

**Asymptotische Eigenschaften  
einer neuen Klasse von  
nichtparametrischen Zwei-Stichproben-Tests**

Vom Fachbereich Mathematik  
der Universität Hannover  
zur Erlangung des Grades  
Doktor der Naturwissenschaften  
Dr. rer. nat.

genehmigte Dissertation

von

Dipl.-Math. CARSTEN FRANZ  
geboren am 7. Februar 1976 in Hannover

Referent: Prof. Dr. Ludwig Baringhaus, Universität Hannover  
Korreferent: Prof. Dr. Norbert Henze, Universität Karlsruhe  
Tag der Promotion: 25. Mai 2004

## Zusammenfassung

Diese Arbeit beschäftigt sich mit einer neuen Familie von nichtparametrischen, multivariaten Zwei-Stichproben-Tests. Die Gestalt dieser Familie sowie zwei unterschiedliche Darstellungsweisen der Testgrößen werden in Kapitel 2 untersucht. Anschließend werden im folgenden Kapitel beide Darstellungen genutzt, um zu zeigen, dass die asymptotische Verteilung der Testgröße unter der Hypothese eine gewichtete Summe von Quadraten unabhängiger standardnormalverteilter Zufallsvariablen ist. Die Gewichte sind in beiden Fällen die Eigenwerte von linearen Integraloperatoren, die von der den beiden Stichproben gemeinsamen zu Grunde liegenden, unbekanntem Verteilung abhängen. Methoden zur Gewinnung des kritischen Wertes werden in Kapitel 4 vorgestellt. Neben dem klassischen Monte Carlo-Bootstrap-Verfahren wird als Alternative die Approximation der Gewichte der Grenzverteilung durch die Eigenwerte eines linearen Operators, bei dem die zu Grunde liegende Verteilung durch die empirische Verteilung der zusammengelegten Beobachtung ersetzt wurde, vorgeschlagen. Um zu zeigen, dass das Verfahren gegen die Grenzverteilung unter der Hypothese konvergiert, wird die spezielle Struktur der Testgröße ausgenutzt. In Kapitel 5 wird ein zweiter Weg für dieses Resultat präsentiert. Unter stärkeren Voraussetzungen erhält man eine Aussage, die sich auf Teststatistiken, deren Aufbau der Grenzverteilung mit dem bei der neuen Familie von Tests übereinstimmt, übertragen lässt. Die Wirksamkeit der Tests im Vergleich zu bekannten Verfahren wird in Kapitel 6 bei einigen konkreten Problemstellungen theoretisch mittels der approximativen Bahadur-Effizienz untersucht, von der gezeigt wird, dass sie lokal mit der Pitman-Effizienz übereinstimmt. Rechnergestützte Güteuntersuchungen unterstreichen im folgenden Kapitel die theoretischen Ergebnisse. Im letzten Kapitel wird für einen der neuen Tests die Vorgehensweise beim Zwei-Stichproben-Fall auf das Anpassungsproblem übertragen.

**Schlagwörter.** multivariater Zwei-Stichproben-Test, Bootstrap-Approximation der Eigenwerte linearer Operatoren, Zufallselemente im Hilbertraum

## Abstract

This thesis describes a new family of nonparametric, multivariate two-sample tests. The structure of the family as well as two different representations of the test statistics are discussed in chapter 2. Starting with each of the representations, the following chapter shows that the asymptotic null distribution is a weighted sum of squares of independent standard normally distributed random variables. For each representation the weights are the eigenvalues of linear integral operators depending on the common underlying, unknown distribution of both samples. Methods for obtaining the critical value are introduced in chapter 4. Besides the classical Monte Carlo-bootstrap-procedure an alternative method is suggested: the approximation of the weights in the limit distribution by substituting them with the eigenvalues of a linear integral operator where the unknown underlying distribution is replaced with the empirical distribution of the pooled sample. To prove the convergence of this method to the limiting null distribution, the special structure of the test statistics is used. In chapter 5 a second possibility for showing the convergence is presented. Under stronger pre-requisites one achieves a result which can be applied to general test statistics with a structure of the limiting distribution similar to that of the new tests. For selected testing problems the performance of the new tests in comparison with well-known tests is investigated in chapter 6 by means of approximate Bahadur-efficiency which is shown to coincide locally with the Pitman-efficiency. A simulation study on the power of the new tests emphasizes the theoretical results in the subsequent chapter. The last chapter exemplarily transfers the two-sample approach to the one-sample goodness-of-fit problem.

**Keywords.** multivariate two-sample test, bootstrap approximation of eigenvalues of linear operators, random elements in Hilbert-spaces

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Entstehung . . . . .	2
1.2	Inhalt und Aufbau der Arbeit . . . . .	4
1.3	Dank . . . . .	7
<b>2</b>	<b>Gestalt der neuen Tests</b>	<b>9</b>
2.1	Abstände zwischen Verteilungen . . . . .	9
2.2	Darstellung als V-Statistik . . . . .	12
2.3	Darstellung mittels empirischer charakteristischer Funktionen . . . . .	15
<b>3</b>	<b>Die asymptotische Verteilung</b>	<b>21</b>
3.1	Asymptotik degenerierter V-Statistiken . . . . .	21
3.1.1	Konvergenz der Differenz von V- und U-Statistik . . . . .	22
3.1.2	Konvergenz der U-Statistik . . . . .	23
3.2	Spezialfälle beim Cramér-Kern . . . . .	32
3.2.1	Die Gleichverteilung . . . . .	32
3.2.2	Die Normalverteilung . . . . .	33
3.3	Asymptotik Hilbertraum-wertiger Zufallselemente . . . . .	36
3.4	Vergleich der beiden Ansätze . . . . .	41
<b>4</b>	<b>Approximation der Grenzverteilung</b>	<b>45</b>
4.1	Das Monte Carlo-Bootstrap-Verfahren . . . . .	46
4.2	Bootstrap in der Grenzverteilung . . . . .	54
4.2.1	Reduktion des Eigenwertproblems . . . . .	56
4.2.2	Einhaltung des Niveaus . . . . .	58

4.3	Anwendung der neuen Tests . . . . .	60
<b>5</b>	<b>Approximation linearer Operatoren</b>	<b>63</b>
5.1	Diskrete Approximation von Räumen . . . . .	64
	Anwendung in der Statistik . . . . .	65
5.2	Diskrete Kompaktheit und Konsistenz . . . . .	67
	Anwendung auf $L_2^{\mathcal{C}}(R, \mathcal{R}, P^Z)$ -Integraloperatoren . . . . .	68
5.2.1	Nachweis für die neue Familie von Tests . . . . .	73
5.2.2	Nachweis für den Cramér-von Mises-Test . . . . .	75
5.3	Konvergenz der Eigenwerte . . . . .	76
	Anwendung auf $L_2^{\mathcal{C}}(R, \mathcal{R}, P^Z)$ -Integraloperatoren . . . . .	77
5.3.1	Nachweis für die neue Familie von Tests . . . . .	80
5.3.2	Anwendung auf den Cramér-von Mises-Test . . . . .	80
<b>6</b>	<b>Effizienz</b>	<b>83</b>
6.1	Bahadur-Effizienz . . . . .	83
6.1.1	Bahadur-Steigungen für die neuen Tests . . . . .	87
6.1.2	Lage-Alternativen bei eindimensionalen Normalverteilungen . . . . .	88
6.1.3	Skalenalternativen bei eindimensionalen Normalverteilungen . . . . .	89
6.1.4	Ein Kontaminationsmodell . . . . .	90
6.2	Pitman-Effizienz . . . . .	91
6.2.1	Lokale Gleichheit bei Lage-Alternativen . . . . .	93
6.2.2	Lokale Gleichheit beim Kontaminationsmodell . . . . .	95
6.2.3	Lokale Gleichheit bei Skalenalternativen . . . . .	96
6.3	Vergleich der Effizienzen einiger Tests der neuen Familie . . . . .	98
<b>7</b>	<b>Simulationen</b>	<b>105</b>
7.1	Lage- und Skalenalternativen bei Normalverteilungen . . . . .	105
7.1.1	Der eindimensionale Fall . . . . .	105
7.1.2	Der mehrdimensionale Fall . . . . .	111
7.2	Mischung mit Lehmann-Alternativen . . . . .	114

---

7.2.1	Der eindimensionale Fall . . . . .	115
7.2.2	Der mehrdimensionale Fall . . . . .	117
<b>8</b>	<b>Die neuen Tests als Anpassungstests</b>	<b>119</b>
8.1	Anpassung an eine vollständig festgelegte Verteilung . . . . .	119
8.2	Anpassung an eine Familie von Verteilungen . . . . .	122
<b>A</b>	<b>Vollständig monotone Funktionen</b>	<b>133</b>
<b>B</b>	<b>U- und V-Statistiken</b>	<b>135</b>
<b>C</b>	<b>Lineare Operatoren auf Hilberträumen</b>	<b>139</b>
C.1	Operatoren der Hilbert-Schmidt-Klasse . . . . .	141
C.2	Operatoren der Spurklasse . . . . .	142
<b>D</b>	<b>Zufallselemente mit Werten im Hilbertraum</b>	<b>143</b>
D.1	Gaußsche Zufallselemente und der zentrale Grenzwertsatz . . . . .	145
	<b>Symbolverzeichnis</b>	<b>147</b>
	<b>Literaturverzeichnis</b>	<b>149</b>



# Kapitel 1

## Einleitung

Diese Arbeit beschäftigt sich mit einer neuen Familie von multivariaten, nichtparametrischen Zwei-Stichproben-Tests. Bei dem Zwei-Stichproben-Problem sind zwei Beobachtungssätze gegeben, auf deren Basis man eine Aussage über die Hypothese treffen möchte, dass beide Stichproben aus derselben Grundgesamtheit stammen. Sind die Beobachtungen reellwertig, so spricht man vom univariaten Zwei-Stichproben-Problem, entstammen sie dem mehrdimensionalen euklidischen Raum, so spricht man vom multivariaten Zwei-Stichproben-Problem. Zahllose Anwendungsbeispiele für solche Tests finden sich bei medizinischen Untersuchungen: „Hat das Medikament einen Einfluss auf ausgewählte Körperfunktionen?“, der Biologie: „Wird das Wachstum einer bestimmten Art von Zellen, zum Beispiel Krebszellen, durch Bestrahlung verändert?“ oder bei sozialwissenschaftlichen Themen: „Existieren Unterschiede in der Einkommensverteilung von weiblichen und männlichen Arbeitnehmern?“ Bei Zwei-Stichproben-Tests wird unterschieden zwischen parametrischen Tests, die ein bestimmtes Vorwissen über die möglichen Verteilungen unter der Hypothese und die Verteilungen in den Alternativen für eine geeignete Testgröße ausnutzen, und nichtparametrischen Tests, denen keine Verteilungsannahmen zu Grunde liegen. Eine Vielzahl von nichtparametrischen Tests, vorwiegend für den univariaten Fall, wurde vorgeschlagen. Genannt seien hierfür neben dem bekannten Kolmogorov-Smirnov- und Cramér-von Mises-Test auch der Rangsummen-Test von Mann-Whitney sowie der Test auf Skalenalternativen von Mood. Ein nichtparametrischer Test, den man auch bei multivariaten Daten anbringen kann, ist der Nächst-Nachbar-Test von Henze. Diese Arbeit beschäftigt sich mit der Einführung und Analyse einer neuen Klasse von nichtparametrischen Zwei-Stichproben-Tests, die sowohl im eindimensionalen als auch im mehrdimensionalen Fall anwendbar sind.

## 1.1 Entstehung

Der Ursprung der neuen Tests liegt in einer Aussage, die in etwas speziellerer Form unter anderem bereits von Deuber und Morgenstern betrachtet wurde.

(A) Bei einer gleichen Anzahl schwarzer und weißer Punkte im euklidischen Raum ist die Summe der Abstände von Punkten gleicher Farbe kleiner gleich der Summe der paarweisen Abstände von Punkten unterschiedlicher Farbe.

(B) Gleichheit tritt genau dann ein, wenn die schwarzen und weißen Punkte übereinander liegen.

Erst vor kurzem bewies Morgenstern [Mor01] mit Hilfe einer geometrischen Betrachtung die Vermutung Deubers [Deu98] für die Gleichheit. Wählt man für  $F$  und  $G$  empirische Maße mit der Gleichverteilung auf den schwarzen beziehungsweise weißen Punkten, so lässt sich die Aussage in der folgenden, verallgemeinerten Form wiederfinden:

$$\frac{1}{2}\mathbb{E}|X_1 - X_2| + \frac{1}{2}\mathbb{E}|Y_1 - Y_2| \leq \mathbb{E}|X_1 - Y_1|, \quad (1.1)$$

wobei  $X_1, X_2, Y_1$  und  $Y_2$  unabhängige  $\mathbb{R}^d$ -wertige Zufallsvektoren sind,  $X_1, X_2$  mit Verteilungsfunktion  $F$  und  $Y_1, Y_2$  mit Verteilungsfunktion  $G$ . Dabei bezeichne  $\mathbb{R}^d$  den  $d$ -dimensionalen euklidischen Raum und  $|\cdot|$  die euklidische Norm auf demselben. Gleichheit liegt genau dann vor, wenn  $F = G$  ist.

Ausgehend von dieser Ungleichung wurde in der Diplomarbeit des Autors [Fra00], siehe auch Baringhaus und Franz [BF04], eine empirische Variante der Ungleichung,

$$\begin{aligned} T_{m,n}^{\text{Cramér}}(x_1, \dots, x_m; y_1, \dots, y_n) & \quad (1.2) \\ &= \frac{mn}{m+n} \left( \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n |x_j - y_k| \right. \\ & \quad \left. - \frac{1}{2m^2} \sum_{j=1}^m \sum_{k=1}^m |x_j - x_k| - \frac{1}{2n^2} \sum_{j=1}^n \sum_{k=1}^n |y_j - y_k| \right) \end{aligned}$$

als Testgröße für das nichtparametrische Zwei-Stichproben-Problem verwendet. Dabei bezeichnen  $x_1, \dots, x_m$  die  $m$  Beobachtungen der ersten Stichprobe,  $y_1, \dots, y_n$  bezeichnen die  $n$  Beobachtungen der zweiten Stichprobe, wobei alle Beobachtungen bei unabhängigen Versuchswiederholungen gewonnen worden seien und die Reihenfolge der Bezeichnung der Stichproben (erste, zweite) nicht von Bedeutung ist. Der angebrachte Faktor  $\frac{mn}{m+n}$  ist notwendig, um unter der Hypothese, dass beide Stichproben

einer gemeinsamen Grundgesamtheit entspringen, eine asymptotische Verteilung der Größe

$$T_{m,n}^{\text{Cramér}} = T_{m,n}^{\text{Cramér}}(X_1, \dots, X_m; Y_1, \dots, Y_n)$$

zu erhalten, wenn die Stichprobenumfänge  $m, n$  gegen unendlich streben. Dabei seien  $X_1, \dots, X_m, Y_1, \dots, Y_n$  unabhängig und identisch verteilt. Mit Hilfe einer Integraldarstellung von  $T_{m,n}^{\text{Cramér}}$  kann die asymptotische Verteilung unter der Hypothese als Integral über das Quadrat einer geeignet indizierten Brownschen Brücke dargestellt werden. Die Kovarianzfunktion hängt über den Zeitparameter von der unbekannt gemeinsamen Verteilung  $F = G$  ab. Im univariaten Fall lautet diese Integraldarstellung

$$T_{m,n}^{\text{Cramér}} = \frac{mn}{m+n} \int_{-\infty}^{+\infty} [F_m(u) - G_n(u)]^2 du \quad (1.3)$$

und es liegt Konvergenz in Verteilung für  $m, n \rightarrow \infty$  gegen die asymptotische Verteilung

$$T_{m,n}^{\text{Cramér}} \xrightarrow[m, n \rightarrow \infty]{\mathcal{D}} \int_{-\infty}^{+\infty} [\mathbb{B}(F(u))]^2 du$$

vor, wobei  $F_m$  die empirische Verteilungsfunktion der Zufallsvariablen  $X_1, \dots, X_m$  sei,  $G_n$  die empirische Verteilungsfunktion der Zufallsvariablen  $Y_1, \dots, Y_n$  bezeichne und  $\mathbb{B}$  einen Brownschen Brückenprozess beschreibt. Da Cramér [Cra28] bereits 1928 eine vergleichbare Testgröße in einer Ein-Stichproben-Variante vorschlug (bevor durch von Mises eine verteilungsfreie Form eingeführt wurde), bekam der Test den Namen *Cramér-Test*.

Die asymptotische Verteilung des neuen Tests ist weder im univariaten noch im multivariaten Fall verteilungsfrei, so dass zur Gewinnung des kritischen Wertes Monte Carlo-Bootstrap-Methoden verwendet wurden.

Ein weiterer Test für das Zwei-Stichproben-Problem wurde 1996 von Bahr in seiner Dissertation [Bah96] vorgeschlagen. Ausgangsbasis war hier der quadratische Abstand zwischen den empirischen Fourier-Transformierten, integriert bezüglich der  $d$ -dimensionalen Standardnormalverteilung

$$\begin{aligned} T_{m,n}^{\text{Bahr}}(x_1, \dots, x_m; y_1, \dots, y_n) \\ = \frac{mn}{m+n} \int_{\mathbb{R}^d} \left| \frac{1}{m} \sum_{j=1}^m \exp(it'x_j) - \frac{1}{n} \sum_{k=1}^n \exp(it'y_k) \right|^2 dN_d(0, I_d)(t). \end{aligned}$$

Es bezeichne  $N_d(\mu, \Sigma)$  die  $d$ -dimensionale Normalverteilung mit Erwartungswert  $\mu \in \mathbb{R}^d$  und Kovarianzmatrix  $\Sigma \in \mathbb{R}^{d \times d}$ ,  $t'$  bezeichne den transponierten Vektor  $t$ ,  $I_d$  die

$d$ -dimensionale Einheitsmatrix und  $i$  die komplexe Einheit. Es wurde gezeigt, dass sich diese Statistik darstellen lässt als

$$\begin{aligned} T_{m,n}^{\text{Bahr}}(x_1, \dots, x_m; y_1, \dots, y_n) & \quad (1.4) \\ &= \frac{mn}{m+n} \left( \frac{2}{mn} \sum_{j=1}^m \sum_{k=1}^n (-1) \cdot \exp\left(-\frac{1}{2}|x_j - y_k|^2\right) \right. \\ & \quad - \frac{1}{m^2} \sum_{j=1}^m \sum_{k=1}^m (-1) \cdot \exp\left(-\frac{1}{2}|x_j - x_k|^2\right) \\ & \quad \left. - \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n (-1) \cdot \exp\left(-\frac{1}{2}|y_j - y_k|^2\right) \right). \end{aligned}$$

Auch hier wurden die (nicht verteilungsfreie) asymptotische Verteilung unter der Hypothese über die Integraldarstellung gewonnen und Monte Carlo-Methoden zur Approximation des kritischen Wertes verwendet.

## 1.2 Inhalt und Aufbau der Arbeit

Vergleicht man die Cramér-Statistik (1.2) und die Statistik von Bahr (1.4), so ist bemerkenswert, dass beide als Spezialfälle einer allgemeineren Statistik

$$\begin{aligned} T_{m,n}^{\phi}(x_1, \dots, x_m; y_1, \dots, y_n) & \quad (1.5) \\ &= \frac{mn}{m+n} \left( \frac{2}{mn} \sum_{j=1}^m \sum_{k=1}^n \phi(|x_j - y_k|^2) \right. \\ & \quad \left. - \frac{1}{m^2} \sum_{j=1}^m \sum_{k=1}^m \phi(|x_j - x_k|^2) - \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \phi(|y_j - y_k|^2) \right) \end{aligned}$$

verstanden werden können, wobei  $\phi : [0, +\infty) \rightarrow \mathbb{R}$  ist.

Für die Cramér-Statistik (1.2) setze man  $\phi(z) := \sqrt{z}/2$  und für die Statistik (1.4) verwende man  $\phi(z) := -e^{-z/2}$ . Die sich nun aufdrängende Frage, *welche Funktionen  $\phi$  sonst noch in Frage kommen, damit sich ein konsistenter und orthogonal invarianter Test ergibt*, wird zu Beginn von Kapitel 2 dieser Arbeit geklärt. Spezialfälle der in Frage kommenden Funktionen sind die bereits von Székely [Szé02] untersuchten Funktionen  $\phi(z) = z^\alpha$ ,  $0 < \alpha < 1$ . Im weiteren Verlauf des Kapitels werden für die geeigneten Funktionen  $\phi$  zwei Darstellungsweisen der Teststatistiken vom Typ (1.5) erläutert. Neben der ersichtlichen Auffassung als verallgemeinerte Zwei-Stichproben-V-Statistik mit Kernlänge 2 in jeder Stichprobe wird auf eine Darstellung als  $L_2$ -Abstand der beiden zu den Stichproben gehörenden empirischen charakteristischen Funktionen in einem von  $\phi$  abhängenden Hilbertraum eingegangen.

Kapitel 3 beleuchtet die asymptotische Verteilung der Testgröße unter der Hypothese, dass beiden Stichproben dieselbe Verteilung zu Grunde liegt. Betrachtet man die Darstellung als V-Statistik, so führt die Anwendung eines Resultats von Neuhaus [Neu77] zu einer Darstellung der Grenzverteilung als gewichtete Summe von unabhängigen  $\chi_1^2$ -verteilten Zufallsvariablen. Besonderes Augenmerk wird auf die Vereinfachung dieses Resultats gelegt, welches den Startpunkt für weitere Überlegungen in den folgenden Kapiteln bilden soll: Die Gewichte der Summe sind die Eigenwerte des Operators

$$H : \begin{array}{ll} L_2(\mathbb{R}^d, \mathcal{B}^d, \gamma) & \rightarrow L_2(\mathbb{R}^d, \mathcal{B}^d, \gamma) \\ f & \mapsto \int h(\cdot, y) f(y) d\gamma(y) \end{array} ,$$

mit  $h(x, y) := E[\phi(|x - Y|^2) + \phi(|y - X|^2) - \phi(|x - y|^2) - \phi(|X - Y|^2)]$ ,  $x, y \in \mathbb{R}^d$ , und  $X, Y \sim \gamma$ , also  $X$  und  $Y$  die Verteilung  $\gamma$  haben mögen, wobei  $\gamma$  die unbekannte Verteilung unter der Hypothese sei.  $\mathcal{B}^d$  bezeichne die Borelsche  $\sigma$ -Algebra auf  $\mathbb{R}^d$ . Wählt man die Darstellung mittels empirischer charakteristischer Funktionen zur Ermittlung der Grenzverteilung unter der Hypothese, so ergibt sich wieder eine gewichtete Summe von  $\chi_1^2$ -verteilten Zufallsvariablen. Die Gewichte sind jedoch die Eigenwerte des Kovarianzoperators, der zu dem Gaußschen Zufallselement gehört, gegen das die empirischen charakteristischen Funktionen mittels des zentralen Grenzwertsatzes im Hilbertraum unter der Hypothese in Verteilung konvergieren. Das Kapitel schließt mit einem Vergleich der beiden Ansätze, der unter anderem zeigt, dass die Operatoren dieselben nicht verschwindenden Eigenwerte besitzen.

Wegen der erwähnten Abhängigkeit der Grenzverteilung unter der Hypothese von der unbekanntem, beiden Stichproben zu Grunde liegenden Verteilung bedarf es numerischer Methoden, um einen kritischen Wert angeben zu können. Zwei Approximationsmethoden werden in Kapitel 4 untersucht. Zum einen ist dies das bei nichtparametrischen Tests häufig verwendete Monte Carlo-Bootstrap-Verfahren, für das bewiesen wird, dass die damit gewonnene Verteilung tatsächlich die unbekannte Grenzverteilung approximiert. Das zweite Verfahren startet direkt von der Grenzverteilung, indem man die unbekannte Verteilung  $\gamma$  in der Definition des Operators  $H$  durch die empirische Verteilung der zusammengelegten Beobachtungen schätzt. In der Praxis ist dies von besonderem Interesse, da aus der Integralgleichung zur Eigenwertgewinnung ein lineares Gleichungssystem wird und es somit möglich ist, Approximationen für die Eigenwerte und mittels dieser auch Approximationen für den kritischen Wert des Tests auf deterministischem Wege anzugeben, ohne auf Monte Carlo-Methoden angewiesen zu sein. Mathematisch gesehen verursacht dies jedoch einige Probleme, da sich die Räume, auf denen die approximierenden Operatoren definiert sind, von dem des interessierenden Operators unterscheiden. Um die Konvergenz der Eigenwerte zu zeigen, wird daher die am Ende des vorangehenden Kapitels angesprochene Gleichheit der positiven Eigenwerte von  $H$  und der Eigenwerte des Kovarianzoperators im Hilbertraum ausgenutzt. Der Umweg über empirische

charakteristische Funktionen hilft beim Nachweis der Konvergenz der mit Eigenwerten von Operatoren, die auf empirischen Maßen beruhen, gewichteten Summe von unabhängigen  $\chi_1^2$ -verteilten Zufallsvariablen gegen die Grenzverteilung unter der Hypothese. Dazu findet der zentrale Grenzwertsatz für Dreiecksschemata im Hilbertraum Verwendung. Im Anschluss an die theoretische Untersuchung der Konvergenz wird die numerische Durchführung erläutert sowie auf die Einhaltung des Niveaus in verschiedenen Fällen eingegangen, wobei die „Eigenwert-Methode“ zur Gewinnung des kritischen Wertes verwendet wird. Abschließend werden die für die Anwendung bei Tests relevanten Ergebnisse kurz zusammengefasst und eine Konsistenzaussage bewiesen.

Die Problematik der Konvergenz der Eigenwerte von Operatoren, die auf unterschiedlichen Räumen definiert sind, wird in Kapitel 5 mit Hilfe von Stummels Theorie zur Störung von linearen Operatoren [Stu70, Stu71, Stu72] bearbeitet. Damit bietet sich zumindest für den Teil der zulässigen Funktionen  $\phi$ , die beschränkt sind, eine Alternative, die statt des im letzten Kapitel gewählten Weges über empirische charakteristische Funktionen verwendet werden kann. Insbesondere die Anwendbarkeit dieses Grundgedankens auf andere Probleme ist von Bedeutung. Diesem wird mit einem recht allgemein gehaltenen Satz Rechnung getragen, der als Beispiel auf die Approximation der (bekannten) Grenzverteilung der Cramér-von Mises-Testgröße angewendet wird. Dabei wird die Wirksamkeit des Verfahrens zur Gewinnung einer Näherung der Eigenwerte mit Simulationen unterstrichen.

In Kapitel 6 wird die approximative Bahadur-Effizienz der neuen Tests bei ausgewählten Testproblemen bestimmt und die lokale Gleichheit mit der Pitman-Effizienz gezeigt. Die betrachteten Testprobleme umfassen ein Kontaminationsmodell sowie Tests auf Lage- und Skalenunterschiede. Es wird nachgewiesen, dass lokale Gleichheit nicht nur für reine Lage- und Skalenalternativen vorliegt, sondern für alle linearen Transformationen, die geeignet gegen die Identität streben. Das Kapitel schließt mit einem graphischen Vergleich der approximativen Bahadur-Steigung für unterschiedliche Kernfunktionen, unter anderem dem Cramér-Kern und dem Kern von Bahr bei Alternativen in Erwartungswert und Streuung zu der Standardnormalverteilung. Die anhand von Effizienzen theoretisch gewonnenen Eindrücke dieses Abschnitts werden in Kapitel 7 mit Simulationsergebnissen verglichen. Neben Tests auf Lage- und Skalenalternativen bei ein- und mehrdimensionalen Normalverteilungen wird ein Kontaminationsmodell mit Lehmann-Alternativen untersucht. Bei allen Simulationen schneiden die ausgewählten Beispiele der neuen Familie von Tests im Vergleich zu bekannten parametrischen und nichtparametrischen Tests gut ab. Insbesondere bei Lage-Alternativen und auch bei Kontaminationsalternativen liefern einige von ihnen ein hervorragendes Güteverhalten. Für Skalenalternativen ist die Güte zwar verhältnismäßig weit von der der parametrischen Tests entfernt, im Vergleich zu den standardmäßig verwendeten nichtparametrischen Tests, wie dem Kolmogorov-Smirnov-Test oder dem Cramér-von Mises-Test, lässt sich aber auch

hier eine ausgezeichnete Güte feststellen.

Das letzte Kapitel widmet sich der Übertragung der vorgestellten Vorgehensweise beim Zwei-Stichproben-Problem auf Anpassungstests. Sowohl für den Fall des Tests gegen eine vollständig festgelegte Verteilung, der Standardnormalverteilung, als auch für den Fall des Tests gegen eine ganze Klasse von Verteilungen, hier der Test auf Normalverteilung, wird ein neues Testverfahren vorgeschlagen und es werden die asymptotischen kritischen Werte ermittelt. Abgerundet werden auch diese Betrachtungen durch einige Güteuntersuchungen im Vergleich mit anderen Anpassungstests für diese Probleme. Es stellt sich heraus, dass der Cramér-Test auch für das Anpassungsproblem ein sehr gutes Güteverhalten zeigt.

Alle Simulationen in dieser Arbeit wurden mit dem Statistikpaket **R** durchgeführt. Dieses Programm ist unter allen momentan gängigen Betriebssystemen frei verfügbar und kann im Internet unter der Adresse <http://www.r-project.org> heruntergeladen werden. Die für diese Dissertation verwendeten Methoden sind in dem Paket `cramer` zusammengestellt, welches ebenfalls unter der angegebenen Adresse erhältlich ist.

## 1.3 Dank

Mein besonderer Dank gilt Herrn Prof. Dr. L. Baringhaus für die Anregung dieser Arbeit sowie für Rat und Unterstützung bei ihrer Entstehung. Herrn Prof. Dr. N. Henze danke ich herzlich für die Übernahme des Korreferats.



# Kapitel 2

## Gestalt der neuen Tests

Dieses Kapitel befasst sich mit der Suche nach Funktionen  $\phi : [0, +\infty) \rightarrow \mathbb{R}$ , die geeignet sind mittels der Testgröße (1.5) ein konsistentes Testverfahren zu liefern. Ferner behandelt es zwei unterschiedliche Darstellungsweisen für die neuen Teststatistiken. Die erste Darstellung als V-Statistik, die sich besonders gut zur numerischen Bestimmung der Testgröße eignet und die bereits in der Einleitung erwähnt wurde, wird in Abschnitt 2.2 erläutert. Anschließend wird in Abschnitt 2.3 gezeigt, dass sich die Testgröße auch als Normquadrat eines Zufallselements in einem geeigneten  $L_2$ -Raum schreiben lässt. Beide Darstellungsarten werden in Kapitel 3 näher untersucht, um die Verteilung der Teststatistik unter der Hypothese zu ermitteln.

### 2.1 Abstände zwischen Verteilungen

Die neue, allgemeine Teststatistik (1.5) entspricht bis auf den Standardisierungsfaktor der empirischen Fassung der Größe

$$d_\phi^2(P^X, P^Y) := 2\mathbb{E}\phi(|X_1 - Y_1|^2) - \mathbb{E}\phi(|X_1 - X_2|^2) - \mathbb{E}\phi(|Y_1 - Y_2|^2), \quad (2.1)$$

wobei  $X_1, X_2 \sim P^X, Y_1, Y_2 \sim P^Y$  unabhängige  $d$ -dimensionale Zufallsvektoren seien und die auftretenden Erwartungswerte endlich seien mögen. Wir werden sehen, dass man diese Größe als Abstand zwischen den beiden Verteilungen  $P^X$  und  $P^Y$  verstehen kann. Bezeichnet man mit  $P_m^X(x_1, \dots, x_m)$  die empirische Verteilung von  $m$  Beobachtungen  $x_1, \dots, x_m$  der ersten Stichprobe und mit  $P_n^Y(y_1, \dots, y_n)$  die empirische Verteilung von  $n$  Beobachtungen  $y_1, \dots, y_n$  der zweiten Stichprobe, so lässt sich die Testgröße schreiben als

$$T_{m,n}^\phi(x_1, \dots, x_m; y_1, \dots, y_n) := \frac{mn}{m+n} d_\phi^2(P_m^X(x_1, \dots, x_m), P_n^Y(y_1, \dots, y_n)). \quad (2.2)$$

Im Folgenden werden geeignete Funktionen  $\phi$  bestimmt, für welche die Teststatistik auf einen gegen jede Alternative konsistenten Zwei-Stichproben-Test führt.

Die Wahl der Funktion  $\phi : [0, +\infty) \rightarrow \mathbb{R}$  unterliegt dabei einigen Beschränkungen. Zunächst muss  $\phi$  stetig sein. Diese auf den ersten Blick starke Forderung erscheint sinnvoll, da nur in diesem Fall der durch (2.1) definierte Abstand zwischen zwei Dirac-Maßen eine stetige Funktion des euklidischen Abstandes der zugehörigen (Einpunkt-)Träger ist. Für die Konsistenz eines Tests mittels (2.2) gegen jede Alternative ist es unerlässlich, dass  $d_\phi(P^X, P^Y)$  genau dann verschwindet, wenn  $X$  und  $Y$  dieselbe Verteilung haben (Schreibweise:  $X \stackrel{\mathcal{D}}{=} Y$ ).

Der folgende Satz besagt, dass durch diese Forderungen die Klasse der „zulässigen“ Funktionen  $\phi$  festgelegt wird.

**Satz 2.1.** *Sei  $\phi : [0, +\infty) \rightarrow \mathbb{R}$  eine stetige Funktion. Dann sind die folgenden beiden Aussagen äquivalent:*

(a) *Für jede natürliche Zahl  $d \in \mathbb{N}$  und jede Wahl von Wahrscheinlichkeitsmaßen  $P^X, P^Y$  auf der Borelschen  $\sigma$ -Algebra  $\mathcal{B}^d$  des  $\mathbb{R}^d$  mit*

$$E\phi(|X|^2) < \infty \quad \text{und} \quad E\phi(|Y|^2) < \infty, \quad (2.3)$$

*wobei  $X \sim P^X, Y \sim P^Y$  ist, gilt*

$$d_\phi^2(P^X, P^Y) \geq 0, \quad (2.4)$$

*und es ist  $d_\phi^2(P^X, P^Y) = 0$  genau dann, wenn  $P^X = P^Y$ .*

(b)  *$\phi$  hat eine nicht-konstante erste Ableitung auf  $(0, +\infty)$ , welche vollständig monoton ist (zur Definition von vollständiger Monotonie siehe Anhang A, Definition A.1).*

**Beweis.** (b) $\implies$ (a). Unter der Voraussetzung, dass die Ausdrücke in (2.3) endlich sind, hat Mattner [Mat90, Satz 6, S.27] gezeigt, dass die Ungleichung (2.4) gilt, falls  $\phi'$  vollständig monoton ist. Gleichheit gilt nach Mattner genau dann, wenn  $P^X = P^Y$  oder  $\phi' \equiv 0$  oder sowohl  $\phi'$  konstant als auch  $EX = EY$  ist. Da  $\phi'$  nach Voraussetzung nicht-konstant ist, folgt die Behauptung.

(a) $\implies$ (b). Seien  $N \in \mathbb{N}$ ,  $c_j \in \mathbb{R}$ ,  $x_j \in \mathbb{R}^d$ ,  $j = 1, \dots, N$  mit  $\sum_{j=1}^N c_j = 0$  beliebig, wobei nicht alle  $c_j$  verschwinden mögen. Sei  $C := \sum_{j=1}^N c_j^+ = \sum_{j=1}^N c_j^-$ , wobei  $c_j^+$  ( $c_j^-$ ) den Positiv- (Negativ-) Teil von  $c_j$  bezeichne. Definiert man  $P^X := \frac{1}{C} \sum_{j=1}^N c_j^+ \delta_{x_j}$  und  $P^Y := \frac{1}{C} \sum_{j=1}^N c_j^- \delta_{x_j}$ , so sind  $P^X, P^Y$  Wahrscheinlichkeitsmaße auf  $\mathcal{B}^d$ , welche offensichtlich die Bedingung (2.3) erfüllen. Dabei bezeichne  $\delta_x$  das Dirac-Maß im

Punkt  $x \in \mathbb{R}^d$ . Nach Voraussetzung gilt (2.4), also

$$\begin{aligned}
 0 &\leq d_\phi^2(P^X, P^Y) \\
 &= 2 \frac{1}{C^2} \sum_{j=1}^N \sum_{k=1}^N c_j^+ c_k^- \phi(|x_j - x_k|^2) \\
 &\quad - \frac{1}{C^2} \sum_{j=1}^N \sum_{k=1}^N c_j^+ c_k^+ \phi(|x_j - x_k|^2) - \frac{1}{C^2} \sum_{j=1}^N \sum_{k=1}^N c_j^- c_k^- \phi(|x_j - x_k|^2) \\
 &= \frac{1}{C^2} \sum_{j=1}^N \sum_{k=1}^N \underbrace{(c_j^+ - c_j^-)}_{c_j} \underbrace{(c_k^+ - c_k^-)}_{c_k} [-\phi(|x_j - x_k|^2)]. \tag{2.5}
 \end{aligned}$$

Es handelt sich somit bei der Funktion  $-\phi(|\cdot|^2)$  um eine bedingt positiv definite Funktion der Ordnung  $k = 1$  auf dem  $\mathbb{R}^d$ . Näheres dazu findet sich in Anhang A: Da  $f := -\phi(|\cdot|^2)$  reellwertig und gerade ist, das heißt  $f(x) = f(-x)$ , genügt es für die bedingte positive Definitheit die Ungleichung (A.2) lediglich für  $N$  reelle Zahlen  $c_1, \dots, c_N$  nachzuweisen, was in (2.5) geschah. Guo, Hu und Sun zeigen, dass dann  $\phi'$  auf  $(0, +\infty)$  vollständig monoton ist, vergleiche Satz A.3. Um einzusehen, dass  $\phi'$  nicht konstant sein kann, nehme man an, dass  $\phi'$  konstant ist, und wähle Verteilungen mit demselben Erwartungswert, beispielsweise im eindimensionalen Fall,  $d = 1$ , zentrierte Normalverteilungen mit Varianz 1 und 2, also die Verteilungen  $P^X = N(0, 1)$ ,  $P^Y = N(0, 2)$ . Dann ist  $d_\phi^2(P^X, P^Y) = 0$ , aber  $P^X \neq P^Y$ . Dies widerspricht der Voraussetzung (a).  $\square$

Der Satz liefert mit  $\phi(z) = \sqrt{z}/2$  beziehungsweise  $\phi(z) = -\exp(-z/2)$  einen weiteren Beweis für die Ungleichung (1.1) und die entsprechende Aussage bei Bahr, denn man zeigt leicht die vollständige Monotonie der Ableitungen dieser Funktionen  $\phi$ .

Offensichtlich gilt  $d_\phi = d_\psi$ , falls  $\phi - \psi \equiv c \in \mathbb{R}$  konstant ist. Daher fordern wir ohne Beschränkung der Allgemeinheit  $\phi(0) = 0$ . Da uns die stetigen Funktionen  $\phi : [0, +\infty) \rightarrow \mathbb{R}$  mit  $\phi(0) = 0$ , deren erste Ableitung vollständig monoton auf  $(0, +\infty)$  und nicht konstant ist, durch die gesamte Arbeit begleiten werden, bezeichnen wir die Menge all dieser Funktionen mit  $\Phi$ .

Der eingangs verwendete Begriff „Abstand“ zwischen zwei Wahrscheinlichkeitsmaßen für  $d_\phi^2$  ist gerechtfertigt, da Mattner [Mat90, Satz 6, S.27] bewies, dass  $d_\phi := (d_\phi^2)^{\frac{1}{2}}$  für alle  $\phi \in \Phi$  eine Metrik auf der Menge aller Wahrscheinlichkeitsmaße auf  $\mathcal{B}^d$  definiert, für die (2.3) endlich ist.

Es sei nun noch eine Hilfsaussage aufgeführt, die wir in den folgenden Kapiteln benötigen werden.

**Lemma 2.2.** *Seien  $\phi \in \Phi$  und  $X_1, X_2 \sim P^X$ ,  $Y_1, Y_2 \sim P^Y$  unabhängige Zufallsvariablen mit Werten in  $\mathbb{R}^d$ . Gilt  $E\phi(|X_1|^2) < \infty$  und  $E\phi(|Y_1|^2) < \infty$ , so folgt*

$$E\phi(|X_1 - X_2|^2) < \infty \quad \text{und} \quad E\phi(|X_1 - Y_1|^2) < \infty.$$

Gilt  $E\phi^2(|X_1|^2) < \infty$  und  $E\phi^2(|Y_1|^2) < \infty$ , so folgt ferner

$$E\phi^2(|X_1 - X_2|^2) < \infty \quad \text{und} \quad E\phi^2(|X_1 - Y_1|^2) < \infty.$$

Gilt sogar  $E\phi^4(|X_1|^2) < \infty$  und  $E\phi^4(|Y_1|^2) < \infty$ , so folgt überdies

$$E\phi^4(|X_1 - X_2|^2) < \infty \quad \text{und} \quad E\phi^4(|X_1 - Y_1|^2) < \infty.$$

**Beweis.** Die erste Aussage zeigt Mattner [Mat90, Satz 6, S.27] unter Ausnutzung von  $\phi(s+t) \leq \phi(s) + \phi(t)$  für  $s, t \geq 0$ . Der Beweis lässt sich problemlos auf die zweite Aussage übertragen:

Zunächst gilt  $\phi^2(s+t) \leq 2\phi^2(s) + 2\phi^2(t)$  für  $s, t \geq 0$ . Sei  $E\phi^2(|X_1|^2) < \infty$ ,  $E\phi^2(|Y_1|^2) < \infty$ , dann gilt

$$\begin{aligned} E\phi^2(|X_1 - Y_1|^2) &\leq E\phi^2(2|X_1|^2 + 2|Y_1|^2) \\ &\leq 2E\phi^2(2|X_1|^2) + 2E\phi^2(2|Y_1|^2) \\ &\leq 8E\phi^2(|X_1|^2) + 8E\phi^2(|Y_1|^2) < \infty. \end{aligned}$$

Durch Ersetzen von  $Y_1$  durch  $X_2$  folgt  $E\phi^2(|X_1 - X_2|^2) < \infty$ .

Die letzte Aussage folgt analog mit  $\phi^4(s+t) \leq 8\phi^4(s) + 8\phi^4(t)$ ,  $s, t \geq 0$ .  $\square$

Abkürzend schreiben wir  $\mathcal{F}_d^k(\phi)$ ,  $k \in \mathbb{N}$ ,  $d \in \mathbb{N}$ ,  $\phi \in \Phi$ , für die Menge aller Wahrscheinlichkeitsmaße  $P^X$  auf  $\mathcal{B}^d$ , für die gilt:  $E\phi^k(|X|^2) < \infty$ ,  $X \sim P^X$  und  $P^X$  ist kein Einpunkt-Maß. Es folgt offensichtlich die Inklusion  $\mathcal{F}_d^{2k}(\phi) \subset \mathcal{F}_d^k(\phi)$ ,  $k \in \mathbb{N}$ . Der Ausschluss von Dirac-Maßen erfolgt, um bei der Untersuchung der Asymptotik der Testgröße trivialen Fällen vorzubeugen.

Wir kennen nun die Funktionen  $\phi \in \Phi$ , welche die Ungleichung (2.4) erfüllen. Um die Grenzverteilung der Teststatistik  $T_{m,n}^\phi$  für  $m, n \rightarrow \infty$  zu gewinnen, werden zwei unterschiedliche Wege beschritten. Den Ansatz dazu bilden zwei Darstellungen der Teststatistik, die in den nächsten Abschnitten erläutert werden.

## 2.2 Darstellung als V-Statistik

Vergleicht man die Struktur der Teststatistik (1.5) mit der Struktur bereits bekannter Statistiken, so drängt sich unmittelbar der Gedanke an V-Statistiken auf. Tatsächlich lassen sich die Teststatistiken (1.5) als verallgemeinerte (Zwei-Stichproben-) V-Statistik der Kernlänge (2, 2) schreiben:

$$T_{m,n}^\phi(x_1, \dots, x_m; y_1, \dots, y_n) = \frac{mn}{m+n} \frac{1}{m^2 n^2} \sum_{i,j=1}^m \sum_{k,\ell=1}^n \tilde{h}(x_i, x_j; y_k, y_\ell).$$

Die Kernfunktion ist gegeben durch

$$\begin{aligned} \tilde{h}(x_1, x_2; y_1, y_2) &:= \frac{1}{2} \left( \phi(|x_1 - y_1|^2) + \phi(|x_1 - y_2|^2) \right. \\ &\quad \left. + \phi(|x_2 - y_1|^2) + \phi(|x_2 - y_2|^2) \right) \\ &\quad - \phi(|x_1 - x_2|^2) - \phi(|y_1 - y_2|^2), \quad x_1, x_2, y_1, y_2 \in \mathbb{R}^d. \end{aligned}$$

Diese Kernfunktion erfüllt die in der Literatur häufig geforderte Symmetrie innerhalb jeder Stichprobe, vergleiche z.B. Koroljuk und Borovskich [KB94, S.35]:

$$\tilde{h}(x_1, x_2; y_1, y_2) = \tilde{h}(x_1, x_2; y_2, y_1) = \tilde{h}(x_2, x_1; y_1, y_2).$$

Die angegebene Kernfunktion ist auf Grund ihrer Stetigkeit  $(\mathcal{B}^d)^4/\mathcal{B}$ -messbar. Motiviert durch das später verwendete Resultat von Neuhaus [Neu77] sei die Reihenfolge der Parameter von der gebräuchlichen Reihenfolge  $(x_1, x_2; y_1, y_2)$  im Folgenden abgeändert auf  $(x_1, y_1; x_2, y_2)$ .

Eng verbunden mit der V-Statistik ist die zugehörige U-Statistik

$$\begin{aligned} U_{m,n}(x_1, \dots, x_m; y_1, \dots, y_n) &= \frac{mn}{m+n} \frac{1}{m(m-1)} \frac{1}{n(n-1)} \sum_{i \neq k}^m \sum_{j \neq \ell}^n \tilde{h}(x_i, y_j; x_k, y_\ell) \\ &= \frac{mn}{m+n} \binom{m}{2}^{-1} \binom{n}{2}^{-1} \sum_{i < k}^m \sum_{j < \ell}^n \tilde{h}(x_i, y_j; x_k, y_\ell). \end{aligned} \quad (2.6)$$

Sen [Sen74] zeigte, dass solche U-Statistiken, deren Kern symmetrisch innerhalb beider Stichproben ist, unter der Hypothese asymptotisch normalverteilt sind, falls der Kern nicht-degeneriert ist. Ein Kern  $h$  der Kernlänge  $(2, 2)$  ist nicht-degeneriert im Sinne von Sen, falls unter der Hypothese

$$\mathbb{E} \left[ \mathbb{E} [h(X_1, Y_1; X_2, Y_2) | X_1]^2 \right] \in (0, \infty) \quad (2.7)$$

und

$$\mathbb{E} \left[ \mathbb{E} [h(X_1, Y_1; X_2, Y_2) | Y_1]^2 \right] \in (0, \infty), \quad (2.8)$$

wobei  $X_1, X_2, Y_1, Y_2$  unabhängig und identisch verteilt sind. Da in unserem Fall bei Gültigkeit der Hypothese weder (2.7) noch (2.8) erfüllt sind, ist  $\tilde{h}$  degeneriert im Sinne von Sen.

Die Asymptotik von V-Statistiken mit degeneriertem Kern wurde von Neuhaus [Neu77] untersucht. Er betrachtete Kerne  $h(x_1, y_1; x_2, y_2)$  für U-Statistiken, die folgende Bedingungen erfüllen:

(a) eine Forderung an die Quadratintegrierbarkeit des Kerns,

$$\mathbb{E} [h^2(X_1, Y_1; X_2, Y_2)] < \infty,$$

(b) eine (im Vergleich zur geläufigen Definition etwas schwächere) Symmetrie-Bedingung,  $h(x_1, y_1; x_2, y_2) = h(x_2, y_2; x_1, y_1)$  für alle  $x_1, x_2, y_1, y_2 \in \mathbb{R}^d$ ,

(c) und eine (im Vergleich zu Sen etwas stärkere) Forderung der Degeneriertheit,

$$\mathbb{E} h(X_1, Y_1, \cdot_1, \cdot_2) = 0 \quad P_1^X \otimes P_1^Y\text{-fast überall}$$

(mit  $P_1^X \otimes P_1^Y$  ist das Produktmaß auf dem Produktraum bezeichnet).

Die Kernfunktion  $\tilde{h}$  erfüllt Neuhaus' stärkere Degeneriertheitsforderung (c) nicht. Um sein Resultat dennoch zu nutzen, können wir seine schwächeren Anforderungen an die Symmetrie (b) ausnutzen, um eine Kernfunktion zu definieren, die alle Bedingungen (a)-(c) erfüllt:

$$\begin{aligned} \bar{h}(x_1, y_1; x_2, y_2) &:= \phi(|x_1 - y_2|^2) + \phi(|x_2 - y_1|^2) \\ &\quad - \phi(|x_1 - x_2|^2) - \phi(|y_1 - y_2|^2). \end{aligned}$$

Dabei werden wir die Quadratintegrierbarkeit des Kerns (a) im Folgenden durch geeignete Voraussetzungen an die Verteilungen für  $P^X, P^Y$  stets erfüllen. Eine hinreichende Voraussetzung ist beispielsweise  $P^X, P^Y \in \mathcal{F}_d^2(\phi)$ . Bedingung (b) gilt offensichtlich und Forderung (c) ist erfüllt, da bei Gültigkeit der Hypothese für  $x, y \in \mathbb{R}^d$  gilt

$$\begin{aligned} \mathbb{E} \bar{h}(X_1, Y_1; x, y) &= \mathbb{E} \phi(|X_1 - y|^2) + \mathbb{E} \phi(|x - Y_1|^2) \\ &\quad - \mathbb{E} \phi(|X_1 - x|^2) - \mathbb{E} \phi(|Y_1 - y|^2) = 0. \end{aligned}$$

Wir können also im Folgenden sowohl Resultate nutzen, die für Kerne gedacht sind, die in jeder Stichprobe symmetrisch sind, als auch das Resultat von Neuhaus, denn

$$\begin{aligned} T_{m,n}^\phi(x_1, \dots, x_m; y_1, \dots, y_n) &= \frac{mn}{m+n} \frac{1}{m^2 n^2} \sum_{i,k=1}^m \sum_{j,\ell=1}^n \tilde{h}(x_i, y_j; x_k, y_\ell) \\ &= \frac{mn}{m+n} \frac{1}{m^2 n^2} \sum_{i,k=1}^m \sum_{j,\ell=1}^n \bar{h}(x_i, y_j; x_k, y_\ell). \end{aligned}$$

Eine Konsequenz der schwächeren Symmetrie ist, dass sich U-Statistiken mit den von Neuhaus verwendeten Kernen nicht wie sonst in der Form (2.6) schreiben lassen. Dies lässt sich einfach, zum Beispiel durch Wahl von  $m = n = 2$ , zeigen. Die von Neuhaus benutzten U-Statistiken haben die Form

$$U_{m,n}^\phi(x_1, \dots, x_m; y_1, \dots, y_n) := \frac{mn}{m+n} \frac{1}{m^2 n^2} \sum_{i \neq k}^m \sum_{j \neq \ell}^n h(x_i, y_j; x_k, y_\ell).$$

Neben der oben genannten Eigenschaft weisen die von ihm verwendeten U-Statistiken zusätzlich einen anderen Normierungsfaktor auf.

Der Erwartungswert der Teststatistik lässt sich auch unter Alternativen leicht bestimmen.

**Lemma 2.3.** *Seien  $\phi \in \Phi$ ,  $P^X, P^Y \in \mathcal{F}_d^1(\phi)$  und  $X_1, \dots, X_m \sim P^X$ ,  $Y_1, \dots, Y_n \sim P^Y$  unabhängig. Dann gilt*

$$ET_{m,n}^\phi = \frac{mn}{m+n} \left[ d_\phi^2(P^X, P^Y) + \frac{1}{m} E\phi(|X_1 - X_2|^2) + \frac{1}{n} E\phi(|Y_1 - Y_2|^2) \right].$$

Unter der Hypothese  $P^X = P^Y$  wird dieser Ausdruck zu

$$ET_{m,n}^\phi = E\phi(|X_1 - X_2|^2).$$

**Beweis.** Beachtet man den Fall gleicher Indizes in den Summen in (1.5), so folgt die Behauptung durch kurze Rechnung.  $\square$

## 2.3 Darstellung mittels empirischer charakteristischer Funktionen

Ist  $\phi \in \Phi$ , also  $\phi : [0, +\infty) \rightarrow \mathbb{R}$  eine stetige Funktion mit nicht-konstanter, vollständig monotoner Ableitung auf  $(0, +\infty)$ , so kann  $\phi'$  gemäß eines Satzes von Bernstein als Laplace-Transformierte eines positiven, nicht notwendigerweise endlichen Radon Maßes  $\sigma$  geschrieben werden,

$$\phi'(z) = \int_{[0, +\infty)} \exp(-\lambda z) d\sigma(\lambda), \quad z > 0, \quad (2.9)$$

vergleiche Satz A.4 im Anhang. Die Endlichkeit des Maßes  $\sigma$  ist äquivalent zur Endlichkeit von  $\phi'(0) := \phi'(0+)$ , siehe Berg, Christensen und Ressel [BCR84, Corollary 4.6.14, S.135]. Nach Vereinbarung ist  $\phi(0) = 0$ , also gilt

$$\begin{aligned} \phi(z) &= \int_0^z \int_{[0, +\infty)} \exp(-\lambda s) d\sigma(\lambda) ds \\ &= z\sigma(\{0\}) + \int_{(0, +\infty)} \frac{1}{\lambda} [1 - \exp(-\lambda z)] d\sigma(\lambda), \quad z \geq 0, \end{aligned} \quad (2.10)$$

wobei die Vertauschung der Integrationsreihenfolge auf Grund des nicht-negativen Integranden nach dem Satz von Fubini erlaubt ist. Die Darstellung von  $\phi$  über ein Maß  $\sigma$  ermöglicht die folgende Identität.

**Lemma 2.4.** *Sei  $\phi \in \Phi$ . Seien  $P^X, P^Y \in \mathcal{F}_d^1(\phi)$  zwei Wahrscheinlichkeitsmaße auf  $\mathcal{B}^d$ . Dann gilt mit  $X \sim P^X$ ,  $Y \sim P^Y$*

$$\begin{aligned} d_\phi^2(P^X, P^Y) &= \int_{(0, +\infty)} \int_{\mathbb{R}^d} \frac{1}{\lambda} \left| \varphi_X(\sqrt{2\lambda}z) - \varphi_Y(\sqrt{2\lambda}z) \right|^2 dN_d(0, I_d)(z) d\sigma(\lambda) \\ &\quad + 2\sigma(\{0\}) |EX - EY|^2, \end{aligned}$$

wobei  $\sigma$  ein der Funktion  $\phi$  durch Gleichung (2.9) zugeordnetes Maß sei und mit

$$\varphi_X(z) := \mathbb{E} \exp(iX'z)$$

die **charakteristische Funktion** der Zufallsvariable  $X$  bezeichnet sei. Definiert man das Maß  $\tau$  auf  $\mathcal{B}^d$  durch

$$\tau(B) := \int_{(0,+\infty)} \int_B \frac{1}{\lambda} dN_d(0, 2\lambda I_d) d\sigma(\lambda), \quad B \in \mathcal{B}^d,$$

so gilt

$$d_\phi^2(P^X, P^Y) = \int |\varphi_X(z) - \varphi_Y(z)|^2 d\tau(z) + 2\sigma(\{0\}) |\mathbb{E}X - \mathbb{E}Y|^2.$$

Das Maß  $\tau$  ist  $\sigma$ -endlich.

**Beweis.** Im Fall  $\sigma(\{0\}) > 0$  folgt zunächst für  $P^X, P^Y \in \mathcal{F}_d^1(\phi)$  mit (2.10), dass die Erwartungswerte  $\mathbb{E}|X|^2$  und  $\mathbb{E}|Y|^2$  endlich sind.

Ist  $Z$  mehrdimensional normalverteilt mit Erwartungswertvektor  $\mu \in \mathbb{R}^d$  und Kovarianzmatrix  $\Sigma \in \mathbb{R}^{d \times d}$ ,  $Z \sim N_d(\mu, \Sigma)$ , so lautet die charakteristische Funktion von  $Z$

$$\varphi_Z(t) := \mathbb{E} \exp(it'Z) = \exp\left(it'\mu - \frac{1}{2}t'\Sigma t\right), \quad t \in \mathbb{R}^d.$$

Mit  $\mu = 0$ ,  $\Sigma = I_d$  folgt

$$\exp(-\lambda|x - y|^2) = \int_{\mathbb{R}^d} \cos(\sqrt{2\lambda}(x - y)'z) dN_d(0, I_d)(z).$$

Nutzt man dies nach Anwendung von (2.10), so erhält man mit unabhängigen Zufallsvariablen  $X_1, X_2 \sim P^X$ ,  $Y_1, Y_2 \sim P^Y$

$$\begin{aligned} d_\phi^2(P^X, P^Y) &= 2\mathbb{E}\phi(|X_1 - Y_1|^2) - \mathbb{E}\phi(|X_1 - X_2|^2) - \mathbb{E}\phi(|Y_1 - Y_2|^2) \\ &= \mathbb{E} [\phi(|X_1 - Y_2|^2) + \phi(|X_2 - Y_1|^2) - \phi(|X_1 - X_2|^2) - \phi(|Y_1 - Y_2|^2)] \\ &= \mathbb{E} \int_{(0,+\infty)} \frac{1}{\lambda} [-\exp(-\lambda|X_1 - Y_2|^2) - \exp(-\lambda|X_2 - Y_1|^2) \\ &\quad + \exp(-\lambda|X_1 - X_2|^2) + \exp(-\lambda|Y_1 - Y_2|^2)] d\sigma(\lambda) \\ &\quad + \sigma(\{0\})\mathbb{E} [|X_1 - Y_2|^2 + |X_2 - Y_1|^2 - |X_1 - X_2|^2 - |Y_1 - Y_2|^2] \\ &= \mathbb{E} \int_{(0,+\infty)} \frac{1}{\lambda} \int_{\mathbb{R}^d} \left[ -\cos(\sqrt{2\lambda}(X_1 - Y_2)'z) - \cos(\sqrt{2\lambda}(X_2 - Y_1)'z) \right. \\ &\quad \left. + \cos(\sqrt{2\lambda}(X_1 - X_2)'z) + \cos(\sqrt{2\lambda}(Y_1 - Y_2)'z) \right] dN_d(0, I_d)(z) d\sigma(\lambda) \\ &\quad + 2\sigma(\{0\}) \mathbb{E} [(X_1 - Y_1)'(X_2 - Y_2)] \end{aligned}$$

$$\begin{aligned}
 &= \int_{(0,+\infty)} \frac{1}{\lambda} \mathbb{E} \int_{\mathbb{R}^d} \left[ \exp(i\sqrt{2\lambda}X_1'z) - \exp(i\sqrt{2\lambda}Y_1'z) \right] \\
 &\quad \left[ \exp(-i\sqrt{2\lambda}X_2'z) - \exp(-i\sqrt{2\lambda}Y_2'z) \right] dN_d(0, I_d)(z) d\sigma(\lambda) \\
 &\quad + 2\sigma(\{0\}) |\mathbb{E}X_1 - \mathbb{E}Y_1|^2.
 \end{aligned}$$

Da der Integrand stetig, also produktmessbar und als Differenz nicht-negativer Funktionen darstellbar ist,

$$\begin{aligned}
 &\left[ 1 - \cos\left(\sqrt{2\lambda}(X_1 - Y_2)'z\right) + 1 - \cos\left(\sqrt{2\lambda}(X_2 - Y_1)'z\right) \right] \\
 &- \left[ 1 - \cos\left(\sqrt{2\lambda}(X_1 - X_2)'z\right) + 1 - \cos\left(\sqrt{2\lambda}(Y_1 - Y_2)'z\right) \right],
 \end{aligned}$$

deren Integrale, z.B.

$$\mathbb{E} \int_{(0,+\infty)} \frac{1}{\lambda} \int_{\mathbb{R}^d} \left[ 1 - \cos\left(\sqrt{2\lambda}(X_1 - Y_2)'z\right) \right] dN_d(0, I_d)(z) d\sigma(\lambda) \leq \mathbb{E}\phi(|X_1 - Y_2|^2)$$

endlich sind, lässt sich die Integrationsreihenfolge im letzten Schritt vertauschen (die zusätzlich auftretenden komplexwertigen Ausdrücke verschwinden bei Integration über die sphärisch symmetrische Normalverteilung). Mittels algebraischer Induktion lässt sich zeigen, dass sich das nach einer weiteren Vertauschung der Integrationsreihenfolge ergebende Integral

$$\int_{(0,+\infty)} \frac{1}{\lambda} \int_{\mathbb{R}^d} \left| \varphi_X(\sqrt{2\lambda}z) - \varphi_Y(\sqrt{2\lambda}z) \right|^2 dN_d(0, I_d)(z) d\sigma(\lambda)$$

in der vorletzten Zeile der obigen Rechnung durch das Integral

$$\int |\varphi_X(z) - \varphi_Y(z)|^2 d\tau(z)$$

bezüglich des angegebenen Maßes  $\tau$  ersetzen lässt.

Wir zeigen nun die  $\sigma$ -Endlichkeit von  $\tau$ . Sei  $B_k := \{z \in \mathbb{R}^d : |z| \geq 1/k\}$ ,  $k \in \mathbb{N}$ , und  $B_0 := \{0\}$ . Dann gilt für  $k \in \mathbb{N}$

$$\int_{B_k} dN_d(0, 2\lambda I_d) = P\left(Z_1^2 + \dots + Z_d^2 \geq \frac{1}{2\lambda k^2}\right), \quad \lambda > 0,$$

wobei  $Z_1, \dots, Z_d$  unabhängig und identisch verteilt (u.i.v.) seien mit einer Standardnormalverteilung,  $N(0, 1)$ . Mit u.i.v.  $Z_1, \dots, Z_{2d} \sim N(0, 1)$  gilt

$$\begin{aligned}
 \int_{B_k} dN_d(0, 2\lambda I_d) &\leq P\left(Z_1^2 + \dots + Z_{2d}^2 \geq \frac{1}{2\lambda k^2}\right) \\
 &\leq d P\left(Z_1^2 + Z_2^2 \geq \frac{1}{2\lambda k^2 d}\right) \\
 &= d \exp\left(-\frac{1}{4\lambda k^2 d}\right), \quad \lambda > 0,
 \end{aligned}$$

da  $Z_1^2 + Z_2^2$  exponentialverteilt ist mit Erwartungswert 2. Bekanntlich gilt

$$\log(x) \geq 1 - \frac{1}{x}, \quad x > 0, \quad \text{und} \quad \exp(x) \geq 1 + x, \quad x \in \mathbb{R}.$$

Wendet man diese beiden Ungleichungen nacheinander an, so erhält man

$$\log(1 - \exp(-x)) \geq 1 - \frac{1}{1 - \exp(-x)} = \frac{1}{1 - \exp(x)} \geq -\frac{1}{x}, \quad x > 0,$$

und daher

$$\exp\left(-\frac{1}{x}\right) \leq 1 - \exp(-x), \quad x > 0.$$

Diese Ungleichung ist von eigenständigem Interesse. Sie besagt, dass die Verteilungsfunktion einer exponentialverteilten Zufallsvariable  $X$  mit Erwartungswert 1 stets größer oder gleich der Verteilungsfunktion von  $X^{-1}$  ist. Eine solche exponentialverteilte Zufallsvariable ist also stochastisch kleiner als ihr Kehrwert. In unserem Fall verhilft die Abschätzung zu

$$\int_{B_k} dN_d(0, 2\lambda I_d) \leq d(1 - \exp(-4\lambda k^2 d)), \quad \lambda > 0,$$

und damit

$$\tau(B_k) \leq \int_{(0, +\infty)} \frac{d}{\lambda} (1 - \exp(-4\lambda k^2 d)) d\sigma(\lambda) = d \phi(4dk^2) < \infty.$$

Da die  $B_k$ ,  $k \in \mathbb{N}_0$  aus den natürlichen Zahlen einschließlich 0, den  $\mathbb{R}^d$  ausschöpfen, ist  $\tau$   $\sigma$ -endlich. Insbesondere wird jedem Kompaktum aus  $\mathbb{R}^d$ , welches nicht den Nullpunkt enthält, mittels  $\tau$  ein endlicher Wert zugeordnet, da es vom Nullpunkt weg beschränkt ist.  $\square$

Das Lemma liefert einen alternativen Beweis zur Aussage  $(b) \implies (a)$  von Satz 2.1.

Der Term  $z\sigma(\{0\})$  in Gleichung (2.10) macht für  $\sigma(\{0\}) > 0$  die Behandlung der Testgröße und ihres Grenzverhaltens in einem Hilbertraum umständlich. Theoretisch ist es möglich, den Abstand  $d_\phi^2$  zu schreiben als

$$d_\phi^2(P^X, P^Y) = \int_{[0, +\infty)} \int_{\mathbb{R}^d} \mathbb{E} [f(\lambda, z, X) - f(\lambda, z, Y)]^2 dN_d(0, I_d)(z) d\sigma(\lambda)$$

mit

$$\begin{aligned} f(\lambda, z, X) := & \frac{1}{\sqrt{\lambda}} \left[ \cos(\sqrt{2\lambda} z' X) + \sin(\sqrt{2\lambda} z' X) \right. \\ & \left. - \mathbb{E} \left( \cos(\sqrt{2\lambda} z' X) + \sin(\sqrt{2\lambda} z' X) \right) \right] \mathbb{I}(\lambda > 0) \\ & + \sqrt{2} z' \mathbb{E}[X] \mathbb{I}(\lambda = 0), \quad z \in \mathbb{R}^d, \lambda \in [0, +\infty), \end{aligned}$$

denn Aufspalten des Integrationsbereichs von  $\sigma$  liefert

$$\begin{aligned} d_\phi^2(P^X, P^Y) &= \int_{[0, +\infty)} \int_{\mathbb{R}^d} \mathbb{E}[f(\lambda, z, X) - f(\lambda, z, Y)]^2 \, dN_d(0, I_d)(z) \, d\sigma(\lambda) \\ &= \int |\varphi_X(z) - \varphi_Y(z)|^2 \, d\tau(z) + 2\sigma(\{0\}) \int [z'(\mathbb{E}X - \mathbb{E}Y)]^2 \, dN_d(0, I_d)(z) \\ &= \int |\varphi_X(z) - \varphi_Y(z)|^2 \, d\tau(z) + 2\sigma(\{0\}) |\mathbb{E}X - \mathbb{E}Y|^2. \end{aligned}$$

Dabei bezeichnet  $I(\cdot)$  die Indikatorfunktion. Bei der letzten Umformung wurde die Rotationsinvarianz der Normalverteilung  $N_d(0, I_d)$  ausgenutzt:

$$\int [z'a]^2 \, dN_d(0, I_d)(z) = |a|^2 \int \left[ z' \frac{a}{|a|} \right]^2 \, dN_d(0, I_d)(z) = |a|^2, \quad a \in \mathbb{R}^d \setminus \{0\},$$

wobei  $a/|a|$  im Integranden durch jeden Vektor der Länge 1 ersetzt werden kann. Wählt man hier den ersten Einheitsvektor  $(1, 0, \dots, 0)' \in \mathbb{R}^d$ , so folgt sofort

$$\int [z'a]^2 \, dN_d(0, I_d)(z) = |a|^2 \int [z'e_1]^2 \, dN_d(0, I_d)(z) = |a|^2 \int z_1^2 \, dN(0, 1)(z_1) = |a|^2.$$

Bei der Untersuchung der Asymptotik der empirischen Prozesse auf dem Hilbertraum

$$L_2(\mathbb{R}^d \times [0, +\infty), \mathcal{B}^d \times \mathcal{B}_{\geq 0}, N_d(0, I_d) \otimes \sigma)$$

würde man einen Gaußschen Prozess finden. Mittels des Weges über V-Statistiken kann man zeigen, dass sich unter der Hypothese als Grenzverteilung eine gewichtete Summe von  $\chi_1^2$ -verteilten Zufallsvariablen ergibt. Die Gewichte entsprechen den Eigenwerten des Kovarianzoperators des Gaußschen Prozesses, dessen Kern in dem angegebenen Hilbertraum von jeweils zwei Argumenten, also insgesamt vier Argumenten abhängt und dementsprechend unübersichtlich ist. Zieht man sich auf den Fall  $\sigma(\{0\}) = 0$  zurück, so bietet sich der Hilbertraum  $L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$  an, in welchem die Darstellung der Testgröße und die Darstellung des Kovarianzoperators wesentlich übersichtlicher ist. Ein weiterer Grund für die Betrachtung des Falles  $\sigma(\{0\}) = 0$  ist die Tatsache, dass die Funktionen  $\phi$ , für die  $\sigma := c \delta_0$  mit  $c \in (0, +\infty)$  ist, die Geraden  $\phi(z) = cz$  sind, welche nicht in  $\Phi$  enthalten sind, da sie eine konstante Ableitung besitzen. Mit der Darstellung der Ableitung von  $\phi$  über das Maß  $\sigma$  sieht man sofort, dass eine Funktion, für die  $\sigma(\{0\}) > 0$  ist, eine Linearkombination aus einer Geraden und einer Funktion aus  $\Phi$  ist. Überdies ordnen die Maße aller in dieser Arbeit näher untersuchten Funktionen  $\phi$  dem Nullpunkt keine Masse zu (für eine Übersicht siehe Tabelle 2.1). Daher beschränken wir uns bei diesem Weg über empirische charakteristische Funktionen auf diejenigen Funktionen  $\phi \in \Phi$ , für deren durch (2.9) zugeordnetes Maß  $\sigma(\{0\}) = 0$  gilt. Die Menge dieser Funktionen sei mit  $\Phi_0$  bezeichnet. Für  $\phi \in \Phi_0$  gilt also

$$\phi(z) = \int_{(0, +\infty)} \frac{1}{\lambda} [1 - \exp(-\lambda z)] \, d\sigma(\lambda), \quad z \geq 0.$$

Die **empirische charakteristische Funktion** der Beobachtungen  $x_1, \dots, x_m \in \mathbb{R}^d$  ist definiert durch

$$\varphi_{\{x_1, \dots, x_m\}}(z) := \frac{1}{m} \sum_{k=1}^m \exp(ix'_k z), \quad z \in \mathbb{R}^d.$$

Das vorangegangene Lemma führt unmittelbar zu einer neuen Darstellung für die Testgröße.

**Korollar 2.5.** *Sei  $\phi \in \Phi_0$ . Die Teststatistik  $T_{m,n}^\phi$  der  $m$  Beobachtungen der ersten Stichprobe  $x_1, \dots, x_m \in \mathbb{R}^d$  und der  $n$  Beobachtungen  $y_1, \dots, y_n \in \mathbb{R}^d$  der zweiten Stichprobe kann geschrieben werden als*

$$T_{m,n}^\phi(x_1, \dots, x_m; y_1, \dots, y_n) = \frac{mn}{m+n} \int |\varphi_{\{x_1, \dots, x_m\}}(z) - \varphi_{\{y_1, \dots, y_n\}}(z)|^2 d\tau(z).$$

Mit der Beweisführung von Lemma 2.4 zeigt man die Identität

$$\phi(|x-y|^2) = \int [1 - \cos((x-y)'z)] d\tau(z), \quad x, y \in \mathbb{R}^d, \quad (2.11)$$

die in der Folge von Nutzen sein wird.

Tabelle 2.1: Maße  $\sigma$  für unterschiedliche Kernfunktionen  $\phi$

	Kernfunktion	Zugehöriges Maß $\sigma$
Kern von Bahr	$\phi(z) = 1 - \exp(-z/2)$	$\sigma = \frac{1}{2}\delta_{1/2}$
Cramér-Kern	$\phi(z) = \sqrt{z}/2$	$d\sigma(x) = \frac{1}{4\sqrt{\pi}}x^{-1/2} dx$ für $x > 0$
	$\phi(z) = \log(1+z)$	$d\sigma(x) = \exp(-x) dx$ für $x \geq 0$
	$\phi(z) = 1 - 1/(1+z)$	$d\sigma(x) = x \exp(-x) dx$ für $x \geq 0$
	$\phi(z) = 1 - 1/(1+z)^2$	$d\sigma(x) = x^2 \exp(-x) dx$ für $x \geq 0$

Tabelle 2.1 listet exemplarisch die zugehörigen Maße  $\sigma$  zu solchen Kernfunktionen auf, die in späteren Kapiteln genauer untersucht werden. Dabei sei  $dx$  das Lebesgue-Borelsche Maß auf  $[0, +\infty)$ . Setzt man das zur Kernfunktion  $\phi(z) = 1 - \exp(-z/2)$  gehörige Maß  $\sigma = \frac{1}{2}\delta_{1/2}$  in die Darstellung aus Korollar 2.5 ein, so ergibt sich

$$\begin{aligned} T_{m,n}^{\text{Bahr}}(x_1, \dots, x_m; y_1, \dots, y_n) \\ = \frac{mn}{m+n} \int_{\mathbb{R}^d} |\varphi_{\{x_1, \dots, x_m\}}(z) - \varphi_{\{y_1, \dots, y_n\}}(z)|^2 dN_d(0, I_d)(z). \end{aligned}$$

Mit Hilfe dieser Darstellung hat Bahr die Asymptotik seiner Testgröße untersucht. Eine für den vorliegenden, allgemeineren Fall geeignete Herangehensweise an die Untersuchung findet sich in Abschnitt 3.3.

# Kapitel 3

## Die asymptotische Verteilung

In diesem Kapitel wird die Verteilung von  $T_{m,n}^\phi$  für  $m, n \rightarrow \infty$  unter der Hypothese untersucht. Dazu werden die zwei im letzten Kapitel beschriebenen Wege besprochen.

Wir beginnen mit der Betrachtung von  $T_{m,n}^\phi$  als V-Statistik und verwenden ein Resultat von Neuhaus [Neu77], das die Asymptotik von degenerierten U-Statistiken behandelt. Im Folgenden seien  $X_1, \dots, X_m, Y_1, \dots, Y_n$  unabhängige,  $\mathbb{R}^d$ -wertige Zufallsvariablen. Die  $X_1, \dots, X_m$  mögen die Verteilung  $P^X$  haben, die  $Y_1, \dots, Y_n$  die Verteilung  $P^Y$ .

### 3.1 Asymptotik degenerierter V-Statistiken

Um das bereits in Abschnitt 2.2 erwähnte Resultat von Neuhaus für degenerierte U-Statistiken der Kernlänge  $(2, 2)$  verwenden zu können, zerlegen wir

$$T_{m,n}^\phi = \frac{mn}{m+n} d_\phi^2(P_m^X, P_n^Y) = \frac{mn}{m+n} \frac{1}{m^2 n^2} \sum_{i,k=1}^m \sum_{j,\ell=1}^n \bar{h}(X_i, Y_j; X_k, Y_\ell) = U_{m,n}^\phi + R_{m,n}^\phi$$

in eine U-Statistik

$$U_{m,n}^\phi := \frac{mn}{m+n} \frac{1}{m^2 n^2} \sum_{i \neq k}^m \sum_{j \neq \ell}^n \bar{h}(X_i, Y_j; X_k, Y_\ell)$$

und einen Rest

$$R_{m,n}^\phi := \frac{mn}{m+n} \left[ \frac{1}{m^2 n^2} \sum_{i=1}^m \sum_{j,\ell=1}^n \bar{h}(X_i, Y_j; X_i, Y_\ell) + \frac{1}{m^2 n^2} \sum_{i,k=1}^m \sum_{j=1}^n \bar{h}(X_i, Y_j; X_k, Y_j) - \frac{1}{m^2 n^2} \sum_{i=1}^m \sum_{j=1}^n \bar{h}(X_i, Y_j; X_i, Y_j) \right]$$

$$\begin{aligned}
&= \frac{1}{mn(m+n)} \left[ 2(m+n-1) \sum_{i=1}^m \sum_{j=1}^n \phi(|X_i - Y_j|^2) \right. \\
&\quad \left. - n \sum_{i=1}^m \sum_{k=1}^m \phi(|X_i - X_k|^2) - m \sum_{j=1}^n \sum_{\ell=1}^n \phi(|Y_j - Y_\ell|^2) \right]. \tag{3.1}
\end{aligned}$$

Das Maß  $P_m^X$  bezeichnet dabei das zufällige empirische Maß mit Gewichten  $m^{-1}$  an den Stellen  $X_1, \dots, X_m$ . Analog bezeichnet  $P_n^Y$  das entsprechende zufällige empirische Maß der Zufallsvariablen  $Y_1, \dots, Y_n$ . Der Faktor  $\frac{mn}{m+n}$  vor  $d_\phi^2(P_m^X, P_n^Y)$  wird angebracht, um unter der Hypothese eine Grenzverteilung für  $m, n \rightarrow \infty$  zu erhalten.

Die nächsten beiden Abschnitte beschäftigen sich mit der Konvergenz der einzelnen Summanden in

$$T_{m,n}^\phi = \frac{mn}{m+n} d_\phi^2(P_m^X, P_n^Y) = U_{m,n}^\phi + R_{m,n}^\phi, \tag{3.2}$$

für  $m, n \rightarrow \infty$  so, dass  $\frac{m}{m+n} \rightarrow \rho \in (0, 1)$ . Zunächst wird die Differenz der V- und U-Statistik,  $R_{m,n}^\phi$ , untersucht.

### 3.1.1 Konvergenz der Differenz von V- und U-Statistik

Der Term  $R_{m,n}^\phi$  kann als Linearkombination von drei V-Statistiken aufgefasst werden. Eine der V-Statistiken ist eine Zwei-Stichproben-V-Statistik der Kernlänge  $(1, 1)$ , die beiden weiteren sind Ein-Stichproben-V-Statistiken der Kernlänge 2. Eichelsbacher [Eic92] gibt Gesetze der großen Zahlen an, welche die fast sichere Konvergenz der einzelnen Statistiken garantieren. Für die Ein-Stichproben-V-Statistiken genügt die Voraussetzung  $P^X, P^Y \in \mathcal{F}_d^1(\phi)$  für die Konvergenz. Bei der Zwei-Stichproben-V-Statistik ist die Voraussetzung  $P^X, P^Y \in \mathcal{F}_d^2(\phi)$  hinreichend.

**Lemma 3.1.** *Es sei  $P^X, P^Y \in \mathcal{F}_d^2(\phi)$ . Es sei  $X_1, X_2, \dots \sim P^X$ ,  $Y_1, Y_2, \dots \sim P^Y$  unabhängig. Die Differenz von V- und U-Statistik konvergiert für  $m, n \rightarrow \infty$  mit  $\frac{m}{m+n} \rightarrow \rho \in (0, 1)$  fast sicher*

$$R_{m,n}^\phi \longrightarrow 2\mathbb{E}\phi(|X_1 - Y_1|^2) - \rho\mathbb{E}\phi(|X_1 - X_2|^2) - (1 - \rho)\mathbb{E}\phi(|Y_1 - Y_2|^2).$$

Unter der Hypothese  $P^X = P^Y$  wird dies zu

$$R_{m,n}^\phi \longrightarrow \mathbb{E}\phi(|X_1 - X_2|^2)$$

fast sicher.

**Beweis.** Die Behauptung folgt mittels Satz B.4, Anhang B, vergleiche auch Koroľjuk und Borovskich [KB94, Theorem 3.1.1, S.93, Theorem 3.2.3, S.101], aus der

Darstellung

$$\begin{aligned}
 R_{m,n}^\phi &= 2 \frac{m+n-1}{m+n} \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \phi(|X_i - Y_j|^2) \\
 &\quad - \frac{m}{m+n} \frac{1}{m^2} \sum_{i=1}^m \sum_{k=1}^m \phi(|X_i - X_k|^2) \\
 &\quad - \frac{n}{m+n} \frac{1}{n^2} \sum_{j=1}^n \sum_{\ell=1}^n \phi(|Y_j - Y_\ell|^2).
 \end{aligned} \tag{3.3}$$

□

Da der Grenzwert des Rests eine Konstante ist, ist im Hinblick auf die Verteilung der Testgröße die getrennte Behandlung der Terme in (3.2) und (3.3) trotz ihrer Abhängigkeit zulässig.

### 3.1.2 Konvergenz der U-Statistik

In diesem Abschnitt werden wir Neuhaus' Resultat auf die vorliegende Situation anwenden und anschließend vereinfachen. Dazu wird mit den Hilberträumen  $\mathcal{L}_2 := L_2(\mathbb{R}^{2d}, \mathcal{B}^{2d}, P^X \otimes P^X)$  und  $\mathcal{L}_2 := L_2(\mathbb{R}^d, \mathcal{B}^d, P^X)$  gearbeitet. Abkürzend sei  $\bar{\gamma} := P^X \otimes P^X$  und  $\gamma := P^X$ . Das jeweilige Innenprodukt sei in beiden Fällen mit  $\langle \cdot, \cdot \rangle$  bezeichnet. Zur Theorie der Operatoren auf Hilberträumen sei auf den Anhang C hingewiesen, in dem die hier verwendeten Begriffe und Sätze zusammengefasst sind. Das bereits zu Beginn dieses Kapitels erwähnte Resultat von Neuhaus ist anwendbar auf U-Statistiken mit Kernen der Länge  $(2, 2)$ , welche die auf S.13f genannten Bedingungen (a)-(c) erfüllen. Diese Bedingungen werden von dem für unsere V-Statistik verwendeten Kern

$$\begin{aligned}
 \bar{h}(x_1, y_1; x_2, y_2) &:= \phi(|x_1 - y_2|^2) + \phi(|x_2 - y_1|^2) \\
 &\quad - \phi(|x_1 - x_2|^2) - \phi(|y_1 - y_2|^2)
 \end{aligned}$$

unter der Hypothese  $P^X = P^Y$  erfüllt, sofern der Kern quadratintegrierbar ist. Die bereits in Lemma 3.1 gestellte Forderung  $P^X, P^Y \in \mathcal{F}_d^2(\phi)$  garantiert dies. Es folgt der Satz von Neuhaus angewendet auf die vorliegende Situation.

**Satz 3.2 (Neuhaus [Neu77, Theorem 1.1, S.425]).** *Sei  $\phi \in \Phi$  und die Verteilung  $P^X \in \mathcal{F}_d^2(\phi)$ . Es seien  $X_1, X_2, \dots, Y_1, Y_2, \dots \sim P^X$  unabhängig. Dann gilt für  $m, n \rightarrow \infty$ , so dass  $\frac{m}{m+n} \rightarrow \rho \in (0, 1)$*

$$U_{m,n}^\phi \xrightarrow[m, n \rightarrow \infty]{\mathcal{D}} U$$

mit

$$U := \sum_{\sigma} \bar{\lambda}_{\sigma} [(a_{\sigma} Z_{\sigma}^{(1)} + b_{\sigma} Z_{\sigma}^{(2)})^2 - (a_{\sigma}^2 + b_{\sigma}^2)]. \quad (3.4)$$

Dabei seien  $(Z_{\sigma}^{(1)})_{\sigma}, (Z_{\sigma}^{(2)})_{\sigma}$  zwei unabhängige Folgen unabhängiger standardnormalverteilter Zufallsvariablen. Die Konstanten  $a_{\sigma}, b_{\sigma}$  sind gegeben durch

$$\begin{aligned} a_{\sigma}^2 &= (1 - \rho) \int (\mathbb{E} \bar{f}_{\sigma}(x, Y_1))^2 d\gamma(x), \\ b_{\sigma}^2 &= \rho \int (\mathbb{E} \bar{f}_{\sigma}(X_1, y))^2 d\gamma(y). \end{aligned}$$

Die Funktionen  $(\bar{f}_{\sigma})_{\sigma}, \bar{f}_{\sigma} \in \bar{\mathcal{L}}_2$  bezeichnen eine Abzählung der orthonormalen Eigenfunktionen mit

$$\int \bar{f}_{\sigma} d\bar{\gamma} = 0$$

des Operators

$$\begin{aligned} \bar{H} : \quad \bar{\mathcal{L}}_2 &\rightarrow \bar{\mathcal{L}}_2, \\ f(\cdot_1, \cdot_2) &\mapsto \int \bar{h}(\cdot_1, \cdot_2; x, y) f(x, y) d\bar{\gamma}(x, y) \end{aligned}$$

zu den entsprechend ihrer Vielfachheit gezählten Eigenwerten  $(\bar{\lambda}_{\sigma})_{\sigma}$ .

Da  $P^X \in \mathcal{F}_d^1(\phi)$ , ist  $P^X$  nicht degeneriert. Der Operator  $\bar{H}$  ist also ungleich dem Nulloperator, so dass wir Konvergenz gegen eine nicht-konstante Zufallsvariable erhalten.

Dieses Resultat kann weitgehend vereinfacht werden. Dazu ist es erforderlich, die Eigenschaften des Operators  $\bar{H}$  zu betrachten.

**Lemma 3.3.** *Es seien die Voraussetzungen von Satz 3.2 erfüllt. Dann ist der Operator  $\bar{H}$  von der Hilbert-Schmidt-Klasse, kompakt, selbstadjungiert und normal. Der Kern des Operators kann geschrieben werden als*

$$\bar{h}(x_1, y_1; x_2, y_2) = \sum_{\sigma} \bar{\lambda}_{\sigma} \bar{f}_{\sigma}(x_1, y_1) \bar{f}_{\sigma}(x_2, y_2), \quad (3.5)$$

wobei  $(\bar{f}_{\sigma})_{\sigma}$  eine Aufzählung der normierten Eigenfunktionen zu den Eigenwerten  $(\bar{\lambda}_{\sigma})_{\sigma}$  sei. Die Konvergenz in der Reihe (3.5) ist in  $L_2((\mathbb{R}^{2d}, \mathcal{B}^{2d}, \bar{\gamma}) \times (\mathbb{R}^{2d}, \mathcal{B}^{2d}, \bar{\gamma}))$ .

**Beweis.** Wegen der Symmetrie des Kerns  $(b)$  ist  $\bar{H}$  selbstadjungiert. Damit ist  $\bar{H}$  ein normaler Operator. Da der Kern  $\bar{h}$  quadratintegrierbar ist, ist nach Dunford und Schwartz, Lemma C.6,  $\bar{H}$  ein Operator der Hilbert-Schmidt-Klasse und somit kompakt. Es folgt mit Lemma C.7 die Darstellung (3.5).  $\square$

Der Kern  $\bar{h}$  hat die wichtige Eigenschaft  $\bar{h}(x_1, y_1; x_2, y_2) = -\bar{h}(y_1, x_1; x_2, y_2)$ . Es folgt somit

$$\begin{aligned}
 \bar{\lambda}_\sigma \bar{f}_\sigma(x, y) &= (\bar{H} \bar{f}_\sigma)(x, y) \\
 &= \int \bar{h}(x, y, u, v) \bar{f}_\sigma(u, v) d\bar{\gamma}(u, v) \\
 &= - \int \bar{h}(y, x, u, v) \bar{f}_\sigma(u, v) d\bar{\gamma}(u, v) \\
 &= -(\bar{H} \bar{f}_\sigma)(y, x) = -\bar{\lambda}_\sigma \bar{f}_\sigma(y, x),
 \end{aligned} \tag{3.6}$$

also  $a_\sigma^2/(1 - \rho) = b_\sigma^2/\rho$ . Da die Wurzelkonvexkombination zweier unabhängiger, standardnormalverteilter Variablen wieder standardnormalverteilt ist, vereinfacht sich die Zufallsvariable (3.4) zu

$$U = \sum_{\sigma} \bar{\lambda}_\sigma c_\sigma^2 [Z_\sigma^2 - 1]$$

mit

$$c_\sigma^2 := a_\sigma^2 + b_\sigma^2 = \int [\mathbb{E} \bar{f}_\sigma(x, Y_1)]^2 d\gamma(x) = \int [\mathbb{E} \bar{f}_\sigma(X_1, y)]^2 d\gamma(y).$$

Die  $c_\sigma$  lassen sich mit folgendem Lemma noch weiter vereinfachen.

**Lemma 3.4.** *Es seien die Voraussetzungen von Satz 3.2 erfüllt. Die Funktionen  $(f_\sigma)_\sigma$  mit  $f_\sigma : \mathbb{R}^d \rightarrow \mathbb{R}$ , definiert durch*

$$f_\sigma(x) := \sqrt{2} \mathbb{E} \bar{f}_\sigma(x, Y_1), \quad x \in \mathbb{R}^d,$$

sind orthonormale Eigenfunktionen des Operators

$$\begin{aligned}
 \mathcal{L}_2 &\rightarrow \mathcal{L}_2, \\
 H : f &\mapsto \int h(\cdot, x) f(x) d\gamma(x)
 \end{aligned}$$

mit  $h(x, y) := \mathbb{E} \bar{h}(x, Y_1; y, Y_2)$  zu den positiven Eigenwerten  $(\bar{\lambda}_\sigma/2)_\sigma$ .

**Beweis.** Es seien  $\bar{f}_\sigma, \bar{f}_\tau$  Eigenfunktionen des Operators  $\bar{H}$  zu den Eigenwerten  $\bar{\lambda}_\sigma, \bar{\lambda}_\tau$  und  $f_\sigma, f_\tau$  wie im Lemma angegeben definiert. Die nützliche Antisymmetrie (3.6) verhilft zur Rechnung

$$\begin{aligned}
 \langle f_\sigma, f_\tau \rangle &:= 2 \int [\mathbb{E} \bar{f}_\sigma(x, Y_1)] [\mathbb{E} \bar{f}_\tau(x, Y_2)] d\gamma(x) \\
 &= 2 \int \bar{f}_\sigma(x, y_1) \bar{f}_\tau(x, y_2) d\gamma(y_1) d\gamma(y_2) d\gamma(x) \\
 &= 2\lambda_\sigma^{-1} \lambda_\tau^{-1} \int [\bar{h}(x, y_1; x', y') \bar{h}(x, y_2; x'', y'')] \\
 &\quad \bar{f}_\sigma(x', y') \bar{f}_\tau(x'', y'') d\gamma^7(x', y', x'', y'', y_1, y_2, x)
 \end{aligned}$$

$$\begin{aligned}
&= 2\lambda_\sigma^{-1}\lambda_\tau^{-1} \int [\phi(|x-y|^2) + \phi(|y_1-x'|^2) - \phi(|x-x'|^2) - \phi(|y_1-y'|^2)] \\
&\quad [\phi(|x-y''|^2) + \phi(|y_2-x''|^2) - \phi(|x-x''|^2) - \phi(|y_2-y''|^2)] \\
&\quad \bar{f}_\sigma(x',y')\bar{f}_\tau(x'',y'') d\gamma^7(x',y',x'',y'',y_1,y_2,x) \\
&\stackrel{(3.6)}{=} 2\lambda_\sigma^{-1}\lambda_\tau^{-1} \int [2\phi(|x-y'|^2) + 2\phi(|y_1-x'|^2)] \\
&\quad [2\phi(|x-y''|^2) + 2\phi(|y_2-x''|^2)] \\
&\quad \bar{f}_\sigma(x',y')\bar{f}_\tau(x'',y'') d\gamma^7(x',y',x'',y'',y_1,y_2,x) \\
&\stackrel{(3.6)}{=} 2\lambda_\sigma^{-1}\lambda_\tau^{-1} \int 4 [\phi(|x-y'|^2)\phi(|x-y''|^2) + \phi(|y-x'|^2)\phi(|x-y''|^2)] \\
&\quad \bar{f}_\sigma(x',y')\bar{f}_\tau(x'',y'') d\gamma^6(x',y',x'',y'',y,x) \\
&\stackrel{(3.6)}{=} 2\lambda_\sigma^{-1}\lambda_\tau^{-1} \int [ 2\phi(|x-y'|^2)\phi(|x-y''|^2) + 2\phi(|y-x'|^2)\phi(|y-x''|^2) \\
&\quad + 2\phi(|y-x'|^2)\phi(|x-y''|^2) + 2\phi(|x-y'|^2)\phi(|y-x''|^2)] \\
&\quad \bar{f}_\sigma(x',y')\bar{f}_\tau(x'',y'') d\gamma^6(x',y',x'',y'',y,x) \\
&= \lambda_\sigma^{-1}\lambda_\tau^{-1} \int [2\phi(|x-y'|^2) + 2\phi(|y-x'|^2)] \\
&\quad [2\phi(|x-y''|^2) + 2\phi(|y-x''|^2)] \\
&\quad \bar{f}_\sigma(x',y')\bar{f}_\tau(x'',y'') d\gamma^6(x',y',x'',y'',y,x) \\
&\stackrel{\text{wie}}{\text{zuvor}} \lambda_\sigma^{-1}\lambda_\tau^{-1} \int [\bar{h}(x,y;x',y')\bar{h}(x,y;x'',y'')] \\
&\quad \bar{f}_\sigma(x',y')\bar{f}_\tau(x'',y'') d\gamma^6(x',y',x'',y'',y,x) \\
&= \langle \bar{f}_\sigma, \bar{f}_\tau \rangle,
\end{aligned}$$

wobei  $\gamma^n$  das  $n$ -fache Produktmaß von  $\gamma$  bezeichnet

$$\gamma^n = \underbrace{\gamma \otimes \dots \otimes \gamma}_{n\text{-mal}}.$$

Die Orthonormalität der  $(\bar{f}_\sigma)_\sigma$  überträgt sich also auf die  $(f_\sigma)_\sigma$ .  $f_\sigma$  ist eine Eigenfunktion von  $H$  zum (positiven) Eigenwert  $\bar{\lambda}_\sigma/2$ , da

$$\begin{aligned}
f_\sigma(x) &= \sqrt{2}\mathbb{E}\bar{f}_\sigma(x, Y_1) \\
&= \sqrt{2}\lambda_\sigma^{-1}\mathbb{E}[(\bar{H}\bar{f}_\sigma)(x, Y_1)] \\
&= \sqrt{2}\lambda_\sigma^{-1} \int \bar{h}(x,y;x',y')\bar{f}_\sigma(x',y') d\gamma^3(x',y',y) \\
&= \sqrt{2}\lambda_\sigma^{-1} \int [\phi(|x-y'|^2) + \phi(|y-x'|^2) - \phi(|x-x'|^2) - \phi(|y-y'|^2)] \\
&\quad \bar{f}_\sigma(x',y') d\gamma^3(x',y',y)
\end{aligned}$$

$$\begin{aligned}
 & \stackrel{(3.6)}{=} 2\sqrt{2}\lambda_\sigma^{-1} \int [-\phi(|x-x'|^2) + \phi(|y-x'|^2)] \bar{f}_\sigma(x', y') d\gamma^3(x', y', y) \\
 & \stackrel{(\star)}{=} 2\sqrt{2}\lambda_\sigma^{-1} \int [-\phi(|x-x'|^2) + \phi(|y_1-x'|^2) + \phi(|x-y_2|^2) - \phi(|y_1-y_2|^2)] \\
 & \quad \bar{f}_\sigma(x', y') d\gamma^4(x', y', y_1, y_2) \\
 & = 2\lambda_\sigma^{-1} \int E\bar{h}(x, Y_1; x', Y_2) f_\sigma(x') d\gamma(x') \\
 & = 2\lambda_\sigma^{-1} (Hf_\sigma)(x),
 \end{aligned}$$

wobei in  $(\star)$  die Eigenschaft  $\int \bar{f} d\bar{\gamma} = 0$  der orthonormalen Eigenfunktionen ausgenutzt wurde, um die letzten beiden Summanden hinzuzufügen. Weiterhin ist dort  $y$  in  $y_1$  umbenannt worden.  $\square$

Dieses Lemma zeigt  $c_\sigma^2 = \langle \frac{1}{\sqrt{2}}f_\sigma, \frac{1}{\sqrt{2}}f_\sigma \rangle = \frac{1}{2}$ . Die Eigenfunktionen  $(\bar{f}_\sigma)_\sigma$  bzw.  $(f_\sigma)_\sigma$  sind somit irrelevant für die Grenzverteilung

$$U = \sum_\sigma \frac{\bar{\lambda}_\sigma}{2} [Z_\sigma^2 - 1]$$

von  $\frac{mn}{m+n} U_{m,n}^\phi$  für  $m, n \rightarrow \infty$  mit  $\frac{m}{m+n} \rightarrow \rho \in (0, 1)$ . Auch der Mischungsparameter  $\rho$  taucht nicht mehr auf.

Das letzte Lemma lässt hoffen, dass sich statt der (halben) Eigenwerte von  $\bar{H}$  auch die Eigenwerte von  $H$  als Gewichte in der U-Statistik verwenden lassen. Dazu ist eine Bijektion zwischen den von Null verschiedenen Eigenwerten erforderlich. Dies lässt sich erreichen, indem man zunächst zeigt, dass alle Eigenwerte von  $H$  nicht-negativ sind. Damit sind auch alle Eigenwerte von  $\bar{H}$  dank des letzten Lemmas nicht-negativ. Man sagt, ein Operator sei von der Spurklasse, wenn die Summe der Beträge seiner Eigenwerte endlich ist. Bei Integraloperatoren des hier verwandten Typs kann man in einem solchen Fall zeigen (vergleiche Brislawn, Lemma C.10), dass die Summe der Eigenwerte dem Integral über die Diagonale des Kerns entspricht. Wäre also  $H$  (und damit auch  $\bar{H}$ ) von der Spurklasse und hätte  $H$  ausschließlich nicht-negative Eigenwerte, so folgt

$$\begin{aligned}
 \sum_\sigma \lambda_\sigma &= \int h(x, x) d\gamma(x) \\
 &= E\phi(|X_1 - X_2|^2) \\
 &= \frac{1}{2} \int \bar{h}(x, y; x, y) d\bar{\gamma}(x, y) = \frac{1}{2} \sum_\sigma \bar{\lambda}_\sigma.
 \end{aligned}$$

Dies ist gleichbedeutend mit einer Bijektion zwischen den positiven Eigenwerten von  $H$  und  $\bar{H}$ . Für die Grenzverteilung der V-Statistik ergäbe sich sodann mit Lemma 3.1 und Satz 3.2 als Summe der asymptotischen Verteilung der U-Statistik und des Grenzwertes des Rests

$$\sum_\sigma \lambda_\sigma [Z_\sigma^2 - 1] + E\phi(|X_1 - X_2|^2) = \sum_\sigma \lambda_\sigma Z_\sigma^2$$

als asymptotische Verteilung der V-Statistik für  $m, n \rightarrow \infty$ .

Wir zeigen nun, dass der Operator  $H$  von der Spurklasse ist, und alle von 0 verschiedenen Eigenwerte positiv sind. Letzteres ist genau dann der Fall, wenn  $H$  positiv ist, d.h.  $\langle Hf, f \rangle \geq 0$  für alle  $f \in \mathcal{L}_2$  ist.

**Satz 3.5.** *Sei  $\phi \in \Phi$  und  $P^X \in \mathcal{F}_d^2(\phi)$ . Dann ist der Operator  $H$  positiv und von der Spurklasse.*

**Beweis.** Für den Nachweis der Spurklasseneigenschaft benötigen wir den auf

$$K = K(k) := [-k, +k]^d, \quad k \in (0, +\infty),$$

eingeschränkten Operator

$$H^K : \begin{array}{l} L_2(K, \mathcal{B}_K, \gamma^K) \rightarrow L_2(K, \mathcal{B}_K, \gamma^K) \\ f \mapsto \int h^K(\cdot, x) f(x) d\gamma^K(x), \end{array}$$

wobei  $h^K(x, y) := \int \bar{h}(x, u; y, v) d(\gamma^K \otimes \gamma^K)(u, v)$  und  $\gamma^K$  die Einschränkung von  $\gamma$  auf die Borelsche Spur- $\sigma$ -Algebra auf  $K$ ,  $\mathcal{B}_K$ , sei. Sei nun die Funktion  $f \in L_2(K, \mathcal{B}_K, \gamma^K)$  beschränkt, dann gilt mit den Bezeichnungen von Lemma 2.4

$$\begin{aligned} \langle H^K f, f \rangle &= \int (H^K f)(x) f(x) d\gamma^K(x) \\ &= \int \bar{h}(x_1, y_1; x_2, y_2) \mathbf{I}(x_1, y_1, x_2, y_2 \in K) f(x_1) f(x_2) d\gamma^4(x_1, y_1, x_2, y_2) \\ &= \mathbf{E} \left[ (\phi(|X_1 - Y_2|^2) + \phi(|X_2 - Y_1|^2) - \phi(|Y_1 - Y_2|^2) - \phi(|X_1 - X_2|^2)) \right. \\ &\quad \left. \mathbf{I}(X_1, Y_1, X_2, Y_2 \in K) f(X_1) f(X_2) \right] \\ &= \mathbf{E} \int [(\exp(-iX_1'z) - \exp(-iY_1'z)) \mathbf{I}(X_1, Y_1 \in K) f(X_1) \\ &\quad (\exp(-iX_2'z) - \exp(-iY_2'z)) \mathbf{I}(X_2, Y_2 \in K) f(X_2)] d\tau(z) \\ &\quad + \sigma(\{0\}) \mathbf{E} \left[ (|X_1 - Y_2|^2 + |X_2 - Y_1|^2 - |Y_1 - Y_2|^2 - |X_1 - X_2|^2) \right. \\ &\quad \left. \mathbf{I}(X_1, Y_1, X_2, Y_2 \in K) f(X_1) f(X_2) \right] \\ &= \int |\mathbf{E}[(\exp(-iX_1'z) - \exp(-iY_1'z)) \mathbf{I}(X_1, Y_1 \in K) f(X_1)]|^2 d\tau(z) \\ &\quad + 2\sigma(\{0\}) |\mathbf{E}[(X_1 - Y_1) \mathbf{I}(X_1, Y_1 \in K) f(X_1)]|^2 \\ &\geq 0, \end{aligned}$$

wobei  $X_1, X_2, Y_1, Y_2 \sim \gamma$  unabhängig. Die Beschränktheit wurde gefordert, um die Vertauschung von Erwartungswert und Integral zu ermöglichen. Eine unbeschränkte

Funktion  $f \in L_2(K, \mathcal{B}_K, \gamma^K)$  approximiere man durch eine Folge beschränkter Funktionen  $(f_n)_n$ , welche in der  $L_2(K, \mathcal{B}_K, \gamma^K)$ -Norm gegen  $f$  streben. Mit der Cauchy-Schwarzschen Ungleichung folgt unter Berücksichtigung der Quadratintegrierbarkeit des Kerns  $h^K$ , dass

$$\langle H^K f_n, f_n \rangle \xrightarrow{n \rightarrow \infty} \langle H^K f, f \rangle.$$

Also ist  $H^K$  positiv.

Als Nächstes zeigen wir, dass  $H^K$  von der Spurklasse ist.  $H^K$  ist ein Hilbert-Schmidt-Operator, vergleiche Lemma C.6, der offensichtlich (siehe Lemma 3.3) beschränkt, kompakt und selbstadjungiert ist. Es existiert daher (vergleiche Lemma C.7) zu jedem  $k \in (0, +\infty)$  ein  $L_2(K, \mathcal{B}_K, \gamma^K)$ -Orthonormalsystem von Eigenfunktionen  $(f_\sigma^K)$  zu den höchstens abzählbar vielen positiven Eigenwerten  $(\lambda_\sigma^K)$  von  $H^K$ , so dass

$$h^K(x, y) = \sum_{\sigma} \lambda_{\sigma}^K f_{\sigma}^K(x) f_{\sigma}^K(y), \quad (3.7)$$

wobei es sich um Konvergenz in  $L_2((K, \mathcal{B}_K, \gamma^K) \times (K, \mathcal{B}_K, \gamma^K))$  handelt. Der Kern  $h^K$  ist stetig auf  $K \times K$ , so dass eine verallgemeinerte Form des Satzes von Mercer (vergleiche Lemma C.8) sogar gleichmäßige Konvergenz in (3.7) liefert. Integration über die Diagonale liefert daher

$$\int h^K(x, x) d\gamma^K(x) = \sum_{\sigma} \lambda_{\sigma}^K.$$

Der Operator  $H^K$  ist von der Spurklasse. Um dies auf den Fall  $k = +\infty$ , also auf den Operator  $H$  zu übertragen, benötigen wir die Konvergenz von  $H^K$  gegen  $H$  in der durch die Operatornorm

$$|H| := \sup_{\|f\| \leq 1} \|Hf\|$$

induzierten Metrik. Die Norm  $\|\cdot\|$  bezeichne die Norm auf  $\mathcal{L}_2$ . Für die Konvergenz fassen wir zunächst  $H^K$  als Operator auf  $\mathcal{L}_2$  auf. Für eine Funktion  $f \in \mathcal{L}_2$  mit  $\|f\| \leq 1$  gilt dann mit der Cauchy-Schwarzschen Ungleichung

$$\begin{aligned} & \| [H - H^K]f \|^2 \\ &= \int \left[ \int [\bar{h}(x_1, y_1; x_2, y_2) - \bar{h}(x_1, y_1; x_2, y_2) \mathbf{I}(x_1, y_1, x_2, y_2 \in K)] f(x_2) \right. \\ & \qquad \qquad \qquad \left. d(\bar{\gamma} \otimes \gamma)(y_1, y_2, x_2) \right]^2 d\gamma(x_1) \\ &\leq \int \bar{h}^2(x_1, y_1; x_2, y_2) [1 - \mathbf{I}(x_1, y_1, x_2, y_2 \in K)] d\gamma^4(x_1, y_1, x_2, y_2). \end{aligned}$$

Für  $k \rightarrow \infty$  strebt dieser Ausdruck auf Grund seiner Endlichkeit (Quadratintegrierbarkeit der Kernfunktion) mit Anwendung des Satzes von der majorisierten

Konvergenz gegen 0. Dann existieren nach Dunford und Schwartz, Satz C.4, geeignete Folgen  $(\lambda_\sigma^K)_\sigma$  von Eigenwerten von  $H^K$ , so dass für jedes  $\sigma$  die Eigenwerte  $\lambda_\sigma^K$  für  $k \rightarrow \infty$  gegen  $\lambda_\sigma$  konvergieren. Dabei sei in allen Folgen jeder Eigenwert so oft aufgeführt, wie es seine Vielfachheit angibt. Die Konvergenz ist gleichmäßig in  $\sigma$ . Sei nun  $\hat{\Lambda}$  eine endliche Teilmenge der Indexmenge  $\Lambda$  der Eigenwerte von  $H$ . Dann gilt

$$\begin{aligned} \sum_{\sigma \in \hat{\Lambda}} \lambda_\sigma &= \lim_{k \rightarrow \infty} \sum_{\sigma \in \hat{\Lambda}} \lambda_\sigma^K \\ &\leq \lim_{k \rightarrow \infty} \sum_{\sigma} \lambda_\sigma^K \\ &= \lim_{k \rightarrow \infty} \int h^K(x, x) d\gamma(x) \\ &= \int h(x, x) d\gamma(x), \end{aligned}$$

wobei schließlich der Satz von der majorisierten Konvergenz Verwendung findet. Mit Lemma 2.2 ist diese Oberschranke endlich. Da  $\hat{\Lambda}$  beliebig war, gilt

$$\sum_{\sigma \in \Lambda} \lambda_\sigma \leq \int h(x, x) d\gamma(x) = \mathbb{E}\phi(|X_1 - X_2|^2) < \infty.$$

Der Operator  $H$  ist somit von der Spurklasse. Mit der zuvor für  $H^K$  angebrachten Argumentation sieht man, dass  $H$  positiv ist, also die Eigenwerte von  $H$  nicht-negativ sind. Alternativ folgt dies auch aus der Konvergenz der positiven Eigenwerte von  $H^K$ .  $\square$

Ohne die eben verwendete Beweistechnik wiederholen zu müssen, erhalten wir wie angekündigt als sofortige Konsequenz das folgende Korollar.

**Korollar 3.6.** *Sei  $\phi \in \Phi$  und  $P^X \in \mathcal{F}_d^2(\phi)$ . Dann ist der Operator  $\bar{H}$  positiv und von der Spurklasse.*

Der nächste Satz fasst die vor Satz 3.5 bereits angekündigten Ergebnisse zur Asymptotik zusammen.

**Satz 3.7.** *Sei  $\phi \in \Phi$  und  $P^X = P^Y \in \mathcal{F}_d^2(\phi)$ . Dann gilt*

$$T_{m,n}^\phi \xrightarrow[m,n \rightarrow \infty]{\mathcal{D}} \sum_{\sigma} \lambda_\sigma Z_\sigma^2 \stackrel{\mathcal{D}}{=} \sum_{\sigma} \frac{\bar{\lambda}_\sigma}{2} Z_\sigma^2 \quad (3.8)$$

für  $m, n \rightarrow \infty$  mit  $\frac{m}{m+n} \rightarrow \rho \in (0, 1)$ , wobei  $(Z_\sigma)_\sigma$  eine Folge unabhängiger standardnormalverteilter Zufallsvariablen sei. Die Folge  $(\bar{\lambda}_\sigma)_\sigma$  bezeichne die Eigenwerte des Operators

$$\begin{aligned} \bar{H} : L_2(\mathbb{R}^{2d}, \mathcal{B}^{2d}, \bar{\gamma}) &\rightarrow L_2(\mathbb{R}^{2d}, \mathcal{B}^{2d}, \bar{\gamma}) \quad , \\ f(\cdot_1, \cdot_2) &\mapsto \int \bar{h}(\cdot_1, \cdot_2; x, y) f(x, y) d\bar{\gamma}(x, y) \end{aligned}$$

mit

$$\begin{aligned} \bar{h}(x_1, y_1; x_2, y_2) &:= \phi(|x_1 - y_2|^2) + \phi(|x_2 - y_1|^2) \\ &\quad - \phi(|x_1 - x_2|^2) - \phi(|y_1 - y_2|^2), \quad x_1, y_1, x_2, y_2 \in \mathbb{R}^d. \end{aligned}$$

Die Folge  $(\lambda_\sigma)_\sigma$  bezeichne die Eigenwerte des Operators

$$\begin{aligned} H : L_2(\mathbb{R}^d, \mathcal{B}^d, \gamma) &\rightarrow L_2(\mathbb{R}^d, \mathcal{B}^d, \gamma) \quad , \\ f &\mapsto \int h(\cdot, x) f(x) d\gamma(x) \end{aligned} \quad (3.9)$$

mit  $h(x, y) := E\bar{h}(x, Y_1; y, Y_2)$ ,  $x, y \in \mathbb{R}^d$ . Die Eigenwerte beider Operatoren sind nicht-negativ und die positiven Eigenwerte von  $H$  entsprechen bis auf den Faktor  $\frac{1}{2}$  den positiven Eigenwerte von  $\bar{H}$  (einschließlich Vielfachheit).

Bei der Untersuchung von Bahadur- und Pitman-Effizienzen sowie bei der Betrachtung des Verfahrens als Anpassungstest wird eine Art Ein-Stichproben-Variante des letzten Satzes benötigt.

**Satz 3.8.** Sei  $\phi \in \Phi$ ,  $P^X \in \mathcal{F}_d^2(\phi)$  und  $X_1, X_2, \dots \sim P^X$  unabhängig. Es bezeichne  $P_m^X$  das zufällige empirische Maß der  $m$  Zufallsvariablen  $X_1, \dots, X_m$ . Dann gilt

$$md_\phi^2(P_m^X, P^X) \xrightarrow{m \rightarrow \infty} \sum_\sigma \lambda_\sigma Z_\sigma^2 \stackrel{\mathcal{D}}{=} \sum_\sigma \frac{\bar{\lambda}_\sigma}{2} Z_\sigma^2,$$

wobei die Variablen  $(\lambda_\sigma)_\sigma$ ,  $(\bar{\lambda}_\sigma)_\sigma$  und  $(Z_\sigma)_\sigma$  in der letzten Gleichung wie im vorangegangenen Satz definiert seien.

**Beweis.** Angelehnt an das vorherige Vorgehen teilt man die Summe

$$\begin{aligned} md_\phi^2(P_m^X, P^X) &= \frac{m}{m^2} \sum_{i,j=1}^m \int \bar{h}(X_i, x_1; X_j, x_2) d(P^X \otimes P^X)(x_1, x_2) \\ &= \frac{1}{m} \sum_{i \neq j} \int \bar{h}(X_i, x_1; X_j, x_2) d(P^X \otimes P^X)(x_1, x_2) \\ &\quad + \frac{1}{m} \sum_{i=1}^m \int \bar{h}(X_i, x_1; X_i, x_2) d(P^X \otimes P^X)(x_1, x_2), \end{aligned}$$

wobei fast sicher

$$\frac{1}{m} \sum_{i=1}^m \int \bar{h}(X_i, x_1; X_i, x_2) d(P^X \otimes P^X)(x_1, x_2) \xrightarrow{m \rightarrow \infty} E\phi(|X_1 - X_2|^2).$$

Die Behauptung folgt nun aus Theorem 1.2 von Neuhaus [Neu77, Theorem 1.2, S.426] und Überlegungen analog zum vorherigen Satz.  $\square$

## 3.2 Spezialfälle beim Cramér-Kern

Es ist schwierig, explizite Ergebnisse für die Eigenwerte der Operatoren  $H, \bar{H}$  zu erhalten. Den Fall des Kerns  $\phi(z) = \sqrt{z}/2$  des Cramér-Tests betrachten wir im Eindimensionalen genauer.

Die Verteilung unter der Hypothese  $P^X \in \mathcal{F}_d^2(\phi)$ ,  $\phi(z) = \sqrt{z}/2$ ,  $z \geq 0$ , besitze eine auf einem (gegebenenfalls unbeschränkten) Intervall  $R$  positive, stetige Dichte  $g$  bezüglich des Lebesgue-Borelschen Maßes, die außerhalb von  $R$  verschwindet. Wir suchen nach den Eigenwerten des Operators  $H$

$$f \mapsto \mathbb{E}[h(\cdot, X_2)f(X_2)] = \mathbb{E}[\bar{h}(\cdot, Y_1; X_2, Y_2)f(X_2)]$$

mit unabhängigen  $X_2, Y_1, Y_2 \sim P^X$ . Dann ist  $f$  eine Eigenfunktion von  $H$ , wenn

$$\begin{aligned} 2\lambda f(x) &= 2\mathbb{E}[h(x, X_2)f(X_2)] & (3.10) \\ &= \int [|y - y_1| - |x - y|] f(y)g(y_1)g(y) \, dy_1 \, dy \\ &= \int |y - y_1| f(y)g(y_1)g(y) \, dy_1 \, dy \\ &\quad - \int_x^\infty (y - x)f(y)g(y) \, dy - \int_{-\infty}^x (x - y)f(y)g(y) \, dy \end{aligned}$$

für alle  $x$  aus dem Innern  $\mathring{R}$  des Trägers  $R$  von  $P^X$ , wobei wir die Eigenschaft  $\mathbb{E}f(X) = 0$  orthogonaler Eigenfunktionen  $f$  von  $H$  ausgenutzt haben. Diese Darstellung erlaubt es,  $f$  auf dem Innern des Trägers von  $P^X$  als stetige und differenzierbare Funktion zu behandeln. Leitet man den Ausdruck zweimal nach  $x$  ab, so folgt

$$2\lambda f''(x) = -2f(x)g(x), \quad (3.11)$$

für  $x \in \mathring{R}$ . Das heißt, dass Eigenfunktionen  $f$  zum Eigenwert  $\lambda$  des Integraloperators  $H$  mit  $\mathbb{E}f(X) = 0$  die vorangestellte Differentialgleichung erfüllen müssen. Hat man also eine allgemeine Lösung der Differentialgleichung gefunden, so kann man durch Einsetzen in die Integralgleichung die Eigenwerte ermitteln. Dieses demonstrieren wir am Beispiel der Gleichverteilung.

### 3.2.1 Die Gleichverteilung

Sei die Verteilung unter der Hypothese die Gleichverteilung auf  $R = [-a/2, +a/2]$

$$P^X = \text{unif}([-a/2, +a/2]), \quad a \in (0, +\infty).$$

Die Differentialgleichung (3.11) lautet dann

$$\lambda f''(x) = -\frac{1}{a}f(x), \quad x \in (-a/2, +a/2).$$

Die linear unabhängigen, orthogonalen Lösungen dieser wohlbekanntes Differentialgleichung sind

$$f_1(x) := \sin\left(\frac{x}{\sqrt{\lambda a}}\right) \quad \text{und} \quad f_2(x) := \cos\left(\frac{x}{\sqrt{\lambda a}}\right).$$

Einsetzen in die ursprüngliche Integralgleichung (3.10) liefert

$$\begin{aligned} 2\lambda f_1(x) &= 2\lambda f_1(x) - 4\lambda \frac{\sqrt{\lambda}}{\sqrt{a}} \sin\left(\frac{\sqrt{a}}{2\sqrt{\lambda}}\right), \\ 2\lambda f_2(x) &= 2\lambda f_2(x) - 2x \frac{\sqrt{\lambda}}{\sqrt{a}} \cos\left(\frac{\sqrt{a}}{2\sqrt{\lambda}}\right) \end{aligned}$$

und somit die positiven Eigenwerte

$$\lambda_\sigma = \frac{a}{\sigma^2 \pi^2}, \quad \sigma \in \mathbb{N}.$$

Wir haben keinen Eigenwert übersehen, da  $H$  nach Satz 3.7 von der Spurklasse ist und daher

$$\sum_{\sigma} \lambda_{\sigma} = \frac{a}{6} \stackrel{!}{=} \mathbb{E} \frac{|X - Y|}{2} = \frac{a}{6}.$$

### 3.2.2 Die Normalverteilung

Jetzt sei die Verteilung unter der Hypothese die Standardnormalverteilung  $P^X = N(0, 1)$ . Die Dichtefunktion sei mit  $g$  bezeichnet. Die Differentialgleichung, die es zu lösen gilt, lautet

$$\lambda f''(x) = -g(x)f(x) = -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) f(x), \quad x \in \mathbb{R}. \quad (3.12)$$

Da sich die Lösung dieser Differentialgleichung mit den bekannten Methoden nicht in geschlossener Form ermitteln lässt, müssen wir selbst in diesem scheinbar einfachen Fall auf numerische Methoden zur Bestimmung der Eigenfunktionen und Eigenwerte zurückgreifen.

Wir nehmen nun an, dass die Integralgleichung (3.10) eine ungerade Eigenfunktion besitzt. Für ungerade Eigenfunktionen  $f$  zum Eigenwert  $\lambda$  muss neben der Differentialgleichung auch folgende Gleichung erfüllt sein, welche sich durch Einsetzen der Differentialgleichung (3.11) in die Integralgleichung (3.10) sowie partielle Integration

ergibt:

$$\begin{aligned}
2\lambda f(x) &= \int_{-x}^{+x} yf(y)g(y) \, dy + 2x \int_x^\infty f(y)g(y) \, dy & (3.13) \\
&= -\lambda \int_{-x}^{+x} yf''(y) \, dy + 2x \int_0^\infty f(y)g(y) \, dy + 2\lambda x \int_0^x f''(y) \, dy \\
&= 2\lambda f(x) - 2\lambda x f'(x) + 2x \int_0^\infty f(y)g(y) \, dy + 2\lambda x f'(x) - 2\lambda x f'(0) \\
&= 2\lambda f(x) + 2x \int_0^\infty f(y)g(y) \, dy - 2\lambda x f'(0), \quad x \in \mathbb{R},
\end{aligned}$$

folglich

$$\lambda f'(0) - \int_0^\infty f(x)g(x) \, dx = 0. \quad (3.14)$$

Ohne Beschränkung der Allgemeinheit kann man voraussetzen, dass, falls  $f'(0)$  nicht verschwindet,  $f'(0) = 1$  ist. Mit Hilfe numerischer Methoden lässt sich nun für gegebenes  $\lambda$  eine Lösung  $f_\lambda$  zur Differentialgleichung (3.12) zu den Anfangsbedingungen  $f(0) = 0$ ,  $f'(0) = 1$  bestimmen. Setzt man die Lösung in (3.14) ein, so lässt sich die Funktion

$$\Delta(\lambda) := \lambda - \int_0^\infty f_\lambda(x)g(x) \, dx$$

definieren, deren Nullstellen es zu finden gilt. Jede Nullstelle entspricht einem Eigenwert zu einer ungeraden Eigenfunktion. Numerisch ergibt sich so als größter Eigenwert

$$\lambda_1 \approx 0.2972739$$

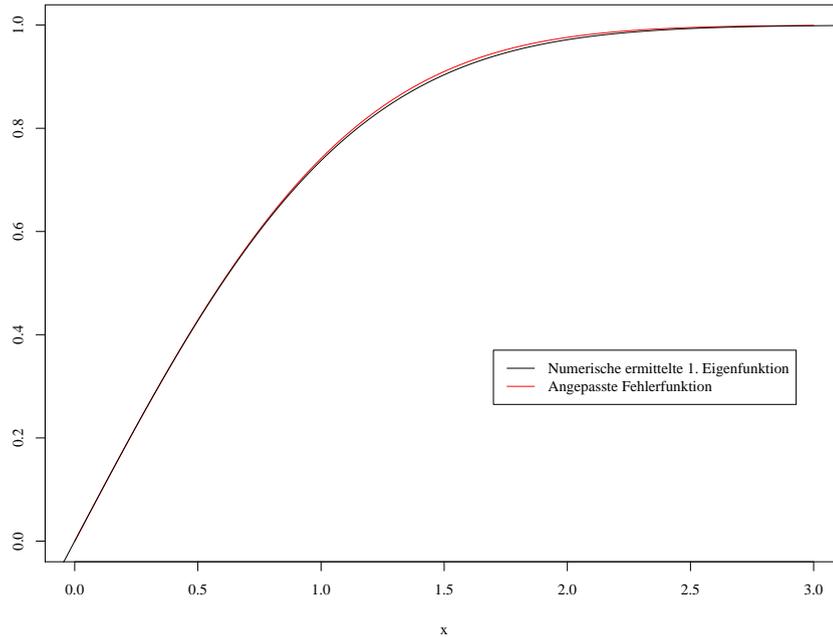
und  $\lambda_3 \approx 0.04484$  als nächster Eigenwert zu einer ungeraden Eigenfunktion. Der Eigenwert  $\lambda_1$  muss einfach sowie der größte sein, denn die Summe der Eigenwerte ist nach dem Resultat des letzten Unterkapitels

$$\sum_\sigma \lambda_\sigma \stackrel{!}{=} \mathbb{E} \frac{|X - Y|}{2} = \frac{1}{\sqrt{\pi}} \approx 0.56189,$$

also kleiner als  $2\lambda_1$ . Durch Diskretisierung des Integraloperators lässt sich nachweisen, dass der so mit Hilfe der Differentialgleichung ermittelte größte Eigenwert samt der zugehörigen Eigenfunktion numerisch mit dem größten Eigenwert des diskretisierten Problems und der zugehörigen Eigenfunktion übereinstimmt. Auch für die geraden Eigenfunktionen lässt sich die Integralgleichung umformen. Statt (3.14) erhält man dann

$$2\lambda \lim_{x \rightarrow \infty} f(x) - \int |y - y_1| f(y)g(y_1)g(y) \, dy_1 \, dy = 0. \quad (3.15)$$

Abbildung 3.1: Die Eigenfunktion zum größten Eigenwert bei Normalverteilung



Auch hier kann man für ein gegebenes  $\lambda$  die Differentialgleichung (3.12) lösen, Gleichung (3.15) als Funktion von  $\lambda$  auffassen und nach Nullstellen suchen. Numerische Experimente lassen allerdings vermuten, dass der Grenzwert  $\lim_{x \rightarrow \infty} f(x)$  nur dann existiert, wenn  $f$  eine Eigenfunktion ist. In Anbetracht dessen, dass für die spätere Untersuchung von approximativen Bahadur-Effizienzen lediglich der bereits gefundene größte Eigenwert von Bedeutung ist, wird auf eine weitere Behandlung dieses Problems verzichtet.

Eine Kuriosität sei jedoch noch bemerkt. Bei der Betrachtung der mit dem beschriebenen Verfahren erhaltenen ersten Eigenfunktion  $f_{\lambda_1}$  (siehe Abbildung 3.1) kann man den Eindruck gewinnen, dass es sich um eine skalierte Fehlerfunktion handelt. Setzt man  $\text{erf}(a \cdot)$ , wobei

$$\text{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt, \quad x \in \mathbb{R},$$

in die umgeformte Integralgleichung (3.13) ein, so kann man zwei Beziehungen zwischen  $\lambda$  und  $a$  ermitteln. Die erste erhält man, indem man  $x \rightarrow \infty$  gehen lässt

$$2\lambda \stackrel{!}{=} \int_{-\infty}^{+\infty} y \text{erf}(ay)g(y) dy = \frac{2a}{\sqrt{\pi}\sqrt{2a^2+1}},$$

und die zweite, indem man die Integralgleichung einmal ableitet und den Grenzwert  $x \rightarrow 0$  betrachtet

$$4 \frac{\lambda a}{\sqrt{\pi}} \stackrel{!}{=} 2 \int_0^{+\infty} \text{erf}(ay)g(y) dy = \frac{2}{\pi} \arctan(a\sqrt{2}).$$

Kombiniert man beide, so erhält man

$$\frac{2a^2}{\sqrt{2a^2 + 1}} = \arctan(a\sqrt{2}),$$

was durch  $a = 0.799177\dots$  gelöst wird. Setzt man nun  $\operatorname{erf}(a \cdot)$  mit dem gefundenen  $a$  in den Rayleigh-Quotienten ein, so erhält man als untere Schranke für den größten Eigenwert von  $H$  den Wert

$$\lambda_1 = \max_f \frac{\langle Hf, f \rangle}{\langle f, f \rangle} \geq \frac{\langle H\operatorname{erf}(a \cdot), \operatorname{erf}(a \cdot) \rangle}{\langle \operatorname{erf}(a \cdot), \operatorname{erf}(a \cdot) \rangle} \approx 0.29727286.$$

Die geringe Abweichung von Fehlerfunktion und tatsächlicher Eigenfunktion (vergleiche Abbildung 3.1) sowie der ermittelten unteren Schranke  $0.2972728\dots$  für den größten Eigenwert  $0.2972739\dots$  ist verblüffend.

Die angestellten Überlegungen zeigen, dass selbst in einfachen Spezialfällen, bei denen sogar die ansonsten unbekannt, beiden Stichproben zu Grunde liegende, gemeinsame Verteilung gegeben ist, die exakte Bestimmung der Eigenwerte schwierig ist. Gesucht sind also numerische Methoden zur Festlegung des kritischen Wertes. Diesen werden wir uns in Kapitel 4 zuwenden.

### 3.3 Asymptotik Hilbertraum-wertiger Zufallselemente

Das Ergebnis dieses Abschnittes wird eine Aussage über die asymptotische Verteilung der Zufallsvariablen  $T_{m,n}^\phi$  für  $m, n \rightarrow \infty$  sein unter der Hypothese, dass beide Stichproben derselben Grundgesamtheit entspringen. Dabei wählen wir eine andere Betrachtungsweise als in Abschnitt 3.1: Wir untersuchen Hilbertraum-wertige Zufallselemente, deren Normquadrat die Testgröße beschreibt.  $T_{m,n}^\phi$  wird wie zuvor angewendet auf  $m$  Zufallsvariablen einer u.i.v. Folge von Zufallsvariablen  $X_1, X_2, \dots$  und  $n$  Zufallsvariablen einer davon unabhängigen u.i.v. Folge von Zufallsvariablen  $Y_1, Y_2, \dots$ , wobei  $X_1, Y_1 \sim P^X$ .

Die in Korollar 2.5 verwendeten, komplexwertigen charakteristischen Funktionen

können umgangen und durch Kosinus-Sinus-Ausdrücke ersetzt werden:

$$\begin{aligned}
 T_{m,n}^\phi &= \frac{mn}{m+n} \int |\varphi_{\{X_1, \dots, X_m\}}(z) - \varphi_{\{Y_1, \dots, Y_n\}}(z)|^2 d\tau(z) \\
 &= \frac{mn}{m+n} \int \left[ \left( \frac{1}{m} \sum_{k=1}^m \cos(X'_k z) - \frac{1}{n} \sum_{\ell=1}^n \cos(Y'_\ell z) \right)^2 \right. \\
 &\quad \left. + \left( \frac{1}{m} \sum_{k=1}^m \sin(X'_k z) - \frac{1}{n} \sum_{\ell=1}^n \sin(Y'_\ell z) \right)^2 \right] d\tau(z) \\
 &= \frac{mn}{m+n} \int \left[ \frac{1}{m} \sum_{k=1}^m (\cos + \sin)(X'_k z) - \frac{1}{n} \sum_{\ell=1}^n (\cos + \sin)(Y'_\ell z) \right]^2 d\tau(z),
 \end{aligned}$$

wobei wir die sphärische Symmetrie von  $\tau$  und die Tatsache, dass  $\cos(\cdot)\sin(\cdot)$  ungerade ist, ausgenutzt haben. Wir können uns somit auf die Betrachtung so genannter **empirischer Kosinus-Sinus-Prozesse**

$$\mathbb{J}_{\{X_1, \dots, X_m\}; P^X}(z) = \sqrt{m} \left( \frac{1}{m} \sum_{k=1}^m (\cos + \sin)(X'_k z) - \mathbb{E}(\cos + \sin)(X'z) \right), \quad z \in \mathbb{R}^d,$$

beschränken. Hierbei sei  $X \sim P^X$ . Unter der Hypothese heben sich die Erwartungswerte in der Teststatistik heraus:

$$T_{m,n}^\phi = \int \left[ \sqrt{\frac{n}{m+n}} \mathbb{J}_{\{X_1, \dots, X_m\}; P^X}(z) - \sqrt{\frac{m}{m+n}} \mathbb{J}_{\{Y_1, \dots, Y_n\}; P^Y}(z) \right]^2 d\tau(z). \quad (3.16)$$

Bahr hat in seiner Arbeit einem Beweis von Giné folgend ein Verteilungskonvergenzresultat für den empirischen Kosinus-Sinus-Prozess als stetigen Prozess auf  $\mathbb{R}^d$  gezeigt. Dafür benötigte er die Endlichkeit des zweiten Moments von  $X$ . Anschließend gelang der Nachweis der Verteilungskonvergenz des Integrals über das Quadrat der Prozesse. Dieser Schritt macht im allgemeineren Fall des nicht notwendigerweise endlichen Maßes  $\tau$  Probleme, da der Satz von der stetigen Abbildung nicht direkt anwendbar ist. Eine Verallgemeinerung der Vorgehensweise von Bahr [Bah96, Satz 4.5, S.43] ist dennoch möglich, wenn man auf eine Technik zurückgreift, die unter anderem von Billingsley allgemein formuliert wurde, und hier wie folgt angewendet werden kann: die Integration über ein Kompaktum, die dortige Anwendung des Satzes von der stetigen Abbildung, der Nachweis, dass sich bei Vergrößerung des Kompaktums eine geeignete Zufallsvariable ergibt (hier ist sogar fast sichere Konvergenz möglich) und das Integral über den Bereich außerhalb des Kompaktums geeignet klein wird.

**Hilfssatz 3.9 (Billingsley).** *Seien  $A_{N,k}, A_N, B_k$  und  $A$  Zufallsvariablen in  $S$ , wobei  $S$  ein separabler, metrischer Raum mit der Metrik  $d$  sei. Falls die Bedingungen*

$$(i) \quad A_{N,k} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} B_k,$$

(ii)  $B_k \xrightarrow[k \rightarrow \infty]{\mathcal{D}} A$  und

(iii) für alle  $\varepsilon > 0$  gilt

$$\lim_{k \rightarrow \infty} \limsup_{N \rightarrow \infty} P(d(A_{N,k}, A_N) \geq \varepsilon) = 0$$

erfüllt sind, gilt  $A_N \xrightarrow[N \rightarrow \infty]{\mathcal{D}} A$ .

**Beweis.** Siehe Billingsley [Bil99, Theorem 3.2, S.28]. □

Unter der Voraussetzung der Endlichkeit des zweiten Moments lässt sich damit die Argumentation von Bahr für die Verteilungskonvergenz der Testgröße sowie die Karhunen-Loève-Entwicklung analog auf unseren Fall übertragen.

Alternativ ist es möglich, den empirischen Kosinus-Sinus-Prozess als Zufallselement im Hilbertraum  $\mathcal{H} := L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$  zu betrachten und den zentralen Grenzwertsatz im Hilbertraum anzubringen, um die Verteilungskonvergenz gegen ein Gaußsches Zufallselement zu zeigen. Die Konvergenz der Testgröße folgt dann aus dem Satz von der stetigen Abbildung und die Karhunen-Loève-Entwicklung aus einer verteilungsgleichen Darstellung des Gaußschen Zufallselementes in der Basis der Eigenfunktionen seines Kovarianzoperators. Wir betrachten nun diesen Weg genauer, da bei dieser Vorgehensweise keine Forderungen an die Endlichkeit der Momente von  $X$  gestellt werden müssen. Die Theorie Hilbertraum-wertiger Zufallselemente ist in Anhang D genauer beleuchtet.

Die Grenzverteilung des empirischen Kosinus-Sinus-Prozesses erhält man ohne Umschweife aus dem zentralen Grenzwertsatz im Hilbertraum.

**Satz 3.10.** *Es sei  $\phi \in \Phi_0$  und  $P^X \in \mathcal{F}_d^1(\phi)$ . Die Zufallsvariablen  $X_1, X_2, \dots$  seien unabhängig und identisch verteilt mit Verteilung  $P^X$ . Dann gilt*

$$\mathbb{J}_{\{X_1, \dots, X_m\}; P^X} \xrightarrow[m \rightarrow \infty]{\mathcal{D}} \mathbb{Z} \quad \text{in } \mathcal{H},$$

wobei  $\mathbb{Z}$  ein zentriertes Gaußsches Zufallselement in  $\mathcal{H}$  ist, welches denselben Kovarianzoperator besitzt wie das Zufallselement

$$\mathbb{X}(z) := (\cos + \sin)(X'_1 z) - E(\cos + \sin)(X'_1 z), \quad z \in \mathbb{R}^d.$$

**Beweis.** Es sei  $\mathcal{A}$  die  $\sigma$ -Algebra des zu Grunde liegenden Wahrscheinlichkeitsraums. Dann ist  $\mathbb{X}$  eine  $\mathcal{A} \otimes \mathcal{B}^d / \mathcal{B}$ -produktmessbare Abbildung, da  $\mathbb{X}$  stetige Pfade besitzt (siehe Dudley [Dud89, Proposition 12.2.1, S.355]).

Der vorliegende Hilbertraum  $L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$  ist separabel (siehe beispielsweise Cohn [Coh80, Proposition 3.4.5, S.110]). Um nachzuweisen, dass eine produktmessbare Abbildung  $\mathbb{X}$  als Hilbertraum-wertiges Zufallselement aufgefasst werden kann, also

$\mathcal{A}/\mathcal{B}(\mathcal{H})$ -messbar ist, wobei  $\mathcal{B}(\mathcal{H})$  die Borelsche  $\sigma$ -Algebra auf  $\mathcal{H}$  ist, genügt es deshalb zu zeigen, dass  $\langle \mathbb{X}, e \rangle$  für alle  $e \in \mathcal{H}$  eine reelle Zufallsvariable ist (siehe Vakhania, Tarieladze und Chobanyan [VTC87, Theorem 1.1, S.89]). Auf Grund der Produktmessbarkeit gilt mit unabhängigen  $X, X_1, X_2 \sim P^X$

$$\begin{aligned} \mathbb{E}\|\mathbb{X}\|^2 &= \mathbb{E} \int [(\cos + \sin)(X'_1 z) - \mathbb{E}(\cos + \sin)(X' z)]^2 d\tau(z) \\ &= \mathbb{E} \int [(\cos + \sin)(X'_1 z) - (\cos + \sin)(X' z)] \\ &\quad [(\cos + \sin)(X'_1 z) - (\cos + \sin)(X'_2 z)] d\tau(z) \\ &= \int \mathbb{E} [1 - \cos((X_1 - X)'z) + 1 - \cos((X_1 - X_2)'z) \\ &\quad - 1 + \cos((X - X_2)'z)] d\tau(z) \\ &= \mathbb{E}\phi(|X_1 - X_2|^2) < \infty. \end{aligned}$$

Fast alle Pfade von  $\mathbb{X}$  sind also Elemente von  $\mathcal{H}$ . Da  $\langle \mathbb{X}, e \rangle$   $\mathcal{A}/\mathcal{B}$ -messbar ist, handelt es sich somit um eine reelle Zufallsvariable, die Erwartungswert 0 hat (Produktmessbarkeit).

Da  $\mathbb{X}$  zentriert ist und  $\mathbb{E}\|\mathbb{X}\|^2 < \infty$  sind die Voraussetzungen des zentralen Grenzwertsatzes im Hilbertraum, Satz D.5 erfüllt und die Behauptung folgt.  $\square$

Der Kovarianzoperator  $C$  von  $\mathbb{X}$  ist definiert durch

$$\langle Ce, e \rangle = \mathbb{E} \left[ \int \mathbb{X}(z)e(z) d\tau(z) \right]^2 = \iint \mathbb{E} [\mathbb{X}(y)\mathbb{X}(z)] e(z) d\tau(z)e(y) d\tau(y),$$

$e \in \mathcal{H}$ . Wegen der Eindeutigkeit des Kovarianzoperators handelt es sich also um den selbstadjungierten, positiven Hilbert-Schmidt-Integraloperator

$$C : \begin{array}{ccc} L_2(\mathbb{R}^d, \mathcal{B}^d, \tau) & \rightarrow & L_2(\mathbb{R}^d, \mathcal{B}^d, \tau) \\ e & \mapsto & \int c(\cdot, z)e(z) d\tau(z), \end{array}$$

wobei für  $y, z \in \mathbb{R}^d$

$$\begin{aligned} c(y, z) &= \mathbb{E} [\mathbb{X}(y)\mathbb{X}(z)] \\ &= \mathbb{E} [(\cos + \sin)(X'_1 y)(\cos + \sin)(X'_1 z)] \\ &\quad - \mathbb{E} [(\cos + \sin)(X'_1 y)] \mathbb{E} [(\cos + \sin)(X'_1 z)]. \end{aligned}$$

Gilt  $\frac{m}{m+n} \rightarrow \rho \in (0, 1)$  für  $m, n \rightarrow \infty$ , so ist die Verteilungskonvergenz des Integranden in (3.16)

$$\sqrt{\frac{n}{m+n}} \mathbb{J}_{\{X_1, \dots, X_m\}; P^{X_1}} - \sqrt{\frac{m}{m+n}} \mathbb{J}_{\{Y_1, \dots, Y_n\}; P^{X_1}} \xrightarrow[m, n \rightarrow \infty]{\mathcal{D}} \mathbb{Z}$$

im Raum  $\mathcal{H}$  eine unmittelbare Konsequenz des vorangegangenen Satzes, da die Wurzelkonvexkombination zweier unabhängiger Gaußscher Zufallselemente mit gleicher Kovarianzstruktur ebenfalls ein Gaußsches Zufallselement mit eben dieser Kovarianzstruktur ist.

Die Karhunen-Loève-Entwicklung folgt nun sofort mit dem folgenden Korollar.

**Korollar 3.11.** *Es sei  $\phi \in \Phi_0$  und  $P^X \in \mathcal{F}_d^1(\phi)$ . Die Zufallsvariablen  $X_1, X_2, \dots, Y_1, Y_2, \dots$  seien unabhängig und identisch verteilt mit Verteilung  $P^X$ . Es gelte  $\frac{m}{m+n} \rightarrow \rho \in (0, 1)$  für  $m, n \rightarrow \infty$ . Dann gilt*

$$T_{m,n}^\phi \xrightarrow[m,n \rightarrow \infty]{\mathcal{D}} \sum_{\sigma} \tilde{\lambda}_{\sigma} Z_{\sigma}^2, \quad (3.17)$$

wobei  $(Z_{\sigma})_{\sigma}$  eine Folge unabhängiger standardnormalverteilter Zufallsvariablen ist und  $(\tilde{\lambda}_{\sigma})_{\sigma}$  die Folge der positiven Eigenwerte des Operators

$$C : \begin{aligned} L_2(\mathbb{R}^d, \mathcal{B}^d, \tau) &\rightarrow L_2(\mathbb{R}^d, \mathcal{B}^d, \tau) \\ e &\mapsto \int c(\cdot, z) e(z) d\tau(z), \end{aligned}$$

ist. Der Kern  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  des Integraloperators ist definiert durch

$$c(y, z) := \begin{aligned} &E [(\cos + \sin)(X_1' y)(\cos + \sin)(X_1' z)] \\ &- E [(\cos + \sin)(X_1' y)] E [(\cos + \sin)(X_1' z)]. \end{aligned}$$

**Beweis.** Mit dem Satz von der stetigen Abbildung folgt aus der Verteilungskonvergenz der Zufallselemente in  $\mathcal{H}$  auch die Verteilungskonvergenz der Quadrate der Normen der Zufallselemente. Es gilt nun das Normquadrat  $\langle \mathbb{Z}, \mathbb{Z} \rangle$  durch die Reihe (3.17) darzustellen. Dazu seien  $(e_{\sigma})_{\sigma}$ ,  $\sigma \in \mathbb{N}$ , ein orthonormales System aus Eigenvektoren zu den Eigenwerten  $(\tilde{\lambda}_{\sigma})_{\sigma}$  von  $C$ . Wir zeigen zuerst, dass  $\mathbb{Y} = \sum_{\sigma=1}^{\infty} \sqrt{\tilde{\lambda}_{\sigma}} Z_{\sigma} e_{\sigma}$  ein Zufallselement in  $\mathcal{H}$  ist, welches dieselbe Verteilung wie  $\mathbb{Z}$  hat. Dabei bezeichne  $(Z_{\sigma})_{\sigma}$  eine Folge unabhängiger standardnormalverteilter Zufallsvariablen.

Es sei  $(\Omega, \mathcal{A}, P)$  der zu Grunde liegende Wahrscheinlichkeitsraum. Die Folge der Partialsummen  $\mathbb{Y}_n = \sum_{\sigma=1}^n \sqrt{\tilde{\lambda}_{\sigma}} Z_{\sigma} e_{\sigma}$ , aufgefasst als Abbildungen von  $\Omega \times \mathbb{R}^d$  nach  $\mathbb{R}$  sind Elemente des  $L_2(\Omega \times \mathbb{R}^d, \mathcal{A} \otimes \mathcal{B}^d, P \otimes \tau)$  und bilden in diesem Raum eine Cauchyfolge, da  $C$  von der Spurklasse ist. Also ist  $\mathbb{Y}$ , aufgefasst als Abbildung von  $\Omega \times \mathbb{R}^d$  nach  $\mathbb{R}$  ein Element aus  $L_2(\Omega \times \mathbb{R}^d, \mathcal{A} \otimes \mathcal{B}^d, P \otimes \tau)$  und damit  $\mathcal{A} \otimes \mathcal{B}^d / \mathcal{B}$ -messbar. Wie im Beweis von Satz 3.10 folgt, dass  $\langle \mathbb{Y}, e \rangle$  für  $e \in L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$  eine reelle Zufallsvariable und damit  $\mathbb{Y}$  ein Hilbertraum-wertiges Zufallselement ist.

In der  $L_2(\Omega, \mathcal{A}, P)$ -Norm konvergiert für jedes  $e \in L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$  die Zufallsvariable  $\langle \mathbb{Y}_n, e \rangle$  gegen  $\langle \mathbb{Y}, e \rangle$ , denn mit der Cauchy-Schwarzschen Ungleichung folgt

$$\begin{aligned} E [\langle \mathbb{Y}, e \rangle - \langle \mathbb{Y}_n, e \rangle]^2 &= E \left[ \int (\mathbb{Y}(z) - \mathbb{Y}_n(z)) e(z) d\tau(z) \right]^2 \\ &\leq \|e\|^2 E \int (\mathbb{Y}(z) - \mathbb{Y}_n(z))^2 d\tau(z) \xrightarrow[n \rightarrow \infty]{} 0, \end{aligned}$$

wegen der  $L_2$ -Konvergenz im Produktraum. Es ist für jedes  $e \in L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$  die Zufallsvariable  $\langle \mathbb{Y}_n, e \rangle$  normalverteilt. Zum Beispiel mittels Betrachtung der zugehörigen charakteristischen Funktionen macht man sich klar, dass  $\langle \mathbb{Y}, e \rangle$  ebenfalls normalverteilt ist sowie den Erwartungswert 0 und die Varianz  $\sum_{\sigma=1}^{\infty} \tilde{\lambda}_{\sigma} (\langle e_{\sigma}, e \rangle)^2$  besitzt. Das Zufallselement  $\mathbb{Y}$  ist also ein Gaußsches Zufallselement des  $\mathcal{H}$  dessen Erwartungswert 0 ist und dessen Kovarianzoperator eindeutig durch die Beziehung

$$\mathbb{E} [\langle \mathbb{Y}, e \rangle^2] = \sum_{\sigma=1}^{\infty} \tilde{\lambda}_{\sigma} (\langle e_{\sigma}, e \rangle)^2$$

festgelegt wird. Für das ebenfalls zentrierte Gaußsche Zufallselement  $\mathbb{Z}$  gilt

$$\begin{aligned} \mathbb{E} [\langle \mathbb{Z}, e \rangle^2] &= \langle Ce, e \rangle = \left\langle C \sum_{\sigma=1}^{\infty} \langle e, e_{\sigma} \rangle e_{\sigma}, \sum_{k=1}^{\infty} \langle e, e_k \rangle e_k \right\rangle \\ &= \sum_{\sigma=1}^{\infty} \sum_{k=1}^{\infty} \langle e, e_{\sigma} \rangle \langle e, e_k \rangle \langle Ce_{\sigma}, e_k \rangle \\ &= \sum_{\sigma=1}^{\infty} (\langle e, e_{\sigma} \rangle)^2 \tilde{\lambda}_{\sigma}. \end{aligned}$$

Also stimmen die Kovarianzoperatoren von  $\mathbb{Y}$  und  $\mathbb{Z}$  und damit auch die Verteilungen in  $\mathcal{H}$  überein.  $\|\mathbb{Y}\|^2$  und  $\|\mathbb{Z}\|^2$  besitzen daher dieselbe Verteilung.

Das Normquadrat von  $\mathbb{Y}_n$  konvergiert im  $L_1$ -Sinn und damit in Verteilung gegen das Normquadrat von  $\mathbb{Y}$ , da mittels zweifacher Anwendung der Cauchy-Schwarzschen Ungleichung

$$\begin{aligned} \mathbb{E} |\langle \mathbb{Y}, \mathbb{Y} \rangle - \langle \mathbb{Y}_n, \mathbb{Y}_n \rangle| &= \mathbb{E} |\langle \mathbb{Y} - \mathbb{Y}_n, \mathbb{Y} + \mathbb{Y}_n \rangle| \\ &\leq \mathbb{E} \sqrt{\|\mathbb{Y} - \mathbb{Y}_n\|^2} \sqrt{\|\mathbb{Y} + \mathbb{Y}_n\|^2} \\ &\leq \sqrt{\mathbb{E} \|\mathbb{Y} - \mathbb{Y}_n\|^2} \sqrt{\mathbb{E} \|\mathbb{Y} + \mathbb{Y}_n\|^2}. \end{aligned}$$

Der erste Faktor der letzten Zeile strebt gegen 0, denn fasst man wie zuvor  $\mathbb{Y}_n$  und  $\mathbb{Y}$  als Abbildungen des Produktraums  $\Omega \times \mathbb{R}^d$  in die reellen Zahlen auf, so konvergieren diese gegeneinander im  $L_2(\Omega \times \mathbb{R}^d, \mathcal{A} \otimes \mathcal{B}^d, P \otimes \tau)$ -Sinn. Der zweite Faktor ist beschränkt, da

$$\begin{aligned} \mathbb{E} \|\mathbb{Y} + \mathbb{Y}_n\|^2 &= \mathbb{E} \|\mathbb{Y} - \mathbb{Y}_n\|^2 + 4\mathbb{E} \langle \mathbb{Y} - \mathbb{Y}_n, \mathbb{Y}_n \rangle + 4\mathbb{E} \|\mathbb{Y}_n\|^2 \\ &\leq \mathbb{E} \|\mathbb{Y} - \mathbb{Y}_n\|^2 + 4\sqrt{\mathbb{E} \|\mathbb{Y} - \mathbb{Y}_n\|^2} \sqrt{\mathbb{E} \|\mathbb{Y}_n\|^2} + 4\mathbb{E} \|\mathbb{Y}_n\|^2, \end{aligned}$$

wobei  $\mathbb{E} \|\mathbb{Y}_n\|^2 = \sum_{\sigma=1}^n \tilde{\lambda}_{\sigma} \leq \sum_{\sigma=1}^{\infty} \tilde{\lambda}_{\sigma} < \infty$ .

Mit der Darstellung  $\|\mathbb{Y}_n\|^2 = \sum_{\sigma=1}^n \tilde{\lambda}_{\sigma} Z_{\sigma}^2$  folgt nun die Behauptung.  $\square$

### 3.4 Vergleich der beiden Ansätze

Es ist zunächst merkwürdig, dass die Gewichte der Grenzverteilung sowohl über den Operator  $\bar{H}$  als auch über den Operator  $H$  bestimmt werden können (siehe

Abschnitt 3.1). Eine Erklärung bietet sich mittels des Ansatzes über empirische charakteristische Funktionen.

Nach (3.16) lässt sich die Teststatistik schreiben als

$$T_{m,n}^\phi = \int \left[ \sqrt{\frac{n}{m+n}} \mathbb{J}_{\{X_1, \dots, X_m\}; P^X}(z) - \sqrt{\frac{m}{m+n}} \mathbb{J}_{\{Y_1, \dots, Y_n\}; P^Y}(z) \right]^2 d\tau(z),$$

wobei  $\tau$  in Korollar 2.5 und  $\mathbb{J}_{\{X_1, \dots, X_m\}; P^X}$  zu Beginn von Abschnitt 3.3 definiert wurde. Im Anschluss an Satz 3.10 wurde bereits bemerkt, dass die sich im Grenzwert ergebende Wurzelkonvexkombination Gaußscher Prozesse auch wieder ein Gaußscher Prozess mit derselben Kovarianzstruktur ist. Die Folge von V-Statistiken

$$\int [\mathbb{J}_{\{X_1, \dots, X_m\}; P^X}(z)]^2 d\tau(z)$$

hat mit der Argumentation von Korollar 3.11 also denselben Grenzwert für  $m \rightarrow \infty$  wie unsere Teststatistik

$$\|\mathbb{Z}\|^2 = \int \mathbb{Z}^2(z) d\tau(z).$$

Diese neue Statistik lässt sich auch schreiben als

$$\begin{aligned} & \int [\mathbb{J}_{\{X_1, \dots, X_m\}; P^X}(z)]^2 d\tau(z) \\ &= m \int \left[ \frac{1}{m} \sum_{k=1}^m (\cos + \sin)(X'_k z) - \mathbb{E}(\cos + \sin)(X' z) \right]^2 d\tau(z) \\ &= m \int [\varphi_{\{X_1, \dots, X_m\}}(z) - \varphi_X(z)]^2 d\tau(z) \\ &= m d_\phi^2(P_m^X, P^X), \end{aligned}$$

wobei  $P_m^X$  das empirische Maß von unabhängigen  $X_1, \dots, X_m \sim P^X$  sei. Das Abstandsquadrat  $d_\phi^2(P_m^X, P^X)$  ist eine Ein-Stichproben-V-Statistik der Kernlänge 2 mit Kern

$$h(x_1, x_2) = \mathbb{E}\phi(|x_1 - X_1|^2) + \mathbb{E}\phi(|x_2 - X_1|^2) - \phi(|x_1 - x_2|^2) - \mathbb{E}\phi(|X_1 - X_2|^2),$$

$x_1, x_2 \in \mathbb{R}^d$ , mit  $X_1, X_2 \sim P^X$  unabhängig. Die Bezeichnung  $h$  für den Kern wurde dabei absichtlich vergeben, da die Kernfunktion mit der aus Satz 3.7 übereinstimmt.

Im Blickwinkel der empirischen charakteristischen Funktionen lässt sich also der Übergang zur Kernfunktion  $h$  ohne Veränderung der Grenzverteilung dadurch erklären, dass die Wurzelkonvexkombination zweier Gaußscher Zufallselemente mit demselben Kovarianzoperator wieder ein solches Gaußsches Zufallselement ist.

Wir werden nun zeigen, dass der Zusammenhang zwischen den Operatoren  $C$  und  $H$  stärker ist, als es bislang deutlich war: Die Operatoren  $C$  und  $H$  besitzen dieselben positiven Eigenwerte.

In Kapitel 3.1.2 wurde gezeigt, dass die Eigenwerte von  $\bar{H}$  bis auf den Faktor  $\frac{1}{2}$  mit denen von  $H$  übereinstimmen. Der dort verwandten Vorgehensweise folgend zeigen wir zunächst, dass die nicht verschwindenden Eigenwerte von  $C$  auch Eigenwerte von  $H$  sind.

**Lemma 3.12.** *Die Funktionen  $(f_j)_j$  mit  $f_j : \mathbb{R}^d \rightarrow \mathbb{R}$ , definiert durch*

$$f_j(x) := \frac{1}{\sqrt{\tilde{\lambda}_j}} \int [\cos(s'x) + \sin(s'x) - \varphi_X(s)] e_j(s) d\tau(s), \quad x \in \mathbb{R}^d,$$

wobei  $(e_j)_j$  die orthonormalen Eigenfunktionen zu den positiven Eigenwerten  $(\tilde{\lambda}_j)_j$  des Operators  $C$  bezeichnen, sind orthonormale Eigenfunktionen des Operators  $H$  zu den Eigenwerten  $(\tilde{\lambda}_j)_j$ .

**Beweis.** Nutzt man aus, dass  $e_j, e_k$  Eigenfunktionen von  $C$  sind, so folgt (die Funktionen  $f_j, f_k$  sind Elemente des Raums  $L_2(\mathbb{R}^d, \mathcal{B}^d, P^X)$ , die Funktionen  $e_j, e_k$  sind Elemente des Raums  $L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$ : dementsprechend ist jeweils das Innenprodukt zu verstehen)

$$\begin{aligned} \langle f_j, f_k \rangle &= \frac{1}{\sqrt{\tilde{\lambda}_j \tilde{\lambda}_k}} \iiint (\cos(s'x) + \sin(s'x) - \varphi_X(s)) e_j(s) \\ &\quad (\cos(t'x) + \sin(t'x) - \varphi_X(t)) e_k(t) d\tau(s) d\tau(t) dP^X(x) \\ &= \frac{1}{\sqrt{\tilde{\lambda}_j \tilde{\lambda}_k}} \iint c(s, t) e_j(s) e_k(t) d\tau(s) d\tau(t) \\ &= \sqrt{\frac{\tilde{\lambda}_j}{\tilde{\lambda}_k}} \int e_j(t) e_k(t) d\tau(t) = \langle e_j, e_k \rangle. \end{aligned}$$

Die Orthonormalität der  $(e_j)$  überträgt sich somit auf die im Lemma definierten  $(f_j)$ . Weiterhin sind die Funktionen  $(f_j)$  Eigenfunktionen zu den Eigenwerten  $(\tilde{\lambda}_j)$ , denn

$$\begin{aligned} \tilde{\lambda}_j f_j(x) &= \tilde{\lambda}_j \frac{1}{\sqrt{\tilde{\lambda}_j}} \int (\cos(s'x) + \sin(s'x) - \varphi_X(s)) e_j(s) d\tau(s) \\ &= \frac{1}{\sqrt{\tilde{\lambda}_j}} \int (\cos(s'x) + \sin(s'x) - \varphi_X(s)) \int c(s, t) e_j(t) d\tau(t) d\tau(s) \\ &= \frac{1}{\sqrt{\tilde{\lambda}_j}} \iiint (\cos(s'x) + \sin(s'x) - \varphi_X(s)) (\cos(s'y) + \sin(s'y) - \varphi_X(s)) \\ &\quad (\cos(t'y) + \sin(t'y) - \varphi_X(t)) e_j(t) dP^X(y) d\tau(t) d\tau(s) \end{aligned}$$

$$\begin{aligned}
&= \iint (\cos(s'x) + \sin(s'x) - \varphi_X(s)) \\
&\quad (\cos(s'y) + \sin(s'y) - \varphi_X(s)) \, d\tau(s) f_j(y) \, dP^X(y) \\
&= \iint \mathbb{E} [ 1 - \cos(s'(x - X_2)) + 1 - \cos(s'(X_1 - y)) \\
&\quad - 1 + \cos(s'(x - y)) - 1 + \cos(s'(X_1 - X_2))] \, d\tau(s) f_j(y) \, dP^X(y) \\
&= \int \mathbb{E} [ \phi(|x - X_2|^2) + \phi(|X_1 - y|^2) \\
&\quad - \phi(|x - y|^2) - \phi(|X_1 - X_2|^2)] f_j(y) \, dP^X(y) \\
&= \int h(x, y) f_j(y) \, dP^X(y) = (H f_j)(x).
\end{aligned}$$

□

Jeder nicht verschwindende Eigenwert von  $C$  ist positiv, also mit dem Lemma auch ein Eigenwert von  $H$ . Wir haben bereits im Beweis von Satz 3.5 festgestellt, dass  $H$  positiv ist. Dabei nutzten wir indirekt aus, dass der Operator  $C$  als Kovarianzoperator positiv ist. Beide Operatoren sind von der Spurklasse und die Spur der Operatoren ist in beiden Fällen dieselbe:

$$\int h(x, x) \, d\gamma(x) = \mathbb{E} \phi(|X - Y|^2) = \mathbb{E} \int [1 - \cos((X - Y)'s)] \, d\tau(s) = \int c(s, s) \, d\tau(s).$$

Daher besitzen die Operatoren  $C$  und  $H$  dieselben von Null verschiedenen Eigenwerte in jeweils derselben Vielfachheit. Diese interessante Eigenschaft, die man bereits auf Grund der Verteilungsgleichheit der mit den Eigenwerten von  $C$ , respektive  $H$  gewichtete Summen von  $\chi_1^2$ -verteilten Zufallsvariablen vermutet, wird es uns in der zweiten Hälfte von Kapitel 4 ermöglichen, ein deterministisches, nicht auf Monte Carlo-Simulationen beruhendes Verfahren zur Bestimmung des kritischen Wertes des Tests zu entwickeln.

# Kapitel 4

## Approximation der Grenzverteilung

Die vorgestellten Testverfahren sind weder verteilungsfrei noch asymptotisch verteilungsfrei, das heißt, die Verteilung von  $T_{m,n}^\phi$  und die asymptotische Verteilung für  $m, n \rightarrow \infty$  hängen unter der Hypothese von der unbekanntem Verteilung ab, die beiden Stichproben zu Grunde liegt. Daher lassen sich die kritischen Werte für eine feste Funktion  $\phi \in \Phi$  nicht tabellieren. Es werden Verfahren benötigt, die bei gegebenen Beobachtungen einen Schätzer für den kritischen Wert liefern. Ein sehr bekanntes Verfahren ist das Monte Carlo-Bootstrap-Verfahren, bei dem die unbekanntem Verteilung durch die empirische Verteilung der zusammengesetzten Beobachtung geschätzt wird („Bootstrap“) und anschließend aus dieser Verteilung wiederholt zufällige Bootstrap-Beobachtungen gezogen werden, welche dann in die Teststatistik eingesetzt werden. Der kritische Wert wird dann mittels der empirischen Verteilung der Realisierungen der Testgröße bestimmt (dies ist der „Monte Carlo“-Vorgang). Eine weitere Möglichkeit, welche auch in Kapitel 5 noch weiter untersucht werden wird, ist, die Eigenwerte des Operators  $C$  beziehungsweise  $H$  in der Grenzverteilung (3.17)

$$\sum_{\sigma} \lambda_{\sigma} Z_{\sigma}$$

beziehungsweise (3.8) zu approximieren, indem man die unbekanntem Verteilung in der Definition des Operators  $C$  beziehungsweise  $H$  durch die empirische Verteilung der zusammengesetzten Beobachtung schätzt. Dieses Kapitel zeigt, dass für die neue Familie von Tests beide Schätzverfahren anwendbar sind.

## 4.1 Das Monte Carlo-Bootstrap-Verfahren

Die Anwendung des von Efron [ET93] popularisierten Bootstrap-Verfahrens auf das Zwei-Stichproben-Problem führt auf die folgende Situation: Es seien  $x_1, x_2, \dots, y_1, y_2, \dots$  Beobachtungen der unabhängigen  $\mathbb{R}^d$ -wertigen Zufallsvariablen  $X_1, X_2, \dots, Y_1, Y_2, \dots$ . Wir nehmen nun  $m$  Beobachtungen  $x_1, \dots, x_m$  der ersten und  $n$  Beobachtungen  $y_1, \dots, y_n$  der zweiten Stichprobe und bezeichnen die zusammengelegte Stichprobe  $x_1, \dots, x_m, y_1, \dots, y_n$  mit  $z_1, \dots, z_{m+n}$ . Unter der Hypothese  $X \stackrel{\mathcal{D}}{=} Y$  stammen diese  $m+n$  Beobachtungen von derselben Verteilung. Es liegt also nahe, die Verteilung der Testgröße unter der Hypothese durch die Verteilung der Testgröße unter der Annahme, dass die empirische Verteilung der Beobachtungen  $z_1, \dots, z_{m+n}$  die gemeinsame unbekannte Verteilung ist, zu approximieren: das so genannte **Bootstrap-Verfahren**. Praktisch geschieht dies, indem man so genannte **Bootstrap-Stichproben**  $\hat{X}_1, \dots, \hat{X}_m, \hat{Y}_1, \dots, \hat{Y}_n$  vom Umfang  $m+n$  aus der zusammengelegten Stichprobe zieht und für jede Bootstrap-Stichprobe die Testgröße

$$\hat{T}_{m,n}^\phi := T_{m,n}^\phi(\hat{X}_1, \dots, \hat{X}_m; \hat{Y}_1, \dots, \hat{Y}_n)$$

bestimmt. Eine Realisierung von  $\hat{T}_{m,n}^\phi$  nennt man auch ein **Bootstrap-Replikat** von  $T_{m,n}^\phi$ . Zieht man nun mehr und mehr Stichproben, insgesamt  $B$  Stück, so kann man nach dem Satz von Glivenko-Cantelli die Verteilung von  $\hat{T}_{m,n}^\phi$ , d.h. die Verteilung der Teststatistik unter der Annahme, dass die zusammengelegte Stichprobe der Verteilung unter der Hypothese entspricht, fast sicher durch die Verteilung von  $\hat{T}_{m,n}^\phi(B)$ , der empirischen Verteilung der erstellten  $B$  Bootstrap-Replikate approximieren. Zur Durchführung des Tests wählt man als approximativen kritischen Wert ein  $1 - \alpha$ -Quantil der empirischen Verteilung  $\hat{T}_{m,n}^\phi(B)$ , wobei  $\alpha$  das Niveau, also die Wahrscheinlichkeit für einen Fehler erster Art, angebe. Diesen kritischen Wert vergleicht man mit der Teststatistik der zu Grunde liegenden Beobachtungen  $x_1, \dots, x_m, y_1, \dots, y_n$ . Ist diese Teststatistik größer als der gewonnene kritische Wert, so verwirft man die Hypothese.

Es sei im Folgenden  $((m_i, n_i))_i$ ,  $i \in \mathbb{N}$  eine Folge von Beobachtungsumfängen der ersten und zweiten Stichprobe mit  $m_i, n_i \rightarrow \infty$  für  $i \rightarrow \infty$  und  $\frac{m_i}{m_i + n_i} \rightarrow \rho \in (0, 1)$  für  $i \rightarrow \infty$ .

In diesem Abschnitt soll nun gezeigt werden, dass für fast jedes Paar von Beobachtungsfolgen die Verteilung von  $\hat{T}_{m_i, n_i}^\phi$  für  $i \rightarrow \infty$  gegen die Grenzverteilung (3.17) aus Korollar 3.11 konvergiert. Da wir das Monte Carlo-Bootstrap-Verfahren auch beim Nachweis der Konsistenz der neuen Tests zulassen wollen, gehen wir davon aus, dass die der ersten Stichprobe zu Grunde liegende Verteilung  $P^X$  nicht zwangsläufig gleich der Verteilung  $P^Y$  ist, die der zweiten Stichprobe zu Grunde liegt.

Betrachtet man das Problem im Hilbertraum  $\mathcal{H} = L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$ , so interessiert

man sich für das Verhalten des Zufallselements

$$\mathbb{J}_{\{Z_1^{(i)}, \dots, Z_m^{(i)}\}; P^{Z^{(i)}}}(z) = \sqrt{m} \left( \frac{1}{m} \sum_{k=1}^m (\cos + \sin)(Z_k^{(i)' } z) - \mathbb{E}(\cos + \sin)(Z^{(i)' } z) \right), \quad (4.1)$$

$z \in \mathbb{R}^d$ , wobei  $Z_1^{(i)}, \dots, Z_m^{(i)}$  unabhängige und identisch verteilte Zufallsvariablen seien mit Verteilung  $P^{Z^{(i)}}$ , der empirischen Verteilung der zusammengesetzten Beobachtung aus  $m_i$  Beobachtungen der ersten Stichprobe und  $n_i$  Beobachtungen der zweiten Stichprobe. Sollte dieses Zufallselement in  $\mathcal{H}$  in Verteilung gegen ein Gaußsches Zufallselement  $Z$  konvergieren, so konvergiert auch

$$\sqrt{\frac{n_i}{m_i + n_i}} \mathbb{J}_{\{Z_1^{(i)}, \dots, Z_m^{(i)}\}; P^{Z^{(i)}}}(z) - \sqrt{\frac{m_i}{m_i + n_i}} \mathbb{J}_{\{Z_{m+1}^{(i)}, \dots, Z_{m+n}^{(i)}\}; P^{Z^{(i)}}}(z), \quad z \in \mathbb{R}^d,$$

mit der Wurzelkonvex-Argumentation in Verteilung gegen dasselbe Gaußsche Zufallselement. Mit dem Satz von der stetigen Abbildung folgt dann die Konvergenz der Normen

$$\hat{T}_{m_i, n_i}^\phi = \left\| \sqrt{\frac{n_i}{m_i + n_i}} \mathbb{J}_{\{Z_1^{(i)}, \dots, Z_m^{(i)}\}; P^{Z^{(i)}}} - \sqrt{\frac{m_i}{m_i + n_i}} \mathbb{J}_{\{Z_{m+1}^{(i)}, \dots, Z_{m+n}^{(i)}\}; P^{Z^{(i)}}} \right\|^2,$$

so dass für die Verteilung von  $\hat{T}_{m_i, n_i}^\phi$  der Kovarianzoperator des Gaußschen-Grenzprozesses entscheidend ist. Wir zeigen nun zunächst, dass der Kovarianzoperator  $C_i$  von  $\mathbb{J}_{\{Z_1^{(i)}, \dots, Z_m^{(i)}\}; P^{Z^{(i)}}}$ , der dem Kovarianzoperator des Zufallselements

$$\mathbb{X}_i(z) := (\cos + \sin)(Z_1^{(i)' } z) - \mathbb{E}(\cos + \sin)(Z^{(i)' } z), \quad z \in \mathbb{R}^d,$$

entspricht, in der Operatornorm gegen den Kovarianzoperator  $C$  von

$$\mathbb{X}(z) := (\cos + \sin)(Z_1' z) - \mathbb{E}(\cos + \sin)(Z' z), \quad z \in \mathbb{R}^d,$$

konvergiert, wobei  $Z, Z_1 \sim \rho P^X + (1 - \rho) P^Y$ .

**Lemma 4.1.** *Seien  $\phi \in \Phi_0$  und  $P^X, P^Y \in \mathcal{F}_d^A(\phi)$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^d, \mathcal{B}^d)$ . Es sei  $((m_i, n_i))_i$  eine Folge von Stichprobenumfängen mit  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$ . Dann gilt für fast jedes Paar von Folgen  $(x_n)_n, (y_n)_n$  von Beobachtungen unabhängiger Zufallsvariablen auf  $\mathbb{R}^d$  mit Verteilung  $P^X$  beziehungsweise  $P^Y$*

$$C_i \rightarrow C$$

in der Operatornorm auf  $\mathcal{H}$ .  $C_i$  sei definiert durch

$$C_i : \begin{array}{l} \mathcal{H} \rightarrow \mathcal{H} \ , \\ e \mapsto \int c_i(\cdot, z) e(z) d\tau(z), \end{array}$$

mit

$$c_i(y, z) := \mathbb{E} \left[ (\cos + \sin)(Z^{(i)'}y)(\cos + \sin)(Z^{(i)'}z) \right] \\ - \mathbb{E} \left[ (\cos + \sin)(Z^{(i)'}y) \right] \mathbb{E} \left[ (\cos + \sin)(Z^{(i)'}z) \right],$$

$y, z \in \mathbb{R}^d$ , wobei  $Z^{(i)}$  die empirische Verteilung von  $m_i$  Beobachtungen der ersten Stichprobe zusammengelegt mit  $n_i$  Beobachtungen der zweiten Stichprobe habe. Der Operator  $C$  sei wie in Korollar 3.11 definiert mit zu Grunde liegender Verteilung  $P^Z := \rho P^X + (1 - \rho)P^Y$ .

**Beweis.** Wir schätzen die Operatornorm der Differenz von  $C_i$  und  $C$  ab.

$$\begin{aligned} |C - C_i|^2 &= \sup_{\|e\| \leq 1, e \in \mathcal{H}} \|(C - C_i)e\|^2 \\ &= \sup_{\|e\| \leq 1, e \in \mathcal{H}} \int \left[ \int [c(s, t) - c_i(s, t)]e(t) d\tau(t) \right]^2 d\tau(s) \\ &\leq \iint [c(s, t) - c_i(s, t)]^2 d\tau(t) d\tau(s) \\ &\leq \int_M \int_M [c(s, t) - c_i(s, t)]^2 d\tau(t) d\tau(s) \end{aligned} \quad (4.2)$$

$$+ 4 \int_{M^c} \int c^2(s, t) d\tau(t) d\tau(s) \quad (4.3)$$

$$+ 4 \int_{M^c} \int c_i^2(s, t) d\tau(t) d\tau(s) \quad (4.4)$$

Dabei sei  $M \subset \mathbb{R}^d \setminus \{0\}$  eine kompakte Menge, die später genauer spezifiziert wird, und  $M^c$  das Komplement von  $M$  in  $\mathbb{R}^d$ .

Dem Satz von Glivenko-Cantelli [GS77, Satz 3.2.2, S.145] folgend konvergiert  $P^{Z^{(i)}}$  für fast jede Folge von Beobachtungen schwach gegen  $P^Z$ . Der Stetigkeitssatz liefert die gleichmäßige Konvergenz der Fourier-Transformierten von  $P^{Z^{(i)}}$  gegen die von  $P^Z$  auf jedem Kompaktum, vergleiche hierzu Bauer [Bau91, Satz 23.8, S.202], also wegen

$$c(s, t) = \operatorname{Re}(\varphi_Z(s-t)) + \operatorname{Im}(\varphi_Z(s+t)) - [(\operatorname{Re} + \operatorname{Im})\varphi_Z(s)][(\operatorname{Re} + \operatorname{Im})\varphi_Z(t)],$$

$s, t \in \mathbb{R}^d$ , die gleichmäßige Konvergenz von  $c_i$  gegen  $c$  auf jedem Kompaktum. Daher konvergiert für fast jede Beobachtungsfolge

$$\int_M \int_M c_i^2(s, t) d\tau(t) d\tau(s) \xrightarrow{i \rightarrow \infty} \int_M \int_M c^2(s, t) d\tau(t) d\tau(s)$$

für jede kompakte Menge  $M \subset \mathbb{R}^d \setminus \{0\}$ .

Wir behandeln nun die in den Ausdrücken (4.3) und (4.4) auftauchende Integration über das Quadrat von  $c$  und  $c_i$ . Die Operatoren  $C$ ,  $H$  und  $\bar{H}$  sind Hilbert-Schmidt-Operatoren. Mit den Beziehungen zwischen den Eigenwerten folgt eine Beziehung zwischen den Hilbert-Schmidt-Normen der Operatoren

$$|C|_{\text{HS}} = |H|_{\text{HS}} = \frac{1}{2}|\bar{H}|_{\text{HS}},$$

wobei die drei Normen auf unterschiedlichen Räumen definiert sind. Es folgt mit Lemma C.6

$$\iint c^2(s, t) d\tau(t) d\tau(s) = \frac{1}{4}\mathbb{E}[\bar{h}^2(X_1, Y_1; X_2, Y_2)],$$

wobei  $X_1, X_2, Y_1, Y_2 \sim P^Z$  unabhängig seien. Dieser Erwartungswert ist endlich, weil nach Voraussetzung  $\mathbb{E}\phi^2(|X_1 - Y_1|^2) < \infty$ . Analog gilt:

$$\iint c_i^2(s, t) d\tau(t) d\tau(s) = \frac{1}{4}\mathbb{E}[\bar{h}^2(X_1^{(i)}, Y_1^{(i)}; X_2^{(i)}, Y_2^{(i)})],$$

$X_1^{(i)}, X_2^{(i)}, Y_1^{(i)}, Y_2^{(i)} \sim P_i^Z$ . Dieser Ausdruck konvergiert für  $i \rightarrow \infty$  nach dem starken Gesetz der großen Zahlen für V-Statistiken unter der Voraussetzung

$$\mathbb{E}\phi^4(|X_1 - Y_1|^2) < \infty$$

für fast jede Beobachtungsfolge gegen  $\iint c^2(s, t) d\tau(t) d\tau(s)$ . Um dies zu zeigen, zerlegt man die Integrationen bezüglich  $P^{Z^{(i)}}$  in ihren Bestandteil aus der ersten Stichprobe und ihren Bestandteil aus der zweiten Stichprobe und wendet dann die Ein- und Zwei-Stichproben-Varianten des starken Gesetzes der großen Zahlen für V-Statistiken, Satz B.4, Anhang B, an.

Sei nun  $\varepsilon > 0$  beliebig. Es sei eine kompakte Teilmenge  $M \subset \mathbb{R}^d \setminus \{0\}$  so groß gewählt, dass

$$\iint c^2(s, t) d\tau(t) d\tau(s) - \int_M \int_M c^2(s, t) d\tau(t) d\tau(s) < \varepsilon.$$

Anschließend wird  $i_0 \in \mathbb{N}$  so groß gewählt, dass für alle  $i \geq i_0$  sowohl

$$\left| \iint c^2(s, t) d\tau(t) d\tau(s) - \iint c_i^2(s, t) d\tau(t) d\tau(s) \right| < \varepsilon$$

als auch

$$\int_M \int_M [c(s, t) - c_i(s, t)]^2 d\tau(t) d\tau(s) < \varepsilon$$

und

$$\left| \int_M \int_M c^2(s, t) d\tau(t) d\tau(s) - \int_M \int_M c_i^2(s, t) d\tau(t) d\tau(s) \right| < \varepsilon$$

gilt. Dann gilt mit den so gewählten  $M$  und  $i \geq i_0$  in (4.2)

$$\int_M \int_M [c(s, t) - c_i(s, t)]^2 d\tau(t) d\tau(s) < \varepsilon,$$

in (4.3)

$$\begin{aligned} & \int_{M^c} \int c^2(s, t) d\tau(t) d\tau(s) \\ & \leq \iint c^2(s, t) d\tau(t) d\tau(s) - \int_M \int_M c^2(s, t) d\tau(t) d\tau(s) < \varepsilon \end{aligned}$$

und in (4.4)

$$\begin{aligned} & \int_{M^c} \int c_i^2(s, t) d\tau(t) d\tau(s) \\ & \leq \iint c_i^2(s, t) d\tau(t) d\tau(s) - \int_M \int_M c_i^2(s, t) d\tau(t) d\tau(s) \\ & \leq \left| \iint c_i^2(s, t) d\tau(t) d\tau(s) - \iint c^2(s, t) d\tau(t) d\tau(s) \right| \\ & \quad + \left| \iint c^2(s, t) d\tau(t) d\tau(s) - \int_M \int_M c^2(s, t) d\tau(t) d\tau(s) \right| \\ & \quad + \left| \int_M \int_M c^2(s, t) d\tau(t) d\tau(s) - \int_M \int_M c_i^2(s, t) d\tau(t) d\tau(s) \right| < 3\varepsilon. \end{aligned}$$

Also gilt  $|C - C_i|^2 < 9\varepsilon$  für jedes  $\varepsilon > 0$  und  $i \geq i_0(\varepsilon)$  für hinreichend großes  $i_0(\varepsilon)$ . Es folgt die Behauptung.  $\square$

Die Kovarianzoperatoren der Zufallselemente (4.1) konvergieren also für  $i \rightarrow \infty$  gegen einen Kovarianzoperator  $C$ . Dies bringt uns bei der Überprüfung der Voraussetzungen des zentralen Grenzwertsatzes für Dreiecksschemata im Hilbertraum, Satz D.6, Anhang D, der die Basis des Nachweises der Verteilungskonvergenz der Verteilungen der Zufallselemente (4.1) in  $\mathcal{H}$  ist, weiter.

**Satz 4.2.** Sei  $\phi \in \Phi_0$  und  $P^X, P^Y \in \mathcal{F}_d^4(\phi)$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^d, \mathcal{B}^d)$ . Es sei  $((m_i, n_i))_i$  eine Folge von Stichprobenumfängen mit  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$ . Dann gilt für fast jedes Paar von Folgen  $(x_n)_n, (y_n)_n$  von Beobachtungen unabhängiger Zufallsvariablen auf  $\mathbb{R}^d$  mit Verteilung  $P^X$  beziehungsweise  $P^Y$

$$\mathbb{J}_{\{Z_1^{(i)}, \dots, Z_{m_i}^{(i)}\}; P^{Z^{(i)}}} \xrightarrow[i \rightarrow \infty]{\mathcal{D}} \mathbb{Z}$$

im Hilbertraum  $\mathcal{H}$ , wobei die Verteilungen  $P^{Z^{(i)}}$  im vorangegangenen Lemma 4.1 definiert und  $Z_1^{(i)}, \dots, Z_{m_i}^{(i)}$  unabhängige Zufallsvariablen mit dieser Verteilung sind.  $\mathbb{Z}$  bezeichnet ein zentriertes Gaußsches Zufallselement in  $\mathcal{H}$  mit dem in Lemma 4.1 beschriebenen Kovarianzoperator  $C$ .

**Beweis.** Im Folgenden sei das Beobachtungsfolgenpaar so gewählt, dass  $C_i$  gemäß Lemma 4.1 gegen  $C$  konvergiert. Wir weisen die Bedingungen von Satz D.6 nach. Wir verwenden als Basis des Hilbertraums  $\mathcal{H}$  eine Orthonormalbasis aus Eigenvektoren  $(e_j)_j$  des Operators  $C$ . Bedingung (i) ist erfüllt, da

$$\langle C_i e_k, e_l \rangle = \langle (C_i - C)e_k, e_l \rangle + \langle C e_k, e_l \rangle$$

zerfällt in die Summe aus

$$\langle (C_i - C)e_k, e_l \rangle^2 \leq \|(C_i - C)e_k\|^2 \xrightarrow{i \rightarrow \infty} 0$$

und

$$\langle C e_k, e_l \rangle = \tilde{\lambda}_k \langle e_k, e_l \rangle.$$

Folglich gilt

$$\langle C_i e_k, e_l \rangle \xrightarrow{i \rightarrow \infty} \begin{cases} \tilde{\lambda}_k & , k = l, \\ 0 & , k \neq l. \end{cases}$$

Bedingung (ii) ist eine Aussage über die Spuren der Operatoren  $C_i, C$ , denn es gilt wegen der Selbstadjungiertheit von  $C_i$

$$\begin{aligned} \sum_k \langle C_i e_k, e_k \rangle &= \sum_k \sum_j \langle C_i e_k, e_j^{(i)} \rangle \langle e_j^{(i)}, e_k \rangle \\ &= \sum_k \sum_j \langle e_k, C_i e_j^{(i)} \rangle \langle e_j^{(i)}, e_k \rangle \\ &= \sum_j \langle C_i e_j^{(i)}, e_j^{(i)} \rangle, \end{aligned}$$

wobei  $(e_j^{(i)})_j$  eine Orthonormalbasis aus Eigenvektoren von  $C_i$  bezeichne. Die Spur von  $C_i$  ist gegeben durch

$$\int c_i(s, s) d\tau(s) = \mathbb{E} \int [1 - \cos((Z_1^{(i)} - Z_2^{(i)})'s)] d\tau(s) = \mathbb{E} \phi(|Z_1^{(i)} - Z_2^{(i)}|^2),$$

$Z_1^{(i)}, Z_2^{(i)} \sim P^{Z^{(i)}}$  unabhängig, vergleiche Lemma C.10, Anhang C, sowie die Bemerkung nach Lemma 3.12. Zerlegt man die Integration bezüglich des Maßes  $P^{Z^{(i)}}$  in den Anteil, der von  $P^X$  stammt, und den Anteil, der von  $P^Y$  stammt, so liefert mit der Voraussetzung  $P^X, P^Y \in \mathcal{F}_d^2(\phi)$  das starke Gesetz der großen Zahlen für Ein-Stichproben-V-Statistiken der Kernlänge 2 sowie für Zwei-Stichproben-V-Statistiken der Kernlänge (1, 1) die Konvergenz der Spur für fast jedes Paar von Beobachtungsfolgen. Die Spur konvergiert dann für  $i \rightarrow \infty$  gegen die Spur von  $C$

$$\int c(s, s) d\tau(s) = \mathbb{E} \int [1 - \cos((Z_1 - Z_2)'s)] d\tau(s) = \mathbb{E} \phi(|Z_1 - Z_2|^2),$$

$Z_1, Z_2 \sim P^Z$  unabhängig. Zum Nachweis von Bedingung (iii) muss gezeigt werden, dass für alle Elemente der Orthonormalbasis  $(e_j)_j$  und alle  $\varepsilon > 0$

$$\lim_{i \rightarrow \infty} \sum_{j=1}^{m_i} \mathbb{E} [\langle \mathbb{X}_{m_i, j}, e_k \rangle^2 \mathbf{I}(|\langle \mathbb{X}_{m_i, j}, e_k \rangle| > \varepsilon)] = 0,$$

wobei

$$\mathbb{X}_{m_i, j}(z) = \frac{1}{\sqrt{m_i}} \left( (\cos + \sin)(Z_j^{(i)'} z) - \mathbb{E} \left[ (\cos + \sin)(Z_j^{(i)'} z) \right] \right), \quad z \in \mathbb{R}^d$$

und  $Z_1^{(i)}, \dots, Z_{m_i}^{(i)} \sim P^{Z^{(i)}}$  unabhängig. Die Zufallselemente  $\mathbb{X}_{m_i, j}$ ,  $j = 1, \dots, m_i$ , haben also dieselbe Verteilung, so dass mit der Cauchy-Schwarzschen Ungleichung

$$\begin{aligned} & \sum_{j=1}^{m_i} \mathbb{E} [\langle \mathbb{X}_{m_i, j}, e_k \rangle^2 \mathbf{I}(|\langle \mathbb{X}_{m_i, j}, e_k \rangle| > \varepsilon)] \\ &= \mathbb{E} [\langle \sqrt{m_i} \mathbb{X}_{m_i, 1}, e_k \rangle^2 \mathbf{I}(|\langle \mathbb{X}_{m_i, 1}, e_k \rangle| > \varepsilon)] \\ &\leq (\mathbb{E} [\langle \sqrt{m_i} \mathbb{X}_{m_i, 1}, e_k \rangle^4] P(|\langle \mathbb{X}_{m_i, 1}, e_k \rangle| > \varepsilon))^{\frac{1}{2}}. \end{aligned}$$

Der Erwartungswert lässt sich durch nochmalige Anwendung der Cauchy-Schwarzschen Ungleichung, diesmal auf das Innenprodukt in  $\mathcal{H}$ , abschätzen

$$\mathbb{E} [\langle \sqrt{m_i} \mathbb{X}_{m_i, 1}, e_k \rangle^4] \leq \mathbb{E} \|\sqrt{m_i} \mathbb{X}_{m_i, 1}\|^4.$$

Diese Oberschranke lässt sich mit (2.11) explizit ausrechnen:

$$\begin{aligned} \mathbb{E} \|\sqrt{m_i} \mathbb{X}_{m_i, 1}\|^4 &= \mathbb{E} \left[ \int \left( (\cos + \sin)(Z_1^{(i)'} z) - \mathbb{E} \left[ (\cos + \sin)(Z_1^{(i)'} z) \right] \right)^2 d\tau(z) \right]^2 \\ &= \mathbb{E} \iint \left( (\cos + \sin)(Z_1^{(i)'} y) - (\cos + \sin)(Z_2^{(i)'} y) \right) \\ &\quad \left( (\cos + \sin)(Z_1^{(i)'} y) - (\cos + \sin)(Z_3^{(i)'} y) \right) \\ &\quad \left( (\cos + \sin)(Z_1^{(i)'} z) - (\cos + \sin)(Z_4^{(i)'} z) \right) \\ &\quad \left( (\cos + \sin)(Z_1^{(i)'} z) - (\cos + \sin)(Z_5^{(i)'} z) \right) d\tau(y) d\tau(z) \\ &= \mathbb{E} \iint \left[ 1 - \cos \left( y'(Z_1^{(i)} - Z_2^{(i)}) \right) \right] \left[ 1 - \cos \left( z'(Z_1^{(i)} - Z_4^{(i)}) \right) \right] d\tau(y) d\tau(z) \\ &= \mathbb{E} \left[ \phi(|Z_1^{(i)} - Z_2^{(i)}|^2) \phi(|Z_1^{(i)} - Z_4^{(i)}|^2) \right]. \end{aligned}$$

Der Ausdruck der letzten Zeile konvergiert mit der oben erläuterten Zerlegung der Integration nach dem starken Gesetz der großen Zahlen für Ein- und Zwei-Stichproben-V-Statistiken für  $i \rightarrow \infty$  für fast jedes Paar von Beobachtungsfolgen gegen

$$\mathbb{E} \left[ \phi(|Z_1 - Z_2|^2) \phi(|Z_1 - Z_4|^2) \right],$$

wobei  $Z_1, Z_2, Z_4 \sim P^Z$  unabhängig, sofern die Voraussetzung des Gesetzes erfüllt ist, was für  $P^X, P^Y \in \mathcal{F}_d^4(\phi)$  der Fall ist. Damit ist  $E\|\sqrt{m_i}\mathbb{X}_{m_i,1}\|^4$  gleichmäßig in  $i$  beschränkt. Von der Wahrscheinlichkeit  $P(|\langle \mathbb{X}_{m_i,1}, e_k \rangle| > \varepsilon)$  lässt sich mit der Markovschen Ungleichung zeigen, dass sie für  $i \rightarrow \infty$  gegen 0 strebt:

$$P(|\langle \mathbb{X}_{m_i,1}, e_k \rangle| > \varepsilon) \leq \frac{1}{\varepsilon^2 m_i} E[\langle \sqrt{m_i} \mathbb{X}_{m_i,1}, e_k \rangle^2] = \frac{1}{\varepsilon^2 m_i} \langle C_i e_k, e_k \rangle$$

Mit der Argumentation zu Beginn des Beweises gilt  $\langle C_i e_k, e_k \rangle \rightarrow \tilde{\lambda}_k$  für  $i \rightarrow \infty$ , also wegen  $m_i \rightarrow \infty$

$$P(|\langle \mathbb{X}_{m_i,1}, e_k \rangle| > \varepsilon) \xrightarrow{i \rightarrow \infty} 0$$

wiederum für fast jedes Paar von Beobachtungsfolgen. Damit ist der zentrale Grenzwertsatz für Dreiecksschemata im Hilbertraum anwendbar und die Behauptung folgt.  $\square$

Wie zu Beginn dieses Abschnitts und im Beweis von Korollar 3.11 erläutert, erhalten wir das folgende Korollar.

**Korollar 4.3.** *Sei  $\phi \in \Phi_0$  und  $P^X, P^Y \in \mathcal{F}_d^4(\phi)$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^d, \mathcal{B}^d)$ . Es sei  $((m_i, n_i))_i$  eine Folge von Stichprobenumfängen mit  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$ . Dann gilt für fast jedes Paar von Folgen  $(x_n)_n, (y_n)_n$  von Beobachtungen unabhängiger Zufallsvariablen auf  $\mathbb{R}^d$  mit Verteilung  $P^X$  bzw.  $P^Y$*

$$\hat{T}_{m_i, n_i}^\phi \xrightarrow{i \rightarrow \infty} \sum_{\sigma} \tilde{\lambda}_{\sigma} Z_{\sigma}^2,$$

wobei  $(Z_{\sigma})_{\sigma}$  eine Folge unabhängiger standardnormalverteilter Zufallsvariablen ist und  $(\tilde{\lambda}_{\sigma})_{\sigma}$  die Folge der positiven Eigenwerte des Operators  $C$  mit zu Grunde liegender Verteilung  $P^Z := \rho P^X + (1 - \rho) P^Y$ . Die Zufallsvariable  $\hat{T}_{m_i, n_i}^\phi$  besitzt die Verteilung der Testgröße, für den Fall, dass die  $m_i$  ( $n_i$ ) Zufallsvariablen der ersten (zweiten) Stichprobe mit der empirischen Verteilung von  $x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}$  verteilt sind.

Haben wir nun ein Paar von Beobachtungsreihen  $x_1, \dots, x_m$  und  $y_1, \dots, y_n$  gegeben, so können wir die Verteilung der zugehörigen Zufallsvariable  $\hat{T}_{m,n}^\phi$  offensichtlich approximieren, indem wir hinreichend oft zufällig  $m + n$  Werte aus der zusammengesetzten Beobachtung ziehen und diese in die Teststatistik einsetzen und die sich ergebende empirische Verteilung betrachten. Dieser Vorgang, das Monte Carlo-Bootstrap-Verfahren, liefert somit zusammen mit dem letzten Korollar eine Approximation für die Grenzverteilung unter der Hypothese. Wir können mit dieser empirischen Verteilung also kritische Werte bestimmen und mit deren Hilfe den Test durchführen.

## 4.2 Bootstrap in der Grenzverteilung

In diesem Kapitel werden wir ein Verfahren entwickeln, um den kritischen Wert bei der Anwendung der neuen Tests zu bestimmen, welches ohne Monte Carlo-Simulationen auskommt.

Die Grenzverteilung der Teststatistik unter der Hypothese ist eine gewichtete Summe von  $\chi_1^2$ -verteilten Zufallsvariablen. Die Gewichte sind die Eigenwerte des Operators

$$H : L_2(\mathbb{R}^d, \mathcal{B}^d, P^Z) \rightarrow L_2(\mathbb{R}^d, \mathcal{B}^d, P^Z) \quad , \\ f \quad \mapsto \int h(\cdot, x) f(x) dP^Z(x)$$

mit  $h(x, y) := E\phi(|x - Z_1|^2) + E\phi(|y - Z_1|^2) - \phi(|x - y|^2) - E\phi(|Z_1 - Z_2|^2)$ ,  $x, y \in \mathbb{R}^d$ ,  $Z_1, Z_2 \sim P^Z$  unabhängig, wobei  $P^Z$  die gemeinsame Verteilung unter der Hypothese ist. Da  $P^Z$  unbekannt ist, liegt es nahe,  $P^Z$  durch einen Schätzer, wie zum Beispiel die empirische Verteilung der zusammgelegten Beobachtung  $x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}$ , zu ersetzen. Das Eigenwertproblem des zugehörigen Operators  $H_i$  lässt sich auf ein Matrixeigenwertproblem reduzieren. Näheres zur Bestimmung der Eigenwerte findet sich im nächsten Abschnitt 4.2.1. Nimmt man die so gewonnenen Eigenwerte von  $H_i$  als Gewichte einer Summe von Quadraten unabhängiger normalverteilter Zufallsvariablen, so zeigen wir, dass sich für wachsende Stichprobenumfänge in der Tat eine Approximation der Grenzverteilung ergibt. Für die Konvergenz der Eigenwerte benötigen wir eine Art Konvergenz der Operatoren  $H_i$ , die auf den zusammgelegten Beobachtungen basieren, gegen den Operator  $H$ , der auf der unbekanntenen gemeinsamen Verteilung beruht, vergleiche Satz C.4 im Anhang C. Da die Operatoren  $H_i, H$  auf unterschiedlichen  $L_2$ -Räumen definiert sind, lässt sich die gewöhnliche Störungstheorie linearer Operatoren nicht verwenden. Eine Alternative dazu bietet zumindest in einigen Fällen die Theorie von Stummel [Stu70, Stu71, Stu72], die Operatoren auf unterschiedlichen Räumen zulässt. In Kapitel 5 werden wir diesen Ansatz genauer erläutern. Für die in dieser Arbeit eingeführten Teststatistiken gelingt der Nachweis der Konvergenz der Eigenwerte mit einem Kunstgriff: In Abschnitt 3.4 wurde gezeigt, dass die nicht verschwindenden Eigenwerte des Operators  $H$  auch in ihrer Vielfachheit mit denen des Operators

$$C : L_2(\mathbb{R}^d, \mathcal{B}^d, \tau) \rightarrow L_2(\mathbb{R}^d, \mathcal{B}^d, \tau) \quad , \\ e \quad \mapsto \int c(\cdot, z)e(z) d\tau(z),$$

übereinstimmen, wobei

$$c(y, z) = E[(\cos + \sin)(Z_1' y)(\cos + \sin)(Z_1' z)] \\ - E[(\cos + \sin)(Z_1' y)] E[(\cos + \sin)(Z_1' z)], \quad y, z \in \mathbb{R}^d,$$

$Z_1 \sim P^Z$ . Natürlich folgt, dass auch die positiven Eigenwerte von  $H_i$  und  $C_i$  übereinstimmen. Bei letzteren Operatoren bleibt der Raum  $\mathcal{H} = L_2(\mathbb{R}^d, \mathcal{B}^d, \tau)$ , auf dem

sie definiert sind, stets derselbe, lediglich der Kern des Operators wandelt sich. Für die Verteilungskonvergenz der gewichteten Summe von  $\chi_1^2$ -Zufallsvariablen mit den Eigenwerten von  $H_i$  als Gewichten genügt also zu zeigen, dass Gaußsche Zufallselemente in  $\mathcal{H}$  mit Kovarianzoperator  $C_i$  für  $i \rightarrow \infty$  in Verteilung gegen ein Gaußsches Zufallselement mit Kovarianzoperator  $C$  konvergieren. Wir führen diesen Nachweis analog zum Beweis von Satz 4.2 mit Hilfe des zentralen Grenzwertsatzes für Dreiecksschemata im Hilbertraum, Satz D.6, Anhang D. Die gewünschte Aussage über die Verteilung der Teststatistik ergibt sich dann mit Hilfe des Satzes von der stetigen Abbildung durch den Übergang zum Normquadrat und den Überlegungen aus Korollar 3.11.

**Satz 4.4.** *Sei  $\phi \in \Phi_0$  und  $P^X, P^Y \in \mathcal{F}_d^4(\phi)$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^d, \mathcal{B}^d)$ . Es sei  $((m_i, n_i))_i$  eine Folge von Stichprobenumfängen mit  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$ . Dann gilt für fast jedes Paar von Folgen  $(x_n)_n, (y_n)_n$  von Beobachtungen unabhängiger Zufallsvariablen auf  $\mathbb{R}^d$  mit Verteilung  $P^X$  bzw.  $P^Y$*

$$\mathbb{Z}_i \xrightarrow[i \rightarrow \infty]{\mathcal{D}} \mathbb{Z}$$

im Hilbertraum  $\mathcal{H}$ , wobei  $\mathbb{Z}_i$  ( $\mathbb{Z}$ ) ein zentriertes Gaußsches Zufallselement in  $\mathcal{H}$  mit Kovarianzoperator  $C_i$  ( $C$ ) sei. Die Operatoren  $C_i, C$  seien wie in Lemma 4.1 definiert.

**Beweis.** Zunächst existieren die Gaußschen Zufallselemente nach einem Satz von Vakhania [Vak81, S.55]. Es seien  $\mathbb{X}_{m_i, j}$ ,  $j = 1, \dots, m_i$ , unabhängige Gaußsche Zufallselemente, jeweils mit Kovarianzoperator  $\frac{1}{m_i}C_i$ . Dann besitzen  $\sqrt{m_i}\mathbb{X}_{m_i, j}$ ,  $j = 1, \dots, m_i$  und  $\sum_{j=1}^{m_i} \mathbb{X}_{m_i, j}$  den Kovarianzoperator  $C_i$ . Der Nachweis der Bedingungen (i) und (ii) erfolgt wie im Beweis von Satz 4.2. Die Lindeberg-Bedingung (iii) zeigen wir ebenfalls ähnlich: Mit der Cauchy-Schwarzschen Ungleichung gilt

$$\begin{aligned} & \sum_{j=1}^{m_i} \mathbb{E} [\langle \mathbb{X}_{m_i, j}, e_k \rangle^2 \mathbf{I}(|\langle \mathbb{X}_{m_i, j}, e_k \rangle| > \varepsilon)] \\ & \leq (\mathbb{E} [\langle \sqrt{m_i} \mathbb{X}_{m_i, 1}, e_k \rangle^4] P(|\langle \mathbb{X}_{m_i, 1}, e_k \rangle| > \varepsilon))^{\frac{1}{2}}. \end{aligned}$$

Bei dem Ausdruck  $\langle \sqrt{m_i} \mathbb{X}_{m_i, 1}, e_k \rangle$  handelt es sich um eine reellwertige Zufallsvariable mit einer Normalverteilung mit Erwartungswert 0. Also ist

$$\mathbb{E} [\langle \sqrt{m_i} \mathbb{X}_{m_i, 1}, e_k \rangle^4] = 3 (\mathbb{E} [\langle \sqrt{m_i} \mathbb{X}_{m_i, 1}, e_k \rangle^2])^2 = 3 \langle C_i e_k, e_k \rangle^2.$$

Dieser Faktor konvergiert für  $i \rightarrow \infty$  gegen  $3\tilde{\lambda}_k^2$ .

Die Wahrscheinlichkeit  $P(|\langle \mathbb{X}_{m_i, 1}, e_k \rangle| > \varepsilon)$  lässt sich mit der Markovschen Ungleichung wie zuvor abschätzen, so dass der zentrale Grenzwertsatz im Hilbertraum für Dreiecksschemata gilt und die Behauptung folgt.  $\square$

Mit dem Satz von der stetigen Abbildung ergibt sich aus dem eben bewiesenen Satz unmittelbar das gesuchte Korollar.

**Korollar 4.5.** *Seien  $\phi \in \Phi_0$  und  $P^X, P^Y \in \mathcal{F}_d^4(\phi)$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^d, \mathcal{B}^d)$ . Es sei  $((m_i, n_i))_i$  eine Folge von Stichprobenumfängen mit  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$ . Es seien  $(\tilde{\lambda}_k^{(i)})$  die Eigenwerte des in Lemma 4.1 definierten Operators  $C_i$  und  $(\tilde{\lambda}_k)$  die Eigenwerte des Operators  $C$ . Dann gilt für fast jedes Paar von Folgen  $(x_n)_n, (y_n)_n$  von Beobachtungen unabhängiger Zufallsvariablen auf  $\mathbb{R}^d$  mit Verteilung  $P^X$  bzw.  $P^Y$*

$$\sum_{\sigma} \tilde{\lambda}_{\sigma}^{(i)} Z_{\sigma}^2 \xrightarrow{i \rightarrow \infty} \sum_{\sigma} \tilde{\lambda}_{\sigma} Z_{\sigma}^2.$$

**Beweis.** Darstellung des Zufallselements  $\sqrt{m_i} \mathbb{X}_{m_i, j}$  aus dem Beweis von Satz 4.4 in der Basis der Eigenvektoren von  $C_i$  wie im Beweis von Korollar 3.11.  $\square$

Dieses Korollar berechtigt uns, die aus den approximierten Eigenwerten  $(\lambda_{\sigma}^{(i)})_{\sigma}$  gewonnene Verteilung zur Bestimmung von kritischen Werten zu nutzen, da sich die Konvergenz der Verteilungen offensichtlich auf deren Quantile überträgt.

Das vorgestellte Verfahren lebt davon, dass die beiden Operatoren  $H_i$  und  $C_i$  dieselben Eigenwerte haben. Ohne diese Eigenschaft wäre die Eigenwertberechnung der Operatoren  $C_i$  schwierig, da hier durch Verwendung des empirischen Maßes im Kovarianzoperator keine nennenswerte Vereinfachung auftritt. Dies ist beim Operator  $H_i$  anders: Im nächsten Abschnitt beschäftigen wir uns mit der Reduktion des Eigenwertproblems des Operators  $H_i$  auf ein lineares Gleichungssystem.

### 4.2.1 Reduktion des Eigenwertproblems

Dieser Abschnitt erläutert die Umwandlung des Eigenwertproblems für den Operator  $H_i$  in ein lineares Gleichungssystem. Gegeben sei die zusammengelegte Beobachtung zweier Stichproben  $x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}$  vom Umfang  $N = m_i + n_i$ , die im Folgenden durch  $z_1, \dots, z_N$  bezeichnet sei. Dabei sei zunächst angenommen, dass es sich um  $N$  verschiedene Beobachtungen handelt.  $P^{Z^{(i)}}$  sei die diskrete Gleichverteilung auf  $z_1, \dots, z_N$ . Jede Funktion  $f$  auf dem Raum  $L_2(\mathbb{R}^d, \mathcal{B}^d, P^{Z^{(i)}})$ , auf dem der Operator  $H_i$  definiert ist, kann als  $N$ -dimensionaler Vektor

$$\vec{f} = (f_1, \dots, f_N)' = (f(z_1), \dots, f(z_N))' \in \mathbb{R}^N$$

aufgefasst werden, der die Funktionswerte an den Stellen  $z_1, \dots, z_N \in \mathbb{R}^d$  enthält. Das Eigenwertproblem

$$\lambda f = H_i f$$

lässt sich also in Form eines linearen Gleichungssystems schreiben

$$\lambda f_j = \sum_{k=1}^N a_{j,k} f_k, \quad j \in \{1, \dots, N\}.$$

Dabei ist die Matrix  $A = (a_{j,k})_{j,k}$ ,  $j, k \in \{1, \dots, N\}$ , gegeben durch

$$a_{j,k} := \frac{1}{N} h(z_j, z_k) = \frac{1}{N^3} \sum_{l=1}^N \sum_{\ell=1}^N \bar{h}(z_j, z_l; z_k, z_\ell).$$

In Matrixform bleibt folgendes Problem zu lösen: *Finde  $\lambda \in \mathbb{R} \setminus \{0\}$  und  $\vec{f} \in \mathbb{R}^N \setminus \{0\}$ , so dass*

$$\lambda \vec{f} = A \vec{f}, \quad \vec{f} \in \mathbb{R}^N. \quad (4.5)$$

Dies ist ein klassisches Matrix-Eigenwertproblem für die  $N \times N$  Matrix  $A$ . Wegen der Selbstadjungiertheit, also Symmetrie von  $H$  ist  $A$  symmetrisch und die Lösung des Eigenwertproblems besteht aus  $N$  reellen Eigenwerten  $(\lambda_\sigma^{(i)})_\sigma$ . Davon ist mindestens einer 0, da die konstante Funktion  $f \equiv 1$  offensichtlich eine Eigenfunktion (hier -vektor) zum Eigenwert 0 ist.

Man hätte das diskretisierte Eigenwertproblem (4.5) statt für den Operator  $H$  auch für den Operator  $\bar{H}$  betrachten können. Bei  $N$  Beobachtungen in der zusammengelegten Stichprobe hätte dies allerdings auf die Bestimmung der Eigenwerte einer  $N^2 \times N^2$ -Matrix geführt. Auch hätte dies wegen der in Abschnitt 3.1.2 gezeigten Bijektion zwischen den nicht verschwindenden Eigenwerte von  $H$  und  $\bar{H}$  keine weiteren Erkenntnisse gebracht,  $N^2 - N + 1$  der ermittelten Eigenwerte wären zwangsläufig 0.

Sind die Beobachtungen nicht, wie zunächst angenommen, paarweise verschieden, so müsste man gleiche Beobachtungen zusammenfassen und das empirische Maß anpassen. Die Matrix  $A$  wäre dann gegeben durch

$$a_{j,k} = p_k h(\tilde{z}_j, \tilde{z}_k), \quad j, k \in \{1, \dots, \tilde{N}\},$$

wobei  $p_k$  die Wahrscheinlichkeit ist, die das empirische Maß der Beobachtungen  $z_1, \dots, z_N$  der Beobachtung  $\tilde{z}_k$  zuordnet.  $\{\tilde{z}_1, \dots, \tilde{z}_{\tilde{N}}\}$  sei dabei die Menge der paarweise verschiedenen Beobachtungen. Damit würden die identischen Zeilen und Spalten vermieden, die sich in der Matrix  $A$  ergeben würden, wenn man die Beobachtungen nicht zusammenfasst. Für die positiven Eigenwerte ist dieser Aspekt allerdings ohne Belang, da durch die identischen Zeilen lediglich die Vielfachheit des Eigenwerts 0 erhöht, nicht aber das ursprüngliche Eigenwertproblem verändert wird. Ein Zusammenfassen der Beobachtungen hat zwar für die numerische Bestimmung der Eigenwerte den Vorteil, dass die Dimension der Matrix kleiner wird, jedoch zugleich den Nachteil, dass die Matrix nicht mehr symmetrisch ist.

Als kritischen Wert eines Tests zum Niveau  $\alpha$  wählt man das  $1 - \alpha$ -Quantil der Verteilung von  $S^{(i)} := \sum_\sigma \lambda_\sigma^{(i)} Z_\sigma^2$ . Dazu müssen wir die Verteilungsfunktion  $F: \mathbb{R} \rightarrow \mathbb{R}$  von  $S^{(i)}$  aus den Eigenwerten berechnen.

Wegen der Unabhängigkeit der Summanden in  $S^{(i)}$  lässt sich die zugehörige charakteristische Funktion als Produkt der charakteristischen Funktionen der Summanden

$$\varphi(t) = \prod_{\sigma} \frac{1}{\sqrt{1 - 2it\lambda_{\sigma}^{(i)}}}, \quad t \in \mathbb{R},$$

sehr leicht bestimmen. Mit Hilfe einer Inversionsformel (siehe Gurland [Gur48]) lässt sich daraus die Verteilungsfunktion

$$F(x) = \frac{1}{2} + \frac{1}{2\pi} \lim_{\varepsilon \downarrow 0, T \rightarrow +\infty} \int_{\varepsilon}^T \frac{\exp(itx)\varphi(-t) - \exp(-itx)\varphi(t)}{it} dt, \quad x \in \mathbb{R},$$

zurückgewinnen. Bestimmt man das Integral durch das Trapez-Verfahren, so lässt sich hervorragend die Fast-Fourier-Transformation zur Summation verwenden. Ein Vorschlag zur Implementation dieses Verfahrens ist in dem in der Einleitung erwähnten **R** Paket enthalten.

## 4.2.2 Einhaltung des Niveaus

Als Probe, ob das Verfahren funktioniert, wird die Einhaltung des Niveaus des Cramér-Tests (Testgröße  $T_{m,n}^{\text{Cramér}}$ , (1.2)) in Spezialfällen überprüft.

Dazu werden bei sieben unterschiedlichen Verteilungen drei unterschiedliche Stichprobenumfänge,  $m = n = 20$ ,  $m = 20, n = 50$  und  $m = n = 50$ , sowie zwei unterschiedliche Testniveaus  $\alpha = 0.05$ ,  $\alpha = 0.1$  untersucht.

Im eindimensionalen Fall beschränken wir uns auf

- ein Beispiel für eine stetige Verteilung, die Standardnormalverteilung  $N(0, 1)$ , und
- ein Beispiel für eine diskrete Verteilung, die Poisson-Verteilung  $\text{Po}(2)$  mit dem Erwartungswert 2.

Im Fall  $d = 2$  betrachten wir als Beispiele stetiger Verteilungen

- die zweidimensionale Standardnormalverteilung  $N_2(0, I_2)$  mit Erwartungswertvektor  $(0, 0)'$  und der zweidimensionalen Einheitsmatrix  $I_2$  als Kovarianzmatrix sowie
- die zweidimensionale logistische Verteilung  $L_2(1)$  mit Parameter 1, wobei  $L_d(\kappa)$  für  $\kappa > 0$  definiert sei als die Verteilung mit Verteilungsfunktion

$$F(x_1, \dots, x_d) = \left( 1 + \sum_{j=1}^d \exp(-x_j) \right)^{-\kappa}, \quad x_1, \dots, x_d \in \mathbb{R},$$

(siehe Johnson [Joh87, S.161]).

Tabelle 4.1: Geschätzte Verwerfungswahrscheinlichkeit in der Hypothese

$m$	$n$	$N(0, 1)$	$N_2(0, I_2)$	$\text{unif}[0, 1]^4$	$L_2$	$L_3$	$\text{Po}(2)$	$\text{Po}_2$
20	20	0.1003	0.0973	0.0903	0.0989	0.0954	0.0932	0.0993
		0.0496	0.0439	0.0389	0.0460	0.0460	0.0462	0.0487
20	50	0.0936	0.1027	0.0970	0.0984	0.0929	0.0979	0.0989
		0.0457	0.0521	0.0457	0.0467	0.0441	0.0459	0.0469
50	50	0.0966	0.0985	0.0994	0.0986	0.0923	0.0961	0.1001
		0.0473	0.0490	0.0478	0.0481	0.0464	0.0489	0.0515

Als Beispiel einer diskreten Verteilung wählen wir hier

- eine zweidimensionale Poisson-Verteilung  $\text{Po}_2$ , also die gemeinsame Verteilung von  $Z_1 := W_1 + W_{12}$  und  $Z_2 := W_2 + W_{12}$ , wobei  $W_1 \sim \text{Po}(2)$ ,  $W_2 \sim \text{Po}(3)$  und  $W_{12} \sim \text{Po}(1)$  (vergleiche Johnson, Kotz und Balakrishnan [JKB97, S.124]).

Für drei- und vierdimensionale Daten betrachten wir jeweils nur eine Verteilung und zwar

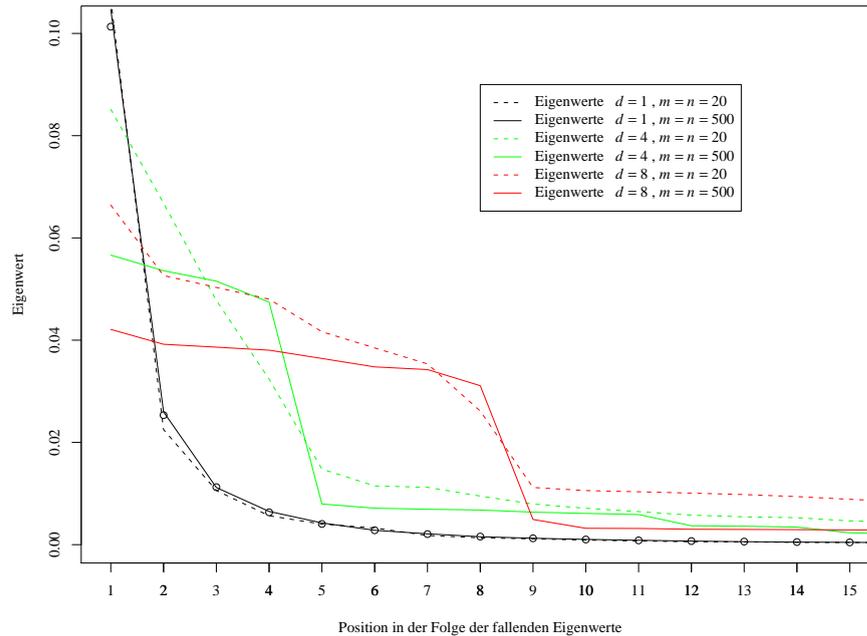
- die dreidimensionale logistische Verteilung  $L_3(1)$  mit Parameter 1 und
- die Gleichverteilung  $\text{unif}[0, 1]^4$  auf dem 4-dimensionalen Einheitswürfel  $[0, 1]^4$ .

Der in der Tabelle 4.1 angegebene Wert ist der Anteil der Verwerfungen von 10000 Wiederholungen, wenn beide Stichproben dieselbe zu Grunde liegende Verteilung besitzen, also eine Approximation des Fehlers 1. Art. Bei festem Stichprobenumfang bezieht sich die obere Zeile auf die Vorgabe  $\alpha = 0.1$ , die untere Zeile auf  $\alpha = 0.05$ .

Die Ergebnisse liegen selbst bei geringen Stichprobenumfängen  $m = n = 20$  für alle Verteilungen der Dimensionen 1 – 3 hinreichend genau am vorgegebenen Niveau und entsprechen damit denen, die für den Fall der Bestimmung des kritischen Wertes mit Monte Carlo-Bootstrap-Methoden von Baringhaus und dem Autor dieser Arbeit [BF04] bereits in derselben Untersuchung gemacht wurden. Bei der Betrachtung der geschätzten Verwerfungswahrscheinlichkeiten bei zu Grunde liegender Verteilung  $\text{unif}[0, 1]^4$  fällt auf, dass diese bei  $m = n = 20$  und Vorgabe von  $\alpha = 0.05$  etwa 20% zu gering geschätzt werden. Der Test ist leicht konservativ. Auf dem Weg zu  $m = n = 50$  verringert sich die Abweichung von der Vorgabe und ist vergleichbar mit den anderen. Simulationen beim Stichprobenumfang  $m = n = 20$  und höheren Dimensionen, beispielsweise der Verteilung  $\text{unif}[0, 1]^8$ , haben gezeigt, dass der Test hier noch konservativer wird.

Eine Erklärung für dieses Verhalten gibt die Betrachtung der Eigenwerte. In Abbildung 4.1 sind die approximierten 15 größten Eigenwerte bei  $m = n = 20$  Beobachtungen und  $m = n = 500$  Beobachtungen bei zu Grunde liegender  $\text{unif}[0, 1]$ ,

Abbildung 4.1: Eigenwertapproximation in unterschiedlichen Dimensionen und bei verschiedenen Stichprobenumfängen



unif  $[0, 1]^4$  und unif  $[0, 1]^8$ -Verteilung dargestellt. Der Übersichtlichkeit halber sind die Eigenwerte zu einer Verteilung und einem Stichprobenumfang in der Abbildung durch Linien miteinander verbunden. Im Fall unif  $[0, 1]$  sind zusätzlich durch Punkte die exakten Eigenwerte der Grenzverteilung angegeben (vergleiche Abschnitt 3.2.1).

Die Abbildung zeigt, dass die Eigenwertapproximation bei zu hoher Dimension und zu geringem Stichprobenumfang zu großen Abweichungen führt. Steht man daher vor dem Problem bei wenigen Daten hoher Dimension einen kritischen Wert der Testgröße ermitteln zu müssen, so ist es ratsam auf das Monte Carlo-Bootstrap-Verfahren auszuweichen, welches diesen Nachteil nicht aufweist (vergleiche [BF04]).

### 4.3 Anwendung der neuen Tests

Mit dem bisher gesammelten Wissen, insbesondere Satz 2.1, Korollar 4.3 und Korollar 4.5 sowie den Überlegungen in Abschnitt 4.2.1, lässt sich nun die Anwendung eines Tests der neuen Familie formulieren.

**Definition 4.6.** Seien  $x_1, \dots, x_m$  und  $y_1, \dots, y_n$  Beobachtungen unabhängiger  $d$ -dimensionaler Zufallsvariablen  $X_1, \dots, X_m \sim P^X$ ,  $Y_1, \dots, Y_n \sim P^Y$ . Sei  $\phi \in \Phi_0$  und

$P^X, P^Y \in \mathcal{F}_d^4(\phi)$ . Für das Testproblem der Hypothese

$$\begin{aligned} H &: X, Y \text{ haben dieselbe Verteilung} && \text{gegen die Alternative} \\ K &: X, Y \text{ haben nicht dieselbe Verteilung} \end{aligned}$$

zum Niveau  $\alpha > 0$  wird mit der Statistik

$$T_{m,n}^\phi(x_1, \dots, x_m; y_1, \dots, y_n) := \frac{mn}{m+n} \left[ \frac{2}{mn} \sum_{j=1}^m \sum_{k=1}^n \phi(|x_j - y_k|^2) - \frac{1}{m^2} \sum_{j=1}^m \sum_{k=1}^m \phi(|x_j - x_k|^2) - \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \phi(|y_j - y_k|^2) \right] \quad (4.6)$$

der Test

$$\varphi_\phi(x_1, \dots, x_m; y_1, \dots, y_n) := \begin{cases} 1 & , \text{ falls } T_{m,n}^\phi > c_{1-\alpha}(x_1, \dots, x_m; y_1, \dots, y_n) \\ 0 & , \text{ falls } T_{m,n}^\phi \leq c_{1-\alpha}(x_1, \dots, x_m; y_1, \dots, y_n) \end{cases}$$

definiert. Dabei sei  $c_{1-\alpha}(x_1, \dots, x_m; y_1, \dots, y_n)$  ein Schätzwert für das  $1 - \alpha$ -Quantil von  $T_{m,n}^\phi$ , der auf den Beobachtungen  $x_1, \dots, x_m, y_1, \dots, y_n$  beruht.

Funktionierende Approximationsmethoden für das  $1 - \alpha$ -Quantil von  $T_{m,n}^\phi$  sind das Monte Carlo-Bootstrap-Verfahren (siehe Korollar 4.3) oder die Approximation der Grenzverteilung durch das Eigenwertverfahren (vergleiche Korollar 4.5).

Das anschließende Kapitel beschäftigt sich mit einer alternativen Möglichkeit, die Aussage des Korollars 4.5 zu erhalten. Abschließend sei die folgende Konsistenzaussage erwähnt:

**Satz 4.7 (Konsistenz).** *Es sei  $\phi \in \Phi_0$ ,  $P^X, P^Y \in \mathcal{F}_d^4(\phi)$  und  $((m_i, n_i))_i$  eine Folge von Stichprobenumfängen mit  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$ . Dann liefert die Folge von Tests  $(\varphi_\phi)_{m_i, n_i}$  ein konsistentes Verfahren, d.h. die Wahrscheinlichkeit für einen Fehler 2. Art konvergiert für  $i \rightarrow \infty$  gegen 0. Dabei werde der kritische Wert mit Hilfe eines der beiden in Kapitel 4 besprochenen Verfahren bestimmt.*

**Beweis.** Das starke Gesetz der großen Zahlen für verallgemeinerte V-Statistiken, Satz B.4, Anhang B, besagt unter den gemachten Voraussetzungen fast sicher

$$\frac{m_i + n_i}{m_i n_i} T_{m_i, n_i}^\phi \xrightarrow{i \rightarrow \infty} d_\phi^2(P^X, P^Y),$$

wobei  $d_\phi^2(P^X, P^Y) > 0$ , da  $X$  und  $Y$  in der Alternative nicht dieselbe Verteilung haben. Die Verteilung, deren  $1 - \alpha$ -Quantil man bei Verwendung eines der beiden Schätzverfahren für die Grenzverteilung als kritischen Wert wählt, konvergiert gegen die asymptotische Grenzverteilung (für  $i \rightarrow \infty$ ) der Testgröße bei Annahme einer gemeinsamen zu Grunde liegenden Verteilung  $\rho P^X + (1 - \rho) P^Y$ . Damit sind

die kritischen Werte  $c_{1-\alpha}(x_1, \dots, x_{m_i}; y_1, \dots, y_{n_i})$  fast sicher, also auch stochastisch beschränkt und mit  $X_1, \dots, X_{m_i} \sim P^X, Y_1, \dots, Y_{n_i} \sim P^Y, P^X \neq P^Y$ , folgt

$$\frac{m_i + n_i}{m_i n_i} c_{1-\alpha}(X_1, \dots, X_{m_i}; Y_1, \dots, Y_{n_i}) \xrightarrow{i \rightarrow \infty} 0.$$

Die Wahrscheinlichkeit für einen Fehler 2. Art,

$$\begin{aligned} & P(T_{m_i, n_i}^\phi \leq c_{1-\alpha}(X_1, \dots, X_{m_i}; Y_1, \dots, Y_{n_i})) \\ &= P\left(\frac{m_i + n_i}{m_i n_i} T_{m_i, n_i}^\phi \leq \frac{m_i + n_i}{m_i n_i} c_{1-\alpha}(X_1, \dots, X_{m_i}; Y_1, \dots, Y_{n_i})\right) \end{aligned}$$

konvergiert also für  $i \rightarrow \infty$  gegen 0. □

# Kapitel 5

## Approximation linearer Operatoren

Im letzten Kapitel, Abschnitt 4.2, wurde eine neue Approximation der Grenzverteilung vorgestellt. Die Grenzverteilung ist eine gewichtete Summe von Quadraten normalverteilter Zufallsvariablen. Als Gewichte ergeben sich die Eigenwerte  $(\lambda_\sigma)_\sigma$  des durch (3.9) definierten Operators  $H$ ,

$$H : \begin{array}{ccc} L_2(\mathbb{R}^d, \mathcal{B}^d, P^Z) & \rightarrow & L_2(\mathbb{R}^d, \mathcal{B}^d, P^Z) \\ f & \mapsto & \int h(\cdot, x) f(x) dP^Z(x) \end{array} ,$$

wobei  $h(x, y) := E\bar{h}(x, Y_1; y, Y_2)$ ,  $x, y \in \mathbb{R}^d$ . Es gilt, die Unkenntnis der zu Grunde liegenden Verteilung unter der Hypothese  $P^Z$  zu überwinden. Wie beim Monte Carlo-Bootstrap-Verfahren erscheint es einleuchtend, die unbekanntes Verteilung durch ihr auf der zusammengelegten Beobachtung beruhendes Gegenstück zu ersetzen. Dieses Vorgehen haben wir in Abschnitt 4.2 gewählt. Dabei wurde die spezielle Struktur des Kerns, der sich als Linearkombination der Funktionen  $\phi$  mit vollständig monotoner Ableitung, angewendet auf bestimmte euklidische Abstände, schreiben lässt, ausgenutzt.

In diesem Kapitel werden wir einen neuen Beweis für die Aussage des Korollars 4.5, welches besagt, dass diese Eigenwertapproximation zulässig ist, finden. Geht man nicht über zu der alternativen Darstellung mittels empirischer charakteristischer Funktionen, so ist man gezwungen, die Integraloperatoren  $H_i$  zu betrachten, die auf  $L_2$ -Räumen, versehen mit der empirischen Verteilung von  $m_i$  Beobachtungen der ersten und  $n_i$  Beobachtungen der zweiten Stichprobe definiert sind. Auch die Kernfunktion  $h_i$  hängt von den Beobachtungen ab. Wir werden nun zeigen, dass die Eigenwerte des Operators  $H_i$  geeignet gegen die Eigenwerte von  $H$  konvergieren. Unglücklicherweise kann man diese Approximation nicht durch die Einschränkung der Operatoren auf kompakte Teilmengen zeigen, die in Abschnitt 3.1.2 für den Beweis von Satz 3.5 verwendet wird. Das große Problem bei diesem Ansatz sind

die unterschiedlichen Räume, auf denen die Operatoren  $H_i$  definiert sind. Es wird in diesem Kapitel gezeigt, dass für fast jede Beobachtungsfolge Verteilungskonvergenz der Summe mit approximierten Gewichten gegen die tatsächlichen Eigenwerte vorliegt, sofern die Kernfunktionen  $h, (h_i)_i$  bestimmte Voraussetzungen erfüllen. Da man das Prinzip auch auf andere Operatoren als den für die asymptotische Verteilung der neuen Teststatistiken relevanten Operator  $H$  anwenden kann, lassen wir Integraloperatoren der folgenden Bauart zu:

$$H : \begin{array}{l} L_2^{\mathbb{C}}(\mathbb{R}^d, \mathcal{B}^d, P^Z) \rightarrow L_2^{\mathbb{C}}(\mathbb{R}^d, \mathcal{B}^d, P^Z) \quad , \\ f \quad \mapsto \int h(\cdot, y) f(y) dP^Z(y), \end{array} \quad (5.1)$$

wobei auf die Bedingungen, die an die noch nicht genauer spezifizierte Kernfunktion  $h$  gestellt werden, später eingegangen wird.  $L_2^{\mathbb{C}}(\mathbb{R}^d, \mathcal{B}^d, P^Z)$  bezeichnet den Raum der Äquivalenzklassen komplexwertiger, quadratisch integrierbarer Funktionen auf  $\mathbb{R}^d$ . Zur Approximation sollen Operatoren

$$H_i : \begin{array}{l} L_2^{\mathbb{C}}(\mathbb{R}^d, \mathcal{B}^d, P_i^Z) \rightarrow L_2^{\mathbb{C}}(\mathbb{R}^d, \mathcal{B}^d, P_i^Z) \quad , \\ f \quad \mapsto \int h_i(\cdot, y) f(y) dP_i^Z(y) \end{array} \quad (5.2)$$

dienen, wobei  $P_i^Z$  das bereits erwähnte empirische Maß sei und  $h_i$  eine Folge von Kernfunktionen, die in geeigneter Weise gegen  $h$  konvergieren mögen.

Wir werden für diese Betrachtung eine Theorie verwenden, die Stummel in den siebziger Jahren zur Störungstheorie linearer Operatoren aufgestellt hat [Stu70, Stu71, Stu72]. Im Folgenden seien alle auftretenden Operatoren als linear vorausgesetzt.

## 5.1 Diskrete Approximation von Räumen

Anstatt eine Folge von Operatoren auf stets ein und demselben normierten Raum  $E$  zu verwenden, hat Stummel zugelassen, dass die Operatoren auf unterschiedlichen, normierten Räumen  $E_\iota$  definiert sind, sofern diese Räume geeignet gegen einen Raum  $E$  streben. Dabei sei  $\iota \in I$  und  $I \subset \mathbb{N}$  eine abzählbare, nicht endliche Indexmenge.

Um Aussagen über die Konvergenz von Funktionen  $u_\iota \rightarrow u$  machen zu können, wobei  $u_\iota \in E_\iota$ ,  $\iota \in I$ , und  $u \in E$  ein Element des „Grenz-“Raumes  $E$  ist, führt man den Begriff von **äquivalenten** Folgen ein. Die Abkürzung  $\iota \in I$ ,  $\iota \rightarrow \infty$  möge in der Folge stets bedeuten, dass die Elemente von  $I$  der Größe nach geordnet, aufsteigend durchlaufen werden.

**Definition 5.1.** Zwei Folgen  $(u_\iota)_{\iota \in I}$ ,  $(v_\iota)_{\iota \in I}$  mit  $u_\iota, v_\iota \in E_\iota$  heißen **äquivalent**, wenn

$$\|u_\iota - v_\iota\|_{E_\iota} \xrightarrow{\iota \rightarrow \infty} 0.$$

Offensichtlich wird durch diesen Begriff eine Äquivalenzrelation auf dem Raum der Folgen  $(u_\iota)_\iota \in \prod_\iota E_\iota$  definiert. Entscheidend für die Anwendung des Konzepts ist die Wahl einer Abbildung  $Q$ , welche die Elemente  $u \in E$  mit der jeweiligen Äquivalenzklasse der gegen  $u$  konvergenten Folgen  $(u_\iota)_\iota, \iota \in I$ , verknüpft.

**Definition 5.2.** Die Folge der normierten Räumen  $(E_\iota)_\iota, \iota \in I$ , nennt man **diskrete Approximation**

$$\mathcal{A} \left( E, \prod_\iota E_\iota, Q \right)$$

des normierten Raumes  $E$ , wenn die folgenden Bedingungen erfüllt sind:

- (i)  $E, E_\iota, \iota \in I$ , sind über demselben skalaren Körper  $\mathbb{K}$  (im Folgenden stets  $\mathbb{K} = \mathbb{C}$ ) definiert und es existiert eine Abbildung  $Q$ , welche jedem Element  $u \in E$  eine Äquivalenzklasse  $Q(u) \subset \prod_\iota E_\iota$  zuordnet.
- (ii) [**Stetigkeit der Normen auf  $Q(u)$** ] Für alle  $u \in E$  und alle Folgen  $(u_\iota)_\iota \in Q(u)$  konvergieren die Normen

$$\|u_\iota\|_{E_\iota} \xrightarrow{\iota \rightarrow \infty} \|u\|_E.$$

- (iii) [**Linearität**] Für alle  $\alpha, \beta \in \mathbb{K}$ , alle  $u, v \in E$  und alle  $(u_\iota)_\iota \in Q(u), (v_\iota)_\iota \in Q(v), (w_\iota)_\iota \in Q(\alpha u + \beta v)$  gilt

$$\|\alpha u_\iota + \beta v_\iota - w_\iota\|_{E_\iota} \xrightarrow{\iota \rightarrow \infty} 0.$$

Hat man eine solche Funktion  $Q$  festgelegt, so spricht man von **diskreter Konvergenz** von  $u_\iota \rightarrow u$  genau dann, wenn  $(u_\iota)_\iota \in Q(u)$ .

**Anwendung in der Statistik** Gesucht ist eine Möglichkeit, die Eigenwerte des durch (5.1) definierten Operators durch die Eigenwerte der durch (5.2) definierten Operatoren zu approximieren. Der Raum, auf dem der Operator  $H$  definiert sein soll, möge der Raum der beschränkten stetigen, komplexwertigen Funktionen  $C_b(R)$  versehen mit der  $L_2^{\mathbb{C}}(R, \mathcal{R}, P^Z)$ -Norm sein, wobei  $R$  der Träger des Maßes  $P^Z$  sei.  $C_b(R)$  ist mit dieser Norm ein normierter Raum.  $\mathcal{R}$  sei im Folgenden die Einschränkung der Borelschen  $\sigma$ -Algebra  $\mathcal{B}^d$  auf  $R$ . Bei der Approximation kommen die Räume der auf  $X_i := \{x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}\}$  definierten, komplexwertigen Funktionen zum Einsatz, die mit der  $L_2^{\mathbb{C}}(X_i, \mathcal{P}(X_i), P_i^Z)$ -Norm versehen werden, wobei  $P_i^Z$  das empirische Maß der Beobachtungen  $x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}$  ist. Mit  $\mathcal{P}(X_i)$  sei die Potenzmenge von  $X_i$  bezeichnet.

Die Betrachtung stetiger Funktionen wird keine große Einschränkung darstellen, da unter den Annahmen, die wir später vornehmen werden, die Eigenfunktionen

der interessierenden Operatoren stetig sind. Die Verwendung beschränkter stetiger komplexwertiger Funktionen, also aus  $C_b(R)$ , führt dazu, dass das auf diesem Weg gewonnene Resultat bei Anwendung auf die für die neuen Teststatistiken relevanten Operatoren lediglich beschränkte Funktionen  $\phi \in \Phi$  zulässt.

Der folgende Satz bezieht sich nicht nur die Situation eines Zwei-Stichproben-Tests, sondern behandelt sowohl den Ein- als auch den Zwei-Stichproben-Fall statistischer Fragestellungen vor diesem Hintergrund.

**Satz 5.3.** (a) Es sei  $E := C_b(R)$  mit der  $L_2^{\mathfrak{C}}(R, \mathcal{R}, P^X)$ -Norm versehen, wobei  $R := \text{supp}(P^X)$  der Träger des Wahrscheinlichkeitsmaßes  $P^X$  auf  $(\mathbb{R}^d, \mathcal{B}^d)$  sei. Dann sind die Räume  $E_N := L_2^{\mathfrak{C}}(X_N, \mathcal{P}(X_N), P_N^X)$  mit  $X_N := \{x_1, \dots, x_N\}$  für fast jede Folge  $(x_n)_n$  von Beobachtungen unabhängiger Zufallsvariablen auf  $\mathbb{R}^d$  mit Verteilung  $P^X$  eine diskrete Approximation von  $E$

$$\mathcal{A} \left( E, \prod_N E_N, Q \right).$$

Das Maß  $P_N^X$  sei das empirische Maß der Beobachtungen  $x_1, \dots, x_N$ . Die Abbildung  $Q$  von  $E$  in den Raum der Äquivalenzklassen auf  $\prod_N E_N$  sei definiert als

$$(u_N)_N \in Q(u) \quad \iff \quad \frac{1}{N} \sum_{n=1}^N |u_N(x_n) - u(x_n)|^2 = \int |u_N - u|^2 dP_N^X \xrightarrow{N \rightarrow \infty} 0$$

für  $u \in E$ .

(b) Es sei  $((m_i, n_i))_i$  eine Folge von Paaren wachsender Stichprobenumfänge, für die  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$  für  $i \rightarrow \infty$ . Es sei  $E := C_b(R)$  mit der  $L_2^{\mathfrak{C}}(R, \mathcal{R}, P^Z)$ -Norm versehen, wobei  $P^Z := \rho P^X + (1 - \rho) P^Y$ ,  $R := \text{supp}(P^X) \cup \text{supp}(P^Y)$  und  $P^X, P^Y$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^d, \mathcal{B}^d)$  sind. Dann sind die Räume  $E_i := L_2^{\mathfrak{C}}(X_i, \mathcal{P}(X_i), P_i^Z)$  mit  $X_i := \{x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}\}$  für fast jedes Paar von Folgen  $(x_n)_n, (y_n)_n$  von Beobachtungen unabhängiger Zufallsvariablen auf  $\mathbb{R}^d$  mit Verteilung  $P^X$  bzw.  $P^Y$  eine diskrete Approximation von  $E$

$$\mathcal{A} \left( E, \prod_i E_i, Q \right).$$

Das Maß  $P_i^Z$  sei das empirische Maß der Beobachtungen  $x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}$ . Die Abbildung  $Q$  von  $E$  in den Raum der Äquivalenzklassen auf  $\prod_i E_i$  sei definiert als

$$\begin{aligned} (u_i)_i \in Q(u) \quad \iff \quad & \frac{1}{m_i + n_i} \left( \sum_{k=1}^{m_i} |u_i(x_k) - u(x_k)|^2 + \sum_{k=1}^{n_i} |u_i(y_k) - u(y_k)|^2 \right) \\ & = \int |u_i - u|^2 dP_i^Z = \|u_i - u\|_i^2 \xrightarrow{i \rightarrow \infty} 0 \end{aligned}$$

für  $u \in E$ .

Für die Integration von  $u \in E$  bezüglich  $P_N^X$  bzw.  $P_i^Z$  wähle man die Restriktion von  $u$  auf  $X_N$  bzw.  $X_i$ .

**Beweis.** (b) Abkürzend seien die jeweiligen  $L_2^c$ -Normen auf  $E_i$  mit  $\|\cdot\|_i$  und auf  $E$  mit  $\|\cdot\|$  bezeichnet. Mit  $\|u\|_i$ ,  $u \in E$ , sei die  $\|\cdot\|_i$ -Norm der Restriktion von  $u$  auf  $X_i$  gemeint. Für fast jede Beobachtungsfolge liegt für  $i \rightarrow \infty$  schwache Konvergenz von  $P_i^Z$  gegen  $P^Z$  vor ( $P_i^Z \implies P^Z$ ), vergleiche Gänsler und Stute [GS77, Satz 3.2.2, S.145]. Sei eine solche Beobachtungsfolge gegeben. Mit Hilfe der Dreiecksungleichung zeigt man leicht, dass  $Q(u)$ ,  $u \in E$ , das Bild von  $u$  unter  $Q$ , aus höchstens einer Äquivalenzklasse besteht. Um zu zeigen, dass  $Q(u)$  nicht-leer ist, wähle man  $u_i$  als Restriktion von  $u$  auf  $X_i$ . Damit ist die Bedingung (i) der Definition 5.2 gezeigt. Für Bedingung (ii) müssen die Normen  $\|u_i\|_i$  von  $(u_i)_i \in Q(u)$  gegen die Norm  $\|u\|$  von  $u$  konvergieren. Mit der Dreiecksungleichung gilt

$$\begin{aligned} \left| \|u_i\|_i - \|u\| \right| &\leq \left| \|u_i\|_i - \|u\|_i \right| + \left| \|u\|_i - \|u\| \right| \\ &\leq \|u_i - u\|_i + \left| \|u\|_i - \|u\| \right|. \end{aligned}$$

Da  $|u|^2$  eine stetige und beschränkte Funktion ist, gilt wegen der Verteilungskonvergenz von  $P_i^Z$  gegen  $P^Z$

$$\|u\|_i \xrightarrow{i \rightarrow \infty} \|u\|.$$

Mit der Definition von  $(u_i)_i \in Q(u)$ ,

$$\|u_i - u\|_i \xrightarrow{i \rightarrow \infty} 0,$$

folgt (ii). Die Linearität (iii) folgt aus der Definition

$$\begin{aligned} \|\alpha u_i + \beta v_i - w_i\|_i &\leq |\alpha| \|u_i - u\|_i + |\beta| \|v_i - v\|_i + \|w_i - (\alpha u + \beta v)\|_i \\ &\xrightarrow{i \rightarrow \infty} 0. \end{aligned}$$

Der Beweis von (a) folgt analog dem von Teil (b). □

## 5.2 Diskrete Kompaktheit und Konsistenz

Im Folgenden sei  $\mathcal{A}(E, \prod_\iota E_\iota, Q)$  eine diskrete Approximation von  $E$ . Hat man eine solche Approximation des interessierenden Raumes  $E$  gefunden, so kann man zur Betrachtung von Operatoren auf diesen Räumen übergehen. Die Operatoren auf den diskreten approximierenden Räumen werden mit dem Operator, dessen Eigenwerte gesucht sind, durch Erfüllen einer „Konsistenzbedingung“ verknüpft.

**Definition 5.4.** Einen beschränkten Operator  $B : E \rightarrow E$  und eine Folge von beschränkten Operatoren  $(B_\iota)_{\iota \in I}$  mit  $B_\iota : E_\iota \rightarrow E_\iota$  nennt man **konsistent**, wenn es eine dichte Teilmenge  $F \subset E$  gibt, so dass für jedes  $f \in F$  eine diskret konvergente Folge  $(f_\iota)_{\iota \in I}$ ,  $f_\iota \in E_\iota$ , existiert mit

$$f_\iota \rightarrow f \quad \text{und} \quad B_\iota f_\iota \rightarrow Bf.$$

Stummel approximierete Eigenwertprobleme der Art

$$Au = \lambda Bu$$

durch „diskrete“ Eigenwertprobleme

$$A_\iota u_\iota = \lambda^\iota B_\iota u_\iota.$$

Für den Nachweis der Konvergenz der Eigenwerte werden Projektionsoperatoren  $P : E \rightarrow E$ ,  $P_\iota : E_\iota \rightarrow E_\iota$  betrachtet, welche mit Hilfe von Integralen über Resolventen  $A(z)^{-1} := (A - zB)^{-1}$  definiert werden.

Für die diskrete Konvergenz der Integrale der Resolventen ist die diskrete Kompaktheit von  $B, B_\iota$  entscheidend.

**Definition 5.5.** Sei  $(B_\iota)_{\iota \in I}$  eine Folge beschränkter Operatoren  $B_\iota : E_\iota \rightarrow E_\iota$ . Diese Folge nennt man **diskret-kompakt**, wenn es zu jeder  $\|\cdot\|_{E_\iota}$ -beschränkten Teilfolge  $(u_\iota)_{\iota \in I'}$  von  $(u_\iota)_{\iota \in I}$ ,

$$\|u_\iota\|_{E_\iota} \leq \gamma \in (0, \infty) \quad \text{für alle } \iota \in I',$$

eine Teil-Folge  $(u_\iota)_{\iota \in I''}$  und ein  $w \in E$  gibt, so dass

$$B_\iota u_\iota \xrightarrow[\iota \in I'', \iota \rightarrow \infty]{} w.$$

Es folgt der Nachweis dieser beiden Eigenschaften für die zu Beginn dieses Kapitels beschriebenen Operatoren. Anwendungen, zum einen auf die in dieser Arbeit vorgestellte Klasse von V-Statistiken und zum anderen auf die Grenzverteilung der Cramér-von Mises-Testgröße, folgen in den Abschnitten 5.2.1 und 5.2.2.

**Anwendung auf  $L_2^{\mathbb{C}}(R, \mathcal{R}, P^Z)$ -Integraloperatoren** Wir betrachten den Zwei-Stichproben-Fall, sind also in der Situation von Satz 5.3, Teil (b) und haben auf den dort angegebenen  $L_2^{\mathbb{C}}$ -Räumen die Operatorenfolge  $(H_i)_i$  mit

$$H_i : \begin{array}{ccc} L_2^{\mathbb{C}}(X_i, \mathcal{P}(X_i), P_i^Z) & \rightarrow & L_2^{\mathbb{C}}(X_i, \mathcal{P}(X_i), P_i^Z) \\ f & \mapsto & \int h_i(\cdot, y) f(y) dP_i^Z(y) \end{array} \quad (5.3)$$

gegeben und als Operator  $H$ , dessen Eigenwerte wir approximieren wollen,

$$H : \begin{array}{l} E \rightarrow E \\ f \mapsto \int h(\cdot, y)f(y) dP^Z(y), \end{array}$$

wobei  $P^Z := \rho P^X + (1 - \rho)P^Y$  und  $R := \text{supp}(P^X) \cup \text{supp}(P^Y)$ . Der Raum  $E$  ist der Raum  $C_b(R)$  der beschränkten, stetigen, komplexwertigen Funktionen auf  $R$ , versehen mit der  $L_2^{\mathbb{C}}(R, \mathcal{R}, P^Z)$ -Norm. Der Operator  $H$  lässt sich auch als Operator auf dem Hilbertraum  $L_2^{\mathbb{C}}(R, \mathcal{R}, P^Z)$  verstehen. Beide Versionen werden unter den Voraussetzungen, die wir machen werden, dieselben stetigen Eigenfunktionen zu denselben positiven Eigenwerten liefern und seien in der Folge beide mit  $H$  bezeichnet, wobei sich die Interpretation aus dem Kontext ergibt.

Wir werden nun nachweisen, dass Operatoren dieser Bauart unter bestimmten Bedingungen *konsistent* und *diskret-kompakt* sind.

**Lemma 5.6.** *Es seien die Voraussetzungen von Satz 5.3, Teil (b) erfüllt und es seien Beobachtungsfolgen gewählt, die die Räume  $(E_i)_i$ ,  $i \in I$  zu einer diskreten Approximation von  $E$  machen, d.h. es möge gelten*

$$P_i^Z \xrightarrow{i \in I, i \rightarrow \infty} P^Z.$$

*Es sei  $(h_i)_i$  eine Folge von in der Supremumsnorm gleichmäßig beschränkten, in ihren Argumenten symmetrischen Funktionen*

$$h_i : (R \times R, \mathcal{R} \otimes \mathcal{R}) \rightarrow (\mathbb{R}, \mathcal{B}).$$

*Es gebe für jedes  $\varepsilon > 0$  ein  $\delta > 0$ , so dass für jedes  $i \in I$  gilt*

$$\sup\{|h_i(x, y) - h_i(x', y)| : |x - x'| < \delta, x, x' \in R, y \in R\} < \varepsilon. \quad (5.4)$$

*Dann ist die oben definierte Folge der Operatoren  $(H_i)_i$  diskret-kompakt.*

Da wir fordern, dass die Funktionen  $h_i$  auf ganz  $R \times R$  definiert sind, lassen sich auch für  $x \in R \setminus X_i$  Funktionswerte von  $(H_i u_i)(x)$  mit  $u_i \in L_2^{\mathbb{C}}(X_i, \mathcal{P}(X_i), P_i^Z)$  durch (5.3) angeben:

$$(H_i u_i)(x) = \int h_i(x, y)u_i(y) dP_i^Z(y), \quad x \in R.$$

**Beweis.** Abkürzend sei für diesen Beweis  $\mu := P^Z$  und  $\mu_i := P_i^Z$  für  $i \in I$ . Wegen der Beschränktheit der Kernfunktion,  $\|h_i\|_{\infty} < c$  (für alle  $i \in I$ ), und der Integration derselben über ein endliches Maß, handelt es sich bei  $H_i$  um einen beschränkten Operator:

$$\|H_i u_i\|_i = \left( \int \left| \int h_i(x, y)u_i(y) d\mu_i(y) \right|^2 d\mu_i(x) \right)^{\frac{1}{2}} \leq c \|u_i\|_i, \quad u_i \in E_i.$$

Es sei  $(R_k)_k$ ,  $k \in \mathbb{N}$  eine Folge von aufsteigenden Kompakta der Form  $R_k := R \cap [-\kappa, +\kappa]^d$  für ein  $\kappa > k$  mit  $\mu(R \setminus R_k) \leq \frac{1}{2^k}$  und  $\mu_i(R \setminus R_k) \leq \frac{1}{2^k}$  für alle  $i \in I$ . Die Gleichmäßigkeit in dieser Aussage ist wegen der Verteilungskonvergenz von  $\mu_i$  gegen  $P^Z$  möglich.

Sei nun  $(u_i)_{i \in I'}$  eine beliebige Teilfolge der Folge  $(u_i)_{i \in I}$ , welche durch  $\gamma$  beschränkt sei,

$$\|u_i\|_i \leq \gamma \quad \text{für alle } i \in I'.$$

Der Beweis ist in drei Abschnitte aufgeteilt:

1. Zeige, dass die Funktionenfolge  $(g_i)_i$  mit

$$g_i(x) := (H_i u_i)(x), \quad x \in R,$$

gleichgradig stetig und gleichmäßig beschränkt in der Supremumsnorm ist.

2. Suche eine Teilfolge  $(u_i)_{i \in I''}$  von  $(u_i)_{i \in I'}$  dergestalt, dass die Funktionenfolge  $(g_i)_{i \in I''}$  gleichmäßig auf  $R_k$  gegen eine stetige Funktion  $w_k$  konvergiert.
3. Konstruktion einer Teilfolge  $(u_i)_{i \in I''}$ , so dass  $(g_i)_{i \in I''}$  auf ganz  $R$  gegen eine stetige und beschränkte Funktion  $w$  konvergiert, die auf jedem Kompaktum  $R_k$  mit der Funktion  $w_k$  übereinstimmt.

In Schritt 1 zeigen wir zuerst, dass  $(g_i)_i$  eine gleichgradig stetige Folge von Funktionen ist. Für ein gegebenes  $\varepsilon > 0$  sind wir nach Voraussetzung in der Lage ein  $\delta > 0$  anzugeben, so dass für  $x, x' \in R$  mit  $|x - x'| < \delta$  und jedes  $y \in R$  gilt

$$|h_i(x, y) - h_i(x', y)| < \varepsilon \quad \text{für alle } i.$$

Mit Hilfe der Cauchy-Schwarzschen Ungleichung erhalten wir damit

$$\begin{aligned} |g_i(x) - g_i(x')| &= \left| \int [h_i(x, y) - h_i(x', y)] u_i(y) \, d\mu_i(y) \right| \\ &\leq \gamma \left( \int [h_i(x, y) - h_i(x', y)]^2 \, d\mu_i(y) \right)^{\frac{1}{2}} < \gamma \varepsilon. \end{aligned}$$

Also ist die Folge der  $(g_i)_i$  eine gleichgradig stetige Folge auf ganz  $R$ . Die Beschränktheit folgt ebenfalls aus der Cauchy-Schwarzschen Ungleichung: Sei  $x \in R$ , dann gilt

$$|g_i(x)| = \left| \int h_i(x, y) u_i(y) \, d\mu_i(y) \right| \leq \gamma \left( \int h_i^2(x, y) \, d\mu_i(y) \right)^{\frac{1}{2}} \leq \gamma c,$$

womit die gleichmäßige Beschränktheit der Folge  $(g_i)_i$  folgt.

Wir betrachten nun für  $k \in \mathbb{N}$  im Schritt 2 das Verhalten der Funktionen  $(g_i)_{i \in I'}$  auf der kompakten Menge  $R_k$ . Dazu benutzen wir den Satz von Arzela und Ascoli.

Mit dem Ergebnis von Schritt 1 garantiert er die Existenz einer Teilfolge  $(g_i)_{i \in I''_k}$  von  $(g_i)_{i \in I'}$  und einer stetigen, komplexwertigen Funktion  $w_k \in C_b(R_k)$ , so dass

$$\max_{x \in R_k} |g_i(x) - w_k(x)| \xrightarrow{i \in I''_k, i \rightarrow \infty} 0.$$

Durch sukzessive Wahl der  $(I''_k)_k$  kann erreicht werden, dass  $I''_{k+1} \subset I''_k$  für alle  $k \in \mathbb{N}$ .

Im letzten Schritt, Schritt 3, wird eine stetige Funktion  $w$  konstruiert, so dass schließlich  $\|H_i u_i - w\|_i$  für eine Teilfolge  $(u_i)_{i \in I''}$  von  $(u_i)_{i \in I'}$  gegen 0 konvergiert. Sei  $(u_i)_{i \in I''}$  die Diagonalfolge in den Folgen  $(u_i)_{i \in I''_k}$ ,  $k \in \mathbb{N}$ , d.h. das  $k$ -te Element der Folge  $(u_i)_{i \in I''}$  entspricht dem  $k$ -ten Element der Folge  $(u_i)_{i \in I''_k}$ . Die Grenzfunktion  $w$  wird definiert

$$w(x) := \lim_{i \in I'', i \rightarrow \infty} g_i(x) = \lim_{i \in I'', i \rightarrow \infty} H_i u_i(x), \quad x \in R.$$

Es bleiben folgende Fragen zu klären:

- (i) *Ist  $w$  wohldefiniert und eine auf ganz  $R$  in der  $\|\cdot\|_\infty$ -Norm stetige Funktion?*  
 Sei  $x \in R$ , dann ist  $x \in R_k$  für alle  $k \geq k_0 \in \mathbb{N}$ . Da  $(u_i)_{i \in I'', i \geq i_0(k)}$  eine Teilfolge von  $(u_i)_{i \in I''_k}$  ist, folgt die Konvergenz von  $H_i u_i(x)$  gegen ein eindeutig bestimmtes  $w(x)$ . Da  $w$  auf jedem Kompaktum  $R_k$  mit der stetigen Funktion  $w_k$  übereinstimmt und die  $R_k$  ganz  $R$  ausschöpfen, ist  $w$  stetig.
- (ii) *Ist  $w$  eine beschränkte Funktion?*  
 Dies ist der Fall, da die  $(g_i)_{i \in I'}$  eine gleichmäßig beschränkte Folge sind. Damit sind die Funktionen  $(w_k)_{k \in \mathbb{N}}$  aus Schritt 2 gleichmäßig beschränkt und somit die Funktion  $w$  beschränkt.

Es folgt die Konvergenz von  $(H_i u_i)_{i \in I''}$  in der  $L_2^{\mathbb{C}}$ -Norm, denn für jedes  $k \in \mathbb{N}$  und  $i \in I''$  gilt

$$\|H_i u_i - w\|_i^2 = \underbrace{\int_{R \setminus R_k} |H_i u_i(x) - w(x)|^2 d\mu_i(x)}_{\leq (2\gamma c)^2 \mu_i(R \setminus R_k) \leq (2\gamma c)^2 2^{-k}} + \underbrace{\int_{R_k} |H_i u_i(x) - w_k(x)|^2 d\mu_i(x)}_{\xrightarrow{i \in I'', i \rightarrow \infty} 0}$$

Mit  $k \rightarrow \infty$  folgt die diskrete Kompaktheit. □

Das vorangegangene Lemma ist auch im Ein-Stichproben-Fall, Satz 5.3, Teil (a), gültig.

Die anschließende Konsistenzbetrachtung stützt sich im Wesentlichen auf die Überlegungen zur diskreten Kompaktheit.

**Lemma 5.7.** *Es seien die Voraussetzungen des vorangegangenen Lemmas erfüllt. Für jedes  $x \in R$  möge darüber hinaus die Folge von Funktionen  $(h_i(x, \cdot))_i$  gleichmäßig auf jedem Kompaktum gegen die in ihren Argumenten symmetrische Funktion  $h(x, \cdot)$  konvergieren, wobei die Kernfunktion*

$$h : (R \times R, \mathcal{R} \otimes \mathcal{R}) \rightarrow (\mathbb{R}, \mathcal{B})$$

*eine in der Supremumsnorm beschränkte Funktion sei mit der Eigenschaft, dass es für jedes  $\varepsilon > 0$  ein  $\delta > 0$  gibt, so dass*

$$\sup\{|h(x, y) - h(x', y)| : |x - x'| < \delta, x, x' \in R, y \in R\} < \varepsilon. \quad (5.5)$$

*Dann sind der Operator  $H$  und die Folge von Operatoren  $(H_i)_i$  konsistent. Ferner ist  $H$  ein kompakter Operator in  $C_b(R)$  versehen mit der  $L_2^c(R, \mathcal{R}, P^Z)$ -Norm.*

**Beweis.** Es seien die Bezeichnungen des Beweises des letzten Lemmas übernommen. Analog zum dortigen Vorgehen zeigt man die Beschränktheit des Operators  $H$ . Zum Nachweis der Konsistenz zeigen wir, dass zu jeder Funktion  $u \in E = C_b(R)$  eine Folge von Funktionen  $(u_i)_i$  mit  $u_i \rightarrow u$  und  $H_i u_i \rightarrow Hu$  existiert. Dazu betrachten wir als Funktionen  $u_i$  die Restriktion von  $u$  auf das jeweilige  $X_i$ . Die Folge  $(u_i)_i$  ist  $\|\cdot\|_i$ -beschränkt. Im letzten Beweis haben wir gesehen, dass  $(g_i)_i$  mit  $g_i := H_i u_i$  für jedes  $k \in \mathbb{N}$  eine gleichmäßig beschränkte, gleichgradig stetige Folge von Funktionen auf  $R_k$  ist. Es gilt für  $x \in R$

$$\begin{aligned} g_i(x) &:= \int h_i(x, y) u_i(y) \, d\mu_i(y) \\ &= \int h_i(x, y) u(y) \, d\mu_i(y) - \int h(x, y) u(y) \, d\mu_i(y) \end{aligned} \quad (5.6)$$

$$+ \int h(x, y) u(y) \, d\mu_i(y) \quad (5.7)$$

$$\xrightarrow{i \rightarrow \infty} \int h(x, y) u(y) \, d\mu(y) = Hu(x).$$

Der Ausdruck (5.6) konvergiert für  $i \rightarrow \infty$  gegen 0, da der Integrationsbereich aufgeteilt werden kann in das Kompaktum  $R_k$ , in dem die Integranden nach Voraussetzung gleichmäßig gegeneinander konvergieren, und den Bereich außerhalb des Kompaktums  $R \setminus R_k$ , dessen  $\mu_i$ -Inhalt für  $k \rightarrow \infty$  gleichmäßig in  $i$  gegen 0 geht und somit das Integral über die beschränkten Funktionen  $|h_i(x, \cdot)u(\cdot)|$  und  $|h(x, \cdot)u(\cdot)|$  gegen Null treibt. Zum Nachweis der Konvergenz (5.7) genügt die in Lemma 5.6 geforderte Verteilungskonvergenz. Aus der punktweisen Konvergenz folgt zusammen mit der gleichgradigen Stetigkeit die gleichmäßige Konvergenz auf Kompakta (allerdings nicht nur für Teilfolgen, da der Satz von Arzela und Ascoli hier nicht notwendig ist), womit, wie im Beweis des letzten Lemmas, die diskrete Konvergenz von  $(g_i)_i$  gegen die beschränkte, stetige Funktion  $w = Hu$  folgt.

Es folgt der Nachweis der Kompaktheit von  $H$  in der durch die  $L_2^{\mathbb{C}}$ -Norm induzierten Topologie auf  $C_b(R)$ . Die Definition der Kompaktheit linearer, beschränkter Operatoren entspricht auch auf normierten Räumen der in Definition C.1 angegebenen. Offensichtlich ist  $H$  beschränkt und linear. Es bleibt zu zeigen (vgl. Dunford und Schwartz [DS88a, Theorem 15, S.22]), dass für eine Folge  $(f_n)_n$  von in ihrer Norm durch 1 beschränkten Funktionen  $f_n \in C_b(R)$  die Bildfolge  $(Hf_n)_n$  einen Häufungspunkt  $f \in C_b(R)$  besitzt. Wie im Beweis der diskreten Kompaktheit, Lemma 5.6, Schritt 1, zeigt man die gleichgradige Stetigkeit und gleichmäßige Beschränktheit der Folge  $(Hf_n)_n$  in der Supremumsnorm. Der Satz von Arzela und Ascoli (vgl. Schritt 2) liefert mit dem Diagonalfolgenargument (vgl. Schritt 3) eine Teilfolge  $(Hf_n)_{n \in I''}$  der Folge  $(Hf_n)_n$ , die auf jedem Kompaktum  $R_k$  (definiert wie im letzten Beweis) gleichmäßig gegen eine stetige und beschränkte Funktion  $f$  konvergiert. Mit  $k \rightarrow \infty$  folgt aus

$$\|Hf_n - f\|^2 = \underbrace{\int_{R \setminus R_k} |Hf_n(x) - f(x)|^2 d\mu(x)}_{\leq (2c)^2 \mu(R \setminus R_k) \leq (2c)^2 2^{-k}} + \underbrace{\int_{R_k} |Hf_n(x) - f(x)|^2 d\mu(x)}_{\xrightarrow{n \in I'', n \rightarrow \infty} 0}$$

wobei  $\|h\|_{\infty} < c$ , die Kompaktheit von  $H$ .  $\square$

Auch der Nachweis der Konsistenz und Kompaktheit lässt sich direkt auf den Ein-Stichproben-Fall übertragen.

### 5.2.1 Nachweis für die neue Familie von Tests

Als erste Anwendung betrachten wir den Integraloperator (3.9),

$$H : \begin{array}{l} E \rightarrow E \\ f \mapsto \int h(\cdot, y) f(y) dP^Z(y), \end{array}$$

mit  $h(x_1, x_2) := \int \bar{h}(x_1, y_1; x_2, y_2) d(P^Z \otimes P^Z)(y_1, y_2)$  und  $E := C_b(R)$  versehen mit der  $L_2^{\mathbb{C}}(R, \mathcal{R}, P^Z)$ -Norm, dessen Eigenwerte sich in der Grenzverteilung der neuen Teststatistiken unter der Hypothese wiederfinden. Wir approximieren diesen Operator, wie in Abschnitt 4.2 beschrieben, durch die Integraloperatoren

$$H_i : \begin{array}{l} E_i \rightarrow E_i \\ f \mapsto \int h_i(\cdot, y) f(y) dP_i^Z(y) \end{array}$$

mit  $h_i(x_1, x_2) := \int \bar{h}(x_1, y_1; x_2, y_2) d(P_i^Z \otimes P_i^Z)(y_1, y_2)$  und  $E_i := L_2^{\mathbb{C}}(X_i, \mathcal{P}(X_i), P_i^Z)$ . Dabei gehen wir wie folgt vor: Zunächst wählen wir eine feste Folge von Stichprobenumfängen. Dann wählen wir Folgen von Beobachtungen der beiden Stichproben,

so dass die Räume  $E_i$  aus Satz 5.3 (b) eine diskrete Approximation des dort angegebenen Raumes  $E$  bilden. Aus diesen Beobachtungen bilden wir nun die Folge von Maßen  $(P_i^Z)_i$  und damit die Operatorenfolge  $(H_i)_i$ . Von dieser Folge weisen wir nach, dass sie diskret-kompakt und konsistent ist. Dazu müssen lediglich die Voraussetzungen des Lemmas 5.6 und des Lemmas 5.7 nachgewiesen werden.

**Satz 5.8.** *Es sei  $\phi \in \Phi$  eine beschränkte Funktion. Für fast jedes Paar von Beobachtungsfolgen bilden dann die Operatoren  $(H_i)_i$  mit den angegebenen Kernfunktionen eine diskret-kompakte Folge, die mit dem in der  $L_2^{\mathbb{C}}(R, \mathcal{R}, P^Z)$ -Topologie auf  $C_b(R)$  kompakten Operator  $H$  konsistent ist.*

**Beweis.** Offensichtlich sind die Kernfunktionen von  $H_i$  und  $H$  gleichmäßig beschränkt. Zunächst zeigen wir die Forderung (5.4). Es sei  $\phi(z) \leq C$  für alle  $z \geq 0$ . Da  $d_\phi$  eine Metrik ist, folgt durch Einsetzen von Dirac-Maßen

$$\sqrt{\phi(|x - y|^2)} \leq \sqrt{\phi(|x - x'|^2)} + \sqrt{\phi(|x' - y|^2)} \quad \text{für alle } x, x', y \in \mathbb{R}^d$$

und somit

$$\phi(|x - y|^2) \leq \phi(|x - x'|^2) + \phi(|x' - y|^2) + 2\sqrt{\phi(|x - x'|^2)}\sqrt{\phi(|x' - y|^2)}$$

also

$$|\phi(|x - y|^2) - \phi(|x' - y|^2)| \leq 3\sqrt{C}\sqrt{\phi(|x - x'|^2)}.$$

Sei  $y \in R$  und  $x, x' \in R$  so, dass  $|x - x'| < \delta$ . Dann gilt mit  $X^{(i)} \sim P_i^Z$

$$\begin{aligned} & |h_i(x, y) - h_i(x', y)| \\ &= |\mathbb{E}\phi(|x - X^{(i)}|^2) - \mathbb{E}\phi(|x' - X^{(i)}|^2) - \phi(|x - y|^2) + \phi(|x' - y|^2)| \\ &\leq (3\sqrt{C} + 3\sqrt{C})\sqrt{\phi(|x - x'|^2)}. \end{aligned}$$

Für ein gegebenes  $\varepsilon > 0$  wähle also  $\delta$  so, dass  $\sqrt{\phi(\delta^2)}/(6\sqrt{C}) \leq \varepsilon$  ist. Dies ist wegen der Stetigkeit von  $\phi$  und  $\phi(0) = 0$  möglich. Die Bedingung (5.4) und die Stetigkeitsbedingung in Lemma 5.7 sind also erfüllt.

Zum Nachweis der Konsistenz sei  $x \in R$  fest gewählt. Dann konvergiert  $(h_i(x, y))_i$  wegen der Verteilungskonvergenz von  $(P_i^Z)_i$  für alle  $y \in R$  gegen  $h(x, y)$ . Wir benötigen statt dieser punktwisen Konvergenz die gleichmäßige Konvergenz. Da gilt:  $(h_i(x, y))_i$  konvergiert gegen  $h(x, y)$  für alle  $y \in R \cap \mathbb{Q}^d$  wird mit der gleichgradigen Stetigkeit (5.4) der Folge  $(h_i(x, \cdot))_i$  (Symmetrie der Kernfunktion) aus der punktwisen Konvergenz auf einer dichten Teilmenge die gleichmäßige Konvergenz auf einem Kompaktum.

Damit sind die Voraussetzungen der Lemmata 5.6 und 5.7 erfüllt und es folgt die Behauptung.  $\square$

Der vorangegangene Satz lässt sich analog auf den Ein-Stichproben-Fall, Satz 5.3, Teil (a), übertragen. Dies ist hilfreich, wenn man beispielsweise die Gewichte der Grenzverteilung der Ein-Stichproben-Statistik aus Satz 3.8 approximieren möchte.

### 5.2.2 Nachweis für den Cramér-von Mises-Test

Als weiteres Beispiel für die Anwendung der Theorie von Stummel betrachten wir die wohlbekannte Cramér-von Mises-Statistik für den Anpassungstest

$$\omega_n^2 = n \int_{-\infty}^{+\infty} [F_n(x) - F_0(x)]^2 dF_0(x),$$

wobei  $F_n$  die empirische Verteilungsfunktion der  $n$  u.i.v. Zufallsvariablen  $X_1, \dots, X_n$  sei und  $F_0$  eine stetige Verteilungsfunktion, auf die getestet werden soll. Diese Statistik lässt sich als V-Statistik

$$\omega_n^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n h_{\text{CvM}}(X_i, X_j)$$

mit Kernfunktion

$$h_{\text{CvM}}(x, y) = \int_{-\infty}^{+\infty} [\mathbb{I}(x \leq t) - F_0(t)] [\mathbb{I}(y \leq t) - F_0(t)] dF_0(t), \quad x, y \in R, \quad (5.8)$$

und  $R = \text{supp}(F_0)$  schreiben. Im Fall, dass  $F_0$  die Gleichverteilung auf  $R = [0, 1]$  ist, ergibt sich beispielsweise die Kernfunktion

$$h_{\text{CvM}}(x, y) = \frac{1}{3} - \max(x, y) + \frac{1}{2}x^2 + \frac{1}{2}y^2, \quad x, y \in R.$$

Die Grenzverteilung der Statistik unter der Hypothese  $X_1, X_2, \dots \sim F_0$  ist bekanntlich eine gewichtete Summe von Quadraten standardnormalverteilter Zufallsvariablen. Die Gewichte sind die Eigenwerte des Operators

$$H^{\text{CvM}} : \begin{array}{ll} C_b(R) & \rightarrow C_b(R) \\ f & \mapsto \int h_{\text{CvM}}(\cdot, y) f(y) dF_0(y), \end{array}$$

Als Beispiel soll gezeigt werden, dass die Eigenwerte der (diskreten) Integraloperatoren, bei denen die Verteilung  $F_0$  durch die empirische Verteilungsfunktion einer vorgegebenen Anzahl  $N$  von Beobachtungen ersetzt ist, die Eigenwerte der Grenzverteilung unter der Hypothese approximieren.

Satz 5.3, Teil (a) lässt sich auf diese Problemstellung anwenden. Die approximierenden Operatoren definieren wir durch

$$H_N^{\text{CvM}} : \begin{array}{ll} L_2^{\mathbb{C}}(X_N, \mathcal{P}(X_N), P_N^X) & \rightarrow L_2^{\mathbb{C}}(X_N, \mathcal{P}(X_N), P_N^X) \\ f & \mapsto \int h_{\text{CvM}}(\cdot, y) f(y) dP_N^X(y). \end{array}$$

Wir zeigen die Konsistenz und diskrete Kompaktheit der Folge der Operatoren  $(H_N^{\text{CvM}})_N$ , um später in Kapitel 5.3.2 auf die Konvergenz der Eigenwerte eingehen zu können.

**Satz 5.9.** Für fast jede Folge von Beobachtungen  $(x_n)_n$  von u.i.v. reellwertigen Zufallsvariablen mit stetiger Verteilungsfunktion  $F_0$  sind die Operatoren  $(H_N^{\text{CvM}})_N$  diskret-kompakt,  $(H_N^{\text{CvM}})_N, H^{\text{CvM}}$  konsistent und der Operator  $H^{\text{CvM}}$  kompakt in der  $L_2^{\mathbb{C}}(R, \mathcal{R}, F_0)$ -Topologie auf  $C_b(R)$ .

**Beweis.** Da die Kernfunktionen aller Operatoren  $(H_N^{\text{CvM}})_N, H^{\text{CvM}}$  übereinstimmen, ist Bedingung (5.4) hier leicht zu zeigen. Seien  $x, x', y \in R$  mit  $R := \text{supp}(P^Z)$  und  $|x - x'| < \delta, x < x'$ , dann gilt

$$|h_{\text{CvM}}(x, y) - h_{\text{CvM}}(x', y)| \leq \int_{-\infty}^{+\infty} 2 \mathbf{I}(x \leq t < x') dF_0(t) = 2(F_0(x') - F_0(x)),$$

womit die Bedingung erfüllt ist (stetige Verteilungsfunktionen sind gleichmäßig stetig). Überdies ist die Kernfunktion gleichmäßig beschränkt.

Damit sind die Voraussetzungen der Lemmata 5.6 und 5.7 erfüllt und es folgt die Behauptung.  $\square$

### 5.3 Konvergenz der Eigenwerte

Es sei das Eigenwertproblem  $Au = \lambda Bu, \lambda \in \mathbb{C}, u \in E$ , vorgelegt. Mit den beiden Voraussetzungen Konsistenz und diskrete Kompaktheit kann man den Hauptsatz über die Eigenwertkonvergenz diskreter Operatoren von Stummel [Stu71, II.3.2.(8), S.257] verwenden:

**Satz 5.10.** Die Operatoren  $A, (A_\iota)_{\iota \in I}$  sowie  $B, (B_\iota)_{\iota \in I}$  seien konsistent. Die Operatoren  $B, B_\iota, \iota \in I$ , seien kompakt und die Folge  $(B_\iota), \iota \in I$ , sei diskret-kompakt. Die Resolventenmenge und das Beschränktheitsgebiet von  $A, B$  seien nicht leer. Sei  $\lambda$  ein Eigenwert von  $A, B$  mit der algebraischen Vielfachheit  $m$ . Weiter sei  $U$  eine beliebige beschränkte abgeschlossene Umgebung von  $\lambda$ , welche mit dem Spektrum  $\Sigma(A, B)$  von  $A, B$  lediglich den Eigenwert  $\lambda$  gemeinsam hat. Dann gibt es ein  $\iota_0 \in \mathbb{N}$ , so dass für alle  $\iota \geq \iota_0, \iota \in I$ , genau  $m$  Eigenwerte  $\lambda_1^\iota, \dots, \lambda_m^\iota$  von  $A_\iota, B_\iota$  existieren, wenn diese ihrer algebraischen Vielfachheit entsprechend oft wiederholt werden,

$$\Sigma(A_\iota, B_\iota) \cap U = \{\lambda_1^\iota, \dots, \lambda_m^\iota\}, \quad \iota \in I, \quad \iota \geq \iota_0,$$

und

$$\lambda_k^\iota \xrightarrow{\iota \rightarrow \infty} \lambda, \quad k = 1, \dots, m.$$

Die beiden Begriffe Resolventenmenge und Beschränktheitsgebiet seien noch erklärt: Die **Resolventenmenge** ist die Menge aller Zahlen  $z \in \mathbb{C}$ , für die die Resolvente  $A(z)^{-1} = (A - zB)^{-1}$  als beschränkte lineare Abbildung existiert. Diese Menge ist das Komplement des Spektrums von  $A, B$ . Das **Beschränktheitsgebiet** ist die

Menge aller  $z \in \mathbb{C}$ , für die es eine Konstante  $\gamma > 0$  und ein  $\iota_0 \in I$  gibt, so dass für alle  $\iota \geq \iota_0$  die Zahl  $z$  in der Resolventenmenge von  $A_\iota, B_\iota$  liegt und die Operatornorm von  $A_\iota(z)^{-1}$  gleichmäßig in  $\iota \geq \iota_0$  durch  $\gamma$  beschränkt ist. Da wir für unsere Anwendungen als  $A$  bzw.  $A_\iota$  stets die Identität wählen werden, ist die Null sowohl in der Resolventenmenge als auch im Beschränktheitsgebiet enthalten, so dass die Erfüllung dieser Voraussetzung uns nicht mehr beschäftigen muss.

Unter den gemachten Voraussetzungen zeigt Stummel die Gleichheit von Resolventenmenge und Beschränktheitsgebiet sowie die Konvergenz der Resolventen [Stu71, II.2.2.(1), S.245]. Er zeigt, dass der Rand einer beliebigen, kompakten Umgebung eines Eigenwertes, welche nur diesen Eigenwert enthält, für  $\iota$  größer einem  $\iota_0$  stets vollständig in der Resolventenmenge von  $A_\iota, B_\iota$  liegt [Stu71, II.2.1.(3), S.243]. Damit sind die positiv orientierten Kurvenintegrale der Resolventen über diesen Rand, die so genannten Resolventenintegrale, wohldefiniert und konvergieren diskret [Stu71, II.2.3.(4), S.249]. Durch sie können Projektionsoperatoren  $P, P_\iota$  definiert werden, welche ebenfalls diskret konvergieren. Ein solcher Projektionsoperator ist eine Projektion des Ausgangsraumes auf den algebraischen Eigenraum des „umringten“ Eigenwertes. Als Konsequenz folgt die diskrete Approximation von  $P(E)$  durch  $P_\iota(E_\iota)$ , welche schließlich dieselbe Dimension wie  $P(E)$  haben. Damit müssen in der anfangs gewählten Umgebung schließlich  $m$  Eigenwerte von  $A_\iota, B_\iota$  gelegen haben. Durch immer kleinere Wahl der Umgebung um den Eigenwert  $\lambda$  folgt die Konvergenz. Auch die diskrete Konvergenz der Eigenfunktionen kann man auf diese Weise folgern.

**Anwendung auf  $L_2^{\mathbb{C}}(R, \mathcal{R}, P^Z)$ -Integraloperatoren** Für die Anwendung des Hauptsatzes auf die aus Abschnitt 5.2 bekannten Integraloperatoren definieren wir die Operatoren  $A, A_i, i \in \mathbb{N}$ , als Identität auf dem jeweiligen Raum  $E, E_i, i \in \mathbb{N}$ . Wählt man die Beobachtungsfolgen so, dass eine diskrete Approximation vorliegt, so sind diese Operatoren konsistent. Seien weiter  $B := H$  und  $B_i := H_i$ . Wir haben gezeigt, dass unter den Voraussetzungen der Lemmata 5.6 und 5.7 Konsistenz von  $(B_i)_i, B$  und diskrete Kompaktheit von  $(B_i)_i$  vorliegt. Weiterhin sind die Operatoren  $B_i$  Hilbert-Schmidt-Operatoren und somit kompakt in  $E_i$  (vergleiche Lemma C.6). Der Operator  $B$  ist nach Lemma 5.7 kompakt in  $E$ . Damit sind die Voraussetzungen des Hauptsatzes erfüllt. Man überlegt sich leicht, dass wegen der Stetigkeitsforderungen an  $h$  die Eigenfunktionen zu den nicht verschwindenden Eigenwerten des Operators  $B$  aufgefasst als Operator auf  $L_2^{\mathbb{C}}(R, \mathcal{R}, P^Z)$  stetig sind. Die Einschränkung von  $B$  auf  $E$  ändert also nichts an den von Null verschiedenen Eigenwerten und deren Vielfachheit. Da die Operatoren  $H_i, H$  selbstadjungiert sind, besitzen sie ausschließlich reelle Eigenwerte, zu denen reellwertige Eigenfunktionen existieren, so dass  $H_i, H$  auch als Operatoren über  $L_2(X_i, \mathcal{P}(X_i), P_i^Z)$  beziehungsweise dem Raum der reellwertigen, stetigen, beschränkten Funktionen, versehen mit der  $L_2(R, \mathcal{R}, P^Z)$ -Norm aufgefasst werden können, ohne dadurch die nicht verschwindenden Eigenwerte und deren Vielfachheit zu verändern.

Es sei bemerkt, dass die zu Beginn dieses Abschnittes beschriebene Definition der Eigenwerte von  $A_i, B_i$  genau die Inversen von den Eigenwerten liefert, die wir zuvor betrachtet haben. Denn hier betrachten wir das Eigenwertproblem

$$u = Au \stackrel{!}{=} \lambda Bu = \lambda Hu$$

und zuvor betrachteten wir

$$\lambda u = Hu.$$

Theoretisch wäre ein Austausch von  $A$  und  $B$  bei der Anwendung des Hauptsatzes denkbar, jedoch liegt die diskrete Kompaktheit bei der Identität nicht vor (Stetigkeitsprobleme), so dass die hier gewählte Herangehensweise erforderlich ist. Im Folgenden wollen wir wieder zu unserer „bewährten“ Definition eines Eigenwertes zurückkehren. Die Konvergenz der von Null verschiedenen Eigenwerte von  $A_i, B_i$  im Hauptsatz liefert natürlich auch die Konvergenz der inversen Eigenwerte, so dass uns dieses technische Detail nicht weiter belasten muss. Der Hauptsatz liefert mit dem eben gesagten folgendes Lemma:

**Lemma 5.11.** *Es seien die Voraussetzungen der Lemmata 5.6 und 5.7 erfüllt. Dann folgt für jeden von Null verschiedenen Eigenwert  $\lambda$  von  $H$  mit Vielfachheit  $m \in \mathbb{N}$  und jede hinreichend kleine Umgebung  $U$  um  $\lambda$ , dass für alle  $i$  größer als ein bestimmtes  $i_0$  der Operator  $H_i$  genau  $m$  Eigenwerte  $\lambda_1^i, \dots, \lambda_m^i$  (Vielfachheit beachtend) in  $U$  besitzt.*

Abschließendes Ziel dieses Abschnittes ist es nun, die Verteilungskonvergenz der gewichteten Summe von Quadraten standardnormalverteilter Zufallsvariablen mit den approximierten Eigenwerten als Gewichten gegen jene gewichtete Summe von Quadraten standardnormalverteilter Zufallsvariablen zu zeigen, deren Gewichte die Eigenwerte des Operators  $H$  sind.

Um dieses Resultat zu erhalten, fordern wir zusätzlich zu den Voraussetzungen des letzten Lemmas,

- (1.) dass die Operatoren  $(H_i)_i, H$  positiv sein mögen und
- (2.) dass der Operator  $H$  von der Spurklasse ist.

Unter diesen Voraussetzungen erhalten wir das gesuchte Ergebnis:

**Satz 5.12.** *Es seien die Voraussetzungen der Lemmata 5.6 und 5.7 sowie die genannten Voraussetzungen (1.) und (2.) erfüllt. Dann gilt*

$$S^i := \sum_{\sigma \in \Sigma_i} \lambda_\sigma^i Z_\sigma^2 \xrightarrow{i \rightarrow \infty} \sum_{\sigma \in \Sigma} \lambda_\sigma Z_\sigma^2 =: S,$$

wobei  $(Z_\sigma)_\sigma$  jeweils eine Folge unabhängiger, standardnormalverteilter Zufallsvariablen sei. Die Folgen  $(\lambda_\sigma^i)_\sigma$ ,  $\sigma \in \Sigma_i$ , bezeichnen die positiven Eigenwerte der Operatoren  $H_i$  und die Folge  $(\lambda_\sigma)_\sigma$ ,  $\sigma \in \Sigma \subset \mathbb{N}$ , beschreibe die positiven Eigenwerte des Operators  $H$ . In allen Folgen seien die Eigenwerte entsprechend ihrer Vielfachheit oft genannt.

**Beweis.** Die Operatoren  $(H_i)_i$  sind von der Spurklasse, da sie jeweils nur endlich viele Eigenwerte besitzen. Mittels (5.4) und (5.5) überlegt man sich, dass  $h_i$  auf jedem Kompaktum gleichmäßig gegen  $h$  konvergiert. Wegen der Beschränktheit von  $h_i$  und  $h$  sowie der in den Voraussetzungen von Lemma 5.6 geforderten Verteilungskonvergenz konvergiert die Spur von  $H_i$  gegen die Spur von  $H$ .

Es sei im Folgenden  $\lambda_\sigma := 0$  für  $\sigma \in \mathbb{N}$ , falls  $\sigma \notin \Sigma$ , und  $\lambda_\sigma^i := 0$  für  $\sigma \in \mathbb{N}$ , falls  $\sigma \notin \Sigma_i$ . Um Konvergenz in Verteilung zu zeigen, weisen wir die  $L_1$ -Konvergenz von  $S^i$  gegen  $S$  nach. Sei  $\varepsilon > 0$  beliebig. Wegen der Endlichkeit der folgenden Summe können wir dann ein  $\sigma_0 \in \mathbb{N}$  so wählen, dass

$$\sum_{\substack{\sigma > \sigma_0 \\ \sigma \in \Sigma}} \lambda_\sigma \leq \varepsilon.$$

Es seien die positiven Eigenwerte  $(\lambda_\sigma^i)$  von  $H_i$  so durch eine neue Indexmenge  $\Sigma_i \subset \mathbb{N}$  indiziert, dass für alle  $\sigma \leq \sigma_0$ ,  $\sigma \in \mathbb{N}$ , gilt  $\lambda_\sigma^i \rightarrow \lambda_\sigma$  für  $i \rightarrow \infty$ . Dies gilt mit einer geeigneten Umindizierung nach Satz 5.10. Entsprechend dieser Indizierung seien die  $(Z_\sigma)_\sigma$  den  $\lambda_\sigma^i$  bzw.  $\lambda_\sigma$  zugeordnet. Wegen der punktweisen Konvergenz der Eigenwerte gibt es ein  $i_0 \in \mathbb{N}$ , so dass für alle  $i > i_0$

$$|\lambda_\sigma^i - \lambda_\sigma| \leq \frac{\varepsilon}{2^\sigma} \quad \text{für } 1 \leq \sigma \leq \sigma_0$$

und wegen der Konvergenz der Summen können wir  $i_0$  vergrößern, bis auch

$$\left| \sum_{\sigma \in \Sigma_i} \lambda_\sigma^i - \sum_{\sigma \in \Sigma} \lambda_\sigma \right| \leq \varepsilon \quad \text{für alle } i > i_0.$$

Nun gilt

$$\begin{aligned} \mathbb{E}|S^i - S| &= \mathbb{E} \left| \sum_{\sigma \in \Sigma_i} \lambda_\sigma^i Z_\sigma^2 - \sum_{\sigma \in \Sigma} \lambda_\sigma Z_\sigma^2 \right| \leq \mathbb{E} \sum_{\sigma} |\lambda_\sigma^i - \lambda_\sigma| Z_\sigma^2 \\ &= \sum_{\sigma \leq \sigma_0} |\lambda_\sigma^i - \lambda_\sigma| + \sum_{\sigma > \sigma_0} |\lambda_\sigma^i - \lambda_\sigma| \\ &\leq 2\varepsilon + \sum_{\sigma > \sigma_0} \lambda_\sigma^i + \sum_{\sigma \leq \sigma_0} \lambda_\sigma^i - \sum_{\sigma \leq \sigma_0} \lambda_\sigma^i + \sum_{\sigma \leq \sigma_0} \lambda_\sigma + \sum_{\sigma > \sigma_0} \lambda_\sigma - \sum_{\sigma} \lambda_\sigma \\ &\leq 5\varepsilon. \end{aligned}$$

Es folgt  $L_1$ -Konvergenz, also Konvergenz in Verteilung.  $\square$

### 5.3.1 Nachweis für die neue Familie von Tests

Wir befinden uns in der Situation von Abschnitt 5.2.1 und es sei wie zuvor  $\phi \in \Phi$  beschränkt. Dann sind die Operatoren  $(H_i)_i$ ,  $H$  nach Satz 3.5 positiv und von der Spurklasse und erfüllen somit Forderungen (1.) und (2.).

Damit folgt aus Satz 5.12 der folgende Satz, der die Verteilungskonvergenz der durch die Approximation der Eigenwerte gewonnenen Verteilung gegen die tatsächliche Grenzverteilung zeigt.

**Satz 5.13.** *Es sei  $((m_i, n_i))_i$  eine Folge von Paaren von Stichprobenumfängen mit  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$ . Sei  $\phi \in \Phi$  beschränkt,  $P^X, P^Y$  nicht degenerierte Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^d, \mathcal{B}^d)$  und  $P^Z := \rho P^X + (1 - \rho)P^Y$ . Dann gilt für fast jedes Paar von Folgen  $(x_n)_n, (y_n)_n$  von Beobachtungen unabhängiger Zufallsvariablen auf  $\mathbb{R}^d$  mit Verteilung  $P^X$  beziehungsweise  $P^Y$*

$$S^i := \sum_{\sigma \in \Sigma_i} \lambda_\sigma^i Z_\sigma \xrightarrow{i \rightarrow \infty} \sum_{\sigma \in \Sigma} \lambda_\sigma Z_\sigma =: S,$$

wobei  $(Z_\sigma)_\sigma$  jeweils eine Folge unabhängiger, standardnormalverteilter Zufallsvariablen sei.

Eine Approximation der in  $S$  verwendeten Eigenwerte durch die Eigenwerte des diskretisierten Operators  $H_i$  ist also zulässig. Der Satz lässt sich analog auf den Einstichproben-Fall übertragen. Die Aussage des Satzes haben wir bereits auf anderem Weg durch Korollar 4.5 in Abschnitt 4.2 erhalten. Auf die numerische Behandlung zur Bestimmung des kritischen Wertes bei den neuen Testverfahren sind wir in Abschnitt 4.2.1 eingegangen.

Wir zeigen nun exemplarisch die Wirksamkeit des Verfahrens bei der Cramér-von Mises-Testgröße aus Abschnitt 5.2.2, angewendet auf die Anpassung an die Gleichverteilung auf dem Intervall  $[0, 1]$ .

### 5.3.2 Anwendung auf den Cramér-von Mises-Test

Wir betrachten den Anpassungstest an eine Gleichverteilung auf  $[0, 1]$ . Die Grenzverteilung der Cramér-von Mises-Testgröße ist unter der Hypothese eine gewichtete Summe von  $\chi_1^2$ -verteilten Zufallsvariablen, wobei die Gewichte die Eigenwerte des Integraloperators  $H^{\text{CvM}}$  mit Kern

$$h_{\text{CvM}}(x, y) = \frac{1}{3} - \max(x, y) + \frac{1}{2}x^2 + \frac{1}{2}y^2, \quad x, y \in [0, 1],$$

sind. Mit den Vorüberlegungen zu Konsistenz und diskreter Kompaktheit, die in Satz 5.9 festgehalten wurden, ist Lemma 5.11 sofort anwendbar und wir erhalten (für fast jede Beobachtungsfolge) Konvergenz der Eigenwerte der Operatoren  $(H_N^{\text{CvM}})_N$

gegen die des Operators  $H^{\text{CvM}}$ . Darüber hinaus folgt aus der Darstellung (5.8) die Positivität der Operatoren  $(H_N^{\text{CvM}})_N, H^{\text{CvM}}$ . Die exakten Eigenwerte des Operators  $H^{\text{CvM}}$  sind  $\lambda_\sigma = (\pi^2 \sigma^2)^{-1}$ ,  $\sigma \in \mathbb{N}$  (vergleiche z.B. Durbin und Knott [DK72]), womit der Operator  $H^{\text{CvM}}$  zur Spurklasse gehört. Die Bedingungen (1.) und (2.) sind also erfüllt. Wir erhalten also nicht nur die Konvergenz der Eigenwerte sondern darüber hinaus sogar die Verteilungskonvergenz der mit den Eigenwerten gewichteten Summe von unabhängigen  $\chi_1^2$ -verteilten Zufallsvariablen, der Grenzverteilung der Cramér-von Mises-Teststatistik unter der Hypothese, dass  $P^X$  die Gleichverteilung auf  $[0, 1]$  ist.

Im diskreten Fall für  $u \in L_2(X_N, \mathcal{P}(X_N), \mu_N)$  genügt es, wie in Abschnitt 4.2.1 erläutert, den diskreten Integraloperator

$$(H^{\text{CvM}}u)(x) = \frac{1}{N} \sum_{j=1}^N h_{\text{CvM}}(x, x_j)u(x_j)$$

an den Stellen  $x = x_1, \dots, x_N$  auszuwerten, wodurch die Eigenwertgleichung zum Linearen Gleichungssystem

$$\lambda \vec{u} = A \vec{u}$$

wird mit  $\vec{u} = (u(x_1), \dots, u(x_N))'$  und der Matrix  $A = (a_{j,k})_{j,k}$ ,  $j, k \in \{1, \dots, N\}$ , gegeben durch

$$a_{j,k} := \frac{1}{N} h_{\text{CvM}}(x_j, x_k), \quad j, k \in \{1, \dots, N\}.$$

Am Beispiel von  $N = 10, 50, 100, 500, 1000$  Beobachtungen von unabhängigen auf  $[0, 1]$  gleichverteilten Zufallsvariablen wurde dieses Vorgehen durchgeführt und die Approximationen der ersten vier Eigenwerte mit den wohlbekannten exakten Eigenwerten verglichen. Die Ergebnisse in Tabelle 5.1 verdeutlichen, dass tatsächlich Konvergenz vorliegt.

Tabelle 5.1: Approximation der Eigenwerte durch diskrete Operatoren mit zufälligen Beobachtungen

$N$	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$
10	0.11729	0.03107	0.00742	0.00634
50	0.10124	0.02927	0.00963	0.00627
100	0.10781	0.02935	0.01095	0.00674
500	0.10086	0.02694	0.01103	0.00633
1000	0.10289	0.02642	0.01092	0.00628
exakt	0.10132	0.02533	0.01126	0.00633

Die Abweichungen von den exakten Werten scheinen stärker von den zufälligen Daten als von der Diskretisierung herzurühren: Wählt man für  $N = 50, 100$  die Werte

$x_j = j/N - 1/(2N)$ ,  $j = 1, \dots, N$ , also die  $(j/N - 1/(2N))$ -Quantile,  $j = 1, \dots, N$ , der Gleichverteilung als „Beobachtungsdaten“, so ergibt sich beim Blick auf die Ergebnisse in Tabelle 5.2 eine deutlich bessere Approximation.

Tabelle 5.2: Approximation der Eigenwerte durch diskrete Operatoren mit Quantilen als vorgegebenen Beobachtungen

$N$	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$
50	0.10135	0.02536	0.01129	0.00637
100	0.10133	0.02534	0.01127	0.00634
exakt	0.10132	0.02533	0.01126	0.00633

Da in dem Fall der neuen Tests eine unbekannte Verteilung zu Grunde liegt, die bestenfalls durch die empirische Verteilung der zusammgelegten Beobachtung approximiert werden kann, darf man hoffen, dass die Approximationsqualität auch für kleinere Größen der zusammgelegten Stichprobe hinreichend gut sein wird.

# Kapitel 6

## Effizienz

Dieses Kapitel beginnt mit der Ermittlung von approximativen Bahadur-Effizienzen für einige Tests der neuen Familie von Teststatistiken. Neben dem Cramér-Test (mit  $\phi(\cdot) = \sqrt{\cdot}/2$ ) und dem Test von Bahr (mit  $\phi(\cdot) = 1 - \exp(-\cdot/2)$ ) werden einige weitere Beispiele betrachtet. Drei Testprobleme werden in der Folge untersucht: Die Tests auf Lage- und Skalenalternativen bei eindimensionalen Normalverteilungen und ein Kontaminationsmodell.

Zusätzliche Bedeutung wird den approximativen Bahadur-Effizienzen dadurch zuteil, dass sie lokal mit den Pitman-Effizienzen übereinstimmen. Für dieses Resultat bedarf es des Nachweises der *Wieandschen Bedingung* III\*, welche in Abschnitt 6.2 für die erwähnten Testprobleme überprüft wird. Der letzte Abschnitt dieses Kapitels veranschaulicht die Resultate zu Lage- und Skalenalternativen bei Normalverteilungen.

### 6.1 Bahadur-Effizienz

Bahadur [Bah60] veröffentlichte 1960 ein Konzept zum Vergleich von zwei Folgen von Ein-Stichproben-Teststatistiken. Im Folgenden übertragen wir seine Vorgehensweise zur Definition und Betrachtung von *approximativen Bahadur-Steigungen* auf den Zwei-Stichproben-Fall, vergleiche dazu auch Perng und Littell [PL76].

Seien  $X_1, X_2, \dots$  eine Folge von  $\mathbb{R}^d$ -wertigen Zufallsvariablen mit Verteilungsfunktion  $F$  und  $Y_1, Y_2, \dots$  eine Folge von  $\mathbb{R}^d$ -wertigen Zufallsvariablen mit Verteilungsfunktion  $G$ . Dabei sei  $(F, G) \in \Theta := \mathcal{F} \times \mathcal{F}$ , wobei  $\mathcal{F}$  eine Teilmenge aller Verteilungsfunktionen  $\mathbb{R}^d$ -wertiger Zufallsvariablen ist. Die Zufallsvariablen  $X_1, X_2, \dots, Y_1, Y_2, \dots$  seien unabhängig.

Im Fall der neuen Familie von Tests wählt man zunächst eine spezielle Folge von Teststatistiken, also ein  $\phi \in \Phi_0$  und definiert  $\mathcal{F} := \mathcal{F}_d^4(\phi)$ , also die Menge aller

nicht-degenerierten Verteilungsfunktionen  $\mathbb{R}^d$ -wertiger Zufallsvariablen  $X$ , für die  $E\phi^4(|X|^2) < \infty$ , womit die in den vorangegangenen Kapiteln gewonnenen Verfahren zur Bestimmung des kritischen Wertes anwendbar sind.

Es sei  $((m_i, n_i))_i$  eine feste Folge von Paaren von Stichprobenumfängen, für die  $m_i/(m_i + n_i) \rightarrow \rho \in (0, 1)$  und  $m_i, n_i \rightarrow \infty$  für  $i \rightarrow \infty$ . Wir interessieren uns nun für den Test der Hypothese  $H$ , bezeichnet durch  $(F, G) \in \Theta_0 \subset \Theta$ , gegen die Alternative  $\Theta_1 := \Theta \setminus \Theta_0$ . Es sei  $(t_{m_i, n_i})_i$  eine Folge von reellwertigen Teststatistiken für den Test der Hypothese  $H$ . Verwerfung möge für große Werte der Testgröße erfolgen. Die Indizes  $m_i, n_i$  geben wie bisher an, dass  $t_{m_i, n_i}$  auf den ersten  $m_i$  Zufallsvariablen der Folge  $X_1, X_2, \dots$  und den ersten  $n_i$  Zufallsvariablen der Folge  $Y_1, Y_2, \dots$  beruht.

**Definition 6.1.** Die Folge  $(t_{m_i, n_i})_i$  wird **Standardfolge** (zum Test von  $H$ ) genannt, wenn die folgenden drei Bedingungen erfüllt sind:

(a) Es gibt eine stetige Verteilungsfunktion  $F_0$ , so dass für alle  $\theta \in \Theta_0$

$$\lim_{i \rightarrow \infty} P_\theta(t_{m_i, n_i} \leq x) = F_0(x) \quad \text{für jedes } x \in \mathbb{R}.$$

(b) Es existiert eine Konstante  $a \in (0, \infty)$  mit

$$\lim_{t \rightarrow \infty} \frac{-2 \log(1 - F_0(t))}{t^2} = a.$$

(c) Es gibt eine reellwertige Funktion  $b$  von  $\Theta_1$  nach  $(0, \infty)$ , so dass für jedes  $\theta \in \Theta_1$  gilt:

$$\lim_{i \rightarrow \infty} P_\theta \left( \left| \sqrt{\frac{m_i + n_i}{m_i n_i}} t_{m_i, n_i} - b(\theta) \right| > x \right) = 0$$

für jedes  $x > 0$ .

In Bedingung (a) ist die Unabhängigkeit der sich asymptotisch ergebenden Verteilungsfunktion  $F_0$  von  $\theta \in \Theta_0$  entscheidend. Da die neuen Teststatistiken nicht asymptotisch verteilungsfrei sind, also die Grenzverteilungen von der zu Grunde liegenden Verteilung abhängen, muss zur Untersuchung von Effizienzen mit Standardfolgen das allgemeine Testproblem

$$\begin{aligned} H & : \theta \in \Theta_0 := \{(F, F) : F \in \mathcal{F}\} && \text{gegen} \\ K & : \theta \in \Theta_1 := \{(F, G) : F, G \in \mathcal{F}, F \neq G\} \end{aligned}$$

eingeschränkt werden auf den Fall einer einfachen Hypothese.

Eine Standardfolge  $(t_{m_i, n_i})_i$  hat unter  $\theta \in \Theta_0$  die asymptotische Verteilung  $F_0$ . Wegen Bedingung (c) strebt die Statistik für  $\theta \in \Theta_1$  in Wahrscheinlichkeit gegen unendlich. Somit sind bei Standardfolgen große Werte signifikant für die Verwerfung

von  $H$ . Für eine Beobachtung  $x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}$  ist der exakte p-Wert, d.h. die Wahrscheinlichkeit, dass

$$t_{m_i, n_i}(X_1, \dots, X_{m_i}, Y_1, \dots, Y_{n_i}),$$

wobei die  $m_i + n_i$  Zufallsvariablen  $X_1, \dots, X_{m_i}, Y_1, \dots, Y_{n_i}$  u.i.v. mit der Verteilung der einfachen Hypothese  $H$  seien mögen, größer ist als

$$t_{m_i, n_i}(x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i}),$$

gegeben durch

$$1 - F_{0; m_i, n_i}(t_{m_i, n_i}(x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i})),$$

wobei  $F_{0; m_i, n_i}$  die Verteilungsfunktion von  $t_{m_i, n_i}(X_1, \dots, X_{m_i}, Y_1, \dots, Y_{n_i})$  ist.

Im Allgemeinen lässt sich der exakte p-Wert leider nicht bestimmen, da notwendige Resultate über die Verteilung und die Tail-Wahrscheinlichkeiten von  $t_{m_i, n_i}$  für endliches  $m_i, n_i$  nicht vorhanden sind. Daher beschäftigte sich Bahadur im Wesentlichen mit dem Verhalten approximativer p-Werte, d.h. die Wahrscheinlichkeit, dass eine Zufallsvariable mit der Grenzverteilung der Teststatistik unter der Hypothese größer ist, als die zur Beobachtung gehörende Teststatistik:

$$1 - F_0(t_{m_i, n_i}(x_1, \dots, x_{m_i}, y_1, \dots, y_{n_i})).$$

Selbstverständlich stimmen diese p-Werte im Allgemeinen nicht mit den exakten p-Werten überein, schlimmer noch, die Gestalt und die Eigenschaften von  $t_{m_i, n_i}$  bei endlichem  $m_i, n_i$  gehen in die Überlegung überhaupt nicht ein. Auf dieses Problem und die teilweise irreführenden relativen Effizienzen wies Bahadur [Bah60] bereits hin, zugleich unterstrich er aber die Bedeutung der häufig vorliegenden lokalen Gleichheit mit der Pitman-Effizienz.

Die approximative relative Effizienz von zwei Standardfolgen von Teststatistiken  $(t_{m_i, n_i}^{(1)})_i, (t_{m_i, n_i}^{(2)})_i$  im Sinn von Bahadur, die so genannte approximative relative Bahadur-Effizienz, wird wie folgt ermittelt: Zuerst werden eine Alternative  $\theta \in \Theta_1$  und zwei Folgen von Paaren von Beobachtungsumfängen  $((m_i^{(j)}, n_i^{(j)}))_i, j = 1, 2$ , so gewählt, dass einerseits das Mischungsverhältnis beider Folgen gegen denselben Wert  $\rho \in (0, 1)$  konvergiert

$$\lim_{i \rightarrow \infty} \frac{m_i^{(j)}}{m_i^{(j)} + n_i^{(j)}} = \rho \in (0, 1), \quad j = 1, 2,$$

und andererseits das Verhältnis der Logarithmen der approximativen p-Werte, aufgefasst als (unabhängige) Zufallsvariablen, in Wahrscheinlichkeit gegen 1 konvergiert

$$\frac{-2 \log \left( 1 - F_0^{(1)} \left( t_{m_i^{(1)}, n_i^{(1)}}^{(1)} \right) \right)}{-2 \log \left( 1 - F_0^{(2)} \left( t_{m_i^{(2)}, n_i^{(2)}}^{(2)} \right) \right)} \xrightarrow{i \rightarrow \infty} \frac{P_\theta}{1}, \quad (6.1)$$

wobei  $F_0^{(j)}$  die Grenzverteilung der Standardfolge  $j = 1, 2$  aus Definition 6.1 ist. Anschaulich lässt sich die Bedingung (6.1) so deuten, dass in der fest gewählten Alternative  $\theta$  beide Tests die Hypothese asymptotisch gleich erfolgreich verwerfen (Zähler und Nenner streben mit gleicher Geschwindigkeit gegen  $+\infty$ ), d.h. der Fehler 2. Art soll gleich schnell gegen 0 streben.

Gilt (6.1), so existiert die **approximative relative Bahadur-Effizienz**  $E_{12}(\theta)$  als Grenzwert der Verhältnisse der Stichprobenumfänge von Folge (1) relativ zu Folge (2)

$$E_{12}(\theta) := \lim_{i \rightarrow \infty} \frac{N_i^{(2)}}{N_i^{(1)}} = \lim_{i \rightarrow \infty} \frac{m_i^{(2)}}{m_i^{(1)}} = \lim_{i \rightarrow \infty} \frac{n_i^{(2)}}{n_i^{(1)}},$$

wobei  $N_i^{(j)} := m_i^{(j)} + n_i^{(j)}$ ,  $j = 1, 2$  (vergleiche Bahadur [Bah60, S.279]). Eine Untersuchung der **exakten relativen Bahadur-Effizienz** (definiert mit Hilfe von exakten p-Werten) wäre von Interesse; die vorhandenen Resultate, siehe unter anderem Nikitin und Ponikarov [NP02], lassen sich aber nicht in aller Allgemeinheit auf unsere Zwei-Stichproben-Situation von degenerierten, teilweise unbeschränkten Kernen verallgemeinern. Die approximative relative Bahadur-Effizienz hingegen lässt sich leicht bestimmen.

Bahadur hat gezeigt, dass für jedes  $\theta \in \Theta_0$  der Logarithmus des p-Wertes einer jeden Standardfolge, aufgefasst als Zufallsvariable, dieselbe asymptotische Verteilung

$$\lim_{i \rightarrow \infty} P_\theta(-2 \log(1 - F_0(t_{m_i, n_i})) \leq x) = 1 - \exp\left(-\frac{1}{2}x\right), \quad x > 0,$$

hat. Weiterhin gilt mit

$$c(\theta) := \begin{cases} 0 & \text{für } \theta \in \Theta_0, \\ ab^2(\theta) & \text{für } \theta \in \Theta_1, \end{cases} \quad (6.2)$$

wobei  $a, b$  in Bedingung (b) und (c) von Definition 6.1 gegeben sind, für jedes  $\theta \in \Theta$

$$\frac{m_i + n_i}{m_i n_i} (-2 \log(1 - F_0(t_{m_i, n_i}))) = c(\theta) + \varepsilon_i. \quad (6.3)$$

Dabei strebt  $\varepsilon_i = \varepsilon_i(\theta, \omega)$  für  $i \rightarrow \infty$  in Wahrscheinlichkeit gegen 0. Die Größe  $c(\theta)$  wird **approximative Bahadur-Steigung** genannt. Nach Ausnutzen von (6.3) und (6.1) in der Definition von  $E_{12}$  erhalten wir

$$E_{12}(\theta) = \frac{c_1(\theta)}{c_2(\theta)}, \quad \theta \in \Theta_1.$$

Ein Weg, die approximative relative Bahadur-Effizienz zu erhalten, führt also über die approximative Bahadur-Steigung, für welche man die Konstante  $a$  und die Funktion  $b(\theta)$  aus Definition 6.1 benötigt.

### 6.1.1 Bahadur-Steigungen für die neuen Tests

In diesem Abschnitt wird die approximative Bahadur-Steigung der neuen Tests für drei typische Probleme bestimmt, auf die Zwei-Stichproben-Tests angewendet werden. Dies sind

- der Test auf Lage-Alternativen,
- der Test auf Skalen-Alternativen sowie
- ein Test auf Verschmutzungen in einem Kontaminationsmodell.

Für den Fall, dass die Hypothese aus der Standardnormalverteilung besteht, wird für Lage- und Skalenalternativen der numerische Wert der Steigung ausgerechnet.

Zunächst gilt es, den Begriff der Standardfolge mit den neuen Teststatistiken  $(T_{m_i, n_i}^\phi)_i$  in Verbindung zu bringen. Diese bilden nicht direkt eine Standardfolge (vergleiche Bedingung (c) der Definition). Im Folgenden weisen wir jedoch nach, dass die Wurzel aus den Teststatistiken eine Standardfolge bildet.

Es seien die Voraussetzungen von Satz 3.7 erfüllt. Dann ist mit diesem und dem Satz von der stetigen Abbildung offensichtlich Bedingung (a) für die Folge

$$\left( \sqrt{T_{m_i, n_i}^\phi} \right)_i$$

erfüllt. Bedingung (b) folgt mit einem Resultat von Zolotarev [Zol61]

$$\begin{aligned} 1 - F_0(t) &= P \left( \sum_{\sigma=1}^{\infty} \lambda_\sigma Z_\sigma^2 > t^2 \right) \\ &= \frac{\prod_{\sigma=n_1+1}^{\infty} \left( 1 - \frac{\lambda_\sigma}{\lambda_1} \right)}{\Gamma \left( \frac{n_1}{2} \right)} \left( \frac{t^2}{2\lambda_1} \right)^{(n_1/2)-1} \exp \left( -\frac{t^2}{2\lambda_1} \right) (1 + o(1)) \end{aligned}$$

für  $t \rightarrow \infty$ , wobei  $\Gamma$  die Gamma-Funktion bezeichnet und  $F_0$  die Verteilungsfunktion zur asymptotischen Verteilung der Statistik  $(T_{m_i, n_i}^\phi)^{\frac{1}{2}}$  für  $i \rightarrow \infty$  sei. Wie in Satz 3.7 bezeichnet  $(Z_\sigma)_\sigma$  eine Folge unabhängiger, standardnormalverteilter Zufallsvariablen und  $(\lambda_\sigma)_\sigma$  die fallende Folge der positiven Eigenwerte des Operators  $H$ . Die Vielfachheit von  $\lambda_1$ , dem größten Eigenwert, sei  $n_1$

$$\lambda_1 = \lambda_2 = \dots = \lambda_{n_1} > \lambda_{n_1+1} \geq \dots \geq 0.$$

Wendet man nun den Logarithmus an, so erhält man

$$\frac{-2 \log(1 - F_0(t))}{t^2} = \frac{1}{\lambda_1} + o(1)$$

für  $t \rightarrow \infty$ , so dass Bedingung (b) erfüllt ist mit  $a = \lambda_1^{-1}$ . Die Existenz der Funktion  $b$  in Bedingung (c) folgt aus der angenehmen Darstellung

$$b(\theta) := d_\phi(F, G), \quad \theta = (F, G) \in \Theta_1. \quad (6.4)$$

Für die folgenden drei Problemstellungen wird  $\Theta_1$  umparametrisiert, um den Lage-, Skalen- beziehungsweise Kontaminationsparameter zu repräsentieren.

### 6.1.2 Lage-Alternativen bei eindimensionalen Normalverteilungen

Gegeben sei das Testproblem der einfachen Hypothese

$$H : \theta \in \Theta_0 = \{(F_{N(0,1)}, F_{N(0,1)})\}$$

gegen die Alternative

$$K : \theta \in \Theta_1 = \{(F_{N(\vartheta,1)}, F_{N(\vartheta,1)}) : \vartheta \in \mathbb{R} \setminus \{0\}\},$$

wobei  $F_{N(\vartheta,\sigma^2)}$  die Verteilungsfunktion der eindimensionalen Standardnormalverteilung mit Mittelwert  $\vartheta$  und Varianz  $\sigma^2 > 0$  sei. Die Beschränkung auf eine einfache Hypothese ist nach den Vorbemerkungen in unserem Fall bei der Verwendung von Bahadurs Standardfolgen unumgänglich. Der größte Eigenwert  $\lambda_1$ , der für die Bahadur-Steigung mittels  $a = \lambda_1^{-1}$  von Bedeutung ist, ist abhängig von der gewählten Funktion  $\phi \in \Phi$ . Für die in dieser Arbeit betrachteten Kerne finden sich die jeweils größten Eigenwerte bei Normalverteilungen in Tabelle 6.1, S.99. Die Funktion  $b(\theta)$ ,  $\theta \in \Theta_1$ , sei für dieses Problem parametrisiert durch den in der Alternative von 0 verschiedenen Erwartungswert  $\vartheta$  der zweiten Stichprobe.

Aus (6.4) folgt für diesen Fall

$$b^2(\vartheta) = 2\mathbb{E}\phi(|Z_2 + \vartheta|^2) - 2\mathbb{E}\phi(|Z_2|^2) \quad (6.5)$$

mit  $Z_2 \sim N(0, 2)$ . Häufig interessiert man sich für die führenden Terme von  $\vartheta$  in der Entwicklung von  $b^2$ , also das lokale Verhalten von  $b^2(\vartheta)$  nahe der Hypothese  $\vartheta = 0$ . Lässt sich  $b^2$  darstellen als

$$b^2(\vartheta) = 2 \sum_{\ell=1}^{\infty} \vartheta^{2\ell} \sum_{j=0}^{\ell} \frac{2^{2j}}{(2j)!(\ell-j)!} \mathbb{E} [Z_2^{2j} \phi^{(\ell+j)}(Z_2^2)], \quad (6.6)$$

wobei wieder  $Z_2 \sim N(0, 2)$ , so sind die führenden Terme von  $\vartheta$  gegeben durch

$$b^2(\vartheta) = 2 (\mathbb{E} [\phi'(Z_2^2)] + 2\mathbb{E} [Z_2^2 \phi''(Z_2^2)]) \vartheta^2 + \mathcal{O}(\vartheta^4) \quad (6.7)$$

für  $\vartheta \rightarrow 0$ .

Anwendung von (6.6) auf den Kern von Bahr,  $\phi(z) = 1 - \exp(-z/2)$ , liefert

$$\begin{aligned} b^2(\vartheta) &= -2 \sum_{\ell=1}^{\infty} \left( \frac{-\vartheta^2}{2} \right)^{\ell} \sum_{j=0}^{\ell} \frac{(-1)^j 2^j}{(2j)!(\ell-j)!} \mathbb{E} [Z_2^{2j} \exp(-Z_2^2/2)] \\ &= -\frac{2}{\sqrt{3}} \sum_{\ell=1}^{\infty} \left( \frac{-\vartheta^2}{6} \right)^{\ell} = \frac{2}{\sqrt{3}} \left( 1 - \exp\left(-\frac{\vartheta^2}{6}\right) \right) \\ &= \frac{\vartheta^2}{3\sqrt{3}} + \mathcal{O}(\vartheta^4) \end{aligned}$$

für  $\vartheta \rightarrow 0$ . Dieses Ergebnis stimmt mit dem überein, welches Bahr [Bah96] durch die Darstellung mit Hilfe der Differenz von charakteristischen Funktionen gewonnen hat. Für den Kern des Cramér-Tests,  $\phi(\cdot) = \sqrt{\cdot}/2$ , ist die vorangegangene Entwicklung nicht hilfreich, da  $\mathbb{E}[\phi'(Z_2^2)] = \infty$ . Über einen direkten Ansatz mit Hilfe von Gleichung (6.5) ergibt sich

$$\begin{aligned} b^2(\vartheta) &= \mathbb{E}|Z_2 + \vartheta| - \mathbb{E}|Z_2| = 2\vartheta \int_0^{\vartheta} f_{Z_2}(z) dz - 2 \int_0^{\vartheta} z f_{Z_2}(z) dz \\ &= \vartheta \operatorname{erf}\left(\frac{\vartheta}{2}\right) + \frac{2}{\sqrt{\pi}} \left( \exp\left(-\frac{\vartheta^2}{4}\right) - 1 \right) \\ &= \frac{\vartheta^2}{2\sqrt{\pi}} + \mathcal{O}(\vartheta^4) \end{aligned} \tag{6.8}$$

für  $\vartheta \rightarrow 0$ , wobei  $f_{Z_2}$  die Lebesgue-Dichte der  $N(0, 2)$ -Verteilung bezeichne.

### 6.1.3 Skalenalternativen bei eindimensionalen Normalverteilungen

Wir beschränken uns auf den Test der einfachen Hypothese

$$H : \theta \in \Theta_0 = \{(F_{N(0,1)}, F_{N(0,1)})\}$$

gegen die Alternative

$$K : \theta \in \Theta_1 = \{(F_{N(0,1)}, F_{N(0,\sigma^2)}) : \sigma^2 \in \mathbb{R}_{>0} \setminus \{1\}\}.$$

In diesem Fall parametrisieren wir die Alternative durch  $1 - \sigma$ . Die Funktion  $b^2$  lautet

$$b^2(1 - \sigma) = 2\mathbb{E}\phi\left(\frac{1 + \sigma^2}{2} Z_2^2\right) - \mathbb{E}\phi(Z_2^2) - \mathbb{E}\phi(\sigma^2 Z_2^2), \quad \sigma^2 \in \mathbb{R}_{>0}, \tag{6.9}$$

mit  $Z_2 \sim N(0, 2)$ . Die führenden Terme von  $b^2$  in  $1 - \sigma$  sind gegeben durch

$$b^2(1 - \sigma) = -\mathbb{E}[Z_2^4 \phi''(Z_2^2)] (1 - \sigma)^2 + \mathcal{O}(1 - \sigma)^3 \tag{6.10}$$

für  $\sigma \rightarrow 1$ , falls die Entwicklung von  $b^2$  in der Nähe von  $\sigma = 1$  in eine Potenzreihe und damit die Darstellung (6.10) zulässig ist. Für den Kern von Bahr folgt somit

$$b^2(1 - \sigma) = \frac{1}{3\sqrt{3}}(1 - \sigma)^2 + \mathcal{O}((1 - \sigma)^3)$$

für  $\sigma \rightarrow 1$ , was Bahrs mit Hilfe der Darstellung von charakteristischen Funktionen gewonnenes Ergebnis verifiziert. Für den Kern des Cramér-Tests ergibt sich auf diesem Weg

$$b^2(1 - \sigma) = \frac{1}{4\sqrt{\pi}}(1 - \sigma)^2 + \mathcal{O}((1 - \sigma)^3)$$

für  $\sigma \rightarrow 1$ . Die letzte Gleichung kann auch durch Entwicklung von (6.9) nach Berechnung der in diesem Fall leicht zu bestimmenden Erwartungswerte erfolgen

$$b^2(1 - \sigma) = \frac{1}{\sqrt{\pi}} \left( \sqrt{2(1 + \sigma^2)} - 1 - \sigma \right), \quad \sigma > 0.$$

### 6.1.4 Ein Kontaminationsmodell

Tukey [OGH<sup>+</sup>60, S.448] untersuchte 1960 ein Modell, bei dem eine Normalverteilung mit einer weiteren Normalverteilung gemischt, „kontaminiert“, wurde, um Erkenntnisse über das Verhalten der relativen Effizienz der mittleren Abweichung zu gewinnen. Wir verzichten hier auf eine spezielle Wahl von Verteilungen und testen, ob eine Stichprobe einer bekannten Verteilung  $F$  signifikant verschieden ist von einer zweiten Stichprobe, die von einer Mischung aus der Verteilung  $F$  und einer weiteren, bekannten Verteilung  $G$  stammt.

Dazu seien  $F, G \in \mathcal{F}$  mit  $F \neq G$  fest gewählt. Das Testproblem hat die Form

$$H : \theta \in \Theta_0 = \{(F, F)\}$$

gegen die Alternative

$$K : \theta \in \Theta_1 = \{(F, (1 - \vartheta)F + \vartheta G) : \vartheta \in (0, 1)\}.$$

Wie zuvor ist die Wurzel der neuen Teststatistiken eine Standardfolge mit  $a = \lambda_1^{-1}$ ,  $\lambda_1$  der größte Eigenwert des Integraloperators  $H$  unter der Hypothese, und

$$b^2(\vartheta) = d_\phi^2(F, (1 - \vartheta)F + \vartheta G) = \vartheta^2 d_\phi^2(F, G).$$

Um eine große approximative Bahadur-Steigung zu erlangen, sollte man bei bekannter zu Grunde liegender Verteilung  $F$  und bekannter „Verschmutzungsverteilung“  $G$  eine Kernfunktion  $\phi$  wählen, bei welcher der Quotient aus dem Abstandsquadrat von  $F$  und  $G$ ,  $d_\phi^2(F, G)$ , und dem größten Eigenwert  $\lambda_1$  von  $H$  (unter der Hypothese) maximal ist.

Betrachtet man das „umgekehrte“ Testproblem, also den Test einer Verschmutzung von  $G$  durch  $F$ , so braucht die Funktion  $\phi$ , welche die Steigung maximiert, trotz der Symmetrie von  $d_\phi^2$  nicht dieselbe zu sein wie bei der ersten Fragestellung, da nun  $G$  und nicht  $F$  die dem Operator  $H$  zu Grunde liegende Verteilung ist und sich somit auch die Eigenwerte, speziell  $\lambda_1$ , ändern können.

## 6.2 Pitman-Effizienz

Bahadur hat 1960 in seinem Artikel [Bah60] auf die Verbindung zwischen der lokalen approximativen Bahadur-Effizienz (wenn sich die Alternative  $\theta \in \Theta_1$  geeignet der Hypothese  $\Theta_0$  nähert) und der lokalen Pitman-Effizienz (wenn das Niveau des Tests gegen 0 geht) hingewiesen. Wieand [Wie74] bewies 1974 in seiner Dissertation eine Erweiterung dieses Ergebnisses, welches die Gleichheit der beiden lokalen Effizienzen unter bestimmten Voraussetzungen sichert. Für den Begriff der Pitman-Effizienz verwenden wir die von Wieand [Wie74] vorgeschlagene Erweiterung der Definition von Fraser [Fra57], bei der wie bei der Bahadur-Effizienz die Formulierung auf den Zwei-Stichproben-Fall angepasst wurde.

**Definition 6.2.** *Es seien  $((m_i^{(j)}, n_i^{(j)}))_i$ ,  $j = 1, 2$ , zwei Folgen von Paaren von wachsenden Stichprobenumfängen mit*

$$\frac{m_i^{(j)}}{m_i^{(j)} + n_i^{(j)}} \rightarrow \rho \in (0, 1) \quad \text{für } i \rightarrow \infty, \quad j = 1, 2.$$

Weiter seien

$$\left( t_{m_i^{(j)}, n_i^{(j)}}^{(j)} \right)_i, \quad j = 1, 2,$$

zwei Folgen von Teststatistiken zum Test der Hypothese  $H : \theta = \theta_0$  gegen die Alternative  $K : \theta = \theta_k$ , wobei  $\theta_k \neq \theta_0$ . Für gegebene Folgen  $\theta_k \rightarrow \theta_0$ ,  $\beta_k^{(1)} \rightarrow \beta$  und  $\beta_k^{(2)} \rightarrow \beta$  mit  $0 < \beta < 1$  sei  $i_0^{(j)}(k)$  die kleinste natürliche Zahl, so dass für jedes  $i \geq i_0^{(j)}(k)$  gilt

$$P_{\theta_k} \left( t_{m_i^{(j)}, n_i^{(j)}}^{(j)} > q_{m_i^{(j)}, n_i^{(j)}}^{(j)} \right) \geq \beta_k^{(j)}, \quad j = 1, 2, \quad k = 1, 2, \dots, \quad (6.11)$$

wobei der kritische Wert des Test ( $j$ ) zum Niveau  $\alpha$  bei Stichprobenumfängen  $m_i^{(j)}$  und  $n_i^{(j)}$ ,  $q_{m_i^{(j)}, n_i^{(j)}}^{(j)}$  gegeben ist durch

$$P_{\theta_0} \left( t_{m_i^{(j)}, n_i^{(j)}}^{(j)} > q_{m_i^{(j)}, n_i^{(j)}}^{(j)} \right) = \alpha, \quad j = 1, 2, \quad i = 1, 2, \dots \quad (6.12)$$

Die **Pitman-Effizienz** der Folge  $\left( t_{m_i^{(1)}, n_i^{(1)}}^{(1)} \right)_i$  bezüglich  $\left( t_{m_i^{(2)}, n_i^{(2)}}^{(2)} \right)_i$  wird definiert durch

$$e_{12}(\alpha, \beta) := \lim_{k \rightarrow \infty} \frac{m_{i_0^{(2)}(k)}^{(2)} + n_{i_0^{(2)}(k)}^{(2)}}{m_{i_0^{(1)}(k)}^{(1)} + n_{i_0^{(1)}(k)}^{(1)}}$$

sofern dieser Grenzwert existiert und unabhängig von der Wahl der Folgen  $\theta_k, \beta_k^{(1)}, \beta_k^{(2)}$  ist. Ist dies nicht der Fall, so sei

$$e_{12}^+(\alpha, \beta) := \sup_{\{\Pi\}} \limsup_{k \rightarrow \infty} \frac{m_{i_0^{(2)}(k)}^{(2)} + n_{i_0^{(2)}(k)}^{(2)}}{m_{i_0^{(1)}(k)}^{(1)} + n_{i_0^{(1)}(k)}^{(1)}}$$

und

$$e_{12}^-(\alpha, \beta) := \inf_{\{\Pi\}} \liminf_{k \rightarrow \infty} \frac{m_{i_0^{(2)}(k)}^{(2)} + n_{i_0^{(2)}(k)}^{(2)}}{m_{i_0^{(1)}(k)}^{(1)} + n_{i_0^{(1)}(k)}^{(1)}}.$$

Hierbei bezeichne  $\sup_{\{\Pi\}}$  bzw.  $\inf_{\{\Pi\}}$  das Supremum bzw. Infimum über alle der Definition entsprechenden Folgen  $\theta_k, \beta_k^{(1)}, \beta_k^{(2)}$ .

**Bemerkung.** In Formel (6.11) liegt die Erweiterung Frasers Definition, da hier eine ausreichend große Güte für alle genügend großen Stichprobenumfänge statt nur für einen bestimmten Stichprobenumfang garantiert wird.

Gleichung (6.12) ist eine vereinfachte Darstellung der Forderung, dass der Test (j) bei Stichprobenumfängen  $m_i^{(j)}$  und  $n_i^{(j)}$  das Niveau  $\alpha$  exakt einhalten soll. Diese Forderung kann durch Randomisierung stets erfüllt werden.

Die schwierigste Aufgabe, vor der man steht, will man Wieands Resultat anwenden, ist der Nachweis seiner bekannten Bedingung III\*.

**Definition (Bedingung III\*).** Es sei  $(t_{m_i, n_i})_i$  eine Standardfolge. Dann sagt man, dass diese Folge die Bedingung III\* erfüllt, falls gilt:

Es gibt ein  $\theta^* > 0$ , so dass für jedes  $\varepsilon > 0$  und  $\delta \in (0, 1)$  ein  $C > 0$  existiert mit der Eigenschaft, dass für jedes  $\theta \in \dot{U}_{\theta^*}(\theta_0) := (\theta_0 - \theta^*, \theta_0 + \theta^*) \setminus \{\theta_0\}$  und jedes  $i$  mit  $m_i, n_i > C/b^2(\theta)$  gilt

$$P_\theta \left( \left| \sqrt{\frac{m_i + n_i}{m_i n_i}} t_{m_i, n_i} - b(\theta) \right| < \varepsilon b(\theta) \right) > 1 - \delta. \quad (6.13)$$

Hat man Bedingung III\* nachgewiesen, so sind die weiteren Voraussetzungen des folgenden Satzes in aller Regel leicht zu zeigen.

**Satz 6.3.** Es seien zwei Folgen von Teststatistiken gegeben, welche die nachfolgenden Bedingungen erfüllen.

- (a) Beide Folgen sind Standardfolgen und besitzen unter der Hypothese asymptotische Verteilungen, deren Verteilungsfunktionen  $F_0^{(j)}(x)$ ,  $x \in \mathbb{R}$ ,  $j = 1, 2$ , für  $x \rightarrow \infty$  streng monoton wachsen.
- (b) Beide Folgen erfüllen Bedingung III\*.

(c) Es gilt  $\lim_{\theta \rightarrow \theta_0} b_j(\theta) = 0$  für  $j = 1, 2$ .

(d)  $\lim_{\theta \rightarrow \theta_0} \frac{\sqrt{a_1} b_1(\theta)}{\sqrt{a_2} b_2(\theta)}$  existiert und ist endlich.

Dann gilt für  $\beta \in (0, 1)$

$$\lim_{\theta \rightarrow \theta_0} E_{12}(\theta) = \lim_{\alpha \rightarrow 0} e_{12}^+(\alpha, \beta) = \lim_{\alpha \rightarrow 0} e_{12}^-(\alpha, \beta),$$

wobei die Existenz des ersten Grenzwerts die Existenz der beiden anderen impliziert, und

$$\lim_{\theta \rightarrow \theta_0} E_{12}(\theta) = \lim_{\alpha \rightarrow 0} e_{12}(\alpha, \beta),$$

falls  $e_{12}(\alpha, \beta)$  für alle  $\alpha$  aus einem Intervall der Form  $(0, \alpha')$  existiert.

**Beweis.** Siehe Wieand [Wie76, S.1005f]. □

**Bemerkung.** Wieand [Wie74, S.6f] bemerkt, dass sich sein für den Ein-Stichproben-Fall bewiesenes Resultat direkt auf den Zwei-Stichproben-Fall übertragen lässt. In seiner ursprünglichen Bedingung III\* fand statt Ungleichung (6.13) die folgende Ungleichung Verwendung:

$$P_\theta \left( \left| \frac{1}{\sqrt{m_i + n_i}} t_{m_i, n_i} - b(\theta) \right| < \varepsilon b(\theta) \right) > 1 - \delta.$$

Dies hat zur Folge, dass in den Ausdrücken für  $b(\theta)$  der Mischungsparameter in Form des Faktors  $\sqrt{\rho(1-\rho)}$  auftaucht (siehe dazu beispielsweise Tabelle VIII [Wie74, S.38]). Durch die Verwendung des Normierungsfaktors  $\sqrt{\frac{m_i + n_i}{m_i n_i}}$  in der obigen Formulierung der Wieandschen Bedingung wird dies vermieden.

Die folgenden Resultate zeigen, dass die im vorangegangenen Abschnitt auf Bahadur-Steigungen untersuchten Tests die Bedingung III\* erfüllen. Bildet man also die approximative relative Bahadur-Effizienz zweier Tests, so stimmt diese wegen des letzten Satzes lokal mit der Pitman-Effizienz überein.

### 6.2.1 Lokale Gleichheit bei Lage-Alternativen

Für Lage-Alternativen liefert Wieands Resultat ein bemerkenswert allgemeines Ergebnis.

**Satz 6.4.** Sei  $\phi \in \Phi$ ,  $F \in \mathcal{F}$  und  $((m_i, n_i))_i$  eine Folge von Paaren von Stichprobenumfängen mit  $m_i, n_i \rightarrow \infty$ ,  $\frac{m_i}{m_i + n_i} \rightarrow \rho \in (0, 1)$  für  $i \rightarrow \infty$ . Dann erfüllt  $((T_{m_i, n_i}^\phi)^{\frac{1}{2}})$  die Bedingung III\* für den Test auf Lage-Alternativen, also der Hypothese

$$H : \theta \in \Theta_0 = \{(F, F)\}$$

gegen die Alternative

$$K : \theta \in \Theta_1 = \{(F, F(\cdot - \vartheta)) : \vartheta \in \mathbb{R}^d \setminus \{0\}\}.$$

Die Annäherung von  $\vartheta \in \mathbb{R}^d$  an die Hypothese  $\vartheta = 0$  kann dabei auf einer beliebigen Kurve erfolgen.

**Beweis.** Es bezeichne abkürzend  $\mu$  die der ersten Stichprobe zu Grunde liegende Verteilung und  $\mu_m$  die (zufällige) empirische Verteilung von  $m$  unabhängigen Zufallsvariablen mit der Verteilung  $\mu$ . Es möge analog  $\nu_n$  unabhängig von  $\mu_m$  für die zweite Stichprobe mit zu Grunde liegender Verteilung  $\nu$  definiert seien. Wir erinnern uns:  $(T_{m,n}^\phi)^{\frac{1}{2}} = \sqrt{\frac{mn}{m+n}} d_\phi(\mu_m, \nu_n)$ . Da  $d := d_\phi$  als Metrik die Dreiecksungleichung erfüllt, gilt

$$\begin{aligned} d(\mu_m, \nu_n) &\leq d(\mu_m, \mu) + d(\mu, \nu) + d(\nu, \nu_n) \quad \text{und} \\ d(\mu, \nu) &\leq d(\mu_m, \mu) + d(\mu_m, \nu_n) + d(\nu, \nu_n). \end{aligned}$$

Die in der Wieandschen Bedingung III\* auftauchende Wahrscheinlichkeit kann damit nach unten abgeschätzt werden:

$$\begin{aligned} P_\vartheta \left( \left| \frac{(T_{m,n}^\phi)^{\frac{1}{2}}}{\sqrt{\frac{mn}{m+n}}} - b(\vartheta) \right| < \varepsilon b(\vartheta) \right) & \tag{6.14} \\ &= P_\vartheta (-\varepsilon b(\vartheta) < d(\mu_m, \nu_n) - d(\mu, \nu) < \varepsilon b(\vartheta)) \\ &= P_\vartheta (d(\mu_m, \nu_n) - d(\mu, \nu) < \varepsilon b(\vartheta), d(\mu, \nu) - d(\mu_m, \nu_n) < \varepsilon b(\vartheta)) \\ &\geq P_\vartheta (d(\mu_m, \mu) + d(\nu_n, \nu) < \varepsilon b(\vartheta)) \\ &\geq P_\vartheta \left( d(\mu_m, \mu) < \frac{\varepsilon}{2} b(\vartheta), d(\nu_n, \nu) < \frac{\varepsilon}{2} b(\vartheta) \right) \\ &= P_\vartheta \left( d(\mu_m, \mu) < \frac{\varepsilon}{2} b(\vartheta) \right) P_\vartheta \left( d(\nu_n, \nu) < \frac{\varepsilon}{2} b(\vartheta) \right). \end{aligned}$$

Die Wahrscheinlichkeiten in der letzten Zeile lassen sich mit Hilfe der Markovschen Ungleichung abschätzen:

$$P_\vartheta \left( d(\mu_m, \mu) < \frac{\varepsilon}{2} b(\vartheta) \right) \geq 1 - \frac{4}{\varepsilon^2 b^2(\vartheta)} \mathbb{E} d^2(\mu_m, \mu) = 1 - \frac{4}{\varepsilon^2 m b^2(\vartheta)} \mathbb{E} \phi(|X_1 - X_2|^2),$$

wobei  $X_1, X_2 \sim F$ . Analog

$$P_\vartheta \left( d(\nu_n, \nu) < \frac{\varepsilon}{2} b(\vartheta) \right) \geq 1 - \frac{4}{\varepsilon^2 b^2(\vartheta)} \mathbb{E} d^2(\nu_n, \nu) = 1 - \frac{4}{\varepsilon^2 n b^2(\vartheta)} \mathbb{E}_\vartheta \phi(|Y_1 - Y_2|^2),$$

wobei  $Y_1, Y_2 \sim F(\cdot - \vartheta)$ . Da sich eine Verschiebung des Erwartungswertes bei der Differenzbildung heraushebt, gilt für alle  $\vartheta \in \mathbb{R}^d \setminus \{0\}$

$$\mathbb{E} \phi(|X_1 - X_2|^2) = \mathbb{E}_\vartheta \phi(|Y_1 - Y_2|^2).$$

Wählt man nun

$$C = \frac{8}{\varepsilon^2 \delta} \mathbb{E} \phi(|X_1 - X_2|^2),$$

so folgt für  $m, n > C/b^2(\vartheta)$

$$P_\vartheta \left( d(\mu_m, \mu) < \frac{\varepsilon}{2} b(\vartheta) \right) \geq 1 - \frac{C\delta}{2mb^2(\vartheta)} > 1 - \frac{\delta}{2} > \sqrt{1 - \delta} \quad (6.15)$$

und analog

$$P_\vartheta \left( d(\nu_n, \nu) < \frac{\varepsilon}{2} b(\vartheta) \right) \geq 1 - \frac{C\delta}{2nb^2(\vartheta)} > 1 - \frac{\delta}{2} > \sqrt{1 - \delta}. \quad (6.16)$$

Für alle  $i$  mit  $m_i, n_i > C/b^2(\vartheta)$  einer Folge  $((m_i, n_i))_i$  ist damit durch Einsetzen dieser beiden Abschätzungen in (6.14) die Wieandsche Bedingung nachgewiesen.  $\square$

Für Lage-Alternativen ist die Wieandsche Bedingung somit nicht nur für Normalverteilungen erfüllt sondern allgemein für jede Verteilung  $F \in \mathcal{F}$ . Dank Wieands Resultat stimmt die approximative Bahadur-Effizienz, sofern sie der Regularitätsforderung ( $d$ ) in Satz 6.3 entspricht, mit der Pitman-Effizienz lokal überein. Folgt man dem Beweis, so ist die Forderung der Konvergenz der Differenz der Erwartungswerte gegen 0 nicht direkt für die Wieandsche Bedingung erforderlich, wird aber in Wieands Satz über die lokale Gleichheit, Satz 6.3, für Bedingung ( $c$ ) benötigt.

Der Nachweis von Bedingung III\* bei Skalenalternativen und beim Kontaminationsmodell kann in gleicher Weise geführt werden. Wir beginnen mit dem Kontaminationsmodell.

### 6.2.2 Lokale Gleichheit beim Kontaminationsmodell

Wir übertragen die Beweisführung von Satz 6.4: Die Abschätzung 6.14 kann dazu übernommen werden, lediglich die einzelnen Faktoren der letzten Zeile der Abschätzung erfordern eine etwas aufwändigere Betrachtung: Beim Test auf Lage-Alternativen stimmt der Erwartungswert  $\mathbb{E} \phi(|X_1 - X_2|^2)$ ,  $X_1, X_2 \sim F$  für jedes beliebige  $\vartheta \in \mathbb{R}^d \setminus \{0\}$  mit dem Erwartungswert  $\mathbb{E}_\vartheta \phi(|Y_1 - Y_2|^2)$  mit  $Y_1, Y_2 \sim G_\vartheta$  überein, wobei  $G_\vartheta = F(\cdot - \vartheta)$ . Dies macht die Ermittlung der Konstanten  $C$  besonders einfach. Es wäre allerdings bereits hinreichend, wenn die Erwartungswerte

$$\mathbb{E}_\vartheta \phi(|Y_1 - Y_2|^2) \xrightarrow{\vartheta \rightarrow 0} \mathbb{E} \phi(|X_1 - X_2|^2) \quad (6.17)$$

konvergieren würden, denn dann könnte man beim Nachweis der Wieandschen Bedingung eine Umgebung  $\dot{U}_{\vartheta^*}(0)$ , also ein  $\vartheta^* > 0$ , so klein wählen, dass

$$\mathbb{E}_\vartheta \phi(|Y_1 - Y_2|^2) < 2\mathbb{E} \phi(|X_1 - X_2|^2) \quad \text{für alle } \vartheta \in \dot{U}_{\vartheta^*}(0).$$

Eine geringfügige Korrektur, genauer: eine Verdoppelung der Konstante  $C$  erlaubt dann ähnliche Abschätzungen wie (6.15) und (6.16):

$$P_{\vartheta} \left( d(\mu_m, \mu) < \frac{\varepsilon}{2} b(\vartheta) \right) \geq 1 - \frac{C\delta}{4mb^2(\vartheta)} > 1 - \frac{\delta}{4} > \sqrt{1-\delta},$$

$$P_{\vartheta} \left( d(\nu_n, \nu) < \frac{\varepsilon}{2} b(\vartheta) \right) \geq 1 - \frac{C\delta}{2nb^2(\vartheta)} > 1 - \frac{\delta}{2} > \sqrt{1-\delta},$$

womit die Wieandsche Bedingung gezeigt ist.

Es genügt daher, in diesem Fall und beim Test auf Skalenalternativen lediglich die Konvergenz (6.17) zu zeigen.

Mit diesen Vorbemerkungen lässt sich nun leicht die lokale Gleichheit zwischen approximativer Bahadur-Effizienz und Pitman-Effizienz für das multivariate Kontaminationsmodell nachweisen.

**Satz 6.5.** *Sei  $\phi \in \Phi$ ,  $F, G \in \mathcal{F}$  und  $((m_i, n_i))_i$  eine Folge von Paaren von Stichprobenumfängen mit  $m_i, n_i \rightarrow \infty$ ,  $\frac{m_i}{m_i+n_i} \rightarrow \rho \in (0, 1)$  für  $i \rightarrow \infty$ . Dann erfüllt die Folge von Teststatistiken  $((T_{m_i, n_i}^{\phi})^{\frac{1}{2}})$  die Bedingung III\* für den Test der Hypothese*

$$H : \theta \in \Theta_0 = \{(F, F)\}$$

gegen die Alternative

$$K : \theta \in \Theta_1 = \{(F, (1-\vartheta)F + \vartheta G) : \vartheta \in (0, 1]\}.$$

**Beweis.** Für unabhängige  $Y_1, Y_2 \sim (1-\vartheta)F + \vartheta G$  gilt

$$E_{\vartheta} \phi(|Y_1 - Y_2|^2) = (1-\vartheta)^2 E\phi(|X_1 - X_2|^2) + 2\vartheta(1-\vartheta) E\phi(|X_1 - Z_1|^2) + \vartheta^2 E\phi(|Z_1 - Z_2|^2)$$

mit unabhängigen  $Z_1, Z_2 \sim G$ . Offensichtlich liegt für  $\vartheta \rightarrow 0$  die Konvergenz (6.17) dieses Erwartungswerts gegen  $E\phi(|X_1 - X_2|^2)$  vor. Es folgt die Behauptung. (Analog zeigt man  $E_{\vartheta} \phi(|X_1 - Y_1|^2) \rightarrow E\phi(|X_1 - X_2|^2)$  für  $\vartheta \rightarrow 0$ , womit Voraussetzung (c) in Satz 6.3 erfüllt ist.)  $\square$

### 6.2.3 Lokale Gleichheit bei Skalenalternativen

Auch bei Skalenalternativen funktioniert die am Anfang des letzten Abschnitts erläuterte Beweismethode. Allgemein gilt der folgende Satz, der die lokale Gleichheit von Pitman-Effizienz und approximativer Bahadur-Effizienz für Alternativen garantiert, die durch eine Drehstreckung aus der Verteilung unter der Hypothese hervorgegangen sind. Als hinreichendes Kriterium wird die Konvergenz der Abbildungsmatrix der Drehstreckung gegen die Identität in der Schur-Norm gefordert. Diese Forderung ist für den bereits betrachteten Fall von Skalenalternativen eindimensionaler Normalverteilungen aus Abschnitt 6.1.3 offensichtlich erfüllt. Durch

die Möglichkeit, andere Verteilungen unter der Hypothese als die Normalverteilung zu verwenden, sowie der bis auf die Konvergenzforderung freien Wahl der Folge der Drehmatrizen, ist der Satz von ähnlich allgemeiner Aussagekraft wie bereits Satz 6.4. Auf eine mögliche Verbindung der beiden Sätze wird am Ende dieses Abschnittes eingegangen.

**Satz 6.6.** *Sei  $\phi \in \Phi$ . Es seien  $X_1, X_2$  unabhängige  $d$ -dimensionale Zufallsvektoren mit Verteilung  $F \in \mathcal{F}$ . Wir betrachten den Test der Hypothese*

$$H : \theta \in \Theta_0 = \{(F, F)\}$$

gegen die Alternative

$$K : \theta \in \Theta_1 = \{(F, G_\vartheta) : \vartheta \in \{\vartheta_k : k \in \mathbb{N}\}\}$$

für eine Folge von Alternativen  $(\vartheta_k)_k$ ,  $k \in \mathbb{N}$ . Für alle  $\vartheta_k$  möge es Matrizen  $A^{(k)} = (a_{j,\ell}^{(k)})_{j,\ell} \in \mathbb{R}^{d \times d}$  geben, so dass die Zufallsvektoren

$$Y_1^{(k)} := A^{(k)} X_1 \quad \text{und} \quad Y_2^{(k)} := A^{(k)} X_2 \tag{6.18}$$

die Verteilung  $G_{\vartheta_k}$  haben. Sei weiter  $((m_i, n_i))_i$  eine Folge von Paaren von Stichprobenumfängen mit  $m_i, n_i \rightarrow \infty$ ,  $\frac{m_i}{m_i + n_i} \rightarrow \rho \in (0, 1)$  für  $i \rightarrow \infty$ . Wenn in dieser Situation  $A^{(k)}$  für  $k \rightarrow \infty$  in der Schur-Norm  $\|\cdot\|_S$  gegen die  $d$ -dimensionale Einheitsmatrix konvergiert, also

$$\|A^{(k)} - 1_d\|_S^2 := \sum_{j,\ell=1}^d \left( a_{j,\ell}^{(k)} - \mathbb{I}(j = \ell) \right)^2 \xrightarrow[k \rightarrow \infty]{} 0,$$

dann gilt  $G_\vartheta \in \mathcal{F}$  für alle  $\vartheta \in \{\vartheta_k : k \geq k_0\}$  für ein  $k_0 \in \mathbb{N}$  und die Folge von Teststatistiken  $((T_{m_i, n_i}^\phi)^{\frac{1}{2}})$  erfüllt die Bedingung III\* für das genannte Testproblem.

**Beweis.** Zunächst gilt  $G_{\vartheta_k} \in \mathcal{F}$  für  $k$  groß genug, denn mit der Monotonie von  $\phi$  und der Ungleichung  $\phi(s+t) \leq \phi(s) + \phi(t)$  für  $s, t \geq 0$  sowie speziell  $\phi(cs) \leq c\phi(s)$  für  $c \in \mathbb{N}$  aus dem Beweis von Lemma 2.2 gilt

$$\mathbb{E}\phi(|Y_1^{(k)}|^2) = \mathbb{E}\phi(|A^{(k)} X_1|^2) \leq \mathbb{E}\phi(\|A^{(k)}\|_S^2 |X_1|^2) \leq c\mathbb{E}\phi(|X_1|^2) < \infty$$

für  $c = \lceil \|A^{(k)}\|_S^2 \rceil$ . Wegen der Konvergenz der Abbildungsmatrizen ist die Verteilung  $G_{\vartheta_k}$  für alle  $k \geq k_0$  mit einem hinreichend großen  $k_0 \in \mathbb{N}$  nicht degeneriert. Es genügt zu zeigen, dass  $\mathbb{E}\phi(|Y_1^{(k)} - Y_2^{(k)}|^2)$  für  $k \rightarrow \infty$  gegen  $\mathbb{E}\phi(|X_1 - X_2|^2)$  konvergiert. Dies

ist der Fall, da

$$\begin{aligned}
\mathbb{E}\phi\left(\left|Y_1^{(k)} - Y_2^{(k)}\right|^2\right) &= \mathbb{E}\phi\left(\left|(A^{(k)} - 1_d)(X_1 - X_2) + X_1 - X_2\right|^2\right) \\
&\leq \mathbb{E}\phi\left(\|A^{(k)} - 1_d\|_S^2 |X_1 - X_2|^2 + |X_1 - X_2|^2\right. \\
&\quad \left.+ 2|(X_1 - X_2)'(A^{(k)} - 1_d)(X_1 - X_2)|\right) \\
&\leq \mathbb{E}\phi\left(\|A^{(k)} - 1_d\|_S^2 |X_1 - X_2|^2\right) + \mathbb{E}\phi\left(|X_1 - X_2|^2\right) \\
&\quad + \mathbb{E}\phi\left(2|X_1 - X_2|^2 \|A^{(k)} - 1_d\|_S\right),
\end{aligned}$$

wobei in der letzten Ungleichung die Cauchy-Schwarzsche Ungleichung zweifach Verwendung fand. Aus dem Satz von der majorisierten Konvergenz folgt

$$\limsup_{k \rightarrow \infty} \mathbb{E}\phi\left(\left|Y_1^{(k)} - Y_2^{(k)}\right|^2\right) \leq \mathbb{E}\phi\left(|X_1 - X_2|^2\right).$$

Ebenso lässt sich für alle  $k \in \mathbb{N}$  mit  $\|A^{(k)} - 1_d\|_S \leq \frac{1}{2}$  eine Abschätzung nach unten vornehmen. Mit  $\phi(s+t) \geq \phi(s) - \phi(|t|)$  für alle  $s \geq 0$  und  $t \geq -s$  gilt

$$\begin{aligned}
\mathbb{E}\phi\left(\left|Y_1^{(k)} - Y_2^{(k)}\right|^2\right) &\geq \mathbb{E}\phi\left(|X_1 - X_2|^2 + 2(X_1 - X_2)'(A^{(k)} - 1_d)(X_1 - X_2)\right) \\
&\geq \mathbb{E}\phi\left(|X_1 - X_2|^2\right) - \mathbb{E}\phi\left(2|X_1 - X_2|^2 \|A^{(k)} - 1_d\|_S\right),
\end{aligned}$$

also mit dem Satz von der majorisierten Konvergenz

$$\liminf_{k \rightarrow \infty} \mathbb{E}\phi\left(\left|Y_1^{(k)} - Y_2^{(k)}\right|^2\right) \geq \mathbb{E}\phi\left(|X_1 - X_2|^2\right).$$

Es folgt die Konvergenz (6.17) und damit die Wieandsche Bedingung. Analog zeigt man  $b^2(\vartheta_k) \rightarrow 0$  für  $k \rightarrow \infty$ , so dass Bedingung (c) des Satzes 6.3 erfüllt ist.  $\square$

Es ist leicht möglich die Sätze 6.4 und 6.6 zu einem allgemeineren Satz zu verknüpfen, der in der Alternative zugleich Veränderungen des Erwartungswerts und der Kovarianzstruktur zulässt. Da eine Veränderung des Erwartungswerts sich durch die Differenzbildung hinaushebt, lässt sich Gleichung (6.18) unmittelbar durch

$$Y_1^{(k)} := A^{(k)}X_1 + c^{(k)} \quad \text{und} \quad Y_2^{(k)} := A^{(k)}X_2 + c^{(k)}$$

ersetzen, wobei  $(c^{(k)})_k$ ,  $k \in \mathbb{N}$ ,  $c^{(k)} \in \mathbb{R}^d$  eine gegen 0 konvergente Folge ist.

### 6.3 Vergleich der Effizienzen einiger Tests der neuen Familie

Zum Abschluss dieses Kapitels stellen wir die ermittelten approximativen Bahadur-Steigungen für die Lage- und Skalenalternativen-Probleme zusammen, die wir im

nächsten Kapitel untersuchen werden. Wir betrachten fünf Kernfunktionen: den Kern des Cramér-Tests, den Kern von Bahr und die Kernfunktionen  $\phi(z) = \log(1 + z)$ ,  $\phi(z) = 1 - 1/(1 + z)$  und  $\phi(z) = 1 - 1/(1 + z)^2$ . Die Konstante  $a$  in der Definition (6.2) der approximativen Bahadur-Steigung, also der Kehrwert des größten Eigenwerts, ist bei festgehaltener Kernfunktion für die beiden Testprobleme, die Untersuchung von Lage- und Skalenalternativen, bei Normalverteilungen derselbe, da in beiden Fällen die eindimensionale Standardnormalverteilung als Verteilung in der Hypothese gewählt wird. Für den Cramér-Kern wurde der größte Eigenwert numerisch ermittelt (vergleiche Abschnitt 3.2.2) und beim Kern von Bahr ist er  $\lambda_1 = ((\sqrt{5} - 1)/2)^3$  [Bah96, S.75]. Die jeweils größten Eigenwerte zu den drei anderen Kernfunktionen wurden mit der in Abschnitt 4.2.1 vorgestellten Methode über „empirische“ Operatoren mit Quantilen der Normalverteilung als „Beobachtungen“ gewonnen (vgl. Abschnitt 5.3.2 für diese spezielle Diskretisierung).

Tabelle 6.1: Bahadur-Steigungen bei Lage-Alternativen eindimensionaler Standardnormalverteilungen

Kernfunktion	Größter EW	$\lim_{\vartheta \rightarrow 0} b^2(\vartheta)/\vartheta^2$	$\lim_{\vartheta \rightarrow 0} ab^2(\vartheta)/\vartheta^2$
$\sqrt{z}/2$	$a^{-1} = 0.29727$	$\frac{1}{2\sqrt{\pi}}$	0.949
$1 - \exp(-z/2)$	$a^{-1} = 0.23607$	$\frac{1}{3\sqrt{3}}$	0.815
$\log(1 + z)$	$a^{-1} \approx 0.49493$	0.45436	0.918
$1 - 1/(1 + z)$	$a^{-1} \approx 0.20361$	0.15923	0.782
$1 - 1/(1 + z)^2$	$a^{-1} \approx 0.22492$	0.15340	0.682

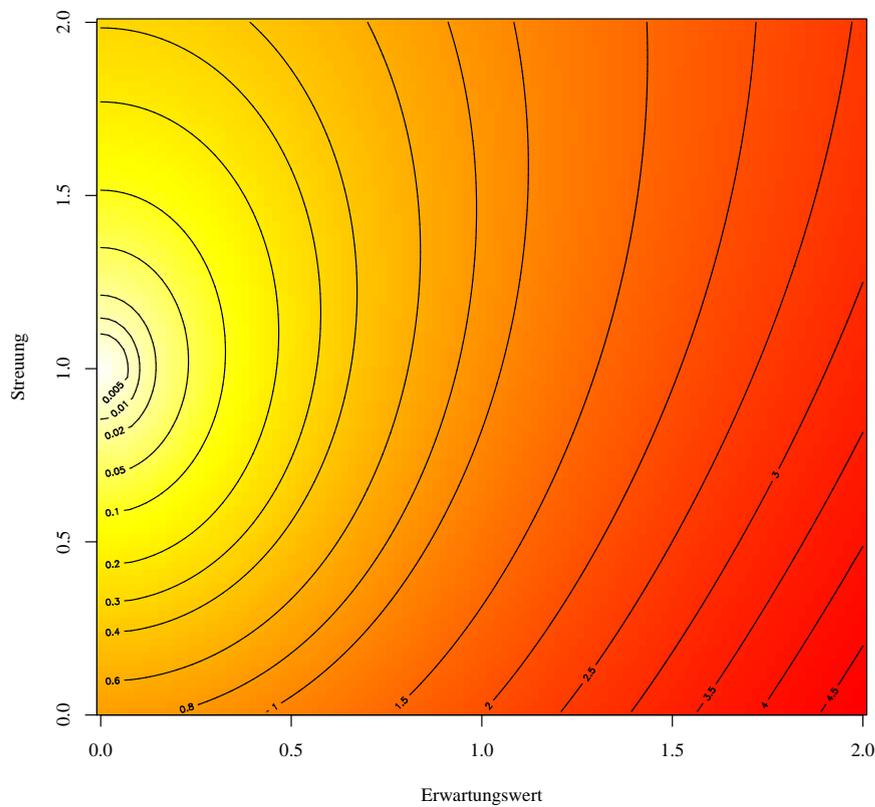
Bei Lage-Alternativen mit Lage-Parameter  $\vartheta$  ist die Funktion  $b^2(\vartheta)$  für den Cramér-Kern in der Nähe der Hypothese  $\vartheta = 0$  gegeben durch (6.8). Bei den anderen Kernfunktionen lässt sich nahe der Hypothese die Darstellung (6.7) verwenden. Die Eigenwerte, die Funktionen  $b^2$  und die resultierenden lokalen approximativen Bahadur-Steigungen sind in Tabelle 6.1 zusammengefasst. Von den vorgestellten Tests hat der Cramér-Test die beste, also größte Bahadur-Steigung, gefolgt von  $\phi(z) = \log(1 + z)$ , Bahrs Kern und den Kernfunktionen  $\phi(z) = 1 - 1/(1 + z)$  und  $\phi(z) = 1 - 1/(1 + z)^2$ . Die Berechnung für den Logarithmus und die beiden letztgenannten Kerne liefert hier, wie auch im Fall von Skalenalternativen, recht komplexe Ausdrücke, die der Übersichtlichkeit halber nur in Dezimaldarstellung angegeben sind.

Im Fall von Skalenalternativen, das heißt dem Test einer Stichprobe aus der  $N(0, \sigma^2)$ -Verteilung gegen eine Stichprobe der Standardnormalverteilung, lässt sich zur Bestimmung der Funktionen  $b^2(\theta) = b^2(1 - \sigma)$  für alle fünf Kerne die Darstellung (6.10) verwenden. Die Ergebnisse für  $b^2$  sowie die lokale Bahadur-Steigung finden sich in Tabelle 6.2. Dieses Testproblem liefert ein verdrehtes Bild: Die bei Lage-Alternativen schlechteren Kerne liefern hier deutlich bessere Steigungen als die Kernfunktion von Bahr, der  $\log(1 + z)$ -Kern oder der im Vergleich schlechteste Cramér-Kern.

Tabelle 6.2: Bahadur-Steigungen bei Skalenalternativen eindimensionaler Normalverteilungen

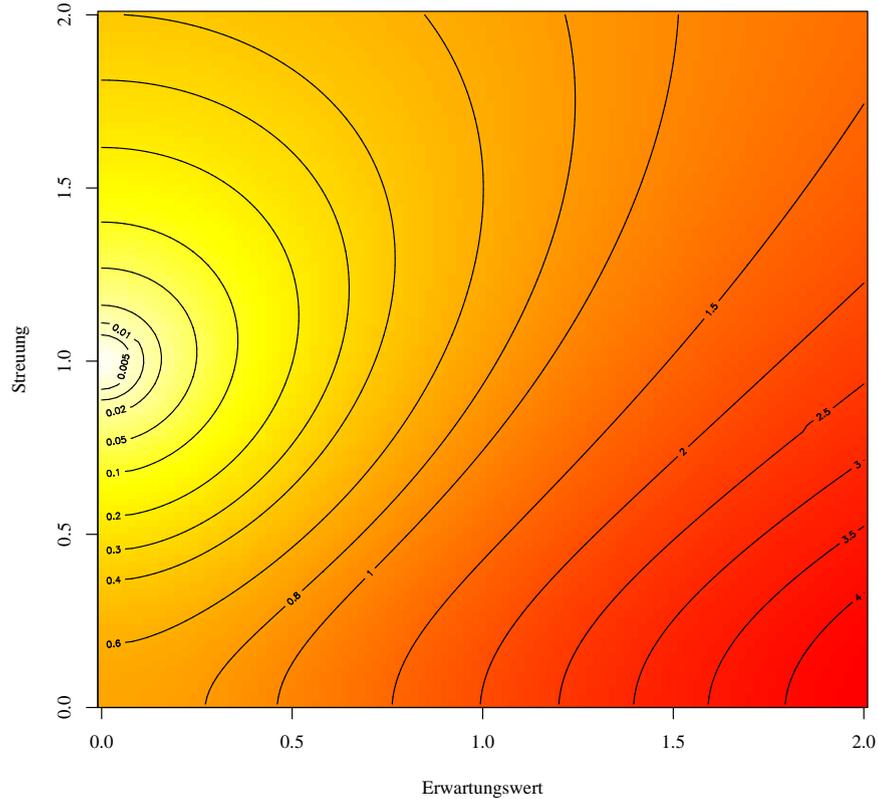
Kernfunktion	Größter EW	$\lim_{\sigma \rightarrow 1} \frac{b^2(1-\sigma)}{(1-\sigma)^2}$	$\lim_{\sigma \rightarrow 1} \frac{ab^2(1-\sigma)}{(1-\sigma)^2}$
$\sqrt{z}/2$	$a^{-1} = 0.29727$	$\frac{1}{4\sqrt{\pi}}$	0.474
$1 - \exp(-z/2)$	$a^{-1} = 0.23607$	$\frac{1}{3\sqrt{3}}$	0.815
$\log(1+z)$	$a^{-1} \approx 0.49493$	0.29513	0.596
$1 - 1/(1+z)$	$a^{-1} \approx 0.20361$	0.16506	1.234
$1 - 1/(1+z)^2$	$a^{-1} \approx 0.22492$	0.18888	1.191

Abbildung 6.1: Bahadur-Steigung beim Cramér-Kern

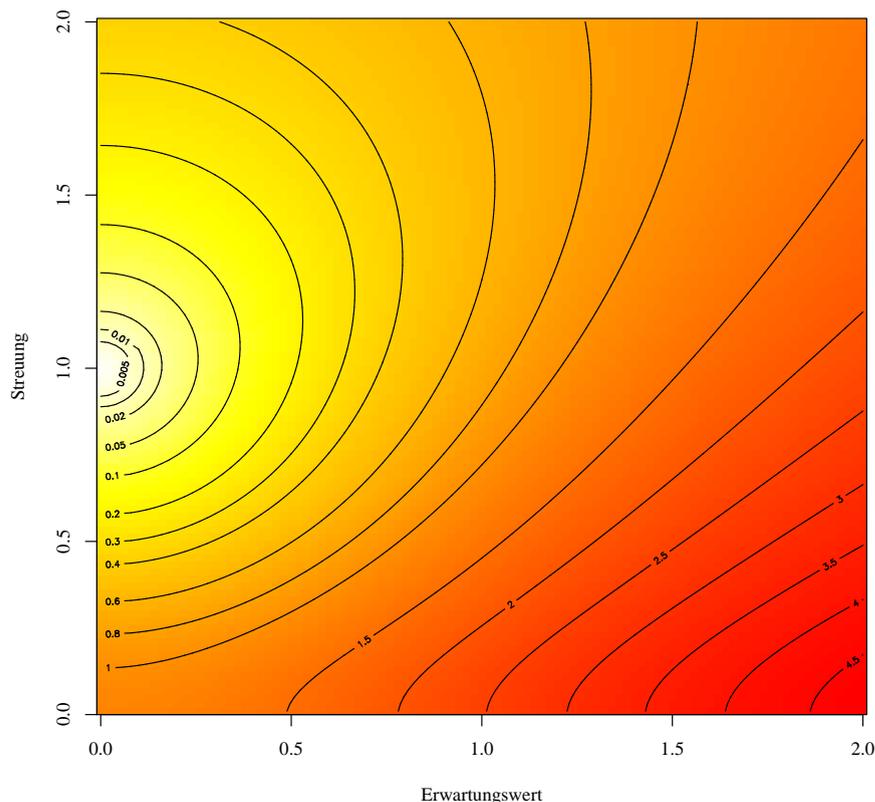


Eine sehr gute Vergleichsmöglichkeit unterschiedlicher Kernfunktionen liefert eine graphische Veranschaulichung der approximativen Bahadur-Steigung  $ab^2(\theta) = d_\phi^2(F, G)/\lambda_1$ . Dabei bezeichnet  $F$  die Verteilungsfunktion der Standardnormalverteilung und  $G$  die Verteilungsfunktion einer Normalverteilung der Alternative, deren Erwartungswert  $\mu$  und Streuung  $\sigma$  durch die Abszisse bzw. Ordinate angegeben sind. Unterschiedliche Farben markieren unterschiedliche Werte von  $ab^2(\theta)$ . Die quanti-

Abbildung 6.2: Bahadur-Steigung beim Kern von Bahr



tativen Werte sind durch Niveaulinien abzulesen. Einleuchtenderweise ergibt sich in allen Fällen in der Hypothese  $\mu = 0$ ,  $\sigma = 1$  eine Steigung von 0 und für jede Alternative ein positiver Wert. Die Niveaulinien geben bei fester Kernfunktion an, welche Verteilungen der Alternative von der Standardnormalverteilung denselben Abstand im Sinne der Bahadur-Steigung (also im Sinne von  $d_\phi^2$ ) haben. So hat beim Cramér-Kern, Abbildung 6.1, zum Beispiel eine Normalverteilung mit Erwartungswert 0 und Streuung 1/4 (Varianz 1/16) ungefähr denselben Abstand von der Normalverteilung wie eine Normalverteilung mit Erwartungswert 0.65 und Varianz 1. Die in Tabelle 6.1 bzw. 6.2 in der letzten Spalte angegebenen Werte kann man als Dichte der Niveaulinien für wachsende Erwartungswerte bzw. wachsende Streuung bei  $\mu = 0$ ,  $\sigma = 1$  interpretieren. Die vergleichsweise hohe Liniendichte in Abbildung 6.1 bei sich änderndem Erwartungswert findet sich in Tabelle 6.1 in der höchsten Bahadur-Steigung nahe der Hypothese wieder. Andererseits führen die elliptisch wirkenden Niveaulinien mit relativ großem Durchmesser in  $\sigma$ -Richtung zur im Vergleich schlechtesten Bahadur-Steigung für Skalenalternativen. Den Gegenpol bildet der Kern  $\phi(z) = 1 - 1/(1+z)$ , Abbildung 6.3. Hier führen elliptische Bahnen mit in  $\sigma$ -Richtung geringem Durchmesser zur besten Bahadur-Steigung für Skalen-

Abbildung 6.3: Bahadur-Steigung beim Kern  $\phi(z) = 1 - \frac{1}{1+z}$ 

alternativen und umgekehrt der hohe Durchmesser in  $\mu$ -Richtung zur schlechtesten Steigung für Lage-Alternativen. Ein ausgeglichenes Bild bietet der Kern von Bahr, Abbildung 6.2, bei dem die Niveaulinien kreisförmig scheinen. Dieser Eindruck wird durch die gleichen Steigungen in den Tabellen 6.1 bzw. 6.2 gefestigt.

Unterstützt durch die lokale Gleichheit von der ansonsten unter Umständen wenig aussagekräftigen approximativen Bahadur-Effizienz und der Pitman-Effizienz geben diese Ergebnisse Anlass zur Vermutung, dass sich beim Vergleich der Güten ein ähnliches Bild einstellen wird. Da sich dies mit Hilfe der Simulationen des nächsten Kapitels bestätigen wird, sei an dieser Stelle noch einmal auf die Aussagekraft dieser Visualisierungen von  $ab^2(\theta)$  hingewiesen. Bei der Untersuchung von neuen Kernfunktionen oder auch anderen Testproblemen lässt sich so ein schneller Eindruck vom Güteverhalten gewinnen. Der größte Eigenwert  $\lambda_1$  kann mit dem Verfahren aus Kapitel 4.2 problemlos approximiert werden; die Funktion  $d_\phi^2(F, G)$  lässt sich in aller Regel durch einfache Integration bzw. Summation bestimmen, so dass die Erstellung eines solchen ersten Bildes nur wenige Sekunden dauert. Dies soll allerdings nicht über die Notwendigkeit der theoretischen Untersuchung, insbesondere des Nachwei-

ses der lokalen Gleichheit der beiden erwähnten Effizienzbeurteilungen hinwegtäuschen.



# Kapitel 7

## Simulationen

In diesem Kapitel wird die Güte einiger Tests der neuen Klasse mit vielfach angewendeten, parametrischen und nichtparametrischen Tests verglichen. Zunächst betrachten wir die klassische Problemstellung der Erkennung von Lage- und Skalenalternativen bei uni- und multivariaten Normalverteilungen. Anschließend beleuchten wir das Güteverhalten beim Kontaminationsmodell mit Lehmann-Alternativen, sowohl ein- als auch mehrdimensional, näher. Für jede Güteapproximation werden 10000 Paare von Beobachtungssätzen generiert und anschließend die angegebenen Tests auf jedes Paar angewendet. Als Güteapproximation dient der jeweilige Anteil der Verwerfungen. Im gesamten Kapitel wird den Tests als Fehlerwahrscheinlichkeit erster Art das Niveau 0.05 vorgeschrieben.

### 7.1 Lage- und Skalenalternativen bei Normalverteilungen

Wir betrachten vier der bereits in Abschnitt 6.3 vorgestellten Tests aus der neuen Klasse mit den Kernfunktionen:

des Cramér-Tests, von Bahr,  $\log(1 + z)$  und  $1 - 1/(1 + z)$ .

Die Approximation des kritischen Wertes erfolgt ausschließlich mit dem Eigenwertverfahren aus Abschnitt 4.2.1.

#### 7.1.1 Der eindimensionale Fall

Die bekannten Tests, mit denen die vier neuen Verfahren verglichen werden, sind die beiden parametrischen Tests

1. der zweiseitige t-Test bei unbekannter gemeinsamer Varianz (ein gleichmäßig

bester, unverfälschter Test für den eindimensionalen Test auf Lage-Alternativen bei Normalverteilungen, siehe z.B. Lehmann [Leh86, S.201ff]) und

2. der zweiseitige F-Test bei unbekanntem Erwartungswert (bei geeigneter Wahl des Ablehnungsbereichs, welche für unsere Simulationen durchgeführt wurde, ist dies ein gleichmäßig bester, unverfälschter Test für den eindimensionalen Test auf Skalenalternativen bei Normalverteilungen, siehe z.B. Lehmann [Leh86, Kapitel 5.3, S.200])

sowie die zweiseitigen Versionen der nichtparametrischen Tests

1. der Kolmogorov-Smirnov (KS) Test (der exakte p-Wert wird mit dem vorgegebenen Niveau verglichen),
2. der Cramér-von Mises (CvM) Test (als kritischer Wert wird das  $1 - \alpha$ -Quantil der sich unter der Hypothese asymptotisch ergebenden Verteilung verwendet, der tabellierte Wert findet sich z.B. bei Owen [Owe62, S.444]),
3. Henzes Nächst-Nachbar (NN) Test (der kritische Wert wird jeweils durch 200 Monte Carlo-Bootstrap-Replikate ermittelt),
4. der Mann-Whitney (MW) Test für Lage-Alternativen (der exakte p-Wert wird mit dem vorgegebenen Niveau verglichen) und
5. der Test von Mood auf Skalenalternativen (der kritische Wert wird aus der Grenzverteilung unter der Hypothese, einer Normalverteilung, bestimmt).

Der t-Test, F-Test, der Kolmogorov-Smirnov-Test und der Cramér-von Mises-Test sind hinlänglich bekannt (siehe z.B. Lehmann [Leh86] für die beiden parametrischen Tests und Durbin [Dur73] für die beiden nichtparametrischen Tests). Die lokalen approximativen Bahadur-Steigungen des t-, F-, Kolmogorov-Smirnov- und Cramér-von Mises-Tests beim Testproblem auf Lage-Alternativen und die beim Test auf Skalenalternativen finden sich bei Wieand [Wie76]. In dieser Arbeit findet sich auch der Nachweis der Wieandschen Bedingung für die beiden Testprobleme. Die Teststatistik des Nächst-Nachbar-Tests von Henze [Hen84, Hen88] ist durch

$$\sum_{i=1}^m \mathbb{I}(\text{der nächste Nachbar von } X_i \text{ gehört zur ersten Stichprobe}) \\ + \sum_{j=1}^n \mathbb{I}(\text{der nächste Nachbar von } Y_j \text{ gehört zur zweiten Stichprobe})$$

gegeben (der nächste Nachbar ist jeweils die Beobachtung, die den geringsten Abstand in der euklidischen Norm aufweist). Standardisiert man die Testgröße, so konvergiert sie bei Gültigkeit der Hypothese gegen eine Normalverteilung. Mit den Ergebnissen von Henze [Hen84] zeigt man durch Bestimmung der Funktion  $b$  in Definition 6.1, (c), leicht, dass der Test sowohl für Lage-, als auch für Skalenalternativen

eine approximative Bahadur-Steigung der Größenordnung  $\mathcal{O}(\vartheta^4)$  für  $\vartheta \rightarrow 0$  bzw.  $\mathcal{O}((1 - \sigma)^4)$  für  $\sigma \rightarrow 1$  hat. Die approximative relative Bahadur-Effizienz bezüglich jedes im Weiteren für Lage-/Skalen-Alternativen betrachteten Tests ist daher 0. Für Lage-Alternativen betrachten wir zusätzlich die zum Rangsummen-Test von Wilcoxon äquivalente Teststatistik von Mann und Whitney,

$$\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \mathbb{I}(X_i \leq Y_j).$$

Die Berechnung des exakten p-Wertes folgt Hollander und Wolfe [HW73, Chapter 4]. Der zweiseitige Mann-Whitney Test hat sich als nichtparametrischer Test bei Lage-Alternativen bewährt und verfügt über die vergleichsweise hohe lokale approximative Bahadur-Steigung von  $3/\sqrt{\pi}\vartheta^2 \approx 0.955\vartheta^2$  (siehe dazu und zur lokalen Gleichheit der Effizienzen die Arbeit von Hollander [Hol67]). Der Test von Mood [Moo54] ist ein nichtparametrischer Test auf Skalenalternativen. Als Teststatistik dient

$$\sum_{j=1}^n \left( r_{y_j} - \frac{m+n+1}{2} \right)^2,$$

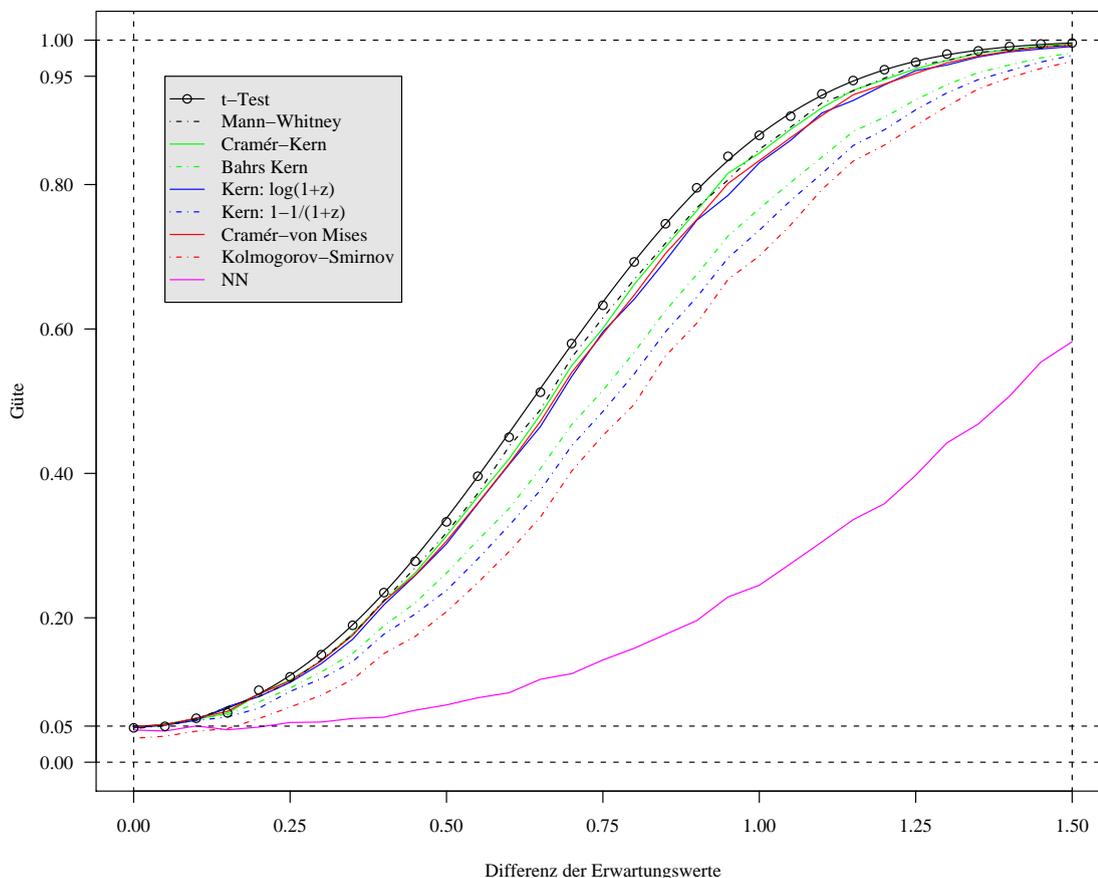
wobei  $r_{y_j}$  den Rang der  $j$ -ten Beobachtung der zweiten Stichprobe innerhalb der zusammengesetzten Stichprobe bezeichnet. Die zugehörige lokale approximative Bahadur-Steigung  $c(\sigma) = \frac{15}{\pi^2}(1 - \sigma)^2$  kann man aus Mood's Arbeit [Moo54] ableiten.

Die Tabellen 7.1 und 7.2 fassen die Ergebnisse zur approximativen Bahadur-Steigung der zu Beginn dieses Kapitels erwähnten Referenztests zusammen. Dabei gibt die zweite Spalte die lokale approximative Bahadur-Steigung für  $\vartheta \rightarrow 0$  in Tabelle 7.1 und für  $\sigma \rightarrow 1$  in Tabelle 7.2 an.

Für den Test auf Lage-Alternativen werden für die erste Stichprobe 20 Beobachtungen  $N(0, 1)$ -verteilter Zufallsvariablen und für die zweite Stichprobe 20 Beobachtungen  $N(\vartheta, 1)$ -verteilter Zufallsvariablen generiert, wobei alle Zufallsvariablen unabhängig seien und  $\vartheta$  in Schritten von 0.05 von 0 bis 1.5 läuft. In Abbildung 7.1 ist die approximierten Güte gegen den Lage-Parameter  $\vartheta$  aufgetragen. Lediglich für den t-Test ist die exakte Güte durchgezogen dargestellt, die approximierten Werte sind gepunktet und verlaufen in geringem Abstand zur exakten Güte. Die Anzahl der Wiederholungen zur Approximation der Güte erscheint daher angebracht.

Die Abbildung bestätigt das durch die Tabellen 6.1 und 7.1 suggerierte Bild. Der Mann-Whitney Test, der Cramér-von Mises Test und die zwei Tests der neuen Familie mit den größten approximativen Bahadur-Steigungen, der Cramér Test und der Test mit dem Kern  $\phi(z) = \log(1 + z)$ , haben nicht nur annähernd die gleiche approximative Bahadur-Steigung, sondern darüber hinaus ungefähr denselben Verlauf der Güte. Bei sorgfältiger Betrachtung kann man für Lagedifferenzen zwischen 0.5 und 1.5 erkennen, dass die Güte des Cramér-von Mises Tests und dem neuen Test mit Kern  $\phi(z) = \log(1 + z)$  etwas unter der Güte des Mann-Whitney Tests und dem

Abbildung 7.1: Vergleich der Güte bekannter Testverfahren bei Lage-Alternativen



Cramér-Test liegen, was zumindest in der Nähe der Hypothese durch die Werte in den Tabellen gestützt wird. Mit schlechterer Güte folgen diesen Tests die Tests mit Bahrs Kern, dem Kern  $\phi(z) = 1 - 1/(1+z)$  und des Kolmogorov-Smirnov-Tests. Auch diese Reihenfolge entspricht der Anordnung der Steigungen in den Tabellen 6.1 und 7.1. Die im Vergleich kleinste Güte hat Henzes Nächst-Nachbar Test, dessen Bahadur-Steigung für dieses Problem 0 ist. Die schlechte Ausschöpfung des Niveaus des Kolmogorov-Smirnov-Tests lässt sich durch die geringe Anzahl von unterschiedlichen Werten, die die Teststatistik annehmen kann, sowie die nicht durchgeführte Randomisierung begründen. Angenehm fällt auf, dass der bei diesem Testproblem bewährte Mann-Whitney Test keine erkennbar bessere Güte besitzt als der Cramér-Test. Betrachtet man lediglich die Tests der neuen Familie, so ist qualitativ der direkte Zusammenhang zu den am Ende des letzten Kapitels erläuterten Abbildungen 6.1, 6.2 und 6.3 erkennbar.

Das vergleichsweise gute Abschneiden der Tests der neuen Klasse beim Test auf Lage-Alternativen lässt sich durch Gleichung (2.1) erklären, welche der Teststatistik zu Grunde liegt. Man sieht, dass die Verschiebung einer von zwei identischen Verteilungen ausschließlich und unmittelbar den ersten Term beeinflusst (denn die

Tabelle 7.1: Bahadur-Steigungen beim Test auf Lage-Alternativen bei Normalverteilungen

Test	$\lim_{\vartheta \rightarrow 0} c(\vartheta)/\vartheta^2$
t-Test	1
MW	0.955
CvM	0.907
KS	0.637
NN	0

Tabelle 7.2: Bahadur-Steigungen beim Test auf Skalenalternativen bei Normalverteilungen

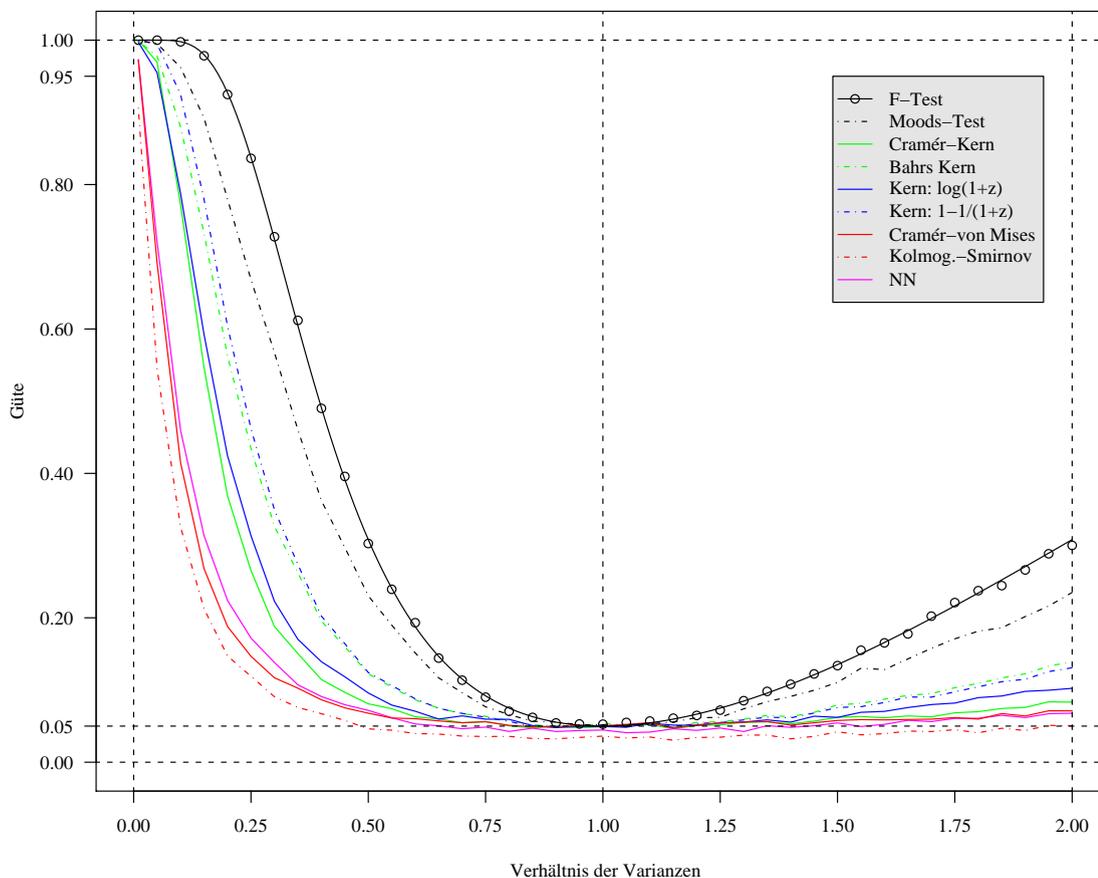
Test	$\lim_{\sigma \rightarrow 1} \frac{c(\sigma)}{(1-\sigma)^2}$
F-Test	2
Mood	1.520
CvM	0.302
KS	0.234
NN	0

beiden anderen Terme sind invariant gegenüber einer solchen Verschiebungen). Das Argument von  $\phi$  wird im Mittel größer und dies führt zu einer größeren Teststatistik (siehe dazu auch die Herleitung der zugehörigen Bahadur-Steigungen).

Auch beim Test auf Skalenalternativen betrage der Stichprobenumfang je Stichprobe 20. Die erste Stichprobe entspringe wieder einer  $N(0, 1)$ -Verteilung und die zweite einer  $N(0, \sigma^2)$  Verteilung. Die Darstellungsweise der Ergebnisse in Abbildung 7.2 ist dieselbe wie in der vorangegangenen Abbildung: die approximierte Güte ist aufgetragen gegen den Parameter der Alternative, hier  $\sigma^2 = 0.01, 0.05, 0.1, 0.15, \dots, 1.95, 2$ . Die exakte Güte des F-Tests ist wiederum durchgezogen gezeichnet. Die in geringem Abstand verlaufenden, gepunkteten Approximationen derselben geben Grund zu der Annahme, dass ausreichend viele Simulationen durchgeführt wurden.

Wie im Fall von Lage-Alternativen wird der Eindruck gefestigt, der sich durch die Tabellen 6.2 und 7.2 gewinnen lässt. Demnach hat der F-Test als gleichmäßig bester, unverfälschter Test für dieses Problem die größte Güte, gefolgt von Mood's nichtparametrischen Test auf Skalenalternativen. Es folgen die vier Beispiele der Klasse neuer Tests und mit den im Vergleich kleinsten Güten der Nächst-Nachbar Test vor den beiden klassischen nichtparametrischen Tests, dem Cramér-von Mises- und dem Kolmogorov-Smirnov-Test. Die Beispiele der neuen Klasse lassen sich in zwei Paare aufteilen. Zum einen sind dies die für Lage-Alternativen etwas besseren und bei dem vorliegenden Test auf Skalenalternativen etwas schlechteren Tests mit durchgezogen

Abbildung 7.2: Vergleich der Güte bekannter Testverfahren bei Skalenalternativen



gezeichneter Güte, der Cramér-Test und der Test mit Kern  $\phi(z) = \log(1+z)$ . Zum anderen sind dies die bei Lage-Alternativen etwas schlechteren, dafür beim vorliegenden Problem etwas besseren Tests mit gestrichelt gezeichneter Güte, der Test von Bahr und der Test mit Kern  $\phi(z) = 1 - 1/(1+z)$ . Für den Bereich  $\sigma^2 \in (0, 1]$  ergibt sich die durch die Tabellen vorgegebene Reihenfolge. Im Bereich  $\sigma^2 > 1$  ist interessanterweise die Güte des Tests von Bahr etwas höher als die des Tests mit Kern  $\phi(z) = 1 - 1/(1+z)$ , obwohl die Bahadur-Steigung des letzteren mit Größenordnung  $1.234(1-\sigma)^2$  deutlich höher ist als die des Tests von Bahr mit  $0.815(1-\sigma)^2$ . Die Erklärung dafür lässt sich unmittelbar den Abbildungen 6.2 und 6.3 entnehmen. Während für eine Streuung der zweiten Stichprobe kleiner 1 die Niveaulinien von  $\phi(z) = 1 - 1/(1+z)$  dichter liegen, ist es bei einer Streuung größer 1 der Test von Bahr, welcher (mittelfristig) schneller steigt. Dies untermauert den Charakter der lokalen Bahadur-Steigung mit starker Aussagekraft nur nahe der Hypothese. Entfernt man sich von dieser, so spielen die Terme größerer Ordnung wieder eine nicht zu vernachlässigende Rolle.

Für die vorgestellten, klassischen, eindimensionalen Problemstellungen lässt sich zusammenfassend festhalten, dass die neuen Tests einen erfolgversprechenden Mit-

telweg zwischen hervorragenden Tests auf Lage-Alternativen, wie dem von Mann-Whitney, und sehr guten Tests für Skalenalternativen, wie z.B. dem Test von Mood, beschreiten. Möchte man bei der Wahl einer Kernfunktion den beiden Möglichkeiten für Alternativen gleichzeitig Rechnung tragen, so empfiehlt sich die Verwendung des Kerns von Bahr,  $\phi(z) = 1 - \exp(-z/2)$ , bei stärkerem Gewicht auf Lage-Alternativen bietet sich der Cramér Test bzw. die Verwendung des Logarithmus als Kernfunktion an.

### 7.1.2 Der mehrdimensionale Fall

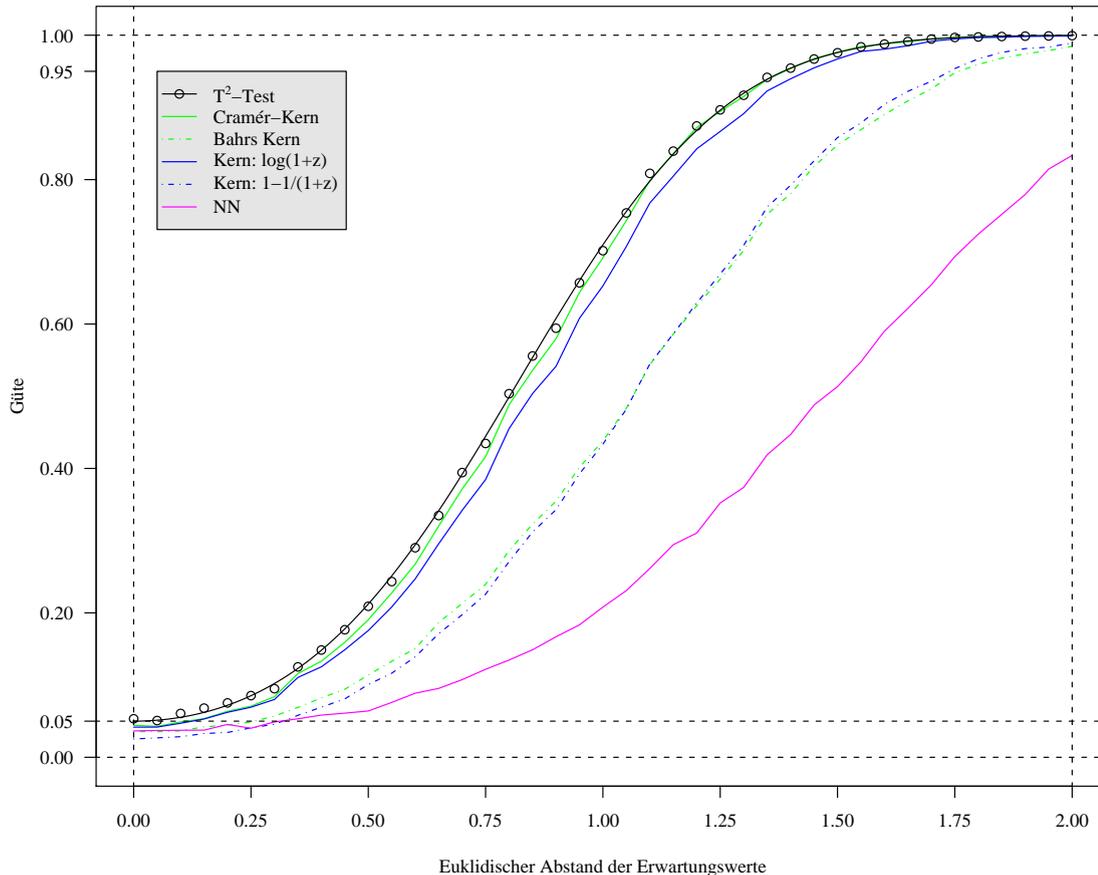
Für multivariat normalverteilte Daten bieten sich folgende parametrische Tests für die Untersuchung auf Lage- bzw. Skalenalternativen an:

1. Hotellings  $T^2$ -Test bei unbekannter gemeinsamer Kovarianzmatrix (als kritischer Wert kommt das entsprechend modifizierte Quantil einer F-Verteilung zum Einsatz) und
2. Bartletts Modifikation des Likelihood-Quotienten-Tests (der kritische Wert kann durch Entwicklung der asymptotischen Verteilung gewonnen werden; dies wird beispielsweise bei Muirhead [Mui82, Theorem 8.2.7, S.309] vorgeführt).

Als nichtparametrischer Test, der für mehrdimensionale Daten geeignet ist, kommt HENZES NÄCHST-NACHBAR TEST zum Einsatz.

Alle Elemente der neuen Familie von Zwei-Stichproben-Tests sind invariant gegenüber orthogonalen Transformationen und Translationen aller Beobachtungen. Diese Eigenschaft erfüllt auch der Test von Hotelling sowie der Nächst-Nachbar Test. Beschränkt man sich beim Test auf Lage-Alternativen bei der Annäherung an die Hypothese auf eine feste Richtung, so ist es daher ausreichend, lediglich genau eine Richtung zu betrachten. Im vorliegenden Fall wird ausschließlich die erste Komponente verändert. Abbildung 7.3 zeigt die Güte des  $T^2$ -Tests, des Nächst-Nachbar Tests sowie der vier Beispiele aus der neuen Familie aufgetragen gegen den euklidischen Abstand der Erwartungswerte bei Dimension  $d = 3$ , wobei die Güte im Abstand von 0.05 approximiert wird. Für andere Dimensionen ergeben sich vergleichbare Ergebnisse.

Wie im Eindimensionalen beträgt der Stichprobenumfang jeweils 20. Dies erklärt auch die schlechte Ausschöpfung des Niveaus bei Bahrs Kern und dem Kern  $1 - 1/(1+z)$ , da sich mit dem Eigenwertverfahren bei höheren Dimensionen und geringem Stichprobenumfang Approximationsprobleme ergeben (siehe dazu auch Kapitel 4.2.2). Abgesehen von dieser Eigenschaft ergibt sich dasselbe Bild wie im Eindimensionalen. Die beiden letztgenannten Tests liegen mit ihrer Güte hinter dem Cramér Test und dem Test mit Kern  $\log(1+z)$ , aber noch über der des Nächst-Nachbar Tests. Der gleichmäßig beste Test, der invariant gegenüber Multiplikation

Abbildung 7.3: Vergleich der Güte bei Lage-Alternativen im Mehrdimensionalen,  $d = 3$ 

aller Daten mit einer regulären Matrix und Translationen ist, der  $T^2$ -Test von Hotelling, hat nur eine unwesentlich größere Güte als der Cramér Test und der neue Test mit Logarithmus-Kern. Die in Abbildung 7.3 gepunktete, approximative Güte des  $T^2$ -Tests liegt auch hier sehr nah bei der exakten, durchgezogen gezeichneten Güte des  $T^2$ -Tests.

Es sei noch bemerkt, dass die neuen Tests nicht über die angenehme affine Invarianz des  $T^2$ -Tests verfügen. Zu jedem der Tests kann man jedoch eine affin invariante Variante konstruieren, indem man die Beobachtungen empirisch standardisiert. Dazu definiert man

$$U_i := (S_{m,n})^{-\frac{1}{2}} X_i,$$

$$V_j := (S_{m,n})^{-\frac{1}{2}} Y_j,$$

wobei

$$S_{m,n} := \frac{1}{m+n} \left( \sum_{i=1}^m (X_i - \bar{Z})(X_i - \bar{Z})' + \sum_{j=1}^n (Y_j - \bar{Z})(Y_j - \bar{Z})' \right)$$

und  $\bar{Z}$  der Mittelwert der zusammengelegten Beobachtung ist. In der bisherigen Testgröße ersetzt man nun die Beobachtungen der ersten bzw. zweiten Stichprobe durch ihre standardisierten Entsprechungen  $U_i$  bzw.  $V_j$ . Man überlegt sich leicht, dass der resultierende Test affin invariant ist.

Bei der Untersuchung von Skalenalternativen bei mehrdimensionalen Normalverteilungen verwenden wir als Referenztest die modifizierte Fassung des Likelihood-Quotienten-Tests auf identische Kovarianzmatrix von Bartlett [Bar37]. Der Test ist invariant gegenüber Translationen des Koordinatensystems (insbesondere sogar innerhalb einer Stichprobe, da lediglich die Kovarianzmatrix eine Rolle spielt) und gegenüber der Multiplikation aller Beobachtungen mit einer regulären Matrix. Auch ist der Test unverfälscht. Leider handelt es sich nicht um einen gleichmäßig besten unverfälschten oder einen gleichmäßig besten invarianten Test.

Wir untersuchen nun zwei spezielle Skalenalternativen für den Fall vierdimensional normalverteilter Daten. Dazu sei in beiden Fällen die erste Stichprobe aus der

$N_4(0, \text{diag}(1, 1, 1, 1))$ -Verteilung.

Für die zweite Stichprobe betrachten wir zum einen

(a) die  $N_4(0, \text{diag}(\Delta, \Delta, \Delta, \Delta))$ -Verteilung und zum anderen

(b) die  $N_4(0, \text{diag}(\Delta, \Delta, 1, 1))$ -Verteilung,

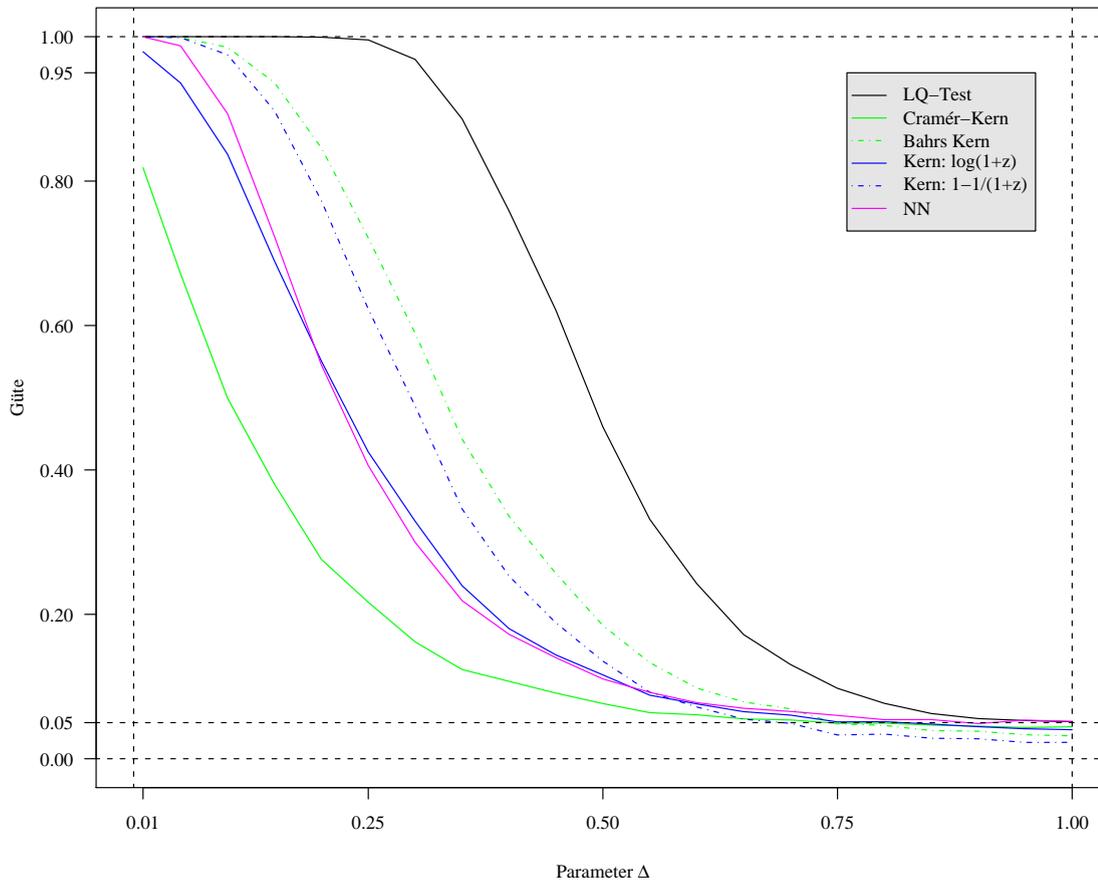
wobei jeweils  $\Delta \in (0, 1]$ . Es bezeichne  $\text{diag}(a_1, \dots, a_d)$  die Diagonalmatrix des  $\mathbb{R}^{d \times d}$  mit den Einträgen  $a_1, \dots, a_d \in \mathbb{R}$  in der Diagonalen.

Für die Approximation der Güte werden nun für jede Verteilung 40 Beobachtungen unabhängiger Zufallsvariablen mit der entsprechenden Verteilung generiert und das Testresultat ermittelt. Die Güteapproximation ist der Anteil der Verwerfungen bei 10000 Durchführungen. Der Parameter  $\Delta$  durchläuft dazu die Werte

$$\{0.01, 0.05, 0.1, 0.15, \dots, 0.95, 1\}.$$

Im Fall (a) ergibt sich ungefähr dasselbe Bild wie in Abbildung 7.2. Die Güte des LQ-Tests ist am größten. Es folgen die beiden auch im Eindimensionalen für Skalenalternativen bewährten Tests von Bahr und der mit Kern  $1 - 1/(1+z)$ . Mit einigem Abstand schließen sich die eher für Lage-Alternativen geeigneten Tests mit Kern  $\log(1+z)$  und der Cramér Test an. An letzter Stelle folgt der Nächst-Nachbar Test, für welchen in diesem Fall der von Henze angegebene asymptotische kritische Wert verwendet wurde.

Erwähnenswert ist das Ergebnis von Untersuchung (b), Abbildung 7.4. Die Verwerfungswahrscheinlichkeit ist dabei gegen unterschiedliche Werte  $\Delta$  aufgetragen. Entgegen aller bisherigen Untersuchungen schneidet nun der Nächst-Nachbar-Test

Abbildung 7.4: Skalernalternativen im Mehrdimensionalen, Fall (b),  $d = 4$ 

von Henze speziell bei Werten  $\Delta$ , die weit von der Hypothese entfernt liegen, besser ab als die beiden bei Lage-Alternativen besseren Tests der neuen Klasse. Die für Skalernalternativen favorisierten Kerne bleiben jedoch besser. Die vergleichsweise schlechte Ausschöpfung des Niveaus ist auf den in Abschnitt 4.2.2 beschriebenen Effekt zurückzuführen.

Zusammenfassend lässt sich auch im multivariaten Fall sagen, dass die neuen Tests ein akzeptables Güteverhalten zeigen, welches dem des eindimensionalen Falles ähnelt. Die für die verschiedenen Testprobleme bevorzugten Tests bewähren sich auch hier.

## 7.2 Mischung mit Lehmann-Alternativen

Eine häufig untersuchte Form von Alternativen in der nichtparametrischen Statistik sind die so genannten Lehmann-Alternativen, welche 1953 von Lehmann [Leh53] im Zusammenhang mit Rang-Tests untersucht wurden. Man versteht darunter Verteilungsfunktionen als Alternative, die eine Potenz der Verteilungsfunktion der Hy-

pothese sind, also die Form  $F^k$  hat,  $k \in \mathbb{N}$ ,  $k \geq 2$ , bekannt. Wir betrachten nun ein Kontaminationsmodell, bei dem die Verteilung, mit der kontaminiert wird, eine Lehmann-Alternative ist. Für das Kontaminationsmodell haben wir bereits in den Abschnitten 6.1.4 und 6.2.2 die approximative Bahadur-Effizienz bzw. die lokale Gleichheit derselben mit der Pitman-Effizienz gezeigt. Konkret untersuchen wir hier mit Hilfe einer Simulation den Fall  $k = 2$ , also den Test der Hypothese

$$H : X, Y \text{ haben dieselbe Verteilungsfunktion } F$$

gegen die Alternative

$$K : X \sim F, Y \sim (1 - \theta)F + \theta F^2, \quad \theta \in (0, 1].$$

### 7.2.1 Der eindimensionale Fall

Lehmann zeigte in seiner Arbeit, dass der (einseitige) Test von Wilcoxon lokal bester Rang-Test für dieses Testproblem ist. Da wir wollen, dass man die Rollen der ersten und zweiten Stichprobe vertauschen kann, ohne ein anderes Ergebnis zu erhalten, verwenden wir den in Abschnitt 7.1.1 vorgestellten, zweiseitigen Rangsummen-Test von Wilcoxon bzw. Mann-Whitney. Daneben werden die weiteren nichtparametrischen Tests in den Vergleich einbezogen: der Kolmogorov-Smirnov-Test, der Cramér-von Mises-Test, der Nächst-Nachbar-Test sowie natürlich die bereits vorgestellten unterschiedlichen Beispiele der neuen Familie von Tests. Als Verteilung für den eindimensionalen Fall wählen wir die Gleichverteilung auf dem Einheitsintervall  $[0, 1]$ . Um zu verhindern, dass die Güte selbst in der Nähe von  $\theta = 1$  sehr klein ist, werden jeweils Stichproben vom Umfang 50 untersucht. Neben der Hypothese  $\theta = 0$  simulieren wir die Fälle  $\theta \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$  jeweils 10000 Mal und verwenden den Verwerfungsanteil als Approximation für die Güte. Die Ergebnisse sind in Tabelle 7.3 zusammengefasst.

Tabelle 7.3: Güte für unterschiedliche  $\theta$  im Kontaminationsmodell

$\theta$	Cramér	$\log(1 + z)$	Bahr	$1 - 1/(1 + z)$	MW	KS	CvM	NN
0.0	0.053	0.054	0.054	0.053	0.051	0.038	0.053	0.045
0.2	0.089	0.093	0.094	0.092	0.090	0.066	0.089	0.047
0.4	0.207	0.212	0.213	0.211	0.210	0.153	0.204	0.055
0.6	0.420	0.433	0.433	0.430	0.416	0.326	0.408	0.080
0.8	0.650	0.666	0.667	0.661	0.635	0.521	0.629	0.112
1.0	0.859	0.873	0.875	0.869	0.833	0.738	0.829	0.177

Unter den genannten Tests haben die vier Beispiele der neuen Testfamilie, die untereinander ein nahezu identisches Güteverhalten besitzen, insbesondere für größere  $\theta$  die beste Güte. Es folgt der Mann-Whitney-Rangsummen-Test. Fast genauso

gut schneidet der Cramér-von Mises-Test ab. Eine etwas schlechtere Güte weist der Kolmogorov-Smirnov-Test auf. Das Schlusslicht in diesem Vergleich bildet der Nächst-Nachbar-Test.

Dieses Bild bestätigt sich lediglich in groben Zügen, wenn man die approximativen Bahadur-Steigungen ausrechnet. Der Cramér-Test, der im eindimensionalen Fall die alternative Darstellung

$$T_{m,n}^{\text{Cramér}} = \frac{mn}{m+n} \int_{-\infty}^{+\infty} [F_m(u) - G_n(u)]^2 du$$

hat ( $F_m, G_n$  die empirische Verteilungsfunktion der ersten, zweiten Stichprobe), besitzt unter der Hypothese, dass beiden Stichproben eine Gleichverteilung auf  $[0, 1]$  zu Grunde liegt, die asymptotische Verteilung

$$\int_0^1 [\mathbb{B}(u)]^2 du,$$

wobei  $\mathbb{B}$  eine Brownsche Brücke bezeichnet (vergleiche Kapitel 1.1). Dies ist auch die Grenzverteilung der Cramér-von Mises-Statistik, die bekanntlich als Summe von  $\chi_1^2$ -verteilten Zufallsvariablen mit den Vorfaktoren  $(j\pi)^{-2}$ ,  $j \in \mathbb{N}$ , geschrieben werden kann (vergleiche dazu Kapitel 5.3.2). Die Konstante  $a$  aus der Definition der Standardfolge ist also für beide Verteilungen  $\pi^2$ . Der Wert  $b^2(\theta)$  lässt sich im Fall des Cramér Tests schreiben als

$$b^2(\theta) = \int_0^1 (F(x) - (1-\theta)F(x) - \theta F^2(x))^2 dx = \int_0^1 (x - (1-\theta)x - \theta x^2)^2 dx = \frac{\theta^2}{30}.$$

Im Fall des Cramér-von Mises-Tests ergibt sich

$$\begin{aligned} b^2(\theta) &= \int_0^1 (F(x) - (1-\theta)F(x) - \theta F^2(x))^2 d\left(\frac{F(x) + (1-\theta)F(x) + \theta F^2(x)}{2}\right)(x) \\ &= \frac{\theta^2}{30}. \end{aligned}$$

Die zunächst verblüffende Gleichheit lässt sich unter anderem mit Hilfe der Hoffding Formel für die Berechnung der Verteilung der Ränge unter Alternativen [WN70, S.126] erklären. Für diese beiden Tests ergibt sich  $ab^2(\theta) = \pi^2\theta^2/30$ . Für den Kolmogorov-Smirnov-Test ergibt sich mit dem Ergebnis von Bahadur [Bah60] die Steigung  $ab^2(\theta) = 4(\theta/4)^2 = \theta^2/4$ . Beim Nächst-Nachbar-Test lässt sich direkt die Formel des Erwartungswertes der Statistik von Henze [Hen84] verwenden, um zu zeigen, dass die Steigung von der Größenordnung  $\mathcal{O}(\theta^4)$  für  $\theta \rightarrow 0$  ist. Die approximative Bahadur-Steigung der drei verbleibenden Kernfunktionen ist als Produkt des Reziproken einer Approximation des größten Eigenwertes und dem explizit berechenbaren Ausdruck  $\theta^2 d_\phi^2(F, F^2)$  leicht numerisch zu ermitteln. Der Rangsummen-Test von Mann-Whitney, hat die approximative Bahadur-Steigung  $c(\theta) = \frac{1}{3}\theta^2$ , wie man

Tabelle 7.4: Bahadur-Steigungen im Kontaminationsmodell

Test	$\lim_{\theta \rightarrow 0} c(\theta)/\theta^2$
Mann-Whitney	0.3333
Cramér	0.3290
$\log(1+z)$	0.3326
Bahr	0.3330
$1 - 1/(1+z)$	0.3309
CvM	0.3290
KS	0.2500
NN	0

leicht aus den Ergebnissen von Hodges und Lehmann [HL56] ableitet. Zusammengefasst ergibt sich Tabelle 7.4.

Obwohl der Mann-Whitney-Test in diesem Vergleich die größte Steigung und somit auch die bei Vergleichen beste relative Effizienz hat, sind die Güten der vier neuen Tests bei  $\theta = 0.8$  beziehungsweise insbesondere  $\theta = 1$  teilweise größer. Eine Erklärung bietet die lediglich lokale Gleichheit, d.h. Gleichheit in der Nähe der Hypothese, der approximativen Bahadur-Effizienz mit der aussagekräftigeren Pitman-Effizienz. Für kleinere Werte von  $\theta$  lässt sich mit Hilfe von Simulationen nur schwierig der Nachweis für die aus der Effizienz-Theorie erhofften Güten erbringen, da bereits die Unsicherheit in den kritischen Werten das Ergebnis für eine derart präzise Analyse zu stark verfälscht (vergleiche dazu auch Abschnitt 4.2.2 über die Einhaltung des Niveaus bei Verwendung der Eigenwertapproximation).

Wir halten fest, dass sich die neuen Tests erfolgreich auf andere Probleme als Lage- und Skalenalternativen anwenden lassen. Als letztes Simulationsbeispiel folgt nun eine mehrdimensionale Variante des eben geschilderten Tests.

### 7.2.2 Der mehrdimensionale Fall

Wir verallgemeinern das soeben dargestellte Modell auf den mehrdimensionalen Fall, speziell den Fall der Dimension 3, bei dem die Gleichverteilung auf dem dreidimensionalen Einheitswürfel mit Verteilungsfunktion  $F(x, y, z) = xyz$ ,  $x, y, z \in [0, 1]$ , durch die Verteilung mit Verteilungsfunktion  $G(x, y, z) = F^2(x, y, z) = x^2y^2z^2$ ,  $x, y, z \in [0, 1]$ , kontaminiert wird. Die Komponenten des Zufallsvektors  $(X, Y, Z)$  werden dabei sukzessive durch Bestimmung ihrer Randverteilung beziehungsweise bedingten Randverteilung generiert. Neben den Beispielen für die neue Testfamilie betrachten wir den Nächst-Nachbar-Test. Wie zuvor beträgt der Stichprobenumfang für beide Stichproben jeweils 50. Für  $\theta \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$  werden jeweils 10000 Paare von Beobachtungssätzen erstellt, auf welche die Tests angewendet wer-

den, und als Güte der Verwerfungsanteil verwendet. Eine Übersicht über die Approximationen gibt Tabelle 7.5.

Tabelle 7.5: Güte für unterschiedliche  $\theta$  im mehrdimensionalen Kontaminationsmodell

$\theta$	Cramér	$\log(1 + z)$	Bahr	$1 - 1/(1 + z)$	NN
0.0	0.050	0.051	0.050	0.051	0.054
0.2	0.117	0.121	0.121	0.120	0.070
0.4	0.358	0.362	0.361	0.364	0.108
0.6	0.715	0.719	0.718	0.718	0.192
0.8	0.944	0.948	0.948	0.945	0.355
1.0	0.998	0.999	0.998	0.998	0.625

Zusammengefasst lässt sich sagen, dass sich für dieses Testproblem das Güteverhalten der neuen Tests sehr ähnelt.

Bei allen näher untersuchten Testproblemen hat sich gezeigt, dass die neuen Tests gut gegenüber den bekannten und bislang verwendeten Tests abschneiden. Speziell im Mehrdimensionalen liefern sie hoffnungsvoll erscheinende neue Möglichkeiten zur Bearbeitung des Zwei-Stichproben-Problems. Die gemachten Untersuchungen erstrecken sich über die Dimensionen 1 bis 4. Es sei noch bemerkt, dass sich für höhere Dimensionen und größere Stichprobenumfänge durchaus numerische Probleme bei den Additionen in der Testgrößenbestimmung ergeben können, denen man in diesen Fällen besonderes Augenmerk widmen sollte.

# Kapitel 8

## Die neuen Tests als Anpassungstests

In den letzten Kapiteln wurde die Familie der neuen Zwei-Stichproben-Tests auf Effizienz- und Güteverhalten abgeklopft. Es stellte sich im Vergleich zu bewährten Standardverfahren ein mehr als akzeptables Bild heraus. Es liegt nun nahe zu versuchen, die neuen Tests auch auf den Ein-Stichproben-Fall, das so genannte Anpassungsproblem anzuwenden.

### 8.1 Anpassung an eine vollständig festgelegte Verteilung

Es sei  $F_0$  eine gegebene Verteilungsfunktion eines  $d$ -dimensionalen Zufallsvektors. Weiter seien  $x_1, \dots, x_m \in \mathbb{R}^d$  Realisierungen von unabhängigen Zufallsvariablen  $X_1, \dots, X_m$  mit unbekannter Verteilung  $F$ . Wir betrachten das Testproblem

$$\begin{aligned} H &: F = F_0 \quad \text{gegen die Alternative} \\ K &: F \neq F_0. \end{aligned}$$

Ist  $\phi \in \Phi$  so gewählt, dass  $F_0 \in \mathcal{F}_d^2(\phi)$ , und bezeichnet man die zufällige empirische Verteilungsfunktion der  $m$  Zufallsvariablen  $X_1, \dots, X_m$  mit  $F_m$ , so hat die Größe

$$md_\phi^2(F_m, F_0) \tag{8.1}$$

unter der Hypothese  $H$  asymptotisch für  $m \rightarrow \infty$  dieselbe Verteilung wie

$$\sum_{\sigma} \lambda_{\sigma} Z_{\sigma}^2.$$

Dabei bezeichnet  $(\lambda_{\sigma})_{\sigma}$  die Folge der Eigenwerte des Operators

$$\begin{aligned} L_2(\mathbb{R}^d, \mathcal{B}^d, F_0) &\rightarrow L_2(\mathbb{R}^d, \mathcal{B}^d, F_0) \quad , \\ H : f &\mapsto \int h(\cdot, y) f(y) dF_0(y) \end{aligned}$$

mit Kern  $h(x, y) := E\bar{h}(x, Y_1; y, Y_2)$ ,  $Y_1, Y_2 \sim F_0$ , und  $(Z_\sigma)_\sigma$  eine Folge unabhängiger standardnormalverteilter Zufallsvariablen. Dies ist die Aussage von Satz 3.8. Die Eigenwerte hängen nur von dem anfangs gewählten  $\phi$  und der bekannten Verteilung  $F_0$  ab, auf die getestet wird, und können mit Hilfe des in Abschnitt 4.2.1 vorgestellten Approximationsverfahrens ermittelt werden. Verwendet man (8.1) als Testgröße für das gestellte Ein-Stichproben-Problem, so lassen sich also die asymptotischen kritischen Werte für jede Verteilung  $F_0$ , die den genannten Regularitätsbedingungen genügt, problemlos tabellieren. Offensichtlich ergibt sich ein gegen jede Alternative konsistentes Verfahren, von dem man nach den bisherigen Resultaten für den Zwei-Stichproben-Fall ein gutes Abschneiden beim Vergleich der Güte mit der klassischen Testverfahren erwarten darf. Die Berechnung der approximativen Bahadur-Effizienzen gestaltet sich nicht nur ähnlich einfach wie im Zwei-Stichproben-Fall sondern erübrigt sich teilweise, da sich die Ergebnisse beispielsweise für Lage- und Skalenalternativen direkt übernehmen lassen.

Ein Hemmnis, welches der Anwendung der verlockend erscheinenden Tests im Weg steht, ist die nun aufwändige Berechnung der Testgröße. Während im Zwei-Stichproben-Fall die Testgröße in einfach zu bestimmende Summen zerfiel, vergleiche Gleichung (4.6), so ergibt sich hier mit  $X_1, X_2 \sim F_0$  unabhängig und  $\hat{F}_m$  als der empirischen Verteilungsfunktion der Beobachtungen  $x_1, \dots, x_m$

$$md_\phi^2(\hat{F}_m, F_0) = 2 \sum_{i=1}^m E\phi(|x_i - X_1|^2) - \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m \phi(|x_i - x_j|^2) - mE\phi(|X_1 - X_2|^2).$$

Der zweite Summand ist leicht zu berechnen, der dritte ist eine Konstante, die gegebenenfalls tabelliert werden könnte, aber für den ersten Summanden  $E\phi(|x_i - X_1|^2)$  bedarf es einer expliziten Darstellung des Integrals, wenn man nicht auf numerische Methoden zurückgreifen möchte.

Ein Fall, in dem dies möglich ist, der bereits in Székelys Arbeit [Szé02] (siehe auch Székely und Rizzo [SR04]) berührt wird und auf den wir im folgenden Abschnitt noch detaillierter eingehen werden, ist der Test auf eine eindimensionale Standardnormalverteilung bei Anwendung des Cramér-Kerns,  $\phi(z) = \sqrt{z}/2$ :

$$\begin{aligned} E\phi(|x - Z_1|^2) &= \frac{1}{2}E|x - Z_1| = f_Z(x) + x \left( F_Z(x) - \frac{1}{2} \right), \quad x \in \mathbb{R}, \\ E\phi(|Z_1 - Z_2|^2) &= \frac{1}{2}E|Z_1 - Z_2| = \frac{1}{\sqrt{\pi}} \end{aligned}$$

mit  $Z_1, Z_2 \sim N(0, 1)$  unabhängig und  $f_Z, F_Z$  als Dichte bzw. Verteilungsfunktion der Standardnormalverteilung. Die Testgröße lautet

$$\begin{aligned} md_\phi^2(\hat{F}_m, F_Z) & \tag{8.2} \\ &= 2 \sum_{i=1}^m \left( f_Z(x_i) + x_i \left( F_Z(x_i) - \frac{1}{2} \right) \right) - \frac{1}{2m} \sum_{i=1}^m \sum_{j=1}^m |x_i - x_j| - \frac{m}{\sqrt{\pi}}. \end{aligned}$$

Tabelle 8.1: Kritische Werte für den Cramér-Test (8.2) beim Anpassungsproblem an eine Standardnormalverteilung

Niveau $\alpha$	kritischer Wert
0.1	1.111
0.05	1.444
0.01	2.269

Als kritische Werte ergeben sich über die Approximation der Grenzverteilung durch das Eigenwertverfahren die in Tabelle 8.1 angegebenen Werte.

Um einen Eindruck von der Güte zu gewinnen, vergleichen wir diesen neuen Test mit drei wohlbekanntem Ein-Stichproben-Tests: dem Cramér-von Mises-Test, dem Kolmogorov-Smirnov-Test und dem Test von Anderson und Darling. Einen Überblick über diese und weitere Testverfahren samt kritischer Werte findet sich bei D'Agostino und Stephens [DS86, S.97 ff.]. Als Alternativen wählen wir nur einige wenige Verteilungen aus: die Normalverteilung mit Erwartungswert 0 und Varianz 2, die Normalverteilung mit Erwartungswert  $\frac{1}{2}$  und Varianz 1, die logistische Verteilung mit Dichte

$$f_{\text{logis}}(x) = \frac{\exp(x)}{(1 + \exp(x))^2}, \quad x \in \mathbb{R},$$

sowie die zweiseitige Exponentialverteilung mit Dichte

$$f_{2\text{-Exp}}(x) = \frac{1}{2} \exp(-|x|), \quad x \in \mathbb{R}.$$

Zusätzlich betrachten wir noch Zufallsvariablen der Form

$$\frac{1}{\lambda}(U^\lambda - (1 - U)^\lambda),$$

$U \sim \text{unif}(0, 1)$ , mit so genannten symmetrischen Lambda Verteilungen von Tukey (siehe Joiner und Rosenblatt[JR71]) für  $\lambda = 1, 0.1, 0.01$ .

Als Stichprobenumfang für alle Tests verwenden wir  $m = 50$  und als Niveau  $\alpha = 0.05$ . Die in der Tabelle 8.2 angegebene Größe ist der Anteil der Verwerfungen bei 10000 Simulationen.

Die Tabelle zeigt, dass alle vier Tests das vorgegebene Niveau einhalten. Der Ein-Stichproben-Cramér-Test schneidet gegen die klassischen nichtparametrischen Tests hervorragend ab. Die Güte bleibt zwar stets etwas hinter der des Anderson-Darling-Tests zurück, ist aber z.B. beim Test gegen die Alternativen mit Tukey-Verteilung dem Cramér-von Mises- und Kolmogorov-Smirnov-Test deutlich überlegen.

Gut zu erkennen ist, dass die Güte bei Tukey-Alternativen für kleinere Werte von  $\lambda$  in die Nähe der Güte der logistischen Verteilung kommt, was mit der Verteilungskonvergenz der Tukey-Verteilung für  $\lambda \rightarrow 0$  gegen die logistische Verteilung zu erklären ist.

Tabelle 8.2: Verwerfungswahrscheinlichkeit beim Test auf die Standardnormalverteilung

Alternative	Cramér	C.-von Mises	Kolm.-Smirnov	And.-Darling
N(0, 1)	0.048	0.048	0.041	0.048
N(0, 2)	0.571	0.274	0.244	0.664
N(0.5, 1)	0.929	0.916	0.856	0.933
2-Exp	0.237	0.071	0.054	0.376
Tukey, $\lambda = 1$	0.718	0.354	0.185	0.825
Tukey, $\lambda = 0.1$	0.755	0.391	0.338	0.835
Tukey, $\lambda = 0.01$	0.917	0.550	0.466	0.961
logis	0.932	0.577	0.497	0.966

Nach diesen hervorragenden Ergebnissen sei nochmals auf die Problematik der Bestimmung der Testgröße hingewiesen, die eine Übertragung aller neuen Tests auf das Anpassungsproblem erschwert.

## 8.2 Anpassung an eine Familie von Verteilungen

Eine weitere Möglichkeit eines Anpassungstests besteht im Test auf eine ganze Klasse von Verteilungen. Wir betrachten wieder den Spezialfall des Ein-Stichproben-Cramér-Tests, diesmal jedoch auf die gesamte Klasse von Normalverteilungen.

Um auf eine Normalverteilung zu testen, standardisieren wir die Beobachtungen durch Subtraktion des Stichprobenmittels

$$\hat{\mu} = \frac{1}{m} \sum_{j=1}^m x_j$$

und Division dieser Differenz durch die Stichprobenstreuung

$$\hat{\sigma} = \sqrt{\frac{1}{m} \sum_{j=1}^m (x_j - \hat{\mu})^2}$$

und testen mit dem Verfahren des letzten Abschnitts, ob die den empirisch standardisierten Beobachtungen zu Grunde liegende Verteilung eine Standardnormalverteilung ist. Die standardisierte Testgröße lautet somit

$$\hat{T}_m = 2 \sum_{i=1}^m \left( f_Z \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) + x_i \left( F_Z \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) - \frac{1}{2} \right) \right) - \frac{1}{2m} \sum_{i=1}^m \sum_{j=1}^m \frac{1}{\hat{\sigma}} |x_i - x_j| - \frac{m}{\sqrt{\pi}}$$

und kann analog zur Darstellung (1.3), welche für die asymptotische Behandlung des Cramér-Tests in [Fra00] benutzt wurde, auch geschrieben werden als

$$\begin{aligned}\hat{T}_m &= m \int_{-\infty}^{+\infty} \left[ \frac{1}{m} \sum_{i=1}^m \mathbb{I} \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \leq t \right) - F_Z(t) \right]^2 dt \\ &= \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m \kappa(x_i, x_j; \hat{\mu}, \hat{\sigma})\end{aligned}\quad (8.3)$$

mit

$$\kappa(x, y; \gamma_1, \gamma_2) := \int_{-\infty}^{+\infty} \left[ \mathbb{I} \left( \frac{x - \gamma_1}{\gamma_2} \leq t \right) - F_Z(t) \right] \left[ \mathbb{I} \left( \frac{y - \gamma_1}{\gamma_2} \leq t \right) - F_Z(t) \right] dt.$$

Unter der Hypothese sind die  $x_i$  Beobachtungen unabhängiger normalverteilter Zufallsvariablen, deren Erwartungswert mit  $\mu$  und Varianz mit  $\sigma^2$  bezeichnet sei.

De Wet und Randles [dWR87] haben die Auswirkung dieser Methode auf die Grenzverteilung, mit deren Hilfe wir den kritischen Wert bestimmen wollen, untersucht. Leider lässt sich ihr Resultat nicht direkt übernehmen, da sie eine Darstellung der Art (8.3) erfordern, bei der die Integration über ein endliches Maß ausgeführt wird. Mit dem in dieser Arbeit mehrfach erwähnten Hilfssatz 3.9 von Billingsley ist es jedoch möglich, das Ergebnis auf den vorliegenden Fall zu übertragen.

**Satz 8.1.** *Seien  $X_1, X_2, \dots$  u.i.v. Zufallsgrößen mit der  $N(\mu, \sigma^2)$ -Verteilung. Dann gilt*

$$T_m := \hat{T}_m(X_1, \dots, X_m) \xrightarrow[m \rightarrow \infty]{\mathcal{D}} T := \sum_{\sigma} \lambda_{\sigma} Z_{\sigma}^2,$$

wobei  $(Z_{\sigma})_{\sigma}$  eine Folge unabhängiger standardnormalverteilter Zufallsvariablen sei und  $(\lambda_{\sigma})_{\sigma}$  die Folge der Eigenwerte des Operators

$$\begin{aligned}L_2(\mathbb{R}, \mathcal{B}, F_Z) &\rightarrow L_2(\mathbb{R}, \mathcal{B}, F_Z) \quad , \\ A : f &\mapsto \int \kappa_*(\cdot, x) f(x) dF_Z(x)\end{aligned}\quad (8.4)$$

mit

$$\begin{aligned}\kappa_*(x, y) &= \int \left[ \mathbb{I}(x \leq t) - F_Z(t) + f_Z(t) \left( x + \frac{x^2}{2}t - \frac{t}{2} \right) \right] \\ &\quad \left[ \mathbb{I}(y \leq t) - F_Z(t) + f_Z(t) \left( y + \frac{y^2}{2}t - \frac{t}{2} \right) \right] dt.\end{aligned}$$

Die Grenzverteilung hängt nicht von  $\mu, \sigma$  ab.

**Beweis.** Zum Beweis betrachten wir zunächst eine Testgröße (8.3), bei der die Integration lediglich über ein Kompaktum ausgeführt wird:

$$\begin{aligned} T_m^k &= m \int_{-k}^{+k} \left[ \frac{1}{m} \sum_{i=1}^m \mathbb{I} \left( \frac{X_i - \hat{\mu}}{\hat{\sigma}} \leq t \right) - F_Z(t) \right]^2 dt \\ &= \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m \kappa^k(X_i, X_j; \hat{\mu}, \hat{\sigma}) \end{aligned}$$

mit

$$\kappa^k(x, y; \gamma_1, \gamma_2) := \int_{-k}^{+k} \underbrace{\left[ \mathbb{I} \left( \frac{x - \gamma_1}{\gamma_2} \leq t \right) - F_Z(t) \right]}_{=: g(x, t; \gamma_1, \gamma_2)} \left[ \mathbb{I} \left( \frac{y - \gamma_1}{\gamma_2} \leq t \right) - F_Z(t) \right] dt.$$

Diese Einschränkung ermöglicht es, das letzte Integral als Integral über ein endliches Maß zu betrachten. Damit sind alle Voraussetzungen des Satzes von de Wet und Randles [dWR87, Theorem 2.16, S.402] erfüllt, welcher zeigt, dass  $T_m^k$  für  $m \rightarrow \infty$  in Verteilung gegen eine Zufallsvariable  $T^k$  konvergiert.  $T^k$  lässt sich als gewichtete Summe von  $\chi_1^2$ -verteilten Zufallsvariablen schreiben, deren Gewichte die Eigenwerte des Operators

$$\begin{aligned} L_2(\mathbb{R}, \mathcal{B}, F_Z) &\rightarrow L_2(\mathbb{R}, \mathcal{B}, F_Z) \quad , \\ A^k : f &\mapsto \int \kappa_*^k(\cdot, x) f(x) dF_Z(x) \end{aligned}$$

mit

$$\begin{aligned} \kappa_*^k(x, y) &= \int_{-k}^{+k} \left[ \mathbb{I}(x \leq t) - F_Z(t) + f_Z(t) \left( x + \frac{x^2}{2}t - \frac{t}{2} \right) \right] \\ &\quad \left[ \mathbb{I}(y \leq t) - F_Z(t) + f_Z(t) \left( y + \frac{y^2}{2}t - \frac{t}{2} \right) \right] dt \end{aligned}$$

sind. Die Terme, die im Vergleich zu  $\kappa^k$  in jedem der beiden Faktoren addiert werden, sind das Skalarprodukt der (Gâteaux-)Ableitung des Erwartungswerts eines Faktors nach den beiden geschätzten Größen  $\gamma_1$  und  $\gamma_2$  an der Stelle des wahren Parameters  $\mu, \sigma$

$$dEg(X_1, t; \gamma_1, \gamma_2) = \left( \begin{array}{c} \frac{d}{d\gamma_1} F_Z(\gamma_2 t + \gamma_1) \\ \frac{d}{d\gamma_2} F_Z(\gamma_2 t + \gamma_1) \end{array} \bigg|_{\gamma_1=\mu, \gamma_2=\sigma} \right) = \frac{1}{\sigma} f_Z(\sigma t + \mu) \begin{pmatrix} 1 \\ t \end{pmatrix}$$

und dem Vektor  $(\alpha_1(x), \alpha_2(x))'$ . Die Funktionen  $\alpha_1$  bzw.  $\alpha_2$  gewinnt man aus einer bis auf ein  $o_P(m^{-1/2})$  genauen Darstellung der Differenz von Schätzer und zu schätzendem Wert als Summe der Beobachtungen, welche mit dieser beliebigen, gemeinsamen Funktion modifiziert sein dürfen. Es gilt

$$\hat{\mu} - \mu = \frac{1}{m} \sum_{j=1}^m \alpha_1(X_j) + o_P(m^{-1/2})$$

mit  $\alpha_1(x) = x - \mu$  und gänzlich verschwindendem  $o_P$ . Für die Stichprobenstreuung gilt bekanntlich

$$\hat{\sigma} - \sigma = \frac{1}{m} \sum_{j=1}^m \alpha_2(X_j) + o_P(m^{-1/2})$$

mit  $\alpha_2(x) = \frac{1}{2\sigma} [(x - \mu)^2 - \sigma^2]$ . Es ergibt sich

$$\begin{aligned} \kappa_*^k(x, y) &= \int_{-k}^{+k} \left[ \mathbb{I} \left( \frac{x - \mu}{\sigma} \leq t \right) - F_Z(t) + f_Z(\sigma t + \mu) \left( \frac{x - \mu}{\sigma} + \frac{(x - \mu)^2}{2\sigma^2} t - \frac{t}{2} \right) \right] \\ &\quad \left[ \mathbb{I} \left( \frac{y - \mu}{\sigma} \leq t \right) - F_Z(t) + f_Z(\sigma t + \mu) \left( \frac{y - \mu}{\sigma} + \frac{(y - \mu)^2}{2\sigma^2} t - \frac{t}{2} \right) \right] dt. \end{aligned}$$

De Wet und Randles integrieren bei der Definition von  $A^k$  bezüglich der Verteilung, die den  $X_1, X_2, \dots$  zu Grunde liegt. Durch Substitution sieht man sofort, dass die Grenzverteilung von  $\mu$  und  $\sigma$  unabhängig ist und man sich auf den Fall  $\mu = 0, \sigma = 1$  beschränken kann. Damit ist die Bedingung (i) zur Anwendung des Hilfssatzes 3.9 erfüllt.

Als Nächstes betrachten wir den Zusammenhang zwischen den Operatoren  $A^k$  und  $A$ . Man zeigt leicht, dass der Operator  $A$  ein Hilbert-Schmidt-Operator von der Spurklasse ist. Insbesondere ist  $A$  kompakt. Darüber hinaus ist  $A$  positiv, da sich der Kern grob gesprochen als Produkt eines Bestandteils für die erste Komponente und eines identischen Bestandteils für die zweite Komponente schreiben lässt,

$$\begin{aligned} &\int Aq(x)q(x) dF_Z(x) \\ &= \iiint \left[ \mathbb{I}(x \leq t) - F_Z(t) + f_Z(t) \left( x + \frac{x^2}{2}t - \frac{t}{2} \right) \right] q(x) \\ &\quad \left[ \mathbb{I}(y \leq t) - F_Z(t) + f_Z(t) \left( y + \frac{y^2}{2}t - \frac{t}{2} \right) \right] q(y) dt dF_Z(x) dF_Z(y) \\ &= \int \left\{ \int \left[ \mathbb{I}(x \leq t) - F_Z(t) + f_Z(t) \left( x + \frac{x^2}{2}t - \frac{t}{2} \right) \right] q(x) dF_Z(x) \right\}^2 dt \geq 0. \end{aligned}$$

Diese Eigenschaften übertragen sich unmittelbar auf die Operatoren  $A^k$ . Die Spur von  $A^k$  konvergiert für  $k \rightarrow \infty$  gegen die Spur von  $A$  (Satz von der monotonen Konvergenz). Analog zeigt man, dass die Summe der Quadrate der Eigenwerte von  $A^k$  gegen die entsprechende Summe von  $A$  konvergiert (Lemma C.6, Anhang C). Wie im Beweis von Satz 3.5 weist man für  $k \rightarrow \infty$  die Konvergenz von  $A^k$  gegen  $A$  in der Operatornorm nach und folgert die gleichmäßige Konvergenz der Eigenwerte von  $A^k$  gegen die Eigenwerte von  $A$ . Dem Beweis von Satz 5.13 folgend zeigt man die Verteilungskonvergenz der  $T^k$  gegen  $T$  durch Nachweis der  $L_1$ -Konvergenz.

Die zweite Voraussetzung des Hilfssatzes von Billingsley ist erfüllt und es bleibt die dritte Voraussetzung zu zeigen:

$$\lim_{k \rightarrow \infty} \limsup_{m \rightarrow \infty} P \left( m \int_{[-k, +k]^c} \left[ \frac{1}{m} \sum_{i=1}^m \mathbf{I} \left( \frac{X_i - \hat{\mu}}{\hat{\sigma}} \leq t \right) - F_Z(t) \right]^2 dt \geq \varepsilon \right) = 0.$$

Wir vergrößern die Wahrscheinlichkeit

$$\begin{aligned} & P \left( m \int_{[-k, +k]^c} \left[ \frac{1}{m} \sum_{i=1}^m \mathbf{I} \left( \frac{X_i - \hat{\mu}}{\hat{\sigma}} \leq t \right) - F_Z(t) \right]^2 dt \geq \varepsilon \right) \\ & \leq P \left( 2m \int_{[-k, +k]^c} \left[ \frac{1}{m} \sum_{i=1}^m \mathbf{I} \left( \frac{X_i - \hat{\mu}}{\hat{\sigma}} \leq t \right) - F(\hat{\sigma}t + \hat{\mu}) \right]^2 dt \geq \frac{\varepsilon}{2} \right) \quad (8.5) \end{aligned}$$

$$+ P \left( 2m \int_{[-k, +k]^c} [F(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{2} \right) \quad (8.6)$$

und schätzen nun die Wahrscheinlichkeiten (8.5) und (8.6) einzeln ab.  $F$  bezeichne dabei die Verteilungsfunktion der  $X_i$  unter der Hypothese, also eine Normalverteilung mit Erwartungswert  $\mu$  und Varianz  $\sigma^2 > 0$ . Auch hier können wir uns im Folgenden auf den Fall  $\mu = 0$ ,  $\sigma^2 = 1$  beschränken.

Wir betrachten zunächst die Wahrscheinlichkeit (8.5) und unterteilen diese

$$\begin{aligned} & P \left( 2m \int_{[-k, +k]^c} \left[ \frac{1}{m} \sum_{i=1}^m \mathbf{I} \left( \frac{X_i - \hat{\mu}}{\hat{\sigma}} \leq t \right) - F_Z(\hat{\sigma}t + \hat{\mu}) \right]^2 dt \geq \frac{\varepsilon}{2} \right) \\ & \leq P \left( \frac{2m}{\hat{\sigma}} \int_{[-\hat{\sigma}k + \hat{\mu}, +\hat{\sigma}k + \hat{\mu}]^c} [F_m(t) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{2}, |\sqrt{m}(\hat{\sigma} - 1)| \leq c_1, |\sqrt{m}\hat{\mu}| \leq c_2 \right) \\ & \quad + P(|\sqrt{m}(\hat{\sigma} - 1)| > c_1) + P(|\sqrt{m}\hat{\mu}| > c_2), \quad c_1, c_2 > 0, \end{aligned}$$

wobei  $F_m(t) := \frac{1}{m} \sum_{i=1}^m \mathbf{I}(X_i \leq t)$ . Für jede vorgegebene Schranke lässt sich durch genügend große Wahl von  $c_1$  und  $c_2$  erreichen, dass die beiden letzteren Wahrscheinlichkeiten für alle  $m$  kleiner sind als die vorgegebene Schranke. Mit der Markovschen Ungleichung folgt für  $\sqrt{m} > 2c_1$  und  $\sqrt{m} > 2c_2$

$$\begin{aligned} & P \left( \frac{2m}{\hat{\sigma}} \int_{[-\hat{\sigma}k + \hat{\mu}, +\hat{\sigma}k + \hat{\mu}]^c} [F_m(t) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{2}, |\sqrt{m}(\hat{\sigma} - 1)| \leq c_1, |\sqrt{m}\hat{\mu}| \leq c_2 \right) \\ & \leq \frac{4}{\varepsilon} \mathbf{E} \left[ \frac{m}{\hat{\sigma}} \int_{[-\hat{\sigma}k + \hat{\mu}, +\hat{\sigma}k + \hat{\mu}]^c} [F_m(t) - F_Z(t)]^2 dt \mathbf{I}(|\sqrt{m}(\hat{\sigma} - 1)| \leq c_1, |\sqrt{m}\hat{\mu}| \leq c_2) \right] \\ & \leq \frac{4}{\varepsilon} \mathbf{E} \left[ \frac{m}{1 - c_1/\sqrt{m}} \int_{[-(1-c_1/\sqrt{m})k + c_2/\sqrt{m}, +(1-c_1/\sqrt{m})k - c_2/\sqrt{m}]^c} [F_m(t) - F_Z(t)]^2 dt \right] \\ & \leq \frac{8}{\varepsilon} \int_{[-(k-1)/2, +(k-1)/2]^c} \mathbf{E} \left[ m (F_m(t) - F_Z(t))^2 \right] dt. \end{aligned}$$

Der Erwartungswert

$$\mathbb{E} \left[ m (F_m(t) - F_Z(t))^2 \right] = F_Z(t)(1 - F_Z(t))$$

ist integrierbar bezüglich  $t$ , womit das letzte Integral der Abschätzung für  $k \rightarrow \infty$  gegen 0 strebt. Wir haben gezeigt, dass der Grenzwert der Wahrscheinlichkeit (8.5), wenn man erst  $m$  und anschließend  $k$  gegen unendlich laufen lässt, 0 ist.

Bei der Abschätzung des Terms (8.6) ist es günstig, den Integrationsbereich aufzuteilen:

$$\begin{aligned} & P \left( 2m \int_{[-k, +k]^c} [F(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{2} \right) \\ & \leq P \left( 2m \int_{-\infty}^{-k} [F(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{4} \right) \\ & \quad + P \left( 2m \int_{+k}^{+\infty} [F(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{4} \right) \end{aligned}$$

und beide Terme getrennt zu behandeln. Dazu zerlegen wir den Ausdruck der letzten Zeile

$$\begin{aligned} & P \left( 2m \int_{(+k, +\infty)} [F_Z(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{4} \right) \\ & \leq P \left( 2m \int_{(+k, +\infty)} [F_Z(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{4}, |\sqrt{m}(\hat{\sigma} - 1)| \leq c_1, |\sqrt{m}\hat{\mu}| \leq c_2 \right) \\ & \quad + P(|\sqrt{m}(\hat{\sigma} - 1)| > c_1) + P(|\sqrt{m}\hat{\mu}| > c_2). \end{aligned}$$

Mit den letzteren beiden Wahrscheinlichkeiten verfahren wir wie bei Term (8.5). Die erste Wahrscheinlichkeit lässt sich mit Hilfe der Markovschen Ungleichung abschätzen

$$\begin{aligned} & P \left( 2m \int_{(+k, +\infty)} [F_Z(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \geq \frac{\varepsilon}{4}, |\sqrt{m}(\hat{\sigma} - 1)| \leq c_1, |\sqrt{m}\hat{\mu}| \leq c_2 \right) \\ & \leq \frac{8}{\varepsilon} \mathbb{E} \left[ \mathbb{I}(|\sqrt{m}(\hat{\sigma} - 1)| \leq c_1, |\sqrt{m}\hat{\mu}| \leq c_2) m \int_{(+k, +\infty)} [F_Z(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \right]. \end{aligned}$$

Mit  $g_m$  als der Dichte von  $\sqrt{m}(\hat{\sigma} - 1)$  und  $f_Z$  als der Dichte von  $\sqrt{m}\hat{\mu}$ , also der Dichte der Standardnormalverteilung, ist

$$\begin{aligned} & \mathbb{E} \left[ \mathbb{I}(|\sqrt{m}(\hat{\sigma} - 1)| \leq c_1, |\sqrt{m}\hat{\mu}| \leq c_2) m \int_{(+k, +\infty)} [F_Z(\hat{\sigma}t + \hat{\mu}) - F_Z(t)]^2 dt \right] \\ & = \int_{-c_1}^{+c_1} \int_{-c_2}^{+c_2} \int_{+k}^{+\infty} m \left[ F_Z \left( \frac{y}{\sqrt{m}}t + t + \frac{z}{\sqrt{m}} \right) - F_Z(t) \right]^2 dt f_Z(z) dz g_m(y) dy. \quad (8.7) \end{aligned}$$

Der Integrand

$$m \left[ F_Z \left( \frac{y}{\sqrt{m}}t + t + \frac{z}{\sqrt{m}} \right) - F_Z(t) \right]^2$$

konvergiert punktweise für jedes feste  $y, z, t$  für  $m \rightarrow \infty$  gegen  $(yt + z)^2 f_Z^2(t)$ . Das Integral über  $t \in (+k, +\infty)$  konvergiert mit, da der Integrand eine integrierbare Majorante hat: Mit dem Mittelwertsatz der Differentialrechnung gilt mit  $\xi_t(m) \in [0, 1]$  für  $t \in (+k, +\infty)$

$$m \left[ F_Z \left( \frac{y}{\sqrt{m}}t + t + \frac{z}{\sqrt{m}} \right) - F_Z(t) \right]^2 = \left[ (yt + z) f_Z \left( t + (yt + z) \frac{\xi_t(m)}{\sqrt{m}} \right) \right]^2.$$

Für  $m > 9y^2$  und  $m > 9z^2$  für alle  $y, z$  aus einem beliebigen, festen Kompaktum und  $k > 1$  lässt sich das Argument in der Dichte der Normalverteilung verkleinern und damit der Integrand punktweise vergrößern durch

$$\left| t + (yt + z) \frac{\xi_t(m)}{\sqrt{m}} \right| \geq t (1 - |y/\sqrt{m}| - |z/\sqrt{m}|) \geq \frac{t}{3}.$$

Das Integral über die Majorante

$$[(yt + z) f_Z(t/3)]^2$$

ist endlich. Der Ausdruck

$$e_m(y, z) := \int_{+k}^{+\infty} m \left[ F_Z \left( y \frac{t}{\sqrt{m}} + t + \frac{z}{\sqrt{m}} \right) - F_Z(t) \right]^2 dt$$

konvergiert also für jedes feste  $y, z$  gegen

$$e(y, z) := \int_{+k}^{+\infty} (z + yt)^2 f_Z^2(t) dt.$$

Überdies ist  $e_m$  stetig und die Folge  $(e_m)_m$  auf jedem Kompaktum gleichmäßig beschränkt. Es bleibt zu zeigen, dass

$$\int_{-c_1}^{+c_1} \int_{-c_2}^{+c_2} e_m(y, z) g_m(y) f_Z(z) dz dy$$

gegen

$$\int_{-c_1}^{+c_1} \int_{-c_2}^{+c_2} e(y, z) g(y) f_Z(z) dz dy$$

strebt, wobei  $g$  die Dichte einer Zufallsvariablen ist. Es gilt

$$\left| \int_{-c_1}^{+c_1} \int_{-c_2}^{+c_2} e(y, z) g(y) f_Z(z) dz dy - \int_{-c_1}^{+c_1} \int_{-c_2}^{+c_2} e_m(y, z) g_m(y) f_Z(z) dz dy \right|$$

$$\leq \int_{-c_1}^{+c_1} \int_{-c_2}^{+c_2} |e(y, z) - e_m(y, z)| g(y) f_Z(z) dz dy \quad (8.8)$$

$$+ \int_{-c_1}^{+c_1} \int_{-c_2}^{+c_2} e_m(y, z) |g(y) - g_m(y)| f_Z(z) dz dy. \quad (8.9)$$

Der Term (8.8) konvergiert wegen des Satzes von der majorisierten Konvergenz gegen 0, bei (8.9) überzeugt man sich davon, dass  $g_m$  punktweise gegen die Dichte einer Normalverteilung mit Erwartungswert 0 und Varianz 1/2 konvergiert,

$$\begin{aligned}
 g_m(x) &= \frac{(x + \sqrt{m})^{m-2} \exp(-(x + \sqrt{m})^2/2)}{(\sqrt{2})^{m-3} \Gamma(\frac{m-1}{2})} \quad \text{für } x \geq -\sqrt{m} \quad \text{und 0 sonst} \\
 &\stackrel{\text{Stirling}}{\approx} \frac{1}{\sqrt{\pi}} \exp\left(-\frac{x^2}{2}\right) \underbrace{\left(\frac{x + \sqrt{m}}{\sqrt{m}}\right)^{m-2} \exp(-\sqrt{m}x)}_{\xrightarrow{m \rightarrow \infty} \exp(-\frac{x^2}{2})} \underbrace{\left(\frac{\sqrt{m}}{\sqrt{m-3}}\right)^{m-2}}_{\xrightarrow{m \rightarrow \infty} \exp(\frac{3}{2})} \exp\left(-\frac{3}{2}\right) \\
 &\xrightarrow{m \rightarrow \infty} \frac{1}{\sqrt{\pi}} \exp(-x^2)
 \end{aligned}$$

und wendet dann den Satz von Scheffé an (vgl. Billingsley [Bil86, Theorem 16.11, S.218]). Das sich als Grenzwert von (8.7) für  $m \rightarrow \infty$  ergebende Integral

$$\int_{-c_1}^{+c_1} \int_{-c_2}^{+c_2} \int_{+k}^{+\infty} (z + yt)^2 f_Z^2(t) dt f_Z(z) dz g(y) dy$$

ist endlich, strebt für  $k \rightarrow \infty$  also gegen 0. Mit dem Integrationsbereich  $(-\infty, -k)$  verfährt man analog. Somit ist gezeigt, dass die Wahrscheinlichkeit (8.6) beliebig klein wird, also gegen 0 strebt, wenn man zunächst  $m \rightarrow \infty$  und dann  $k \rightarrow +\infty$  laufen lässt.

Es folgt mit dem Hilfssatz von Billingsley die Verteilungskonvergenz von  $T_m$  gegen  $T$ . □

Die Kernfunktion  $\kappa_*(x, y)$  lässt sich explizit angeben. Approximiert man die Normalverteilung, bezüglich der in der Darstellung (8.4) des Operators  $A$  integriert wird, durch ein diskretes Maß, so lässt sich das Eigenwertproblem wie zuvor in Abschnitt 4.2.1 als Matrix-Eigenwertproblem auffassen. Auf diesem Wege erhält man die in Tabelle 8.3 angegebenen, asymptotischen kritischen Werte.

Tabelle 8.3: Asymptotische kritische Werte für den Cramér-Test beim Anpassungsproblem an eine Normalverteilung

Niveau $\alpha$	kritischer Wert
0.1	0.349
0.05	0.418
0.01	0.580

Die Arbeit schließt mit einem Vergleich dieses neuen Anpassungstests mit den Versionen des Kolmogorov-Smirnov-Tests, des Cramér-von Mises-Tests und des Tests von Anderson-Darling für den Test auf Normalverteilung. Diese drei Tests haben

wir bereits in ihrer Form zur Anpassung an eine vollständig bestimmte Verteilung verwendet. D'Agostino und Stephens [DS86, S.123] geben für diese Tests approximative kritische Werte an, die bei Anderson-Darling und Cramér-von Mises auf der asymptotischen Verteilung und beim Kolmogorov-Smirnov-Test auf Monte Carlo-Simulationen beruhen.

Wie zuvor wird beim Vergleich zum Niveau  $\alpha = 0.05$  getestet und wir approximieren die Güte durch den Anteil der Verwerfungen bei 10000 Tests. Jedem Test liegen jeweils 50 Beobachtungen u.i.v. Zufallsvariablen mit einer der folgenden Verteilungen zu Grunde:

- der Standardnormalverteilung,
- der zweiseitigen Exponentialverteilung,
- der symmetrischen Lambda-Verteilung von Tukey für  $\lambda = 1, 0.1, 0.01$  und
- der logistischen Verteilung.

Nur eine unwesentlich größere Güte als das Niveau ist bei einer symmetrischen Lambda-Verteilung für  $\lambda = 0.1$  zu erwarten, da  $\lambda = 0.1$  in der Nähe von  $\lambda = 0.135$  liegt, welches gewählt wird, will man eine Normalverteilung mit einer symmetrischen Lambda-Verteilung approximieren (siehe Joiner und Rosenblatt [JR71]).

Die Simulationsergebnisse in Tabelle 8.4 entsprechen diesen Erwartungen. Es bestätigen sich die Ergebnisse, die bereits im Fall des Tests auf eine Standardnormalverteilung gemacht wurden: Der Cramér-Test hat ein besseres Güteverhalten als der Cramér-von Mises-Test und ein deutlich besseres Güteverhalten als der Kolmogorov-Smirnov-Test. Im Vergleich mit dem Test von Anderson-Darling scheint letzterer nur unwesentlich überlegen zu sein.

Tabelle 8.4: Verwerfungswahrscheinlichkeit beim Test auf eine Normalverteilung

Alternative	Cramér	C.-von Mises	Kolm.-Smirnov	And.-Darling
N(0, 1)	0.051	0.050	0.046	0.049
2-Exp	0.518	0.509	0.392	0.516
Tukey, $\lambda = 1$	0.590	0.470	0.256	0.602
Tukey, $\lambda = 0.1$	0.062	0.060	0.049	0.058
Tukey, $\lambda = 0.01$	0.138	0.126	0.091	0.141
logis	0.143	0.131	0.095	0.147

Insgesamt lässt sich festhalten, dass der Cramér-Test beim Anpassungsproblem sowohl an eine Standardnormalverteilung als auch an die Klasse aller Normalverteilungen sehr gut abschneidet. Man würde ein ähnliches Verhalten für andere Verteilungen und andere Kernfunktionen  $\phi \in \Phi$  erwarten, jedoch dürfte bedauerlicherweise

in diesen Fällen neben den Schwierigkeiten bei der Bestimmung der Testgröße (vergleiche Abschnitt 8.1) auch der Beweis der asymptotischen Aussage für die mit Hilfe von Schätzern standardisierte Version der Testgröße, Satz 8.1, ein großes Hindernis darstellen.



# Anhang A

## Vollständig monotone Funktionen

In diesem Anhang wird kurz auf die in der Arbeit, vornehmlich in Kapitel 2.1, benötigten Zusammenhänge zwischen vollständig monotonen Funktionen und (bedingt) positiv definiten Funktionen eingegangen. Eine genaue Behandlung von positiv definiten Funktionen ist neben der fundamentalen Arbeit von Schoenberg [Sch38] auch in dem Buch von Berg, Christensen und Ressel [BCR84] zu finden. Das verwendete Resultat zu bedingt positiv definiten Funktionen findet sich zwar bereits bei Schoenberg [Sch38], allgemeiner und deutlicher wird es jedoch bei Guo, Hu und Sun [GHS93].

**Definition A.1.** Die Funktion  $f : I \rightarrow \mathbb{R}$ , definiert auf einem Intervall  $I$ , heißt **vollständig monoton**, wenn  $f \in C^\infty(I)$ , d.h.  $f$  ist auf  $I$  beliebig oft stetig differenzierbar, und  $(-1)^j f^{(j)}(t) \geq 0$  für alle  $j \in \mathbb{N}_0$ ,  $t \in I$ .

Bei der Definition von bedingt positiv definiten Funktionen lehnen wir uns an die Terminologie von Guo, Hu und Sun an.

**Definition A.2.** Eine Funktion  $g : \mathbb{R}^d \rightarrow \mathbb{C}$  nennt man **bedingt positiv definit** von der Ordnung  $k \in \mathbb{N}$ , falls für beliebige Zahlen  $c_1, \dots, c_N \in \mathbb{C}$  und beliebige Punkte  $x_1, \dots, x_N \in \mathbb{R}^d$ ,  $N \in \mathbb{N}$ , mit

$$\sum_{j=1}^N c_j p(x_j) = 0 \tag{A.1}$$

für alle Polynome  $p$  mit  $d$  Variablen und Grad kleiner gleich  $k - 1$  folgt

$$\sum_{j=1}^N \sum_{k=1}^N c_j \bar{c}_k g(x_j - x_k) \geq 0. \tag{A.2}$$

( $\bar{c}_k$  bezeichnet hier den komplex-konjugierten Wert zu  $c_k$ )

Die zusätzliche Bedingung (A.1) unterscheidet bedingt positiv definite Funktionen von positiv definiten Funktionen. Bei letzteren unterliegen die  $c_j$  keiner Einschränkung. Bei dem für uns relevanten Fall von bedingt positiv definiten Funktionen der Ordnung  $k = 1$  wird die Bedingung (A.1) zu  $\sum_{j=1}^N c_j = 0$ .

Zwischen vollständig monotonen Funktionen und bedingt positiv definiten Funktionen besteht der folgende Zusammenhang:

**Satz A.3 (Guo, Hu, Sun [GHS93, Theorem 2.1]).** *Es sei  $f : [0, +\infty) \rightarrow \mathbb{R}$  eine stetige Funktion und  $g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(t) := f(t^2)$ . Dann sind die folgenden Aussagen äquivalent:*

- (a)  $g(| \cdot |)$  ist als Funktion von  $\mathbb{R}^d$  nach  $\mathbb{R}$  bedingt positiv definit von der Ordnung  $k \in \mathbb{N}$  für alle Dimensionen  $d \in \mathbb{N}$ .
- (b)  $f \in C^\infty((0, +\infty))$  und  $(-1)^k f^{(k)}$  ist vollständig monoton auf  $(0, +\infty)$ .

Eine Besonderheit von vollständig monotonen Funktionen, die letztlich in dieser Arbeit den Weg zur Behandlung der Asymptotik über empirische charakteristische Funktionen öffnet, ist das folgende Resultat, das auf Bernstein zurückgeht.

**Satz A.4 (Berg, Christensen, Ressel [BCR84, Corollary 4.6.14, S.135]).** *Für eine Funktion  $f : (0, +\infty) \rightarrow \mathbb{R}$  sind die folgenden Bedingungen äquivalent:*

- (a) *Es existiert ein positives, von innen reguläres Maß, das Kompakta endliche Werte zuordnet, also ein positives Radon Maß  $\sigma$  auf der  $\sigma$ -Algebra  $\mathcal{B}_{\geq 0}$  auf den nicht-negativen reellen Zahlen  $\mathbb{R}_{\geq 0}$ , so dass  $f$  die Laplace-Transformierte von  $\sigma$  ist:*

$$f(z) = \int_{[0, +\infty)} \exp(-\lambda z) d\sigma(\lambda), \quad z > 0.$$

- (b)  $f$  ist unendlich oft stetig differenzierbar und vollständig monoton.

Zur Definition der maßtheoretischen Begriffe siehe auch Bauer [Bau90, Definition 25.2 und 25.3, S.174ff].

# Anhang B

## U- und V-Statistiken

Dieser Abschnitt soll die verwendeten Ergebnisse zu V-Statistiken zusammenfassen. Eine Referenz für die den V-Statistiken verwandten U-Statistiken ist das Buch von Koroljuk und Borovskich [KB94]. Aussagen zur Asymptotik von Mehr-Stichproben-U-Statistiken finden sich unter anderem in dem Artikel von Sen [Sen74] für den nicht-degenerierten Fall sowie in dem Artikel von Neuhaus [Neu77] für den in dieser Arbeit auftauchenden Fall von degenerierten Zwei-Stichproben-U-Statistiken der Kernlänge  $(2, 2)$ . Die folgenden Ausführungen zum starken Gesetz der großen Zahlen orientieren sich an der Arbeit von Eichelsbacher [Eic92], der insbesondere auf den Mehr-Stichproben-Fall bei V-Statistiken eingeht.

Die starken Gesetze der großen Zahlen für U- und V-Statistiken gehen von folgender Definition dieser Teststatistiken aus.

**Definition B.1.** Seien  $X_{\ell;1}, \dots, X_{\ell;n_\ell}$ ,  $\ell = 1, \dots, k$ ,  $n_\ell \in \mathbb{N}$  unabhängige  $\mathbb{R}^d$ -wertige Zufallsvektoren, welche für festes  $\ell$  dieselbe Verteilung haben mögen, und

$$h : ((\mathbb{R}^d)^{m_1} \times \dots \times (\mathbb{R}^d)^{m_k}, (\mathcal{B}^d)^{m_1} \otimes \dots \otimes (\mathcal{B}^d)^{m_k}) \rightarrow (\mathbb{R}, \mathcal{B})$$

eine wie angegeben messbare Kernfunktion mit  $n_\ell \geq m_\ell$ ,  $\ell = 1, \dots, k$ , die in jeder Stichprobe symmetrisch ist. Dann nennt man die Statistik

$$U_{n_1, \dots, n_k}(h) := \frac{1}{\prod_{\ell=1}^k \binom{n_\ell}{m_\ell}} \sum_c h(X_{1;j_{1;1}}, \dots, X_{1;j_{1;m_1}}; \dots; X_{k;j_{k;1}}, \dots, X_{k;j_{k;m_k}})$$

eine **U-Statistik mit Kern  $h$  und Grad  $(m_1, \dots, m_k) \in \mathbb{N}^k$** .

Dabei sei  $\{j_{\ell;1}, \dots, j_{\ell;m_\ell}\}$ ,  $j_{\ell;1} < \dots < j_{\ell;m_\ell}$ , eine  $m_\ell$  elementige Teilmenge der Menge  $\{1, \dots, n_\ell\}$  für  $\ell = 1, \dots, k$  und  $c$  bezeichne die Summation über alle Möglichkeiten, solche Teilmengen zu bilden.

Symmetrie innerhalb jeder Stichprobe bedeutet, dass für jedes  $\ell = 1, \dots, k$  und jede Permutation  $\pi$  der Zahlen  $1, \dots, m_\ell$  gilt

$$\begin{aligned} & h(x_{1;1}, \dots, x_{1;m_1}; \dots; x_{\ell;1}, \dots, x_{\ell;m_\ell}; \dots; x_{k;1}, \dots, x_{k;m_k}) \\ &= h(x_{1;1}, \dots, x_{1;m_1}; \dots; x_{\ell;\pi(1)}, \dots, x_{\ell;\pi(m_\ell)}; \dots; x_{k;1}, \dots, x_{k;m_k}). \end{aligned}$$

Eng verknüpft mit einer U-Statistik ist die V-Statistik.

**Definition B.2.** Seien  $X_{\ell;1}, \dots, X_{\ell;n_\ell}$ ,  $\ell = 1, \dots, k$ ,  $n_\ell \in \mathbb{N}$  unabhängige  $\mathbb{R}^d$ -wertige Zufallsvektoren, welche für festes  $\ell$  dieselbe Verteilung haben mögen, und

$$h : ((\mathbb{R}^d)^{m_1} \times \dots \times (\mathbb{R}^d)^{m_k}, (\mathcal{B}^d)^{m_1} \otimes \dots \otimes (\mathcal{B}^d)^{m_k}) \rightarrow (\mathbb{R}, \mathcal{B})$$

eine wie angegeben messbare Kernfunktion, die in jeder Stichprobe symmetrisch ist. Dann nennt man die Statistik

$$V_{n_1, \dots, n_k}(h) := \frac{1}{\prod_{\ell=1}^k n_\ell^{m_\ell}} \sum_{j_{1;1}, \dots, j_{1;m_1}=1}^{n_1} \dots \sum_{j_{k;1}, \dots, j_{k;m_k}=1}^{n_k} h(X_{1;j_{1;1}}, \dots, X_{1;j_{1;m_1}}; \dots; X_{k;j_{k;1}}, \dots, X_{k;j_{k;m_k}})$$

eine **V-Statistik mit Kern  $h$  und Grad  $(m_1, \dots, m_k) \in \mathbb{N}^k$** .

In dieser Arbeit tauchen U- und V-Statistiken im Mehr-Stichproben-Fall auf. Von den interessanten Aussagen zu U- und V-Statistiken wird im Wesentlichen nur eine benötigt: um die fast sichere Konvergenz geeignet normierter Größen bei der Untersuchung der Asymptotik sowie der Konsistenz zu zeigen benötigt man starke Gesetze der großen Zahlen. Das starke Gesetz der großen Zahlen für U-Statistiken lautet wie folgt.

**Satz B.3 (Eichelsbacher [Eic92, Satz 3.2, S.47]).** Sei  $h$  der in jeder Stichprobe symmetrische Kern einer U-Statistik vom Grad  $(m_1, \dots, m_k) \in \mathbb{N}^k$ . Es gelte

$$\mathbb{E} [h^k(X_{1;1}, \dots, X_{1;m_1}; \dots; X_{k;1}, \dots, X_{k;m_k})] < \infty.$$

Wenn ferner alle Stichprobenumfänge  $n_\ell$ ,  $\ell = 1, \dots, k$  gegen unendlich streben, dann gilt fast sicher und in  $L_1$

$$U_{n_1, \dots, n_k}(h) \rightarrow \mathbb{E}h(X_{1;1}, \dots, X_{1;m_1}; \dots; X_{k;1}, \dots, X_{k;m_k}).$$

Im Fall von V-Statistiken benötigt man stärkere Voraussetzungen für das starke Gesetz der großen Zahlen:

**Satz B.4 (Eichelsbacher [Eic92, Satz 3.4, S.48]).** Sei  $h$  der in jeder Stichprobe symmetrische Kern einer V-Statistik vom Grad  $(m_1, \dots, m_k) \in \mathbb{N}^k$ . Es gelte

$$\mathbb{E} \left[ h^k(X_{1;j_{1;1}}, \dots, X_{1;j_{1;m_1}}; \dots; X_{k;j_{k;1}}, \dots, X_{k;j_{k;m_k}}) \right] < \infty \quad (\text{B.1})$$

für alle  $j_{\ell;1}, \dots, j_{\ell;m_\ell} \in \{1, \dots, m_\ell\}$ ,  $\ell = 1, \dots, k$ . Die Stichprobenumfänge mögen so gegen unendlich streben, dass  $\lim_{n \rightarrow \infty} \frac{n_\ell}{n} = \tau_\ell \in (0, 1)$ ,  $\ell = 1, \dots, k$ , wobei  $n = n_1 + \dots + n_k$ . Dann gilt fast sicher

$$V_{n_1, \dots, n_k}(h) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}h(X_{1;1}, \dots, X_{1;m_1}; \dots; X_{k;1}, \dots, X_{k;m_k}).$$

Den Ein-Stichproben-Fall, vergleiche Koroljuk, Borovskich [KB94, Theorem 3.1.1, S.93, Theorem 3.2.3, S.101], erhält man als Spezialfall, indem man  $k = 1$  wählt.

In Satz B.4 kann die Forderung der Symmetrie der Kernfunktion fallen gelassen werden. Um dies einzusehen, macht man sich klar, dass sich die V-Statistik einer nicht-symmetrischen Kernfunktion als V-Statistik gleichen Grades mit einer symmetrischen Kernfunktion schreiben lässt, indem man über alle Permutationen der Argumente summiert. Bedingung (B.1) muss dann für die symmetrisierte Kernfunktion nachgewiesen werden. Dazu wendet man die verallgemeinerte Höldersche Ungleichung (siehe beispielsweise Hewitt und Stromberg [HS65, ex. (13.26), S.200]) auf jeden Summanden der ausmultiplizierten Bedingung (B.1) für die symmetrisierte Kernfunktion an und sieht, dass die Forderung der Bedingung (B.1) für die nicht symmetrisierte Kernfunktion genügt, um den Voraussetzungen des Satzes B.4 zu genügen.



# Anhang C

## Lineare Operatoren auf Hilberträumen

Bei der Behandlung der Asymptotik degenerierter V-Statistiken stößt man auf die Eigenwerte von Operatoren auf Hilberträumen. Dieser Anhang fasst die im Zusammenhang mit solchen Operatoren auftauchenden Begriffe und Aussagen zusammen. Alle Aussagen finden sich im Standardwerk von Dunford und Schwartz [DS88a, DS88b]. Als weitere Referenz sei das Buch von Heuser [Heu92] genannt.

In der vorliegenden Arbeit werden lediglich der Hilberträume  $\mathcal{H} = L_2(R, \mathcal{R}, \mu)$  über dem skalaren Körper  $\mathbb{R}$  und  $\mathcal{H} = L_2^{\mathbb{C}}(R, \mathcal{R}, \mu)$  über dem skalaren Körper  $\mathbb{C}$  benötigt, wobei  $\mathcal{R}$  die Borelsche  $\sigma$ -Algebra auf  $R \in \mathcal{B}^d$ ,  $d \in \mathbb{N}$ , und  $\mu$  ein positives,  $\sigma$ -endliches Maß auf  $\mathcal{R}$  ist. Deswegen sind die folgenden Ausführungen auf diesen Fall beschränkt. Das innere Produkt auf  $\mathcal{H}$  sei bezeichnet mit  $\langle \cdot, \cdot \rangle$  und die kanonische Norm mit  $\|f\| := \sqrt{\langle f, f \rangle}$ ,  $f \in \mathcal{H}$ .

**Definition C.1.** *Ein linearer Operator  $T$  auf  $\mathcal{H}$  ist eine Abbildung von  $\mathcal{H}$  in  $\mathcal{H}$  mit der Eigenschaft  $T(\alpha f + \beta g) = \alpha T(f) + \beta T(g)$  mit  $\alpha, \beta \in \mathbb{R}$  (bzw.  $\mathbb{C}$ ),  $f, g \in \mathcal{H}$ . Die Norm eines linearen Operators ist gegeben durch*

$$|T| := \sup\{\|T(f)\| : f \in \mathcal{H}, \|f\| \leq 1\}.$$

Der Operator heißt **beschränkt**, falls  $|T| < +\infty$ . Ein beschränkter Operator heißt **kompakt**, falls der Abschluss von

$$\{T(f) : f \in \mathcal{H}, \|f\| \leq 1\}$$

in der durch die Metrik  $\|\cdot - \cdot\|$  auf  $\mathcal{H}$  induzierten Topologie kompakt in der durch  $\|\cdot - \cdot\|$  induzierten Topologie ist.

Jedem beschränkten, linearen Operator  $T$  auf dem Hilbertraum  $\mathcal{H}$  wird durch die Beziehung

$$\langle Tf, g \rangle = \langle f, T^*g \rangle, \quad f, g \in \mathcal{H}$$

der so genannte **adjungierte Operator**  $T^*$  zugeordnet. Ein linearer Operator  $T$  auf  $\mathcal{H}$  heißt **selbstadjungiert**, falls

$$\langle Tf, g \rangle = \langle f, Tg \rangle, \quad f, g \in \mathcal{H}.$$

Ist  $T$  selbstadjungiert, so heißt  $T$  **positiv**, falls

$$\langle Tf, f \rangle \geq 0, \quad f \in \mathcal{H}.$$

Ein beschränkter, linearer Operator  $T$  heißt **normal**, falls  $TT^* = T^*T$ .

Für unsere Betrachtungen ist die Theorie der Eigenwerte von linearen Operatoren von großer Wichtigkeit.

**Definition C.2.** Sei  $T$  ein beschränkter, linearer Operator auf  $\mathcal{H}$ . Das **Spektrum**  $\sigma(T)$  von  $T$  ist die Menge aller  $\lambda \in \mathbb{C}$ , für welche die Abbildung  $\lambda \mathbf{1} - T$ , wobei  $\mathbf{1}$  die Identität auf  $\mathcal{H}$  bezeichne, keine beschränkte Inverse besitzt. Ein  $\lambda \in \sigma(T)$ , für das  $\lambda \mathbf{1} - T$  keine Bijektion ist, heißt **Eigenwert** von  $T$  und jedes  $f \in \mathcal{H}$ ,  $f \neq 0$ , mit  $Tf = \lambda f$  heißt **Eigenvektor** zum Eigenwert  $\lambda$ . Offensichtlich sind die Eigenwerte eines positiven Operators nicht-negativ. Die Umkehrung, dass ein Operator positiv ist, falls alle Eigenwerte nicht-negativ sind, gilt für beschränkte, normale Operatoren ebenfalls (vergleiche Dunford und Schwartz [DS88b, Theorem X.4.2, S.906]).

Anders als im Endlich-dimensionalen Fall gibt es bei allgemeinen Hilberträumen auch Punkte im Spektrum, die keine Eigenwerte sind. Der folgende Satz schließt diesen Fall für die in dieser Arbeit behandelten Operatoren allerdings aus.

**Satz C.3 (Dunford, Schwartz [DS88b, Corollary X.3.5, S.905]).** Das Spektrum  $\sigma(T)$  eines kompakten, normalen Operators  $T$  auf  $\mathcal{H}$  ist abzählbar und besitzt höchstens einen Häufungspunkt im Nullpunkt der komplexen Ebene. Jeder von Null verschiedene Wert  $\lambda \in \sigma(T)$  ist ein Eigenwert und die Anzahl linear unabhängiger Eigenvektoren zu  $\lambda$  ist endlich. Es existiert eine Orthonormalbasis von  $\mathcal{H}$ , die aus Eigenvektoren von  $T$  besteht.

Für die Konvergenz der Eigenwerte einer Operatorfolge gegen die Eigenwerte eines bestimmten Operators ist der folgende Satz hilfreich.

**Satz C.4 (Dunford, Schwartz [DS88b, Lemma XI.9.5, S.1091]).** Seien  $T_n$ ,  $T$ ,  $n \in \mathbb{N}$ , kompakte Operatoren auf  $\mathcal{H}$  und es gelte

$$\|T - T_n\| = \sup\{\|(T - T_n)(f)\| : f \in \mathcal{H}, \|f\| \leq 1\} \xrightarrow{n \rightarrow \infty} 0.$$

Es sei  $(\lambda_m(T))_m$  eine die Vielfachheit berücksichtigende Abzählung der nicht verschwindenden Eigenwerte von  $T$ . Dann existieren die Vielfachheit berücksichtigende Abzählungen  $(\lambda_m(T_n))_m$  der nicht verschwindenden Eigenwerte von  $T_n$ , so dass gleichmäßig in  $m$  gilt

$$\lambda_m(T_n) \xrightarrow{n \rightarrow \infty} \lambda_m(T).$$

## C.1 Operatoren der Hilbert-Schmidt-Klasse

Die als Kovarianzoperatoren in Kapitel 3 auftauchenden Abbildungen sind Integraloperatoren der Hilbert-Schmidt-Klasse.

**Definition C.5.** Sei  $(f_\alpha)_\alpha$  eine Orthonormalbasis von  $\mathcal{H}$ . Ein beschränkter, linearer Operator  $T$  heißt **Hilbert-Schmidt-Operator**, falls

$$|T|_{\text{HS}} := \left( \sum_{\alpha} \|T(f_\alpha)\|^2 \right)^{\frac{1}{2}} < +\infty.$$

$|T|_{\text{HS}}$  wird auch als **Hilbert-Schmidt-Norm** von  $T$  bezeichnet. Sie hängt nicht von der Wahl der Orthonormalbasis ab (vergleiche Dunford und Schwartz [DS88b, Lemma XI.6.2, S.1010]). Jeder Hilbert-Schmidt-Operator ist kompakt (vergleiche Dunford und Schwartz [DS88b, Lemma XI.6.6, S.1012]).

Der Nachweis, dass Kovarianzoperatoren zu stochastischen Prozessen von der Hilbert-Schmidt-Klasse sind, nutzt die  $L_2$ -Struktur des Hilbertraums  $\mathcal{H}$  aus. Den Zusammenhang knüpft das folgende Lemma.

**Lemma C.6 (Dunford, Schwartz [DS88b, ex. XI.8.44, S.1083]).** Ein Operator  $A$  auf dem Hilbertraum  $\mathcal{H} = L_2^{(\mathbb{C})}(R, \mathcal{R}, \mu)$  (d.h.  $L_2(R, \mathcal{R}, \mu)$  bzw.  $L_2^{\mathbb{C}}(R, \mathcal{R}, \mu)$ ) ist genau dann ein Hilbert-Schmidt-Operator, wenn eine messbare Abbildung  $a$  von  $R \times R$  nach  $\mathbb{R}$  (bzw.  $\mathbb{C}$ ) existiert, so dass

$$\left( \iint |a(s, t)|^2 d\mu(s) d\mu(t) \right)^{\frac{1}{2}} < +\infty \quad (\text{C.1})$$

und

$$(A(f))(s) = \int a(s, t)f(t) d\mu(t), \quad f \in \mathcal{H}, \quad (\text{C.2})$$

für  $\mu$ -fast alle  $s$ . Wenn eine solche Funktion  $a$  existiert, so ist diese  $\mu \otimes \mu$ -f.ü. eindeutig und die Hilbert-Schmidt-Norm stimmt mit dem Ausdruck in (C.1) überein. Sind umgekehrt (C.1) und (C.2) für eine messbare Abbildung  $a$  von  $R \times R$  nach  $\mathbb{R}$  (bzw.  $\mathbb{C}$ ) erfüllt, so wird durch (C.2) ein Hilbert-Schmidt-Operator auf  $\mathcal{H}$  definiert.

Für die Kerne von Operatoren der Hilbert-Schmidt-Klasse auf  $L_2^{(\mathbb{C})}(R, \mathcal{R}, \mu)$  lässt sich eine Eigenwertzerlegung durchführen.

**Lemma C.7 (Dunford, Schwartz [DS88b, ex. XI.8.56, S.1087]).** Sei  $A$  ein selbstadjungierter Hilbert-Schmidt-Operator auf  $L_2^{(\mathbb{C})}(R, \mathcal{R}, \mu)$ . Sei  $(f_\alpha)_\alpha$  eine Abzählung der normierten Eigenfunktionen von  $A$  zu den nicht verschwindenden Eigenwerten und  $(\lambda_\alpha)_\alpha$  die Folge der zugehörigen Eigenwerte. Dann gilt mit den Bezeichnungen von Lemma C.6

$$a(s, t) = \sum_{\alpha} \lambda_{\alpha} f_{\alpha}(s) \overline{f_{\alpha}(t)}, \quad s, t \in R,$$

wobei die Reihe in  $L_2^{(\mathbb{C})}((R, \mathcal{R}, \mu) \times (R, \mathcal{R}, \mu))$  konvergiert.

Auf Kompakta lässt sich eine noch schärfere Konvergenz erreichen, welche es zusammen mit Satz C.4 ermöglicht, Aussagen von Kompakta auf den ganzen  $\mathbb{R}^d$  zu erweitern.

**Lemma C.8 (Dunford, Schwartz [DS88b, ex. XI.8.58, S.1088]).** Sei  $R \in \mathcal{B}^d$  kompakt und  $\mu$  ein endliches Maß auf  $\mathcal{R}$ . Mit den Voraussetzungen und Bezeichnungen von Lemma C.6 sei  $a(\cdot, \cdot)$  eine stetige Funktion mit der Eigenschaft

$$a(s, t) = \overline{a(t, s)}, \quad s, t \in R.$$

Der zugehörige Operator  $A$  sei positiv. Dann gilt

$$a(s, t) = \sum_{\alpha} \lambda_{\alpha} f_{\alpha}(s) \overline{f_{\alpha}(t)}, \quad s, t \in R,$$

wobei die Reihe gleichmäßig konvergiert. ( $\bar{x}$  bezeichnet hier den zu  $x \in \mathbb{C}$  komplex-konjugierten Wert)

## C.2 Operatoren der Spurklasse

In der vorliegenden Arbeit werden die Eigenwerte von Operatoren als Gewichte einer Summe von unabhängigen  $\chi_1^2$ -verteilten Zufallsvariablen benutzt. Um aus der Konvergenz der Eigenwerte die Verteilungskonvergenz der mit diesen Eigenwerten gewichteten Summen von unabhängigen  $\chi_1^2$ -verteilten Zufallsvariablen zu folgern, ist es erforderlich, Aussagen über die Summe der Eigenwerte zu machen. Operatoren der Spurklasse erlauben eine besonders elegante Darstellung.

**Definition C.9.** Es sei  $A$  ein Operator der Hilbert-Schmidt-Klasse. Dann sagt man, dass  $A$  zur **Spurklasse** gehört, falls die Summe der Beträge der Eigenwerte  $(\lambda_{\alpha})_{\alpha}$  von  $A$  endlich ist,

$$\sum_{\alpha} |\lambda_{\alpha}| < +\infty.$$

In der Folge  $(\lambda_{\alpha})_{\alpha}$  tauche jeder von Null verschiedene Eigenwert entsprechend seiner Vielfachheit auf. Die **Spur** von  $A$  ist definiert als die Summe der Eigenwerte

$$\operatorname{tr}(A) := \sum_{\alpha} \lambda_{\alpha}.$$

Die folgende Aussage macht es möglich, die schwer zugängliche Summe der Eigenwerte durch ein Integral über die Diagonale des Kerns zu ersetzen.

**Lemma C.10 (Brislaw [Bri91, Corollary 3.2, S.237]).** Es sei  $A$  ein Operator der Spurklasse. Mit den Bezeichnungen von Lemma C.6 sei  $a(\cdot, \cdot)$  der stetige Kern von  $A$ . Dann gilt

$$\operatorname{tr}(A) = \sum_{\alpha} \lambda_{\alpha} = \int a(s, s) d\mu(s).$$

# Anhang D

## Zufallselemente mit Werten im Hilbertraum

Es sei  $\mathcal{H} := L_2(R, \mathcal{R}, \mu)$  der zu Beginn von Anhang C beschriebene Hilbertraum mit Innenprodukt  $\langle \cdot, \cdot \rangle$  und Norm  $\|\cdot\|$ . Die Definition von und die Voraussetzungen an  $R$ ,  $\mathcal{R}$  und  $\mu$  seien übernommen. Es bezeichne  $\mathcal{B}(\mathcal{H})$  die durch die Norm auf  $\mathcal{H}$  induzierte Borelsche  $\sigma$ -Algebra. Weiter bezeichne  $(e_j)_j$  eine Orthonormalbasis von  $\mathcal{H}$ . Für die folgenden Aussagen ist insbesondere von Wichtigkeit zu notieren, dass  $\mathcal{H}$  ein separabler Hilbertraum ist (siehe Cohn [Coh80, Proposition 3.4.5, S.110]). Mit dem Satz von Riesz-Fischer besteht damit ein Isomorphismus, der das innere Produkt erhält, so dass wesentliche Elemente der Theorie von  $\ell_2$ -wertigen Zufallselementen übertragen werden können. Die sich im Fall  $L_2$ -wertiger Zufallselemente ergebenden Resultate sind hier kurz zusammengefasst. Die Darstellung lehnt sich eng an die von Vakhania [Vak81] und Parthasarathy [Par67] an. Die Aussagen zum zentralen Grenzwertsatz sind dem Buch von Ledoux und Talagrand [LT91] und der Arbeit von Kundu, Majumdar und Mukherjee [KMM00] entnommen.

**Definition D.1.** *Ein Zufallselement  $\mathbb{X}$  mit Werten im separablen Hilbertraum  $\mathcal{H}$  ist eine  $\mathcal{A}/\mathcal{B}(\mathcal{H})$ -messbare Abbildung, wobei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum sei.*

Gilt für ein Zufallselement  $\mathbb{X}$  im Hilbertraum  $E\|\mathbb{X}\| := E\sqrt{\langle \mathbb{X}, \mathbb{X} \rangle} < \infty$ , so ist  $E\langle \mathbb{X}, \cdot \rangle$  eine lineare, reellwertige Abbildung auf  $\mathcal{H}$ , deren Operatornorm wegen der Cauchy-Schwarzschen Ungleichung durch  $E\|\mathbb{X}\|$  nach oben beschränkt ist. Mit dem Darstellungssatz von Fréchet-Riesz existiert damit ein eindeutiges Element  $a \in \mathcal{H}$ , so dass  $\langle a, f \rangle = E\langle \mathbb{X}, f \rangle$  für alle  $f \in \mathcal{H}$ . Dieses Element bezeichnet man auch als **Erwartungswert**  $E\mathbb{X} := a$  von  $\mathbb{X}$  [Par67, Definition 3.2, S.168]. Ist  $a = 0$ , so sagt man auch, das Zufallselement sei **zentriert**. Gilt überdies  $E\|\mathbb{X}\|^2 < \infty$ , so definiert man den **Kovarianzoperator**  $C$  von  $\mathbb{X}$  als den selbstadjungierten, linearen Operator, der durch die quadratische Form

$$\langle Cf, f \rangle := E[\langle \mathbb{X} - a, f \rangle^2] \quad \text{für alle } f \in \mathcal{H}$$

festgelegt wird [Par67, Definition 2.2, S.154]. Offensichtlich ist  $C$  positiv und ein Hilbert-Schmidt-Operator, denn mit der verallgemeinerten Cauchy-Schwarzschen Ungleichung, vergleiche Heuser [Heu92, Satz 29.2, S.195], gilt

$$\begin{aligned} |C|_{\text{HS}}^2 &= \sum_j \langle Ce_j, Ce_j \rangle = \sum_j \mathbb{E} [\langle \mathbb{X} - a, C(e_j) \rangle^2] \\ &\leq \left( \sum_j \langle Ce_j, e_j \rangle \right) \mathbb{E} [\langle C(\mathbb{X} - a), \mathbb{X} - a \rangle], \end{aligned}$$

wobei mit  $\mathbb{X}, \mathbb{Y}$  unabhängig und identisch verteilt

$$\mathbb{E} [\langle C(\mathbb{X} - a), \mathbb{X} - a \rangle] = \mathbb{E} [\langle \mathbb{X} - a, \mathbb{Y} - a \rangle^2] \leq (\mathbb{E} [\|\mathbb{X} - a\|^2])^2 < +\infty$$

und wegen  $\sum_j \langle f, e_j \rangle^2 = \langle \sum_j \langle f, e_j \rangle e_j, \sum_k \langle f, e_k \rangle e_k \rangle = \langle f, f \rangle = \|f\|^2$  für alle  $f \in \mathcal{H}$

$$\sum_j \langle Ce_j, e_j \rangle = \sum_j \mathbb{E} [\langle \mathbb{X} - a, e_j \rangle^2] = \mathbb{E} [\|\mathbb{X}\|^2] - \|a\|^2 < \infty.$$

Aus der letzten Zeile folgt zusätzlich die Spurklassen-Eigenschaft von  $C$ .

Der Begriff der Verteilungskonvergenz beziehungsweise schwachen Konvergenz von Wahrscheinlichkeitsmaßen ist bei Hilbertraum-wertigen Zufallselementen definiert wie im Fall reeller Zufallsvariablen.

**Definition D.2.** *Es seien  $\mu, (\mu_n)$  Wahrscheinlichkeitsmaße auf  $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ . Dann sagt man, die Folge  $(\mu_n)$  **konvergiert schwach** gegen  $\mu$ , wenn für alle beschränkten und stetigen Funktionen  $f : \mathcal{H} \rightarrow \mathbb{R}$  gilt*

$$\lim_{n \rightarrow \infty} \int f \, d\mu_n = \int f \, d\mu.$$

Schwache Konvergenz von  $(\mu_n)$  gegen  $\mu$  liegt vor, wenn die Folge  $(\mu_n)$  **relativ kompakt** ist, d.h. jede Teilfolge enthält eine Teil-Teilfolge, die schwach gegen ein Wahrscheinlichkeitsmaß auf  $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$  konvergiert, und dieses Wahrscheinlichkeitsmaß gleich  $\mu$  ist. Relative Kompaktheit liegt dem Satz von Prohorov folgend vor, wenn die Folge  $(\mu_n)$  **straff** ist, was bedeutet, dass es zu jedem  $\varepsilon > 0$  ein Kompaktum  $K$  in  $\mathcal{H}$  gibt, so dass

$$\mu_n(K) > 1 - \varepsilon \quad \text{für alle } n \in \mathbb{N}.$$

**Satz D.3 (Vakhania [Vak81, Theorem 3.1.2, S.54]).** *Sind die folgenden zwei Bedingungen für eine Folge  $(\mu_n)$  von Wahrscheinlichkeitsmaßen auf  $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$  erfüllt, so ist die Folge straff und damit auch relativ kompakt.*

$$(a) \sup_n \int \|x\|^2 \, d\mu_n(x) \leq \text{Konst.} < +\infty,$$

$$(b) \lim_{m \rightarrow \infty} \sup_n \sum_{k=m}^{\infty} \langle x, e_k \rangle^2 \, d\mu_n(x) = 0.$$

## D.1 Gaußsche Zufallselemente und der zentrale Grenzwertsatz

Bei der Entwicklung des zentralen Grenzwertsatzes für Hilbertraum-wertige Zufallselemente spielen die Gaußschen Zufallselemente wie gewohnt eine zentrale Rolle.

**Definition D.4.** Ein Zufallselement  $\mathbb{Z}$  auf  $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$  heißt **Gaußsches Zufallselement**, falls  $\langle f, \mathbb{Z} \rangle$  für alle  $f \in \mathcal{H}$  eine Normalverteilung besitzt. Ist  $\mathbb{Z}$  ein Gaußsches Zufallselement, so existiert ein Element  $a \in \mathcal{H}$  und ein linearer, selbstadjungierter, positiver Operator  $C$  der Spurklasse mit

$$\langle a, h \rangle = \mathbb{E}\langle \mathbb{Z}, h \rangle \quad \text{und} \quad \langle Ch, h \rangle = \mathbb{E}\langle \mathbb{Z} - a, h \rangle^2$$

für alle  $h \in \mathcal{H}$ . Das Element  $a$  heißt der **Erwartungswert** und der Operator  $C$  der **Kovarianzoperator** von  $\mathbb{Z}$ .

Für den gewöhnlichen zentralen Grenzwertsatz im Hilbertraum benötigt man lediglich recht schwache Voraussetzungen.

**Satz D.5 (Ledoux, Talagrand [LT91, Theorem 10.5, S.281]).** Es sei  $\mathbb{X}_1, \mathbb{X}_2, \dots$  eine Folge unabhängiger und identisch verteilter Zufallselemente in  $\mathcal{H}$  mit  $\mathbb{E}\|\mathbb{X}_1\|^2 < +\infty$  und  $\mathbb{E}\mathbb{X}_1 = 0$ . Dann gilt der zentrale Grenzwertsatz

$$\frac{\mathbb{X}_1 + \dots + \mathbb{X}_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathbb{Z} \quad \text{in } \mathcal{H},$$

wobei  $\mathbb{Z}$  ein zentriertes Gaußsches Zufallselement in  $\mathcal{H}$  mit demselben Kovarianzoperator wie  $\mathbb{X}_1$  ist.

Auch der zentrale Grenzwertsatz für Dreiecksschemata lässt sich für Hilbertraumwertige Zufallselemente formulieren. Dabei werden neben der gewohnten Endlichkeit der zweiten Momente  $\mathbb{E}\|\cdot\|^2 < \infty$ , eine Lindeberg-Bedingung sowie schwache Konvergenzanforderungen an den Kovarianzoperator benötigt. Die Darstellung des folgenden Satzes folgt der Arbeit von Kundu, Majumdar und Mukherjee [KMM00].

**Satz D.6 (Kundu et al. [KMM00, Theorem 1.1, S.266]).** Für jedes  $n \in \mathbb{N}$  sei  $\mathbb{X}_{n,1}, \dots, \mathbb{X}_{n,n}$  eine Folge unabhängiger, zentrierter Zufallselemente in  $\mathcal{H}$ , für die gilt  $\mathbb{E}\|\mathbb{X}_{n,j}\|^2 < +\infty$ ,  $1 \leq j \leq n$ . Es sei  $C_n$  der Kovarianzoperator von  $S_n := \sum_{j=1}^n \mathbb{X}_{n,j}$  und es seien folgende Voraussetzungen erfüllt

$$(i) \quad \lim_{n \rightarrow \infty} \langle C_n e_k, e_l \rangle = c_{kl}, \quad k, l \in \mathbb{N}.$$

$$(ii) \quad \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} \langle C_n e_k, e_k \rangle = \sum_{k=1}^{\infty} c_{kk} < \infty.$$

(iii) Für jedes  $\varepsilon > 0$  und jedes  $k \in \mathbb{N}$  gelte

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n \mathbb{E} [\langle \mathbb{X}_{n,j}, e_k \rangle^2 \mathbf{I}(|\langle \mathbb{X}_{n,j}, e_k \rangle| > \varepsilon)] = 0.$$

Dann gilt  $S_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathbb{Z}$  in  $\mathcal{H}$ , wobei  $\mathbb{Z}$  ein zentriertes, Gaußsches Zufallselement in  $\mathcal{H}$  ist, dessen Kovarianzoperator charakterisiert wird durch

$$\langle C e_k, e_k \rangle = c_{kk}, \quad k \in \mathbb{N}.$$

# Symbolverzeichnis

Allgemeines	Die neue Testfamilie
$\mathbb{R}^d$ . . . . .	$\phi$ . . . . .
$\mathcal{B}^d$ . . . . .	$\Phi$ . . . . .
$ \cdot $ . . . . .	$\Phi_0$ . . . . .
$M^{\mathbb{C}}$ . . . . .	$T_{m,n}^{\phi}$ . . . . .
$\mathcal{P}(M)$ . . . . .	$d_{\phi}^2$ . . . . .
$x'$ . . . . .	$\mathcal{F}_d^k(\phi)$ . . . . .
$\ \cdot\ _S$ . . . . .	$\rho$ . . . . .
$i$ . . . . .	$\mathcal{L}_2, \bar{\mathcal{L}}_2$ . . . . .
$I_d$ . . . . .	$H, h, \lambda, f$ . . . . .
$\mathbb{N}$ . . . . .	$\bar{H}, \bar{h}, \bar{\lambda}, \bar{f}$ . . . . .
$\mathbb{N}_0$ . . . . .	$\gamma, \bar{\gamma}$ . . . . .
$C_b$ . . . . .	$X_i$ . . . . .
$I(\cdot)$ . . . . .	$\mathcal{H}$ . . . . .
$\text{erf}(\cdot)$ . . . . .	$C, c, \tilde{\lambda}, e$ . . . . .
$\langle \cdot, \cdot \rangle, \ \cdot\ $ . . . . .	$\tau$ . . . . .
$ \cdot _{\text{HS}}$ . . . . .	$\mathbb{J}_{\{x_1, \dots, x_m\}; P^X}$ . . . . .
<hr/>	
Wahrscheinlichkeitstheorie	$b(\theta)$ . . . . .
$N_d(\mu, \Sigma)$ . . . . .	$c(\theta)$ . . . . .
unif . . . . .	
$\delta_x$ . . . . .	
$X \sim \mu$ . . . . .	
$X \stackrel{\mathcal{D}}{=} Y$ . . . . .	
u.i.v. . . . . .	
$\stackrel{\mathcal{D}}{\rightarrow}$ . . . . .	
$\implies$ . . . . .	
$P^X \otimes P^Y$ . . . . .	
$\varphi_X$ . . . . .	
$\varphi_{\{x_1, \dots, x_m\}}$ . . . . .	



# Literaturverzeichnis

- [Bah60] Bahadur, R. R. (1960). *Stochastic comparison of tests*. Ann. Math. Statist. **31**, 276–295.
- [Bah96] Bahr, R. (1996). *Ein neuer Test für das mehrdimensionale Zwei-Stichproben-Problem bei allgemeiner Alternative*. Dissertation, Universität Hannover.
- [BF04] Baringhaus, L., Franz, C. (2004). *On a new multivariate two-sample test*. J. Multivariate Analysis **88**, 190–206.
- [Bar37] Bartlett, M. S. (1937). *Properties of sufficiency and statistical tests*. Proc. Roy. Soc. London Ser. A **168**, 268–282.
- [Bau90] Bauer, H. (1990). *Maß- und Integrationstheorie*. Zweite Auflage. Walter de Gruyter & Co., Berlin.
- [Bau91] Bauer, H. (1991). *Wahrscheinlichkeitstheorie*. Vierte Auflage. Walter de Gruyter & Co., Berlin.
- [BCR84] Berg, C., Christensen, J. P. R., Ressel, P. (1984). *Harmonic analysis on semigroups*, Band 100 aus *Graduate Texts in Mathematics*. Springer-Verlag, New York.
- [Bil86] Billingsley, P. (1986). *Probability and measure*. Zweite Auflage. John Wiley & Sons Inc., New York.
- [Bil99] Billingsley, P. (1999). *Convergence of probability measures*. Zweite Auflage. John Wiley & Sons Inc., New York.
- [Bri91] Brislawn, C. (1991). *Traceable integral kernels on countably generated measure spaces*. Pacific J. Math. **150**, 229–240.
- [Coh80] Cohn, D. L. (1980). *Measure theory*. Birkhäuser, Boston.
- [Cra28] Cramér, H. (1928). *On the composition of elementary errors, 2nd paper*. Skand. Aktuarietidskr. **11**, 141–180.

- [DS86] D'Agostino, R. B., Stephens, M. A., editors (1986). *Goodness-of-fit techniques*, Band 68 aus *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York.
- [dWR87] de Wet, T., Randles, R. H. (1987). *On the effect of substituting parameter estimators in limiting  $\chi^2$   $U$  and  $V$  statistics*. Ann. Statist. **15**, 398–412.
- [Deu98] Deuber, W. (1998). *2. Problem of W. Deuber, Problem 303*. Discrete Math. **192**, 348.
- [Dud89] Dudley, R. M. (1989). *Real analysis and probability*. Wadsworth & Brooks/Cole, Pacific Grove.
- [DS88a] Dunford, N., Schwartz, J. T. (1988). *Linear operators. Part I*. John Wiley & Sons Inc., New York.
- [DS88b] Dunford, N., Schwartz, J. T. (1988). *Linear operators. Part II*. John Wiley & Sons Inc., New York.
- [Dur73] Durbin, J. (1973). *Distribution theory for tests based on the sample distribution function*. Society for Industrial and Applied Mathematics, Philadelphia.
- [DK72] Durbin, J., Knott, M. (1972). *Components of Cramér-von Mises Statistics I*. J. Roy. Statist. Soc. B **34**, 290–307.
- [ET93] Efron, B., Tibshirani, R. J. (1993). *An introduction to the bootstrap*, Band 57 aus *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.
- [Eic92] Eichelsbacher, P. (1992). *Varadhans Prinzip großer Abweichungen für spezielle Statistikklassen*. Dissertation, Universität Bielefeld.
- [Fra00] Franz, C. (2000). *Ein statistischer Test für das mehrdimensionale Zweistichproben-Problem*. Diplomarbeit, Universität Hannover.
- [Fra57] Fraser, D. A. S. (1957). *Nonparametric methods in statistics*. John Wiley & Sons Inc., New York.
- [GS77] Gänszler, P., Stute, W. (1977). *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin.
- [GHS93] Guo, K., Hu, S., Sun, X. (1993). *Conditionally positive definite functions and Laplace-Stieltjes integrals*. J. Approx. Theory **74**, 249–265.
- [Gur48] Gurland, J. (1948). *Inversion formulae for the distribution of ratios*. Ann. Math. Statist. **19**, 228–237.

- [Hen84] Henze, N. (1984). *Über die Anzahl von Zufallspunkten mit typ-gleichem nächsten Nachbarn und einen multivariaten Zwei-Stichproben-Test*. *Metrika* **31**, 259–273.
- [Hen88] Henze, N. (1988). *A multivariate two-sample test based on the number of nearest neighbor type coincidences*. *Ann. Statist.* **16**, 772–783.
- [Heu92] Heuser, H. (1992). *Funktionalanalysis*. Dritte Auflage. B. G. Teubner, Stuttgart.
- [HS65] Hewitt, E., Stromberg, K. (1965). *Real and abstract analysis. A modern treatment of the theory of functions of a real variable*. Springer-Verlag, New York.
- [HL56] Hodges, J. L., Lehmann, E. L. (1956). *The efficiency of some nonparametric competitors of the  $t$ -test*. *Ann. Math. Statist.* **27**, 324–335.
- [Hol67] Hollander, M. (1967). *Asymptotic efficiency of two nonparametric competitors of Wilcoxon's two sample test*. *J. Amer. Statist. Assoc.* **62**, 939–949.
- [HW73] Hollander, M., Wolfe, D. A. (1973). *Nonparametric statistical methods*. John Wiley & Sons, New York-London-Sydney.
- [Joh87] Johnson, M. E. (1987). *Multivariate Statistical Simulation*. John Wiley & Sons Inc., New York.
- [JKB97] Johnson, N. L., Kotz, S., Balakrishnan, N. (1997). *Discrete multivariate distributions*. John Wiley & Sons Inc., New York.
- [JR71] Joiner, B. L., Rosenblatt, J. R. (1971). *Some Properties of the Range in Samples from Tukey's Symmetric Lambda Distribution*. *J. Amer. Statist. Assoc.* **66**, 394–399.
- [KB94] Koroljuk, V. S., Borovskich, Y. V. (1994). *Theory of  $U$ -statistics*, Band 273 aus *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.
- [KMM00] Kundu, S., Majumdar, S., Mukherjee, K. (2000). *Central limit theorems revisited*. *Statist. Probab. Lett.* **47**, 265–275.
- [LT91] Ledoux, M., Talagrand, M. (1991). *Probability in Banach spaces*, Band 23 aus *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*. Springer-Verlag, Berlin.
- [Leh53] Lehmann, E. L. (1953). *The power of rank tests*. *Ann. Math. Statist.* **24**, 23–43.

- [Leh86] Lehmann, E. L. (1986). *Testing statistical hypotheses*. Zweite Auflage. John Wiley & Sons Inc., New York.
- [Mat90] Mattner, L. (1990). *Behandlung einiger Extremalprobleme für Wahrscheinlichkeitsverteilungen*. Dissertation, Universität Hannover.
- [Moo54] Mood, A. M. (1954). *On the asymptotic efficiency of certain non-parametric two-sample tests*. Ann. Math. Statist. **25**, 514–522.
- [Mor01] Morgenstern, D. (2001). *Proof of a conjecture by Walter Deuber concerning the distances between points of two types in  $\mathbf{R}^d$* . Discrete Math. **226**, 347–349.
- [Mui82] Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York.
- [Neu77] Neuhaus, G. (1977). *Functional limit theorems for  $U$ -statistics in the degenerate case*. J. Multivariate Analysis **7**, 424–439.
- [NP02] Nikitin, Y. Y., Ponikarov, E. (2002). *Large deviations of non-degenerate two-sample von Mises functionals*. ICER Working Paper Ser. 11/2002 .
- [OGH<sup>+</sup>60] Olkin, I., Ghurye, S. G., Hoeffding, W., Madow, W. G., Mann, H. B. (1960). *Contributions to probability and statistics: Essays in honor of Harold Hotelling*, Band 2 aus *Stanford Studies in Mathematics and Statistics*. Stanford University Press, Stanford.
- [Owe62] Owen, D. B. (1962). *Handbook of statistical tables*. Addison-Wesley Publishing Co., Inc., Reading.
- [Par67] Parthasarathy, K. R. (1967). *Probability measures on metric spaces*. Academic Press Inc., New York.
- [PL76] Perng, S. K., Littell, R. C. (1976). *A test of equality of two normal population means and variances*. J. Amer. Statist. Assoc. **71**, 968–971.
- [Sch38] Schoenberg, I. J. (1938). *Metric Spaces And Completely Monotone Functions*. Ann. Math. **39**, 811–841.
- [Sen74] Sen, P. K. (1974). *Weak convergence of generalized  $U$ -statistics*. Ann. Prob. **2**, 90–102.
- [Stu70] Stummel, F. (1970). *Diskrete Konvergenz linearer Operatoren. I*. Math. Ann. **190**, 45–92.
- [Stu71] Stummel, F. (1971). *Diskrete Konvergenz linearer Operatoren. II*. Math. Z. **120**, 231–264.

- [Stu72] Stummel, F. (1972). *Diskrete Konvergenz linearer Operatoren. III*. In *Linear operators and approximation (Proc. Conf., Oberwolfach, 1971)*. Birkhäuser, Basel, 196–216. Internat. Ser. Numer. Math., Vol. 20.
- [Szé02] Székely, G. (2002).  *$\mathcal{E}$ -statistics: the Energy of Statistical Samples*. Ann. Statist. (eingereicht) .
- [SR04] Székely, G. J., Rizzo, M. L. (2004). *A new test for multivariate normality*. J. Multivariate Analysis .
- [Vak81] Vakhania, N. N. (1981). *Probability distributions on linear spaces*. North-Holland Publishing Co., New York.
- [VTC87] Vakhania, N. N., Tarieladze, V. I., Chobanyan, S. A. (1987). *Probability distributions on Banach spaces*, Band 14 aus *Mathematics and its Applications (Soviet Series)*. D. Reidel Publishing Co., Dordrecht.
- [Wie74] Wieand, H. S. (1974). *A condition under which the Pitman and Bahadur approaches to efficiency coincide*. Dissertation, University of Maryland.
- [Wie76] Wieand, H. S. (1976). *A condition under which the Pitman and Bahadur approaches to efficiency coincide*. Ann. Statist. **4**, 1003–1011.
- [WN70] Witting, H., Nölle, G. (1970). *Angewandte mathematische Statistik. Optimale finite und asymptotische Verfahren*, Band 14 aus *Leitfäden der angewandten Mathematik und Mechanik*. B. G. Teubner, Stuttgart.
- [Zol61] Zolotarev, V. M. (1961). *Concerning a certain probability problem*. Theor.Prob.Appl. **6**, 201–202.



## Lebenslauf

Name Carsten Franz  
geboren am 7.2.1976 in Hannover  
Familienstand ledig

### Bildungsweg

1982 - 1986 Grundschule Arnum  
1986 - 1988 Orientierungsstufe der Carl-Friedrich-Gauß-Schule Hemmingen  
1988 - 1995 Gymnasialzweig der Carl-Friedrich-Gauß-Schule Hemmingen

### Studium

1995 - 1998 Studium der Mathematik mit Nebenfach Physik  
an der Universität Hannover  
1998 - 1999 Studium an der Brunel University (London)  
Abschluss als „*Master of Science in Computational Mathematics*“  
1999 - 2001 Fortsetzung des Mathematik Studiums  
Abschluss als „*Diplom-Mathematiker*“

### Hochschultätigkeit

1998, 2000 studentische Hilfskraft am Institut für Mathematische Stochastik  
der Universität Hannover  
2001- wissenschaftlicher Mitarbeiter des Instituts für Mathematische  
Stochastik der Universität Hannover