentraler Grenzwertsatz

$$d_1(\mathcal{L}(\left(\frac{a_n \log(F_n)}{V_n}\right)((X_{\tau(i)})_{i \le n})|X_1^2, \dots, X_n^2), N(0, 1)) \xrightarrow{P} 0.$$
 (5.40)

Beweis. Zeigt man, dass unter H_0

$$V_n^2((X_{\tau(i)})_{i \le n}) \xrightarrow{P \otimes \tilde{P}} \frac{p\rho_1^4 + q\rho_2^4 - \sigma^4}{\sigma^4}$$

$$\tag{5.41}$$

gilt, so folgt hieraus nach dem bedingten Lemma von Slutzky und Satz 5.12 gerade die Behauptung. Hierfür teilt man zunächst die Statistik \tilde{V}_n^2 wie folgt in

zwei Teile auf

$$\begin{split} \tilde{V}_n^2(((X_{\tau(i)})_{i \leq n}) &= \frac{n_2}{n} \hat{\rho}_1^4((X_{\tau(i)})_{i \leq n}) + \frac{n_1}{n} \hat{\rho}_2^4((X_{\tau(i)})_{i \leq n}) \\ &- \left(\frac{n_2}{n} \hat{\sigma}_{n_1}^2((X_{\tau(i)})_{i \leq n}) + \frac{n_1}{n} \hat{\sigma}_{n_2}^2((X_{\tau(i)})_{i \leq n})\right)^2. \end{split}$$

Schreibt man den ersten Teil analog zu JANSSEN [29] mittels Regressionskoeffizienten

$$d_{n,i} := \begin{cases} n_2/n_1 n & i \le n_1 \\ n_1/n_2 n & i > n_1 \end{cases}$$

als $\sum_{i=1}^n d_{n,i} X_{\tau(i)}^4$ um, so erhält man hierfür wegen $\sum_{i=1}^n d_{n,i} = 1$.

$$E\left(\sum_{i=1}^{n} d_{n,i} X_{\tau(i)}^{4} | X_{1}, \dots, X_{n}\right) = \sum_{i=1}^{n} d_{n,i} \left(\frac{n_{1}}{n} \frac{1}{n_{1}} \sum_{j=1}^{n_{1}} X_{j}^{4} + \frac{n_{2}}{n} \frac{1}{n_{2}} \sum_{k=1}^{n_{2}} Y_{k}^{4}\right)$$

$$\xrightarrow{f.s.} p\rho_{1}^{4} + q\rho_{2}^{4}. \tag{5.42}$$

Falls nun die achten Momente $E(X_1^8) + E(Y_1^8) < \infty$ exisiteren, so folgt aufgrund von $\sum_{i=1}^n (d_{n,i} - \overline{d}_n)^2 \to 0$ (vgl. HÁJEK ET. AL. [22, S.61, Theorem 3])

$$\operatorname{Var}\left(\sum_{i=1}^{n} d_{n,i} X_{\tau(i)}^{4} | X_{1}, \dots, X_{n}\right) = \sum_{k=1}^{n} (d_{n,k} - \overline{d}_{n})^{2} \frac{1}{n-1} \sum_{i=1}^{n} (X_{i}^{4} - \frac{1}{n} \sum_{j=1}^{n} X_{j}^{4})^{2}$$

$$\xrightarrow{f.s.} 0$$

und hieraus

$$\sum_{i=1}^{n} d_{n,i} X_{\tau(i)}^4 \xrightarrow{P \otimes \tilde{P}} p\rho_1^4 + q\rho_2^4. \tag{5.43}$$

Falls die achten Momente nicht existieren, definiert man abgeschnittene Zufallsvariablen $X_{i,k} := X_i \mathbb{1}_{[-k,k]}(X_i)$ und $Z_{i,k} := X_i - X_{i,k}$. Aus (5.43) erhält man so für $n \to \infty$

$$\sum_{i=1}^{n} d_{n,i} X_{\tau(i),k}^{4} \xrightarrow{P \otimes \tilde{P}} \rho_{k}^{4} := p \operatorname{E}(X_{1}^{4} \mathbb{1}_{[-k,k]}(X_{1})) + q \operatorname{E}(Y_{1}^{4} \mathbb{1}_{[-k,k]}(Y_{1})).$$

Aufgrund der Konvergenz

$$\frac{1}{\epsilon} \limsup_{n \to \infty} E(\sum_{i=1}^{n} d_{n,i} Z_{\tau(i),k}^{4}) = \frac{1}{\epsilon} (p\rho_{1}^{4} + q\rho_{2}^{4} - \rho_{k}^{4}) \to 0$$

für $k \to \infty$ folgt (5.43) mit Hilfe der Markoff Ungleichung nach Theorem 4.2. von BILLINGSLEY [6]. Aufgrund von $\hat{\sigma}_{n_1} = \frac{n_1}{n_1-1}(\frac{1}{n_1}\sum_{i=1}^{n_1}X_i^2 - \overline{X}_{n_1}^2)$, kann man die $P \otimes \tilde{P}$ - Konvergenz von $\frac{n_2}{n}\hat{\sigma}_{n_1}^2((X_{\tau(i)})_{i \le n}) + \frac{n_1}{n}\hat{\sigma}_{n_2}^2((X_{\tau(i)})_{i \le n})$ gegen σ^2 analog zeigen und erhält somit gerade (5.41), d.h. die Behauptung.

Satz 5.15 liefert somit (unter dem Rand der Hypothesen) die asymptotische Äquivalenz des unbedingten asymptotischen α -Niveau-Tests $\varphi_n = \mathbb{1}_{(u_{1-\alpha},\infty)}(a_n \log(F_n) \mathbb{1}_{\{V_n^2>0\}}/V_n)$ und des Permutationstests (5.22)

$$\varphi_{n,PF}^* = \begin{cases} 1 & \text{,falls } \frac{a_n \log(F_n) \mathbb{1}_{\{V_n^2 > 0\}}}{V_n} > c_{n,PF}^*(\alpha). \end{cases}$$

Der Permutationstest besitzt aber wieder den Vorteil, dass er bei identisch verteilten Stichproben sogar exakt das Niveau hält.

Das asymptotische Verhalten der Gütefunktion wird durch den nächsten Satz erschlossen.

Satz 5.16

 $Sei \ r^2 := \sigma_1^2/\sigma_2^2 \ das \ Varianzverhältnis \ der \ beiden \ Gruppen, \ dann \ gilt$

(i)
$$E(\varphi_{n,PF}) \to 1$$
, falls $r > 1$ und

(ii)
$$E(\varphi_{n,PF}) \to 0$$
, falls $r < 1$.

Beweis. Es liegen die folgenden fast sicheren Konvergenzen vor

$$F_n \xrightarrow{f.s.} r^2$$

$$V_n^2 \xrightarrow{f.s.} \frac{q\rho_1^4 + p\rho_2^4 - (q\sigma_1^2 + p\sigma_2^2)^2}{(q\sigma_1^2 + p\sigma_2^2)^2}$$

$$= \frac{q^2 \operatorname{Var}(X_1^2) + p^2 \operatorname{Var}(Y_1^2) + qp \operatorname{E}((X_1^2 - Y_1^2)^2)}{(q\sigma_1^2 + p\sigma_2^2)^2} =: \varsigma^2.$$

Da E $((X_1^2-Y_1^2)^2)$ nach der Jensenschen Ungleichung positiv ist, erhält man $\varsigma>0$ und somit

$$\frac{\log(F_n)}{V_n} \xrightarrow{f.s.} \frac{2\log(r)}{\varsigma} \begin{cases} > 0 \\ < 0 \end{cases}, \text{falls } r > 1. \tag{5.44}$$

Analog zum Beweis von Satz 5.15 zeigt man auch

$$V_n^2((X_{\tau(i)})_{i \le n}) \xrightarrow{P \otimes \tilde{P}} \frac{p\rho_1^4 + q\rho_2^4 - (p\sigma_1^2 + q\sigma_2^2)^2}{(p\sigma_1^2 + q\sigma_2^2)^2}$$

$$= \frac{p^2 \operatorname{Var}(X_1^2) + q^2 \operatorname{Var}(Y_1^2) + pq \operatorname{E}((X_1^2 - Y_1^2)^2)}{(p\sigma_1^2 + q\sigma_2^2)^2} =: \tilde{\varsigma}^2$$

sowie

$$\log(F_n((X_{\tau(i)})_{i \le n})) \xrightarrow{P \otimes \tilde{P}} \log\left(\frac{p\sigma_1^2 + q\sigma_2^2}{p\sigma_1^2 + q\sigma_2^2}\right) = \log(1) = 0.$$

Weil auch hier nach der Jensenschen Ungleichung $\tilde{\zeta} > 0$ gilt, folgt die Behauptung mit Hilfe von (5.44).

Das nächste Zweistichprobenproblem befasst sich mit Testen auf Unterschiede in den Kovarianzen. Analog zur obigen Konstruktion erhält man aus dem abschliessenden Satz 5.19 einen studentisierten Permutationstest, der sich wieder ohne asysmptotischen Güteverlust ausführen lässt.

Permutationstest zum Vergleich zweier Kovarianzen

Analog zum oben behandelten Einstichprobenfall, werden in diesem Anwendungsbeispiel zwei voneinander unabhängige \mathbb{R}^2 -wertige Stichproben

$$X_1, \dots, X_{n_1}$$
 i.i.d. P^{X_1} , (5.45)
 Y_1, \dots, Y_{n_2} i.i.d. P^{Y_1}

auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit unbekannten Erwartungswerten $E(X_1) = \mu_1, E(Y_1) = \mu_2 \in \mathbb{R}^2$ und unbekannten Kovarianzmatrizen

$$\operatorname{Cov}(X_1) = \begin{pmatrix} \sigma_{11}^2 & c_1 \\ c_1 & \sigma_{12}^2 \end{pmatrix}, \operatorname{Cov}(Y_1) = \begin{pmatrix} \sigma_{21}^2 & c_2 \\ c_2 & \sigma_{22}^2 \end{pmatrix}$$

mit $\sigma_{ij}^2>0$ für $1\leq i,j\leq 2$ betrachtet. Von Interesse ist nun das Testproblem identischer Kovarianzen⁵

$$H: \{c_1 = c_2\} \text{ gegen } K: \{c_1 > c_2\}.$$

Als mögliche Teststatistik kommt

$$T_n := \sqrt{\frac{n_1 n_2}{n}} (\hat{C}_1(X) - \hat{C}_2(Y))$$

in Frage, wobei

$$\hat{C}_1 := \hat{C}_1(X) := \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i^{(1)} - \overline{X}_{n_1}^{(1)}) (X_i^{(2)} - \overline{X}_{n_1}^{(2)})$$

die empirische Kovarianz der ersten Stichprobe ist. Analog definiert man $\hat{C}_2:=\hat{C}_2(Y)$. Da die Teststatistik offensichtlich unabhängig vom Erwartungswert der

⁵Das zweiseitge Testproblem mit der Alternative $\tilde{K}:\{c_1\neq c_2\}$ erhält man wieder durch Betrachtung des Betrages der Teststatistik.

einzelnen Koordinaten ist, wird im Weiteren ohne Einschränkung angenommen, dass $E(X_1) = E(Y_1) = 0$ gilt.

Satz 5.17

Existieren in den einzelnen Gruppen vierte Momente $E(||X_1||^4) + E(||Y_1||^4) < \infty$, so konvergiert die Teststatistik unter H in Verteilung gegen eine normalverteilte Zufallsvariable X

$$T_n \xrightarrow{\mathcal{D}} X \sim N\left(0, q \operatorname{Var}(X_1^{(1)} X_1^{(2)}) + p \operatorname{Var}(Y_1^{(1)} Y_1^{(2)})\right).$$
 (5.46)

Beweis. Da der zweite Term auf der rechten Seite der Gleichung

$$\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} (X_i^{(1)} - \overline{X}_{n_1}^{(1)})(X_i^{(2)} - \overline{X}_{n_1}^{(2)}) = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} X_i^{(1)} X_i^{(2)} - \sqrt{n_1} \overline{X}_{n_1}^{(1)} \overline{X}_{n_1}^{(2)}$$

nach Slutzky stochastisch gegen 0 konvergiert, folgt

$$T_{n} = \sqrt{\frac{n_{2}}{n}} \frac{1}{\sqrt{n_{1}}} \sum_{i=1}^{n_{1}} X_{i}^{(1)} X_{i}^{(2)} - \sqrt{\frac{n_{1}}{n}} \frac{1}{\sqrt{n_{2}}} \sum_{i=1}^{n_{2}} Y_{i}^{(1)} Y_{i}^{(2)} + o_{P}(1) \quad (5.47)$$

$$= \sqrt{\frac{n_{2}}{n}} \left(\frac{1}{\sqrt{n_{1}}} \sum_{i=1}^{n_{1}} X_{i}^{(1)} X_{i}^{(2)} - \sqrt{n_{1}} c_{1} \right)$$

$$-\sqrt{\frac{n_{1}}{n}} \left(\frac{1}{\sqrt{n_{2}}} \sum_{i=1}^{n_{2}} Y_{i}^{(1)} Y_{i}^{(2)} - \sqrt{n_{2}} c_{2} \right) + o_{P}(1).$$

Aufgrund der Unabhängigkeit der beiden Stichproben erhält man die Behauptung aus dem Zentralen Grenzwertsatz.

Wie schon beim obigen F-Test wird als Nächstes die permutierte Version von T_n untersucht. Dabei sei τ wieder die auf der symmetrischen Gruppe S_n gleichverteilte Permutation, $X_{n_1+i} := Y_i$ für $1 \le i \le n_2$ und $X_{n,i} := \frac{1}{\sqrt{n}} X_i^{(1)} X_i^{(2)}$.

Satz 5.18

Unter den Voraussetzungen von Satz 5.17 gilt für die Permuationsversion der Teststatistik unter H ein bedingter Zentraler Grenzwertsatz

$$d_1\Big(\mathcal{L}(T_n((X_{\tau(i)})_{i\leq n})|X_{n,1},\ldots,X_{n,n}),N(0,p\tilde{\sigma}_1^2+q\tilde{\sigma}_2^2)\Big)\stackrel{P}{\longrightarrow} 0.$$

Dabei gilt für die Grenzvarianz $\tilde{\sigma}_1^2 := \text{Var}(X_1^{(1)}X_1^{(2)}) \ und \ \tilde{\sigma}_2^2 := \text{Var}(Y_1^{(1)}Y_1^{(2)})$

Beweis. Aufgrund von $\mu_i = 0$ für i = 1, 2 erhält man wie in (5.34) mit dem Lemma von Slutzky

$$\left(\sqrt{n_1} \frac{1}{n_1} \sum_{j=1}^{n_1} X_{\tau(j)}^{(1)}\right) \left(\frac{1}{n_1} \sum_{i=1}^{n_1} X_{\tau(i)}^{(2)}\right) \stackrel{P \otimes \tilde{P}}{\longrightarrow} 0.$$

Somit besitzt die Permutationsversion von T_n eine Entwicklung der Form (5.47), d.h. es gilt

$$T_{n}((X_{\tau(i)})_{i \leq n}) = \sqrt{\frac{n_{2}}{n}} \frac{1}{\sqrt{n_{1}}} \sum_{i=1}^{n_{1}} X_{\tau(i)}^{(1)} X_{\tau(i)}^{(2)} - \sqrt{\frac{n_{1}}{n}} \frac{1}{\sqrt{n_{2}}} \sum_{i=1}^{n_{2}} X_{\tau(i)}^{(1)} X_{\tau(i)}^{(2)} + o_{P \otimes \tilde{P}}(1)$$

$$= \sqrt{n} \sum_{i=1}^{n} c_{n,\tau(i)} X_{n,i} + o_{P \otimes \tilde{P}}(1).$$

Dabei sind $(c_{n,i})_{i\leq n}$ die Regressionskoeffizienten des Zweistichprobenproblems, vgl. (2.9). Da das Dreiecksschema die Maximumsbedingung (3.6) erfüllt und unter H die Konvergenz

$$\sum_{i=1}^{n} (X_{n,i} - \overline{X}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i^{(1)} X_i^{(2)})^2 - \left(\frac{1}{n} \sum_{i=1}^{n} X_i^{(1)} X_i^{(2)}\right)^2$$

$$\xrightarrow{P} p \operatorname{E}((X_1^{(1)} X_1^{(2)})^2) + q \operatorname{E}((Y_1^{(1)} Y_1^{(2)})^2)$$

$$- \left(p \operatorname{E}(X_1^{(1)} X_1^{(2)}) + q \operatorname{E}(Y_1^{(1)} Y_1^{(2)})\right)^2$$

$$= p \operatorname{Var}(X_1^{(1)} X_1^{(2)}) + q \operatorname{Var}(Y_1^{(1)} Y_1^{(2)})$$

vorliegt, folgt die Behauptung somit aus Satz 3.3.

Damit also die Grenzverteilung von T_n und der zugehörigen Permutationsversion unter H übereinstimmen, benötigt man auch in diesem Fall entweder einen balancierten Stichprobenumfang, d.h. p=q=1/2, oder identische Kovarianzen. Um dies zu umgehen, kann man die Statistik analog zu oben mit Hilfe von

$$V_n^2 := a_n^2 \left(\frac{1}{n_1} \hat{\sigma}_{n_1}^2 ((X_i^{(1)} X_i^{(2)})_{i \le n_1}) + \frac{1}{n_2} \hat{\sigma}_{n_2}^2 ((Y_i^{(1)} Y_i^{(2)})_{i \le n_2}) \right)$$

studentisieren. Dabei ist $\hat{\sigma}_n((z_i)_{i\leq n})$ wieder die empirische Standardabweichung der Daten z_1, \ldots, z_n .

Mit dieser Studentisierung konvergiert T_n/V_n in Verteilung gegen eine standardnormalverteilte Zufallsvariable, so dass $\varphi_n = \mathbb{1}_{(u_{1-\alpha},\infty)}(\mathbb{1}_{\{V_n^2>0\}}T_n/V_n)$ ein asymptotischer α -Niveau-Test ist. Dasselbe gilt auch für die permutierte Version.

Satz 5.19

Unter den Voraussetzungen von Satz 5.17 gilt unter H_0

$$d_1\left(\mathcal{L}(\frac{T_n}{V_n}((X_{\tau(i)})_{i\leq n})|X_1,\dots,X_n),N(0,1)\right) \stackrel{P}{\longrightarrow} 0.$$
 (5.48)

Beweis. Analog zum Beweis von Satz 5.15 zeigt man

$$V_n^2((X_{\tau(i)})_{i \le n}) \xrightarrow{P \otimes \tilde{P}} p \operatorname{Var}(X_1^{(1)} X_1^{(2)}) + q \operatorname{Var}(Y_1^{(1)} Y_1^{(2)}),$$

so dass die Behauptung mit Hilfe des Lemmas von Slutzky aus Satz 5.18 folgt.

Somit ist der Permutationstest in der Teststatistik T_n/V_n asymptotisch effektiv im Bezug auf φ_n .

Bemerkung 5.20

- 1. Auf das Verhalten der Gütefunktion kann man durch Diskussion der Gleichung (5.47) auf die gleiche Weise schließen wie beim Permutations-F-Test.
- 2. Kombiniert man die Aussagen aus den beiden obigen Anwendungen, so kann man wiederum die asymptotische Permutationsverteilung der Differenz der empirischen Korrelationen der beiden Gruppen bestimmen und so einen konsistenten Permutationstest zum Korrelationsvergleich konstruieren. Da die hierzugehörigen Rechnungen jedoch etwas länger sind und keine neuen Beweistechniken verwenden, wird dieses Beispiel an dieser Stelle nicht ausgeführt, sondern erst in einer zukünftigen Arbeit vorgestellt.

Zweistichproben-Permutationstest in asymptotisch effizienten Schätzern

In diesem Beispiel wird die Theorie aus Kapitel 4.1 angewendet.

Sei dazu zunächst $\kappa_0: \mathcal{M}_1(\Omega, \mathcal{A}) \to \mathbb{R}$ ein Funktional, das im Einstichprobenfall von Interesse ist. Weiter wird angenommen, dass hierfür ein asymptotisch effizienter Schätzer $\hat{\kappa}_n := \hat{\kappa}_n(X_1, \dots, X_n)$ in unabhängigen Zufallsvariablen X_1, \dots, X_n mit identischer Verteilung P existiere. Dabei bedeutet asymptotisch effizient, dass für den Schätzer die folgende, asymptotisch lineare Entwicklung

vorliegt

$$\sqrt{n}(\hat{\kappa}_n - \kappa_0(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{g}(X_i) + o_P(1).$$
 (5.49)

Dabei ist \tilde{g} eine messbare Funktion mit $E(\tilde{g}(X_1)) = 0$ und $Var(\tilde{g}(X_1)) = \sigma^2 \in (0, \infty)$. Dieses \tilde{g} heisst auch effektive score-Funktion oder kanonischer Gradient. Unter geeigneten Regularitätsvoraussetzungen gelten solche Entwicklungen beispielsweise für Maximum-Likelihood-Schätzer, vgl. VAN DER VAART [63, Theorem 5.39].

Diese schöne Eigenschaft effizienter Schätzer wird nun für die Entwicklung eines Permutationstest für Zweistichprobenprobleme verwendet. Um dabei die Schreibweise aus Kapitel 4.1 zu verwenden, sei für $n = n_1 + n_2$

$$\mathcal{P}_n = \{ P^{n_1} \otimes Q^{n_2} : P, Q \in \mathcal{M}_1(\mathbb{R}, \mathcal{B}) \}$$
 (5.50)

die Menge der Verteilungen für das Zweistichprobenproblem. In diesem Fall ist das Funktional

$$\kappa_n(P^{n_1} \otimes Q^{n_2}) := \kappa_0(P) - \kappa_0(Q)$$

von Interesse und typische Teststatistiken für die Hypothese

$$H: \{P=Q\}$$

sind von der Form

$$T_n := \hat{\kappa}_{n_1}(X_1, \dots, X_{n_1}) - \hat{\kappa}_{n_2}(X_{n_1+1}, \dots, X_n), \tag{5.51}$$

wobei $(X_i)_{i\leq n}$ die kanonischen Projektionen auf $(\Omega^n, \mathcal{A}^n, P^{n_1} \otimes Q^{n_2})$ sind. Hierfür gilt unter der Voraussetzung (5.49) mit der Normierungskonstante $a_n = \sqrt{\frac{n_1 n_2}{n}}$ unter H das Folgende

$$a_n T_n = a_n \left((\hat{\kappa}_{n_1} - \kappa_0(P)) - (\hat{\kappa}_{n_2} - \kappa_0(Q)) \right)$$
$$= a_n \left(\frac{1}{n_1} \sum_{i=1}^{n_1} \tilde{g}(X_i) + \frac{o_P(1)}{\sqrt{n_1}} - \frac{1}{n_2} \sum_{i=n_1+1}^{n} \tilde{g}(X_i) + \frac{o_P(1)}{\sqrt{n_2}} \right).$$

Wegen

$$\left(\frac{a_n}{\sqrt{n_1}} - \frac{a_n}{\sqrt{n_2}}\right) o_P(1) = \left(\sqrt{\frac{n_2}{n}} - \sqrt{\frac{n_1}{n}}\right) o_P(1) = o_P(1)$$

erhält man nun, dass die obige Teststatistik bis auf einen kleinen Fehler unter H die folgende Darstellung als lineare Statistik besitzt

$$a_n T_n = \sqrt{n} \sum_{i=1}^n c_{n,i} X_{n,i} + o_P(1).$$
 (5.52)

Dabei sind $(c_{n,i})_i$ die Regressionskoeffizienten des Zweistichprobenproblems, vgl. (2.9), und $X_{n,i} := \tilde{g}(X_i)/\sqrt{n}$ ein zeilenweise unabhängiges Dreiecksschema. Ausserdem liegt aufgrund von (5.49) und der Unabhängigkeit der beiden Stich-

Ausserdem liegt aufgrund von (5.49) und der Unabhängigkeit der beiden Stichproben Verteilungskonvergenz gegen eine Normalverteilung vor

$$a_n T_n \xrightarrow{\mathcal{D}} Z \sim N(0, (p+q)\sigma^2) = N(0, \sigma^2).$$
 (5.53)

Hieraus erhält man den asymptotischen α -Niveau Test

$$\varphi_{n,PM} = \begin{cases} 1 & \text{, falls } a_n T_n > \\ 0 & \leq \end{cases}$$
 (5.54)

dessen Ausführung jedoch die Kenntnis der Varianz σ^2 voraussetzt.

Es wird nun wieder gezeigt, dass die Permutationsversion der obigen Teststatistik dieselbe Grenzverteilung besitzt. Dabei verwendet man - wie in den beiden vorangegangenen Anwendungen - die auf der symmetrischen Gruppe S_n gleichverteilte Permutation τ .

Satz 5.21

Für die Permuationsversion der Teststatistik a_nT_n gilt unter H ein bedingter Zentraler Grenzwertsatz

$$d_1\Big(\mathcal{L}\big(a_nT_n((X_{\tau(i)})_{i\leq n})|\tilde{g}(X_1),\ldots,\tilde{g}(X_n)\big),N(0,\sigma^2)\Big) \stackrel{P}{\longrightarrow} 0.$$
 (5.55)

Beweis. Es werden zunächst wieder die Permutationsgewichte $W_{n,i} := c_{n,\tau(i)}$ definiert. Da diese die Generalvoraussetzungen erfüllen und das Dreiecksschema unter H zeilenweise i.i.d. ist, folgt die Verteilungskonvergenz

$$d_1\left(\mathcal{L}(T_n^*|X_{n,1},\dots,X_{n,n}),N(0,\sigma^2)\right) \xrightarrow{P} 0 \tag{5.56}$$

der Permutationsstatistik $T_n^* := \sqrt{n} \sum_{i=1}^n W_{n,i} X_{n,i} \stackrel{\mathcal{D}}{=} \sqrt{n} \sum_{i=1}^n c_{n,i} X_{n,\tau(i)}$ wegen $Var(\tilde{g}(X_1)) = \sigma^2 < \infty$ aus Korollar 3.7 bzw. Bemerkung 3.8. Weil die Daten unter H insbesondere austauschbar sind, ergibt sich so aufgrund von Lemma 4.5 bzw. Lemma A.7 und der Darstellung (5.52) die Behauptung.

Dies führt auf den nach Lemma 4.4 zu $\varphi_{n,PM}$ asymptotisch äquivalenten Permutationstest

$$\varphi_{n,PM}^* = \begin{cases} 1 & \text{, falls } a_n T_n > c_{n,PM}^*(\alpha), \\ 0 & \leq \end{cases}$$
 (5.57)

wobei $c_{n,PM}^*(\alpha)$ das bedingte $(1-\alpha)$ -Quantil der Permutationsverteilung der Teststatistik ist.

In vielen Fällen erhält man Entwicklungen der Form (5.49) entlang von L_2 differenzierbaren Kurven $(P_\theta)_{\theta\in\Theta}$, vgl. wieder VAN DER VAART [63, Theorem
5.39] für Maximum-Likelihood-Schätzer. Aus diesem Grunde wird für die folgenden Betrachtungen der Gütefunktion vorausgesetzt, dass $(P_\theta)_{\theta\in\Theta}$ eine in $\theta_0\in\stackrel{\circ}{\Theta}L^2$ -differenzierbare Familie mit score-Funktion g und Fisher-Information $Var_{\theta_0}(g)=I_0$ ist. Die Darstellung in (5.49) schreibt man dann häufig wie folgt
um

$$\sqrt{n}(\hat{\kappa}_n - \kappa_0(P_{\theta_0})) = \frac{1}{\sqrt{n}I_0} \sum_{i=1}^n g(X_i) + o_{P_{\theta_0}}(1).$$
 (5.58)

Unter diesen Voraussetzungen gibt der folgende Satz Auskunft über das Verhalten der obigen Tests unter lokalen Alternativen⁶

$$P_{n,\theta} := \bigotimes_{i=1}^{n} P_{\theta_0 + d_{ni}\theta} = \mathcal{L}((X_i)_{i \le n}) \text{ mit } d_{ni} := \begin{cases} d_n, & \text{für } 1 \le i \le n_1, \\ \tilde{d}_n, & \text{für } n_1 < i \le n \end{cases}$$
(5.59)

und $d_n, \tilde{d}_n \in \mathbb{R}$.

Satz 5.22

Erfüllen die Koeffizienten in (5.59) die Bedingungen

$$\max_{i \le n} |d_{ni}| \to 0, \quad \sum_{i=1}^{n} d_{ni}^2 \to c^2 \in \mathbb{R}_{>0}$$
 (5.60)

 $mit \Theta_n \subset \{\theta : d_n\theta, \tilde{d}_n\theta \in \Theta\}, \Theta_n \uparrow \mathbb{R}, \text{ so gelten die folgenden Aussagen:}$

a)
$$E_{P_{n,\theta}}(|\varphi_{n,PM} - \varphi_{n,PM}^*|) \longrightarrow 0.$$

b)
$$\lim_{n\to\infty} \mathbb{E}_{P_{n,\theta}}(\varphi_{n,PM}^*) = \lim_{n\to\infty} \mathbb{E}_{P_{n,\theta}}(\varphi_{n,PM}) = \Phi(\theta \frac{\sigma_{12}}{\sigma_1} - u_{1-\alpha}).$$

c)
$$ARE(\varphi_{n,PM}) = ARE(\varphi_{n,PM}^*) = I_0^2 \left(\frac{\sigma_{12}}{c}\right)^2$$
.

d) Für Alternativen in den Koeffizienten
$$d_{ni} = b_{ni} := a_n I_0^{-1}(\frac{1}{n_1} \mathbb{1}_{\{1,...,n_1\}}(i) + \frac{1}{n_2} \mathbb{1}_{\{n_1+1,...,n\}}(i))$$
 gilt $ARE(\varphi_{n,PM}) = 1$.

Dabei gilt
$$\sigma_{12} = \sum_{i=1}^{\infty} d_{ni}b_{ni}$$
 und $\sigma_1^2 = c^2I_0$.

⁶Eine allgemeinere Wahl ist hier nicht möglich, da die Entwicklung (5.58) in den beiden Stichproben nur funktioniert, falls diese jeweils i.i.d. sind.

Beweis. Aufgrund von Satz 5.21 sind die beiden Tests nach Lemma 4.4 asymptotisch äquivalent, d.h. es gilt

$$\operatorname{E}_{P_{\theta_0}^n}(\left|\varphi_{n,PM} - \varphi_{n,PM}^*\right|) \to 0. \tag{5.61}$$

Da $(P_{\theta})_{\theta \in \Theta}$ in θ_0 L^2 -differenzierbar ist und (5.60) gilt, folgt nach dem Satz von Le Cam, vgl. z.B. Van der Vaart [63, Kapitel 7] oder Janssen [30, Satz 14.8], dass das Experiment $\{P_{n,\theta}: \theta \in \Theta_n\}$ die LAN-Bedingung erfüllt mit zentraler Folge

$$\frac{1}{c^2 I_0} Z_n := \frac{1}{c^2 I_0} \sum_{i=1}^n d_{ni} g(X_i). \tag{5.62}$$

Dies impliziert gerade die wechselseitige Benachbartheit $P_{n,\theta} \triangleleft \triangleright P_{\theta_0}^n$, so dass aus (5.61)

$$\left|\varphi_{n,PM} - \varphi_{n,PM}^*\right| \stackrel{P_{n,\theta}}{\longrightarrow} 0$$

folgt. Dies impliziert aufgrund von $\left|\varphi_{n,PM}-\varphi_{n,PM}^*\right|\leq 2$ nach dem Satz von der dominierten Konvergenz gerade a).

Aufgrund von (5.52) und (5.58) besitzt die Teststatistik die Darstellung

$$a_n T_n = \sum_{i=1}^n b_{ni} g(X_i) + o_P(1).$$

Seien nun $\lambda_1, \lambda_2 \in \mathbb{R}$ beliebig, so gilt

$$\lambda_1 Z_n + \lambda_2 \sum_{i=1}^n b_{ni} g(X_i) = \sum_{i=1}^n (\lambda_1 d_{ni} + \lambda_2 b_{ni}) g(X_i) =: \sum_{i=1}^n a_{ni} g(X_i).$$

Aus

$$\max_{i \le n} |b_{ni}| \to 0, \quad \sum_{i=1}^{n} b_{ni}^2 = \frac{1}{I_0^2} a_n^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{1}{I_0^2}$$
 (5.63)

folgt

$$\max_{i \le n} |a_{ni}| \to 0, \quad \sum_{i=1}^{n} a_{ni}^{2} \to \lambda_{1}^{2} c^{2} + \lambda_{2}^{2} \frac{1}{I_{0}^{2}} + 2\lambda_{1} \lambda_{2} \sum_{i=1}^{\infty} d_{ni} b_{ni} =: \tilde{\sigma}^{2}.$$

Dabei ist der Grenzwert auf der rechten Seite endlich, da $\sigma_{12} = \sum_{i=1}^{\infty} d_{ni}b_{ni}$ nach der Cauchy-Schwarz-Ungleichung absolut konvergent ist. Der Satz von Le Cam impliziert deshalb zusammen mit dem Lemma von Slutzky

$$\lambda_1 Z_n + \lambda_2 a_n T_n \xrightarrow{\mathcal{D}} N(0, I_0 \tilde{\sigma}^2),$$

so dass sich die gemeinsame Konvergenz

$$\begin{pmatrix} Z_n \\ a_n T_n \end{pmatrix} \xrightarrow{\mathcal{D}} N \left(0, I_0 \begin{pmatrix} c^2 & \sigma_{12} \\ \sigma_{12} & I_0^{-2} \end{pmatrix} \right)$$

aus dem klassischen Cramér-Wold-Device ergibt. Aus dem dritten Lemma von Le Cam folgen hieraus die Aussagen b) und c) zunächst für $\varphi_{n,PM}$ (vgl. z.B. Janssen [30, Satz 15.2]) und aufgrund von a) auch für $\varphi_{n,PM}^*$. Aussage d) erhält man nun aus Gleichung (5.63).

Das folgende Beispiel soll verdeutlichen, dass die Äquivalenz des Permutationstests und des parametrischen Gegenstücks von der richtigen Parametrisierung abhängt. Ausgezeichnet ist hierbei eine Mittelpunktsparametrisierung, die etwa von Janssen und Völker [36] für bedingte Tests diskutiert wird. Für Abweichungen von dieser Mittelpunktsparametrisierung ergeben sich unterschiedliche Grenzverteilungen für die Teststatistik und die zugehörige Permutationsstatistik.

Beispiel 5.23

Gegeben sei eine Exponentialfamilie

$$\frac{dP_{\vartheta}}{dP_0} = C(\vartheta)e^{\vartheta T}, \quad \vartheta \in \Theta \subset \mathbb{R}$$

mit $0 \in \overset{\circ}{\Theta}$, so dass $E_{P_0}(T) = 0$ und $Var_{P_0}(T) = \sigma^2 \in (0,\infty)$ gilt. Für die Kurve

$$\vartheta \mapsto Q_{n,\vartheta} := P_{\vartheta/\sqrt{n_1}}^{n_1} \otimes P_0^{n_2}$$

betrachtet man unter der Voraussetzung $n_1/n \to p \in (0,1)$ nun das (eingeschränkte) Zweistichprobenproblem

$$H = \{\vartheta = 0\}$$
 gegen $K = \{\vartheta > 0\}.$

Wegen

$$\frac{dQ_{n,\vartheta}}{dP_0^n}(x_1,\ldots,x_n) = C(\vartheta)^{n_1} \exp\left(\vartheta \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} T(x_i)\right)$$

erhält man hierfür den parametrischen score-Test in der Teststatistik

$$T_n := \sqrt{n_1} \cdot \sum_{i=1}^{n_1} (T(X_i) - E_{P_0^n}(T(X_i))) = \sqrt{n_1} \cdot \sum_{i=1}^{n_1} T(X_i),$$

vgl. WITTING [65, Beispiel 2.42], wobei $X_i, 1 \leq i \leq n$, wieder die kanonischen Projektionen sind. Die Verteilung von T_n konvergiert nach dem klassischen Zentralen Grenzwertsatz schwach gegen eine $N(0, \sigma^2)$ -Verteilung. Da hierbei die Varianzen mitkonvergieren, ist für die Äquivalenz des score-Tests und des zugehörigen Permutationstests nach Lemma 4.8 notwendig, dass (4.22) gilt. Weil aber für $Y_i := T(X_i), 1 \leq i \leq n$, und die Regressionskoeffizienten $c_{n,i}$ aus (2.9)

$$T_{n} - E_{P_{0}^{n}}(T_{n}|X_{1:n}, \dots, X_{n:n}) = \sqrt{n_{1}} \cdot \overline{Y}_{n_{1}} - \sqrt{n_{1}} \left(\frac{n_{1}}{n} \overline{Y}_{n_{1}} + \frac{n_{2}}{n} \overline{Y}_{n_{2}}\right)$$

$$= \sqrt{\frac{n_{2}}{n}} \sqrt{\frac{n_{1}n_{2}}{n}} (\overline{Y}_{n_{1}} - \overline{Y}_{n_{2}}) = \sqrt{\frac{n_{2}}{n}} \sum_{i=1}^{n} c_{n,i} Y_{i}$$

gilt, ist diese Bedingung wegen der Konvergenz

$$Var_{P_0^n}(T_n - E_{P_0^n}(T_n | X_{1:n}, \dots, X_{n:n})) = \frac{n_2}{n} \sum_{i=1}^n c_{n,i}^2 Var_{P_0^n}(Y_i)$$
$$= \frac{n_2}{n} \sigma^2 \to (1-p)\sigma^2$$

nicht erfüllt, so dass keine Äquivalenz der beiden Tests vorliegen kann.

Zweistichproben-Permutationstest in endlichdimensionalen Quadratischen Formen

Wie im Einstichprobenfall kann man auch im Zweistichprobenfall die Erwartungswerte von zwei mehrdimensionalen Stichproben vergleichen. Aus Gründen der Übersichtlichkeit beschränken sich die nachfolgenden Betrachtungen dabei auf einfache Dreiecksschema von k-dimensionalen Zufallsvektoren

$$X_{n,i} = \begin{cases} Y_i & \text{für } i \le n_1 \\ Z_i, & \text{für } n_1 < i \le n. \end{cases}$$
 (5.64)

Dabei seien Y_1, \ldots, Y_{n_1} i.i.d. mit $E(Y_1) = \mu_1$ sowie $Cov(Y_1) =: \Gamma_1 \in \mathbb{R}^{k \times k}_{p.d.}$ und davon unabhängig Z_{n_1+1}, \ldots, Z_n i.i.d. mit $E(Z_n) = \mu_2$ sowie $Cov(Z_n) =: \Gamma_2 \in \mathbb{R}^{k \times k}_{p.d.}$.

Das zugehörige Testproblem ist nun gegeben durch

$$H = \{\mu_1 = \mu_2\} \text{ gegen } K = \{\mu_1 \neq \mu_2\}.$$
 (5.65)

Allgemein gilt dann unter H für $\overline{X}_{n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{n,i}$ und $\overline{X}_{n_2} = \frac{1}{n_2} \sum_{i=n_1+1}^{n} X_{n,i}$

$$T_{n} := \sqrt{\frac{n_{1}n_{2}}{n}} \left(\overline{X}_{n_{1}} - \overline{X}_{n_{2}} \right)$$

$$= \sqrt{\frac{n_{2}}{n}} \sqrt{n_{1}} (\overline{Y}_{n_{1}} - \mu_{1}) - \sqrt{\frac{n_{1}}{n}} \sqrt{n_{2}} (\overline{Z}_{n_{2}} - \mu_{2}) + \sqrt{\frac{n_{1}n_{2}}{n}} (\mu_{1} - \mu_{2})$$

$$\xrightarrow{\mathcal{D}} Z \sim N(0, \Gamma)$$

$$(5.66)$$

mit $\Gamma = (1-p)\Gamma_1 + p \ \Gamma_2 \in \mathbb{R}^{k \times k}_{p.d.}$. Sind Γ_1 und Γ_2 bekannt, so bietet sich als Teststatistik in Analogie zum Einstichprobenfall die Quadratische Form

$$Q_n := \frac{n_1 n_2}{n} \left(\overline{X}_{n_1} - \overline{X}_{n_2} \right)^t \Gamma^{-1} \left(\overline{X}_{n_1} - \overline{X}_{n_2} \right)$$
 (5.67)

an, da diese nach Satz 4.9 asymptotisch χ^2_k -verteilt ist. Dies führt auf den asymptotischen α -Niveau-Test

$$\varphi_n = \begin{cases} 1 & \text{falls } Q_n > \chi_{k,\alpha}^2, \\ 0 & \leq \end{cases}$$
 (5.68)

der aber den Nachteil besitzt, dass der kritische Wert nicht von n abhängt und man so für kleine Stichprobenumfänge keine Kontrolle über die wirkliche Fehlerwahrscheinlichkeit hat. Eine Lösung hierfür stellt nun wieder die Berechnung von datenabhängigen kritischen Werten mittels Resampling-Verfahren dar. Da es sich um ein Zweistichprobenproblem handelt, werden auch hier Permutationstests verwendet. Seien dazu wie in den vorigen Beispielen $(c_{n,i})_i$ die Regressionskoeffizienten des Zweistichprobenproblems und $W_{n,i} = c_{n,\tau(i)}$ die zugehörigen Permutationsgewichte aus Beispiel 2.6.

In diesem Fall erfüllt das Dreiecksschema $X_{n,i}/n^{1/2}$, $i \leq n$, unter $H = \{\mu_1 = \mu_2\}$ die Bedingungen des Multivariaten Zentralen Grenzwertsatzes 3.20. Dabei ergibt sich die Maximumsbedingung (3.17) aus Korollar 3.7 sowie der darauf folgenden Bemerkung 3.8 und für die empirische Kovarianzmatrix erhält man

$$\Gamma_{n}\left(\frac{X_{n,i}}{\sqrt{n}}\right) = \left(\frac{n_{1}}{n}\frac{1}{n_{1}}\sum_{i=1}^{n_{1}}Y_{i}^{(j)}Y_{i}^{(r)} + \frac{n_{2}}{n}\frac{1}{n_{2}}\sum_{i=n_{1}+1}^{n}Z_{i}^{(j)}Z_{i}^{(r)} - \overline{X}_{n}^{(j)}\overline{X}_{n}^{(r)}\right)_{1 \leq j,r \leq k}$$

$$\xrightarrow{f.s.} p\Gamma_{1} + q\Gamma_{2} =: \tilde{\Gamma}. \tag{5.70}$$

Dies impliziert nach Satz 3.20 unter H die Konvergenz

$$d_k(\mathcal{L}(\sum_{i=1}^n W_{n,i} X_{n,i} | (X_{n,i})_{i \le n}), N(0, \tilde{\Gamma})) \xrightarrow{P} 0.$$
 (5.71)

Damit also die Grenzverteilung von T_n , vgl. (5.66), und der zugehörigen Permutationsversion $T_n^* = \sqrt{n} \sum_{i=1}^n W_{n,i} X_{n,i}$ unter H übereinstimmen, benötigt man wie beim F-Test entweder einen balancierten Stichprobenumfang, d.h. p = 1/2, oder identische Kovarianzmatrizen $\Gamma_1 = \Gamma_2$. Ist eine dieser Bedingungen erfüllt, so bietet sich der folgende Permutationstest an

$$\varphi_{n,1}^* = \begin{cases} 1 & \text{falls } Q_n > c_{n,1}^*(\cdot, \alpha). \\ 0 & \leq \end{cases}$$
 (5.72)

Dabei wird $c_{n,1}^*(\cdot,\alpha)$ als das bedingte $(1-\alpha)$ -Quantil von $Q_n^*:=(T_n^*)^t\tilde{\Gamma}^{-1}T_n^*$ bei gegebenen Daten bestimmt.

Allerdings sind die obigen Einschränkungen i.a. nicht wünschenswert. Aus diesem Grund wird nun gezeigt, wie man diese in dem interessanteren Fall unbekannter und verschiedener positiv definiter Kovarianzmatrizen Γ_1 und Γ_2 umgehen kann. Hierzu verwendet man wiederum Permuationsversionen von studentisierten Statistiken. Aus diesem Grunde benötigt man noch einen geeigneten Schätzer für die unbekannte Kovarianzmatrix Γ , der gegenüber unbalancierten Stichproben robust ist. Dies wird von

$$\Gamma_{n,1} = \frac{n_1 n_2}{n} \left(\frac{1}{n_1 (n_1 - 1)} \sum_{i=1}^{n_1} (X_{n,i}^{(j)} - \overline{X}_{n_1}^{(j)}) (X_{n,i}^{(r)} - \overline{X}_{n_1}^{(r)}) \right)$$
(5.73)

$$+ \frac{1}{n_2(n_2 - 1)} \sum_{i=n_1+1}^{n} (X_{n,i}^{(j)} - \overline{X}_{n_2}^{(j)}) (X_{n,i}^{(r)} - \overline{X}_{n_2}^{(r)}) \right)_{1 < j,r < k} (5.74)$$

erfüllt, da $\Gamma_{n,1} \xrightarrow{f.s.} \Gamma$ gilt.

Analog zu (5.68) erhält man so den asymptotischen α -Niveau-Test $\mathbb{1}_{(\chi_{k,\alpha}^2,\infty)}(\hat{Q}_n)$ in der quadratischen Form

$$\hat{Q}_{n} = \frac{n_{1}n_{2}}{n} \left(\overline{X}_{n_{1}} - \overline{X}_{n_{2}} \right)^{t} \Gamma_{n,1}^{-1} \left(\overline{X}_{n_{1}} - \overline{X}_{n_{2}} \right) \mathbb{1}_{\left\{ \Gamma_{n,1}^{-1} \text{ existiert} \right\}}.$$
 (5.75)

Der nächste Satz gibt Aufschluss über das Konvergenzverhalten der zugehörigen Permutationsstatistik.

Satz 5.24

Für die Permuationsversion der Teststatistik \hat{Q}_n gilt unter H ein bedingter Zentraler Grenzwertsatz

$$d_1\left(\mathcal{L}(\hat{Q}_n((X_{n,\tau(i)})_{i\leq n})|X_{n,1},\dots,X_{n,n}),\chi_k^2\right) \stackrel{P}{\longrightarrow} 0, \tag{5.76}$$

falls vierte Momente $E(||Y_1||^4 + ||Z_n||^4) < \infty$ existieren.

Beweis. Aufgrund von

$$\hat{Q}_n((X_{n,\tau(i)})_{i\leq n}) = (T_n^*)^t \Gamma_{n,1}^{-1}((X_{n,\tau(i)})_{i\leq n}) T_n^*$$

reicht es nach (5.71) aus, zu zeigen, dass $\Gamma_{n,1}((X_{n,\tau(i)})_{i\leq n})$ stochastisch gegen $\tilde{\Gamma}$ konvergiert. Hierzu betrachtet man für beliebige $j,r\leq k$ den Matrixeintrag an der Stelle (j,r)

$$\frac{n_2}{n}C_{n,1}^{(j,r)}(X_{n,i}) + \frac{n_1}{n}C_{n,2}^{(j,r)}(X_{n,i}) := \frac{n_2}{n}\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{n,i}^{(j)} - \overline{X}_{n_1}^{(j)})(X_{n,i}^{(r)} - \overline{X}_{n_1}^{(r)}) \\
+ \frac{n_1}{n}\frac{1}{n_2-1}\sum_{i=n_1+1}^{n}(X_{n,i}^{(j)} - \overline{X}_{n_2}^{(j)})(X_{n,i}^{(r)} - \overline{X}_{n_2}^{(r)}).$$

Für den ersten Summanden gilt

$$\frac{n_1 - 1}{n_1} C_{n,1}^{(j,r)}(X_{n,i}) = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{n,i}^{(j)} X_{n,i}^{(r)} - \overline{X}_{n_1}^{(j)} \overline{X}_{n_1}^{(r)}.$$
 (5.77)

Da das Dreiecksschema zeilenweise unabhängig mit jeweils endlicher Varianz ist, konvergiert $\frac{1}{n_1}\sum_{i=1}^{n_1}X_{n,\tau(i)}^{(j)}$ nach dem schwachen Gesetz der grossen Zahlen unter H stochastisch gegen $\mu_1^{(j)}$, so dass nur noch der erste Term von $C_{n,1}^{(j,r)}((X_{n,\tau(i)})_{i\leq n})$ untersucht werden muss. Dazu definiert man Regressionskoeffizienten

$$c_{n,i} = \begin{cases} \frac{1}{n_1}, & \text{für } 1 \le i \le n_1, \\ 0, & \text{für } n_1 < i \le n. \end{cases}$$

Hierfür gilt

$$\frac{1}{n_1} \sum_{i=1}^{n_1} X_{n,\tau(i)}^{(j)} X_{n,\tau(i)}^{(r)} = \sum_{i=1}^{n} c_{n,i} X_{n,\tau(i)}^{(j)} X_{n,\tau(i)}^{(r)}$$
(5.78)

sowie

$$\sum_{i=1}^{n} c_{n,i} = 1 \text{ und } \sum_{i=1}^{n} (c_{n,i} - \overline{c}_n)^2 \to 0.$$

Für den bedingten Erwartungswert der Statisitk (5.78) gilt unter H

$$E\left(\sum_{i=1}^{n} c_{n,i} X_{n,\tau(i)}^{(j)} X_{n,\tau(i)}^{(r)} | X_{n,1}, \dots, X_{n,n} \right) = \frac{1}{n} \sum_{i=1}^{n} X_{n,i}^{(j)} X_{n,i}^{(r)}$$

$$\xrightarrow{f.s.} p E(Y_1^{(j)} Y_1^{(r)}) + q E(Z_1^{(j)} Z_1^{(r)}).$$

Da $\mathrm{E}((X_{n,i}^{(j)}X_{n,i}^{(r)})^2) < \infty$ exisitiert, gilt für die bedingte Varianz nach der Varianzformel von HÁJEK UND ŠIDAK, vgl. [22, S.61, Theorem 3],

$$\operatorname{Var}\left(\sum_{i=1}^{n} c_{n,i} X_{n,\tau(i)}^{(j)} X_{n,\tau(i)}^{(r)} | X_{n,1}, \dots, X_{n,n}\right)$$

$$= \sum_{i=1}^{n} (c_{n,i} - \overline{c}_n)^2 \frac{1}{n-1} \sum_{i=1}^{n} (X_{n,i}^{(j)} X_{n,i}^{(r)} - \frac{1}{n} \sum_{k=1}^{n} X_{n,k}^{(j)} X_{n,k}^{(r)})^2 \xrightarrow{f.s.} 0.$$

Hieraus folgt

$$\sum_{i=1}^{n} c_{n,i} X_{n,\tau(i)}^{(j)} X_{n,\tau(i)}^{(r)} \xrightarrow{P \otimes \tilde{P}} p \operatorname{E}(Y_1^{(j)} Y_1^{(r)}) + q \operatorname{E}(Z_1^{(j)} Z_1^{(r)})$$
 (5.79)

und somit

$$\begin{split} \frac{n_2}{n} C_{n,1}^{(j,r)}(X_{n,\tau(i)}) & \stackrel{P \otimes \tilde{P}}{\longrightarrow} & q \Big(p \, \mathrm{E}(Y_1^{(j)} Y_1^{(r)}) + q \, \mathrm{E}(Z_1^{(j)} Z_1^{(r)}) - \mu_1^{(j)} \mu_1^{(r)} \Big) \\ & = & q \Big(p \, \mathrm{Cov} \, (Y_1^{(j)}, Y_1^{(r)}) + q \, \mathrm{Cov} \, (Z_1^{(j)} Z_1^{(r)}) \Big). \end{split}$$

Da für $C_{n,1}^{(j,r)}$ analoge Aussagen folgen, schließt dies den Beweis.

Mit Hilfe des obigen Satzes lässt sich nun der zugehörige Permutationstest als

$$\varphi_{n,PQ} = \begin{cases} 1 & \text{falls } \hat{Q}_n > c_{n,PQ}(\cdot, \alpha) \\ 0 & \leq \end{cases}$$
 (5.80)

konstruieren, wobei $c_{n,PQ}(\cdot,\alpha)$ das bedingte $(1-\alpha)$ -Quantil von $\hat{Q}_n((X_{n,\tau(i)})_{i\leq n})$ gegeben den Daten ist.

Über die Güte unter Alternativen gibt der folgende Satz Auskunft. Dabei seien $\mu_1 = (\mu_1^{(1)}, \dots, \mu_1^{(k)})$ und $\mu_2 = (\mu_2^{(1)}, \dots, \mu_2^{(k)})$ die Erwartungswerte der ersten bzw. zweiten Stichprobe

Satz 5.25

Unter
$$K = \{\mu_1 \neq \mu_2\}$$
 gilt
$$E_K(\varphi_{n,PO}) \to 1. \tag{5.81}$$

Beweis. Da Γ und Γ^{-1} positiv definit sind und $\Gamma_{n,1} \xrightarrow{f.s.} \Gamma$ gilt, konvergiert die Teststatistik \hat{Q}_n aufgrund von (5.66) stochastisch gegen $+\infty$, da dies für den dominierenden Term

$$a_n(\mu_1-\mu_2)^t\Gamma_{n,1}^{-1}(\mu_1-\mu_2)\stackrel{P}{\longrightarrow} +\infty$$

gilt. Aus Gleichung (5.69) liest man die Konvergenz $\Gamma_n\left(\left(\frac{X_{n,i}}{\sqrt{n}}\right)_{i\leq n}\right) \xrightarrow{f.s.} \tilde{\Gamma} + C$ mit $C = (c_{j,r})_{j,r\leq k}$ und

$$c_{j,r} = p\mu_1^{(j)}\mu_1^{(r)} + q\mu_2^{(j)}\mu_2^{(r)} - (p\mu_1^{(j)} + q\mu_2^{(j)})(p\mu_1^{(r)} + q\mu_2^{(r)})$$

ab, so dass die bedingte Verteilung von T_n^* gegeben den Daten - analog zu (5.71) - schwach gegen $N(0, \tilde{\Gamma} + C)$ konvergiert.

Zeigt man nun, dass $\Gamma_{n,1}((X_{n,\tau(i)})_{i\leq n})$ unter K stochastisch gegen $\tilde{\Gamma}+C$ konvergiert, so folgt die Behauptung. Dies kann man aber gerade an den Gleichungen (5.77) und (5.79) ablesen.

Bemerkung 5.26

1. Die meisten der in diesem Kapitel vorgestellten Aussagen für Zweistichprobenpermutationstests lassen sich mit ähnlichen Methoden auch für ZweistichprobenBootstrap- oder Double-Bootstrap-Tests beweisen. Dazu zieht man aus den einzelnen Stichproben jeweils unabhängig voneinander Bootstrap-Stichproben $X_1^*, \ldots, X_{n_1}^*$ bzw. $Y_1^*, \ldots, Y_{n_2}^*$ und berechnet wie im Einstichprobenfall zu den
oben gegebenen Teststatistiken

$$T_n(X_1,\ldots,X_{n_1},Y_1,\ldots,Y_{n_2})$$

die bedingten Quantile der bedingten Verteilung der zugehörigen Bootstrap-Version

$$T_n(X_1^*,\ldots,X_{n_1}^*,Y_1^*,\ldots,Y_{n_2}^*).$$

Da die beiden Bootstrap-Gruppen wieder voneinander unabhängig sind, lässt sich so mit Hilfe von leichten Modifikationen die Konsistenz der Bootstrap-Version des Zweistichproben-Tests beweisen⁷. Allerdings besitzen diese Tests - wie oben bereits erwähnt - im Vergleich zum vorgestellten Permutationstest den Nachteil, dass sie i.a. unter der kleineren Nullhypothese⁸

$$\tilde{H}_0: \{P^{n_1} \otimes Q^{n_2} = P^n\}$$

nicht exakt sind.

 $^{^7\}mathrm{Dies}$ funktioniert beispielsweise bei den obigen Tests zum Varianz- oder Kovarianzvergleich.

⁸Hierbei sei P die Verteilung der 1. und Q die Verteilung der 2. Stichprobe.

2. Genauso lässt sich mit Hilfe von Satz 3.11 und dem sich anschließenden Kommentar beweisen, dass die meisten der oben gezeigten Behauptungen für Permutationstests auch in dem Fall extrem unbalancierter Stichproben, d.h. $n_1/n \to 0, n_1 \to \infty$, gültig bleiben.

Teil II

Exkurs: Implementierbare Resampling-Verfahren

Kapitel 6

Resampling-Verfahren für das weighted Resampling

In Teil I wurden u.a. bedingte Zentrale Grenzwertsätze für die gewichtete Resampling-Statistik

$$T_n^* = T_n^*(X_{n,i}, W_{n,i}) := \sqrt{k(n)} \sum_{i=1}^{k(n)} W_{n,i}(X_{n,i} - \overline{X}_n)$$

und deren Erweiterungen behandelt. Um dabei Aussagen über das (bedingte) Konvergenzverhalten in Verteilung treffen zu können, benötigte man dort, dass die Gewichte $(W_{n,i})_{i \leq k(n)}$ den Generalvoraussetzungen genügen. Da diese im vorliegenden Kapitel genauer untersucht werden, werden sie im Folgenden nochmals wiederholt.

$$(W_{n,1}, \dots, W_{n,k(n)})$$
 sind austauschbar, (6.1)

$$\max_{1 \le i \le k(n)} |W_{n,i} - \overline{W}_n| \to 0 \ \tilde{P}\text{-stochastisch}, \tag{6.2}$$

$$\sum_{i=1}^{k(n)} (W_{n,i} - \overline{W}_n)^2 \to 1 \ \tilde{P}\text{-stochastisch}$$
 (6.3)

und es liegt Verteilungskonvergenz

$$\sqrt{k(n)}(W_{n,1} - \overline{W}_n) \stackrel{\mathcal{D}}{\to} W_1 \tag{6.4}$$

gegen eine Zufallsvariable W_1 vor.

Dabei ist
$$\overline{W}_n := \frac{1}{k(n)} \sum_{i=1}^{k(n)} W_{n,i}$$
.

Bemerkung 6.1

- 1. Aus den Voraussetzungen folgt, dass $E(W_1) = 0$ und $Var(W_1) \le 1$ für die Zufallsvariable W_1 aus (6.4) gelten, vgl. JANSSEN UND PAULS [33, Abschnitt 3] in Verbindung mit JANSSEN [35, Lemma 6.1].
- 2. Erfüllt ein Dreiecksschema von reellen Gewichtsfunktionen $(W_{n,i})_{i \leq k(n)}$ die Bedingungen (6.2) und (6.3), so erhält man hierfür die Austauschbarkeit (6.1), indem man mit Hilfe einer auf $S_{k(n)}$ gleichverteilten Permutation τ die Gewichte $\tilde{W}_{n,i} := W_{n,\tau(i)}$ definiert.

Für die beiden, in den Arbeiten von Janssen [35] und Del Barrio et.al. [12] behandelten Fälle

- 1. $Var(W_1) = 1$,
- 2. $Var(W_1) = 0$,

werden nun verschiedene Gewichte angegeben, die den Generalvoraussetzungen genügen. Dabei gelte stets $\min(k(n), m(n)) \to \infty$ für $n \to \infty$.

Beispiel 6.2 (Der m(n)-Bootstrap)

Die m(n)-Bootstrap-Gewichte (2.7) aus Kapitel 2.4 erfüllen die obigen Voraussetzungen. Dabei liegt

- (a) für $m(n)/k(n) \to c \in (0,\infty]$ der Fall $\mathrm{Var}(W_1) = 1$ vor, wobei W_1 im Fall $0 < c < \infty$ eine zentrierte Poisson- und im Fall $c = \infty$ eine standardnormalverteilte Zufallsvariable ist.
- (b) für $m(n)/k(n) \to 0$ der Fall $Var(W_1) = 0$ vor.

Für den Spezialfall k(n) = m(n) = n erhält man insbesondere den klassischen Bootstrap von Efron.

Beweis. Für den ersten Fall $Var(W_1) = 1$ findet man den Beweis von (6.1)-(6.3) bei Janssen und Pauls [33, Lemma 4 und Theorem 6]. Die Verteilungskonvergenz in (6.4) erhält man wegen $M_{n,1} \sim B(m(n), \frac{1}{k(n)})$ für $c = \infty$ aus dem klassischen Zentralen Grenzwertsatz und für $0 < c < \infty$ aus dem Poissonschen Grenzwertsatz.

Der Fall c = 0 mit $Var(W_1) = 0$ wird bei Del Barrio et al. [12, Example 2.1 a)] behandelt.

Beispiel 6.3 (Der m(n)-double Bootstrap)

Die m(n)-double Bootstrap-Gewichte aus (2.8) erfüllen die Generalvoraussetzungen

- (a) mit $\text{Var}(W_1)=1$, falls zusätzlich zu $\frac{m(n)}{k(n)}\to c>0$ noch $\frac{m(n)^2}{k(n)^3}\to 0$ gilt.
- (b) des Low-Resampling Falls mit $W_1 = 0$ fast sicher, falls hingegen $\frac{m(n)}{k(n)} \to 0$ und $\frac{k(n)}{m(n)^2} \to 0$ gelten.

Beweis. Den Beweis für den Fall $Var(W_1) = 1$ findet man bei JANSSEN [35, Example 3(b)]).

Im Folgenden wird deshalb der Low-Resampling Fall (b) behandelt.

Die Austauschbarkeit (6.1) ergibt sich aus der Definition. Zum Nachweis der Voraussetzungen (6.2) und (6.3) geht man die Beweise zum m(n)-double-Bootstrap von JANSSEN [35] und PAULS [49] mit den neuen Eigenschaften kurz durch. Einerseits gilt nach JANSSEN [35, S.520-521]

$$\tilde{P}\left(\max_{i \le k(n)} |W_{n,i}| \ge \epsilon\right) \le \frac{4}{m(n)^2 \epsilon^4} \sum_{i=1}^{k(n)} E\left((M'_{n,i} - \frac{m(n)}{k(n)})^4\right) \\
\le C \cdot \frac{k(n)}{m(n)^2} \cdot \frac{m(n)^4}{k(n)^4} = \frac{Cm(n)^2}{k(n)^3} \to 0$$

für eine geeignete Konstante $C \ge 0$, woraus sich (6.2) ergibt. Andererseits folgt der Nachweis von (6.3) analog zu PAULS [49, S.82-84] durch die Aufteilung¹

$$\frac{1}{m(n)} \sum_{i=1}^{k(n)} \left(M'_{n,i} - \frac{m(n)}{k(n)} \right)^2 = \frac{k(n)}{m(n)} \operatorname{E}(M'_{n,1}) - \frac{m(n)}{k(n)} + \frac{1}{m(n)} \sum_{i=1}^{k(n)} M'_{n,i} - \operatorname{E}(M'_{n,i}).$$

Dort wird gezeigt, dass für den ersten Term die Konvergenz

$$\frac{k(n)}{m(n)} \operatorname{E}(M'_{n,1}^2) - \frac{m(n)}{k(n)} \to 2$$

vorliegt und für den zweiten Term

$$\tilde{P}\left(\left|\frac{1}{m(n)}\sum_{i=1}^{k(n)}M_{n,i}^{'2} - \operatorname{E}(M_{n,i}^{'2})\right| \ge \epsilon\right) \le K \cdot \frac{k(n)}{m(n)^2} \to 0$$

¹Nach Definition gilt $\overline{W}_n = 0$.

für eine geeignete Konstante $K \geq 0$ gilt. Um den Beweis abzuschließen, fehlt somit noch der Nachweis von (6.4) mit $Z_1 = 0$. Hierzu beachte man, dass $M_{n,1} \sim \mathcal{B}(m(n), \frac{1}{k(n)})$ binomialverteilt ist und somit

$$\left(\frac{k(n)}{m(n)}\right)^{1/2} \cdot \operatorname{E}\left(\left|M'_{n,1}\right|\right) = \left(\frac{k(n)}{m(n)}\right)^{1/2} \cdot \operatorname{E}\left(\operatorname{E}\left(M'_{n,1}|M_{n,1},\dots,M_{n,k(n)}\right)\right)$$

$$= \left(\frac{k(n)}{m(n)}\right)^{1/2} \cdot \operatorname{E}\left(M_{n,1}|M_{n,1},\dots,M_{n,k(n)}\right)$$

gilt. Hieraus folgt (6.4) mit $W_1 = 0$.

Genauso kann man auch allgemeine Permutationsgewichte betrachten. Sei hierzu $\pi = (\pi(i))_{i \leq k(n)}$ eine auf der symmetrischen Gruppe $\mathcal{S}_{k(n)}$ Laplace-verteilte Permutation, die unabhängig von dem vorliegenden Datensatz $X_{n,1}, \ldots, X_{n,k(n)}$ ist und $(c_{n,i})_{i \leq k(n)}$ eine Familie reeller Zahlen. Dann erfüllen die Permutationsgewichte

$$W_{n,i} = c_{n,\pi(i)}, \quad 1 \le i \le k(n),$$
 (6.5)

die Generalvoraussetzungen, falls (6.4) gilt und die Regressionskoeffizienten $c_{n,i}$ mit arithmetischem Zeilenmittel $\overline{c}_n := 1/k(n) \sum_{i=1}^{k(n)} c_{n,i}$ den Bedingungen $\max_{i \le k(n)} |c_{n,i} - \overline{c}_n| \to 0$ und $\sum_{i=1}^{k(n)} (c_{n,i} - \overline{c}_n)^2 \to 1$ genügen.

Als Beispiel erhält man die Gewichte aus Hilfssatz 5.13 oder die Zweistichprobenpermutationsgewichte.

Beispiel 6.4 (Die Zweistichprobenpermutationsgewichte) Seien

$$c_{n,i} = \left(\frac{n_1 \cdot n_2}{k(n)}\right)^{1/2} \cdot \begin{cases} \frac{1}{n_1}, & \text{für } 1 \le i \le n_1, \\ -\frac{1}{n_2}, & \text{für } n_1 < i \le k(n) \end{cases}$$
(6.6)

die Regressionskoeffizienten des Zweistichprobenproblems, vgl. (2.9). Dann erfüllen die Gewichte $W_{n,i} := c_{n,\pi(i)}, i \leq k(n) = n = n_1 + n_2$, vgl. (2.10), die Generalvoraussetzungen, falls $\min(n_1, n_2) \to \infty$ und $n_1/n \to p \in [0, 1]$ gilt. Hierbei ist

- (a) $W_1 = \sqrt{\frac{p}{q}} \mathbb{1}_A \sqrt{\frac{q}{p}} \mathbb{1}_{A^c}$ im Fall $p \in (0,1)$ eine Zufallsvariable mit 2-Punktverteilung und $Var(W_1) = 1$, wobei $\tilde{P}(A) = q$ gelte².
- (b) $W_1 = 0$ fast sicher in den anderen Fällen.

 $[\]overline{\ ^2\text{F\"{u}r den Fall }0<\lim\inf_{n\to\infty}\frac{n_1}{n_2}}\leq \lim\sup_{n\to\infty}\frac{n_1}{n_2}<1\text{ erh\"{a}lt man Grenzwerts\"{a}tze f\"{u}r die}}$ Resampling-Statistik T_n^* aus Teil I durch Betrachtung konvergenter Teilfolgen.

Beweis. Da die Permutation gleichverteilt ist, sind die Gewichte austauschbar. Wegen $\bar{c}_n = 0$ folgt Bedingung (6.2) aus

$$\max_{i \le k(n)} |W_{n,i}| = \max \left(\frac{n_2}{n_1 + n_2} \cdot \frac{1}{n_1}, \frac{n_1}{n_1 + n_2} \cdot \frac{1}{n_2} \right)^{1/2} \to 0.$$

Die dritte Voraussetzung (6.3) folgt wieder durch Nachrechnen

$$\sum_{i=1}^{k(n)} W_{n,i}^2 = \sum_{i=1}^{k(n)} c_{n,i}^2 = \frac{n_1}{n} \frac{n_2}{n_1} + \frac{n_2}{n} \frac{n_1}{n_2} = 1.$$

Für die letzte Voraussetzung werden die beiden Fälle getrennt behandelt. Sei dazu zunächst $n_1/n_2 \to 0$. Hierfür zeigt man (6.4) wie folgt

$$\tilde{P}\left(\left|\sqrt{n}W_{n,1}\right| \ge \epsilon\right) = \tilde{P}\left(\left(\frac{n_2}{n_1}\right)^{1/2} \ge \epsilon\right) \cdot \frac{n_1}{n} + \tilde{P}\left(\left(\frac{n_1}{n_2}\right)^{1/2} \ge \epsilon\right) \cdot \frac{n_2}{n}$$

$$\le \frac{n_1}{n_2} + \tilde{P}\left(\left(\frac{n_1}{n_2}\right)^{1/2} \ge \epsilon\right) \to 0.$$

Durch Vertauschung von n_1 und n_2 erhält man, dass $Var(W_1) = 0$ auch im Fall $n_2/n_1 \to 0$ gilt. Den anderen Fall mit $Var(W_1) = 1$ beweist man wie im Beweis zu Hilfssatz 5.13.

Die bisherigen Resampling-Verfahren haben alle gemeinsam, dass der Umfang der Resampling-Stichprobe deterministisch vorgegeben wird. Aus diesem Grunde haben sich einige Autoren (vgl. u.a. Mammen [45][46]) mit Bedingungen befasst, unter denen Konsistenz des Bootstrap-Verfahrens auch bei zufälligem Stichprobenumfang vorliegt. Die kommenden Beispiele zeigen, dass sich diese Verfahren auch in das weighted Resampling einordnen lassen.

Beispiel 6.5 (Der Wild-Bootstrap)

Sei $Z_{n,i}$ ein Dreiecksschema von zeilenweise i.i.d. Zufallsvariablen mit $\mathrm{E}(Z_{n,1})=0$ und $\mathrm{Var}(Z_{n,1})=1$, dann sind die **Wild-Bootstrap-Gewichte** definiert durch

$$W_{n,i} := \frac{1}{\sqrt{k(n)}} Z_{n,i}, \quad 1 \le i \le k(n), \tag{6.7}$$

vgl. Pauls [49, Beispiel 2.10.(c)]. Diese erfüllen die Voraussetzungen (6.1) - (6.3). Die Generalvoraussetzungen, d.h. Voraussetzung (6.4), sind somit erfüllt, falls $Z_{n,1}$ zusätzlich noch in Verteilung gegen eine zentrierte Zufallsvariable W_1 mit $Var(W_1) \in \{0,1\}$ konvergiert. Diese ist dann auch die Grenzvariable in (6.4).

Beweis. Bedingung (6.1) ist nach Definition erfüllt. Da nach der Tchebyscheff Ungleichung $\mathbb{1}_{\{|Z_{n,1}| \geq \epsilon \sqrt{k(n)}\}} \xrightarrow{\tilde{P}} 0$ gilt, erfüllt das Dreicksschema $(W_{n,i})_{i \leq k(n)}$ nach dem Lemma von Pratt die Lindeberg-Bedingung

$$\frac{1}{k(n)} \sum_{i=1}^{k(n)} \int_{\{|Z_{n,i}| \geq \epsilon \sqrt{k(n)}\}} Z_{n,i}^2 d\tilde{P} = \int \mathbb{1}_{\{|Z_{n,1}| \geq \epsilon \sqrt{k(n)}\}} Z_{n,1}^2 d\tilde{P} \to 0.3$$

Da das Dreiecksschema somit insbesondere infinitesimal ist und sich hieraus auch $\sqrt{k(n)}\overline{W}_n \stackrel{\tilde{P}}{\longrightarrow} 0$ ergibt, folgen die Voraussetzungen (6.2) und (6.3) aus dem Satz von Raikov.

Der obige Beweis zeigt insbesondere, dass man in Lemma 5.3. von Pauls auf die Bedingung $k(n)^{-1/2} \sum_{i=1}^{k(n)} Z_{n,i} \xrightarrow{\mathcal{D}} Z \sim N(0,1)$ verzichten kann.

Als Spezialfälle des Wild-Bootstraps erhält man einen normalisierten Bootstrap sowie einen erweiterten Poisson-Bootstrap.

Beispiel 6.6 (Der normalisierte Bootstrap)

Für W_i i.i.d. $\mathcal{N}(1,1)$ -verteilte Zufallsvariablen definiert das Dreiecksschema

$$W_{n,i} := \frac{1}{\sqrt{k(n)}}(W_i - 1), \quad 1 \le i \le k(n), \tag{6.8}$$

die Gewichte des **normalisierten Bootstraps**. Dies erfüllen offensichtlich die Generalvoraussetzungen mit $Var(W_1) = 1$.

Beispiel 6.7 (Der m(n)-Poisson-Bootstrap)

Es wird zunächst der klassische **Poisson-Bootstrap** von MAMMEN [45][46] vorgestellt. Sei dazu $N \sim \mathcal{P}(k(n))$ eine Poisson-verteilte Zufallsvariable, die unabhängig von der Ausgangsstichprobe $(X_{n,1},\ldots,X_{n,k(n)})$ ist. Man zieht nun bedingt unter $X_{n,1},\ldots,X_{n,k(n)}$ und N=k mit Zurücklegen die Poisson-Bootstrap-Stichprobe $X_{n,1}^*,\ldots,X_{n,k}^*$. Unter der Voraussetzung, dass die Zufallsvariablen $X_{n,1},\ldots,X_{n,k(n)}$ (fast sicher) paarweise verschieden sind, wird durch

$$N_j := \left| \left\{ 1 \le i \le N : X_{n,i}^* = X_{n,j} \right\} \right|, \quad 1 \le j \le k(n) \tag{6.9}$$

eine unabhängige Folge von Poisson- $\mathcal{P}(1)$ -verteilten Zufallsvariablen mit $N = \sum_{j=1}^{k(n)} N_j$ definiert, vgl. Mammen [46, S.15f.]. Die zu diesem Verfahren gehörenden Gewichte sind dann durch

$$W_{n,i} := \frac{1}{\sqrt{k(n)}} (N_i - 1), \quad 1 \le i \le k(n), \tag{6.10}$$

³Dies impliziert gerade Bedingung (5.2) bei PAULS [49].

gegeben und erfüllen die Generalvoraussetzungen mit $Var(W_1) = 1$.

Dieser Ansatz kann nun erweitert werden, indem man für eine reelle Teilfolge $(a(n))_n \subset \mathbb{R}_{>0}$ mit $\lim_{n\to\infty} a(n) = a \in [0,\infty]$ i.i.d. Poisson- $\mathcal{P}(a(n))$ -verteilte Zufallsvariablen $N_1,\ldots,N_{k(n)}$ betrachtet. Dann sind die **erweiterten Poisson-Bootstrap-Gewichte** definiert durch

$$W_{n,i} := \frac{1}{\sqrt{k(n)}} \left(\frac{N_i - a(n)}{\sqrt{a(n)}} \right), \quad 1 \le i \le k(n).$$
 (6.11)

Diese erfüllen die Generalvoraussetzungen mit $Var(W_1) = 1$, falls c > 0 gilt. Andernfalls liegt der Fall mit $Var(W_1) = 0$ vor.

Die Wahl a(n) = m(n)/k(n) führt zu einem sog. m(n)-Poisson-Bootstrap.

Beweis. Da das Dreiecksschema $Z_{n,i} := \frac{N_i - a(n)}{\sqrt{a(n)}}$ zeilenweise i.i.d. mit Erwartungswert 0 und Varianz 1 ist, handelt es sich um Wild-Bootstrap-Gewichte, so dass nach Beispiel 6.5 nur noch die Konvergenz der charakteristischen Funktion von $Z_{n,1}$

$$\hat{\varphi}_{Z_{n,1}}(t) = \exp\left(a(n)\left(\frac{-it}{\sqrt{a(n)}} + \exp\left(\frac{it}{\sqrt{a(n)}}\right) - 1\right)\right)$$

diskutiert werden muss. Für a=0 ergibt sich $\hat{\varphi}_{Z_{n,1}}(t) \to 1=\hat{\varepsilon}_0(t)$

Die beiden anderen Fällen weist man analog zum Beweis von Lemma 6.4 bei PAULS [49] nach.

DEL BARRIO ET AL. [12] haben in diesem Zusammenhang noch ein weiteres Verfahren vorgeschlagen, dass mehr Gewicht in die Flanken legt. Dieser sogenannte heavy-tailed Wild-Bootstrap wird im folgenden Beispiel vorgestellt.

Beispiel 6.8 (Der heavy-tailed Wild-Bootstrap)

Sei $(\xi_i)_{i \leq k(n)}$ eine i.i.d. Fogle von 0-symmetrischen Zufallsvariablen mit $\text{Var}(\xi_1) = \infty$. Für $\xi_{n,i} = \xi_i \mathbb{1}_{[0,k(n)^{1/2}]}(\xi_i)$ definiert man dann für n groß genug die heavytailed Wild-Bootstrap-Gewichte als

$$W_{n,i} = \frac{\xi_{n,i}}{(k(n) \cdot \text{Var}(\xi_{n,i}))^{1/2}}, \quad 1 \le i \le k(n).$$
 (6.12)

Diese erfüllen die Generalvoraussetzungen 3.2 mit $W_1=0$, vgl. Del Barrio et al. [12, Example 2.1 c)].

Die Bayesianische Variante des klassischen Bootstraps von Efron wurde 1981 von Rubin [56] eingeführt und lässt sich wie folgt in das gewichtete Verfahren integrieren.

Beispiel 6.9 (Der Bayesian-Bootstrap)

Für i.i.d. standard exponential-verteilte Zufallsvariable $W_1, \ldots, W_{k(n)}$ definiert man die Bayesian-Bootstrap-Gewichte durch

$$W_{n,i} := \sqrt{k(n)} \left(\frac{W_i}{\sum_{j=1}^{k(n)} W_j} - \frac{1}{k(n)} \right), \quad 1 \le i \le k(n).$$
 (6.13)

Diese erfüllen die Generalvoraussetzungen mit $Var(W_1) = 1$, vgl. JANSSEN [35, Example 4]).

Als Verallgemeinerung hiervon erhält man den zuerst von Praestgaard und Wellner [53] vorgestellten i.i.d. weighted Bootstrap.

Beispiel 6.10 (Der i.i.d. weighted Bootstrap)

Sei $(\nu_i)_i$ eine Folge strikt positiver i.i.d. Zufallsvariablen mit endlicher Varianz $0 < \text{Var}(\nu_1) < \infty$ und $\rho^2 = \frac{\text{Var}(\nu_1)}{\text{E}(\nu_1)^2}$, dann erfüllen die Gewichte des i.i.d. weighted Bootstraps

$$W_{n,i} := \frac{\sqrt{k(n)}}{\rho} \left(\frac{\nu_i}{\sum_{j=1}^{k(n)} \nu_j} - \frac{1}{k(n)} \right), \quad 1 \le i \le k(n), \tag{6.14}$$

die Generalvoraussetzungen mit $Var(W_1) = 1$, vgl. Janssen [35, Lemma 6.1].

Den bisherigen Verfahren ist - mit Ausnahme des Permutationsverfahrens - gemein, dass die Resampling-Stichprobe durch Ziehen mit Zurücklegen aus der gesamten Stichprobe erzeugt wird. Mit diesen Bedingungen brechen nun die folgenden drei Verfahren.

Zunächst wird das **Subsampling**-Verfahren vorgestellt, bei dem aus einer sehr großen Gesamtstichprobe eine verhältnismäßig kleine Resampling-Stichprobe durch Ziehen ohne Zurücklegen erzeugt wird, vgl. u.a. DEL BARRIO ET AL. [12, Example 2.1 d)].

Beispiel 6.11 (Das Subsampling)

Man definiert die Subsampling-Gewichte durch

$$W_{n,i} = m(n)^{1/2} \left(\frac{1}{m(n)} H_{n,i} - \frac{1}{k(n)} \right), \quad 1 \le i \le k(n), \tag{6.15}$$

wobei $H_n = (H_{n,1}, \dots, H_{n,k(n)})$ so mehrdimensional hypergeometrisch verteilt ist, dass $P(H_n = x_n) = (\binom{k(n)}{m(n)})^{-1}$ für $x_n = (x_n^{(i)})_{i \leq k(n)} \in \{0,1\}^{k(n)}$ mit

 $\sum_{i=1}^{k(n)} x_n^{(i)} = m(n)$ gilt. Unter der Bedingung $\frac{m(n)}{k(n)} \to 0$ erfüllen diese Gewichte die Generalvoraussetzungen mit $Var(W_1) = 0$.

Beweis. Austauschbarkeit (6.1) und Zentrierung ($\overline{W}_n = 0$) der Gewichte erhält man unmittelbar aus der Definition. Für den Beweis der drei anderen Bedingungen sei bemerkt, dass die Zufallsvariablen $H_{n,i}$ einer hypergeometrischen- $\mathcal{H}(k(n), 1, m(n))$ -Verteilung unterliegen und deshalb mit Wahrscheinlichkeit $p_n = \binom{k(n)-1}{m(n)-1}/\binom{k(n)}{m(n)} = m(n)/k(n)$ den Wert 1 und mit Wahrscheinlichkeit $1-p_n$ den Wert 0 annehmen. So erhält man die Konvergenzen in (6.2)

$$\max_{i \le k(n)} |W_{n,i}| = \max \left(\frac{m(n)^{1/2}}{k(n)}, \frac{1}{m(n)^{1/2}} - \frac{m(n)^{1/2}}{k(n)} \right) \to 0,$$

(6.3)

$$\sum_{i=1}^{k(n)} W_{n,i}^2 = \frac{1}{m(n)} \sum_{i=1}^{k(n)} H_{n,i}^2 - \frac{2\sum_{i=1}^{k(n)} H_{n,i} + m(n)}{k(n)} = 1 - \frac{m(n)}{k(n)} \to 0$$

und (6.4)

$$\tilde{P}\left(\left|\sqrt{k(n)}W_{n,1}\right| \ge \epsilon\right) = p_n \cdot \tilde{P}\left(\left|\left(\frac{k(n)}{m(n)}\right)^{1/2} - \left(\frac{m(n)}{k(n)}\right)^{1/2}\right| \ge \epsilon\right) + (1 - p_n) \cdot \tilde{P}\left(\frac{m(n)}{k(n)} \ge \epsilon^2\right) \\
\le p_n + \tilde{P}\left(\frac{m(n)}{k(n)} \ge \epsilon^2\right) \to 0.$$

Das folgende Verfahren ist eine Erweiterung des klassischen Jackknife-Verfahrens, welches vor den Zeiten des Bootstraps recht populär war, diesem jedoch in der Regel unterlegen ist, vgl. z.B. Kapitel 11, insbesondere Abschnitt 11.6., bei Efron und Tibshirani [15]. Dort findet man auch eine schöne Einführung in die Methode des klassischen jackknifes. Die folgende Erweiterung findet man u.a. bei Praestgaard und Wellner [53].

Beispiel 6.12 (Der gruppierte jackknife)

Beim gruppierten jackknife wird die Resampling-Stichprobe erzeugt, indem man aus den Daten zufällig eine feste Anzahl $h_n \in \mathbb{N}, h_n \leq n$, von Beobachtungen weglässt, vgl. Praestgaard und Wellner [53, S.2061]. Dies kann mit

Hilfe der Koeffizienten

$$j_{n,i} = \begin{cases} \frac{n}{n - h_n}, & \text{falls } i \le n - h_n \\ 0 & \text{sonst} \end{cases},$$

beschrieben werden. Die gruppierten jackknife-Gewichte sind dann durch

$$W_{n,i} := \sqrt{\frac{n - h_n}{n}} \frac{(j_{n,\tau(i)} - 1)}{\sqrt{h_n}}, \quad 1 \le i \le n$$
 (6.16)

definiert, wobei τ wieder die auf der symmetrischen Gruppe \mathcal{S}_n gleichverteilte Permutation aus Beispiel 2.6 ist. Diese Gewichte erfüllen die Generalvoraussetzungen mit $\operatorname{Var}(W_1) = 1$, falls $h_n/n \to c \in (0,1)$ für $n \to \infty$ gilt.

Beweis. Die Gewichte sind aufgrund der Definition austauschbar. Die anderen Bedingungen erhält man aufgrund von $\overline{W}_n = 0$ wie folgt. Zunächst erhält man wegen der Konvergenz

$$\max_{i \le n} |W_{n,i}| = \frac{1}{\sqrt{n}} \max\{\sqrt{\frac{h_n}{n - h_n}}, \sqrt{\frac{n - h_n}{h_n}}\} \to 0$$

(6.2). (6.3) folgt aus der Gleichung

$$\sum_{i=1}^{n} W_{n,i}^{2} = \frac{n - h_{n}}{n h_{n}} \left((n - h_{n}) \left(\frac{h_{n}}{n - h_{n}} \right)^{2} + h_{n} \right) = 1.$$

Weiter gilt

$$\tilde{P}(j_{n,\tau(1)} - 1 = -1) = \frac{h_n}{n} \to c$$

sowie

$$\tilde{P}(j_{n,\tau(1)} - 1 = \frac{h_n}{n - h_n}) = \frac{n - h_n}{n} \to 1 - c$$

und somit aufgrund von $\frac{h_n}{n-h_n} \to \frac{c}{1-c}$ auch

$$\sqrt{n}W_{n,1} \xrightarrow{\mathcal{D}} W_1 \text{ mit } \tilde{P}(W_1 = -\sqrt{\frac{1-c}{c}}) = c \text{ und } \tilde{P}(W_1 = \sqrt{\frac{c}{1-c}}) = 1-c,$$

so dass (6.4) mit
$$E(W_1) = 0$$
 und $Var(W_1) = 1$ gilt.

Der obige Beweis zeigt insbesondere, dass sich der klassische jackknife, bei dem nur eine Beobachtung weggelassen wird, nicht in die vorliegende gewichtete Theorie einsortieren lässt, da u.a. Bedingung (6.2) verletzt ist.

Das nächste Verfahren scheint auf den ersten Blick mit der Subsampling-Prozedur verwandt zu sein, da die Gewichte zu derselben Verteilungsklasse gehören. Jedoch ist der Ansatzpunkt dieses Verfahrens ein anderer.

Beispiel 6.13 (Der erweiterte multivariate hypergeometrische Bootstrap)

Beim multivariaten hypergeometrischen Bootstrap erzeugt man von jeder Stichprobe K Kopien und zieht dann aus den nK Elementen n Mal ohne Zurücklegen. Der Vorteil liegt dabei darin, dass die Masse gleichmässiger über die gesamte Stichprobe verteilt wird als beim gewöhnlichen Bootstrap, vgl. PRAEST-GAARD UND WELLNER [53, S.2060]. Dieser Ansatz wird nun erweitert, indem man zum Einen zulässt, dass die Anzahl der Kopien h(n) mit n anwächst und zum anderen m(n) mal ohne Zurücklegen zieht. Dies wird durch einen multivariat hypergeometrisch verteilten Zufallsvektor $H_n = (H_{n,i})_{i \leq n}$ mit

$$\tilde{P}(H_{n,1} = k_1, \dots, H_{n,n} = k_n) = \frac{\binom{h(n)}{k_1} \cdots \binom{h(n)}{k_n}}{\binom{nh(n)}{m(n)}}, \quad \sum_{i=1}^n k_i = m(n), 0 \le k_j \le h(n),$$

modelliert, vgl. Johnson und Kotz [38, S.91].

Die hierzu gehörenden Gewichte

$$W_{n,i} := \sqrt{m(n)} \left(\frac{1}{m(n)} H_{n,i} - \frac{1}{n}\right), \quad 1 \le i \le n, \tag{6.17}$$

erfüllen die Generalvoraussetzungen mit $Var(W_1) = 1$, falls $min(n, h(n)) \to \infty$ und $m(n)/n \to c \in (0, \infty)$ gilt.

Um dies zu beweisen, wird der nächste Hilfssatz benötigt.

Hilfssatz 6.14

Sei $\mathcal{H}(N,R,n)$ eine hypergeometrische-Verteilung⁴ zu den Parametern $N=N(n), R=R(n), n\in\mathbb{N}$, die für $n\to\infty$ den Bedingungen $\min(R,N)\to\infty$ sowie $nR/N\to\lambda\in(0,\infty)$ genügen. Dann konvergiert $\mathcal{H}(N,R,n)$ schwach gegen eine Poissonverteilung zum Parameter λ , i.Z.

$$\mathcal{H}(N,R,n) \xrightarrow{w} \mathcal{P}(\lambda).$$
 (6.18)

Beweis. Der Beweis erfolgt durch Betrachtung der Totalvariationsabstände⁵. Nach Beispiel 25 aus dem Buch von LINDVALL [42, S.65ff.] gilt

$$\|\mathcal{H}(N,R,n) - \mathcal{P}(n\frac{R}{N})\|_{TV} \le 2\frac{R+n}{N-1} \to 0.$$

 $^{^4}$ Zieht man n mal ohne Zurücklegen aus einer Urne mit N Kugeln, von denen R verschiedene rot sind, so beschreibt diese Verteilung gerade die Wahrscheinlichkeit für die Anzahl gezogener roter Kugeln.

⁵Der Totalvariationsabstand zweier Maße P und Q auf einem Messraum (Ω, \mathcal{A}) ist dabei definiert als $||P - Q||_{TV} := \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$.

Da man mit Hilfe der majorisierten Konvergenz auch die Konvergenz

$$\|\mathcal{P}(n\frac{R}{N}) - \mathcal{P}(\lambda)\|_{TV} \to 0$$

erhält und die Konvergenz in der Norm der Totalvariation schon Verteilungskonvergenz impliziert, folgt die Behauptung aus der Dreiecksungleichung.

Hiermit kann nun die Aussage aus Beispiel 6.13 bewiesen werden.

Beweis. Die Austauschbarkeit der Gewichte folgt unmittelbar aus der Definition. Weiter gilt $\overline{W}_n = 0$. Aus Hilfsatz 6.14 folgt aufgrund von $\min(n, h(n)) \to \infty$ die vierte Voraussetzung (6.4)

$$\sqrt{n}W_{n,1} = \sqrt{\frac{n}{m(n)}}H_{n,1} - \sqrt{\frac{m(n)}{n}} \xrightarrow{\mathcal{D}} \sqrt{\frac{1}{c}}\tilde{Z} - \sqrt{c},$$

wobei $\tilde{Z} \sim \mathcal{P}(c)$ gilt.

Für die beiden anderen Bedingungen benötigt man nun noch einige Aussagen über die Momente der hypergeometrischen Verteilung. Nach JOHNSON UND KOTZ [38, S.92f] gilt für die faktoriellen Momente der hypergeometrischen Verteilung 6

$$\mathrm{E}\left(\prod_{i=1}^n H_{n,i}^{(r_i)}\right) = \frac{m(n)^{(r)}}{(nh(n))^{(r)}} \prod_{i=1}^n h(n)^{(r_i)} \text{ für } r = \sum_{i=1}^n r_i, \ r_i \in \mathbb{N}_0.$$

Dies impliziert für feste $r_i \in \mathbb{N}_0$

$$\mathrm{E}\left(\prod_{i=1}^n H_{n,i}^{(r_i)}\right) \to c^r < \infty.$$

Insbesondere erhält man die Konvergenzen

$$E(H_{n,1}) \to c, \quad E(H_{n,1}^2) \to c(c+1), \quad E(H_{n,1}^4) \to p(c) < \infty,$$
 (6.19)

wobei p ein Polynom vom Grad 4 ist, sowie für $i \neq j$

$$E(H_{n,i}H_{n,j}^2) \to c^3 + c^2.$$
 (6.20)

⁶Für $r \in \mathbb{N}$ und $x \in \mathbb{R}$ definiert man $x^{(r)} := x(x-1) \cdots (x-r+1)$.

Hiermit lassen sich nun die Bedingungen (6.2) und (6.3) nachweisen. Zum einen erhält man aus der allgemeinen Markoff Ungleichung die Konvergenz in (6.2)

$$\tilde{P}(\max_{i \le n} |W_{n,i}| \ge \epsilon) \le n\tilde{P}(|H_{n,1} - \frac{m(n)}{n}| \ge \sqrt{m(n)}\epsilon)$$

$$\le \frac{n}{\epsilon^4 m(n)^2} \tilde{E}(|H_{n,1} - \frac{m(n)}{n}|^4)$$

$$= \frac{1}{m(n)} \mathcal{O}(1) \to 0.$$

Andererseits folgt

$$E(\sum_{i=1}^{n} W_{n,i}^{2}) = \frac{1}{m(n)} \sum_{i=1}^{n} E(H_{n,i}^{2}) - \frac{2}{n} \sum_{i=1}^{n} E(H_{n,i}) + \frac{m(n)}{n}$$

$$\to (c+1) - 2c + c = 1.$$

Beweist man nun, dass die Varianz der Summe der Quadrate gegen 0 konvergiert, so folgt auch Bedingung (6.3) und somit die Behauptung. Hierfür zeigt man zunächst, dass für hinreichend großes n

$$\operatorname{Var}(\sum_{i=1}^{n} W_{n,i}^{2}) = \sum_{i=1}^{n} \operatorname{Var}(W_{n,i}^{2}) + \sum_{j=j} \operatorname{Cov}(W_{n,i}^{2}, W_{n,j}^{2}) \le \sum_{i=1}^{n} \operatorname{Var}(W_{n,i}^{2}) + o(1)$$

gilt. Dafür teilt man den zu behandelnden Term wie folgt auf

$$\sum_{j=j} \operatorname{Cov}(W_{n,i}^2, W_{n,j}^2) = \frac{1}{m(n)^2} \sum_{j=j} \operatorname{Cov}(H_{n,i}^2, H_{n,j}^2) + \frac{4}{n^2} \sum_{j=j} \operatorname{Cov}(H_{n,i}, H_{n,j}) - \frac{4}{nm(n)} \sum_{j=j} \operatorname{Cov}(H_{n,i}^2, H_{n,j}).$$

Da nach JOHNSON UND KOTZ [38, S.92f] $\operatorname{Cov}(H_{n,i}, H_{n,j}) \leq 0$ gilt und der Beweis von Praestgaard und Wellner [53, S.2083] auch in dieser Situation $\operatorname{Cov}(H_{n,i}^2, H_{n,j}^2) \leq 0$ impliziert, kann man in der Abschätzung die beiden ersten Summen weglassen. Die Konvergenz der letzten Summe gegen 0 ergibt sich mit Hilfe der Aussagen (6.19) und (6.20)

$$Cov(H_{n,i}, H_{n,j}^2) = E(H_{n,i}H_{n,j}^2) - E(H_{n,i}) E(H_{n,j}^2) \to 0.$$

Dies impliziert nach (6.19) und (6.20) die Behauptung

$$\sum_{i=1}^{n} \operatorname{Var}(W_{n,i}^{2}) = \frac{n}{m(n)^{2}} \operatorname{Var}(H_{n,1}^{2}) + \frac{4}{n^{2}} \operatorname{Var}(H_{n,1}) - \frac{4}{m(n)n} \operatorname{Cov}(H_{n,1}^{2}, H_{n,1}) \to 0.$$

Bemerkung 6.15

Wählt man im obigen Beispiel 6.13 - wie Praestgaard und Wellner [53] - eine konstante Anzahl von Kopien $h(n) \equiv h \in \mathbb{N}_{\geq \lceil c \rceil}$, so erhält man auch für diese Gewichte mit obigem Beweis die Bedingungen (6.1)-(6.3). Allerdings liegt dann für die charakteristische Funktion von $H_{n,1}$

$$\hat{\varphi}_{H_{n,1}}(t) = \sum_{k=0}^{h} e^{itk} \frac{\binom{nh-h}{m(n)-k}\binom{h}{k}}{\binom{nh}{m(n)}} \to \sum_{k=0}^{h} e^{itk} \binom{h}{k} \left(\frac{c}{h}\right)^k \left(\frac{h-c}{h}\right)^{h-k} = \hat{\varphi}_Y(t)$$

Konvergenz gegen die charakteristische Funktion einer $B(h, \frac{c}{h})$ -verteilten Zufallsvariable Y vor. Somit ist Bedingung (6.4) erfüllt

$$\sqrt{n}W_{n,1} \xrightarrow{\mathcal{D}} \sqrt{\frac{1}{c}}Y - \sqrt{c} =: Z,$$

wobei $\operatorname{Var}(Z) = \frac{h-c}{h} < 1$ gilt. Dieser Fall mit einer Grenzvariablen Z mit 0 < Var(Z) < 1 wird jedoch noch nicht von den in Kapitel 3 aufgeführten Grenzwertsätzen abgedeckt, so dass an dieser Stelle noch weiterer Forschungsbedarf besteht. Detaillierte Untersuchungen existieren nur für die oben angesprochenen Fälle Var(Z) = 1, vgl. Janssen [35] sowie Janssen und Pauls [33], und Var(Z) = 0, vgl. Del Barrio et.al. [12], [13].

Wie schon zu Beginn dieser Arbeit in Beispiel 2.3 erwähnt wurde, versagt das klassische Bootstrap-Verfahren bei abhängigen Daten, da dieses keine Abhängigkeitsstruktur berücksichtigt. In vielen Fällen kann dies aber wieder durch das folgende Verfahren behoben werden.

Das Block-Bootstrap-Verfahren

Bei dieser Resampling-Methode besteht die Grundidee darin, die Stichprobe in verschiedene Blöcke der Länge $L(n) \geq 1$ zu unterteilen. Die Block-Bootstrap-Stichprobe erhält man dann durch Ziehen mit Zurücklegen dieser Blöcke. Der Vorteil dieses Verfahrens besteht darin, dass die Abhängigkeitsstruktur der Daten durch Blockbildung nicht zerstört wird.

In der Literatur finden sich viele Versionen dieser Vorgehensweise, die sich durch die Wahl der Blöcke bzw. der Blocklänge (deterministische oder stochastische Blocklänge) unterscheiden. Jedoch haben alle Verfahren aus Konsistenzgründen die Forderung gemein, dass mit wachsendem Stichprobenumfang auch die Blocklänge wächst, vgl. HÄRDLE ET. AL. [20, S.440]. Diese Forderung wird bei abhängigen Daten oft dadurch ergänzt, dass die Blöcke nicht zu groß werden, d.h. man benötigt

 $L(n) \to \infty \text{ und } \frac{L(n)}{k(n)} \to 0,$ (6.21)

vgl. Lahiri [39, S.25]. Im Folgenden werden zunächst die einzelnen Verfahren vorgestellt, um dann in die Terminologie der gewichteten Resampling-Verfahren eingeordnet zu werden.

Beispiel 6.16 (Der Nonoverlapping-Block-Bootstrap)

Wie der Name schon suggeriert, teilt man beim

Nonoverlapping-Block-Bootstrap (NBB) die Stichprobe $X_{n,1}, \ldots, X_{n,k(n)}$ in b(n) überschneidungsfreie Blöcke

$$B_{n,i}^{NBB} := \{ X_{n,(i-1)\cdot L(n)+1}, \dots, X_{n,i\cdot L(n)} \}, \quad 1 \le i \le b(n), \tag{6.22}$$

der Länge L(n) ein. Dabei wird durch die Wahl der Blocklänge L(n), vgl. hierzu Bemerkung 6.19, die Anzahl der Blöcke b(n) durch $b(n) = \lfloor \frac{k(n)}{L(n)} \rfloor$ bestimmt. Aus der Aufteilung (6.22) zieht man nun r(n) Blöcke $B_{n,1}^*, \ldots, B_{n,r(n)}^*$ mit Zurücklegen und erhält so durch Hintereinanderschreiben die NBB-Stichprobe $X_{n,1}^*, \ldots, X_{n,m(n)}^*$ mit $m(n) = r(n) \cdot L(n)$. Da für den Stichprobenumfang nicht immer $k(n) = b(n) \cdot L(n) =: \tilde{k}(n)$ gilt, werden bei dieser Methode $k(n) - \tilde{k}(n)$ Beobachtungen nicht berücksichtigt.

Um dieses Verfahren nun in die Theorie des gewichteten Resamplings einzuordnen, summiert man die Beobachtungen innerhalb der Blöcke wie folgt auf

$$Y_{n,i}^{NBB} := \sum_{j=1}^{L(n)} X_{n,L(n)\cdot(i-1)+j}, \quad 1 \le i \le b(n), \tag{6.23}$$

und zieht nun aus dieser neuen Stichprobe r(n) Mal mit Zurücklegen. Die zugehörigen Resampling-Gewichte sind dann analog zum m(n)-Bootstrap (2.7) definiert als

$$W_{n,i}^{NBB} = \sqrt{r(n)} \left(\frac{1}{r(n)} M_{n,i}^{NBB} - \frac{1}{b(n)} \right), \quad 1 \le i \le b(n). \tag{6.24}$$

Dabei ist $M^{NBB}:=(M_{n,1}^{NBB},\ldots,M_{n,b(n)}^{NBB})\sim \mathcal{M}(r(n),\frac{1}{b(n)})$ multinomialverteilt zum Stichprobenumfang $r(n)=\sum_{i=1}^{b(n)}M_{n,i}^{NBB}$ mit jeweils gleicher Auswahlwahrscheinlichkeit $\frac{1}{b(n)}$. Analog zu Beispiel 6.2 sind die Generalvoraussetzungen

erfüllt, wenn $\min(r(n), b(n)) \to \infty$ gilt. Dabei zählt der Fall $\frac{r(n)}{b(n)} \to 0$ zu den low-Resampling-Gewichten mit $Var(W_1) = 0$.

Die hierzu gehörende Resampling-Statistik ist

$$T_n^*(Y_{n,i}^{NBB}) = \sqrt{b(n)L(n)} \sum_{i=1}^{b(n)} \sum_{j=1}^{L(n)} W_{n,i}^{NBB} X_{n,L(n)\cdot(i-1)+j}, \tag{6.25}$$

für die nach Kapitel 3 bekanntlich bedingte zentrale Grenzwersätze existieren⁷, falls das Dreiecksschema $(Y_{n,i}^{NBB})_{i \leq b(n)}$ den dortigen Bedingungen⁸ genügt.

Ein Nachteil des NBB besteht darin, dass, je nach Wahl der Blocklänge L(n), nicht der gesamte Stichprobenumfang k(n) ausgeschöpft werden kann. Um dies zu umgehen, kann man von der Forderung der Überschneidungsfreiheit Abstand nehmen. In diesem Fall spricht man vom sog. Overlapping-Block-Bootstrap.

Beispiel 6.17 (Der Overlapping-Block-Bootstrap)

Anders als beim NBB wählt man die Blöcke beim Overlapping-Block-

Bootstrap (OBB) so, dass sich zwei aufeinander folgende Blöcke nur um genau eine Beobachtung unterscheiden. Zu geeignet gewählter Blocklänge L(n) definiert man hierbei die Blöcke durch

$$B_{n,i}^{OBB} := \{X_{n,i}, \dots, X_{n,i+L(n)-1}\}, \quad 1 \le i \le B(n),$$
 (6.26)

für B(n)=k(n)-L(n)+1. Analog zum NBB zieht man nun hieraus r(n) Blöcke $B_{n,1}^*,\ldots,B_{n,r(n)}^*$ mit Zurücklegen und erhält so durch Hintereinanderschreiben die OBB-Stichprobe $X_{n,1}^*,\ldots,X_{n,m(n)}^*$ mit $m(n)=r(n)\cdot L(n)$. Offensichtlich ist das klassische m(n)-Bootstrap-Verfahren für die Blocklänge L(n)=1 enthalten. Die Übertragung ins gewichtete Resampling erfolgt wie oben durch r(n)-maliges Ziehen mit Zurücklegen aus den aufsummierten Beobachtungen

$$Y_{n,i}^{OBB} := \sum_{j=1}^{L(n)} X_{n,i+j-1}, \quad 1 \le i \le B(n), \tag{6.27}$$

innerhalb der Blöcke. Die hierzu gehörenden Gewichte sind dann

$$W_{n,i}^{OBB} = \sqrt{r(n)} \left(\frac{1}{r(n)} M_{n,i}^{OBB} - \frac{1}{B(n)} \right), \quad 1 \le i \le B(n), \tag{6.28}$$

⁷Dabei gelten die Grenzwertsätze bedingt nach den zugehörigen Blöcken.

⁸Die Maximumsbedingung ist beispielsweise erfüllt, falls $L(n) \cdot \max_{i \leq \tilde{k}(n)} |X_{n,i}| \xrightarrow{P} 0$ gilt, was zu Bedingungen an L(n) führt.

wobei $M^{OBB}:=(M_{n,1}^{OBB},\ldots,M_{n,B(n)}^{OBB})\sim \mathcal{M}(r(n),\frac{1}{B(n)})$ diesmal multinomialverteilt zum Stichprobenumfang $r(n)=\sum_{i=1}^{B(n)}M_{n,i}^{OBB}$ mit jeweils gleicher Auswahlwahrscheinlichkeit $\frac{1}{B(n)}$ ist und die Generalvoraussetzungen - wie oben - für den Fall $\min(r(n),B(n))\to\infty$ erfüllt sind.

Durch die Methode des OBB werden die verschiedenen Beobachtungen $X_{n,1}, \ldots, X_{n,k(n)}$ - abhängig von der Wahl der Blocklänge L(n)- unterschiedlich gewichtet, da sie nicht alle in gleich vielen Blöcken enthalten sind. Um dies zu vermeiden, haben Politits und Romano [41, S.263 ff.] 1992 ausgehend vom OBB den Circular-Block-Bootstrap entwickelt.

Beispiel 6.18 (Der Circular-Block-Bootstrap)

Das Verfahren des Circular-Block-Bootstraps (CBB) wurde ursprünglich für stationäre Zufallsvariablen konzipiert, lässt sich aber auch auf beliebige Dreieckschema erweitern. Die grundlegende Idee besteht dabei darin, das OBB Verfahren mit dem "balanced bootstrap resampling" von DAVISON, HINKLEY UND SCHECHTMANN [11] zu kombinieren, d.h. man möchte jede Beobachtung $X_{n,i}$ in gleich vielen Blöcken wiederfinden. Dazu vervielfältigt man die Ausgangsstichprobe $X_{n,1}, \ldots, X_{n,k(n)}$ wie folgt

$$Z_{n,0} := X_{n,k(n)}, \ Z_{n,i} := X_{n,i \mod k(n)}, \quad i \in \mathbb{N}.$$
 (6.29)

Analog zum OBB defniert man nun hierzu Blöcke der Länge L(n) durch

$$B_{n,i}^{CBB} := \{ Z_{n,i}, \dots, Z_{n,i+L(n)-1} \}, \quad 1 \le i \le k(n)$$
 (6.30)

mit dem Unterschied, dass die Anzahl der Blöcke hier nicht von der Länge L(n) abhängt. Dies bewirkt gerade den gewünschten Gleichmäßigkeitseffekt⁹. Die CBB-Stichprobe erhält man nun wieder, indem man hieraus r(n) Blöcke $B_{n,1}^*, \ldots, B_{n,r(n)}^*$ mit Zurücklegen zieht und die zugehörigen Beobachtungen $Z_{n,1}^*, \ldots, Z_{n,m(n)}^*$ mit $m(n) = r(n) \cdot L(n)$ hintereinanderschreibt.

Der CBB wird genauso wie die beiden anderen Verfahren ins gewichtete Resampling implementiert, indem man r(n) Mal mit Zurücklegen aus den innerhalb der

⁹An dieser Stelle ist zu bemerken, dass hierdurch weit voneinander entfernte Beobachtungen wie $X_{n,k(n)}$ und $X_{n,1}$, die z.B. im m-dependent Fall auch unabhängig sein können, in mehreren Blöcken aufeinander folgen.

Blöcke aufsummierten Beobachtungen

$$Y_{n,i}^{CBB} := \sum_{j=1}^{L(n)} Z_{n,i+j-1}, \quad 1 \le i \le k(n), \tag{6.31}$$

zieht. Die hierzu gehörigen Gewichte schreiben sich nun als

$$W_{n,i}^{CBB} = \sqrt{r(n)} \left(\frac{1}{r(n)} M_{n,i} - \frac{1}{k(n)} \right), \quad 1 \le i \le k(n).$$
 (6.32)

Dabei ist $M^{CBB}:=(M_{n,1}^{CBB},\ldots,M_{n,k(n)}^{CBB})\sim \mathcal{M}(r(n),\frac{1}{k(n)})$ multinomialverteilt zum Stichprobenumfang $r(n)=\sum_{i=1}^{k(n)}M_{n,i}^{CBB}$ mit gleicher Auswahlwahrscheinlichkeit $\frac{1}{k(n)}$ und die Generalvoraussetzungen sind - wie oben - für den Fall $\min(r(n),k(n))\to\infty$ erfüllt.

Bemerkung 6.19

- 1. Die optimale Wahl der Blocklänge L(n) in den drei obigen Beispielen hängt von der konkreten Situation ab. Will man beispielsweise bei einem Autoregressiven Prozess der Ordnung 1 die Varianz $\operatorname{Var}(e_1)$ des zugehörigen White-Noise-Prozesses $(e_i)_i$ mittels des NBB schätzen, so hat Carlstein (vgl. [20, S.443]) gezeigt, dass eine Wahl von $L(n) \propto n^{1/3}$ optimal ist. Dabei bedeutet optimal in diesem Zusammenhang, dass durch dieses L(n) asymptotisch der kleinste Quadrate Fehler des NBB-Varianzschätzers minimiert wird. Weitere Fälle finden sich bei HÄRDLE ET AL. [20, Abschnitt 3.2]. Dort liegt die Konvergenzrate von L(n) je nach Situation zwischen $n^{1/3}$ und $n^{1/6}$.
- 2. Aufgrund der "universellen" Anwendbarkeit haben einige Autoren vorgeschlagen, das Subsampling Verfahren (s.o.) bei abhängigen Daten anzuwenden, um die Block-Bootstrap-Methode zu umgehen. Allerdings zeigt sich, dass die verschiedenen Block-Bootstrap-Verfahren im Vergleich zum Subsampling in den bisherigen Fällen bessere Konvergenzraten liefern, vgl. HÄRDLE ET AL. [20, Abschnitt 3.3].
- 3. Analog zur obigen Implementierung der bekannten Block-Bootstrap-Verfahren, lassen sich anhand der vorangegangenen Gewichte weitere Block-Bootstrap-Verfahren entwickeln. Beispielsweise könnte man Blockpermutations-, m(n)-double-Block-Bootstrap- oder auch Wild-Block-Bootstrap-Verfahren untersuchen. □

Im Folgenden werden einige Resampling-Verfahren vorgestellt, die zwar praktische Anwendungen besitzen, sich aber (noch) nicht als Gewichte von obigem Typ schreiben lassen.

Beispiel 6.20 (Der Antithetic Bootstrap)

Die Grundidee dieses von Hall [26, S.296f] vorgestellten Verfahrens besteht in der allgemeinen Tatsache, dass zwei negativ korrellierte Schätzer h_1 und h_2 für ein statistisches Funktional g mit gleichen Erwartungswerten $E(h_1) = E(h_2)$ und Varianzen $Var(h_1) = Var(h_2)$ zu einem besseren Schätzer

$$h_3 := \frac{1}{2}(h_1 + h_2)$$

kombiniert werden können. Es gilt nämlich

$$Var(h_3) = \frac{1}{4}(Var(h_1) + Var(h_2) + Cov(h_1, h_2)) < \frac{1}{2}Var(h_1).$$
 (6.33)

Die Übertragung auf den Bootstrap funktioniert wie folgt:

- 1. Ziehe eine Bootstrap-Stichprobe $\mathbb{X}^* := \{X_{n,1}^*, \dots, X_{n,m(n)}^*\}$ aus. $\mathbb{X} := \{X_{n,1}, \dots, X_{n,k(n)}\}$
- 2. Bezeichne mit j_i den zu $X_{n,i}^*$ gehörenden Index, d.h. $X_{n,i}^* = X_{n,j_i}$ und erzeuge mit Hilfe einer Permutation $\tau \in \mathcal{S}_{k(n)}$ eine zweite Resampling-Stichprobe $\mathbb{X}^{**} := \{X_{n,1}^{**}, \dots, X_{n,m(n)}^{**}\}$ mit $X_{n,i}^{**} = X_{n,\tau(j_i)}$. Diese besitzt bedingt unter \mathbb{X} die gleiche Verteilung wie die Bootstrap-Stichprobe.
- 3. Ersetze nun einen gegeben Schätzer $h(\mathbb{X})$ durch $h_A := \frac{1}{2}(h(\mathbb{X}^*) + h(\mathbb{X}^{**}))$.

Dabei ist die Permutation τ zu wählen, die die Kovarianz Cov $(h(\mathbb{X}^*), h(\mathbb{X}^{**}))$ von $h(\mathbb{X}^*)$ und $h(\mathbb{X}^{**})$ minimiert. Hall zeigt, dass diese in den meisten praktischen Situationen durch die Permutation gegeben ist, die für $1 \leq i \leq k(n)$ den it-größten Wert $X_{n,k(n)-i+1:k(n)}$ der Stichprobe mit dem it-kleinsten $X_{n,i:k(n)}$ vertauscht. Sind die Beobachtungen fast sicher verschieden, so lässt sich dies mit Antirängen $d_i := d_i((X_{n,i})_{i \leq k(n)})$ wie folgt formulieren

$$S_n \ni \tau = \tau((X_{n,i})_{i \le k(n)}) : d_i \mapsto d_{k(n)-i+1}.$$

Die hierzu gehörenden Gewichte können mit obiger Permutation wie folgt definiert werden

$$W_{n,i} = \sqrt{2 \cdot m(n)} \left(\frac{M_{n,i} + M_{n,\tau^{-1}(i)}}{2m(n)} - \frac{1}{k(n)} \right), \quad 1 \le i \le k(n).$$
 (6.34)

Dabei ist $M := (M_{n,1}, \ldots, M_{n,k(n)})$, analog zu den m(n)-Bootstrap-Gewichten (2.7), eine multinomial- $\mathcal{M}(m(n), \frac{1}{k(n)})$ -verteilte Zufallsvariable. Die hierzu gehörende Resampling-Statistik ist mit obigen Bezeichnungen gerade

$$T_n^* = k(n)^{1/2} \sum_{i=1}^{k(n)} W_{n,i} X_{n,i} = \sqrt{2m(n)k(n)} \left(\sum_{i=1}^{k(n)} \frac{(M_{n,i} + M_{n,\tau^{-1}(i)}) X_{n,i}}{2m(n)} - \overline{X}_n \right)$$
$$= \sqrt{2m(n)k(n)} \left(\frac{1}{m(n)} \sum_{i=1}^{m(n)} \frac{X_{n,i}^* + X_{n,i}^{**}}{2} - \overline{X}_n \right).$$

Da $M_{n,i}$ und $M_{n,\tau^{-1}(i)}$ jeweils Binomial- $\mathcal{B}(m(n),\frac{1}{k(n)})$ -verteilt sind, kann man analog zu den Gewichten (2.7) des m(n)-Bootstraps zeigen, dass die anthitetischen Gewichte die Bedingungen (6.2)-(6.4) erfüllen. Jedoch ist zum einen die Voraussetzung der Unabhängigkeit von den Daten und zum anderen die Austauschbarkeitsvoraussetzung (6.1) verletzt, da τ eine Involution ist.

Auch das nächste Verfahren scheitert an denselben Voraussetzungen.

Beispiel 6.21 (Importance Resampling)

Dieses Verfahren stellt eine Verallgemeinerung des m(n)-Bootstraps dar und hat bei der Schätzung von Verteilungs- und Quantilfunktionen eine gewisse Bedeutung, vgl. Hall [23, S.298ff].

Dabei gewichtet man die verschiedenen Beobachtungen der Ausgangsstichprobe unterschiedlich, indem man jeder Beobachtung $X_{n,i}$ anstelle von $\frac{1}{k(n)}$ eine bestimmte Auswahlwahrscheinlichkeit $p_{n,i} \in [0,1], \sum_{i=1}^{k(n)} p_{n,i} = 1$, zuordnet. Auf diese Weise erhält man Gewichte der Form (2.7) mit dem Unterschied, dass $M := (M_{n,1}, \ldots, M_{n,k(n)})$ nun multinomial- $\mathcal{M}(m(n), p_{n,1}, \ldots, p_{n,k})$ -verteilt ist. Hierfür lassen sich unter geeigneten Voraussetzungen an die Auswahlwahrscheinlichkeiten zwar die Bedingungen (6.2)-(6.4) ableiten, jedoch ist die Austauschbarkeitsbedingung (6.1) nur dann erfüllt, wenn alle $p_{n,i}$ identisch, d.h. gleich 1/k(n) sind.

Das nächste Beispiel befasst sich mit der Variabilität der Resampling-Stichprobe.

Beispiel 6.22 (Der sequentielle Bootstrap)

Zieht man verschiedene Bootstrap-Stichproben B_1^*, \ldots, B_k^* nach dem klassischen Verfahren von Efron - oder auch nach der m(n)-Bootstrap-Methode - so ist die Information dieser Stichproben, gemessen durch die Anzahl der verschiede-

nen Elemente, nicht identisch. Da diese Variabilität nicht unbedingt wünschenswert ist, vgl. JIMÉNEZ-GAMARO ET. AL. [37], haben einige Autoren Verfahren entwickelt, die diese Informationsschwankungen verhindern sollen. Hierzu zählt u.a. der **sequentielle Bootstrap** von RAO ET. AL. [54]. Dabei wird die Resampling-Stichprobe $\{X_1^*, \ldots, X_{N_m}^*\}$ erzeugt, indem man so lange mit Zurücklegen aus der Ausgangsstichprobe $\{X_1, \ldots, X_n\}$ zieht, bis $m \leq n$ verschiedene Beobachtungen aufgetreten sind. RAO ET AL. [54] zeigen dabei in ihrer Arbeit, dass die Wahl $m := |n(1 - e^{-1})|$ gewisse Vorteile besitzt.

Da sich die Anzahl N_m der Elemente in der Resampling-Stichprobe als Summe von unabhängigen, geometrisch verteilten Zufallsvariablen¹¹

$$N_m = I_1 + \cdots + I_m$$

schreiben lässt, vgl. [54, (2.3)], ist dieses Verfahren in gewisserweise der Wild-Bootstrap-Methodik zuzuordnen. Ob die hierzugehörenden Resampling-Gewichte den Generalvoraussetzungen genügen, muß jedoch erst noch in Nachfolgearbeiten geklärt werden.

Neben den oben bereits aufgezählten Verfahren existieren in der Literatur viele weitere Resampling-Methoden.¹² Ob sich diese allerdings auch in einen noch allgemeineren gewichteten Resampling-Ansatz integrieren lassen, muss sich erst noch zeigen. Die größte Problematik stellt hierbei die Abschwächung der Unabhängigkeits- und Austauschbarkeitsbedingung an die Gewichte dar, da diese für die Beweise¹³ der obigen Verfahren essentiell sind.

¹⁰Ein weiteres Verfahren, das dieselbe Zielsetzung verfolgt, ist das sog. Reduced-Bootstrap-Verfahren von JIMÉNEZ-GAMARO ET. AL. [37].

¹¹Hierbei ist $I_1 = 1$ (vom letzten Zug) sowie $\tilde{P}(I_j = k) = \left(1 - \frac{j-1}{n}\right) \left(\frac{j-1}{n}\right)^{k-1}$ für $2 \leq j \leq m, k \in \mathbb{N}$.

¹²Zu den bekannteren zählen beispielsweise noch das **Balanced Bootstrap-Verfahren**, bei dem man jede Beobachtung gleichhäufig in der Bootstrap-Stichprobe haben möchte, sowie der *prepivoted Bootstrap*, der eine Mischung aus m(n)-double- und m(n)-Bootstrap darstellt und sich auch iterativ fortsetzen lässt, vgl. jeweils Kapitel 9.2.1. bei Pauls.

¹³Diese werden in den oben zitierten Arbeiten aus Zentralen Grenzwersätzen von Rangstatistiken unter der Hypothese gleichverteilter Ränge abgeleitet.

Anhang A

Appendix

A.1 Schwache Konvergenz und bedingte Verteilungen

In diesem Abschnitt werden einige Aussagen über die schwache Konvergenz von bedingten und unbedingten Verteilungen angegeben. Dabei seien stets ein Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) sowie polnische Räume - d.h. vollständig metrisierbare und separable Räume - $(\mathbb{E}, \mathcal{E}, \tilde{d}_{\mathbb{E}})$ und $(\mathbb{F}, \mathcal{F}, \tilde{d}_{\mathbb{F}})$ mit Borel- σ -Algebren \mathcal{E} bzw. \mathcal{F} gegeben.

Der folgende Satz über stetige Abbildungen ist eines der wesentlichen Hilfsmittel zum Nachweis von Verteilungkonvergenz.

Satz A.1

Seien $X_n: (\Omega, \mathcal{A}, P) \to (\mathbb{E}, \mathcal{E})$ Zufallsvariablen und $g: \mathbb{E} \to \mathbb{F}$ eine messbare, P^{X_0} -fast sicher stetige Funktion. Dann folgt aus $X_n \xrightarrow{\mathcal{D}} X_0$ auch $g(X_n) \xrightarrow{\mathcal{D}} g(X_0)$.

Beweis. Vgl. Witting und Müller-Funk [66, Satz 5.43]. ■

Da $(\mathbb{E}, \mathcal{E}, \tilde{d}_{\mathbb{E}})$ und $(\mathbb{F}, \mathcal{F}, \tilde{d}_{\mathbb{F}})$ polnisch sind, ist auch der Produktraum $(\mathbb{E} \times \mathbb{F}, \mathcal{E} \otimes \mathcal{F}, \tilde{d}_{\mathbb{E} \times \mathbb{F}})$ polnisch mit der kanonischen Produktmetrik

$$\tilde{d}_{\mathbb{E}\times\mathbb{F}}(x,y) = \tilde{d}_{\mathbb{E}}(x_1,y_1) + \tilde{d}_{\mathbb{F}}(x_2,y_2)$$

für $x=(x_1,x_2),y=(y_1,y_2)\in\mathbb{E}\times\mathbb{F}$, vgl. Bauer [3, Beispiel 2, S.179].

Das folgende Lemma ist eine Verallgemeinerung des Hilfssatzes 5.80 b) von WITTING UND MÜLLER-FUNK [66, S.74].

Lemma A.2

Sind $Y_n, Y : \Omega \to \mathbb{E}$ und $Z_n : \Omega \to \mathbb{F}$ Zufallsvariablen mit $Y_n \xrightarrow{\mathcal{D}} Y$ und $Z_n \xrightarrow{P} c \in \mathbb{F}$, so liegt auch Verteilungskonvergenz auf dem Produktraum vor

$$(Y_n, Z_n) \xrightarrow{\mathcal{D}} (Y, c).$$

Beweis. Aus den Voraussetzungen folgt zum einen

$$(Y_n,c) \xrightarrow{\mathcal{D}} (Y,c)$$

und zum anderen

$$P(\tilde{d}_{\mathbb{E}\times\mathbb{F}}((Y_n,c),(Y_n,Z_n))>\epsilon)\to 0,$$

so dass die Behauptung durch Theorem 4.1 von Billingsley [6, S.25] impliziert wird.

Als nächstes wird die schwache Konvergenz von bedingten Verteilungen untersucht.

Hierzu seien $T_n: (\Omega, \mathcal{A}) \to (\mathbb{E}, \mathcal{E})$ eine Statistik und $\mathcal{F}_n \subset \mathcal{A}$ eine Teil- σ -Algebra für $n \in \mathbb{N}$. Dann interessiert man sich häufig für das Konvergenzverhalten der bedingten Verteilungen $\mathcal{L}(T_n|\mathcal{F}_n)$. Dabei ist die Existenz dieser bedingten Verteilung gesichert, da T_n eine Statistik mit Werten in dem polnischen Raum \mathbb{E} ist, vgl. z.B. BAUER [4, Satz 44.3].

Unter der P-fast sicheren Verteilungskonvergenz von T_n gegeben \mathcal{F}_n gegen eine Statistik T_0 mit Verteilung μ , i.Z. $\mathcal{L}(T_n|\mathcal{F}_n) \xrightarrow{f.s.} \mathcal{L}(T_0) = \mu$, versteht man die schwache Konvergenz

$$\mathcal{L}(T_n|\mathcal{F}_n)(\omega) \xrightarrow{w} \mu \tag{A.1}$$

für P-fast alle $\omega \in \Omega$. Ersetzt man nun die fast sichere durch stochastische Konvergenz, so ist es aus Gründen der Übersichtlichkeit geschickter (vgl. z.B. GINÉ [19, S.41 f.]), die schwache Konvergenz mittels einer Metrik $d_{\mathbb{E}}$ auf dem Raum der Wahrscheinlichkeitsmaße $\mathcal{M}_1(\mathbb{E}, \mathcal{E})$ darzustellen. Damit ist

$$d_{\mathbb{E}}(\mathcal{L}(T_n|\mathcal{F}_n), \mu) \xrightarrow{P} 0$$
 (A.2)

nach dem Teilfolgenkriterium äquivalent dazu, dass für jede Teilfolge eine weitere Teilfolge existiert, entlang derer (A.1) für P-fast alle $\omega \in \Omega$ gilt. Die Existenz einer solchen Metrik $d_{\mathbb{E}}$ ist allgemein auf polnischen Räumen gesichert, vgl. WITTING UND MÜLLER-FUNK [66, S. 50].

Mit Hilfe dieser Konvention lässt sich nun der folgende Satz beweisen.

Satz A.3

Für alle $n \in \mathbb{N}$ seien $T_n : (\Omega, \mathcal{A}) \to (\mathbb{E}, \mathcal{E})$ eine Statistik, $\mathcal{T}_n \subset \mathcal{F}_n \subset \mathcal{A}$ Teil- σ -Algebren und $\mu \in \mathcal{M}_1(\mathbb{E}, \mathcal{E})$ ein Wahrscheinlichkeitsmaß. Dann folgt aus der unter \mathcal{F}_n bedingten Verteilungskonvergenz (A.2) schon die unter \mathcal{T}_n bedingte Verteilungskonvergenz

$$d_{\mathbb{E}}(\mathcal{L}(T_n|\mathcal{T}_n), \mu) \xrightarrow{P} 0.$$
 (A.3)

Beweis. Durch Anwendung des Teilfolgenkriteriums für stochastische Konvergenz erhält man zu jeder Teilfolge eine weitere Teilfolge - die im folgenden mit $\{n\}$ bezeichnet wird - entlang derer die Konvergenz in (A.2) fast sicher ist. Sei nun $f: \mathbb{E} \to \mathbb{R}$ stetig und beschränkt. Dann gilt folgende Gleichheit fast sicher

$$\int f dP^{T_n|\mathcal{F}_n} = E(f \circ T_n|\mathcal{F}_n),$$

vgl. z.B. Kapitel 10.2 von Dudley [8]. Hieraus folgt aufgrund von (A.2) die folgende fast sichere Konvergenz

$$E(f \circ T_n | \mathcal{F}_n) \xrightarrow{f.s.} \int f d\mu.$$

Da f beschränkt ist, liegt sogar L_1 -Konvergenz vor

$$\int \Big| E(f \circ T_n | \mathcal{F}_n) - \int f d\mu \Big| dP \to 0.$$

Die Behauptung folgt nun mit Hilfe der Jensenschen Ungleichung

$$\int |E(f \circ T_n - \int f d\mu | \mathcal{T}_n)| dP = \int |E(E(f \circ T_n - \int f d\mu | \mathcal{F}_n) | \mathcal{T}_n)| dP$$

$$\leq \int E(|E(f \circ T_n - \int f d\mu | \mathcal{F}_n) | |\mathcal{T}_n) dP$$

$$= \int |E(f \circ T_n - \int f d\mu | \mathcal{F}_n)| dP \to 0$$

und einer erneuten Anwendung des Teilfolgenkriteriums.

Als Korollar erhält man den Zusammenhang zwischen bedingter und unbedingter Konvergenz.

Korollar A.4

Unter den Voraussetzungen von Satz A.3 folgt aus der bedingten schwachen Konvergenz in (A.2) schon die unbedingte schwache Konvergenz

$$\mathcal{L}(T_n) \xrightarrow{w} \mu.$$
 (A.4)

Beweis. Setzt man $\mathcal{T}_n := \{\Omega, \emptyset \}$ für alle $n \in \mathbb{N}$, so folgt dies aus Satz A.3.

Um nun den direkten Bezug zu den behandelten Resampling-Verfahren herzustellen, wird im Folgenden vorausgesetzt, dass $(X_{n,i})_{i \leq k(n)}$ mit $X_{n,i}:(\Omega,\mathcal{A},P) \to (\mathbb{E},\mathcal{E})$ ein Dreiecksschema von Zufallsvariablen ist, welches unabhängig¹ vom reellen Dreiecksschema $(W_{n,i})_{1\leq i\leq k(n)}$ mit $W_{n,i}:(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{P})\to(\mathbb{R},\mathbb{B})$ sei. Man interessiert sich nun für das Konvergenzverhalten der bedingten Verteilungen $\mathcal{L}(T_n|\vec{X}_n)$, wobei $\vec{X}_n:=(X_{n,i})_{i\leq k(n)}$ der Vektor der Daten und $T_n:=T_n((\vec{X}_n,(W_{n,i})_{i\leq k(n)})):(\Omega\times\tilde{\Omega},\mathcal{A}\otimes\tilde{\mathcal{A}},P\otimes\tilde{P})\to(\mathbb{E},\mathcal{E})$ eine Statistik seien². Für diesen Fall fasst das folgende Lemma einige wichtige Aussagen über die schwache Konvergenz der obigen bedingten Verteilungen zusammen. Dabei seien $\mathcal{F}_n:=\sigma(\vec{X}_n)$ die von \vec{X}_n erzeugte σ -Algebra, T_0 eine weitere Statistik mit Verteilung $\mathcal{L}(T_0)=\mu$ und $d_{\mathbb{F}}$ sowie $d_{\mathbb{E}\times\mathbb{F}}$ Metriken, die die schwache Konvergenz auf den Räumen $\mathcal{M}_1(\mathbb{F},\mathcal{F})$ bzw. $\mathcal{M}_1(\mathbb{E}\times\mathbb{F},\mathcal{E}\otimes\mathcal{F})$ metrisieren.

Lemma A.5

Unter der Voraussetzung (A.2) gelten die folgenden Aussagen:

(i) Sei $f : \mathbb{E} \to \mathbb{F}$ eine messbare Abbildung, so dass die Menge der Unstetigkeitsstellen von f eine $\mathcal{L}(T_0)$ Nullmenge ist, so gilt

$$d_{\mathbb{F}}(\mathcal{L}(f \circ T_n | \vec{X}_n), \mathcal{L}(f \circ T_0)) \xrightarrow{P} 0.$$

(ii) Sei $R_n := R_n((X_{n,i})_{i \le k(n)}, (W_{n,i})_{i \le k(n)})$ eine weitere Statistik mit Werten in \mathbb{F} , die stochastisch gegen eine Konstante $c \in \mathbb{F}$ konvergiere, d.h. $R_n \xrightarrow{P \otimes \tilde{P}} c$. Dann gilt

$$d_{\mathbb{E}\times\mathbb{F}}(\mathcal{L}((T_n,R_n)|\vec{X}_n),\mathcal{L}(T_0)\otimes\varepsilon_c)\stackrel{P}{\longrightarrow} 0.$$

Beweis. (i) Aufgrund der fast sicheren Eindeutigkeit der bedingten Verteilung, erhält man aus der definierenden Gleichung für alle $A \in \mathcal{F}$ die Gleichung

$$\mathcal{L}(f \circ T_n | \vec{X}_n = \cdot)(A) = \mathcal{L}(T_n | \vec{X}_n = \cdot)(f^{-1}(A)). \tag{A.5}$$

fast sicher. Nach (A.2) existiert nun für jede Teilfolge eine weitere Teilfolge $\{m\}$ und eine P-1-Menge M, so dass für alle $\omega \in M$ die Konvergenz in (A.1) entlang

¹Dies ist wie in Kapitel 3 als Unabhängigkeit auf dem Produktraum $(\Omega \times \tilde{\Omega}, \mathcal{A} \otimes \tilde{\mathcal{A}}, P \otimes \tilde{P})$ zu verstehen.

 $^{^2}$ Für Resampling-Statistiken $T_n=T_n^*$ benötigt man zusätzlich, dass $\mathbb E$ ein $\mathbb R$ -Vektorraum ist. Dies ist aber z.B. für $\mathbb E=\mathbb R^I$ mit abzählbarer Indexmenge I erfüllt.

 $\{m\}$ vorliegt. Für festes $\omega \in M$ kann man somit hierauf Satz A.1 anwenden und erhält so (i).

Für den Beweis von (ii) findet man nach Voraussetzung für jede Teilfolge eine weitere Teilfolge $\{m\}$ und eine $(P \otimes \tilde{P})$ -1-Menge M, so dass $R_m(\omega, \tilde{\omega})$ für alle $(\omega, \tilde{\omega}) \in M$ gegen c konvergiert. Aufgrund von³

$$1 = P \otimes \tilde{P}(M) = \int \tilde{P}(M_{\omega}) dP(\omega)$$

ist $M_1 := \{\omega : \tilde{P}(M_\omega) = 1\}$ eine P-1-Menge, so dass

$$R_m((X_{m,i}(\omega))_{i < k(m)}, (W_{m,i})_{i < k(m)}) \to c \quad \tilde{P}$$
- fast sicher (A.6)

für alle $\omega \in M_1$ gilt. Genauso findet man für diese Teilfolge $\{m\}$ eine weitere Teilfolge $\{m_1\}$ und eine P-1-Menge M_2 , so dass (A.1) und (A.6) entlang $\{m_1\}$ für jedes feste $\omega \in M_1 \cap M_2$ gelten. Somit kann man auf $\mathcal{L}((T_{m_1}, R_{m_1})|\vec{X}_{m_1} = \vec{X}_{m_1}(\omega))$ das unbedingte Resultat in Form von Lemma A.2 für festes ω anwenden und erhält so die Behauptung mit Hilfe des Teilfolgenkriteriums.

Bemerkung A.6

Der Beweis von (i) funktioniert auch für allgemeine Statistiken T_n , die nicht von der speziellen Gestalt $T_n = T_n(\vec{X}_n, (W_{n,i})_{i \leq n})$ sind.

Außerdem erhält man ein Resultat für das Verhalten von Resttermen bei Permutationsverfahren. Diese verhalten sich nämlich unter Austauschbarkeit genauso wie im unbedingten Fall.

Lemma A.7

Sei $\tau = (\tau_1, \dots, \tau_n) : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}) \to \mathcal{S}_n$ eine gleichverteilte Permutation, welche unabhängig von den zeilenweise austauschbaren Zufallsvariablen $(X_{n,i})_{i \leq n}$ sei. Dann folgt aus $R_n((X_{n,i})_{i \leq n}) \xrightarrow{P} c$ schon

$$R_n((X_{n,\tau_i})_{i\leq n}) \stackrel{P\otimes \tilde{P}}{\longrightarrow} c.$$

Beweis. Ohne Einschränkung sei c=0. Aufgrund der Austauschbarkeit erhält man für jedes feste $\tilde{\omega}\in\tilde{\Omega}$

$$R_n((X_{n,\tau_i(\tilde{\omega})})_{i < n}) \stackrel{\mathcal{D}}{=} R_n((X_{n,i})_{i < n}) \stackrel{P}{\longrightarrow} 0.$$

Hieraus folgt unmittelbar die Behauptung.

³Die Menge $M_{\omega} := \{ \tilde{\omega} \in \tilde{\Omega} : (\omega, \tilde{\omega}) \in M \}$ ist der ω -Schnitt von M.

Bemerkung A.8

Das Lemma bleibt auch für allgemeine Randomisationsverfahren gültig, wenn man die Austauschbarkeitsbedingung durch eine Invarianzbedingung ersetzt. Im Fall von Symmetrietests wäre dies beispielsweise durch die 0-Symmetrie gegeben.

Als Nächstes wird ein bedingtes Lemma von Slutzky bewiesen. Hierfür benötigt man, dass \mathbb{E} zusätzlich ein topologischer \mathbb{R} -Vektorraum ist, d.h. die Addition $\mathbb{E} \times \mathbb{E} \ni (x,y) \mapsto x+y \in \mathbb{E}$ und die Skalarmultiplikation $\mathbb{R} \times \mathbb{E} \ni (\lambda,x) \mapsto \lambda x \in \mathbb{E}$ sind bzgl. der gewählten Topologie stetige Abbildungen.

Lemma A.9

Sei $\mathbb E$ ein topologischer $\mathbb R$ -Vektorraum. Dann erhält man unter den Voraussetzungen von Lemma A.5 die folgenden Aussagen

(i)
$$d_{\mathbb{E}}(\mathcal{L}(T_n + R_n | \vec{X}_n), \mathcal{L}(T_0 + c)) \xrightarrow{P} 0$$
, falls $c \in \mathbb{E} = \mathbb{F}$.

(ii)
$$d_{\mathbb{E}}(\mathcal{L}(R_nT_n|\vec{X}_n), \mathcal{L}(cT_0)) \xrightarrow{P} 0$$
, falls $c \in \mathbb{R} = \mathbb{F}$.

Beweis. Die Behauptungen folgen, wenn man die Aussage (i) aus Lemma A.5 jeweils für die Addition und Skalarmultiplikation auf (ii) anwendet.

Ist \mathbb{E} ein endlich-dimensionaler euklidischer Raum, so wird häufig auch die stärkere Konvergenz in Integralmetriken untersucht, vgl. z.B. BICKEL UND FREED-MAN [5]. Der Einfachheit halber beschränken wir uns an dieser Stelle auf $\mathbb{E} = \mathbb{R}^k$ und die Mallows-Metrik d_{M2} . Diese wird zunächst definiert.

Definition A.10

Seien $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^k, \mathcal{B}^k)$ mit $\int ||x||^2 d(\mu + \nu)(x) < \infty$. Hierzu definiert man die Menge

$$I_{\mu} = \{X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^k, \mathcal{B}^k) \text{ ist Zufallsvariable mit } P^X = \mu\}.$$

Dann ist der Mallows-Abstand von μ und ν definiert als

$$d_{M2}(\mu,\nu) = \inf_{(X,Y)\in I_{\mu}\times I_{\nu}} \left(\int (\|X - Y\|^2 dP)^{1/2}. \right)$$
(A.7)

Der Nachweis der Metrik-Eigenschaft sowie den folgenden Zusammenhang zur schwachen Konvergenz findet man im Appendix von BICKEL UND FREEDMAN [5].

Anhang______ 130

Lemma A.11

Sei $\mu_n \in \mathcal{M}_1(\mathbb{R}^k, \mathcal{B}^k)$ mit $\int ||x||^2 d\mu_n(x) < \infty$ für alle $n \in \mathbb{N}_0$. Dann sind äquivalent:

$$d_{M2}(\mu_n, \mu_0) \to 0, \tag{A.8}$$

$$\mu_n \xrightarrow{w} \mu_0 \ und \int ||x||^2 d\mu_n(x) \to \int ||x||^2 d\mu_0(x).$$
 (A.9)

Beweis. Dies folgt aus Lemma 8.3. von Bickel und Freedman [5].

Weitere Eigenschaften, wie z.B. ein Analogon zum Satz über stetige Abbildung unter geeigneten Lipschitz Bedingungen findet man ebenfalls im Appendix von BICKEL UND FREEDMAN [5].

A.2 Stationäre und ergodische Zufallsvariablen

Der vorliegende zweite Teil des Appendix fasst in Kürze die für das Kapitel 3 benötigte Theorie aus dem Buch von Stout [60] über stationäre und ergodische Zufallsvariablen zusammen. Dabei sei stets ein Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) gegeben.

Definition A.12

(a) Eine Folge von Zufallsvariablen $X_n : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}), n \in \mathbb{N}$, heißt stark (oder auch strikt) $station\ddot{a}r$, falls für jedes $k \in \mathbb{N}$ und $B \in \mathcal{B}^{\mathbb{N}}$ gilt

$$P((X_1, X_2, \dots) \in B) = P((X_k, X_{k+1}, \dots) \in B),$$
 (A.10)

 $d.h. X_1, X_2, \ldots$ und X_k, X_{k+1}, \ldots besitzen die gleiche Verteilung.

- (b) Sei $(X_n)_{n\in\mathbb{N}}$ eine stark stationäre Folge. Eine Menge $A\in\mathcal{A}$, die die Darstellung $A=\{(X_1,X_2,\dots)\in B\}$ für ein $B\in\mathcal{B}^\mathbb{N}$ besitzt, heißt invariant bzgl. $(X_n)_{n\in\mathbb{N}}$, falls $A=\{(X_k,X_{k+1},\dots)\in B\}$ für alle $k\in\mathbb{N}$ gilt. Die hierzu gehörige σ -Algebra⁴ der bzgl. $(X_n)_{n\in\mathbb{N}}$ invarianten Mengen wird mit \mathcal{H} bezeichnet.
- (c) Die stark stationäre Folge $(X_n)_{n\in\mathbb{N}}$ heißt **ergodisch**, falls jedes invariante Ereignis die Wahrscheinlichkeit 0 oder 1 besitzt.

⁴Es lässt sich leicht nachprüfen, dass die Menge aller bzgl. $(X_n)_{n\in\mathbb{N}}$ invarianten Mengen eine σ -Algebra bildet.

Anhang______ 131

Bemerkung A.13

1. Ist $(X_n)_{n\in\mathbb{N}}$ eine stark stationäre (oder auch ergodische) Folge, so gilt dies auch für $(X_n^2)_{n\in\mathbb{N}}$.

2. Insbesondere ist jede i.i.d. Folge $(X_n)_{n\in\mathbb{N}}$ stationär und ergodisch.

Beweis. Der erste Teil folgt aus den Theoremen 3.5.3. und Theorem 3.5.8. bei Stout [60] und der zweite Teil ist eine Folgerung aus dem 0-1-Gesetz von Kolmogoroff, vgl. ebenfalls Stout [60, Lemma 3.5.8.].

Nachdem nun alle notwendigen Begriffe eingeführt wurden, kann das Hauptresultat - ein starkes Gesetz der grossen Zahlen für stationäre und ergodische Zufallsvariablen - zitiert werden.

Satz A.14

Sei $(X_n)_{n\in\mathbb{N}}$ eine stark stationäre Folge mit $E(|X_1|) < \infty$, so gilt das folgende starke Gesetz der grossen Zahlen⁵

$$\frac{1}{n}\sum_{i=1}^{n}X_{i} \to E(X_{1}|\mathcal{H}) \quad P\text{-}fast \ sicher. \tag{A.11}$$

Ist $(X_n)_{n\in\mathbb{N}}$ sogar ergodisch, so gilt

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to E(X_1) \quad P\text{-}fast \ sicher. \tag{A.12}$$

Beweis. Vgl. Theorem 3.5.7. bei Stout [60].

⁵Da \mathcal{H} eine Unter- σ -Algebra von \mathcal{A} und X_1 integrierbar ist, existiert der bedingte Erwartungswert $E(X_1|\mathcal{H})$.

Symbol- und Abkürzungsverzeichnis

Im Folgenden werden die in dieser Arbeit verwendeten Symbole und Abkürzungen erläutert.

Symbole

:=	wird definiert als	
$\forall,\exists,\Leftrightarrow,\Rightarrow,\vee$	übliche logische Symbole	
$\in,\not\in,\subset,\cap,\cup,{}^c,\backslash$	übliche Mengenoperationen	
	Ende einer Bemerkung	
	Ende eines Beweises	
$[\cdot]$	Hinweis auf die Literatur mit Nr. \cdot	
${ m I\!N}$	Menge der natürlichen Zahlen	
\mathbb{Z}	Menge der ganzen Zahlen	
$\mathbb Q$	Menge der rationalen Zahlen	
\mathbb{R}, \mathbb{R}^k	Menge der reellen bzw. k-dimensionalen reellen Zahlen	
$\mathbb{R}^{\mathbb{N}}$	Raum der reellen Folgen $(x_n)_{n\in\mathbb{N}}$	
$\mathcal{B},\mathcal{B}^k,\mathcal{B}^{\mathbb{N}}$	Borel'sche Mengen auf $\mathbb R$ bzw. $\mathbb R^k$ bzw. $\mathbb R^{\mathbb N}$	
Id_k	Identität bzw. Einheitsmatrix auf \mathbb{R}^k	
$\mathbb{R}^{m \times n}$	Menge der reellen Matrizen mit m Zeilen und n Spalten	
$\mathbb{R}_{p.s.}^{k \times k}, \mathbb{R}_{p.d.}^{k \times k}$	Menge der symmetrischen, positiv semidefiniten bzw.	
•	positiv definiten $(k \times k)$ -Matrizen über $\mathbb R$	
(Ω, \mathcal{A}, P)	Wahrscheinlichkeitsraum mit Grundmenge Ω ,	
	$\sigma\text{-Algebra }\mathcal{A}$ und Wahrscheinlichkeitsmaß P	

$\mathcal{M}_1(\Omega,\mathcal{A})$	Menge aller Wahrscheinlichkeitsmaße auf dem Messraum (Ω, \mathcal{A})			
$\lceil \cdot \rceil, \lfloor \cdot \rfloor$	Auf- bzw. Abrundungsklammern			
$o(\cdot), \mathcal{O}(\cdot)$	Landau-Symbole			
$o_P(1)$	steht für eine Folge von Zufallsvariablen, die P -stochastisch			
	gegen 0 konvergiert			
$\mathcal{L}(X P) = P^X$	Verteilung von X unter P			
$\mathcal{L}(X Y)$	Bedingte Verteilung von X unter Y bzgl. des			
	Wahrscheinlichkeitsmaßes P			
$X \sim P$	Die Zufallsvariable X besitzt die Verteilung P			
$X_n \stackrel{\mathcal{D}}{\to} X$	X_n konvergiert in Verteilung gegen X			
$X_n \xrightarrow{P} X$	X_n konvergiert P-stochastisch gegen X			
$X_n \xrightarrow{f.s.} X$	X_n konvergiert P-fast sicher gegen X			
$P_n \xrightarrow{w} P$	P_n konvergiert schwach gegen P			
$\mathrm{E}(X)$	Erwartungswert einer Zufallsvariablen X			
Var(X)	Varianz einer reellen Zufallsvariable X			
Cov(X,Y)	Kovarianz der reellen Zufallsvariablen X und Y			
Cov(X)	Kovarianzmatrix der \mathbb{R}^k -wertigen Zufallsvariable X			
$\mathbb{1}_A$	Indikator funktion der Menge ${\cal A}$			
$X_{i:n}$	i -te Orderstatistik der reellen Zufallsvariablen X_1,\dots,X_n			
\mathcal{S}_n	symmetrische Gruppe der Dimension n			
$arepsilon_x$	${ m Einpunktmaß}$			
B(n,p)	Binomialverteilung mit Erfolgswahrscheinlichkeit p			
$\mathcal{P}(\lambda)$	Poissonverteilung zum Parameter λ			
$\mathcal{M}(n,p)$	k-dimensionale Multinomialverteilung zum Stichprobenumfan			
	n mit Auswahlwahrscheinlichkeiten $p=(p_1,\ldots,p_n)$			
$N(\mu,\Gamma)$	k-dimensionale Normalverteilung mit Erwartungswert μ			
	und Kovarianzmatrix Γ			
Φ	Verteilungsfunktion der Standardnormalverteilung			
χ_k^2	Chi-Quadrat-Verteilung mit k Freiheitsgraden			
$\chi^2_{k;lpha}$	Das $(1 - \alpha)$ -Quantil der Chi-Quadrat-Verteilung			
	mit k Freiheitsgraden			

Abkürzungen

bzgl. bezüglich

bzw. beziehungsweise

d.h. das heißt

et al. et alii (und andere)

f. folgende

ff. fortfolgende

f.s. fast sicher

ggü. gegenüber

i.i.d. unabhängig identisch verteilt

S. Seite(n)

sog. sogenannt(e, -er, -es)

vgl. vergleiche z.B. zum Beispiel

Literaturverzeichnis

- [1] Babu, G. J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. Sankhya 46, Ser. A, 85-93.
- [2] BARBE, P. UND BERTAIL, P. (1995). The weighted bootstrap. Springer Verlag, New York.
- [3] BAUER, H. (1992). Maß- und Integrationstheorie. deGruyter, Berlin.
- [4] BAUER, H. (2002). Wahrscheinlichkeitstheorie. deGruyter, Berlin.
- [5] BICKEL, P.J. UND FREEDMANN, D.A. (1981). Some asymptotic theory for the Bootstrap. Ann. Stat. 9, 1196-1217.
- [6] BILLINGSLEY, P. (1968). Convergence of probability measures. John Wiley & Sons, New York.
- [7] CSÖRGŐ, S. UND ROSALSKY, A. (2003). A Survey of Limit Laws for Bootstrapped Sums. International Journal of Mathematics and Mathematica, No.43, 2835-2861.
- [8] DUDLEY, R.M. (2002). Real Analysis and Probability. Cambridge University Press, New York.
- [9] Datta (1995). On a modified bootstrap for certain asymptotically nonnormal statistics. Stat.Probab.Lett. 24, 91-98.
- [10] DAVISON, A.C. UND HINKLEY, D.V. (1997). Bootstrap Methods and their Application. Cambridge University Press, Cambridge.
- [11] DAVISON, A.C.; HINKLEY, D.V. UND SCHECHTMANN E. (1986). Efficient Bootstrap Simulation. Biometrika 73, 555-566.
- [12] DEL BARRIO, E.; JANSSEN, A. UND MATRÁN, C. (2008). Resampling schemes with low resampling intensity and their applications in testing hypotheses. (erscheint in J.Stat.Plan.Inf.).
- [13] DEL BARRIO, E.; JANSSEN, A. UND MATRÁN, C. (2008). On the low intensity bootstrap for triangular arrays of independent identically distributed random variables (erscheint in TEST).

- [14] EFRON, B. (1979). Bootstrap Methods: Another Look at the Jachnife. Ann. Stat. 7, 1-21.
- [15] EFRON, B. UND TIBSHIRANI, R.J. (1993). An Introduction to the Bootstrap. Chapman & Hall, New York.
- [16] EINMAHL, J.H.J. UND ROSALSKY, A. (2005). General Weak Laws of Large Numbers for Bootstrap Sample Means. Stoch.Anal.Appl. 23, No.4, 853-869.
- [17] EINMAHL, U. UND MASON, D. (1992). Approximations to Permutation and Exchangeable Processes. Jour. Theor. Prob. 5, 1, 101-126.
- [18] GÄNSSLER, P. UND STUTE, W. (1977). Wahrscheinlichkeitstheorie. Springer-Verlag, Berlin.
- [19] Giné, E. (1996). Lectures on some aspects of the Bootstrap. Lectures on Probability Theory and Statistics. Giné et al. [Hrsg.], Ecole d'Eté de probabilités de Saint-Flour XXVI-1996. Lecture Notes in Mathematics 1665, 37-151, Springer, New York.
- [20] HÄRDLE, W.; HOROWITZ, J. UND KREISS, J.-P. (2003). Bootstrap-Methods for Time Series. Int.Stat.Rev. 71, No.2, 435-459.
- [21] HÄUSLER, E.; MASON, D.M. UND NEWTON, M.A. (1991). Weighted bootstrapping of means. CWI Q. 4, No.3, 213-228.
- [22] HÁJEK, J.; ŠIDAK, Z. UND SEN, P.K. (1999). Theory of Rank Tests. Academic Press, San Diego.
- [23] Hall, P. (1992). The Bootstrap and edgeworth expansion. Springer-Verlag, New York.
- [24] HALL, P. UND HEYDE, C.C. (1980). Martingale Limit Theory and its application. Academic Press, New York.
- [25] HALL, P. UND WILSON, SUSAN R. (1991). Two Guidelines for Bootstrap Hypothesis Testing. Biometrics 47, 757-762.
- [26] HALL, P. (1989). Antithetic Resampling for the Bootstrap. Biometrica 76, 713-724.
- [27] HENZE, N. (1996). Empirical-distribution-function goodness-of-fit tests for discrete models. Can. J. Stat. 24, No.1, 81-93.
- [28] HOROWITZ, J.L. (2003). The Bootstrap in Econometrics. Stat. Science Vol.18, No. 2, 211-218.
- [29] Janssen, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem. Stat.Probab.Lett. 36, 9-21.
- [30] Janssen, A. (1998). Zur Asymptotik nichtparametrischer Tests. Skripten zur Mathematischen Statistik Nr.29, Münster.

- [31] Janssen, A. (1999). Nonparametric symmetry tests for statistical functionals. Math. Methods Stat. 8, No. 3, 320-343.
- [32] Janssen, A. und Mayer, C.-D. (2001). Conditional Studentized Survival Tests for Randomly Censored Models. Scand. Jour. Stat. Vol. 28, No. 2, 283-293.
- [33] Janssen, A. und Pauls, T. (2003). How do Bootstrap and permutation tests work? Ann.Stat.31, 768-806.
- [34] Janssen, A. und Pauls, T. (2003). A Monte-Carlo comparison of studentized bootstrap and permutation tests for heteroscedastic two-sample problems. Comput. Stat. 20, No. 3, 369-383.
- [35] Janssen, A. (2005). Resampling Student's t-Type Statistics. Ann.Inst.Statist.Math. Vol.57, No.3, 507-529.
- [36] Janssen, A. und Völker, D. (2007). Most powerful conditional tests. Stat.&Dec. 25, 41-62.
- [37] JIMÉNEZ-GAMERO, M.D.; MUÑOZ-GARCÍA, J. UND CUBILES-DE-LA-VEGA, M.D. (2006). Consistency of the reduced bootstrap for sample means. Stat.Probab.Lett. 76, 689-697.
- [38] JOHNSON, N. UND KOTZ, S. (1977). Urn Models and Their Application. John Wiley & Sons, New York.
- [39] Lahiri, S.N. (2003). Resampling Methods for Dependent Data. Springer-Verlag, New York.
- [40] LEHMANN, E.L. UND ROMANO, J.P. (2005). Testing Statistical Hypotheses. Springer-Verlag, New York.
- [41] LEPAGE, R. UND BILLARD, L. (1992). Exploring the Limits of Bootstrap. John Wiley & Sons, New York.
- [42] LINDVALL, T. (1992). Lectures on the coupling method. John Wiley & Sons, New York.
- [43] Lo, A. Y. UND SAZONOV, V. V. (1995). The von Mises ω^2 -statistic and the bootstrap. Theory Prob. Appl. 40, 653-659.
- [44] Lo, A. Y. UND SAZONOV, V. V. (1997). Von Mises ω^2 -statistic and the generalized Bayesian bootstraps. Ann. Inst. Stat. Math 49, 25-34.
- [45] MAMMEN, E. (1992). Bootstrap, wild bootstrap, and asymptotic normality. Prob.Theor.Rel.Fields 93, 439-455.
- [46] Mammen, E. (1992). When Does Bootstrap Work? Asymptotic Results and Simulations. Lecture Notes in Statistics 77, Springer Verlag.

- [47] MANLY, B.F.J. (1997). Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman & Hall, London.
- [48] MASON, D. M. UND NEWTON, M. A. (1992). A Rank Statistics Approach to the Consistency of a General Bootstrap. Ann. Stat. 20, 1611-1624.
- [49] Pauls, T. (2003). Resampling-Verfahren und ihre Anwendung in der Testtheorie. Dissertation Düsseldorf 2003. Books on Demand GmbH, Norderstedt.
- [50] Pauly, M. (2005). Ein multivariater zentraler Grenzwertsatz f\u00fcr Resampling-Statistiken. Diplomarbeit Universit\u00e4t D\u00fcsseldorf.
- [51] POLITITS, D.N. (2003). The Impact of Bootstrap Methods on Time Series Analysis. Stat. Science 18, 219-230.
- [52] POLITITS, D.N. UND ROMANO, J.P. (1994). The Stationary Bootstrap. J.Am.Stat.A. 89, No.428, 1303-1313.
- [53] PRAESTGAARD, J. UND WELLNER, J.A. (1993). Exchangeably weighted Bootstraps of the general empirical process. Ann. Probab. 21, No.4, 2053-2086.
- [54] RADHAKRISHNA RAO, C.; PATHAK, P.K. UND KOLTCHINSKII, V.I. (1997). Bootstrap by sequential resampling. J. Stat. Plan.Inf. 64, 257-281.
- [55] ROMANO, J.P. (1990). On the Behavior of Randomization Tests without a Group Invariance Assumption. J.Am.Stat.A. 85, No.411, 686-692.
- [56] Rubin, D. (1981). The Bayesian Bootstrap. Ann. Stat. 9, 130-134.
- [57] SERFLING, R.J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley & Sons, New York.
- [58] SHORACK, G.R. UND WELLNER J.A. (1986). Empirical processes with applications to statistics. John Wiley & Sons, New York.
- [59] SINGH, K. (1981). On the asymptotic accuracy of Efron's Bootstrap. Ann. Stat. 9, 1187-1195.
- [60] Stout, W.F. (1974). Almost sure convergence. Academic Press, New York.
- [61] SWANEPOEL, J.W.H. (1986). A note on proving that the modified bootstrap works. Commun. Stat., A15, 3193-3203.
- [62] VAN DER VAART, A.W. UND WELLNER J.A. (1995). Weak Convergence and Empirical Processes. Springer Verlag, New York.
- [63] VAN DER VAART, A.W. (1998). Asymptotic Statistics. Cambridge University Press, Cambridge.

- [64] WITTING, H. UND NÖLLE, G. (1970). Angewandte Mathematische Statistik. B.G. Teubner, Stuttgart.
- [65] WITTING, H. (1985). Mathematische Statistik I. Parametrische Verfahren bei festem Stichprobenumfang. B.G. Teubner, Stuttgart.
- [66] WITTING, H. UND MÜLLER-FUNK, U. (1995). Mathematische Statistik II. Asymptoti $sche\ Statistik:\ Parametrische\ Modelle\ und\ nichtparametrische\ Funktionale.\ B.G.\ Teub-nicht gegen auch gegen auch gegen gege$ $\mathbf{ner},\ \mathbf{Stuttgart}.$

Erklärung

Die hier vorgelegte Dissertation habe ich eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Markus Pauly

Düsseldorf, den 06.01.2009