

**Asymptotische Untersuchungen zur Klasse der  
BHEP–Tests auf multivariate Normalverteilung  
mit festem und variablem Glättungsparameter**

Zur Erlangung des akademischen Grades eines  
DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik der  
Universität Karlsruhe  
genehmigte

DISSERTATION

von

Dipl.-Math. Nora Gürtler  
aus Mannheim

Tag der mündlichen Prüfung:

20.12.2000

Referent:

Prof. Dr. N. Henze

Korreferent:

Prof. Dr. L. Baringhaus

Karlsruhe

2000



# Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftliche Angestellte am Institut für Mathematische Stochastik der Universität Karlsruhe. Mein herzlicher Dank geht an all diejenigen, die mich bei ihrer Anfertigung unterstützt, ermutigt und beraten haben:

- zuerst an Herrn Prof. Dr. Norbert Henze für die Anregung des Themas, sein Interesse am Fortgang der Arbeit, die freundliche Unterstützung im allgemeinen und die angenehme Atmosphäre am Institut
- und an Herrn Prof. Dr. L. Baringhaus für die freundliche Übernahme des Korreferats,
- desweiteren an meine Kollegen Dr. Bernhard Klar und Dipl.-Math. oec. Guido Bader für die vielen angeregten Diskussionen und Ratschläge, die diese Arbeit während ihrer Entstehung begleitet haben; Herrn Bader danke ich überdies für geduldiges Korrekturlesen;
- an Herrn HDoz. Dr. P. Ghanaat für eine hilfreiche Diskussion über Differentialkalkül und Herrn HDoz. Dr. G. Herzog bzw. Herrn HDoz. Dr. A. Müller für verschiedene Hinweise auf dem Gebiet der Analysis bzw. der Stochastik,
- an Herrn Dr. Christian Riede für Unterstützung in vielfältiger Weise, insbesondere für den Zugang zu einer Rechner-Umgebung von seltener Qualität,
- an meine Eltern dafür, daß sie mir auf meinem gesamten bisherigen Lebensweg zur Seite standen und mir jegliche benötigte Unterstützung zukommen ließen, und
- last, but not least, an alle Mitglieder des Instituts für Mathematische Stochastik, insbesondere an Frau Ingrid Voß und Frau Michaela Taßler, für das angenehme Arbeitsklima während der Jahre am Institut.

Karlsruhe, den 20.12.2000



# Inhaltsverzeichnis

<b>Einleitung</b>	<b>1</b>
<b>1. Der BHEP-Test</b>	<b>5</b>
1.1. Konvergenzaussagen	8
1.2. Schätzen der Varianz $\tau_\beta^2$	21
1.3. Anwendungen der Konvergenzaussagen	28
1.3.1. Asymptotischer Konfidenzbereich für $L_\beta$	28
1.3.2. Gütenäherung für den BHEP-Test	28
1.3.3. Inverser Test auf Normalverteilung	30
1.3.4. Betrachtung des Grenzwertes $L_\beta$	33
1.3.5. Übertragbarkeit auf andere Modelle	36
1.4. Simulationen an Beispielverteilungen	40
1.4.1. Die Gleichverteilung	43
1.4.2. Die Laplace-Verteilung	44
1.4.3. Die Exponentialverteilung	45
1.4.4. Die erste Normal Mixture	46
1.4.5. Die zweite Normal Mixture	46
1.4.6. Die dritte Normal Mixture	47
1.4.7. Empirische Resultate in den sechs Beispielen	48
1.5. Technische Lemmata und Beweise	57
<b>2. Die Bowman-Foster Statistik</b>	<b>72</b>
2.1. Historischer Kontext	74
2.1.1. Eigenschaften von Kerndichteschätzern	75
2.1.2. Der Bickel-Rosenblatt-Test	77
2.1.3. Vorhandene Resultate zum Bickel-Rosenblatt-Test	80
2.1.4. Offene Probleme	82
2.2. Situation und Notationen	82
2.3. $H_0$ -Konvergenz der Bowman-Foster-Statistik	91
2.4. $H_1$ -Konvergenz der Bowman-Foster-Statistik	101
2.5. Lokale Alternativen	109
2.5.1. Asymptotik ohne geschätzte Parameter	110
2.5.2. Behandlung der geschätzten Parameter	116
2.6. Technische Lemmata	125
<b>A. Vektor- und Matrixnormen</b>	<b>149</b>

<b>B. Schwache Konvergenz und Gaußprozesse in Hilberträumen</b>	<b>152</b>
B.1. Operatoren auf Hilberträumen . . . . .	152
B.2. Maße und Zufallselemente auf Hilberträumen . . . . .	153
B.2.1. Erwartungswert und Kovarianzoperator . . . . .	153
B.2.2. Straffheit und Verteilungskonvergenz . . . . .	154
B.2.3. Das charakteristische Funktional . . . . .	155
B.2.4. Gaußsche Zufallselemente auf Hilberträumen . . . . .	156
B.3. Zentrale Grenzwertsätze auf Hilberträumen . . . . .	156
B.4. Gaußprozesse und Gaußsche Zufallselemente in $L^2$ . . . . .	157
<b>C. Hilfsmittel aus dem Differentialkalkül</b>	<b>160</b>
C.1. Der allgemeine Differentialkalkül . . . . .	160
C.2. Differentiation spezieller Funktionen . . . . .	165
C.2.1. Die Funktion $\text{sq}$ . . . . .	166
C.2.2. Die Funktion $\text{sqr}$ . . . . .	167
C.2.3. Die Funktion $\Upsilon$ . . . . .	168
C.2.4. Verkettungen der bisher betrachteten Funktionen . . . . .	168
C.3. Taylor-Entwicklung der Funktion $K_h$ . . . . .	169
<b>D. <math>U</math>-Statistiken</b>	<b>180</b>
D.1. Fast sichere Konvergenz von $U$ - bzw. $V$ -Statistiken . . . . .	180
D.2. $U$ -Statistiken mit variablem Kern . . . . .	181
<b>E. Einige Aussagen der asymptotischen Statistik</b>	<b>185</b>
E.1. Benachbarte Alternativen . . . . .	185
E.2. Asymptotische relative Effizienz . . . . .	187
<b>Symbolverzeichnis</b>	<b>189</b>
<b>Literaturverzeichnis</b>	<b>191</b>

# Einleitung

In der induktiven Statistik faßt man Daten  $x_1, \dots, x_n$  als Realisierungen von Zufallsvariablen  $X_1, \dots, X_n$  auf. Um Schlußfolgerungen ziehen zu können, die über die bloße Beschreibung der Daten hinausgehen, sind Annahmen über ein zugrundeliegendes stochastisches Modell nötig, das die Verteilung von  $X_1, \dots, X_n$  beschreibt. Einerseits ist eine sehr genaue Spezifikation dieses Modells wünschenswert, damit möglichst präzise Aussagen getroffen werden können; ein Beispiel hierfür ist der klassische Ansatz unabhängiger und identisch standard-normalverteilter Zufallsvariablen  $X_1, \dots, X_n$ . Andererseits brechen auf einem derartigen Ansatz beruhende statistische Methoden unter Umständen zusammen, falls das Modell die Daten doch nicht genau genug beschreibt. Unter diesem Gesichtspunkt wären möglichst schwache Annahmen zu bevorzugen. Schwächere Voraussetzungen hingegen werden in der Regel nur schwächere Schlußfolgerungen zulassen.

Ein Kompromiß in diesem Dilemma ist die Wahl eines *Modellrahmens*, z.B. einer parametrischen Verteilungsfamilie, und das Überprüfen mit statistischen Methoden, ob dieser Rahmen mit den vorliegenden Daten verträglich ist. Kann dies bejaht werden, so können z.B. Parameter dieses Modellrahmens anhand der Daten angepaßt und auf das Rahmenmodell aufbauende statistische Verfahren angewandt werden.

Im folgenden treffen wir die grundlegende Annahme, daß die Daten  $X_1, \dots, X_n$  unabhängige Realisierungen einer zugrundeliegenden Verteilung sind. Diese Voraussetzung ist in der Praxis bei kontrollierten Experimenten oft zumindest näherungsweise erfüllt. Die Überprüfung, ob ein bestimmter Modellrahmen mit vorliegenden Beobachtungen verträglich ist, geschieht mittels sogenannter statistischer Anpassungstests. Bezeichnet  $\{P_\vartheta | \vartheta \in \Theta\}$  eine Verteilungsfamilie, die durch einen endlichdimensionalen Vektor  $\vartheta$  parametrisiert ist, und schreibt man  $P$  für die den Daten zugrundeliegende unbekannte Wahrscheinlichkeitsverteilung, so testet ein Anpassungstest die Hypothese

$$H_0 : P \in \{P_\vartheta | \vartheta \in \Theta\} \quad \text{gegen die allgemeine Alternative} \quad H_1 : P \notin \{P_\vartheta | \vartheta \in \Theta\} .$$

Ein solches Verfahren beruht auf einer *Teststatistik*, einer Funktion der Daten, und lehnt die Hypothese ab, wenn der Wert dieser Statistik auf den vorliegenden Beobachtungen in einem noch zu definierenden *kritischen Bereich* oder *Ablehnbereich* liegt. Hierfür benötigt man ein Vergleichskriterium, das prüft, ob die vorliegenden Daten aus  $\{P_\vartheta | \vartheta \in \Theta\}$  stammen können oder nicht, also eine Art Abstand zwischen einer Verteilung und der Familie  $\{P_\vartheta | \vartheta \in \Theta\}$ . Zu große Werte dieses Abstandes wird man dem kritischen Bereich zuordnen.

In den in der Praxis interessanten Fällen handelt es sich üblicherweise um eine *zusammengesetzte Hypothese*, d.h. die Familie  $\{P_\vartheta | \vartheta \in \Theta\}$  enthält im Gegensatz zur einfachen Hypothese mehr

als ein Element. Dies macht das Problem mathematisch komplizierter als bei einer einfachen Hypothese, denn es muß zusätzlich der Modellparameter  $\vartheta$  durch einen konsistenten Schätzer  $\hat{\vartheta}_n$  aus den Daten geschätzt werden. Resultat dieser Schätzung sollte der „den Daten am nächsten liegende“ Vertreter  $P_{\hat{\vartheta}_n}$  aus der parametrischen Familie  $\{P_\vartheta | \vartheta \in \Theta\}$  sein. Ein Anpassungstest für  $H_0$  gegen  $H_1$  basiert typischerweise auf einem eine Verteilung charakterisierenden Merkmal und vergleicht dessen Wert auf den Daten mit seinem Wert auf dem Vertreter  $P_{\hat{\vartheta}_n}$  der Verteilungsfamilie. Ist die Abweichung zu groß, hat man Anlaß, an der Beschreibung der Daten durch das Modell  $\{P_\vartheta | \vartheta \in \Theta\}$  zu zweifeln.

Ein wichtiger Modellrahmen in der Statistik ist die Lokations–Skalenfamilie der *multivariaten Normalverteilungen*, also die Familie  $\mathcal{N}_d := \{\mathcal{N}_d(\mu, \Sigma) | \mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d} \text{ positiv definit}\}$ , mit dem Lokations– und Skalenparameter  $(\mu, \Sigma)$  als zu schätzendem Parameter.

Klassische Anpassungstests für univariate Daten verwenden die Tatsache, daß die Verteilung  $P_\vartheta$  durch ihre Verteilungsfunktion  $F(\cdot, \vartheta)$  eindeutig bestimmt ist. Bezeichnen  $\hat{\vartheta}_n$  einen konsistenten Schätzer für den Modellparameter  $\vartheta$  und  $F_n(t) := n^{-1} \sum_{j=1}^n \mathbf{1}\{X_j \leq t\}$ ,  $t \in \mathbb{R}$ , die empirische Verteilungsfunktion der Daten  $X_1, \dots, X_n$ , so basiert eine wichtige Gruppe dieser Tests auf einem geeigneten Abstandsmaß zwischen  $F_n(\cdot)$  und  $F(\cdot, \hat{\vartheta}_n)$ . Der *Kolmogorov–Smirnov–Test*  $K_n$  verwendet als Abstand die Supremumsnorm auf  $\mathbb{R}$ , der *Cramér–von Mises–Test*  $C_n$  den  $L^2$ –Abstand bezüglich dem Maß  $dF(\cdot, \hat{\vartheta}_n)$ . Damit sind die diesen Tests zugrundeliegenden Statistiken gegeben durch

$$K_n := \sqrt{n} \sup_{t \in \mathbb{R}} |F_n(t) - F(t, \hat{\vartheta}_n)|, \quad C_n := n \int_{-\infty}^{\infty} (F_n(t) - F(t, \hat{\vartheta}_n))^2 dF(t, \hat{\vartheta}_n).$$

Modernere Verfahren beruhen auf anderen Merkmalen, die eine Verteilung charakterisieren und die insbesondere einer multivariaten Verallgemeinerung leichter zugänglich sind. So basiert z.B. der *BHEP–Test* auf multivariate Normalverteilung, ein Nachfolger des *Epps–Pulley–Tests*, auf einem gewichteten  $L^2$ –Abstand zwischen der charakteristischen Funktion  $\Psi(t) = \exp\{-\|t\|^2/2\}$  der multivariaten Standard–Normalverteilung und der empirischen charakteristischen Funktion  $\Psi_n(t) = n^{-1} \sum_{j=1}^n \exp\{it^T Y_j\}$  der mit dem Stichprobenmittel  $\bar{X}_n$  und der empirischen Kovarianzmatrix  $S_n$  standardisierten Daten  $Y_j = S_n^{-1/2}(X_j - \bar{X}_n)$ . Dieses Verfahren berücksichtigt die Problematik der zusammengesetzten Hypothese im Spezialfall der Lokations–Skalenfamilie der multivariaten Normalverteilungen in anderer Form als bisher beschrieben: Anstatt einen Schätzer  $\hat{\vartheta}_n$  für die unbekanntes Modellparameter einzusetzen, erfolgt eine *Standardisierung* der Daten, was zur Folge hat, daß unter  $H_0$  ohne Beschränkung der Allgemeinheit eine Standard–Normalverteilung angenommen werden kann. Die dem BHEP–Test zugrundeliegende Statistik ist somit definiert durch

$$T_{n,\beta} := n \int_{\mathbb{R}^d} \left| \Psi_n(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt,$$

wobei  $\varphi_\beta$  die Dichte der Normalverteilung  $\mathcal{N}_d(0, \beta^2 I_d)$  bezeichnet, mit einem Glättungsparameter  $\beta$ , auf dessen Rolle zu einem späteren Zeitpunkt eingegangen wird.

Es gibt verschiedene Kriterien, denen ein „guter“ Anpassungstest genügen sollte. Eine Mindestanforderung ist die *globale Konsistenz*, d.h. die Konsistenz gegenüber einer möglichst großen Klasse von alternativen Verteilungen; Tests mit dieser Eigenschaft nennt man *Omnibus*

*Tests.* Unter der Konsistenz eines Tests gegenüber einer Alternative versteht man, daß er sie erkennen kann: Liegt die Alternative den Daten zugrunde, soll die Wahrscheinlichkeit, daß der Test die Hypothese ablehnt, mit wachsendem Stichprobenumfang  $n$  gegen 1 konvergieren.

Ist das parametrische Modell  $\{P_\vartheta | \vartheta \in \Theta\}$  eine Lokations–Skalen–Familie, so ist die Invarianz gegenüber affinen Transformationen der Daten eine weitere wünschenswerte Eigenschaft einer Teststatistik. Denn die Tatsache, wie gut sich vorliegende Daten durch ein solches Modell beschreiben lassen, sollte nicht von deren Lage bzw. Skalierung abhängen. Eine möglichst hohe Güte auch bei kleineren und mittleren Stichprobenumfängen  $n$  sowie eine nichtdegenerierte Güte gegenüber benachbarten Alternativen, die mit der Geschwindigkeit  $n^{-1/2}$  gegen die Hypothese konvergieren sind weitere Eigenschaften, die einen guten Anpassungstest gegenüber konkurrierenden Verfahren auszeichnen. Desweiteren sollte die Teststatistik für vorliegende Daten numerisch stabil und schnell zu berechnen sein.

Da die Alternativen zu dem Modell  $\{P_\vartheta | \vartheta \in \Theta\}$  in der Regel nichtparametrisch sind, d.h. nicht endlichdimensional parametrisiert werden können, existiert kein Test, der die größte Güte gegenüber jeder Alternative besitzt (in eingeschränkteren Situationen ist dies anders). Dennoch möchte man in einer konkreten Situation mehrere zur Auswahl stehende Testverfahren vergleichen können, und als Kriterium kommt in erster Linie deren Güte gegenüber allgemeinen oder spezielleren Alternativen in Frage. In der vorliegenden Arbeit richtet sich die Aufmerksamkeit deshalb auf das Güteverhalten von Anpassungstests bei *festem, endlichem* Stichprobenumfang  $n$ . Hierzu sind theoretische Aussagen bisher nur sehr eingeschränkt vorhanden. In der Praxis beschränkt man sich daher meist darauf, die Güte eines Anpassungstests für verschiedene endliche Werte von  $n$  durch möglichst sorgfältige und umfangreiche Monte–Carlo–Simulationsstudien zu untersuchen. „Umfangreich“ bedeutet in diesem Zusammenhang, daß möglichst viele Alternativen und Stichprobenumfänge berücksichtigt werden.

Kapitel 1 der vorliegenden Arbeit schlägt ein theoretisch begründetes Verfahren zur Annäherung der Güte des wichtigen BHEP–Tests auf multivariate Normalverteilung vor. Dabei wurde der BHEP–Test unter mehreren zur Verfügung stehenden Anpassungstests ausgewählt, weil er die oben angesprochenen Anforderungen an ein gutes multivariates Testverfahren in besonderem Maße erfüllt. Für die Gütenäherung ist eine Konvergenzaussage der zugehörigen Statistik unter festen zugrundeliegenden Alternativen vonnöten, die in Abschnitt 1.1 hergeleitet wird. Der Beweis dieser Konvergenzaussage beruht auf Grenzwertsätzen für Zufallselemente mit Werten in einem Hilbertraum, die in Anhang B zusammengestellt sind. In Abschnitt 1.2 werden Schätzer für einen von der unbekanntem zugrundeliegenden Verteilung abhängigen Parameter hergeleitet. Abschnitt 1.3 konkretisiert das vorgeschlagene Verfahren zur Gütenäherung und erläutert weitere statistische Anwendungen der theoretischen Resultate. In Abschnitt 1.4 werden die vorgeschlagenen Verfahren an sechs alternativen Beispielverteilungen geprüft.

Die gesamten Untersuchungen in Kapitel 1 wurden unter der Prämisse durchgeführt, daß der Wert des Glättungsparameters  $\beta$  fest bleibt. Läßt man  $\beta$  jedoch mit dem Stichprobenumfang  $n$  *variieren*, gelangt man zu einer zweiten Darstellung der BHEP–Statistik: Bei variierendem Glättungsparameter ist diese Größe identisch mit der *Bowman–Foster–Statistik*, die in Bowman und Foster (1993) zur Realisierung eines Anpassungstests auf multivariate Normalverteilung vorgeschlagen wurde. Interessanterweise entstammt diese Bowman–Foster–Statistik

einem völlig anderen Kontext, nämlich dem der Kerndichteschätzung. Dies führte zu dem zweiten Schwerpunkt der vorliegenden Arbeit, nämlich der Frage nach der Grenzverteilung der Bowman–Foster–Statistik unter der Hypothese  $H_0$  und unter festen bzw. lokalen Alternativen, die in Kapitel 2 untersucht wird. Dort wird zunächst die Verbindung zwischen der BHEP–Statistik und der Bowman–Foster–Statistik präzisiert. Da letztere in engem Zusammenhang mit der *Bickel–Rosenblatt–Statistik* steht, die Gegenstand zahlreicher Veröffentlichungen ist, folgt ein Überblick über bereits vorhandene Resultate in Abschnitt 2.1. Nach einer exakten Formulierung der von dort an zugrundegelegten Situation in Abschnitt 2.2 gelangt man zu den eigentlichen Resultaten in den Abschnitten 2.3, 2.4 und 2.5, d.h. der Asymptotik der Bowman–Foster–Statistik in den drei oben beschriebenen Fällen. Sowohl in Kapitel 1 als auch in Kapitel 2 wurden Rechnungen und Lemmata von vorwiegend technischer Natur in eigene Abschnitte an das Ende des jeweiligen Kapitels ausgelagert.

An die eigentliche Arbeit schließt sich ein mehrteiliger Anhang an, in dem Hilfsmittel aus der Analysis, Maßtheorie und Stochastik zusammengefaßt und Beweise der Resultate angegeben werden, für die keine geeigneten Zitate in der Literatur ausfindig gemacht werden konnten. Anhang A greift kurz den Zusammenhang zwischen den verwendeten Vektor– und Matrixnormen auf. Anhang B faßt grundlegende Definitionen und Sätze der Wahrscheinlichkeitstheorie in Hilberträumen zusammen, die zwar auf einfache Weise aus der allgemeinen Theorie in Banachräumen folgen, für die sich aber leider keine zusammenhängende Darstellung in einem Lehrbuch fand. Anhang C stellt die benötigten Hilfsmittel des allgemeinen Differentialkalküls bereit, und Anhang D diejenigen aus der Theorie der  $U$ –Statistiken. Der Schwerpunkt liegt dabei auf  $U$ –Statistiken mit variablem Kern, die weitaus weniger bekannt und untersucht sind als die klassischen  $U$ –Statistiken mit festem Kern. Anhang E schließlich gibt die in der vorliegenden Arbeit benötigten Tatsachen über benachbarte Alternativen aus Witting und Müller–Funk (1995) wieder. Ein Symbolverzeichnis befindet sich auf Seite 189.

An dieser Stelle soll noch auf einige grundlegende Vereinbarungen und Notationen hingewiesen werden. *i.i.d.* steht für „unabhängig und identisch verteilt“, *o.B.d.A.* für „ohne Beschränkung der Allgemeinheit“, „ $\mathbf{1}$ “ für die Indikatorfunktion einer Menge und „const“ für eine generische Konstante, die stets endliche, aber u.U. unterschiedliche Werte annimmt. Alle betrachteten Zufallsvariablen sind auf einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  definiert. Der  $\mathbb{R}^d$  sei stets mit der Borel– $\sigma$ –Algebra  $\mathfrak{B}^d$  versehen;  $\lambda^d$  bezeichne das  $d$ –dimensionale Lebesgue–Maß. Wenn wir ohne nähere Angaben von einer „Dichte“ sprechen, meinen wir damit eine Lebesgue–Dichte.  $P^X$  gibt die Verteilung einer Zufallsvariablen  $X$  an; wie üblich schreiben wir  $\xrightarrow{D}$ ,  $\xrightarrow{P}$  und  $\xrightarrow{\text{f.s.}}$  für Verteilungs–, stochastische und fast sichere Konvergenz.  $O_P(1)$  bezeichnet einen straffen Term,  $o_P(1)$  einen Term, der stochastisch gegen 0 strebt.  $O_{\text{f.s.}}(1)$  und  $o_{\text{f.s.}}(1)$  werden im fast sicheren Sinn,  $O(1)$  und  $o(1)$  in einem deterministischen Kontext verwendet. Deterministische oder zufällige Vektoren sind, wenn nicht ausdrücklich anders definiert, Elemente des  $\mathbb{R}^d$ , der mit der Euklidnorm  $\|\cdot\|$  versehen wird.  $I_d$  bezeichne die  $d$ –dimensionale Einheitsmatrix; der Raum der  $d \times d$ –Matrizen  $\mathbb{R}^{d \times d}$  sei mit der Schur–Norm  $\|\cdot\|_S$  bzw. mit der Spektralnorm  $\|\cdot\|$  versehen. Zusammenhänge zwischen diesen Normen sind in Anhang A auf Seite 149 zu finden. Transponierte eines Vektors oder einer Matrix werden durch  $(\cdot)^T$  gekennzeichnet. Integrale über Vektoren oder Matrizen sind komponentenweise zu lesen; Integrale ohne spezifizierten Integrationsbereich laufen stets über den gesamten  $\mathbb{R}^d$ .

# 1. Der BHEP-Test

Es sei  $X$  ein  $d$ -dimensionaler Zufallsvektor ( $d \in \mathbb{N}$ ) mit Verteilung  $P^X$ . Bezeichnet  $\mathcal{N}_d$  die Klasse der nichtdegenerierten  $d$ -dimensionalen Normalverteilungen, so besteht ein gängiges Problem der Statistik darin, anhand von  $n$  unabhängigen und identisch verteilten Realisierungen  $X_1, \dots, X_n$  von  $X$  die Hypothese

$$H_0 : P^X \in \mathcal{N}_d ,$$

daß  $X$  eine  $d$ -dimensionale Normalverteilung besitzt, gegen die allgemeine Alternative

$$H_1 : P^X \notin \mathcal{N}_d$$

zu testen. Wünschenswerte Eigenschaften solcher Anpassungstests sind die Konsistenz gegenüber einer möglichst großen Klasse von alternativen Verteilungen („Omnibus Tests“), eine möglichst hohe Güte auch bei kleineren und mittleren Stichprobenumfängen  $n$  sowie eine nichtdegenerierte Güte gegenüber benachbarten Alternativen, die mit der Geschwindigkeit  $n^{-1/2}$  gegen die Hypothese konvergieren. Ferner sollte die Teststatistik auch in höheren Dimensionen  $d$  schnell und numerisch stabil zu berechnen sein. Da die Klasse  $\mathcal{N}_d$  der nichtdegenerierten  $d$ -dimensionalen Normalverteilungen abgeschlossen ist gegenüber regulären affinen Transformationen und da im allgemeinen die Lage und Skalierung konkret vorliegender Daten unbekannt sind, sollte die betrachtete Teststatistik  $T_n$  ebenfalls affin invariant sein, d.h. für nichtsinguläre Matrizen  $A \in \mathbb{R}^{d \times d}$  und Vektoren  $b \in \mathbb{R}^d$  sollte gelten:

$$T_n(AX_1 + b, \dots, AX_n + b) = T_n(X_1, \dots, X_n).$$

Ein Verfahren, das den obigen Anforderungen erstaunlich gut gerecht wird, ist der sogenannte *BHEP-Test*, der wie folgt konstruiert ist: Die charakteristische Funktion

$$\Psi(t) := \mathbb{E} \left[ e^{it^T X} \right] \quad (t \in \mathbb{R}^d, \quad i := \sqrt{-1})$$

von  $X$  nimmt unter der standardisierten Hypothese (d.h. bei Vorliegen von  $X \sim \mathcal{N}_d(0, I_d)$ ) den Wert  $\Psi(t) = \exp\{-\|t\|^2/2\}$  an. Da die Verteilung von  $X$  durch  $\Psi$  charakterisiert wird und da die Lage und Skalierung von  $X$  nicht näher spezifiziert sind, liegt es nahe, die Funktion  $\Psi$  unter  $H_0$  mit der empirischen charakteristischen Funktion

$$\Psi_n(t) := \frac{1}{n} \sum_{j=1}^n e^{it^T Y_j} \quad , \quad t \in \mathbb{R}^d,$$

der standardisierten Daten

$$Y_j := S_n^{-1/2}(X_j - \bar{X}_n) \quad , \quad j = 1, \dots, n,$$

zu vergleichen. Dabei bezeichnen  $\bar{X}_n := \frac{1}{n} \sum_{j=1}^n X_j$  das empirische Mittel und

$$S_n := \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T$$

die empirische Kovarianzmatrix von  $X_1, \dots, X_n$  sowie  $S_n^{-1/2}$  die symmetrische positiv definite Wurzel von  $S_n^{-1}$ . Als Abstandsmaß für den Vergleich der beiden charakteristischen Funktionen bietet sich die kanonische Norm in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t) dt)$  an, wobei

$$\varphi_\beta(t) := \frac{1}{(2\pi\beta^2)^{d/2}} e^{-\frac{\|t\|^2}{2\beta^2}}$$

(d.h. die Dichte von  $\mathcal{N}_d(0, \beta^2 I_d)$ ) eine Gewichtsfunktion ist, die durch den Parameter  $\beta > 0$  zusätzliche Flexibilität erhält. Damit gelangt man zu der Statistik

$$T_{n,\beta} := n \int_{\mathbb{R}^d} \left| \Psi_n(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt, \quad (1.1)$$

die allerdings nur im Falle einer nichtsingulären empirischen Kovarianzmatrix  $S_n$  wohldefiniert ist. Dies ist nach Eaton und Perlman (1973) fast sicher gegeben, sofern alle  $(d-1)$ -dimensionalen Hyperebenen des  $\mathbb{R}^d$  unter der Verteilung von  $X$  das Maß 0 besitzen und der Stichprobenumfang  $n$  echt größer als die Dimension  $d$  ist. Csörgő (1989) schlug deshalb vor, im Falle der Singularität von  $S_n$  den Wert der Teststatistik durch den größtmöglichen, also durch  $4n$ , zu ersetzen, was schließlich zur Definition der BHEP-Statistik führt:

$$D_{n,\beta} := 4n \cdot \mathbf{1}\{S_n \text{ singular}\} + T_{n,\beta} \cdot \mathbf{1}\{S_n \text{ nicht singular}\}.$$

Der auf  $D_{n,\beta}$  basierende BHEP-Test besitzt einen oberen Ablehnbereich; insbesondere bedeutet im Falle der Singularität von  $S_n$  der Wert  $D_{n,\beta} = 4n$  die Ablehnung von  $H_0$ . Da dieser Fall unter den Voraussetzungen von Eaton und Perlman (1973) nur mit Wahrscheinlichkeit 0 eintritt, beschränken wir uns bei den nachfolgenden Untersuchungen auf die Statistik  $T_{n,\beta}$ , mit der stillschweigenden Vereinbarung, daß diese Voraussetzungen stets erfüllt sind.

Berechnet man das Integral in (1.1) für jedes zugrundeliegende  $\omega \in \Omega$ , erhält man die alternative Darstellung

$$T_{n,\beta} = \frac{1}{n} \sum_{j,k=1}^n e^{-\frac{\beta^2}{2} \|Y_j - Y_k\|^2} - \frac{2}{(1 + \beta^2)^{d/2}} \sum_{j=1}^n e^{-\frac{\beta^2 \|Y_j\|^2}{2(1+\beta^2)}} + \frac{n}{(1 + 2\beta^2)^{d/2}}, \quad (1.2)$$

woran man erkennt, daß die standardisierten Daten nur durch

$$\|Y_j\|^2 = (X_j - \bar{X}_n)^T S_n^{-1} (X_j - \bar{X}_n)$$

bzw. durch

$$\|Y_j - Y_k\|^2 = (X_j - X_k)^T S_n^{-1} (X_j - X_k)$$

in die Berechnung von  $T_{n,\beta}$  eingehen. Dies zeigt zum einen, daß man zur Berechnung nur die Inverse  $S_n^{-1}$ , nicht aber deren Quadratwurzel benötigt, und zum anderen die affine Invarianz der BHEP–Statistik. Diese wiederum erlaubt es, die Daten o.B.d.A. als standardisiert anzunehmen, im  $H_0$ –Fall also die Vereinbarung  $X \sim \mathcal{N}_d(0, I_d)$  zu treffen.

Der BHEP–Test wurde in Epps und Pulley (1983) für den Spezialfall  $d = 1$  und  $\beta = 1$  vorgeschlagen. Baringhaus und Henze (1988) verallgemeinerten diesen Ansatz auf die Dimension  $d > 1$ , Csörgő (1989) wies die universelle Konsistenz des resultierenden Tests nach und schlug die Bezeichnung „BHEP–Test“ im Hinblick auf die vier Autoren vor, die den Test entwickelt hatten. Henze und Zirkler (1990) führten schließlich mit dem Gewichtsparameter  $\beta$  zusätzliche Flexibilität ein.

Die Einführung der flexiblen Gewichtsfunktion  $\varphi_\beta$  durch Henze und Zirkler (1990) hat den Vorteil, daß für kleine Werte von  $\beta$  die Masse von  $\varphi_\beta$  in der Nähe des Ursprungs des  $\mathbb{R}^d$  konzentriert ist und daß damit Abweichungen zwischen  $\Psi_n$  und  $\Psi$  in der Nähe des Ursprungs verstärkt bewertet werden. Nach den Sätzen von Abel und Tauber (vgl. z.B. Feller (1966), Chapter VIII.5) ist zu erwarten, daß Abweichungen in den Tails der beiden Verteilungen gut erkannt werden, d.h. daß der BHEP–Test mit kleinem  $\beta$  eine hohe Güte gegenüber Alternativen mit starken Tails aufweist.

Die asymptotische  $H_0$ –Verteilung der Statistik  $T_{n,\beta}$  wurde in Baringhaus und Henze (1988) bzw. in Henze und Zirkler (1990) anhand der Darstellung (1.2) von  $T_{n,\beta}$  als  $V$ –Statistik mit geschätzten Parametern und unter Anwendung der Ergebnisse von De Wet und Randles (1987) hergeleitet. Henze und Wagner (1997) führten den Konvergenznachweis im  $H_0$ –Fall anhand der weiteren Darstellung

$$T_{n,\beta} = \int_{\mathbb{R}^d} Z_n^2(t) \varphi_\beta(t) dt$$

mit

$$Z_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^n \left( \cos(t^T Y_j) + \sin(t^T Y_j) - e^{-\frac{\|t\|^2}{2}} \right),$$

indem sie  $Z_n(\cdot)$  als Zufallselement in dem Fréchet–Raum  $\mathcal{C}(\mathbb{R}^d)$  der stetigen Funktionen auf  $\mathbb{R}^d$ , versehen mit der Metrik

$$\rho(x, y) := \sum_{k=1}^{\infty} 2^{-k} \cdot \frac{\rho_k(x, y)}{1 + \rho_k(x, y)}, \quad \rho_k(x, y) := \max_{\|t\| \leq k} |x(t) - y(t)|$$

betrachteten und Corollar 7.17 aus Araujo und Giné (1980) anwandten.

Henze und Wagner (1997) bestimmten außerdem die Grenzverteilung von  $T_{n,\beta}$  unter einer Folge von benachbarten Alternativen, die mit Geschwindigkeit  $n^{-1/2}$  gegen die Klasse der Normalverteilungen konvergiert, und zeigten, daß der BHEP–Test in diesem Fall eine nichtdegenerierte Limeschärfe aufweist.

Baringhaus und Henze (1988) zeigten im Rahmen des Konsistenznachweises für den BHEP–

Test, daß unter jeder festen (o.B.d.A. standardisierten) Alternative mit existierender Varianz gilt:

$$\frac{1}{n}T_{n,\beta} \xrightarrow{\text{f.s.}} L_\beta := \int_{\mathbb{R}^d} \left| \Psi(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt$$

(vgl. auch Csörgő (1989)). Da der Grenzwert  $L_\beta$  genau im Falle einer zugrundeliegenden Standard-Normalverteilung gleich 0 ist, folgt insbesondere die Konsistenz des Tests gegen Alternativen mit existierender Varianz.

Aufgrund der bisher dargelegten zahlreichen wünschenswerten Eigenschaften der Klasse der BHEP-Tests, die dieses Verfahren aus der umfangreichen Menge von Anpassungstests auf Normalverteilung hervorheben, erscheint es gerechtfertigt, die Statistik  $T_{n,\beta}$  noch genauer zu untersuchen, mit dem Ziel, weitere Aussagen über den zugehörigen Test zu erhalten. Gegenstand von Kapitel 1.1 ist deshalb der Nachweis von

$$S_{n,\beta} := \sqrt{n} \left( \frac{1}{n}T_{n,\beta} - L_\beta \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_\beta^2) \quad (1.3)$$

unter einer festen Alternative mit existierenden vierten Momenten, wobei die Grenzvarianz  $\tau_\beta^2$  in Theorem 5 explizit angegeben wird.

Um dieses theoretische Ergebnis in der Praxis verwenden zu können, muß die von der unbekannt zugrundeliegenden Verteilung abhängende Grenzvarianz  $\tau_\beta^2$  anhand der Daten geschätzt werden. Die Herleitung eines solchen Schätzers ist Gegenstand von Kapitel 1.2. Einige technische Lemmata und Beweise sind ausgelagert in Abschnitt 1.5.

Aus der Konvergenzaussage (1.3) lassen sich mehrere statistische Anwendungen herleiten. Zum einen kann man einen asymptotischen Konfidenzbereich für den von der unbekannt zugrundeliegenden Verteilung abhängenden Wert  $L_\beta$  angeben; die geschieht in Abschnitt 1.3.1. Die wichtigere Folgerung ist jedoch die Angabe einer Näherung für die Güte des BHEP-Tests, die in Abschnitt 1.3.2 ausgearbeitet wird. Ferner wird in Abschnitt 1.3.3 ein „inverser Test auf Normalverteilung“ im Sinne des Testens auf Bioäquivalenz vorgeschlagen. Die Brauchbarkeit der Verfahren, insbesondere der Gütenäherung, in der Praxis wird an sechs Beispielveilteilungen in Kapitel 1.4 untersucht.

## 1.1. Konvergenzaussagen

Im folgenden seien  $X_1, \dots, X_n$  unabhängige und identisch verteilte Realisierungen einer  $d$ -dimensionalen Verteilung, von der wir annehmen, daß sie jede  $(d-1)$ -dimensionale Hyperebene des  $\mathbb{R}^d$  mit Masse 0 belegt und daß die vierten Momente existieren, d.h. daß  $E \|X_1\|^4 < \infty$  gilt. Aufgrund der affinen Invarianz der BHEP-Statistik  $T_{n,\beta}$  können wir  $X_1$  o.B.d.A. als standardisiert voraussetzen, d.h. es gelte  $EX_1 = 0$  und  $\text{Cov}(X_1) = EX_1X_1^T = I_d$ . Ferner gelte für den Stichprobenumfang  $n$  stets  $n > d$ , damit wir die empirische Kovarianzmatrix  $S_n$  als fast sicher nichtsingulär voraussetzen können.

Setzen wir

$$\Phi_1(t) := E [\cos(t^T X_1)] \quad , \quad \Phi_2(t) := E [\sin(t^T X_1)]$$

für den Real- bzw. Imaginärteil der charakteristischen Funktion von  $X_1$  und

$$\Phi_{1,n}(t) := \frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) \quad , \quad \Phi_{2,n}(t) := \frac{1}{n} \sum_{j=1}^n \sin(t^T Y_j) \quad (1.4)$$

für die entsprechenden auf den standardisierten Daten beruhenden empirischen Größen, so ist die *Sinus/Cosinus-Transformierte*

$$\Phi(t) := \Phi_1(t) + \Phi_2(t) = \mathbb{E} [\cos(t^T X_1) + \sin(t^T X_1)]$$

das reelle Analogon der charakteristischen Funktion  $\Psi = \Phi_1 + i\Phi_2$  von  $X_1$  und die *empirische Sinus/Cosinus-Transformierte*

$$\Phi_n(t) := \Phi_{1,n}(t) + \Phi_{2,n}(t) = \frac{1}{n} \sum_{j=1}^n \{\cos(t^T Y_j) + \sin(t^T Y_j)\}$$

dasjenige der auf den standardisierten Daten beruhenden empirischen charakteristischen Funktion  $\Psi_n = \Phi_{1,n} + i\Phi_{2,n}$ .

Beachtet man, daß aus Symmetriegründen  $\int_{\mathbb{R}^d} \sin(t^T Y_k) \left( \cos(t^T Y_j) - \exp\{-\frac{\|t\|^2}{2}\} \right) \varphi_\beta(t) dt = 0$  bzw.  $\int_{\mathbb{R}^d} \mathbb{E} [\sin(t^T X_1)] \left( \mathbb{E} [\cos(t^T X_1)] - \exp\{-\frac{\|t\|^2}{2}\} \right) \varphi_\beta(t) dt = 0$  gilt, so erhält man mit Hilfe der Größen  $\Phi_n$  bzw.  $\Phi$  durch direkte Rechnung eine weitere, rein reelle Darstellung der BHEP-Statistik

$$\frac{1}{n} T_{n,\beta} = \int_{\mathbb{R}^d} \left| \Psi_n(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt = \int_{\mathbb{R}^d} \left( \Phi_n(t) - e^{-\frac{\|t\|^2}{2}} \right)^2 \varphi_\beta(t) dt$$

bzw. des fast sicheren Grenzwertes von  $n^{-1} T_{n,\beta}$  unter der zugrundeliegenden festen Alternative:

$$L_\beta = \int_{\mathbb{R}^d} \left| \Psi(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt = \int_{\mathbb{R}^d} \left( \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right)^2 \varphi_\beta(t) dt .$$

Dies liefert die folgende Zerlegung der uns interessierenden Größe  $S_{n,\beta} = \sqrt{n} \left( \frac{1}{n} T_{n,\beta} - L_\beta \right)$ :

$$\begin{aligned} S_{n,\beta} &= \sqrt{n} \int_{\mathbb{R}^d} \left\{ \left( \Phi_n(t) - e^{-\frac{\|t\|^2}{2}} \right)^2 - \left( \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right)^2 \right\} \varphi_\beta(t) dt \\ &= \sqrt{n} \int_{\mathbb{R}^d} \left\{ (\Phi_n(t) - \Phi(t)) (\Phi_n(t) + \Phi(t)) - 2e^{-\frac{\|t\|^2}{2}} (\Phi_n(t) - \Phi(t)) \right\} \varphi_\beta(t) dt \\ &= \sqrt{n} \int_{\mathbb{R}^d} (\Phi_n(t) - \Phi(t)) \left\{ \Phi_n(t) - \Phi(t) + 2 \left( \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right) \right\} \varphi_\beta(t) dt \\ &= 2 \int_{\mathbb{R}^d} Z_n(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + \frac{1}{\sqrt{n}} \int_{\mathbb{R}^d} Z_n(t)^2 \varphi_\beta(t) dt \end{aligned} \quad (1.5)$$

mit

$$Z_n(t) = \sqrt{n} [\Phi_n(t) - \Phi(t)] = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{\cos(t^T Y_j) + \sin(t^T Y_j) - \Phi(t)\} . \quad (1.6)$$

Um die Konvergenz von  $S_{n,\beta}$  gegen eine zentrierte Normalverteilung zu zeigen, betrachten wir  $Z_n(\cdot)$  als Zufallselement in dem Hilbert-Raum  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  und weisen seine Konvergenz in diesem Raum gegen ein zentriertes Gaußsches Zufallselement  $Z$  mit einer bestimmten Kovarianzstruktur nach. Dabei ist der  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  mit dem kanonischen Skalarprodukt  $\langle \cdot, \cdot \rangle_{L^2}$  und der kanonischen Norm  $\|\cdot\|_{L^2}$  versehen, d.h. für  $g, h \in L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  gelten

$$\langle g, h \rangle_{L^2} = \int_{\mathbb{R}^d} g(t)h(t) \varphi_\beta(t) dt \quad , \quad \|h\|_{L^2}^2 = \int_{\mathbb{R}^d} h(t)^2 \varphi_\beta(t) dt .$$

Das erste Integral in (1.5) ist das Skalarprodukt in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  von  $Z_n(\cdot)$  und der ebenfalls in diesem Raum liegenden Funktion  $h(\cdot) := \Phi(\cdot) - \exp\{-\frac{\|\cdot\|^2}{2}\}$ , das zweite Integral ist gerade das Quadrat der  $L^2$ -Norm von  $Z_n(\cdot)$ . Da sowohl das Skalarprodukt als auch die Norm stetige Funktionen von  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  nach  $\mathbb{R}$  sind, wird sich nach dem Abbildungssatz die Konvergenz von  $Z_n(\cdot)$  auf die beiden Terme in (1.5) übertragen, was das gewünschte Resultat für  $S_{n,\beta}$  liefert. Man beachte, daß der zweite Term in (1.5) aufgrund des Faktors  $n^{-1/2}$  stochastisch gegen 0 konvergiert.

Ein Gaußsches Zufallselement  $Z$  in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  ist dadurch definiert, daß für jedes  $h$  aus diesem Raum das Skalarprodukt  $\langle Z, h \rangle_{L^2}$  eine reelle Gaußsche Zufallsvariable ist (vgl. Definition B.12 in Anhang B). Wie im Reellen ist ein Gaußsches Zufallselement in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  charakterisiert durch seinen Erwartungswert  $EZ \in L^2$  und seinen Kovarianzoperator  $S_Z : L^2 \rightarrow L^2$ , die durch die Gleichungen

$$\langle h, EZ \rangle_{L^2} = E \langle h, Z \rangle_{L^2} \quad , \quad \langle S_Z h, h \rangle_{L^2} = E [\langle h, Z - EZ \rangle_{L^2}^2] ,$$

$h \in L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$ , eindeutig bestimmt sind (vgl. Definition B.4 und Theorem B.13). Nach den Ausführungen in Abschnitt B.4 auf Seite 157 können wir  $Z$  als einen meßbaren Gaußprozeß mit Pfaden in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  betrachten, der im zentrierten Fall charakterisiert ist durch

$$\langle S_Z h, h \rangle_{L^2} = E [\langle h, Z \rangle_{L^2}^2] = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} E [Z(s)Z(t)] h(s)h(t) \varphi_\beta(s) ds \varphi_\beta(t) dt ,$$

d.h. durch den Kovarianzkern

$$k(s, t) := E [Z(s)Z(t)] .$$

Um im vorliegenden Fall die Kovarianzstruktur des Grenzwertes  $Z(\cdot)$  von  $Z_n(\cdot)$  mit vertretbarem Platzaufwand niederschreiben zu können, sind die folgenden Notationen hilfreich. Zunächst setzen wir

$$\tilde{\Phi}(t) := \Phi_1(t) - \Phi_2(t) = E [\cos(t^T X_1) - \sin(t^T X_1)] .$$

Die Ableitungen

$$\Phi_1'(t) := -E [\sin(t^T X_1)X_1] \quad , \quad \Phi_2'(t) := E [\cos(t^T X_1)X_1]$$

von  $\Phi_1$  und  $\Phi_2$  nach  $t$  sind nach dem Satz über die Differentiation unter dem Integral wohldefiniert. Entsprechendes gilt für die Ableitungen von  $\Phi$  und  $\tilde{\Phi}$  nach  $t$ :

$$\begin{aligned} \Phi'(t) &= \Phi_1'(t) + \Phi_2'(t) = E [(\cos(t^T X_1) - \sin(t^T X_1)) X_1] , \\ \tilde{\Phi}'(t) &= \Phi_1'(t) - \Phi_2'(t) = -E [(\cos(t^T X_1) + \sin(t^T X_1)) X_1] . \end{aligned}$$

Ebenso sind die zweiten Ableitungen von  $\Phi_1$  und  $\Phi_2$  nach  $t$  gegeben durch

$$\Phi_1''(t) := -\mathbb{E} [\cos(t^T X_1) X_1 X_1^T] \quad , \quad \Phi_2''(t) := -\mathbb{E} [\sin(t^T X_1) X_1 X_1^T] \quad .$$

Diese sind wiederum nach dem Satz über die Differentiation unter dem Integral wohldefiniert, da die Momente zweiter Ordnung von  $X_1$  nach Voraussetzung existieren. Damit gilt auch

$$\begin{aligned} \Phi''(t) &:= \Phi_1''(t) + \Phi_2''(t) = -\mathbb{E} [(\cos(t^T X_1) + \sin(t^T X_1)) X_1 X_1^T] \quad , \\ \tilde{\Phi}''(t) &:= \Phi_1''(t) - \Phi_2''(t) = \mathbb{E} [(\sin(t^T X_1) - \cos(t^T X_1)) X_1 X_1^T] \quad . \end{aligned}$$

Ferner benötigen wir die speziellen multivariaten Momente dritter Ordnung

$$M'_{3,1}(t) := \mathbb{E} \left[ \left( t^T X_1 X_1^T \Phi_1'(t) \right) X_1 \right] \quad , \quad M'_{3,2}(t) := \mathbb{E} \left[ \left( t^T X_1 X_1^T \Phi_2'(t) \right) X_1 \right] \quad (1.7)$$

sowie vierter Ordnung:

$$M''_{4,1}(t) := \mathbb{E} \left[ \left( t^T X_1 X_1^T \Phi_1'(t) \right) X_1 X_1^T \right] \quad , \quad M''_{4,2}(t) := \mathbb{E} \left[ \left( t^T X_1 X_1^T \Phi_2'(t) \right) X_1 X_1^T \right] \quad . \quad (1.8)$$

Hierbei beachte man, daß  $\Phi_1'(t)$ ,  $\Phi_2'(t)$ ,  $\Phi'(t)$ ,  $\tilde{\Phi}'(t)$ ,  $M'_{3,1}(t)$  sowie  $M'_{3,2}(t)$   $d$ -dimensionale Spaltenvektoren sind, und  $\Phi_1''(t)$ ,  $\Phi_2''(t)$ ,  $\Phi''(t)$ ,  $\tilde{\Phi}''(t)$ ,  $M''_{4,1}(t)$  sowie  $M''_{4,2}(t)$   $d \times d$ -Matrizen. Auch bei später noch zu definierenden Größen werden wir das Prinzip,  $d$ -dimensionale Spaltenvektoren mit dem Zusatz  $'$  und  $d \times d$ -Matrizen mit dem Zusatz  $''$  zu versehen, beibehalten, in der Hoffnung, daß die Vielzahl an Bezeichnungen dadurch ein wenig übersichtlicher wird. Die auf diese Art definierten Vektoren und Matrizen haben allerdings nichts mehr gemeinsam mit den ersten und zweiten Ableitungen, aus denen die Notation ursprünglich hervorging.

**Theorem 1** Es sei  $X_1, X_2, \dots$  eine Folge von unabhängigen und identisch verteilten  $d$ -dimensionalen standardisierten Zufallsvektoren mit nichtdegenerierter Verteilung und endlichen vierten Momenten.

Dann existiert ein zentrierter Gaußprozeß  $Z$  in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  mit durch

$$\begin{aligned} k(s, t) &:= \Phi_1(t-s) + \Phi_2(t+s) - \Phi(t)\Phi(s) + \tilde{\Phi}(t)t^T\tilde{\Phi}'(s) + \tilde{\Phi}(s)s^T\tilde{\Phi}'(t) + \tilde{\Phi}(s)\tilde{\Phi}(t)s^Tt \\ &\quad + \frac{1}{2} \left\{ t^T \Phi''(s) \Phi'(t) + s^T \Phi''(t) \Phi'(s) \right\} + \frac{1}{2} \left\{ \Phi(s)t^T \Phi'(t) + \Phi(t)s^T \Phi'(s) \right\} \\ &\quad + \frac{1}{2} \tilde{\Phi}(s)s^T \left( M'_{3,1}(t) + M'_{3,2}(t) \right) + \frac{1}{2} \tilde{\Phi}(t)t^T \left( M'_{3,1}(s) + M'_{3,2}(s) \right) \\ &\quad + \frac{1}{4} s^T \left( M''_{4,1}(t) + M''_{4,2}(t) \right) \Phi'(s) - \frac{1}{4} s^T \Phi'(s) t^T \Phi'(t) \end{aligned}$$

( $s, t \in \mathbb{R}^d$ ) definierter Kovarianzstruktur, so daß für das in (1.6) erklärte Zufallselement  $Z_n(\cdot)$  gilt:

$$Z_n \xrightarrow{\mathcal{D}} Z \quad \text{in } L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt) \quad .$$

BEWEIS: Der Zentrale Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallselementen in Hilberträumen ist sehr einfach (vgl. Theorem B.16 in Anhang B sowie den Beweis von Lemma 2 weiter unten). Das Problem bei  $Z_n$  liegt jedoch darin, daß aufgrund der Schätzer in den standardisierten Daten  $Y_j$  ( $j = 1, \dots, n$ ) noch keine Summe von unabhängigen Zufallselementen vorliegt. Folglich müssen die Schätzer „herausgerechnet“ werden, was in erster Näherung durch eine Taylor-Entwicklung ersten Grades der Sinus/Cosinus-Terme geschieht, in denen die  $Y_j$  auftreten. Dies führt zu dem ersten Hilfsprozeß

$$\tilde{Z}_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ \cos(t^T X_j) + \sin(t^T X_j) - \Phi(t) + [\cos(t^T X_j) - \sin(t^T X_j)] t^T \Delta_j \right\}, \quad (1.9)$$

wobei

$$\Delta_j := Y_j - X_j = (S_n^{-1/2} - I_d) X_j - S_n^{-1/2} \bar{X}_n \quad (1.10)$$

die Differenz zwischen den standardisierten und den ursprünglichen Daten bezeichnet. Extrahiert man aus dem Restterm  $\Delta_j$  die relevanten Anteile, gelangt man schließlich zu dem zweiten Hilfsprozeß

$$Z_n^*(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ \cos(t^T X_j) + \sin(t^T X_j) - \Phi(t) - t^T X_j \tilde{\Phi}(t) - \frac{1}{2} t^T (X_j X_j^T - I_d) \Phi'(t) \right\}, \quad (1.11)$$

der eine gewichtete Summe von unabhängigen und identisch verteilten Zufallselementen in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t) dt)$  ist, auf die der Zentrale Grenzwertsatz angewandt werden kann.

Das weitere Vorgehen läßt sich in drei Schritte unterteilen: Zeigt man

- (i)  $Z_n^* \xrightarrow{\mathcal{D}} Z$  in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t) dt)$ ,
- (ii)  $\left\| Z_n - \tilde{Z}_n \right\|_{L^2} \xrightarrow{P} 0$  und
- (iii)  $\left\| \tilde{Z}_n - Z_n^* \right\|_{L^2} \xrightarrow{P} 0$

beim Grenzübergang  $n \rightarrow \infty$ , so folgt die Behauptung von Theorem 1 direkt aus Theorem 4.1 in Billingsley (1968). Der Nachweis der Behauptungen (i), (ii) und (iii) ist Gegenstand der drei folgenden Lemmata.

**Lemma 2** Für das in (1.11) definierte Zufallselement  $Z_n^*$  gilt die Konvergenz  $Z_n^* \xrightarrow{\mathcal{D}} Z$  in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t) dt)$ .

BEWEIS: Es gilt  $Z_n^*(t) = n^{-1/2} \sum_{j=1}^n Z_j^{**}(t)$  mit

$$Z_j^{**}(t) := \cos(t^T X_j) + \sin(t^T X_j) - \Phi(t) - t^T X_j \tilde{\Phi}(t) - \frac{1}{2} t^T (X_j X_j^T - I_d) \Phi'(t), \quad (1.12)$$

$j = 1, \dots, n$ . Aufgrund der stetigen Pfade  $Z_j^{**}(\omega, \cdot)$  sind  $Z_1^{**}, \dots, Z_n^{**}$  meßbare stochastische Prozesse (vgl. Wentzell (1981), S. 22) und damit unabhängige und identisch verteilte Zufallselemente in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t) dt)$ . Wegen  $E[Z_1^{**}(t)] = 0$  für alle  $t \in \mathbb{R}^d$  ist das Zufallselement  $Z_1^{**}$  zentriert in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t) dt)$ . Ferner gilt

$$E \|Z_1^{**}\|_{L^2}^2 = \int_{\mathbb{R}^d} E [Z_1^{**}(t)^2] \varphi_\beta(t) dt < \infty$$

aufgrund der Voraussetzung  $E \|X_1\|^4 < \infty$ . Nach Definition B.14, Bemerkung B.15 und Theorem B.16 konvergiert

$$Z_n^* = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j^{**} \xrightarrow{\mathcal{D}} Z \quad \text{in } L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$$

gegen einen zentrierten Gaußprozeß  $Z$ , der dieselbe Kovarianzstruktur wie  $Z_1^{**}(\cdot)$  besitzt. Um den Beweis von Lemma 2 zu vollenden, reicht es also, den Kovarianzkern  $k(s, t) = E [Z_1^{**}(s)Z_1^{**}(t)]$  zu identifizieren. Wegen  $EX_1 = 0$  und  $EX_1X_1^T = I_d$  gilt

$$\begin{aligned} & E [Z_1^{**}(s)Z_1^{**}(t)] \\ &= E \left[ \left\{ (\cos(s^T X_1) + \sin(s^T X_1)) - \Phi(s) - s^T X_1 \tilde{\Phi}(s) - \frac{1}{2}s^T (X_1X_1^T - I_d) \Phi'(s) \right\} \right. \\ & \quad \left. \left\{ (\cos(t^T X_1) + \sin(t^T X_1)) - \Phi(t) - t^T X_1 \tilde{\Phi}(t) - \frac{1}{2}t^T (X_1X_1^T - I_d) \Phi'(t) \right\} \right] \\ &= E \left[ (\cos(s^T X_1) + \sin(s^T X_1)) (\cos(t^T X_1) + \sin(t^T X_1)) \right] \\ & \quad - \Phi(t)E [\cos(s^T X_1) + \sin(s^T X_1)] - \tilde{\Phi}(t)t^T E [(\cos(s^T X_1) + \sin(s^T X_1)) X_1] \\ & \quad - \frac{1}{2}t^T \{E [(\cos(s^T X_1) + \sin(s^T X_1)) X_1X_1^T] - E [\cos(s^T X_1) + \sin(s^T X_1)] I_d\} \Phi'(t) \\ & \quad - \Phi(s)E [\cos(t^T X_1) + \sin(t^T X_1)] + \Phi(s)\Phi(t) \\ & \quad - \tilde{\Phi}(s)s^T E [(\cos(t^T X_1) + \sin(t^T X_1)) X_1] + \tilde{\Phi}(s)\tilde{\Phi}(t)E [s^T X_1 t^T X_1] \\ & \quad + \frac{1}{2}\tilde{\Phi}(s)E [s^T X_1 t^T (X_1X_1^T - I_d) \Phi'(t)] + \frac{1}{2}\tilde{\Phi}(t)E [t^T X_1 s^T (X_1X_1^T - I_d) \Phi'(s)] \\ & \quad - \frac{1}{2}s^T \{E [(\cos(t^T X_1) + \sin(t^T X_1)) X_1X_1^T] - E [\cos(t^T X_1) + \sin(t^T X_1)] I_d\} \Phi'(s) \\ & \quad + \frac{1}{4}E [s^T (X_1X_1^T - I_d) \Phi'(s) t^T (X_1X_1^T - I_d) \Phi'(t)] \\ &= E \left[ \cos \left( (t-s)^T X_1 \right) \right] + E \left[ \sin \left( (t+s)^T X_1 \right) \right] - \Phi(s)\Phi(t) + \tilde{\Phi}(t)t^T \tilde{\Phi}'(s) \\ & \quad + \frac{1}{2}t^T \left\{ \Phi''(s) + \Phi(s)I_d \right\} \Phi'(t) + \tilde{\Phi}(s)s^T \tilde{\Phi}'(t) + \tilde{\Phi}(s)\tilde{\Phi}(t)s^T t \\ & \quad + \frac{1}{2}\tilde{\Phi}(s)E [s^T X_1 t^T X_1 X_1^T \Phi'(t)] + \frac{1}{2}s^T \left\{ \Phi''(t) + \Phi(t)I_d \right\} \Phi'(s) \\ & \quad + \frac{1}{2}\tilde{\Phi}(t)E [t^T X_1 s^T X_1 X_1^T \Phi'(s)] \\ & \quad + \frac{1}{4} \left\{ E [s^T X_1 X_1^T \Phi'(s) t^T X_1 X_1^T \Phi'(t)] - s^T \Phi'(s) t^T \Phi'(t) \right\} \\ &= \Phi_1(t-s) + \Phi_2(t+s) - \Phi(s)\Phi(t) + \tilde{\Phi}(t)t^T \tilde{\Phi}'(s) + \tilde{\Phi}(s)s^T \tilde{\Phi}'(t) + \tilde{\Phi}(s)\tilde{\Phi}(t)s^T t \\ & \quad + \frac{1}{2} \left\{ t^T \Phi''(s) \Phi'(t) + s^T \Phi''(t) \Phi'(s) \right\} + \frac{1}{2} \left\{ \Phi(s)t^T \Phi'(t) + \Phi(t)s^T \Phi'(s) \right\} \\ & \quad + \frac{1}{2}\tilde{\Phi}(s)s^T \left( M'_{3,1}(t) + M'_{3,2}(t) \right) + \frac{1}{2}\tilde{\Phi}(t)t^T \left( M'_{3,1}(s) + M'_{3,2}(s) \right) \\ & \quad + \frac{1}{4}s^T \left( M''_{4,1}(t) + M''_{4,2}(t) \right) \Phi'(s) - \frac{1}{4}s^T \Phi'(s) t^T \Phi'(t) \end{aligned}$$

□

**Lemma 3** Für die in (1.6) und (1.9) definierten Zufallselemente  $Z_n$  und  $\tilde{Z}_n$  gilt:  
 $\|Z_n - \tilde{Z}_n\|_{L^2} \xrightarrow{P} 0.$

BEWEIS: Mit  $\Delta_j$  aus (1.10) gilt  $Y_j = X_j + \Delta_j$ . Die Additionstheoreme für die Sinus- bzw. die Cosinus-Funktion und eine elementare Abschätzung liefern die Identitäten (vgl. Henze und Wagner (1997), S. 8)

$$\cos(t^T Y_j) = \cos(t^T X_j) - t^T \Delta_j \sin(t^T X_j) + \varepsilon_{n,j}(t), \quad (1.13)$$

$$\sin(t^T Y_j) = \sin(t^T X_j) + t^T \Delta_j \cos(t^T X_j) + \eta_{n,j}(t) \quad (1.14)$$

mit  $|\varepsilon_{n,j}(t)| \leq \|t\|^2 \|\Delta_j\|^2$  und  $|\eta_{n,j}(t)| \leq \|t\|^2 \|\Delta_j\|^2$ .

Nach dem Zentralen Grenzwertsatz in  $\mathbb{R}^d$  gilt  $\sqrt{n}\bar{X}_n = O_P(1)$ ; nach dem starken Gesetz großer Zahlen gilt  $S_n \rightarrow I_d$  fast sicher und damit auch  $S_n^{-1/2} \rightarrow I_d$  fast sicher. Dies impliziert die Identitäten (vgl. Henze und Wagner (1997), S. 9)

$$\sqrt{n}(S_n - I_d) = \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) - \sqrt{n}\bar{X}_n \bar{X}_n^T \quad (1.15)$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) + O_P(n^{-1/2}),$$

$$\sqrt{n}(S_n^{-1} - I_d) = -\frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) + O_P(n^{-1/2}),$$

$$\sqrt{n}(S_n^{-1/2} - I_d) = -\frac{1}{2\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) + o_P(1).$$

Mit

$$\|\Delta_j\|^2 = X_j^T (S_n^{-1/2} - I_d)^2 X_j - 2\bar{X}_n^T S_n^{-1/2} (S_n^{-1/2} - I_d) X_j + \bar{X}_n^T S_n^{-1} \bar{X}_n$$

folgt daraus

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{j=1}^n \|\Delta_j\|^2 \\ & \leq \frac{1}{\sqrt{n}} \|\sqrt{n}(S_n^{-1/2} - I_d)\|_S^2 \frac{1}{n} \sum_{j=1}^n \|X_j\|^2 \\ & \quad + 2 \|\bar{X}_n\| \|S_n^{-1/2}\|_S \|\sqrt{n}(S_n^{-1/2} - I_d)\|_S \frac{1}{n} \sum_{j=1}^n \|X_j\| + \frac{1}{\sqrt{n}} \|\sqrt{n}\bar{X}_n\|^2 \|S_n^{-1}\|_S \\ & = o_P(1). \end{aligned}$$

Setzt man die vorangehenden Überlegungen zusammen, so ergibt sich für alle  $t \in \mathbb{R}^d$

$$\left| Z_n(t) - \tilde{Z}_n(t) \right| = \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n [\varepsilon_{n,j}(t) + \eta_{n,j}(t)] \right| \leq \frac{2\|t\|^2}{\sqrt{n}} \sum_{j=1}^n \|\Delta_j\|^2.$$

Da  $\int_{\mathbb{R}^d} \|t\|^4 \varphi_\beta(t) dt < \infty$  gilt, folgt

$$\begin{aligned} \|Z_n - \tilde{Z}_n\|_{L^2}^2 &= \int_{\mathbb{R}^d} \left( Z_n(t) - \tilde{Z}_n(t) \right)^2 \varphi_\beta(t) dt \\ &\leq 4 \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n \|\Delta_j\|^2 \right) \int_{\mathbb{R}^d} \|t\|^4 \varphi_\beta(t) dt \xrightarrow{P} 0 \end{aligned}$$

und damit die Behauptung.  $\square$

**Lemma 4** Für die in (1.9) und (1.11) definierten Zufallselemente  $\tilde{Z}_n$  und  $Z_n^*$  gilt:  
 $\|\tilde{Z}_n - Z_n^*\|_{L^2} \xrightarrow{P} 0$ .

**BEWEIS:** Setzt man  $\Delta_j = (S_n^{-1/2} - I_d)X_j - S_n^{-1/2}\bar{X}_n$  ein, so ergibt sich für die punktweise Differenz von  $\tilde{Z}_n$  und  $Z_n^*$

$$\begin{aligned} \tilde{Z}_n(t) - Z_n^*(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ [\cos(t^T X_j) - \sin(t^T X_j)] t^T \Delta_j + t^T X_j \tilde{\Phi}(t) + \frac{1}{2} t^T (X_j X_j^T - I_d) \Phi'(t) \right\} \\ &= t^T (S_n^{-1/2} - I_d) \frac{1}{\sqrt{n}} \sum_{j=1}^n [\cos(t^T X_j) - \sin(t^T X_j)] X_j \\ &\quad - t^T S_n^{-1/2} \bar{X}_n \frac{1}{\sqrt{n}} \sum_{j=1}^n [\cos(t^T X_j) - \sin(t^T X_j)] \\ &\quad + \tilde{\Phi}(t) t^T \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j + \frac{1}{2} t^T \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) \Phi'(t) \\ &= A_n(t) + t^T \sqrt{n} (S_n^{-1/2} - I_d) \Phi'(t) + B_n(t) - \tilde{\Phi}(t) t^T S_n^{-1/2} \sqrt{n} \bar{X}_n \\ &\quad + \tilde{\Phi}(t) t^T \sqrt{n} \bar{X}_n + t^T \frac{1}{2\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) \Phi'(t) \\ &= A_n(t) + B_n(t) + C_n(t) + D_n(t) \end{aligned}$$

mit

$$\begin{aligned} A_n(t) &:= t^T (S_n^{-1/2} - I_d) \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ \cos(t^T X_j) X_j - \sin(t^T X_j) X_j - \Phi'(t) \right\}, \\ B_n(t) &:= -t^T S_n^{-1/2} \bar{X}_n \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ \cos(t^T X_j) - \sin(t^T X_j) - \tilde{\Phi}(t) \right\}, \\ C_n(t) &:= \tilde{\Phi}(t) t^T (I_d - S_n^{-1/2}) \sqrt{n} \bar{X}_n, \\ D_n(t) &:= t^T \left[ \sqrt{n} (S_n^{-1/2} - I_d) + \frac{1}{2\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) \right] \Phi'(t) = t^T \tilde{D}_n \Phi'(t). \end{aligned}$$

Dabei ergibt sich die Matrix  $\tilde{D}_n$  durch Einsetzen von (1.15) in  $\sqrt{n}(S_n^{-1/2} - I_d) = -\sqrt{n}(S_n - I_d)S_n^{-1}(S_n^{-1/2} + I_d)^{-1}$  zu

$$\begin{aligned}\tilde{D}_n &:= \sqrt{n}(S_n^{-1/2} - I_d) + \frac{1}{2\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j X_j^T - I_d) \left\{ \frac{1}{2} I_d - S_n^{-1} (S_n^{-1/2} + I_d)^{-1} \right\} + \sqrt{n} \bar{X}_n \bar{X}_n^T S_n^{-1} (S_n^{-1/2} + I_d)^{-1}.\end{aligned}$$

Da nach der Dreiecksungleichung

$$\left\| \tilde{Z}_n - Z_n^* \right\|_{L^2} \leq \|A_n\|_{L^2} + \|B_n\|_{L^2} + \|C_n\|_{L^2} + \|D_n\|_{L^2}$$

gilt, reicht es, die stochastische Konvergenz von  $\|A_n\|_{L^2}^2$ ,  $\|B_n\|_{L^2}^2$ ,  $\|C_n\|_{L^2}^2$  und  $\|D_n\|_{L^2}^2$  gegen 0 zu zeigen, um den Beweis von Lemma 4 zu vollenden.

NACHWEIS VON  $\|A_n\|_{L^2}^2 \xrightarrow{P} 0$ : Es gilt

$$|A_n(t)| \leq \|t\| \|S_n^{-1/2} - I_d\| \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ \cos(t^T X_j) X_j - \sin(t^T X_j) X_j - \Phi'(t) \right] \right\|,$$

also

$$\begin{aligned}\|A_n\|_{L^2}^2 &= \int_{\mathbb{R}^d} A_n(t)^2 \varphi_\beta(t) dt \\ &\leq \|\sqrt{n}(S_n^{-1/2} - I_d)\|^2 \int_{\mathbb{R}^d} \left\| \frac{1}{n} \sum_{j=1}^n \tilde{A}(t, X_j) \right\|^2 \|t\|^2 \varphi_\beta(t) dt \quad (1.16)\end{aligned}$$

mit den unabhängigen, identisch verteilten und zentrierten Zufallsvektoren

$$\tilde{A}(t, X_j) := \cos(t^T X_j) X_j - \sin(t^T X_j) X_j - \Phi'(t) \quad (j = 1, \dots, n).$$

Der erste Faktor in (1.16) ist straff; folglich reicht es, die stochastische Konvergenz des Integrals in (1.16) gegen 0 zu zeigen. Dieses berechnet sich jedoch zu

$$\int_{\mathbb{R}^d} \left\| \frac{1}{n} \sum_{j=1}^n \tilde{A}(t, X_j) \right\|^2 \|t\|^2 \varphi_\beta(t) dt = \frac{1}{n^2} \sum_{i,j=1}^n \int_{\mathbb{R}^d} \tilde{A}(t, X_i)^T \tilde{A}(t, X_j) \|t\|^2 \varphi_\beta(t) dt,$$

also einer  $V$ -Statistik zweiten Grades, die nach Theorem D.2 aus Anhang D fast sicher gegen

$$\begin{aligned}\mathbb{E} \left[ \int_{\mathbb{R}^d} \tilde{A}(t, X_1)^T \tilde{A}(t, X_2) \|t\|^2 \varphi_\beta(t) dt \right] \\ = \int_{\mathbb{R}^d} \mathbb{E} \left[ \tilde{A}(t, X_1) \right]^T \mathbb{E} \left[ \tilde{A}(t, X_2) \right] \|t\|^2 \varphi_\beta(t) dt = 0\end{aligned}$$

konvergiert.

NACHWEIS VON  $\|B_n\|_{L^2}^2 \xrightarrow{P} 0$ : Es ist

$$\begin{aligned} \|B_n\|_{L^2}^2 &= \int_{\mathbb{R}^d} B_n(t)^2 \varphi_\beta(t) dt \\ &= (\sqrt{n}\bar{X}_n)^T S_n^{-1/2} \int_{\mathbb{R}^d} tt^T \left( \frac{1}{n} \sum_{j=1}^n \tilde{B}(t, X_j) \right)^2 \varphi_\beta(t) dt S_n^{-1/2} (\sqrt{n}\bar{X}_n) \end{aligned} \quad (1.17)$$

mit den unabhängigen, identisch verteilten und zentrierten Zufallsvariablen

$$\tilde{B}(t, X_j) := \cos(t^T X_j) - \sin(t^T X_j) - \tilde{\Phi}(t) \quad (j = 1, \dots, n).$$

Aufgrund der Straffheit von  $\sqrt{n}\bar{X}_n$  und  $S_n^{-1/2}$  reicht es, die stochastische Konvergenz des Integrals in (1.17) gegen 0 zu zeigen. Dieses ist gleich

$$\frac{1}{n^2} \sum_{i,j=1}^n \int_{\mathbb{R}^d} tt^T \tilde{B}(t, X_i) \tilde{B}(t, X_j) \varphi_\beta(t) dt,$$

also einer  $V$ -Statistik zweiten Grades, die nach Theorem D.2 aus Anhang D fast sicher gegen

$$\begin{aligned} \mathbb{E} \left[ \int_{\mathbb{R}^d} tt^T \tilde{B}(t, X_1) \tilde{B}(t, X_2) \varphi_\beta(t) dt \right] \\ = \int_{\mathbb{R}^d} tt^T \mathbb{E} \left[ \tilde{B}(t, X_1) \right] \mathbb{E} \left[ \tilde{B}(t, X_2) \right] \varphi_\beta(t) dt = 0 \end{aligned}$$

konvergiert.

NACHWEIS VON  $\|C_n\|_{L^2}^2 \xrightarrow{P} 0$ : Es gilt

$$|C_n(t)| \leq \left| \tilde{\Phi}(t) \right| \cdot \|t\| \|I_d - S_n^{-1/2}\| \|\sqrt{n}\bar{X}_n\|,$$

also

$$\begin{aligned} \|C_n\|_{L^2}^2 &= \int_{\mathbb{R}^d} C_n(t)^2 \varphi_\beta(t) dt \\ &\leq \|I_d - S_n^{-1/2}\|^2 \|\sqrt{n}\bar{X}_n\|^2 \int_{\mathbb{R}^d} \left| \tilde{\Phi}(t) \right|^2 \|t\|^2 \varphi_\beta(t) dt \xrightarrow{P} 0, \end{aligned}$$

denn aufgrund der Beschränktheit von  $\tilde{\Phi}$  ist das obige Integral endlich. Ferner ist  $\sqrt{n}\bar{X}_n$  straff, und  $S_n^{-1/2}$  strebt fast sicher gegen  $I_d$ .

NACHWEIS VON  $\|D_n\|_{L^2}^2 \xrightarrow{P} 0$ : Ebenso gilt

$$|D_n(t)| \leq \|t\| \left\| \tilde{D}_n \right\| \left\| \Phi'(t) \right\|,$$

also

$$\|D_n\|_{L^2}^2 = \int_{\mathbb{R}^d} D_n(t)^2 \varphi_\beta(t) dt \leq \|\tilde{D}_n\|^2 \int_{\mathbb{R}^d} \|\Phi'(t)\|^2 \|t\|^2 \varphi_\beta(t) dt \xrightarrow{P} 0,$$

denn aufgrund der Beschränktheit von  $\|\Phi'(t)\|$  ist das obige Integral endlich. Ferner gilt  $\|\tilde{D}_n\|^2 \xrightarrow{P} 0$  aufgrund der Straffheit von  $n^{-1/2} \sum_{j=1}^n (X_j X_j^T - I_d)$  bzw. von  $\sqrt{n} \bar{X}_n$  und aufgrund der fast sicheren Konvergenz von  $\bar{X}_n$  gegen 0 bzw. von  $S_n^{-1}$  gegen  $I_d$ .  $\square$

**Theorem 5** Es sei  $X_1, X_2, \dots$  eine Folge von unabhängigen und identisch verteilten  $d$ -dimensionalen standardisierten Zufallsvektoren mit nichtdegenerierter Verteilung und endlichen vierten Momenten.

Dann konvergiert die in Gleichung (1.3) auf Seite 8 definierte Zufallsvariable  $S_{n,\beta}$  nach Verteilung gegen ein Grenzelement  $S_\beta$ , das gegeben ist durch

$$S_\beta := 2 \int_{\mathbb{R}^d} Z(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt$$

mit dem Gaußprozeß  $Z$  aus Theorem 1. Es gilt  $S_\beta \sim \mathcal{N}(0, \tau_\beta^2)$  mit

$$\begin{aligned} \tau_\beta^2 := & 4 \cdot \left\{ K_{\beta 1}^1 + K_{\beta 2}^1 + (K_{\beta 1}^2 + K_{\beta 2}^2) (K_{\beta 1}^3 + K_{\beta 2}^3 - (K_{\beta 1}^2 + K_{\beta 2}^2)) - \frac{1}{4} (K_{\beta 1}^3 + K_{\beta 2}^3)^2 \right. \\ & + \left\| K_{\beta 1}^{5'} - K_{\beta 2}^{5'} \right\|^2 + (K_{\beta 1}^{5'} - K_{\beta 2}^{5'})^T (2(K_{\beta 1}^{4'} - K_{\beta 2}^{4'}) + K_{\beta 1}^{6'} + K_{\beta 2}^{6'}) \\ & \left. + (K_{\beta 1}^9 + K_{\beta 2}^9) \left( K_{\beta 1}^{7''} + K_{\beta 2}^{7''} + \frac{1}{4} (K_{\beta 1}^{8''} + K_{\beta 2}^{8''}) \right) \right\}. \end{aligned} \quad (1.18)$$

Dabei bezeichnen

$$\begin{aligned} K_{\beta 1}^1 &:= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi_1(t-s) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt, \\ K_{\beta 2}^1 &:= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi_2(t+s) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt, \\ K_{\beta 1}^2 &:= \int_{\mathbb{R}^d} \Phi_1(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 2}^2 &:= \int_{\mathbb{R}^d} \Phi_2(t)^2 \varphi_\beta(t) dt, \\ K_{\beta 1}^3 &:= \int_{\mathbb{R}^d} t^T \Phi_1'(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 2}^3 &:= \int_{\mathbb{R}^d} t^T \Phi_2'(t) \Phi_2(t) \varphi_\beta(t) dt \end{aligned}$$

Skalare,

$$K_{\beta 1}^{4'} := \int_{\mathbb{R}^d} \Phi_1'(t) \Phi_2(t) \varphi_\beta(t) dt, \quad K_{\beta 2}^{4'} := \int_{\mathbb{R}^d} \Phi_2'(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt,$$

$$\begin{aligned}
K_{\beta 1}^{5'} &:= \int_{\mathbb{R}^d} t \Phi_1(t) \Phi_2(t) \varphi_\beta(t) dt, & K_{\beta 2}^{5'} &:= \int_{\mathbb{R}^d} t \Phi_2(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, \\
K_{\beta 2}^{6'} &:= \int_{\mathbb{R}^d} M'_{3,2}(t) \Phi_2(t) \varphi_\beta(t) dt, & K_{\beta 1}^{6'} &:= \int_{\mathbb{R}^d} M'_{3,1}(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt
\end{aligned}$$

Vektoren und

$$\begin{aligned}
K_{\beta 1}^{7''} &:= \int_{\mathbb{R}^d} \Phi_1''(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, & K_{\beta 2}^{7''} &:= \int_{\mathbb{R}^d} \Phi_2''(t) \Phi_2(t) \varphi_\beta(t) dt, \\
K_{\beta 1}^{8''} &:= \int_{\mathbb{R}^d} M''_{4,1}(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, & K_{\beta 2}^{8''} &:= \int_{\mathbb{R}^d} M''_{4,2}(t) \Phi_2(t) \varphi_\beta(t) dt
\end{aligned}$$

Matrizen. Für  $A \in \mathbb{R}^{d \times d}$  notieren wir

$$\begin{aligned}
K_{\beta 1}^9(A) &:= \int_{\mathbb{R}^d} t^T A \Phi_1'(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, \\
K_{\beta 2}^9(A) &:= \int_{\mathbb{R}^d} t^T A \Phi_2'(t) \Phi_2(t) \varphi_\beta(t) dt
\end{aligned}$$

sowie  $(K_{\beta 1}^9 + K_{\beta 2}^9)(A) := K_{\beta 1}^9(A) + K_{\beta 2}^9(A)$ .

BEWEIS: Nach (1.5) ist

$$S_{n,\beta} = 2 \int_{\mathbb{R}^d} Z_n(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + \frac{1}{\sqrt{n}} \int_{\mathbb{R}^d} Z_n(t)^2 \varphi_\beta(t) dt,$$

und nach Theorem 1 konvergiert  $Z_n(\cdot)$  schwach in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t) dt)$  gegen einen zentrierten Gaußprozeß  $Z$  mit dem in Theorem 1 angegebenen Kovarianzkern  $k$ . Wie schon auf Seite 10 erwähnt, sind die in der Darstellung von  $S_{n,\beta}$  auftretenden Integrale als Skalarprodukt von  $Z_n(\cdot)$  mit der Funktion  $h(\cdot) := \Phi(\cdot) - e^{-\frac{\|\cdot\|^2}{2}} \in L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t) dt)$  bzw. als Norm von  $Z_n(\cdot)$  stetige Funktionale, und nach dem Abbildungssatz folgt

$$S_{n,\beta} \xrightarrow{\mathcal{D}} S_\beta = 2 \int_{\mathbb{R}^d} Z(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt.$$

Dabei ist das auf der rechten Seite auftretende Integral punktweise auf  $\Omega$  definiert. Die reelle Zufallsvariable  $S_\beta$  ist als Skalarprodukt zwischen dem Gaußprozeß  $Z$  und der Funktion  $h$  nach Definition normalverteilt (vgl. Definition B.12 in Anhang B). Da  $Z$  zentriert ist, ist auch  $S_\beta$  zentriert. Folglich verbleibt es, die Varianz  $\tau_\beta^2$  zu berechnen:

$$\begin{aligned}
\tau_\beta^2 &= \mathbb{E} \left[ \left( 2 \int_{\mathbb{R}^d} Z(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right)^2 \right] \\
&= 4 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} k(s, t) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) \varphi_\beta(t) ds dt \\
&= 4 \left\{ K_{\beta 1}^1 + K_{\beta 2}^1 - \left( \int \Phi(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right)^2 \right\}
\end{aligned}$$

$$\begin{aligned}
& +2 \left( \int t \tilde{\Phi}(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right)^T \left( \int \tilde{\Phi}'(s) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \varphi_\beta(s) ds \right) \\
& + \left\| \int t \tilde{\Phi}(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right\|^2 \\
& + \int t^T \left( \int \Phi''(s) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \varphi_\beta(s) ds \right) \Phi'(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\
& + \left( \int \Phi(s) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \varphi_\beta(s) ds \right) \left( \int t^T \Phi'(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right) \\
& + \int \tilde{\Phi}(s) s^T \left( \int [M'_{3,1}(t) + M'_{3,2}(t)] \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \varphi_\beta(s) ds \\
& + \frac{1}{4} \int s^T \left( \int [M''_{4,1}(t) + M''_{4,2}(t)] \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right) \Phi'(s) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \varphi_\beta(s) ds \\
& - \frac{1}{4} \left( \int t^T \Phi'(t) \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right)^2 \Big\} .
\end{aligned}$$

Beachtet man nun, daß aus Symmetriegründen Skalare wie  $\int \Phi_1(t)\Phi_2(t)\varphi_\beta(t)dt$  und Vektoren wie  $\int t\Phi_1(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t)dt$  verschwinden, so erhält man:

$$\begin{aligned}
\tau_\beta^2 & = 4 \cdot \left\{ K_{\beta 1}^1 + K_{\beta 2}^1 - (K_{\beta 1}^2 + K_{\beta 2}^2)^2 + 2 \left( K_{\beta 1}^{5'} - K_{\beta 2}^{5'} \right)^T \left( K_{\beta 1}^{4'} - K_{\beta 2}^{4'} \right) + \left\| K_{\beta 1}^{5'} - K_{\beta 2}^{5'} \right\|^2 \right. \\
& \quad + \int t^T \left( K_{\beta 1}^{7''} + K_{\beta 2}^{7''} \right) \Phi_1'(t) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\
& \quad + \int t^T \left( K_{\beta 1}^{7''} + K_{\beta 2}^{7''} \right) \Phi_2'(t)\Phi_2(t) \varphi_\beta(t) dt \\
& \quad + (K_{\beta 1}^2 + K_{\beta 2}^2) (K_{\beta 1}^3 + K_{\beta 2}^3) + \left( K_{\beta 1}^{5'} - K_{\beta 2}^{5'} \right)^T \left( K_{\beta 1}^{6'} + K_{\beta 2}^{6'} \right) \\
& \quad + \frac{1}{4} \int s^T \left( K_{\beta 1}^{8''} + K_{\beta 2}^{8''} \right) \Phi_1'(s) \left[ \Phi_1(s) - e^{-\frac{\|s\|^2}{2}} \right] \varphi_\beta(s) ds \\
& \quad \left. + \frac{1}{4} \int s^T \left( K_{\beta 1}^{8''} + K_{\beta 2}^{8''} \right) \Phi_2'(s)\Phi_2(s) \varphi_\beta(st) ds - \frac{1}{4} (K_{\beta 1}^3 + K_{\beta 2}^3)^2 \right\} \\
& = 4 \cdot \left\{ K_{\beta 1}^1 + K_{\beta 2}^1 + (K_{\beta 1}^2 + K_{\beta 2}^2) (K_{\beta 1}^3 + K_{\beta 2}^3 - (K_{\beta 1}^2 + K_{\beta 2}^2)) - \frac{1}{4} (K_{\beta 1}^3 + K_{\beta 2}^3)^2 \right. \\
& \quad + \left\| K_{\beta 1}^{5'} - K_{\beta 2}^{5'} \right\|^2 + \left( K_{\beta 1}^{5'} - K_{\beta 2}^{5'} \right)^T \left( 2(K_{\beta 1}^{4'} - K_{\beta 2}^{4'}) + K_{\beta 1}^{6'} + K_{\beta 2}^{6'} \right) \\
& \quad \left. + (K_{\beta 1}^9 + K_{\beta 2}^9) \left( K_{\beta 1}^{7''} + K_{\beta 2}^{7''} + \frac{1}{4}(K_{\beta 1}^{8''} + K_{\beta 2}^{8''}) \right) \right\} . \quad \square
\end{aligned}$$

**Bemerkung 6** Die Methode, die empirische Größe  $Z_n(\cdot)$  als Zufallselement in dem Hilbert-Raum  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  zu betrachten, ihre Konvergenz in diesem Raum gegen einen zentrierten Gaußprozeß  $Z$  nachzuweisen und zum Nachweis der Konvergenz der zugehörigen  $L^2$ -Norm den Abbildungssatz zu verwenden, läßt sich auch in dem von Henze und Wagner (1997)

untersuchten  $H_0$ -Kontext anwenden. Damit ist eine dritte Möglichkeit gegeben, die Konvergenz der BHEP-Statistik unter der Nullhypothese zu zeigen und die Grenzverteilung zu bestimmen. Die ersten beiden Methoden hierfür waren die Anwendung des Theorems von De Wet und Randles (1987) durch Baringhaus und Henze (1988) im Kontext der  $V$ -Statistiken mit geschätzten Parametern und das Betrachten von  $Z_n(\cdot)$  als Zufallselement in dem Fréchet-Raum  $\mathcal{C}(\mathbb{R}^d)$  der stetigen Funktionen auf  $\mathbb{R}^d$  durch Henze und Wagner (1997). Um den Konvergenznachweis auf lokale Alternativen auszuweiten, kann man entweder so vorgehen wie in Henze und Wagner (1997) und die z.B. in Witting und Müller-Funk (1995), Kapitel 6.3, dargestellte und in Anhang E der vorliegenden Arbeit kurz zusammengefaßte Theorie benachbarter Alternativen verwenden, oder man hat die Möglichkeit, direkt die Konvergenz des dann auf Dreiecksschemata von unabhängigen Zufallsvariablen beruhenden Zufallselementes  $Z_n(\cdot)$  in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$  zu zeigen, unter Verwendung von Theorem B.17 aus Anhang B.

## 1.2. Schätzen der Varianz $\tau_\beta^2$

Um aus der Konvergenzaussage von Theorem 5 verwendbare statistische Folgerungen herzuleiten, muß die bei unbekannter zugrundeliegender Verteilung ebenfalls unbekannte Grenzvarianz  $\tau_\beta^2$  geschätzt werden. Diese in Gleichung (1.18) auf Seite 18 definierte Größe ist die Summe von Integralen über die Bestandteile  $\Phi_1$  und  $\Phi_2$  der Sinus/Cosinus-Transformierten der zugrundeliegenden Verteilung. Folglich liegt es nahe, die Unbekannten  $\Phi_1$  und  $\Phi_2$  sowie ihre Ableitungen nach  $t \in \mathbb{R}^d$  durch konsistente Schätzer zu ersetzen und als Schätzer für  $\tau_\beta^2$  einen zu Gleichung (1.18) auf Seite 18 analogen empirischen Ausdruck zu bilden.

Als Schätzer für  $\Phi_1$ ,  $\Phi_2$  und ihre Ableitungen bieten sich arithmetische Mittelwerte über die entsprechenden empirischen Größen an. Da die zugrundeliegende Verteilung stets als standardisiert vorausgesetzt wurde, müssen auch die standardisierten Daten  $Y_j$  anstelle der ursprünglichen Daten  $X_j$  in die Schätzer eingesetzt werden. Diese sind gegeben durch

$$\begin{aligned}\Phi_{1,n}(t) &= \frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j), & \Phi_{2,n}(t) &= \frac{1}{n} \sum_{j=1}^n \sin(t^T Y_j), \\ \Phi'_{1,n}(t) &:= -\frac{1}{n} \sum_{j=1}^n \sin(t^T Y_j) Y_j, & \Phi'_{2,n}(t) &:= \frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) Y_j, \\ \Phi''_{1,n}(t) &:= -\frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) Y_j Y_j^T, & \Phi''_{2,n}(t) &:= -\frac{1}{n} \sum_{j=1}^n \sin(t^T Y_j) Y_j Y_j^T\end{aligned}$$

bzw. durch

$$\Phi_n(t) = \Phi_{1,n}(t) + \Phi_{2,n}(t), \quad \Phi'_n(t) = \Phi'_{1,n}(t) + \Phi'_{2,n}(t), \quad \Phi''_n(t) = \Phi''_{1,n}(t) + \Phi''_{2,n}(t)$$

(vgl. auch Gleichung (1.4) auf Seite 9 und die Definitionen auf Seite 11). Ferner sind

$$M'_{3,1,n}(t) := -\frac{1}{n^2} \sum_{j,k=1}^n \sin(t^T Y_j) (Y_k^T Y_j) (t^T Y_k) Y_k,$$

$$\begin{aligned}
M'_{3,2,n}(t) &:= \frac{1}{n^2} \sum_{j,k=1}^n \cos(t^T Y_j) (Y_k^T Y_j) (t^T Y_k) Y_k, \\
M''_{4,1,n}(t) &:= -\frac{1}{n^2} \sum_{j,k=1}^n \sin(t^T Y_j) (t^T Y_k) (Y_k^T Y_j) Y_k Y_k^T, \\
M''_{4,2,n}(t) &:= \frac{1}{n^2} \sum_{j,k=1}^n \cos(t^T Y_j) (t^T Y_k) (Y_k^T Y_j) Y_k Y_k^T
\end{aligned}$$

naheliegende Schätzer für die in den Gleichungen (1.7) und (1.8) auf Seite 11 definierten dritten Momente  $M'_{3,1}(t)$ ,  $M'_{3,2}(t)$  und vierten Momente  $M''_{4,1}(t)$ ,  $M''_{4,2}(t)$ . Ihre Herleitung erfolgt leicht durch Ersetzen der dort befindlichen Erwartungswerte durch empirische Analoga.

Zunächst muß man sich davon überzeugen, daß die empirischen Größen  $\Phi_{1,n}(t), \Phi_{2,n}(t), \dots, M''_{4,2,n}(t)$  für jeden festen Parameter  $t \in \mathbb{R}^d$  fast sicher gegen  $\Phi_1(t), \Phi_2(t), \dots, M''_{4,2}(t)$  konvergieren, obwohl sie aufgrund der in den standardisierten Daten  $Y_j$  enthaltenen Schätzer  $\bar{X}_n$  und  $S_n^{-1/2}$  für die unbekannt Parameter keine Summen von unabhängigen und identisch verteilten Zufallsvariablen sind. Den ersten Schritt in diese Richtung stellt Lemma 7 dar, in dem gezeigt wird, daß die Differenzen zwischen den anhand der standardisierten Daten  $Y_j$  definierten und den anhand der ursprünglichen Daten  $X_j$  in analoger Weise gebildeten Schätzern vernachlässigbar sind. Letztere sind erklärt durch

$$\begin{aligned}
\Phi_{1,n}^\diamond(t) &:= \frac{1}{n} \sum_{j=1}^n \cos(t^T X_j), & \Phi_{2,n}^\diamond(t) &:= \frac{1}{n} \sum_{j=1}^n \sin(t^T X_j), \\
\Phi'_{1,n}^\diamond(t) &:= -\frac{1}{n} \sum_{j=1}^n \sin(t^T X_j) X_j, & \Phi'_{2,n}^\diamond(t) &:= \frac{1}{n} \sum_{j=1}^n \cos(t^T X_j) X_j, \\
\Phi''_{1,n}^\diamond(t) &:= -\frac{1}{n} \sum_{j=1}^n \cos(t^T X_j) X_j X_j^T, & \Phi''_{2,n}^\diamond(t) &:= -\frac{1}{n} \sum_{j=1}^n \sin(t^T X_j) X_j X_j^T
\end{aligned}$$

und  $\Phi_n^\diamond(t) := \Phi_{1,n}^\diamond(t) + \Phi_{2,n}^\diamond(t)$  sowie

$$\begin{aligned}
M'_{3,1,n}^\diamond(t) &:= -\frac{1}{n^2} \sum_{j,k=1}^n \sin(t^T X_j) (X_k^T X_j) (t^T X_k) X_k, \\
M'_{3,2,n}^\diamond(t) &:= \frac{1}{n^2} \sum_{j,k=1}^n \cos(t^T X_j) (X_k^T X_j) (t^T X_k) X_k, \\
M''_{4,1,n}^\diamond(t) &:= -\frac{1}{n^2} \sum_{j,k=1}^n \sin(t^T X_j) (t^T X_k) (X_k^T X_j) X_k X_k^T, \\
M''_{4,2,n}^\diamond(t) &:= \frac{1}{n^2} \sum_{j,k=1}^n \cos(t^T X_j) (t^T X_k) (X_k^T X_j) X_k X_k^T.
\end{aligned}$$

Damit liefert Lemma 7 den folgenden Zusammenhang zwischen den auf den standardisierten Daten basierenden Schätzern  $\Phi_{1,n}(t), \Phi_{2,n}(t), \dots, M_{4,2,n}''(t)$  und den auf den ursprünglichen Daten beruhenden Schätzern  $\Phi_{1,n}^\diamond(t), \Phi_{2,n}^\diamond(t), \dots, M_{4,2,n}''^\diamond(t)$ .

**Lemma 7** Für jedes feste  $t \in \mathbb{R}^d$  gilt mit jeweils  $i = 1, 2$

1. (i)  $\Phi_{1,n}(t) = \Phi_{1,n}^\diamond(t) + R_{1,n}(t)$   
(ii)  $\Phi_{2,n}(t) = \Phi_{2,n}^\diamond(t) + R_{2,n}(t)$  } mit  $|R_{i,n}(t)| \leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}$ ,
2. (i)  $\Phi'_{1,n}(t) = \Phi'^\diamond_{1,n}(t) + R'_{1,n}(t)$   
(ii)  $\Phi'_{2,n}(t) = \Phi'^\diamond_{2,n}(t) + R'_{2,n}(t)$  } mit  $\|R'_{i,n}(t)\| \leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}$ ,
3. (i)  $\Phi''_{1,n}(t) = \Phi''^\diamond_{1,n}(t) + R''_{1,n}(t)$   
(ii)  $\Phi''_{2,n}(t) = \Phi''^\diamond_{2,n}(t) + R''_{2,n}(t)$  } mit  $\|R''_{i,n}(t)\|_S \leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}$ ,
4. (i)  $M'_{3,1,n}(t) = M'^\diamond_{3,1,n}(t) + R'_{3,1,n}(t)$   
(ii)  $M'_{3,2,n}(t) = M'^\diamond_{3,2,n}(t) + R'_{3,2,n}(t)$  } mit  $\|R'_{3,i,n}(t)\| \leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}$ ,
5. (i)  $M''_{4,1,n}(t) = M''^\diamond_{4,1,n}(t) + R''_{4,1,n}(t)$   
(ii)  $M''_{4,2,n}(t) = M''^\diamond_{4,2,n}(t) + R''_{4,2,n}(t)$  } mit  $\|R''_{4,i,n}(t)\|_S \leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}$ .

Dabei bezeichnen  $R_{1,n}(t)$  und  $R_{2,n}(t)$  Zufallsvariablen,  $R'_{1,n}(t)$ ,  $R'_{2,n}(t)$ ,  $R'_{3,1,n}(t)$  und  $R'_{3,2,n}(t)$   $d$ -dimensionale Zufallsvektoren sowie  $R''_{1,n}(t)$ ,  $R''_{2,n}(t)$ ,  $R''_{4,1,n}(t)$  und  $R''_{4,2,n}(t)$   $d \times d$ -Zufallsmatrizen, die den oben angegebenen Abschätzungen genügen. Die Terme  $o_{f.s.}(1)$  auf der rechten Seite dieser Abschätzungen bezeichnen jeweils eine Zufallsvariable, die fast sicher gegen 0 konvergiert und nicht mehr von dem Parameter  $t$  abhängt.

**BEWEIS:** Der Beweis ist sehr technisch und wurde deshalb in Abschnitt 1.5 ausgelagert, wo er ab Seite 60 zu finden ist. Im wesentlichen müssen die in den standardisierten Daten  $Y_j$  auftretenden Schätzer  $\bar{X}_n$  und  $S_n^{-1/2}$  „herausgerechnet“ werden, was durch eine Taylor-Entwicklung ersten Grades der Sinus/Cosinus-Terme geschieht, in denen die  $Y_j$  vorkommen. Hierzu werden die Entwicklungen (1.13) und (1.14) von Seite 14 verwendet.  $\square$

**Bemerkung 8** Die fast sichere Konvergenz der Schätzer  $\Phi_{1,n}(t), \Phi_{2,n}(t), \dots, M_{4,2,n}''(t)$  gegen  $\Phi_1(t), \Phi_2(t), \dots, M_{4,2}''(t)$  für jeden festen Parameter  $t \in \mathbb{R}^d$  folgt anhand der Darstellungen aus Lemma 7 unmittelbar nach dem starken Gesetz großer Zahlen, denn die auf den ursprünglichen Daten basierenden Größen  $\Phi_{1,n}^\diamond(t), \Phi_{2,n}^\diamond(t), \dots, M_{4,2,n}''^\diamond(t)$  sind gewichtete Summen von unab-

hängigen und identisch verteilten Zufallsvariablen, –vektoren oder –matrizen. So zerfällt z.B. die Doppelsumme  $M_{4,1,n}''\diamond(t)$  in die beiden Bestandteile

$$M_{4,1,n}''\diamond(t) = - \sum_{l=1}^d \left( \frac{1}{n} \sum_{j=1}^n \sin(t^T X_j) X_{j,l} \right) \left( \frac{1}{n} \sum_{k=1}^n X_{k,l} (t^T X_k) X_k X_k^T \right),$$

wobei  $X_{j,l}$  bzw.  $X_{k,l}$  ( $l = 1, \dots, d$ ) die  $d$  Komponenten der Vektoren  $X_j$  bzw.  $X_k$  bezeichnen. Man beachte, daß in allen zehn vorliegenden Fällen die eingangs getroffene Annahme der Existenz von Momenten vierter Ordnung für die fast sichere Konvergenz nach dem starken Gesetz großer Zahlen ausreicht.

Mit Hilfe der Schätzer  $\Phi_{1,n}, \Phi_{2,n}, \dots, M_{4,2,n}''$  können wir nun die empirischen Analoga der in Theorem 5 auf Seite 18 erklärten Bestandteile von  $\tau_\beta^2$  definieren. Durch anschließendes Zusammenfügen dieser Bestandteile gelangt man zu einem Schätzer für  $\tau_\beta^2$  selbst.

**Definition 9** Es bezeichnen

$$\begin{aligned} K_{\beta 1n}^1 &:= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi_{1,n}(t-s) \left[ \Phi_n(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi_n(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt, \\ K_{\beta 2n}^1 &:= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi_{2,n}(t+s) \left[ \Phi_n(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi_n(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt, \\ K_{\beta 1n}^2 &:= \int_{\mathbb{R}^d} \Phi_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 2n}^2 &:= \int_{\mathbb{R}^d} \Phi_{2,n}(t)^2 \varphi_\beta(t) dt, \\ K_{\beta 1n}^3 &:= \int_{\mathbb{R}^d} t^T \Phi'_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 2n}^3 &:= \int_{\mathbb{R}^d} t^T \Phi'_{2,n}(t) \Phi_{2,n}(t) \varphi_\beta(t) dt \end{aligned}$$

Skalare,

$$\begin{aligned} K_{\beta 1n}^{4'} &:= \int_{\mathbb{R}^d} \Phi'_{1,n}(t) \Phi_{2,n}(t) \varphi_\beta(t) dt, & K_{\beta 2n}^{4'} &:= \int_{\mathbb{R}^d} \Phi'_{2,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 1n}^{5'} &:= \int_{\mathbb{R}^d} t \Phi_{1,n}(t) \Phi_{2,n}(t) \varphi_\beta(t) dt, & K_{\beta 2n}^{5'} &:= \int_{\mathbb{R}^d} t \Phi_{2,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 2n}^{6'} &:= \int_{\mathbb{R}^d} M'_{3,2,n}(t) \Phi_{2,n}(t) \varphi_\beta(t) dt, & K_{\beta 1n}^{6'} &:= \int_{\mathbb{R}^d} M'_{3,1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \end{aligned}$$

Vektoren und

$$\begin{aligned} K_{\beta 1n}^{7''} &:= \int_{\mathbb{R}^d} \Phi''_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, & K_{\beta 2n}^{7''} &:= \int_{\mathbb{R}^d} \Phi''_{2,n}(t) \Phi_{2,n}(t) \varphi_\beta(t) dt, \\ K_{\beta 1n}^{8''} &:= \int_{\mathbb{R}^d} M''_{4,1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt, & K_{\beta 2n}^{8''} &:= \int_{\mathbb{R}^d} M''_{4,2,n}(t) \Phi_{2,n}(t) \varphi_\beta(t) dt \end{aligned}$$

Matrizen. Für  $A \in \mathbb{R}^{d \times d}$  notieren wir

$$K_{\beta 1n}^9(A) := \int_{\mathbb{R}^d} t^T A \Phi'_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt,$$

$$K_{\beta 2n}^9(A) := \int_{\mathbb{R}^d} t^T A \Phi'_{2,n}(t) \Phi_{2,n}(t) \varphi_\beta(t) dt$$

sowie  $(K_{\beta 1n}^9 + K_{\beta 2n}^9)(A) := K_{\beta 1n}^9(A) + K_{\beta 2n}^9(A)$ . Damit wird der Schätzer  $\tau_{\beta n}^2$  für  $\tau_\beta^2$  analog zu (1.18) erklärt:

$$\begin{aligned} \tau_{\beta n}^2 := & 4 \cdot \left\{ K_{\beta 1n}^1 + K_{\beta 2n}^1 + (K_{\beta 1n}^2 + K_{\beta 2n}^2) (K_{\beta 1n}^3 + K_{\beta 2n}^3 - (K_{\beta 1n}^2 + K_{\beta 2n}^2)) - \frac{1}{4} (K_{\beta 1n}^3 + K_{\beta 2n}^3)^2 \right. \\ & + \left\| K_{\beta 1n}^{5'} - K_{\beta 2n}^{5'} \right\|^2 + (K_{\beta 1n}^{5'} - K_{\beta 2n}^{5'})^T \left( 2(K_{\beta 1n}^{4'} - K_{\beta 2n}^{4'}) + K_{\beta 1n}^{6'} + K_{\beta 2n}^{6'} \right) \\ & \left. + (K_{\beta 1n}^9 + K_{\beta 2n}^9) \left( K_{\beta 1n}^{7''} + K_{\beta 2n}^{7''} + \frac{1}{4}(K_{\beta 1n}^{8''} + K_{\beta 2n}^{8''}) \right) \right\}. \end{aligned} \quad (1.19)$$

Zunächst ist die Konsistenz des Schätzers  $\tau_{\beta n}^2$  zu zeigen, was durch Nachweis der fast sicheren Konvergenz der Bestandteile  $K_{\beta 1n}^1, K_{\beta 2n}^1, \dots, K_{\beta 2n}^9$  geschieht.

**Satz 10** I) Es gelten

$$\begin{array}{ll} 1.(i) & K_{\beta 1n}^1 \rightarrow K_{\beta 1}^1 \quad \text{f.s.} & (ii) & K_{\beta 2n}^1 \rightarrow K_{\beta 2}^1 \quad \text{f.s.} \\ 2.(i) & K_{\beta 1n}^2 \rightarrow K_{\beta 1}^2 \quad \text{f.s.} & (ii) & K_{\beta 2n}^2 \rightarrow K_{\beta 2}^2 \quad \text{f.s.} \\ 3.(i) & K_{\beta 1n}^3 \rightarrow K_{\beta 1}^3 \quad \text{f.s.} & (ii) & K_{\beta 2n}^3 \rightarrow K_{\beta 2}^3 \quad \text{f.s.} \\ 4.(i) & K_{\beta 1n}^{4'} \rightarrow K_{\beta 1}^{4'} \quad \text{f.s.} & (ii) & K_{\beta 2n}^{4'} \rightarrow K_{\beta 2}^{4'} \quad \text{f.s.} \\ 5.(i) & K_{\beta 1n}^{5'} \rightarrow K_{\beta 1}^{5'} \quad \text{f.s.} & (ii) & K_{\beta 2n}^{5'} \rightarrow K_{\beta 2}^{5'} \quad \text{f.s.} \\ 6.(i) & K_{\beta 1n}^{6'} \rightarrow K_{\beta 1}^{6'} \quad \text{f.s.} & (ii) & K_{\beta 2n}^{6'} \rightarrow K_{\beta 2}^{6'} \quad \text{f.s.} \\ 7.(i) & K_{\beta 1n}^{7''} \rightarrow K_{\beta 1}^{7''} \quad \text{f.s.} & (ii) & K_{\beta 2n}^{7''} \rightarrow K_{\beta 2}^{7''} \quad \text{f.s.} \\ 8.(i) & K_{\beta 1n}^{8''} \rightarrow K_{\beta 1}^{8''} \quad \text{f.s.} & (ii) & K_{\beta 2n}^{8''} \rightarrow K_{\beta 2}^{8''} \quad \text{f.s.} \end{array}$$

Dabei gilt die Konvergenz der Vektoren in 4., 5. und 6. bzw. die Konvergenz der Matrizen in 7. und 8. komponentenweise oder (äquivalent dazu) in einer beliebigen Norm.

II) Außerdem gilt

$$\begin{aligned} & (K_{\beta 1n}^9 + K_{\beta 2n}^9) \left( K_{\beta 1n}^{7''} + K_{\beta 2n}^{7''} + \frac{1}{4}(K_{\beta 1n}^{8''} + K_{\beta 2n}^{8''}) \right) \\ & \rightarrow (K_{\beta 1}^9 + K_{\beta 2}^9) \left( K_{\beta 1}^{7''} + K_{\beta 2}^{7''} + \frac{1}{4}(K_{\beta 1}^{8''} + K_{\beta 2}^{8''}) \right) \quad \text{f.s.} \end{aligned}$$

III) Aus I) und II) folgt  $\tau_{\beta n}^2 \rightarrow \tau_\beta^2$  f.s.

BEWEIS: Auch dieser Beweis ist sehr technisch und wurde deshalb in Abschnitt 1.5 ausgelagert, wo er ab Seite 63 zu finden ist. Das Beweisprinzip besteht darin, daß die in Lemma 7 hergeleiteten Zerlegungen der Bestandteile der empirischen Sinus/Cosinus-Transformierten in die Darstellungen der Terme  $K_{\beta 1n}^1, K_{\beta 2n}^1$ , etc. eingesetzt werden und daß die fast sichere Konvergenz der Restterme gegen 0 gezeigt wird. Auf die aus Summen von unabhängigen und identisch verteilten Zufallsvariablen, -vektoren oder -matrizen bestehenden Hauptteile können das Starke Gesetz großer Zahlen bzw. Theorem D.2 auf Seite 181 angewandt werden.  $\square$

Die in Definition 9 angegebene Integraldarstellung des Schätzers  $\tau_{\beta n}^2$  für  $\tau_\beta^2$  ist für statistische Anwendungen in der Praxis nicht geeignet, da dort eine möglichst einfache und schnelle Berechnung der Schätzwerte anhand vorliegender Daten benötigt wird. Eine numerische Berechnung der in den Bestandteilen  $K_{\beta 1n}^1, K_{\beta 2n}^1, \dots, K_{\beta 2n}^9$  auftretenden Integrale ist in dem vorliegenden Kontext zwar prinzipiell möglich, aber nicht praktikabel. Deshalb leiten wir in dem folgenden Satz elementare Darstellungen dieser Bestandteile als Funktionen der standardisierten Daten her, die eine Berechnung von  $\tau_{\beta n}^2$  mit tolerierbarem numerischen Aufwand auch in höheren Dimensionen zulassen.

**Satz 11** Gemäß der Definition in Lemma 18 auf Seite 57 seien

$$\begin{aligned} I_1(Y_j) &:= \frac{1}{(1 + \beta^2)^{d/2}} e^{-\frac{\beta^2}{2(1+\beta^2)} \|Y_j\|^2}, \\ I_2(Y_j, Y_k) &:= \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|Y_j - Y_k\|^2} + e^{-\frac{\beta^2}{2} \|Y_j + Y_k\|^2} \right], \\ I_3(Y_j, Y_k) &:= \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|Y_j - Y_k\|^2} - e^{-\frac{\beta^2}{2} \|Y_j + Y_k\|^2} \right], \\ I'_4(Y_j, Y_k) &:= \frac{\beta^2}{2} \left[ (Y_k - Y_j) e^{-\frac{\beta^2}{2} \|Y_k - Y_j\|^2} + (Y_k + Y_j) e^{-\frac{\beta^2}{2} \|Y_k + Y_j\|^2} \right], \\ I'_5(Y_j) &:= Y_j \frac{\beta^2}{(1 + \beta^2)^{\frac{d}{2}+1}} e^{-\frac{\beta^2}{2(1+\beta^2)} \|Y_j\|^2}. \end{aligned}$$

Man beachte, daß  $I_1, I_2$  bzw.  $I_3$  Skalare und  $I'_4$  bzw.  $I'_5$   $d$ -dimensionale Vektoren sind. Im Gegensatz zu  $I_2$  und  $I_3$  ist  $I'_4$  nicht symmetrisch in seinen beiden Argumenten. Damit gelten

$$\begin{aligned} K_{\beta 1n}^1 &= \frac{1}{n} \sum_{j=1}^n \left\{ \left( \frac{1}{n} \sum_{k=1}^n I_2(Y_j, Y_k) - I_1(Y_j) \right)^2 + \left( \frac{1}{n} \sum_{k=1}^n I_3(Y_j, Y_k) \right)^2 \right\}, \\ K_{\beta 2n}^1 &= \frac{2}{n} \sum_{j=1}^n \left\{ \frac{1}{n} \sum_{k=1}^n I_3(Y_j, Y_k) \right\} \left\{ \frac{1}{n} \sum_{k=1}^n I_2(Y_j, Y_k) - I_1(Y_j) \right\}, \\ K_{\beta 1n}^2 &= \frac{1}{n^2} \sum_{j,k=1}^n I_2(Y_j, Y_k) - \frac{1}{n} \sum_{j=1}^n I_1(Y_j), \end{aligned}$$

$$\begin{aligned}
K_{\beta 2n}^2 &= \frac{1}{n^2} \sum_{j,k=1}^n I_3(Y_j, Y_k), \\
K_{\beta 1n}^3 &= -\frac{1}{n^2} \sum_{j,k=1}^n Y_j^T I_4'(Y_k, Y_j) + \frac{1}{n} \sum_{j=1}^n Y_j^T I_5'(Y_j), \\
K_{\beta 2n}^3 &= \frac{1}{n^2} \sum_{j,k=1}^n Y_j^T I_4'(Y_j, Y_k), \\
K_{\beta 1n}^{4'} &= -\frac{1}{n^2} \sum_{j,k=1}^n I_3(Y_j, Y_k) Y_j, \\
K_{\beta 2n}^{4'} &= \frac{1}{n^2} \sum_{j,k=1}^n I_2(Y_j, Y_k) Y_j - \frac{1}{n} \sum_{j=1}^n I_1(Y_j) Y_j, \\
K_{\beta 1n}^{5'} &= \frac{1}{n^2} \sum_{j,k=1}^n I_4'(Y_j, Y_k), \\
K_{\beta 2n}^{5'} &= \frac{1}{n^2} \sum_{j,k=1}^n I_4'(Y_k, Y_j) - \frac{1}{n} \sum_{j=1}^n I_5'(Y_j), \\
K_{\beta 1n}^{6'} &= \frac{1}{n^2} \sum_{j,k=1}^n (Y_k^T Y_j) (Y_k^T I_5'(Y_j)) Y_k - \frac{1}{n^3} \sum_{i,j,k=1}^n (Y_k^T Y_j) (Y_k^T I_4'(Y_i, Y_j)) Y_k, \\
K_{\beta 2n}^{6'} &= \frac{1}{n^3} \sum_{i,j,k=1}^n (Y_k^T Y_j) (Y_k^T I_4'(Y_j, Y_i)) Y_k, \\
K_{\beta 1n}^{7''} &= -\frac{1}{n^2} \sum_{j,k=1}^n I_2(Y_j, Y_k) Y_j Y_j^T + \frac{1}{n} \sum_{j=1}^n I_1(Y_j) Y_j Y_j^T, \\
K_{\beta 2n}^{7''} &= -\frac{1}{n^2} \sum_{j,k=1}^n I_3(Y_j, Y_k) Y_j Y_j^T, \\
K_{\beta 1n}^{8''} &= \frac{1}{n^2} \sum_{j,k=1}^n (Y_k^T Y_j) (Y_k^T I_5'(Y_j)) Y_k Y_k^T - \frac{1}{n^3} \sum_{i,j,k=1}^n (Y_k^T Y_j) (Y_k^T I_4'(Y_i, Y_j)) Y_k Y_k^T, \\
K_{\beta 2n}^{8''} &= \frac{1}{n^3} \sum_{i,j,k=1}^n (Y_k^T Y_j) (Y_k^T I_4'(Y_j, Y_i)) Y_k Y_k^T, \\
K_{\beta 1n}^9(A) &= \frac{1}{n} \sum_{j=1}^n I_5'(Y_j)^T A Y_j - \frac{1}{n^2} \sum_{j,k=1}^n I_4'(Y_k, Y_j)^T A Y_j, \quad A \in \mathbb{R}^{d \times d}, \\
K_{\beta 2n}^9(A) &= \frac{1}{n^2} \sum_{j,k=1}^n I_4'(Y_j, Y_k)^T A Y_j, \quad A \in \mathbb{R}^{d \times d}.
\end{aligned}$$

BEWEIS: Einsetzen der Definitionen und Anwenden von Lemma 18. □

### 1.3. Anwendungen der Konvergenzaussagen

In den folgenden Abschnitten sollen einige statistische Anwendungen diskutiert werden, die sich aus der in Theorem 5 auf Seite 18 hergeleiteten Konvergenzaussage ergeben.

#### 1.3.1. Asymptotischer Konfidenzbereich für $L_\beta$

Zunächst kann man anhand der bisherigen Betrachtungen einen Vertrauensbereich für den fast sicheren Grenzwert

$$L_\beta = \int_{\mathbb{R}^d} \left| \Psi(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt$$

von  $n^{-1}T_{n,\beta}$  angeben, auf dessen Bedeutung in Abschnitt 1.3.4 näher eingegangen wird.

Nach Theorem 5 gilt für jede Alternative zur Normalverteilung, die  $E \|X_1\|^4 < \infty$  erfüllt,

$$\frac{S_{n,\beta}}{\tau_\beta} = \frac{\sqrt{n}}{\tau_\beta} \left( \frac{1}{n} T_{n,\beta} - L_\beta \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad (n \rightarrow \infty),$$

und nach Kapitel 1.2 bleibt diese Aussage richtig, falls  $\tau_\beta^2$  durch den konsistenten Schätzer  $\tau_{\beta n}^2$  aus Gleichung (1.19) und entsprechend  $\tau_\beta = \sqrt{\tau_\beta^2}$  durch  $\tau_{\beta n} = \sqrt{\tau_{\beta n}^2}$  ersetzt wird:

$$\frac{S_{n,\beta}}{\tau_{\beta n}} = \frac{\sqrt{n}}{\tau_{\beta n}} \left( \frac{1}{n} T_{n,\beta} - L_\beta \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad (n \rightarrow \infty).$$

Bezeichnet  $\Phi^{-1}(\alpha)$ ,  $\alpha \in (0, 1)$ , das  $\alpha$ -Quantil der Standard-Normalverteilung, so folgt

$$\begin{aligned} 1 - \alpha &= \lim_{n \rightarrow \infty} P \left( \left| \frac{\sqrt{n}}{\tau_{\beta n}} \left( \frac{1}{n} T_{n,\beta} - L_\beta \right) \right| \leq \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right) \\ &= \lim_{n \rightarrow \infty} P \left( \frac{1}{n} T_{n,\beta} - \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tau_{\beta n}}{\sqrt{n}} \leq L_\beta \leq \frac{1}{n} T_{n,\beta} + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tau_{\beta n}}{\sqrt{n}} \right). \end{aligned}$$

Somit ist ein asymptotischer  $(1 - \alpha)$ -Konfidenzbereich für  $L_\beta$  gegeben durch

$$I_{\beta,n,1-\alpha} := \left[ \frac{1}{n} T_{n,\beta} - \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tau_{\beta n}}{\sqrt{n}}, \frac{1}{n} T_{n,\beta} + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tau_{\beta n}}{\sqrt{n}} \right]. \quad (1.20)$$

$I_{\beta,n,1-\alpha}$  hat die Länge  $2 \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tau_{\beta n}}{\sqrt{n}}$ . Um diese zu halbieren, muß also der Stichprobenumfang in etwa vervierfacht werden.

#### 1.3.2. Gütenäherung für den BHEP-Test

Zu  $\alpha \in (0, 1)$  bezeichne  $q_{\beta,1-\alpha}$  den kritischen Wert des auf der BHEP-Statistik basierenden asymptotischen Anpassungstests auf multivariate Normalverteilung zum Niveau  $\alpha$ , d.h. das  $(1 - \alpha)$ -Quantil der Grenzverteilung der BHEP-Statistik  $T_{n,\beta}$  unter der Hypothese  $H_0 : P^X \in$

$\mathcal{N}_d$  beim Grenzübergang  $n \rightarrow \infty$ . Der *asymptotische Test* lehnt  $H_0$  nach Konstruktion genau dann ab, wenn der Wert der Teststatistik auf den Daten  $x_1, \dots, x_n$  diesen kritischen Wert überschreitet:

$$\text{Ablehnen von } H_0 \iff T_{n,\beta}(x_1, \dots, x_n) > q_{\beta,1-\alpha}.$$

Da die Quantile  $q_{\beta,1-\alpha}$  zum einen analytisch schwer faßbar sind und da man zum anderen die Diskrepanz zwischen der Verteilung der BHEP–Statistik  $T_{n,\beta}$  bei endlichem Stichprobenumfang  $n$  und ihrer Grenzverteilung nach dem Übergang  $n \rightarrow \infty$  kompensieren möchte, verwendet man zur konkreten Durchführung des Tests anstelle der Grenzquantile  $q_{\beta,1-\alpha}$  oft die durch Monte–Carlo–Simulation der Verteilung von  $T_{n,\beta}$  bei endlichem Stichprobenumfang  $n$  empirisch ermittelten Quantile  $q_{\beta,n,1-\alpha}$ , was auf den zugehörigen *Test bei endlichem Stichprobenumfang* führt:

$$\text{Ablehnen von } H_0 \iff T_{n,\beta}(x_1, \dots, x_n) > q_{\beta,n,1-\alpha}.$$

Obwohl die Quantile  $q_{\beta,1-\alpha}$  und  $q_{\beta,n,1-\alpha}$  nicht explizit ausgerechnet werden können, werden sie nicht nur für die Durchführung des BHEP–Tests benötigt, sondern auch für die im folgenden vorgestellte Gütenäherung. Deshalb begründen wir zunächst, warum man sie dennoch als bekannt und ihre tabellierten Werte als gesichert annehmen kann. Zum einen sind die Werte  $q_{\beta,n,1-\alpha}$  bei endlichem Stichprobenumfang  $n$  in mehreren umfangreichen Monte–Carlo–Simulationsstudien evaluiert worden, deren Ergebnisse übereinstimmen, so z.B. in Baringhaus und Henze (1988), Henze und Zirkler (1990) und Henze und Wagner (1997). Henze (1990) führte im Fall  $d = 1$ ,  $\beta = 1$  und  $\alpha \in \{0.01, 0.025, 0.05, 0.1\}$  nicht nur eine aufwendige Simulationsstudie durch, in der jedes Quantil  $q_{\beta,n,1-\alpha}$  anhand von 200 000 Monte–Carlo–Wiederholungen ermittelt wurde. In dieser Arbeit wurden vielmehr auch die ersten vier Momente der Grenzverteilung der BHEP–Statistik unter  $H_0$  berechnet und diese durch Anpassen der ersten vier Momente durch jeweils einen Vertreter der Johnson–Verteilungsfamilie bzw. der Pearson–Verteilungsfamilie angenähert. Diese Verteilungsfamilien sind eindeutig durch die ersten vier Momente bestimmt und z.B. in Johnson und Kotz (1970) in Kapitel 12 beschrieben. Henze (1990) zeigte, daß für großes  $n$  die anhand der Monte–Carlo–Simulation erhaltenen Quantile  $q_{\beta,n,1-\alpha}$  und die anhand der beiden Anpassungen erhaltenen Näherungen  $q_{\beta,J,1-\alpha}$  (Johnson) bzw.  $q_{\beta,P,1-\alpha}$  (Pearson) für die Quantile  $q_{\beta,1-\alpha}$  der Grenzverteilung von  $T_{n,\beta}$  perfekt übereinstimmen und daß überdies die Monte–Carlo–Quantile mit wachsendem Stichprobenumfang  $n$  erstaunlich schnell gegen diesen Grenzwert streben ( $n \geq 15$  reicht aus). Die hier relevanten Werte ( $\alpha = 0.1$ ) aus dieser Studie sind in Tabelle 1.1 wiedergegeben.

n	4	5	6	7	8	9	10	15	20	30	50	100	200	Johns.	Pears.
$q_{\beta,n,0.9}$	0.249	0.244	0.263	0.266	0.271	0.275	0.279	0.284	0.287	0.288	0.290	0.291	0.290	0.289	0.292

Tabelle 1.1.: 90%–Quantile der BHEP–Statistik nach Henze (1990)

Eine wichtige Problemstellung in der Statistik besteht nun darin, Aussagen über die Güte von Testverfahren zu gewinnen, in unserem Fall also über die Wahrscheinlichkeit  $P_F(T_{n,\beta} > q_{\beta,n,1-\alpha})$ , daß der BHEP–Test bei Vorliegen einer Alternative  $F \in H_1$  diesen Sachverhalt erkennt und die Hypothese ablehnt.

Nach Theorem 5 auf Seite 18 gilt für jede Alternative zur Normalverteilung, die  $E \|X_1\|^4 < \infty$  erfüllt,

$$\frac{S_{n,\beta}}{\tau_\beta} = \frac{T_{n,\beta} - n L_\beta}{\sqrt{n} \tau_\beta} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad (n \rightarrow \infty).$$

Bezeichnet  $\Phi$  die Verteilungsfunktion der Standard-Normalverteilung, so liefert diese Grenzwertaussage die folgende Näherung für die Güte des BHEP-Tests unter der zugrundeliegenden festen Alternative  $F$  :

$$\begin{aligned} P_F(T_{n,\beta} > q_{\beta,n,1-\alpha}) &= 1 - P_F(T_{n,\beta} \leq q_{\beta,n,1-\alpha}) \\ &= 1 - P_F\left(\frac{T_{n,\beta} - n L_\beta}{\sqrt{n} \tau_\beta} \leq \frac{q_{\beta,n,1-\alpha} - n L_\beta}{\sqrt{n} \tau_\beta}\right) \\ &\approx 1 - \Phi\left(\frac{q_{\beta,n,1-\alpha} - n L_\beta}{\sqrt{n} \tau_\beta}\right). \end{aligned} \quad (1.21)$$

Dabei hängen sowohl  $L_\beta$  als auch  $\tau_\beta$  von der Alternative ab und sind a priori unbekannt.

Berechnet man die Näherung (1.21) bei einer *bekannten* festen Alternative mit endlichen vierten Momenten und damit bekannten Konstanten  $L_\beta$  und  $\tau_\beta$ , so erhält man nicht nur eine Kontrolle der bisher nur durch Simulationsstudien ermittelbaren Monte-Carlo-Gütwerte, sondern kann für derartige alternative Verteilungen die Güte des BHEP-Tests ganz ohne Simulationsstudie annähern. Dadurch entfällt insbesondere die Abhängigkeit von der Qualität der Zufallszahlengeneratoren. Zwar müssen die in der Definition von  $\tau_\beta$  auftretenden Integrale berechnet werden; hierfür existieren aber geprüfte und numerisch stabile Integrationsroutinen, die sich bei den untersuchten und in Abschnitt 1.4 vorgestellten Beispielveilungen als sehr zuverlässig und schnell erwiesen haben.

Desweiteren ist die Näherung aus (1.21) erstaunlich gut, wenn für  $L_\beta$  und  $\tau_\beta$  die Schätzer  $n^{-1}T_{n,\beta}$  und  $\tau_{\beta n}^2$  eingesetzt werden; vgl. hierzu die Ergebnisse in Abschnitt 1.4. Man hat also ohne Kenntnis der zugrundeliegenden Alternative eine Vorstellung davon, wie sich der BHEP-Test ihr gegenüber verhält.

### 1.3.3. „Inverser Test“ auf Normalverteilung

Auf dem Raum  $L^2(\mathbb{R}^d, \mathfrak{B}, \varphi_\beta(t) dt)$  komplexwertiger Funktionen  $f : \mathbb{R}^d \rightarrow \mathbb{C}$  mit  $\int |f(t)|^2 \varphi_\beta(t) dt < \infty$ , versehen mit dem Skalarprodukt  $\langle f, g \rangle := \int f(t) \overline{g(t)} \varphi_\beta(t) dt$ , ist ein Abstand zwischen einer beliebigen Verteilung  $F$  und der Standard-Normalverteilung  $\mathcal{N}_d(0, I_d)$  gegeben durch den  $L^2$ -Abstand ihrer charakteristischen Funktionen

$$d_\beta(F, \mathcal{N}_d(0, I_d)) := L_\beta(F)^{1/2} := L_\beta^{1/2} = \left( \int_{\mathbb{R}^d} \left| \Psi(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt \right)^{1/2}$$

mit  $\Psi(t) = E \left[ e^{it^T X_1} \right]$ . Dabei gilt

$$d_\beta(F, \mathcal{N}_d(0, I_d)) = 0 \iff L_\beta(F) = 0 \iff F = \mathcal{N}_d(0, I_d),$$

d.h.  $L_\beta$  charakterisiert die Standard-Normalverteilung. Klassische Anpassungstests testen die einfache Hypothese  $H_0 : F = \mathcal{N}_d(0, I_d)$  gegen  $H_1 : F \neq \mathcal{N}_d(0, I_d)$ , oder durch  $L_\beta$  ausgedrückt  $H_0 : L_\beta = 0$  gegen  $H_1 : L_\beta \neq 0$ . Eine Möglichkeit zur Konstruktion eines solchen Tests beruht auf der BHEP-Statistik

$$T_{n,\beta} = n \int_{\mathbb{R}^d} \left| \Psi_n(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt ,$$

falls  $\Psi_n(t)$  die empirische charakteristische Funktion der Daten bezeichnet. Ist  $\Psi_n(t)$  sogar die empirische charakteristische Funktion der *standardisierten* Daten (was bei unseren Betrachtungen stets der Fall war), so testet die BHEP-Statistik die zusammengesetzte Hypothese  $H_0 : F \in \mathcal{N}_d$  gegen die allgemeine Alternative  $H_1 : F \notin \mathcal{N}_d$ , wobei  $\mathcal{N}_d$  die Klasse der nichtdegenerierten  $d$ -dimensionalen Normalverteilungen bezeichnet. Auch dieser Test beruht auf dem Abstandsmaß  $L_\beta$  und prüft  $H_0 : L_\beta = 0$  gegen  $H_1 : L_\beta \neq 0$ .

Das Problem bei den klassischen Testverfahren besteht darin, daß nach Konstruktion nur der Fehler 1. Art kontrolliert werden kann, d.h. bei Gültigkeit der Hypothese (Vorliegen einer Normalverteilung) wird diese nur mit einer kleinen Wahrscheinlichkeit von höchstens  $\alpha$  fälschlicherweise verworfen. Erhebt der Test hingegen keinen Einwand gegen  $H_0$ , so ist das zu prüfende Modell dadurch keineswegs gesichert.

In der Praxis ist man jedoch genau an dieser Fragestellung interessiert: Man möchte ein Modell validieren, d.h. in einem „inversen“ Testproblem  $H_0 : L_\beta \neq 0$  gegen  $H_1 : L_\beta = 0$  bei einem kontrollierten Fehler 1. Art die Alternative sichern. Dies ist in der angegebenen Form nicht möglich, aber man kann die Alternative durch eine geeignete Umgebung ersetzen, innerhalb derer man Unterschiede als nicht relevant erachtet. Dieser Ansatz der nichtparametrischen Modell-Validierung führt auf das sogenannte „Precise Testing Problem“

$$H_\Delta : L_\beta \geq \Delta \quad \text{gegen} \quad K_\Delta : L_\beta < \Delta$$

mit einer hinreichend kleinen Konstanten  $\Delta > 0$ . Für  $L_\beta < \Delta$  ist man der Meinung, daß die Daten hinreichend genau durch das Normalverteilungsmodell beschrieben werden, und ein zugehöriger Test dient dazu, diese Alternative bei einem kleinen Fehler 1. Art zu sichern. Um den kritischen Wert dieses Tests zu ermitteln, benötigt man Kenntnisse über das Verhalten der zugrundeliegenden Teststatistik unter festen Alternativen. Das entscheidende Problem liegt in der Wahl von  $\Delta$ .

Derartige Fragestellungen werden seit längerem z.B. in der Biometrie untersucht, wo sie unter dem Stichwort „Äquivalenznachweis“ bzw. „Bioequivalence Testing“ (Testen auf Bioäquivalenz) zu finden sind. Die Problemstellung ist dabei nicht auf Anpassungstests beschränkt, sondern läßt sich auf eine Vielzahl statistischer Fragestellungen übertragen, wie z.B. den Vergleich von Mittelwerten oder den Zusammenhang zwischen verschiedenen Kenngrößen in einem Regressionsmodell. Im konkreten Fall müssen geeignete Abstandsmaße für die zu untersuchenden Größen und eine adäquate Umgebung der zu validierenden Alternative gefunden werden, d.h. es ist mit den Anwendern abzuklären, welche Werte dieser Abstandsmaße noch mit dem Modell vereinbar bzw. tolerierbar sind. Da man neben dem Schätzen von Störparametern Grenzwertsätze unter festen Alternativen benötigt, ist die mathematische Behandlung dieses Ansatzes in der Regel schwieriger als die der klassischen Testprobleme. Auf dem Gebiet des Testens auf

Bioäquivalenz sind bereits zahlreiche Arbeiten erschienen, die derartige Fragestellungen in verschiedenen Kontexten untersuchen, z.B. Brown, Hwang und Munk (1997), Dette und Munk (1998) in einem Regressionskontext, Munk und Czado (1998) und Czado und Munk (1998) mit einem Anpassungstest, der auf dem Mallows-Abstand beruht, um nur einige zu nennen.

Im folgenden setzen wir voraus, daß die vierten Momente der zugrundeliegenden Verteilung  $F$  existieren. Aus der nach Theorem 5 auf Seite 18 und Satz 10 auf Seite 25 gültigen Grenzwertaussage

$$\frac{\sqrt{n}}{\tau_{\beta n}} \left( \frac{1}{n} T_{n,\beta} - L_{\beta}(F) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad (n \rightarrow \infty) \quad (1.22)$$

mit dem konsistenten Schätzer  $\tau_{\beta n}^2$  für  $\tau_{\beta}^2$  kann ein Test für  $H_{\Delta}$  gegen  $K_{\Delta}$  hergeleitet werden. Da die affin invariante Teststatistik  $T_{n,\beta}$  auf den standardisierten Daten beruht, wird damit nicht nur die einfache, sondern die zugehörige zusammengesetzte Hypothese behandelt, d.h. es wird geprüft, ob die von  $F$  erzeugte Lokations-Skalen-Familie  $F(\Sigma^{-1/2}(\cdot - \mu))$  mit der Lokations-Skalen-Familie der  $d$ -dimensionalen Normalverteilungen vereinbar ist. Da der resultierende Test die Hypothese bei zu kleinen Werten von  $n^{-1}T_{n,\beta}$  ablehnt, lautet die Bedingung für ein Verfahren zum Niveau  $\alpha$  mit kritischem Wert  $q$

$$P_F \left( \frac{1}{n} T_{n,\beta} \leq q \right) = P_F \left( \frac{\sqrt{n} \left( \frac{1}{n} T_{n,\beta} - L_{\beta}(F) \right)}{\tau_{\beta n}} \leq \frac{\sqrt{n} (q - L_{\beta}(F))}{\tau_{\beta n}} \right) \stackrel{!}{\leq} \alpha, \quad F \in H_{\Delta}.$$

Anhand der Grenzwertaussage (1.22) ergibt sich für den entsprechenden approximativen Test die Bedingung

$$\frac{\sqrt{n} (q - L_{\beta}(F))}{\tau_{\beta n}} \stackrel{!}{=} \Phi^{-1}(\alpha),$$

also für den kritischen Wert

$$q := \frac{\tau_{\beta n}}{\sqrt{n}} \Phi^{-1}(\alpha) + L_{\beta}(F).$$

Wegen  $\inf_{F \in H_{\Delta}} L_{\beta}(F) = \Delta$  definiert man den gesuchten Test durch

$$H_{\Delta} \text{ ablehnen} \iff \frac{1}{n} T_{n,\beta} \leq \frac{\tau_{\beta n}}{\sqrt{n}} \Phi^{-1}(\alpha) + \Delta.$$

Es gilt dann für jedes  $F \in H_{\Delta}$

$$\limsup_{n \rightarrow \infty} P_F („H_{\Delta} \text{ ablehnen}“) \leq \alpha,$$

mit Gleichheit im Fall  $L_{\beta}(F) = \Delta$ . Dieser Test ist allerdings nur in der Klasse der Verteilungen mit existierenden vierten Momenten sinnvoll, da andernfalls die Grenzwertaussage (1.22) nicht gesichert ist. Ein weiteres Problem besteht darin, einen sinnvollen Wert für  $\Delta$  zu bestimmen. Dazu muß das Abstandsmaß  $L_{\beta}$  genauer beleuchtet werden, was Gegenstand des nächsten Abschnittes ist.

### 1.3.4. Betrachtung des Grenzwertes $L_\beta$

Nachdem einige statistische Verfahren vorgeschlagen worden sind, die auf dem Abstand  $L_\beta$  zwischen einer Alternative  $F$  und der Normalverteilung beruhen, soll im folgenden die Größe  $L_\beta$  näher untersucht werden. Diese ist, wie schon in Abschnitt 1.3.3 erwähnt, das Quadrat des  $L^2$ -Abstandes zwischen der charakteristischen Funktion  $\Psi(t) = \mathbb{E} [\exp\{it^T X_1\}]$  der zugrundeliegenden Verteilung  $F$ , von der wir wie immer die Existenz der vierten Momente voraussetzen, und der charakteristischen Funktion  $\Psi_{\mathcal{N}}(t) = \exp\{-\frac{\|t\|^2}{2}\}$  der Standard-Normalverteilung  $\mathcal{N}_d(0, I_d)$ . Da alle betrachteten statistischen Verfahren auf den standardisierten Daten  $Y_j$  beruhen, setzen wir  $F$  o.B.d.A. als standardisiert voraus, d.h. mit  $\mathbb{E}X_1 = 0$  und  $\mathbb{E}X_1 X_1^T = I_d$ .

Da die charakteristische Funktion eine Verteilung eindeutig bestimmt und da nach dem Stetigkeitssatz von Lévy–Cramér die schwache Konvergenz von  $d$ -dimensionalen Verteilungen äquivalent ist zur punktweisen Konvergenz der zugehörigen charakteristischen Funktionen (gegen eine charakteristische Funktion), erscheint die Verwendung von  $L_\beta$  sinnvoll zum Zwecke der Charakterisierung einer Umgebung der Normalverteilung.

#### Obere Schranken für $L_\beta$

Um die Rolle von  $L_\beta$  genauer zu beleuchten, stellen wir dieselben Überlegungen an, anhand derer Henze und Zirkler (1990) den Zusammenhang zwischen der BHEP-Statistik und der Bowman–Foster-Statistik herleiteten (vgl. hierzu die Ausführungen zu Beginn von Kapitel 2 auf Seite 72): Die Fourier-Transformierte  $\hat{g}(t) := \int \exp\{it^T x\} g(x) dx$  einer Funktion  $g \in L^2(\mathbb{R}^d, \mathfrak{B}, dx)$  liegt ebenfalls in  $L^2(\mathbb{R}^d, \mathfrak{B}, dx)$ , und die *Formel von Plancherel* aus der Fourier-Analyse besagt, daß zwischen den beiden  $L^2$ -Normen der folgende Zusammenhang besteht:

$$\int |\hat{g}(t)|^2 dt = (2\pi)^d \int |g(t)|^2 dt$$

(vgl. z.B. Rudin (1987)). Wir nehmen ab jetzt zusätzlich an, daß die (standardisierte) Verteilung  $F$  die Dichte  $f \in L^2(\mathbb{R}^d, \mathfrak{B}, dx)$  besitzt. Zunächst stellt man fest, daß mit der Dichte  $\varphi_\beta(t) = (2\pi\beta^2)^{-d/2} \exp\{-\frac{\|t\|^2}{2\beta^2}\}$  von  $\mathcal{N}_d(0, \beta^2 I_d)$  die Darstellung

$$L_\beta = \int_{\mathbb{R}^d} \left| \Psi(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt = \frac{1}{(2\pi\beta^2)^{d/2}} \int_{\mathbb{R}^d} \left| \Psi(t) e^{-\frac{\|t\|^2}{4\beta^2}} - e^{-\frac{1}{2} \left[1 + \frac{1}{2\beta^2}\right] \|t\|^2} \right|^2 dt$$

gilt. Berücksichtigt man ferner, daß mit  $\sigma^2 := \frac{1}{2\beta^2}$  die Funktion  $\Psi(t) \exp\{-\frac{\|t\|^2}{4\beta^2}\}$  die Fourier-Transformierte der Faltung  $F * \mathcal{N}_d(0, \sigma^2 I_d)$  ist und daß mit  $\tau^2 := 1 + \frac{1}{2\beta^2} = 1 + \sigma^2$  die Funktion  $\exp\left\{-\frac{1}{2} \left[1 + \frac{1}{2\beta^2}\right] \|t\|^2\right\}$  die Fourier-Transformierte der Normalverteilung  $\mathcal{N}_d(0, \tau^2 I_d) = \mathcal{N}_d(0, I_d) * \mathcal{N}_d(0, \sigma^2 I_d)$  ist, so liefert die Formel von Plancherel

$$L_\beta = \frac{(2\pi)^{d/2}}{\beta^d} \int_{\mathbb{R}^d} |f * \varphi_\sigma(t) - \varphi_1 * \varphi_\sigma(t)|^2 dt = \frac{(2\pi)^{d/2}}{\beta^d} \int_{\mathbb{R}^d} |(f - \varphi_1) * \varphi_\sigma(t)|^2 dt .$$

Nach Aufgabe 4, S. 174 aus Rudin (1987) gilt wegen  $f \in L^2(\mathbb{R}^d, \mathfrak{B}, dx)$  und  $\varphi_1, \varphi_\sigma \in L^1(\mathbb{R}^d, \mathfrak{B}, dx) \cap L^2(\mathbb{R}^d, \mathfrak{B}, dx)$  die Abschätzung

$$L_\beta \leq \frac{(2\pi)^{d/2}}{\beta^d} \left( \int_{\mathbb{R}^d} |\varphi_\sigma(t)| dt \right)^2 \int_{\mathbb{R}^d} |f(t) - \varphi_1(t)|^2 dt = \frac{(2\pi)^{d/2}}{\beta^d} \int_{\mathbb{R}^d} |f(t) - \varphi_1(t)|^2 dt.$$

$L_\beta$  läßt sich also durch den quadratischen Abstand der zugrundeliegenden (standardisierten) Dichte  $f$  von der Dichte der  $d$ -dimensionalen Standard-Normalverteilung in  $L^2(\mathbb{R}^d, \mathfrak{B}, dx)$  abschätzen. Außerdem liegt  $L_\beta$  als Funktion von  $\beta$  unterhalb der Hyperbel  $\frac{\text{const}}{\beta^d}$ , was durch die in Abbildung 1.2 auf Seite 50 festgehaltenen empirischen Ergebnisse bestätigt wird.

Da die Dichte  $f$  auch in  $L^1(\mathbb{R}^d, \mathfrak{B}, dx)$  liegt, kann man in Aufgabe 4, S. 174 aus Rudin (1987) die Rollen von  $f - \varphi_1$  und  $\varphi_\sigma$  vertauschen und gelangt zu der Abschätzung

$$L_\beta \leq \frac{(2\pi)^{d/2}}{\beta^d} \left( \int_{\mathbb{R}^d} |f(t) - \varphi_1(t)| dt \right)^2 \int_{\mathbb{R}^d} |\varphi_\sigma(t)|^2 dt = \left( \int_{\mathbb{R}^d} |f(t) - \varphi_1(t)| dt \right)^2,$$

denn es ist

$$\int_{\mathbb{R}^d} \varphi_\sigma(t)^2 dt = \frac{1}{(4\pi\sigma^2)^{d/2}} = \frac{\beta^d}{(2\pi)^{d/2}}.$$

Der quadratische  $L^1$ -Abstand der zugrundeliegenden Dichte  $f$  von der Dichte der  $d$ -dimensionalen Standard-Normalverteilung ist also eine von  $\beta$  unabhängige obere Schranke für  $L_\beta$ . Wir haben damit den folgenden Satz gezeigt:

**Satz 12** Besitzt die zugrundeliegende standardisierte Verteilung  $F$  endliche Momente vierter Ordnung und eine Dichte  $f \in L^2(\mathbb{R}^d, \mathfrak{B}, dx)$ , so gelten für  $L_\beta$  die Abschätzungen

$$L_\beta \leq \frac{(2\pi)^{d/2}}{\beta^d} \int_{\mathbb{R}^d} |f(t) - \varphi_1(t)|^2 dt, \quad L_\beta \leq \left( \int_{\mathbb{R}^d} |f(t) - \varphi_1(t)| dt \right)^2.$$

### Taylor-Entwicklung von $L_\beta$ für $\beta \rightarrow 0$

Weiteren Aufschluß über das Verhalten von  $L_\beta$  beim Grenzübergang  $\beta \rightarrow 0$  gibt eine Taylor-Entwicklung dieser Größe. Bei den gesamten nachfolgenden Ausführungen setzen wir stets  $EX_1 = 0$  und  $EX_1X_1^T = I_d$  voraus. Zusätzlich benötigen wir die Existenz von Momenten sechster Ordnung, d.h.  $E \|X_1\|^6 < \infty$ .

Ausgangspunkt der Betrachtungen ist die Beobachtung von Henze (1997), daß die BHEP-Statistik

$$T_{n,\beta} = \frac{1}{n} \sum_{j,k=1}^n e^{-\frac{\beta^2}{2} \|Y_j - Y_k\|^2} - \frac{2}{(1 + \beta^2)^{d/2}} \sum_{j=1}^n e^{-\frac{\beta^2 \|Y_j\|^2}{2(1+\beta^2)}} + \frac{n}{(1 + 2\beta^2)^{d/2}} \quad (1.23)$$

(vgl. (1.2) auf Seite 6) für festes  $n$  punktweise in  $\omega \in \Omega$  beim Grenzübergang  $\beta \rightarrow 0$  die Darstellung

$$\frac{1}{n} T_{n,\beta} = \beta^6 \left\{ \frac{1}{6} b_{1,d} + \frac{1}{4} \tilde{b}_{1,d} \right\} + O(\beta^8)$$

besitzt, welche man durch direkte Entwicklung der Brüche und Exponentialterme nach  $\beta$  beweist. Dabei ist  $b_{1,d}$  die auf den standardisierten Daten  $Y_j$  beruhende *empirische multivariate Schiefe von Mardia (1970)*, der die Kenngröße  $\beta_{1,d}$  der zugrundeliegenden Verteilung entspricht.  $\tilde{b}_{1,d}$  ist ein von Henze (1997) betrachtetes, ebenfalls auf den standardisierten Daten  $Y_j$  beruhendes empirisches Analogon der von Móri, Rohatgi und Székely (1993) eingeführten Schiefe  $\tilde{\beta}_{1,d}$ . Diese vier Größen sind definiert durch

$$\begin{aligned}\beta_{1,d} &:= \mathbb{E} \left[ (X_1^T X_2)^3 \right], & b_{1,d} &:= \frac{1}{n^2} \sum_{j,k=1}^n (Y_j^T Y_k)^3, \\ \tilde{\beta}_{1,d} &:= \left\| \mathbb{E} \left[ \|X_1\|^2 X_1 \right] \right\|^2, & \tilde{b}_{1,d} &:= \frac{1}{n^2} \sum_{j,k=1}^n Y_j^T Y_k \|Y_j\|^2 \|Y_k\|^2.\end{aligned}$$

Durch Einsetzen der Definition von  $\Psi$ , Anwenden des Satzes von Fubini und von Lemma 18 auf Seite 57 leitet man die (1.23) entsprechende Darstellung von  $L_\beta$  her:

$$\begin{aligned}L_\beta &= \int_{\mathbb{R}^d} \left| \Psi(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt \\ &= \mathbb{E} \left[ e^{-\frac{\beta^2}{2} \|X_1 - X_2\|^2} \right] - \frac{2}{(1 + \beta^2)^{d/2}} \mathbb{E} \left[ e^{-\frac{\beta^2}{2(1+\beta^2)} \|X_1\|^2} \right] + \frac{1}{(1 + 2\beta^2)^{d/2}}.\end{aligned}\quad (1.24)$$

Wie in Henze (1997) überzeugt man sich leicht von den Entwicklungen

$$\begin{aligned}\frac{2}{(1 + \beta^2)^{d/2}} &= 2 - d\beta^2 + \frac{d(d+2)}{4} \beta^4 - \frac{d(d+2)(d+4)}{24} \beta^6 + O(\beta^8), \\ \frac{1}{(1 + 2\beta^2)^{d/2}} &= 1 - d\beta^2 + \frac{d(d+2)}{2} \beta^4 - \frac{d(d+2)(d+4)}{6} \beta^6 + O(\beta^8)\end{aligned}$$

der in (1.24) auftretenden Brüche, von den Entwicklungen

$$\begin{aligned}e^{-\frac{\beta^2}{2} \|X_1 - X_2\|^2} &= 1 - \frac{\beta^2}{2} \|X_1 - X_2\|^2 + \frac{\beta^4}{8} \|X_1 - X_2\|^4 - \frac{\beta^6}{48} \|X_1 - X_2\|^6 + O(\beta^8), \\ e^{-\frac{\beta^2}{2(1+\beta^2)} \|X_1\|^2} &= 1 - \frac{\beta^2}{2} \|X_1\|^2 + \beta^4 \left( \frac{1}{2} \|X_1\|^2 + \frac{1}{8} \|X_1\|^4 \right) \\ &\quad + \beta^6 \left( -\frac{1}{2} \|X_1\|^2 - \frac{1}{4} \|X_1\|^4 - \frac{1}{48} \|X_1\|^6 \right) + O(\beta^8)\end{aligned}$$

der Exponentialterme und von den Beziehungen  $\mathbb{E} \|X_1\|^2 = d$ ,  $\mathbb{E} \|X_1 - X_2\|^2 = 2d$  sowie

$$\begin{aligned}\mathbb{E} \|X_1 - X_2\|^4 &= 2\mathbb{E} \|X_1\|^4 + 2d(d+2), \\ \mathbb{E} \|X_1 - X_2\|^6 &= -8\beta_{1,d} - 12\tilde{\beta}_{1,d} + 6(d+4)\mathbb{E} \|X_1\|^4 + 2\mathbb{E} \|X_1\|^6\end{aligned}$$

zwischen den Momenten der zugrundeliegenden Verteilung. Man beachte, daß an dieser Stelle die eingangs getroffenen Annahmen über die Standardisierung und Existenz von Momenten von  $X_1$  eingehen. Setzt man diese Darstellungen in (1.24) ein, so ergibt sich analog zu Henze (1997) die folgende Aussage.

**Satz 13** Besitzt die zugrundeliegende standardisierte Verteilung  $F$  endliche Momente sechster Ordnung, so gilt für  $L_\beta$  die Entwicklung

$$\begin{aligned} L_\beta &= \beta^6 \left\{ \frac{1}{6} \beta_{1,d} + \frac{1}{4} \tilde{\beta}_{1,d} \right\} + O(\beta^8) \\ &= \beta^6 \left\{ \frac{1}{6} \mathbb{E} \left[ (X_1^T X_2)^3 \right] + \frac{1}{4} \left\| \mathbb{E} \left[ \|X_1\|^2 X_1 \right] \right\|^2 \right\} + O(\beta^8) \quad (\beta \rightarrow 0). \end{aligned} \quad (1.25)$$

Der führende Term von  $L_\beta$  beim Grenzübergang  $\beta \rightarrow 0$  ist also eine gewichtete Summe der Schiefemaße von Mardia und Móri et al., die unter  $\mathcal{N}_d(0, I_d)$  beide verschwinden.

**Bemerkung 14** Im Eindimensionalen sind  $\beta_{1,d}$  und  $\tilde{\beta}_{1,d}$  gleich, und die Entwicklung (1.25) vereinfacht sich zu

$$L_\beta = \frac{5}{12} (\mathbb{E}X_1^3)^2 \beta^6 + \frac{35}{24} \left\{ \frac{1}{8} (\mathbb{E}X_1^4 - 3)^2 - \frac{\mathbb{E}X_1^3 \mathbb{E}X_1^5}{5} \right\} \beta^8 + O(\beta^{10}),$$

wobei man den Koeffizienten von  $\beta^8$  leicht durch direkte Rechnung analog zu den obigen erhält. Der führende Term in  $\beta$  vergleicht also die Schiefe der zugrundeliegenden Verteilung mit der Schiefe 0 der Standard-Normalverteilung.

Ist die zugrundeliegende eindimensionale Verteilung sogar symmetrisch um den Ursprung (d.h. es gilt  $X_1 \sim -X_1$ ), so ergibt sich

$$L_\beta = \frac{35}{192} (\mathbb{E}X_1^4 - 3)^2 \beta^8 + O(\beta^{10}),$$

d.h. der führende Term in  $\beta$  vergleicht die Wölbung der zugrundeliegenden Verteilung mit der Wölbung 3 der Standard-Normalverteilung.

### 1.3.5. Übertragbarkeit auf andere Modelle

Analog zum BHEP-Test für die Modellfamilie der  $d$ -dimensionalen Normalverteilungen existieren eine Reihe von Anpassungstests für andere Modelle, die ebenfalls auf  $L^2$ -Abständen von charakterisierenden Merkmalen einer Verteilung beruhen. Es stellt sich die Frage, ob die im Kontext der BHEP-Statistik bewiesenen Konvergenzsätze unter festen Alternativen und die daraus hergeleiteten statistischen Anwendungen auf andere Fälle übertragbar sind. Dies kann bejaht werden, was am Beispiel der (eindimensionalen) Exponentialverteilung kurz skizziert werden soll.

#### Tests auf Exponentialverteilung

Wir gehen im folgenden von eindimensionalen, positiven, unabhängigen und identisch verteilten Zufallsvariablen  $X_1, \dots, X_n$  aus, deren Laplace-Transformierte mit  $\Psi(t) := \mathbb{E}[\exp\{-tX_1\}]$  bezeichnet werde. Ferner bezeichne  $\mathcal{E} := \{\text{Exp}(\lambda) | \lambda > 0\}$  die Skalenfamilie der Exponentialverteilungen. Zum Testen der Hypothese

$$H_0 : P^{X_1} \in \mathcal{E} \quad \text{gegen} \quad H_1 : P^{X_1} \notin \mathcal{E}$$

schlug Henze (1993) die aus dem  $L^2$ -Abstand der empirischen Laplace-Transformierten  $\Psi_n(t) := n^{-1} \sum_{j=1}^n \exp\{-tY_j\}$  der *skalierten* Daten  $Y_j := X_j/\bar{X}_n$  und der Laplace-Transformierten  $1/(1+t)$  der  $\text{Exp}(1)$ -Verteilung gebildete Statistik

$$\tilde{T}_{n,a} := n \int_0^\infty \left( \Psi_n(t) - \frac{1}{1+t} \right)^2 e^{-at} dt, \quad a > 0,$$

vor. Ähnlich wie dem Parameter  $\beta$  im Falle der BHEP-Statistik kommt  $a > 0$  die Rolle eines Gewichtsparameters zu, der dieser Klasse von Statistiken zusätzliche Flexibilität verleiht. Auch hier liegt die Vorstellung zugrunde, daß bei Gültigkeit der Hypothese die empirische Laplace-Transformierte „nahe“ an der Laplace-Transformierten der  $\text{Exp}(1)$ -Verteilung liegt und daß die in den  $Y_j$  eingebauten Schätzer nichts an diesem Sachverhalt ändern. Dementsprechend besitzt der auf  $\tilde{T}_{n,a}$  basierende Test einen oberen Ablehnbereich. Da diese Statistik auf den skalierten Daten  $Y_j$  beruht, ist sie skaleninvariant, und wir können sowohl unter der Hypothese als auch unter festen Alternativen o.B.d.A.  $EX_1 = 1$  annehmen.

Verbinden wir die Überlegungen zum Verhalten der BHEP-Statistik unter festen Alternativen mit dem von Henze (1993) geführten Nachweis der  $H_0$ -Konvergenz von  $\tilde{T}_{n,a}$ , läßt sich leicht das Verhalten dieser Statistik unter festen Alternativen herleiten: Hierzu nehmen wir im folgenden an, daß  $X_1, \dots, X_n$  unabhängige und identisch verteilte Realisierungen einer Verteilung  $F$  auf  $\mathbb{R}_{>0}$  seien, für die zusätzlich gelte:  $EX_1 = 1$ ,  $\sigma^2 := \text{Var}(X_1) < \infty$  und  $E[X_1^{-\varepsilon}] < \infty$  für ein  $\varepsilon \in (0, 1)$ . Im Rahmen des Konsistenznachweises für den auf  $\tilde{T}_{n,a}$  beruhenden Test zeigte Henze (1993) in Theorem 3.1 die stochastische Konvergenz

$$\frac{1}{n} \tilde{T}_{n,a} \xrightarrow{P} \tilde{L}_a := \int_0^\infty \left( \Psi(t) - \frac{1}{1+t} \right)^2 e^{-at} dt$$

unter  $F$ . Analog zu Gleichung (1.3) auf Seite 8 im BHEP-Fall liegt es nahe, einen Konvergenznachweis der Form

$$\tilde{S}_{n,a} := \sqrt{n} \left( \frac{1}{n} \tilde{T}_{n,a} - \tilde{L}_a \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\tau}_a^2)$$

anzustreben. Dies geschieht am besten anhand der Zerlegung

$$\begin{aligned} \tilde{S}_{n,a} &= 2 \int_0^\infty \tilde{Z}_n(e^{-t}) \left[ \Psi(t) - \frac{1}{1+t} \right] e^{-at} dt + \frac{1}{\sqrt{n}} \int_0^\infty \tilde{Z}_n^2(e^{-t}) e^{-at} dt \\ &= 2 \int_0^1 \tilde{Z}_n(u) \left[ \Psi(-\log u) - \frac{1}{1-\log u} \right] u^{a-1} du + \frac{1}{\sqrt{n}} \int_0^1 \tilde{Z}_n^2(u) u^{a-1} du \end{aligned}$$

mit

$$\tilde{Z}_n(e^{-t}) := \frac{1}{\sqrt{n}} \sum_{j=1}^n (e^{-tY_j} - \Psi(t)) = \frac{1}{\sqrt{n}} \sum_{j=1}^n (u^{Y_j} - \Psi(-\log u)) = \tilde{Z}_n(u).$$

Dabei ist der Übergang zwischen  $t \in (0, \infty)$  und  $u \in [0, 1]$  durch die Transformation  $t = -\log u$  geregelt.  $\tilde{Z}_n$  ist ein Zufallselement mit Werten in dem Raum  $\mathcal{C}[0, 1]$  der stetigen Funktionen  $[0, 1] \rightarrow \mathbb{R}$ , dessen schwache Konvergenz in diesem Raum sich analog zu

Henze (1993) durch Anwenden von Corollar 7.17 aus Araujo und Giné (1980) und „Herausrechnen“ des in den Daten enthaltenen Schätzers anhand von Taylorentwicklungen zeigen läßt. Der Grenzwert ist ein zentrierter Gaußprozeß  $\tilde{Z}$  in  $\mathcal{C}[0, 1]$  mit fast sicher stetigen Pfaden, der durch die Kovarianzfunktion

$$\tilde{k}(e^{-s}, e^{-t}) := \Psi(s+t) - \Psi(s)\Psi(t) + \Psi'(s)\Psi'(t) [\sigma^2 st + s+t] + t\Psi'(t)\Psi(s) + s\Psi'(s)\Psi(t)$$

charakterisiert ist. Nach dem Abbildungssatz folgt direkt die Verteilungskonvergenz

$$\tilde{S}_{n,a} \xrightarrow{\mathcal{D}} \tilde{S}_a := 2 \int_0^\infty \tilde{Z}(e^{-t}) \left[ \Psi(t) - \frac{1}{1+t} \right] e^{-at} dt \sim \mathcal{N}(0, \tilde{\tau}_a^2)$$

mit

$$\begin{aligned} \tilde{\tau}_a^2 = & 4 \left\{ \int_0^\infty \int_0^\infty \Psi(s+t) \left[ \Psi(s) - \frac{1}{1+s} \right] \left[ \Psi(t) - \frac{1}{1+t} \right] e^{-a(s+t)} ds dt \right. \\ & - \left( \int_0^\infty \Psi(s) \left[ \Psi(s) - \frac{1}{1+s} \right] e^{-as} ds \right)^2 + \sigma^2 \left( \int_0^\infty s\Psi'(s) \left[ \Psi(s) - \frac{1}{1+s} \right] e^{-as} ds \right)^2 \\ & \left. + 2 \left( \int_0^\infty s\Psi'(s) \left[ \Psi(s) - \frac{1}{1+s} \right] e^{-as} ds \right) \left( \int_0^\infty (\Psi'(t) + \Psi(t)) \left[ \Psi(t) - \frac{1}{1+t} \right] e^{-at} dt \right) \right\}. \end{aligned}$$

Dabei ist  $\Psi'(t) := -E[X_1 e^{-tX_1}]$  die Ableitung der Laplace-Transformierten  $\Psi$  nach  $t$ .

Wie im BHEP-Fall kann man die Grenzvarianz  $\tilde{\tau}_a^2$  in ihre Bestandteile zerlegen und Schätzer für diese definieren. Das geschieht wiederum durch Ersetzen der Laplace-Transformierten  $\Psi(t)$  sowie ihrer Ableitung  $\Psi'(t)$  durch die auf den standardisierten Daten  $Y_j$  beruhende empirische Laplace-Transformierte  $\Psi_n(t)$  sowie deren Ableitung nach  $t$ . Ähnlich wie in Lemma 7 auf Seite 23 und Satz 10 auf Seite 25 weist man die Konsistenz dieser Schätzer nach und gelangt durch Zusammensetzen der Einzelteile zu einem konsistenten Schätzer  $\tilde{\tau}_{an}^2$  für  $\tilde{\tau}_a^2$ .

Nun stehen alle theoretischen Hilfsmittel bereit, um die statistischen Anwendungen aus den Abschnitten 1.3.1, 1.3.2 und 1.3.3 auf den Test auf Exponentialverteilung zu übertragen. Wie in Abschnitt 1.3.1 erhält man ein asymptotisches  $(1 - \alpha)$ -Konfidenzintervall

$$\tilde{I}_{a,n,1-\alpha} := \left[ \frac{1}{n} \tilde{T}_{n,a} - \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tilde{\tau}_{an}}{\sqrt{n}}, \frac{1}{n} \tilde{T}_{n,a} + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tilde{\tau}_{an}}{\sqrt{n}} \right]$$

der Breite  $2 \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tilde{\tau}_{an}}{\sqrt{n}}$  für  $\tilde{L}_a$ . Bezeichnet  $\tilde{q}_{a,n,1-\alpha}$  das durch Monte-Carlo-Simulation der Verteilung von  $\tilde{T}_{n,a}$  unter der Hypothese bei endlichem Stichprobenumfang  $n$  gewonnene  $(1 - \alpha)$ -Quantil der Statistik  $\tilde{T}_{n,a}$ , so ergibt sich analog zu (1.21) auf Seite 30 die Näherung

$$P_F \left( \tilde{T}_{n,a} > \tilde{q}_{a,n,1-\alpha} \right) \approx 1 - \Phi \left( \frac{\tilde{q}_{a,n,1-\alpha} - n \tilde{L}_a}{\sqrt{n} \tilde{\tau}_a} \right)$$

für die Güte des zugehörigen Tests unter einer festen Alternative  $F$ . Auch in dem Kontext der Exponentialverteilung muß über die Zuverlässigkeit dieser Quantile diskutiert werden (vgl. die Ausführungen in Abschnitt 1.3.2). Analog zu Abschnitt 1.3.3 läßt sich ein „inverser“ Test auf Exponentialverteilung angeben.

### Tests auf Poisson–Verteilung

Ein dritter Kontext, in dem die bisher erläuterten Betrachtungen und Folgerungen interessant sein könnten, ist der diskreter Modelle, z.B. der Poisson–Verteilung. Eine Verteilung auf  $\mathbb{N}_0$  ist charakterisiert durch ihre erzeugende Funktion  $g(u) := E[u^X] = \sum_{k=0}^{\infty} P(X = k)u^k$ ,  $|u| \leq 1$ . Es seien  $\mathcal{P} := \{P_o(\vartheta) | \vartheta > 0\}$  die Verteilungsfamilie der Poisson–Verteilungen und  $g(u; \vartheta) := \sum_{k=0}^{\infty} P_o(\vartheta)(X = k)u^k = \exp(\vartheta(u - 1))$  die erzeugende Funktion von  $P_o(\vartheta)$ . Rueda, Pérez-Abreu und O’Reilly (1991) schlugen einen Anpassungstest für die Hypothese

$$H_0 : P^X \in \mathcal{P} \quad \text{gegen} \quad H_1 : P^X \notin \mathcal{P}$$

vor, der auf dem  $L^2$ –Abstand zwischen der auf  $n$  unabhängigen und identisch verteilten Realisierungen  $X_1, \dots, X_n$  von  $X$  basierenden empirischen erzeugenden Funktion  $g_n(u) := n^{-1} \sum_{j=1}^n u^{X_j}$  und der mit dem Maximum–Likelihood–Schätzer  $\hat{\vartheta}_n := n^{-1} \sum_{j=1}^n X_j$  für den Parameter  $\vartheta$  einer Poisson–Verteilung versehenen erzeugenden Funktion  $g(u; \hat{\vartheta}_n)$  unter dem Poisson–Modell beruht:

$$\tilde{T}_n := n \int_0^1 \left( g_n(u) - g(u; \hat{\vartheta}_n) \right)^2 du .$$

Baringhaus, Gürtler und Henze (2000) führten mit dem Gewichtsparameter  $a \geq 0$  zusätzliche Flexibilität ein, was in der Statistik

$$\tilde{T}_{n,a} := n \int_0^1 \left( g_n(u) - g(u; \hat{\vartheta}_n) \right)^2 u^a du$$

resultierte. Hierbei ist zu beachten, daß  $\tilde{T}_{n,a}$  im Gegensatz zu den Statistiken für die stetigen Verteilungsmodelle unter  $H_0$  nicht mehr verteilungsfrei ist. Insbesondere benötigt man bei der praktischen Durchführung des Tests ein parametrisches Bootstrap–Verfahren zur Bestimmung des kritischen Wertes. Dies bedeutet eine zusätzliche Schwierigkeit bei der theoretischen Untersuchung des Tests, denn nun treten beim Nachweis der  $H_0$ –Konvergenz der Statistik  $\tilde{T}_{n,a}$  Dreiecksschemata von Zufallsvariablen auf. Die  $H_0$ –Konvergenz von  $\tilde{T}_{n,a}$  wurde in Gürtler und Henze (1999) bewiesen, wo auch im Rahmen des Konsistenznachweises des zugehörigen Tests gezeigt wurde, daß unter einer festen Alternative mit endlichem Erwartungswert  $EX$  gilt:

$$\frac{1}{n} \tilde{T}_{n,a} \xrightarrow{\text{f.s.}} \tilde{L}_a := \int_0^1 \left( E[u^X] - g(u; EX) \right)^2 u^a du .$$

Ähnlich wie im Fall der bisher betrachteten stetigen Modelle kann man wieder die Differenz  $\sqrt{n}(n^{-1}\tilde{T}_{n,a} - \tilde{L}_a)$  betrachten, eine Aussage bezüglich der Verteilungskonvergenz nachweisen und daraus statistische Schlußfolgerungen ziehen. Dies könnte Gegenstand weiterer Forschung sein, soll aber hier nicht näher ausgeführt werden.

**Bemerkung 15** Angesichts der parallel verlaufenden Untersuchungen in den drei vorgestellten Fällen Normal–, Exponential– und Poisson–Verteilung erscheint es wünschenswert, einen gemeinsamen Rahmen und eine einheitliche Behandlung zu finden. Darauf wurde in dieser Arbeit jedoch verzichtet, da dadurch wahrscheinlich kein struktureller Gewinn zu erwarten ist und da sich aufgrund der unterschiedlichen Definitionsbereiche der betrachteten Größen und der abweichenden Methoden erhebliche notationstechnische Schwierigkeiten ergeben hätten.

## 1.4. Simulationen an Beispielveilteilungen

In diesem Abschnitt sollen die in Kapitel 1.3 hergeleiteten statistischen Aussagen über den BHEP-Test auf Normalverteilung an sechs alternativen Beispielveilteilungen veranschaulicht werden. Für diese werden der asymptotische Konfidenzbereich für  $L_\beta$  und verschiedene Varianten der Gütenäherung für den BHEP-Test berechnet und in Tabellen bzw. als Graphiken dargestellt. Auf den in Abschnitt 1.3.3 auf Seite 30 vorgeschlagenen inversen Test wird nur kurz eingegangen. Aus Gründen der Anschaulichkeit beschränken wir uns auf den Fall  $d = 1$ . Dies hat außerdem den Vorteil, daß sich die Darstellung der Grenzvarianz  $\tau_\beta^2$  und damit auch der zugehörigen Schätzer vereinfacht, wie in den nachfolgenden Bemerkungen 16 und 17 erläutert wird.

**Bemerkung 16** Im Fall  $d = 1$  gelten mit den Bezeichnungen von Seite 18

$$\begin{aligned} K_{\beta 1}^{6'} &= EX_1^3 \cdot K_{\beta 1}^3, & K_{\beta 1}^{8''} &= EX_1^4 \cdot K_{\beta 1}^3, & K_{\beta 1}^9(A) &= A \cdot K_{\beta 1}^3, \\ K_{\beta 2}^{6'} &= EX_1^3 \cdot K_{\beta 2}^3, & K_{\beta 2}^{8''} &= EX_1^4 \cdot K_{\beta 2}^3, & K_{\beta 2}^9(A) &= A \cdot K_{\beta 2}^3. \end{aligned} \quad (1.26)$$

Damit erhält man eine vereinfachte Darstellung der Grenzvarianz  $\tau_\beta^2$ , in der die Terme  $K_{\beta 1}^{6'}$ ,  $K_{\beta 2}^{6'}$ ,  $K_{\beta 1}^{8''}$ ,  $K_{\beta 2}^{8''}$ ,  $K_{\beta 1}^9$  und  $K_{\beta 2}^9$  nicht mehr auftreten:

$$\begin{aligned} \tau_\beta^2 &= 4 \left\{ K_{\beta 1}^1 + K_{\beta 2}^1 - (K_{\beta 1}^2 + K_{\beta 2}^2)^2 + \frac{EX_1^4 - 1}{4} (K_{\beta 1}^3 + K_{\beta 2}^3)^2 + (K_{\beta 1}^{5'} - K_{\beta 2}^{5'})^2 \right. \\ &\quad + (K_{\beta 1}^3 + K_{\beta 2}^3)(K_{\beta 1}^2 + K_{\beta 2}^2 + K_{\beta 1}^{7''} + K_{\beta 2}^{7''}) + EX_1^3 (K_{\beta 1}^3 + K_{\beta 2}^3)(K_{\beta 1}^{5'} - K_{\beta 2}^{5'}) \\ &\quad \left. + 2(K_{\beta 1}^{5'} - K_{\beta 2}^{5'})(K_{\beta 1}^{4'} - K_{\beta 2}^{4'}) \right\} \end{aligned}$$

mit

$$\begin{aligned} K_{\beta 1}^1 &= \int \int \Phi_1(t-s) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt, \\ K_{\beta 2}^1 &= \int \int \Phi_2(t+s) \left[ \Phi(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt, \\ K_{\beta 1}^2 &= \int \Phi_1(t) \left[ \Phi_1(t) - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 2}^2 &= \int (\Phi_2(t))^2 \varphi_\beta(t) dt, \\ K_{\beta 1}^3 &= \int t \Phi_1'(t) \left[ \Phi_1(t) - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 2}^3 &= \int t \Phi_2'(t) \Phi_2(t) \varphi_\beta(t) dt, \\ K_{\beta 1}^{4'} &= \int \Phi_1'(t) \Phi_2(t) \varphi_\beta(t) dt, \\ K_{\beta 2}^{4'} &= \int \Phi_2'(t) \left[ \Phi_1(t) - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 1}^{5'} &= \int t \Phi_1(t) \Phi_2(t) \varphi_\beta(t) dt, \\ K_{\beta 2}^{5'} &= \int t \Phi_2(t) \left[ \Phi_1(t) - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 1}^{7''} &= \int \Phi_1''(t) \left[ \Phi_1(t) - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 2}^{7''} &= \int \Phi_2''(t) \Phi_2(t) \varphi_\beta(t) dt. \end{aligned}$$

Setzt man entweder die in Definition 9 auf Seite 24 erklärten empirischen Analoga  $K_{\beta 1n}^1, K_{\beta 2n}^1, K_{\beta 1n}^2, K_{\beta 2n}^2, K_{\beta 1n}^3, K_{\beta 2n}^3, K_{\beta 1n}^{4'}, K_{\beta 2n}^{4'}, K_{\beta 1n}^{5'}, K_{\beta 2n}^{5'}, K_{\beta 1n}^{7''}$  und  $K_{\beta 2n}^{7''}$  in die obige Darstellung von  $\tau_\beta^2$  ein, oder führt man in der Darstellung von  $\tau_\beta^2$  aus Definition 9 für die Terme  $K_{\beta 1n}^{6'}, K_{\beta 2n}^{6'}, K_{\beta 1n}^{8''}, K_{\beta 2n}^{8''}, K_{\beta 1n}^9$  und  $K_{\beta 2n}^9$  dieselben Vereinfachungen durch wie in (1.26) für deren Grenzwerte, so gelangt man zu dem folgenden Schätzer für  $\tau_\beta^2$ :

$$\begin{aligned} \tau_{\beta n}^2 = & 4 \left\{ K_{\beta 1n}^1 + K_{\beta 2n}^1 - (K_{\beta 1n}^2 + K_{\beta 2n}^2)^2 + \frac{\widehat{m}_4 - 1}{4} (K_{\beta 1n}^3 + K_{\beta 2n}^3)^2 + (K_{\beta 1n}^{5'} - K_{\beta 2n}^{5'})^2 \right. \\ & + (K_{\beta 1n}^3 + K_{\beta 2n}^3)(K_{\beta 1n}^2 + K_{\beta 2n}^2 + K_{\beta 1n}^{7''} + K_{\beta 2n}^{7''}) + \widehat{m}_3 (K_{\beta 1n}^3 + K_{\beta 2n}^3)(K_{\beta 1n}^{5'} - K_{\beta 2n}^{5'}) \\ & \left. + 2(K_{\beta 1n}^{5'} - K_{\beta 2n}^{5'})(K_{\beta 1n}^{4'} - K_{\beta 2n}^{4'}) \right\}. \end{aligned}$$

Dabei sind

$$\widehat{m}_3 := \frac{1}{n} \sum_{j=1}^n Y_j^3 \quad \text{und} \quad \widehat{m}_4 := \frac{1}{n} \sum_{j=1}^n Y_j^4$$

die üblichen auf den standardisierten Daten beruhenden Schätzer für das dritte bzw. vierte Moment der zugrundeliegenden Verteilung, die fast sicher gegen  $EX_1^3$  bzw.  $EX_1^4$  konvergieren, wie man analog zu Lemma 7 auf Seite 23 leicht einsieht.

**Bemerkung 17** Ist die zugrundeliegende Verteilung von  $X_1$  sogar symmetrisch um den Ursprung (d.h. es gilt  $X_1 \sim -X_1$ ), so folgen  $EX_1^3 = 0$  und  $\Phi_2 \equiv \Phi_2' \equiv \Phi_2'' \equiv 0$ . Damit verschwinden die Größen  $K_{\beta 2}^1, K_{\beta 2}^2, K_{\beta 2}^3, K_{\beta 2}^{4'}, K_{\beta 2}^{5'}, K_{\beta 2}^{5'}$  und  $K_{\beta 2}^{7''}$ , und die Varianz  $\tau_\beta^2$  vereinfacht sich zu

$$\tau_\beta^2 = 4 \left\{ K_{\beta 1}^1 - (K_{\beta 1}^2)^2 + \frac{EX_1^4 - 1}{4} (K_{\beta 1}^3)^2 + K_{\beta 1}^3 (K_{\beta 1}^2 + K_{\beta 1}^{7''}) \right\}. \quad (1.27)$$

Analog reduziert sich der Schätzer für  $\tau_\beta^2$  zu

$$\tau_{\beta n}^2 = 4 \left\{ K_{\beta 1n}^1 - (K_{\beta 1n}^2)^2 + \frac{\widehat{m}_4 - 1}{4} (K_{\beta 1n}^3)^2 + K_{\beta 1n}^3 (K_{\beta 1n}^2 + K_{\beta 1n}^{7''}) \right\}.$$

Ferner gilt für symmetrische Verteilungen

$$K_{\beta 1}^1 = 4 \mathbb{E} \left[ \left( \int_0^\infty \cos(tX_1) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right)^2 \right]. \quad (1.28)$$

**BEWEIS:** Die Identität  $\cos(x - y) = \cos(x) \cos(y) + \sin(x) \sin(y)$ , der Satz von Fubini und Symmetrieüberlegungen zeigen

$$\begin{aligned} K_{\beta 1}^1 &= \int \int \mathbb{E} [\cos((t - s)X_1)] \left[ \Phi_1(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) \varphi_\beta(t) ds dt \\ &= \mathbb{E} \left[ \left( \int_{-\infty}^\infty \cos(tX_1) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right)^2 \right. \\ &\quad \left. + \left( \int_{-\infty}^\infty \sin(tX_1) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right)^2 \right] \\ &= \mathbb{E} \left[ \left( 2 \int_0^\infty \cos(tX_1) \left[ \Phi_1(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \right)^2 \right] \end{aligned}$$

und damit die Darstellung (1.28). □

Im folgenden werden die sechs eindimensionalen Beispielverteilungen vorgestellt, an denen die statistischen Aussagen aus Abschnitt 1.3 veranschaulicht werden sollen. Da bei der Herleitung der Parameter  $L_\beta$  und  $\tau_\beta^2$  die zugrundeliegende Verteilung als standardisiert vorausgesetzt wurde, werden auch die Beispielverteilungen in ihrer standardisierten Form verwendet, d.h. mit Erwartungswert  $EX_1 = 0$  und Varianz  $\text{Var}(X_1) = 1$ . Im einzelnen betrachten wir

1. die Gleichverteilung  $\mathcal{U}(-\sqrt{3}, \sqrt{3})$  mit der Dichte  $f_{X_1}(x) = \frac{1}{2\sqrt{3}} \cdot \mathbf{1}_{[-\sqrt{3}, \sqrt{3}]}(x)$ ,
2. die Laplace-Verteilung  $\text{Laplace}(0, \frac{1}{\sqrt{2}})$  mit der Dichte  $f_{X_1}(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}$ ,
3. die Exponentialverteilung  $\text{Exp}(1) - 1$  (d.h.  $X_1 = Z - 1$  mit  $Z \sim \text{Exp}(1)$ ) mit der Dichte  $f_{X_1}(x) = e^{-x-1} \cdot \mathbf{1}_{[x > -1]}$ ,
4. die Normal Mixture  $\frac{1}{2}\mathcal{N}(\frac{\sqrt{3}}{2}, \frac{1}{4}) + \frac{1}{2}\mathcal{N}(-\frac{\sqrt{3}}{2}, \frac{1}{4})$ , im folgenden als *Mix1* bezeichnet, mit der Dichte  $f_{X_1}(x) = \frac{1}{\sqrt{2\pi}} \left( e^{-2(x-\frac{\sqrt{3}}{2})^2} + e^{-2(x+\frac{\sqrt{3}}{2})^2} \right)$ ,
5. die Normal Mixture  $\frac{1}{2}\mathcal{N}(\frac{\sqrt{8}}{3}, \frac{1}{9}) + \frac{1}{2}\mathcal{N}(-\frac{\sqrt{8}}{3}, \frac{1}{9})$ , im folgenden als *Mix2* bezeichnet, mit der Dichte  $f_{X_1}(x) = \frac{3}{2\sqrt{2\pi}} \left( e^{-\frac{9}{2}(x-\frac{\sqrt{8}}{3})^2} + e^{-\frac{9}{2}(x+\frac{\sqrt{8}}{3})^2} \right)$  und
6. die Normal Mixture  $\frac{1}{3}\mathcal{N}(-\frac{2}{\sqrt{3}}, \frac{1}{3}) + \frac{2}{3}\mathcal{N}(\frac{1}{\sqrt{3}}, \frac{1}{3})$ , im folgenden als *Mix3* bezeichnet, mit der Dichte  $f_{X_1}(x) = \frac{1}{\sqrt{3\pi}} \left( \frac{1}{\sqrt{2}} e^{-\frac{3}{2}(x+\frac{2}{\sqrt{3}})^2} + \sqrt{2} e^{-\frac{3}{2}(x-\frac{1}{\sqrt{3}})^2} \right)$ .

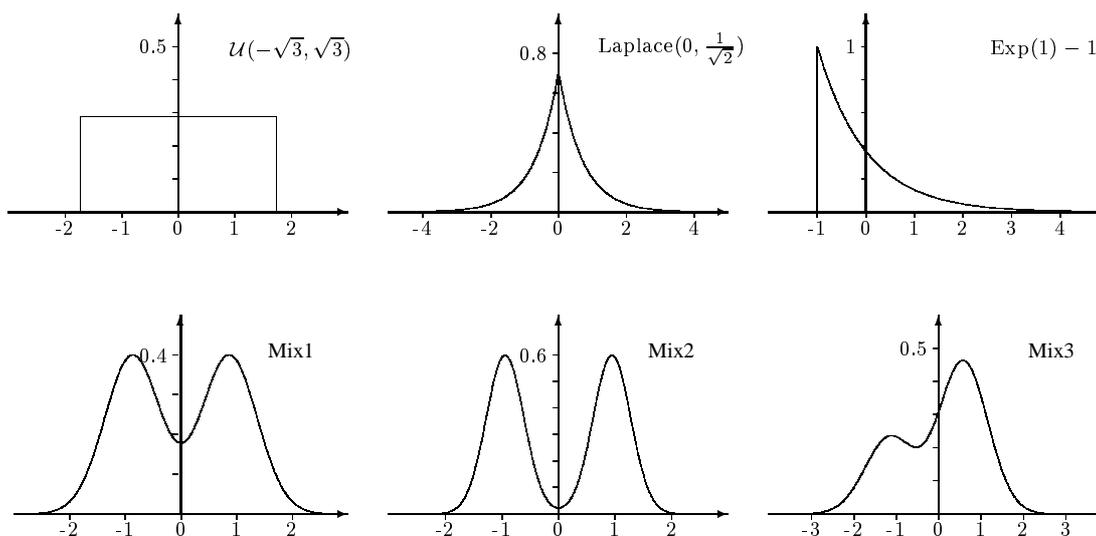


Abbildung 1.1.: Dichten der betrachteten Verteilungen

Mischungen aus Normalverteilungen (Normal Mixtures)  $X_1 \sim p\mathcal{N}(\mu, \sigma^2) + (1-p)\mathcal{N}(\nu, \tau^2)$  mit  $p \in (0, 1)$ ,  $\mu, \nu \in \mathbb{R}$  und  $\sigma^2, \tau^2 > 0$  lassen sich in der Form  $X_1 = UY + (1-U)Z$  mit unabhängigen Zufallsvariablen  $U, Y, Z$  darstellen. Dabei gelten  $P(U = 1) = p = 1 - P(U = 0)$ ,  $Y \sim \mathcal{N}(\mu, \sigma^2)$  und  $Z \sim \mathcal{N}(\nu, \tau^2)$ . Bezeichnen

$$\Psi_Y(t) := e^{it\mu - \frac{1}{2}t^2\sigma^2} \quad \text{und} \quad \Psi_Z(t) := e^{it\nu - \frac{1}{2}t^2\tau^2}$$

die charakteristischen Funktionen von  $Y$  und  $Z$ , so ist die charakteristische Funktion von  $X_1$  gegeben durch

$$\begin{aligned} \Psi_{X_1}(t) &= p\Psi_Y(t) + (1-p)\Psi_Z(t) \\ &= pe^{-\frac{1}{2}t^2\sigma^2} \cos(t\mu) + (1-p)e^{-\frac{1}{2}t^2\tau^2} \cos(t\nu) + i \left[ pe^{-\frac{1}{2}t^2\sigma^2} \sin(t\mu) + (1-p)e^{-\frac{1}{2}t^2\tau^2} \sin(t\nu) \right], \end{aligned}$$

und folglich sind die Anteile der Sinus/Cosinus-Transformierten von  $X_1$  gleich

$$\begin{aligned} \Phi_1(t) &= pe^{-\frac{1}{2}t^2\sigma^2} \cos(t\mu) + (1-p)e^{-\frac{1}{2}t^2\tau^2} \cos(t\nu), \\ \Phi_2(t) &= pe^{-\frac{1}{2}t^2\sigma^2} \sin(t\mu) + (1-p)e^{-\frac{1}{2}t^2\tau^2} \sin(t\nu). \end{aligned}$$

Bezeichnen  $f_Y$  bzw.  $f_Z$  Dichten von  $Y$  bzw.  $Z$ , so hat  $X_1$  die Dichte

$$f_{X_1}(x) = pf_Y(x) + (1-p)f_Z(x).$$

Die ersten beiden Momente von  $X_1$  sind gegeben durch

$$EX_1 = \mu p + \nu(1-p), \quad EX_1^2 = p(\sigma^2 + \mu^2) + (1-p)(\tau^2 + \nu^2).$$

Im Fall  $\nu = -\mu$  gilt speziell

$$EX_1 = 0 \iff p = \frac{1}{2},$$

und falls diese Äquivalenz zusätzlich erfüllt ist, gilt

$$\text{Var}(X_1) = EX_1^2 = 1 \iff \sigma^2 + \tau^2 = 2(1 - \mu^2).$$

In den folgenden Abschnitten werden zunächst die zur Berechnung von  $\tau_\beta^2$  benötigten Größen für die sechs Beispielverteilungen angegeben, bevor die Konfidenzbereiche, die verschiedenen Gütenäherungen und ausgewählte Ergebnisse des inversen Tests vorgestellt werden.

### 1.4.1. Die Gleichverteilung

Im Fall  $X_1 \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$  gelten  $EX_1^3 = 0$  sowie  $EX_1^4 = \frac{9}{5}$ , und es liegt die Dichte

$$f_{X_1}(x) = \frac{1}{2\sqrt{3}} \cdot \mathbf{1}_{[-\sqrt{3}, \sqrt{3}]}(x)$$

zugrunde. Da die Verteilung symmetrisch zum Ursprung ist, gelten  $\Phi_2 \equiv \Phi_2' \equiv \Phi_2'' \equiv 0$ , und  $\tau_\beta^2$  nimmt die vereinfachte Gestalt (1.27) von Seite 41 an. Die nichttrivialen Komponenten der Sinus/Cosinus-Transformierten von  $X_1$  sind gegeben durch

$$\begin{aligned}\Phi_1(t) &= \frac{\sin(\sqrt{3}t)}{\sqrt{3}t}, \\ \Phi_1'(t) &= \frac{\sqrt{3}t \cos(\sqrt{3}t) - \sin(\sqrt{3}t)}{\sqrt{3}t^2}, \\ \Phi_1''(t) &= \frac{-3t^2 \sin(\sqrt{3}t) - 2\sqrt{3}t \cos(\sqrt{3}t) + 2 \sin(\sqrt{3}t)}{\sqrt{3}t^3},\end{aligned}$$

der Grenzwert  $L_\beta$  durch

$$L_\beta = \frac{1}{\sqrt{6\pi}\beta} \int_{-\infty}^{\infty} \left(\frac{\sin t}{t}\right)^2 e^{-\frac{t^2}{6\beta^2}} dt - \sqrt{\frac{2}{3\pi}} \frac{1}{\beta} \int_{-\infty}^{\infty} \frac{\sin t}{t} e^{-\frac{(1+\beta^2)}{6\beta^2}t^2} dt + \frac{1}{\sqrt{1+\beta^2}}$$

und die nichttrivialen Komponenten von  $\tau_\beta^2$  durch

$$\begin{aligned}K_{\beta 1}^1 &= \frac{2}{\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} \left[ \int_0^\infty \cos(tx) \left[ \frac{\sin(\sqrt{3}t)}{\sqrt{3}t} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt \right]^2 dx, \\ K_{\beta 1}^2 &= 2 \int_0^\infty \frac{\sin(\sqrt{3}t)}{\sqrt{3}t} \left[ \frac{\sin(\sqrt{3}t)}{\sqrt{3}t} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 1}^3 &= 2 \int_0^\infty t \left[ \frac{\sqrt{3}t \cos(\sqrt{3}t) - \sin(\sqrt{3}t)}{\sqrt{3}t^2} \right] \left[ \frac{\sin(\sqrt{3}t)}{\sqrt{3}t} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 1}^{7''} &= 2 \int_0^\infty \left[ \frac{-3t^2 \sin(\sqrt{3}t) - 2\sqrt{3}t \cos(\sqrt{3}t) + 2 \sin(\sqrt{3}t)}{\sqrt{3}t^3} \right] \left[ \frac{\sin(\sqrt{3}t)}{\sqrt{3}t} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt.\end{aligned}$$

Damit ist  $\tau_\beta^2 = 4 \left\{ K_{\beta 1}^1 - (K_{\beta 1}^2)^2 + \frac{1}{5} (K_{\beta 1}^3)^2 + K_{\beta 1}^3 (K_{\beta 1}^2 + K_{\beta 1}^{7''}) \right\}$ .

### 1.4.2. Die Laplace-Verteilung

Im Fall  $X_1 \sim \text{Laplace}(0, \frac{1}{\sqrt{2}})$  gelten  $EX_1^3 = 0$  sowie  $EX_1^4 = 6$ , und es liegt die Dichte

$$f_{X_1}(x) = \frac{1}{\sqrt{2}} \exp\{-\sqrt{2}|x|\}$$

zugrunde. Da die Verteilung symmetrisch zum Ursprung ist, gelten  $\Phi_2 \equiv \Phi_2' \equiv \Phi_2'' \equiv 0$ , und  $\tau_\beta^2$  nimmt die vereinfachte Gestalt (1.27) von Seite 41 an. Die nichttrivialen Komponenten der Sinus/Cosinus-Transformierten von  $X_1$  sind gegeben durch

$$\Phi_1(t) = \frac{2}{2+t^2}, \quad \Phi_1'(t) = \frac{-4t}{(2+t^2)^2}, \quad \Phi_1''(t) = 4 \frac{3t^2-2}{(2+t^2)^3},$$

der Grenzwert  $L_\beta$  durch

$$L_\beta = 2 \int_0^\infty \left( \frac{2}{2+t^2} - e^{-\frac{t^2}{2}} \right)^2 \varphi_\beta(t) dt$$

und die nichttrivialen Komponenten von  $\tau_\beta^2$  durch

$$\begin{aligned} K_{\beta 1}^1 &= \frac{8}{\sqrt{2}} \int_0^\infty \left( \int_0^\infty \cos(tx) \left[ \frac{2}{2+t^2} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt \right)^2 e^{-\sqrt{2}x} dx, \\ K_{\beta 1}^2 &= 2 \int_0^\infty \frac{2}{2+t^2} \left[ \frac{2}{2+t^2} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 1}^3 &= -8 \int_0^\infty \frac{t^2}{(2+t^2)^2} \left[ \frac{2}{2+t^2} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt, \\ K_{\beta 1}^{7''} &= 8 \int_0^\infty \frac{3t^2-2}{(2+t^2)^3} \left[ \frac{2}{2+t^2} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt. \end{aligned}$$

Damit ist  $\tau_\beta^2 = 4 \left\{ K_{\beta 1}^1 - (K_{\beta 1}^2)^2 + \frac{5}{4}(K_{\beta 1}^3)^2 + K_{\beta 1}^3(K_{\beta 1}^2 + K_{\beta 1}^{7''}) \right\}$ .

### 1.4.3. Die Exponentialverteilung

Im Fall  $X_1 \sim \text{Exp}(1) - 1$  gelten  $EX^3 = 2$ ,  $EX^4 = 9$ , und  $X_1$  hat die Dichte

$$f_{X_1}(x) = \exp\{-x - 1\} \cdot \mathbf{1}\{x > -1\}.$$

Diese Verteilung ist nicht symmetrisch zum Ursprung; deshalb müssen alle Komponenten der Sinus/Cosinus-Transformierten von  $X_1$  berechnet werden:

$$\begin{aligned} \Phi_1(t) &= \frac{\cos t + t \sin t}{1 + t^2}, \\ \Phi_1'(t) &= \frac{t^3 \cos t - 2t^2 \sin t - t \cos t}{(1 + t^2)^2}, \\ \Phi_1''(t) &= \frac{-t^5 \sin t - 3t^4 \cos t + 4t^3 \sin t + 4t^2 \cos t - 3t \sin t - \cos t}{(1 + t^2)^3} \end{aligned}$$

sowie

$$\begin{aligned} \Phi_2(t) &= \frac{t \cos t - \sin t}{1 + t^2}, \\ \Phi_2'(t) &= \frac{-t^3 \sin t - 2t^2 \cos t + t \sin t}{(1 + t^2)^2}, \\ \Phi_2''(t) &= \frac{-t^5 \cos t + 3t^4 \sin t + 4t^3 \cos t - 4t^2 \sin t - 3t \cos t + \sin t}{(1 + t^2)^3}. \end{aligned}$$

Damit ist  $L_\beta$  gegeben durch

$$L_\beta = \int_{-\infty}^{\infty} \left( \frac{\cos t - \sin t + t(\cos t + \sin t)}{1 + t^2} - e^{-\frac{t^2}{2}} \right)^2 \varphi_\beta(t) dt.$$

Die Komponenten von  $\tau_\beta^2$  sind wie auf Seite 40 definiert, und  $\tau_\beta^2$  ergibt sich zu

$$\begin{aligned} \tau_\beta^2 &= 4 \left\{ K_{\beta 1}^1 + K_{\beta 2}^1 - (K_{\beta 1}^2 + K_{\beta 2}^2)^2 + 2(K_{\beta 1}^3 + K_{\beta 2}^3)^2 + (K_{\beta 1}^{5'} - K_{\beta 2}^{5'})^2 \right. \\ &\quad + (K_{\beta 1}^3 + K_{\beta 2}^3)(K_{\beta 1}^2 + K_{\beta 2}^2 + K_{\beta 1}^{7''} + K_{\beta 2}^{7''}) + 2(K_{\beta 1}^3 + K_{\beta 2}^3)(K_{\beta 1}^{5'} - K_{\beta 2}^{5'}) \\ &\quad \left. + 2(K_{\beta 1}^{5'} - K_{\beta 2}^{5'})(K_{\beta 1}^{4'} - K_{\beta 2}^{4'}) \right\}. \end{aligned}$$

### 1.4.4. Die erste Normal Mixture

Im Fall  $X_1 \sim \frac{1}{2}\mathcal{N}\left(\frac{\sqrt{3}}{2}, \frac{1}{4}\right) + \frac{1}{2}\mathcal{N}\left(-\frac{\sqrt{3}}{2}, \frac{1}{4}\right)$  gelten  $EX^3 = 0$ ,  $EX^4 = \frac{15}{8}$ , und es liegt die Dichte

$$f_{X_1}(x) = \frac{1}{\sqrt{2\pi}} \left( e^{-2\left(x-\frac{\sqrt{3}}{2}\right)^2} + e^{-2\left(x+\frac{\sqrt{3}}{2}\right)^2} \right) = \frac{1}{\sqrt{2\pi}} e^{(-2x^2-\frac{3}{2})} \left( e^{2\sqrt{3}x} + e^{-2\sqrt{3}x} \right)$$

zugrunde. Da die Verteilung symmetrisch zum Ursprung ist, gelten  $\Phi_2 \equiv \Phi_2' \equiv \Phi_2'' \equiv 0$ , und  $\tau_\beta^2$  nimmt die vereinfachte Gestalt (1.27) von Seite 41 an. Die nichttrivialen Komponenten der Sinus/Cosinus-Transformierten von  $X_1$  sind gegeben durch

$$\begin{aligned} \Phi_1(t) &= \cos\left(\frac{\sqrt{3}}{2}t\right) e^{-\frac{t^2}{8}}, \\ \Phi_1'(t) &= -e^{-\frac{t^2}{8}} \left[ \frac{t}{4} \cos\left(\frac{\sqrt{3}}{2}t\right) + \frac{\sqrt{3}}{2} \sin\left(\frac{\sqrt{3}}{2}t\right) \right], \\ \Phi_1''(t) &= e^{-\frac{t^2}{8}} \left[ \frac{t^2}{16} \cos\left(\frac{\sqrt{3}}{2}t\right) + \frac{\sqrt{3}t}{4} \sin\left(\frac{\sqrt{3}}{2}t\right) - \cos\left(\frac{\sqrt{3}}{2}t\right) \right], \end{aligned}$$

der Grenzwert  $L_\beta$  durch

$$L_\beta = 2 \int_0^\infty \left( \cos\left(\frac{\sqrt{3}}{2}t\right) e^{-\frac{t^2}{8}} - e^{-\frac{t^2}{2}} \right)^2 \varphi_\beta(t) dt$$

und die nichttrivialen Komponenten von  $\tau_\beta^2$  durch

$$K_{\beta 1}^1 = \frac{4}{\sqrt{2\pi}} \int_{-\infty}^\infty \left( \int_0^\infty \cos(tx) \left[ \cos\left(\frac{\sqrt{3}}{2}t\right) e^{-\frac{t^2}{8}} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt \right)^2 \left( e^{-2\left(x-\frac{\sqrt{3}}{2}\right)^2} + e^{-2\left(x+\frac{\sqrt{3}}{2}\right)^2} \right) dx,$$

$$K_{\beta 1}^2 = 2 \int_0^\infty \cos\left(\frac{\sqrt{3}}{2}t\right) e^{-\frac{t^2}{8}} \left[ \cos\left(\frac{\sqrt{3}}{2}t\right) e^{-\frac{t^2}{8}} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt,$$

$$K_{\beta 1}^3 = -2 \int_0^\infty t e^{-\frac{t^2}{8}} \left[ \frac{t}{4} \cos\left(\frac{\sqrt{3}}{2}t\right) + \frac{\sqrt{3}}{2} \sin\left(\frac{\sqrt{3}}{2}t\right) \right] \left[ \cos\left(\frac{\sqrt{3}}{2}t\right) e^{-\frac{t^2}{8}} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt,$$

$$K_{\beta 1}^{7''} = 2 \int_0^\infty e^{-\frac{t^2}{8}} \left[ \frac{t^2}{16} \cos\left(\frac{\sqrt{3}}{2}t\right) + \frac{\sqrt{3}t}{4} \sin\left(\frac{\sqrt{3}}{2}t\right) - \cos\left(\frac{\sqrt{3}}{2}t\right) \right] \left[ \cos\left(\frac{\sqrt{3}}{2}t\right) e^{-\frac{t^2}{8}} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt.$$

Damit ist  $\tau_\beta^2 = 4 \left\{ K_{\beta 1}^1 - (K_{\beta 1}^2)^2 + \frac{7}{32} (K_{\beta 1}^3)^2 + K_{\beta 1}^3 (K_{\beta 1}^2 + K_{\beta 1}^{7''}) \right\}$ .

### 1.4.5. Die zweite Normal Mixture

Im Fall  $X_1 \sim \frac{1}{2}\mathcal{N}\left(\frac{\sqrt{8}}{3}, \frac{1}{9}\right) + \frac{1}{2}\mathcal{N}\left(-\frac{\sqrt{8}}{3}, \frac{1}{9}\right)$  gelten  $EX^3 = 0$ ,  $EX^4 = \frac{115}{81}$ , und es liegt die Dichte

$$f_{X_1}(x) = \frac{3}{2\sqrt{2\pi}} \left( e^{-\frac{9}{2}\left(x-\frac{\sqrt{8}}{3}\right)^2} + e^{-\frac{9}{2}\left(x+\frac{\sqrt{8}}{3}\right)^2} \right)$$

zugrunde. Da die Verteilung symmetrisch zum Ursprung ist, gelten  $\Phi_2 \equiv \Phi_2' \equiv \Phi_2'' \equiv 0$ , und  $\tau_\beta^2$  nimmt die vereinfachte Gestalt (1.27) von Seite 41 an. Die nichttrivialen Komponenten der Sinus/Cosinus-Transformierten von  $X_1$  sind gegeben durch

$$\begin{aligned}\Phi_1(t) &= \cos\left(\frac{\sqrt{8}}{3}t\right) e^{-\frac{t^2}{18}}, \\ \Phi_1'(t) &= -e^{-\frac{t^2}{18}} \left[ \frac{t}{9} \cos\left(\frac{\sqrt{8}}{3}t\right) + \frac{\sqrt{8}}{3} \sin\left(\frac{\sqrt{8}}{3}t\right) \right], \\ \Phi_1''(t) &= e^{-\frac{t^2}{18}} \left[ \frac{t^2}{81} \cos\left(\frac{\sqrt{8}}{3}t\right) + \frac{2\sqrt{8}t}{27} \sin\left(\frac{\sqrt{8}}{3}t\right) - \cos\left(\frac{\sqrt{8}}{3}t\right) \right],\end{aligned}$$

der Grenzwert  $L_\beta$  durch

$$L_\beta = 2 \int_0^\infty \left( \cos\left(\frac{\sqrt{8}}{3}t\right) e^{-\frac{t^2}{18}} - e^{-\frac{t^2}{2}} \right)^2 \varphi_\beta(t) dt$$

und die nichttrivialen Komponenten von  $\tau_\beta^2$  durch

$$K_{\beta 1}^1 = \frac{6}{\sqrt{2\pi}} \int_{-\infty}^\infty \left( \int_0^\infty \cos(tx) \left[ \cos\left(\frac{\sqrt{8}}{3}t\right) e^{-\frac{t^2}{18}} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt \right)^2 e^{-\frac{9}{2}x^2 - 4} \left( e^{3\sqrt{8}x} + e^{-3\sqrt{8}x} \right) dx,$$

$$K_{\beta 1}^2 = 2 \int_0^\infty \cos\left(\frac{\sqrt{8}}{3}t\right) e^{-\frac{t^2}{18}} \left[ \cos\left(\frac{\sqrt{8}}{3}t\right) e^{-\frac{t^2}{18}} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt,$$

$$= \frac{3}{2\sqrt{2\beta^2+9}} \left( 1 + e^{-\frac{16\beta^2}{2\beta^2+9}} \right) - \frac{3}{\sqrt{10\beta^2+9}} e^{-\frac{4\beta^2}{10\beta^2+9}},$$

$$K_{\beta 1}^3 = -2 \int_0^\infty t e^{-\frac{t^2}{18}} \left[ \frac{t}{9} \cos\left(\frac{\sqrt{8}}{3}t\right) + \frac{\sqrt{8}}{3} \sin\left(\frac{\sqrt{8}}{3}t\right) \right] \left[ \cos\left(\frac{\sqrt{8}}{3}t\right) e^{-\frac{t^2}{18}} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt,$$

$$K_{\beta 1}^{7''} = 2 \int_0^\infty e^{-\frac{t^2}{18}} \left[ \frac{t^2}{81} \cos\left(\frac{\sqrt{8}}{3}t\right) + \frac{2\sqrt{8}t}{27} \sin\left(\frac{\sqrt{8}}{3}t\right) - \cos\left(\frac{\sqrt{8}}{3}t\right) \right] \left[ \cos\left(\frac{\sqrt{8}}{3}t\right) e^{-\frac{t^2}{18}} - e^{-\frac{t^2}{2}} \right] \varphi_\beta(t) dt.$$

Dabei wurde zum Beweis der zweiten Darstellung von  $K_{\beta 1}^2$  Gleichung (1.30) auf Seite 57 verwendet. Folglich ist  $\tau_\beta^2 = 4 \left\{ K_{\beta 1}^1 - K_{\beta 1}^2 + \frac{17}{162} (K_{\beta 1}^3)^2 + K_{\beta 1}^3 (K_{\beta 1}^2 + K_{\beta 1}^{7''}) \right\}$ .

### 1.4.6. Die dritte Normal Mixture

Im Fall  $X_1 \sim \frac{1}{3}\mathcal{N}\left(-\frac{2}{\sqrt{3}}, \frac{1}{3}\right) + \frac{2}{3}\mathcal{N}\left(\frac{1}{\sqrt{3}}, \frac{1}{3}\right)$  gelten  $EX^3 = -\frac{2}{3\sqrt{3}}$ ,  $EX^4 = \frac{7}{3}$ , und  $X_1$  hat die Dichte

$$f_{X_1}(x) = \frac{1}{\sqrt{3\pi}} \left( \frac{1}{\sqrt{2}} e^{-\frac{3}{2}\left(x+\frac{2}{\sqrt{3}}\right)^2} + \sqrt{2} e^{-\frac{3}{2}\left(x-\frac{1}{\sqrt{3}}\right)^2} \right).$$

Diese Verteilung ist nicht symmetrisch zum Ursprung; deshalb müssen alle Komponenten der Sinus/Cosinus-Transformierten von  $X_1$  berechnet werden:

$$\begin{aligned}\Phi_1(t) &= \frac{1}{3}e^{-\frac{t^2}{6}} \left[ \cos\left(\frac{2t}{\sqrt{3}}\right) + 2\cos\left(\frac{t}{\sqrt{3}}\right) \right], \\ \Phi_1'(t) &= \frac{1}{9}e^{-\frac{t^2}{6}} \left[ -t \left( \cos\left(\frac{2t}{\sqrt{3}}\right) + 2\cos\left(\frac{t}{\sqrt{3}}\right) \right) - 2\sqrt{3} \left( \sin\left(\frac{2t}{\sqrt{3}}\right) + \sin\left(\frac{t}{\sqrt{3}}\right) \right) \right], \\ \Phi_1''(t) &= \frac{1}{27}e^{-\frac{t^2}{6}} \left[ t^2 \left( \cos\left(\frac{2t}{\sqrt{3}}\right) + 2\cos\left(\frac{t}{\sqrt{3}}\right) \right) + 4\sqrt{3}t \left( \sin\left(\frac{2t}{\sqrt{3}}\right) + \sin\left(\frac{t}{\sqrt{3}}\right) \right) \right. \\ &\quad \left. - 15\cos\left(\frac{2t}{\sqrt{3}}\right) - 12\cos\left(\frac{t}{\sqrt{3}}\right) \right]\end{aligned}$$

sowie

$$\begin{aligned}\Phi_2(t) &= \frac{1}{3}e^{-\frac{t^2}{6}} \left[ -\sin\left(\frac{2t}{\sqrt{3}}\right) + 2\sin\left(\frac{t}{\sqrt{3}}\right) \right], \\ \Phi_2'(t) &= \frac{1}{9}e^{-\frac{t^2}{6}} \left[ t \left( \sin\left(\frac{2t}{\sqrt{3}}\right) - 2\sin\left(\frac{t}{\sqrt{3}}\right) \right) - 2\sqrt{3} \left( \cos\left(\frac{2t}{\sqrt{3}}\right) - \cos\left(\frac{t}{\sqrt{3}}\right) \right) \right], \\ \Phi_2''(t) &= \frac{1}{27}e^{-\frac{t^2}{6}} \left[ t^2 \left( -\sin\left(\frac{2t}{\sqrt{3}}\right) + 2\sin\left(\frac{t}{\sqrt{3}}\right) \right) + 4\sqrt{3}t \left( \cos\left(\frac{2t}{\sqrt{3}}\right) - \cos\left(\frac{t}{\sqrt{3}}\right) \right) \right. \\ &\quad \left. + 15\sin\left(\frac{2t}{\sqrt{3}}\right) - 12\sin\left(\frac{t}{\sqrt{3}}\right) \right].\end{aligned}$$

Damit ist  $L_\beta$  gegeben durch

$$L_\beta = \int_{-\infty}^{\infty} \frac{1}{9}e^{-\frac{t^2}{3}} \left( \cos\left(\frac{2t}{\sqrt{3}}\right) - \sin\left(\frac{2t}{\sqrt{3}}\right) + 2\cos\left(\frac{t}{\sqrt{3}}\right) + 2\sin\left(\frac{t}{\sqrt{3}}\right) \right)^2 \varphi_\beta(t) dt.$$

Die Komponenten von  $\tau_\beta^2$  sind wie auf Seite 40 definiert, und  $\tau_\beta^2$  ergibt sich zu

$$\begin{aligned}\tau_\beta^2 &= 4 \left\{ K_{\beta_1}^1 + K_{\beta_2}^1 - (K_{\beta_1}^2 + K_{\beta_2}^2)^2 + \frac{1}{3}(K_{\beta_1}^3 + K_{\beta_2}^3)^2 + (K_{\beta_1}^{5'} - K_{\beta_2}^{5'})^2 \right. \\ &\quad \left. + (K_{\beta_1}^3 + K_{\beta_2}^3)(K_{\beta_1}^2 + K_{\beta_2}^2 + K_{\beta_1}^{7''} + K_{\beta_2}^{7''}) - \frac{2}{3\sqrt{3}}(K_{\beta_1}^3 + K_{\beta_2}^3)(K_{\beta_1}^{5'} - K_{\beta_2}^{5'}) \right. \\ &\quad \left. + 2(K_{\beta_1}^{5'} - K_{\beta_2}^{5'})(K_{\beta_1}^{4'} - K_{\beta_2}^{4'}) \right\}.\end{aligned}$$

### 1.4.7. Empirische Resultate in den sechs Beispielen

Im folgenden soll anhand der sechs ausgewählten Beispiel-Alternativen untersucht werden, inwieweit die statistischen Folgerungen aus den Kapiteln 1.3.1, 1.3.2 und 1.3.3 in der Praxis brauchbar sind. Alle zugehörigen Implementierungen wurden in Fortran 90 durchgeführt, und es wurden so weit wie möglich Routinen der NAG- oder IMSL-Programmbibliothek verwendet. Dabei kamen vor allem numerische Integrationsroutinen zur Auswertung der in den Bestandteilen von  $\tau_\beta^2$  vorhandenen Integrale und Routinen zur Erzeugung von Pseudo-Zufallszahlen für die Monte-Carlo-Simulationen zum Einsatz.

Der Verlauf von  $L_\beta$  und der Näherung  $n^{-1}T_{n,\beta}$  bzw. von  $\tau_\beta^2$  und der Näherung  $\tau_{\beta n}^2$  in Abhängigkeit von  $\beta$  ist im Fall der sechs Beispielverteilungen in den Abbildungen 1.2 und 1.3 für

$n \in \{20, 50, 100\}$  dargestellt. Dabei basiert jeder der Werte von  $n^{-1}T_{n,\beta}$  bzw.  $\tau_{\beta n}^2$  auf 100 000 Monte–Carlo–Wiederholungen. Dies bedeutet, daß 100 000 unabhängige Stichproben vom Umfang  $n$  unter der zugrundeliegenden Alternative gebildet wurden, daß dann anhand jeder Stichprobe der Wert von  $n^{-1}T_{n,\beta}$  bzw.  $\tau_{\beta n}^2$  berechnet wurde, und daß schließlich das arithmetische Mittel dieser 100 000 Werte notiert wurde. An den Schaubildern sieht man, daß für alle sechs zugrundeliegenden Alternativen die theoretischen Kurven ab einem Stichprobenumfang von  $n = 100$  recht gut angenähert werden. Tendenziell sind die Näherungen von  $\tau_{\beta}^2$  besser als die von  $L_{\beta}$ . Für kleine Parameterwerte  $\beta \leq 1.5$ , d.h. den uns hauptsächlich interessierenden Bereich (s.u.), sind die Näherungen bereits für kleine Stichprobenumfänge ( $n = 20$ ) sehr gut.

Alle nachfolgenden empirischen Untersuchungen der statistischen Folgerungen aus den Kapiteln 1.3.1, 1.3.2 und 1.3.3 wurden für die Gewichtsparameter  $\beta \in \{0.5, 0.8, 1.0, 1.2, 1.4, 1.8, 2.0, 2.2, 2.6, 3.0, 3.6, 4.4\}$  und die Stichprobenumfänge  $n \in \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000\}$  durchgeführt. Aus Platzgründen werden in der vorliegenden Arbeit in den Tabellen 1.2 bis 1.7 nur die numerischen Ergebnisse für den Gewichtsparameter  $\beta = 1.0$  angegeben, in den Abbildungen 1.4 bis 1.9 die Gütekurven in den Fällen  $\beta \in \{0.8, 1.0, 3.6\}$ . Für andere Werte von  $\beta$  ergab sich ein sehr ähnliches Bild; auf eventuell abweichende Ergebnisse bei anderen Parametern wird an geeigneter Stelle hingewiesen. Die besondere Betonung von  $\beta = 1.0$  geschah einerseits unter dem Gesichtspunkt, daß in dem Bereich (0.8, 1.2) die empirische Güte des BHEP–Tests in Monte–Carlo–Simulationen bei nahezu allen untersuchten Alternativen am höchsten war (vgl. z.B. Henze und Zirkler (1990)). Andererseits ist der BHEP–Test für  $\beta = 1.0$  identisch mit dem älteren, spezielleren Epps–Pulley–Test, dessen  $H_0$ –Quantile  $q_{1.0,n,\alpha}$  nicht nur in Monte–Carlo–Simulationen, sondern auch theoretisch untersucht worden sind und damit als gesichert angenommen werden können (vgl. Henze (1990) und die Ausführungen auf Seite 29 dieser Arbeit).

Die von den zugrundeliegenden Verteilungen abhängigen Parameter  $L_{\beta}$  und  $\tau_{\beta}^2$  sind in den Abschnitten 1.4.1 bis 1.4.6 für die sechs Beispiel–Alternativen angegeben worden; die konkreten numerischen Werte befinden sich unter den Tabellen 1.2 bis 1.7. In den Fällen, in denen die diesen Zahlenwerten zugrundeliegenden Integraldarstellungen aus den Abschnitten 1.4.1 bis 1.4.6 nicht explizit ausgerechnet werden konnten, wurden die numerischen Integrationsroutinen der IMSL–Programmbibliothek verwendet.  $n^{-1}T_{n,\beta}$  bzw.  $\tau_{\beta,n}^2$  bezeichnen die anhand einer Stichprobe vom Umfang  $n$  berechneten Schätzer für  $L_{\beta}$  bzw.  $\tau_{\beta}^2$ , die (wie bereits oben erklärt) jeweils auf 100 000 Monte–Carlo–Wiederholungen beruhen. Die zweite bzw. vierte Spalte in den Tabellen 1.2 bis 1.7 zeigen deutlich die Konvergenz der Schätzwerte  $n^{-1}T_{n,\beta}$  bzw.  $\tau_{\beta,n}^2$  gegen die theoretischen Größen  $L_{\beta}$  bzw.  $\tau_{\beta}^2$ .

Nach Gleichung (1.20) aus Abschnitt 1.3.1 ist ein asymptotischer  $(1 - \alpha)$ –Konfidenzbereich für  $L_{\beta}$  gegeben durch

$$I_{\beta,n,1-\alpha} = \left[ \frac{1}{n}T_{n,\beta} - \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tau_{\beta,n}}{\sqrt{n}}, \frac{1}{n}T_{n,\beta} + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \frac{\tau_{\beta,n}}{\sqrt{n}} \right].$$

Wir wählen  $\alpha = 0.1$  und geben in der dritten Spalte der Tabellen 1.2 bis 1.7 den jeweiligen anhand von  $n^{-1}T_{n,\beta}$  und  $\tau_{\beta,n}^2$  berechneten 90%–Vertrauensbereich  $I_{\beta,n,0.9}$  für  $L_{\beta}$  an. Dabei wurden (bei kleinen Stichprobenumfängen  $n$ ) auftretende negative Werte der unteren Grenze

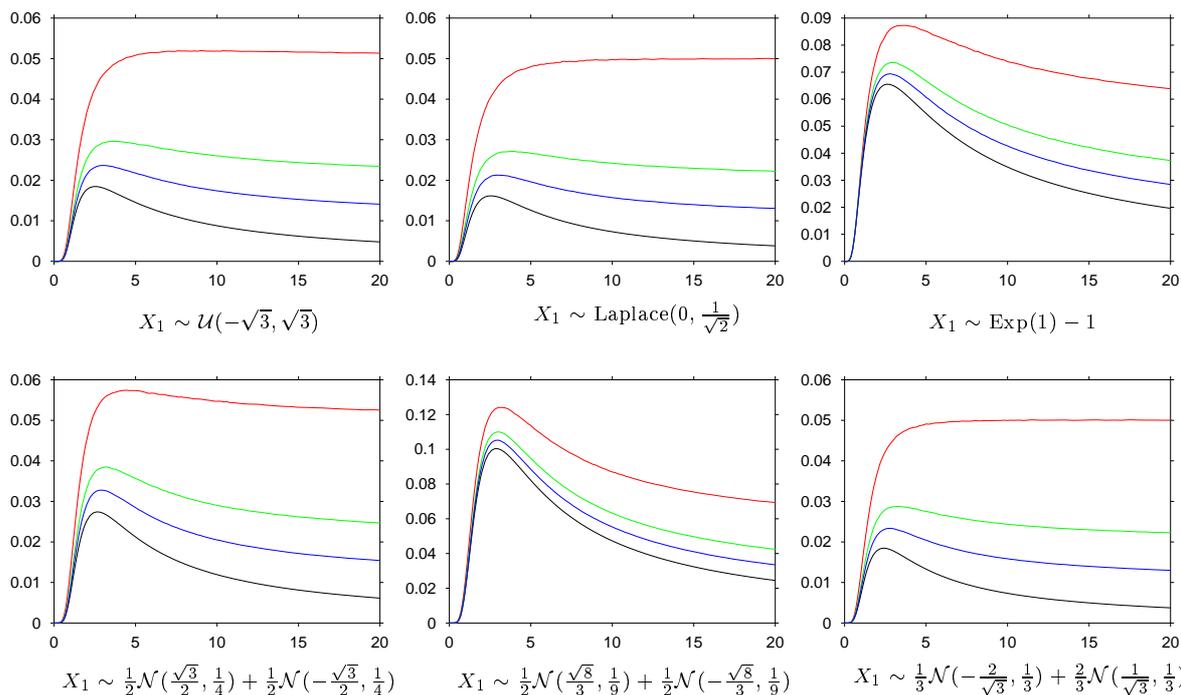


Abbildung 1.2.: Verlauf von  $L_\beta$  und der Näherung  $n^{-1}T_{n,\beta}$  für  $n \in \{20, 50, 100\}$  in Abhängigkeit von  $\beta$  (schwarz:  $L_\beta$ , rot:  $n = 20$ , grün:  $n = 50$ , blau:  $n = 100$ )

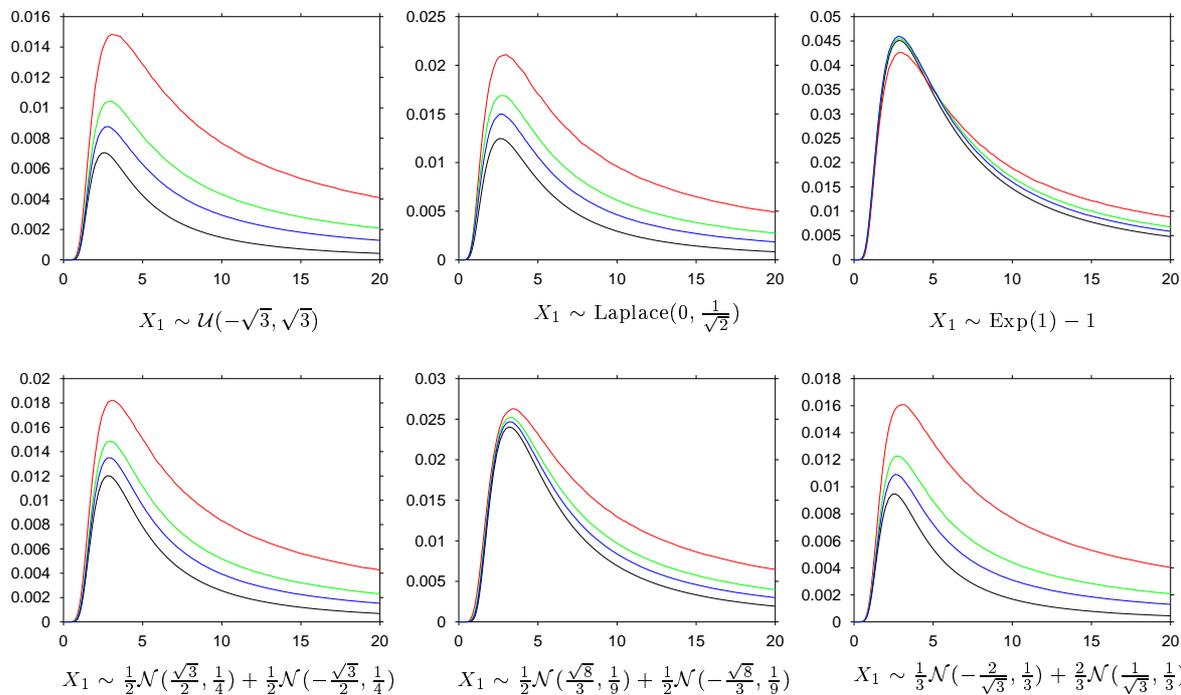


Abbildung 1.3.: Verlauf von  $\tau_\beta^2$  und der Näherung  $\tau_{\beta n}^2$  für  $n \in \{20, 50, 100\}$  in Abhängigkeit von  $\beta$  (schwarz:  $\tau_\beta^2$ , rot:  $n = 20$ , grün:  $n = 50$ , blau:  $n = 100$ )

durch die natürliche untere Schranke 0 für  $L_\beta$  ersetzt. Die Ergebnisse in den Tabellen 1.2 bis 1.7 zeigen allerdings, daß der Vertrauensbereich  $I_{\beta,n,0.9}$  erst bei großen Stichprobenumfängen ( $n \geq 100$ ) brauchbar wird.

Desweiteren sollen die in Abschnitt 1.3.2 vorgeschlagenen Gütenäherungen für den BHEP-Test zum Niveau  $\alpha = 0.1$  untersucht werden. In den unten aufgeführten Tabellen 1.2 bis 1.7 wird zunächst das durch die übliche Monte-Carlo-Simulation ermittelte empirische 90%-Quantil  $q_{\beta,n,0.9}$  von  $T_{n,\beta}$  unter  $H_0$  zum Stichprobenumfang  $n$  angegeben. Jeder dieser Einträge beruht auf 100 000 Monte-Carlo-Wiederholungen und stimmt sehr gut mit den in Tabelle 1.1 auf Seite 29 angegebenen Werten aus Henze (1990) überein. Bei der Auflistung der ermittelten Gütewerte wird zuerst in der Spalte *MC-Güte* der durch die übliche Monte-Carlo-Simulation ermittelte empirische Gütewert zum Stichprobenumfang  $n$  angegeben. Jeder dieser Einträge basiert auf 10 000 Wiederholungen.

Die mit dem in dieser Arbeit vorgestellten neuen Ansatz ermittelten Gütenäherungen bei fester Alternative  $F \in H_1$  sollen in drei Varianten untersucht werden: Bei bekannten  $\tau_\beta^2$  und  $L_\beta$  ist die nach Gleichung (1.21) auf Seite 30 ermittelte Approximation gegeben durch

$$P_F(T_{n,\beta} > q_{\beta,n,0.9}) \approx 1 - \Phi\left(\frac{q_{\beta,n,0.9} - n L_\beta}{\sqrt{n} \tau_\beta}\right);$$

die entsprechenden Einträge in den Tabellen 1.2 bis 1.7 tragen die Überschrift *Güte1*. Ersetzt man den Wert  $\tau_\beta^2$  durch die empirische Größe  $\tau_{\beta,n}^2$ , so erhält man die Gütenäherung

$$P_F(T_{n,\beta} > q_{\beta,n,0.9}) \approx 1 - \Phi\left(\frac{q_{\beta,n,0.9} - n L_\beta}{\sqrt{n} \tau_{\beta,n}}\right),$$

die in den Spalten *Güte2* festgehalten ist. Ersetzt man überdies auch  $L_\beta$  durch den Schätzwert  $n^{-1}T_{n,\beta}$ , gelangt man zu

$$P_F(T_{n,\beta} > q_{\beta,n,0.9}) \approx 1 - \Phi\left(\frac{q_{\beta,n,0.9} - T_{n,\beta}}{\sqrt{n} \tau_{\beta,n}}\right)$$

und den Einträgen der Spalten *Güte3*.

Die Abbildungen 1.4 bis 1.9 zeigen den Verlauf der Gütenäherungen MC-Güte, Güte1, Güte2 und Güte3 in Abhängigkeit von  $n$  für die Werte  $\beta \in \{0.8, 1.0, 3.6\}$  des Gewichtsparameters. Sowohl anhand der Tabellen 1.2 bis 1.7 als auch anhand der Kurven 1.4 bis 1.9 sieht man, daß die dritte Näherung Güte3 am besten mit der Monte-Carlo-Güte übereinstimmt. Ferner zeigen die Kurven, daß der Wert  $\beta = 1.0$  in allen sechs Beispielen mit die besten Näherungen liefert. Schon für  $\beta = 0.8$  entfernen sich die vier Näherungen insbesondere bei kleinen Stichprobenumfängen voneinander; für größere Werte von  $\beta$  ( $\beta \geq 3.6$ ) werden die Ergebnisse in den Fällen der Gleichverteilung, der ersten und dritten Normal Mixture deutlich schlechter, im Fall der zweiten Normal Mixture jedoch etwas besser. Allgemein sieht man, daß sich die Kurven von Güte1 und Güte2 kaum unterscheiden, daß also das Ersetzen von  $\tau_\beta^2$  durch die empirische Größe  $\tau_{\beta,n}^2$  fast keinen Unterschied bewirkt. Ferner verlaufen beide Kurven unterhalb der Monte-Carlo-Güte, d.h. die Approximationen Güte1 und Güte2 liefern in allen untersuchten Beispielen eine konservative Gütenäherung (d.h. eine untere Schranke). Im Fall der Laplace-Verteilung, die

unter den betrachteten Beispiel-Alternativen der Normalverteilung am nächsten kommt, unterscheidet sich der Verlauf der Monte-Carlo-Güte am deutlichsten von den drei in dieser Arbeit ermittelten Gütenäherungen. Dieser Effekt wird erst bei großen Werten des Glättungsparameters ( $\beta = 3.6$ ) abgemildert. Bei den Verteilungen, die sich stärker von der Normalverteilung abheben, stimmen die unterschiedlichen Näherungen besser überein.

Schließlich sind in Tabelle 1.8 die Ergebnisse des in Abschnitt 1.3.3 beschriebenen inversen Tests auf Normalverteilung zum Niveau  $\alpha = 0.1$  und für die Wahl  $\Delta = 0.006$  angegeben. Die Ergebnisse zeigen einerseits die Abhängigkeit des Testergebnisses vom Gewichtsparameter  $\beta$ , und andererseits wird deutlich, daß man ohne eine genauere Interpretation des Abstandes  $L_\beta$  und damit des tolerierbaren Wertes  $\Delta$  nicht weiterkommt.

$n$	$n^{-1}T_{n,1.0}$	$I_{1.0,n,0.9}$	$\tau_{1.0,n}^2$	$q_{1.0,n,0.9}$	MC-Güte	Güte1	Güte2	Güte3
10	0.016059	[0.000000, 0.037592]	0.00171386	0.2783	0.13	0.002	0.051	0.184
20	0.011004	[0.000000, 0.023279]	0.00111386	0.2863	0.26	0.064	0.146	0.329
30	0.009449	[0.000351, 0.018547]	0.00091787	0.2860	0.44	0.234	0.290	0.494
40	0.008684	[0.001240, 0.016128]	0.00081919	0.2884	0.60	0.419	0.435	0.628
50	0.008237	[0.001822, 0.014651]	0.00076049	0.2892	0.74	0.583	0.569	0.735
60	0.007949	[0.002243, 0.013655]	0.00072208	0.2892	0.83	0.710	0.683	0.816
70	0.007722	[0.002541, 0.012903]	0.00069449	0.2912	0.89	0.798	0.768	0.871
80	0.007567	[0.002790, 0.012344]	0.00067473	0.2898	0.94	0.865	0.836	0.913
90	0.007458	[0.003002, 0.011914]	0.00066055	0.2919	0.96	0.907	0.883	0.940
100	0.007343	[0.003163, 0.011523]	0.00064583	0.2887	0.98	0.939	0.921	0.960
200	0.006902	[0.004077, 0.009727]	0.00058991	0.2914	1.00	0.999	0.998	0.999
300	0.006730	[0.004463, 0.008996]	0.00056966	0.2927	1.00	1.000	1.000	1.000
400	0.006680	[0.004733, 0.008627]	0.00056047	0.2893	1.00	1.000	1.000	1.000
500	0.006632	[0.004898, 0.008366]	0.00055563	0.2905	1.00	1.000	1.000	1.000
1000	0.006548	[0.005334, 0.007762]	0.00054470	0.2919	1.00	1.000	1.000	1.000

Tabelle 1.2.:  $X_1 \sim U(-\sqrt{3}, \sqrt{3})$ ,  $L_{1.0} = 0.00646666$ ,  $\tau_{1.0}^2 = 0.00053391$

$n$	$n^{-1}T_{n,1.0}$	$I_{1.0,n,0.9}$	$\tau_{1.0,n}^2$	$q_{1.0,n,0.9}$	MC-Güte	Güte1	Güte2	Güte3
10	0.019797	[0.000000, 0.048229]	0.00298791	0.2783	0.24	0.082	0.110	0.321
20	0.013819	[0.000000, 0.037512]	0.00414984	0.2863	0.36	0.237	0.296	0.486
30	0.011507	[0.000000, 0.030822]	0.00413676	0.2860	0.47	0.369	0.401	0.567
40	0.010311	[0.000000, 0.026666]	0.00395445	0.2884	0.54	0.468	0.476	0.622
50	0.009616	[0.000000, 0.023945]	0.00379438	0.2892	0.63	0.548	0.537	0.670
60	0.009115	[0.000000, 0.021938]	0.00364666	0.2892	0.68	0.613	0.590	0.709
70	0.008793	[0.000000, 0.020480]	0.00353410	0.2912	0.74	0.664	0.634	0.743
80	0.008500	[0.000000, 0.019244]	0.00341296	0.2898	0.80	0.710	0.676	0.772
90	0.008300	[0.000000, 0.018292]	0.00332093	0.2919	0.83	0.746	0.710	0.797
100	0.008150	[0.000000, 0.017532]	0.00325295	0.2887	0.87	0.780	0.743	0.822
200	0.007370	[0.001171, 0.013568]	0.00284008	0.2914	0.99	0.935	0.914	0.942
300	0.007117	[0.002193, 0.012042]	0.00268941	0.2927	1.00	0.979	0.970	0.980
400	0.006964	[0.002778, 0.011150]	0.00259030	0.2893	1.00	0.993	0.990	0.993
500	0.006930	[0.003210, 0.010649]	0.00255637	0.2905	1.00	0.997	0.996	0.998
1000	0.006751	[0.004184, 0.009318]	0.00243480	0.2919	1.00	1.000	1.000	1.000

Tabelle 1.3.:  $X_1 \sim \text{Laplace}(0, \frac{1}{\sqrt{2}})$ ,  $L_{1.0} = 0.00660203$ ,  $\tau_{1.0}^2 = 0.00231562$

$n$	$n^{-1}T_{n,1.0}$	$I_{1.0,n,0.9}$	$\tau_{1.0,n}^2$	$q_{1.0,n,0.9}$	MC-Güte	Güte1	Güte2	Güte3
10	0.035915	[0.000015, 0.071815]	0.00476354	0.2783	0.54	0.544	0.560	0.644
20	0.034465	[0.001577, 0.067352]	0.00799552	0.2863	0.85	0.789	0.799	0.843
30	0.033638	[0.005031, 0.062245]	0.00907425	0.2860	0.97	0.897	0.893	0.917
40	0.032956	[0.007715, 0.058196]	0.00941903	0.2884	0.99	0.947	0.940	0.953
50	0.032687	[0.009931, 0.055443]	0.00956983	0.2892	1.00	0.972	0.966	0.974
60	0.032531	[0.011664, 0.053398]	0.00965652	0.2892	1.00	0.985	0.981	0.986
70	0.032334	[0.013004, 0.051664]	0.00966725	0.2912	1.00	0.992	0.989	0.992
80	0.032141	[0.014084, 0.050199]	0.00964155	0.2898	1.00	0.996	0.994	0.995
90	0.032055	[0.015053, 0.049057]	0.00961591	0.2919	1.00	0.998	0.996	0.997
100	0.031943	[0.015844, 0.048042]	0.00957931	0.2887	1.00	0.999	0.998	0.999
200	0.031574	[0.020314, 0.042835]	0.00937295	0.2914	1.00	1.000	1.000	1.000
300	0.031409	[0.022294, 0.040525]	0.00921435	0.2927	1.00	1.000	1.000	1.000
400	0.031344	[0.023496, 0.039192]	0.00910578	0.2893	1.00	1.000	1.000	1.000
500	0.031323	[0.024299, 0.038346]	0.00911649	0.2905	1.00	1.000	1.000	1.000
1000	0.031214	[0.026308, 0.036121]	0.00889768	0.2919	1.00	1.000	1.000	1.000

Tabelle 1.4.:  $X_1 \sim \text{Exp}(1) - 1$ ,  $L_{1.0} = 0.03110125$ ,  $\tau_{1.0}^2 = 0.00873219$

$n$	$n^{-1}T_{n,1.0}$	$I_{1.0,n,0.9}$	$\tau_{1.0,n}^2$	$q_{1.0,n,0.9}$	MC-Güte	Güte1	Güte2	Güte3
10	0.018581	[0.000000, 0.041716]	0.00197817	0.2783	0.18	0.003	0.071	0.255
20	0.012866	[0.000000, 0.026179]	0.00131015	0.2863	0.37	0.093	0.188	0.429
30	0.010980	[0.001126, 0.020834]	0.00107664	0.2860	0.57	0.294	0.345	0.595
40	0.010020	[0.002005, 0.018035]	0.00094985	0.2884	0.71	0.493	0.494	0.718
50	0.009442	[0.002551, 0.016334]	0.00087769	0.2892	0.82	0.654	0.627	0.809
60	0.009058	[0.002946, 0.015170]	0.00082836	0.2892	0.90	0.771	0.734	0.873
70	0.008793	[0.003248, 0.014338]	0.00079541	0.2912	0.95	0.849	0.812	0.915
80	0.008584	[0.003482, 0.013687]	0.00076975	0.2898	0.97	0.903	0.872	0.945
90	0.008411	[0.003671, 0.013150]	0.00074731	0.2919	0.98	0.937	0.912	0.964
100	0.008286	[0.003840, 0.012733]	0.00073069	0.2887	0.99	0.961	0.942	0.977
200	0.007712	[0.004728, 0.010696]	0.00065829	0.2914	1.00	1.000	0.999	1.000
300	0.007508	[0.005117, 0.009899]	0.00063390	0.2927	1.00	1.000	1.000	1.000
400	0.007455	[0.005401, 0.009509]	0.00062368	0.2893	1.00	1.000	1.000	1.000
500	0.007364	[0.005540, 0.009188]	0.00061467	0.2905	1.00	1.000	1.000	1.000
1000	0.007254	[0.005979, 0.008528]	0.00060057	0.2919	1.00	1.000	1.000	1.000

Tabelle 1.5.:  $X_1 \sim \frac{1}{2}N(\frac{\sqrt{3}}{2}, \frac{1}{4}) + \frac{1}{2}N(-\frac{\sqrt{3}}{2}, \frac{1}{4})$ ,  $L_{1.0} = 0.00714241$ ,  $\tau_{1.0}^2 = 0.00058621$

$n$	$n^{-1}T_{n,1.0}$	$I_{1.0,n,0.9}$	$\tau_{1.0,n}^2$	$q_{1.0,n,0.9}$	MC-Güte	Güte1	Güte2	Güte3
10	0.032367	[0.004180, 0.060553]	0.00293651	0.2783	0.59	0.154	0.325	0.604
20	0.026333	[0.010161, 0.042504]	0.00193329	0.2863	0.96	0.855	0.720	0.889
30	0.024256	[0.012532, 0.035981]	0.00152423	0.2860	1.00	0.991	0.930	0.981
40	0.023216	[0.013835, 0.032597]	0.00130110	0.2884	1.00	1.000	0.988	0.997
50	0.022584	[0.014627, 0.030541]	0.00116994	0.2892	1.00	1.000	0.998	1.000
60	0.022156	[0.015189, 0.029122]	0.00107630	0.2892	1.00	1.000	1.000	1.000
70	0.021863	[0.015617, 0.028109]	0.00100928	0.2912	1.00	1.000	1.000	1.000
80	0.021631	[0.015944, 0.027318]	0.00095627	0.2898	1.00	1.000	1.000	1.000
90	0.021462	[0.016213, 0.026711]	0.00091640	0.2919	1.00	1.000	1.000	1.000
100	0.021315	[0.016418, 0.026211]	0.00088617	0.2887	1.00	1.000	1.000	1.000
200	0.020687	[0.017531, 0.023842]	0.00073612	0.2914	1.00	1.000	1.000	1.000
300	0.020449	[0.017963, 0.022935]	0.00068536	0.2927	1.00	1.000	1.000	1.000
400	0.020363	[0.018247, 0.022478]	0.00066160	0.2893	1.00	1.000	1.000	1.000
500	0.020292	[0.018421, 0.022163]	0.00064676	0.2905	1.00	1.000	1.000	1.000
1000	0.020187	[0.018897, 0.021477]	0.00061515	0.2919	1.00	1.000	1.000	1.000

Tabelle 1.6.:  $X_1 \sim \frac{1}{2}N(\frac{\sqrt{8}}{3}, \frac{1}{9}) + \frac{1}{2}N(-\frac{\sqrt{8}}{3}, \frac{1}{9})$ ,  $L_{1.0} = 0.02004698$ ,  $\tau_{1.0}^2 = 0.00058511$

$n$	$n^{-1}T_{n,1.0}$	$I_{1.0,n,0.9}$	$\tau_{1.0,n}^2$	$q_{1.0,n,0.9}$	MC-Güte	Güte1	Güte2	Güte3
10	0.017577	[0.000000, 0.041960]	0.00219731	0.2783	0.17	0.030	0.080	0.245
20	0.012257	[0.000000, 0.027838]	0.00179462	0.2863	0.33	0.174	0.219	0.414
30	0.010500	[0.000000, 0.022507]	0.00159867	0.2860	0.48	0.344	0.363	0.553
40	0.009606	[0.000000, 0.019666]	0.00149643	0.2884	0.60	0.483	0.485	0.652
50	0.009095	[0.000257, 0.017933]	0.00144345	0.2892	0.70	0.596	0.588	0.731
60	0.008754	[0.000794, 0.016713]	0.00140506	0.2892	0.79	0.684	0.672	0.792
70	0.008475	[0.001182, 0.015768]	0.00137612	0.2912	0.85	0.750	0.738	0.835
80	0.008294	[0.001520, 0.015068]	0.00135683	0.2898	0.90	0.805	0.793	0.872
90	0.008149	[0.001800, 0.014499]	0.00134116	0.2919	0.93	0.845	0.833	0.898
100	0.008031	[0.002038, 0.014024]	0.00132756	0.2887	0.95	0.879	0.869	0.921
200	0.007497	[0.003348, 0.011647]	0.00127294	0.2914	1.00	0.987	0.986	0.992
300	0.007349	[0.003979, 0.010719]	0.00125937	0.2927	1.00	0.999	0.998	0.999
400	0.007249	[0.004344, 0.010153]	0.00124735	0.2893	1.00	1.000	1.000	1.000
500	0.007179	[0.004589, 0.009770]	0.00123990	0.2905	1.00	1.000	1.000	1.000
1000	0.007091	[0.005266, 0.008916]	0.00123130	0.2919	1.00	1.000	1.000	1.000

Tabelle 1.7.:  $X_1 \sim \frac{1}{3}N(-\frac{2}{\sqrt{3}}, \frac{1}{3}) + \frac{2}{3}N(\frac{1}{\sqrt{3}}, \frac{1}{3})$ ,  $L_{1.0} = 0.00697973$ ,  $\tau_{1.0}^2 = 0.00122091$

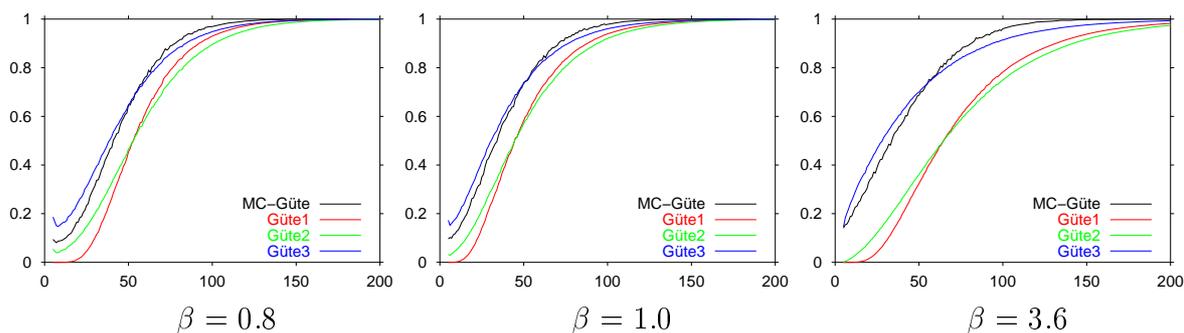


Abbildung 1.4.: Gütenäherung im Fall  $X_1 \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$  in Abhängigkeit von  $n$

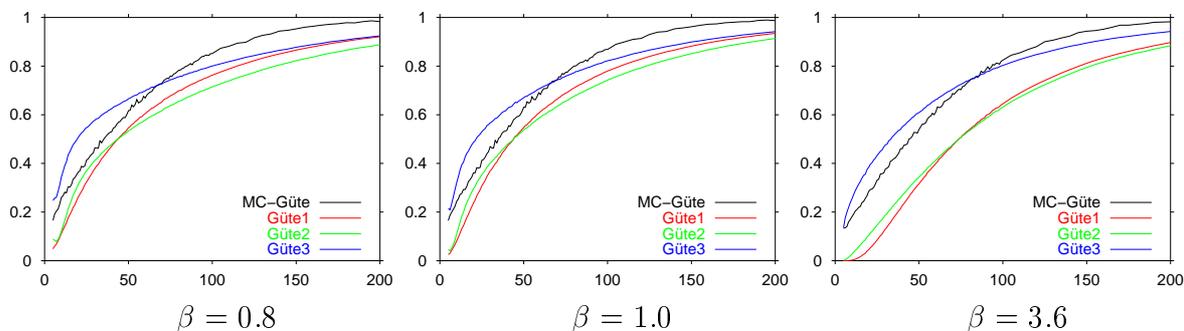


Abbildung 1.5.: Gütenäherung im Fall  $X_1 \sim \text{Laplace}(0, \frac{1}{\sqrt{2}})$  in Abhängigkeit von  $n$

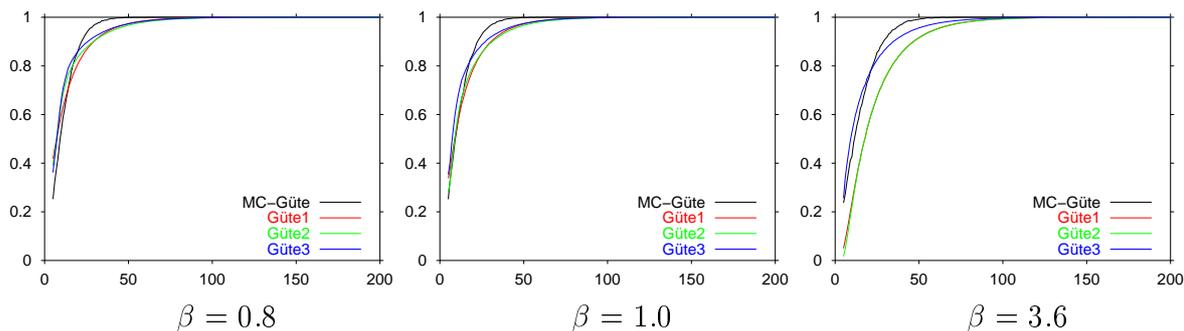


Abbildung 1.6.: Gütenäherung im Fall  $X_1 \sim \text{Exp}(1) - 1$  in Abhängigkeit von  $n$

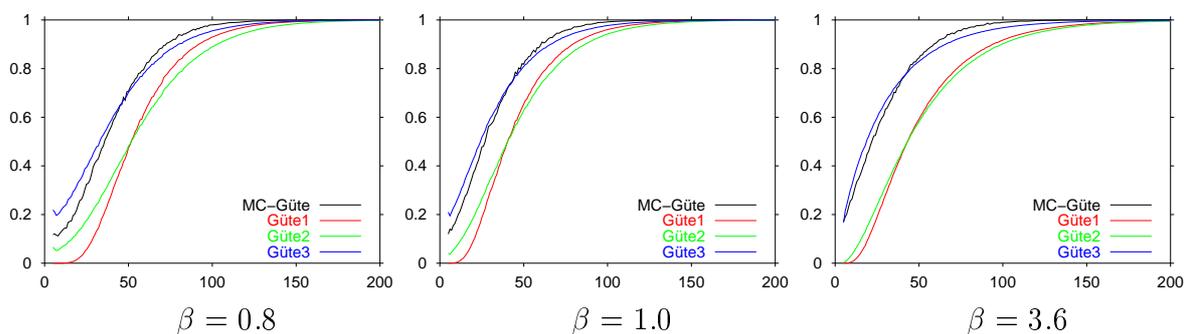


Abbildung 1.7.: Gütenäherung im Fall  $X_1 \sim \frac{1}{2}\mathcal{N}(\frac{\sqrt{3}}{2}, \frac{1}{4}) + \frac{1}{2}\mathcal{N}(-\frac{\sqrt{3}}{2}, \frac{1}{4})$  in Abhängigkeit von  $n$

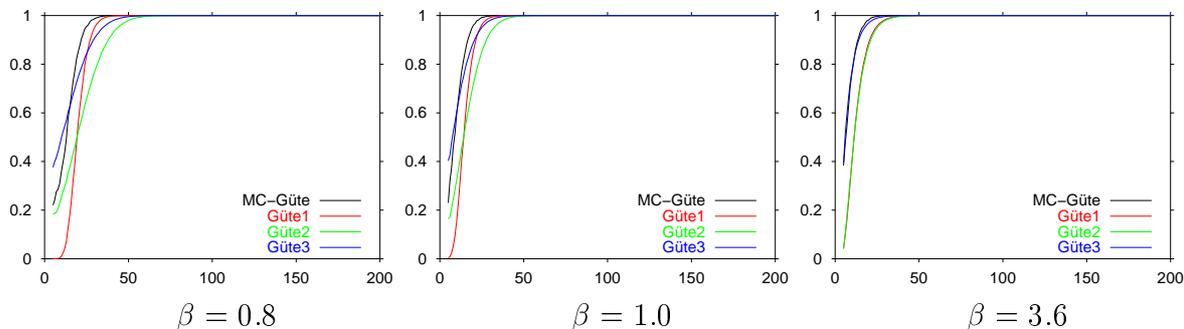


Abbildung 1.8.: Gütenäherung im Fall  $X_1 \sim \frac{1}{2}\mathcal{N}(\frac{\sqrt{8}}{3}, \frac{1}{9}) + \frac{1}{2}\mathcal{N}(-\frac{\sqrt{8}}{3}, \frac{1}{9})$  in Abhängigkeit von  $n$

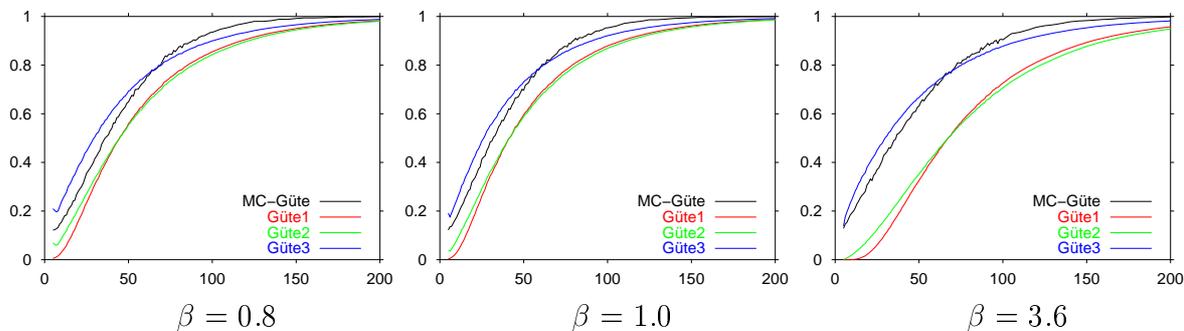


Abbildung 1.9.: Gütenäherung im Fall  $X_1 \sim \frac{1}{3}\mathcal{N}(-\frac{2}{\sqrt{3}}, \frac{1}{3}) + \frac{2}{3}\mathcal{N}(\frac{1}{\sqrt{3}}, \frac{1}{3})$  in Abhängigkeit von  $n$

$U(-\sqrt{3}, \sqrt{3})$					Laplace(0, 1/ $\sqrt{2}$ )					Exp(1) - 1				
$n$	0.5	0.8	1.0	1.2	$n$	0.5	0.8	1.0	1.2	$n$	0.5	0.8	1.0	1.2
10	95.6	8.3	1.5	0.3	10	79.4	13.6	3.3	0.8	10	53.6	2.1	0.4	0.1
20	99.8	20.6	3.2	1.0	20	84.4	23.3	8.2	3.0	20	46.4	1.1	0.2	0.0
30	100.0	31.7	4.0	1.2	30	87.5	26.5	10.7	4.6	30	43.3	0.5	0.1	0.0
40	100.0	42.2	4.3	1.2	40	90.1	28.5	11.6	5.3	40	41.7	0.2	0.0	0.0
50	100.0	51.6	4.4	1.2	50	92.2	29.9	11.9	5.3	50	40.7	0.1	0.0	0.0
60	100.0	59.9	4.5	0.9	60	93.9	31.0	11.9	5.5	60	40.6	0.0	0.0	0.0
70	100.0	66.8	4.5	0.8	70	95.3	32.2	11.7	5.3	70	40.5	0.0	0.0	0.0
80	100.0	73.0	4.6	0.7	80	96.3	33.3	11.6	5.0	80	40.6	0.0	0.0	0.0
90	100.0	77.9	4.5	0.6	90	97.1	34.2	11.5	4.8	90	40.9	0.0	0.0	0.0
100	100.0	82.0	4.6	0.6	100	97.7	35.2	11.4	4.4	100	41.3	0.0	0.0	0.0
200	100.0	98.2	4.3	0.1	200	99.8	44.1	9.8	2.6	200	47.0	0.0	0.0	0.0
300	100.0	99.8	4.0	0.1	300	100.0	51.6	8.6	1.3	300	54.3	0.0	0.0	0.0
400	100.0	100.0	3.5	0.0	400	100.0	57.2	8.5	1.0	400	59.8	0.0	0.0	0.0
500	100.0	100.0	3.2	0.0	500	100.0	62.3	7.5	0.8	500	64.8	0.0	0.0	0.0
1000	100.0	100.0	2.4	0.0	1000	100.0	82.2	6.0	0.1	1000	84.2	0.0	0.0	0.0

Mix1					Mix2					Mix3				
$n$	0.5	0.8	1.0	1.2	$n$	0.5	0.8	1.0	1.2	$n$	0.5	0.8	1.0	1.2
10	93.5	6.1	1.0	0.2	10	83.7	0.2	0.0	0.0	10	90.1	9.2	1.9	0.4
20	99.6	14.9	1.9	0.5	20	97.9	0.1	0.0	0.0	20	97.7	19.9	4.4	1.6
30	100.0	24.5	2.1	0.5	30	99.7	0.1	0.0	0.0	30	99.5	26.6	5.5	2.0
40	100.0	33.6	2.2	0.5	40	100.0	0.0	0.0	0.0	40	99.9	30.5	6.0	2.2
50	100.0	42.4	2.2	0.4	50	100.0	0.0	0.0	0.0	50	100.0	34.4	6.0	1.9
60	100.0	50.4	2.1	0.3	60	100.0	0.0	0.0	0.0	60	100.0	37.1	6.1	1.8
70	100.0	57.8	2.1	0.2	70	100.0	0.0	0.0	0.0	70	100.0	39.7	6.0	1.5
80	100.0	64.0	2.1	0.2	80	100.0	0.0	0.0	0.0	80	100.0	41.9	6.0	1.4
90	100.0	69.8	2.0	0.1	90	100.0	0.0	0.0	0.0	90	100.0	43.8	5.8	1.2
100	100.0	74.8	2.0	0.1	100	100.0	0.0	0.0	0.0	100	100.0	46.0	5.7	1.2
200	100.0	96.5	1.4	0.0	200	100.0	0.0	0.0	0.0	200	100.0	63.1	4.5	0.4
300	100.0	99.6	0.9	0.0	300	100.0	0.0	0.0	0.0	300	100.0	75.3	3.4	0.2
400	100.0	100.0	0.8	0.0	400	100.0	0.0	0.0	0.0	400	100.0	82.8	3.3	0.0
500	100.0	100.0	0.7	0.0	500	100.0	0.0	0.0	0.0	500	100.0	88.8	2.8	0.0
1000	100.0	100.0	0.1	0.0	1000	100.0	0.0	0.0	0.0	1000	100.0	98.5	1.1	0.0

Tabelle 1.8.: Ablehnhäufigkeit (in %) des inversen Tests auf Normalverteilung zum Niveau  $\alpha = 0.1$  und mit der Wahl  $\Delta = 0.006$  für die sechs Beispielverteilungen und Werte  $\beta \in \{0.5, 0.8, 1.0, 1.2\}$  des Gewichtsparameters

## 1.5. Technische Lemmata und Beweise

**Lemma 18** Für beliebige Vektoren  $a, b \in \mathbb{R}^d$  und für  $\beta > 0$  gilt:

$$\begin{aligned} I_1(a) &:= \int_{\mathbb{R}^d} \cos(a^T t) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt = \frac{1}{(1+\beta^2)^{d/2}} e^{-\frac{\beta^2}{2(1+\beta^2)} \|a\|^2}, \\ I_2(a, b) &:= \int_{\mathbb{R}^d} \cos(a^T t) \cos(b^T t) \varphi_\beta(t) dt = \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|a-b\|^2} + e^{-\frac{\beta^2}{2} \|a+b\|^2} \right], \\ I_3(a, b) &:= \int_{\mathbb{R}^d} \sin(a^T t) \sin(b^T t) \varphi_\beta(t) dt = \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|a-b\|^2} - e^{-\frac{\beta^2}{2} \|a+b\|^2} \right], \\ I_4(a, b) &:= \int_{\mathbb{R}^d} t \cos(a^T t) \sin(b^T t) \varphi_\beta(t) dt = \frac{\beta^2}{2} \left[ (b-a) e^{-\frac{\beta^2}{2} \|b-a\|^2} + (b+a) e^{-\frac{\beta^2}{2} \|b+a\|^2} \right], \\ I_5'(a) &:= \int_{\mathbb{R}^d} t \sin(a^T t) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt = a \frac{\beta^2}{(1+\beta^2)^{\frac{d}{2}+1}} e^{-\frac{\beta^2}{2(1+\beta^2)} \|a\|^2}. \end{aligned}$$

Man beachte, daß  $I_1, I_2$  bzw.  $I_3$  Skalare sind und  $I_4$  bzw.  $I_5'$  Vektoren.

BEWEIS: Seien im folgenden  $a, b \in \mathbb{R}^d$  und  $\beta > 0$ . Die Darstellungen von  $I_2(a, b), I_3(a, b), I_4'(a, b)$  und  $I_5'(a)$  werden zunächst für die Dimension  $d = 1$  nachgewiesen und dann per Induktion auf  $d > 1$  verallgemeinert, wobei im Induktionsschritt einige der Integrale aus Symmetriegründen wegfallen. Wir verwenden des öfteren die Additionstheoreme der trigonometrischen Funktionen

$$\begin{aligned} \cos(x+y) &= \cos(x)\cos(y) - \sin(x)\sin(y), \\ \sin(x+y) &= \sin(x)\cos(y) + \cos(x)\sin(y) \end{aligned} \quad (1.29)$$

und schreiben  $\tilde{a}, \tilde{b}, \tilde{t}$  für die ersten  $d-1$  Komponenten der Vektoren  $a, b, t$  sowie  $a_d, b_d, t_d$  für die letzte Komponente, in Vektorschreibweise also  $a = \begin{pmatrix} \tilde{a} \\ a_d \end{pmatrix}, b = \begin{pmatrix} \tilde{b} \\ b_d \end{pmatrix}, t = \begin{pmatrix} \tilde{t} \\ t_d \end{pmatrix}$ . Mit

$$\varphi_{\beta,d}(t) := \frac{1}{(2\pi\beta^2)^{d/2}} e^{-\frac{\|t\|^2}{2\beta^2}}$$

gilt  $\varphi_\beta(t) = \varphi_{\beta,d}(t) = \varphi_{\beta,d-1}(\tilde{t}) \cdot \varphi_{\beta,1}(t_d)$ .

NACHWEIS DER DARSTELLUNG VON  $I_1(a)$ :

Es gilt  $\int_{\mathbb{R}^d} \cos(a^T t) \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|t\|^2}{2}} dt = e^{-\frac{\|a\|^2}{2}}$  (Realteil der charakteristischen Funktion). Substituieren der Parameter liefert die Darstellung von  $I_1(a)$ .

NACHWEIS DER DARSTELLUNG VON  $I_2(a, b)$  UND  $I_3(a, b)$ :

Nach Formel Nr. 3.898 aus Gradshteyn und Ryzhik (1980) gilt im Eindimensionalen für  $\beta > 0$ :

$$\int_0^\infty \cos(at) \cos(bt) e^{-\beta t^2} dt = \frac{1}{4} \sqrt{\frac{\pi}{\beta}} \left( e^{-\frac{(a-b)^2}{4\beta}} + e^{-\frac{(a+b)^2}{4\beta}} \right), \quad (1.30)$$

$$\int_0^{\infty} \sin(at) \sin(bt) e^{-\beta t^2} dt = \frac{1}{4} \sqrt{\frac{\pi}{\beta}} \left( e^{-\frac{(a-b)^2}{4\beta}} - e^{-\frac{(a+b)^2}{4\beta}} \right),$$

und der Wert von  $I_2(a, b)$  bzw.  $I_3(a, b)$  ergibt sich durch Einsetzen der Konstanten und Ausnutzen von Symmetrien. Der Nachweis für Dimensionen  $d > 1$  erfolgt durch Induktion parallel über die Darstellungen von  $I_2(a, b)$  und  $I_3(a, b)$ . Sind diese im  $(d - 1)$ -Dimensionalen gültig, so folgt für  $I_2(a, b)$  mit Hilfe der Additionstheoreme (1.29):

$$\begin{aligned} & \int_{\mathbb{R}^d} \cos(a^T t) \cos(b^T t) \varphi_{\beta}(t) dt \\ &= \int_{\mathbb{R}^d} \left[ \cos(\tilde{a}^T \tilde{t}) \cos(a_d t_d) \cos(\tilde{b}^T \tilde{t}) \cos(b_d t_d) + \sin(\tilde{a}^T \tilde{t}) \sin(a_d t_d) \sin(\tilde{b}^T \tilde{t}) \sin(b_d t_d) \right. \\ & \quad \left. - \cos(\tilde{a}^T \tilde{t}) \cos(a_d t_d) \sin(\tilde{b}^T \tilde{t}) \sin(b_d t_d) - \sin(\tilde{a}^T \tilde{t}) \sin(a_d t_d) \cos(\tilde{b}^T \tilde{t}) \cos(b_d t_d) \right] \varphi_{\beta}(t) dt \\ &= \left( \int_{\mathbb{R}^{d-1}} \cos(\tilde{a}^T \tilde{t}) \cos(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \right) \left( \int_{\mathbb{R}} \cos(a_d t_d) \cos(b_d t_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\ & \quad + \left( \int_{\mathbb{R}^{d-1}} \sin(\tilde{a}^T \tilde{t}) \sin(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \right) \left( \int_{\mathbb{R}} \sin(a_d t_d) \sin(b_d t_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\ &= \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|\tilde{a} - \tilde{b}\|^2} + e^{-\frac{\beta^2}{2} \|\tilde{a} + \tilde{b}\|^2} \right] \cdot \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} (a_d - b_d)^2} + e^{-\frac{\beta^2}{2} (a_d + b_d)^2} \right] \\ & \quad + \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|\tilde{a} - \tilde{b}\|^2} - e^{-\frac{\beta^2}{2} \|\tilde{a} + \tilde{b}\|^2} \right] \cdot \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} (a_d - b_d)^2} - e^{-\frac{\beta^2}{2} (a_d + b_d)^2} \right] \\ &= \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|a - b\|^2} + e^{-\frac{\beta^2}{2} \|a + b\|^2} \right]. \end{aligned}$$

Die Rechnung für  $I_3(a, b)$  verläuft analog.

NACHWEIS DER DARSTELLUNG VON  $I_4'(a, b)$ :

Der Wert von  $I_4'(a, b)$  im Eindimensionalen ergibt sich durch partielle Integration und Einsetzen der Darstellungen von  $I_2(a, b)$  und  $I_3(a, b)$ . Die  $d$ -dimensionale Verallgemeinerung erfolgt wiederum durch Induktion über  $d$ . Setzt man die  $(d - 1)$ -dimensionale Gültigkeit der Darstellung von  $I_4'(a, b)$  voraus, ergibt sich der Induktionsschritt durch Anwenden der Additionstheoreme (1.29) und Einsetzen der Induktionsvoraussetzung. Die aufgeführten Vektoren enthalten in der ersten Zeile die ersten  $d - 1$  Koordinaten und in der zweiten Zeile die  $d$ -te:

$$\begin{aligned} & \int_{\mathbb{R}^d} t \cos(a^T t) \sin(b^T t) \varphi_{\beta}(t) dt \\ &= \left( \int_{\mathbb{R}^{d-1}} \tilde{t} \cos(\tilde{a}^T \tilde{t}) \sin(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} \cos(a_d t_d) \cos(b_d t_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\ & \quad + \left( \int_{\mathbb{R}^{d-1}} \cos(\tilde{a}^T \tilde{t}) \sin(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} t_d \cos(a_d t_d) \cos(b_d t_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\ & \quad + \left( \int_{\mathbb{R}^{d-1}} \tilde{t} \cos(\tilde{a}^T \tilde{t}) \cos(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} \cos(a_d t_d) \sin(b_d t_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\ & \quad + \left( \int_{\mathbb{R}^{d-1}} \cos(\tilde{a}^T \tilde{t}) \cos(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} t_d \cos(a_d t_d) \sin(b_d t_d) \varphi_{\beta, 1}(t_d) dt_d \right) \end{aligned}$$

$$\begin{aligned}
& - \left( \int_{\mathbb{R}^{d-1}} \tilde{t} \sin(\tilde{a}^T \tilde{t}) \sin(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} \sin(at_d) \cos(bt_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\
& - \left( \int_{\mathbb{R}^{d-1}} \sin(\tilde{a}^T \tilde{t}) \sin(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} t_d \sin(at_d) \cos(bt_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\
& - \left( \int_{\mathbb{R}^{d-1}} \tilde{t} \sin(\tilde{a}^T \tilde{t}) \cos(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} \sin(at_d) \sin(bt_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\
& - \left( \int_{\mathbb{R}^{d-1}} \sin(\tilde{a}^T \tilde{t}) \cos(\tilde{b}^T \tilde{t}) \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} t_d \sin(at_d) \sin(bt_d) \varphi_{\beta, 1}(t_d) dt_d \right) \\
& = \left( \frac{\beta^2}{2} \left[ (\tilde{b} - \tilde{a}) e^{-\frac{\beta^2}{2} \|\tilde{b} - \tilde{a}\|^2} + (\tilde{b} + \tilde{a}) e^{-\frac{\beta^2}{2} \|\tilde{b} + \tilde{a}\|^2} \right] \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} (b_d - a_d)^2} + e^{-\frac{\beta^2}{2} (b_d + a_d)^2} \right] \right) \\
& \quad \left( \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|\tilde{b} - \tilde{a}\|^2} + e^{-\frac{\beta^2}{2} \|\tilde{b} + \tilde{a}\|^2} \right] \frac{\beta^2}{2} \left[ (b_d - a_d) e^{-\frac{\beta^2}{2} (b_d - a_d)^2} + (b_d + a_d) e^{-\frac{\beta^2}{2} (b_d + a_d)^2} \right] \right) \\
& - \left( \frac{\beta^2}{2} \left[ (\tilde{a} - \tilde{b}) e^{-\frac{\beta^2}{2} \|\tilde{b} - \tilde{a}\|^2} + (\tilde{b} + \tilde{a}) e^{-\frac{\beta^2}{2} \|\tilde{b} + \tilde{a}\|^2} \right] \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} (b_d - a_d)^2} - e^{-\frac{\beta^2}{2} (b_d + a_d)^2} \right] \right) \\
& \quad \left( \frac{1}{2} \left[ e^{-\frac{\beta^2}{2} \|\tilde{b} - \tilde{a}\|^2} - e^{-\frac{\beta^2}{2} \|\tilde{b} + \tilde{a}\|^2} \right] \frac{\beta^2}{2} \left[ (a_d - b_d) e^{-\frac{\beta^2}{2} (b_d - a_d)^2} + (b_d + a_d) e^{-\frac{\beta^2}{2} (b_d + a_d)^2} \right] \right) \\
& = \left( \frac{\beta^2}{4} \left\{ 2 (\tilde{b} - \tilde{a}) e^{-\frac{\beta^2}{2} \|b-a\|^2} + 2 (\tilde{b} + \tilde{a}) e^{-\frac{\beta^2}{2} \|b+a\|^2} \right\} \right) \\
& \quad \left( \frac{\beta^2}{4} \left\{ 2 (b_d - a_d) e^{-\frac{\beta^2}{2} \|b-a\|^2} + 2 (b_d + a_d) e^{-\frac{\beta^2}{2} \|b+a\|^2} \right\} \right) \\
& = \frac{\beta^2}{2} \left[ (b-a) e^{-\frac{\beta^2}{2} \|b-a\|^2} + (b+a) e^{-\frac{\beta^2}{2} \|b+a\|^2} \right].
\end{aligned}$$

NACHWEIS DER DARSTELLUNG VON  $I'_5(a)$ :

Der Wert von  $I'_5(a)$  im Eindimensionalen ergibt sich durch partielle Integration und Einsetzen von  $I_1(a)$ . Die  $d$ -dimensionale Verallgemeinerung erfolgt wiederum durch Induktion über  $d$ . Die aufgeführten Vektoren enthalten in der ersten Zeile die ersten  $d-1$  Koordinaten und in der zweiten Zeile die  $d$ -te. Setzt man die  $(d-1)$ -dimensionale Gültigkeit der Darstellung von  $I'_5(a, b)$  voraus, ergibt sich wiederum durch Anwenden der Additionstheoreme (1.29) und Einsetzen der Induktionsvoraussetzung:

$$\begin{aligned}
& \int_{\mathbb{R}^d} t \sin(a^T t) e^{-\frac{\|t\|^2}{2}} \varphi_{\beta}(t) dt \\
& = \left( \int_{\mathbb{R}^{d-1}} \tilde{t} \sin(\tilde{a}^T \tilde{t}) e^{-\frac{\|\tilde{t}\|^2}{2}} \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} \cos(at_d) e^{-\frac{t_d^2}{2}} \varphi_{\beta, 1}(t_d) dt_d \right) \\
& \quad \left( \int_{\mathbb{R}^{d-1}} \sin(\tilde{a}^T \tilde{t}) e^{-\frac{\|\tilde{t}\|^2}{2}} \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} t_d \cos(at_d) e^{-\frac{t_d^2}{2}} \varphi_{\beta, 1}(t_d) dt_d \right) \\
& \quad + \left( \int_{\mathbb{R}^{d-1}} \tilde{t} \cos(\tilde{a}^T \tilde{t}) e^{-\frac{\|\tilde{t}\|^2}{2}} \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} \sin(at_d) e^{-\frac{t_d^2}{2}} \varphi_{\beta, 1}(t_d) dt_d \right) \\
& \quad + \left( \int_{\mathbb{R}^{d-1}} \cos(\tilde{a}^T \tilde{t}) e^{-\frac{\|\tilde{t}\|^2}{2}} \varphi_{\beta, d-1}(\tilde{t}) d\tilde{t} \cdot \int_{\mathbb{R}} t_d \sin(at_d) e^{-\frac{t_d^2}{2}} \varphi_{\beta, 1}(t_d) dt_d \right)
\end{aligned}$$

$$\begin{aligned}
&= \left( \tilde{a} \frac{\beta^2}{(1+\beta^2)^{\frac{d-1}{2}+1}} e^{-\frac{\beta^2}{2(1+\beta^2)} \|\tilde{a}\|^2} \cdot \frac{1}{(1+\beta^2)^{1/2}} e^{-\frac{\beta^2}{2(1+\beta^2)} a_d^2} \right) \\
&\quad \left( \frac{1}{(1+\beta^2)^{(d-1)/2}} e^{-\frac{\beta^2}{2(1+\beta^2)} \|\tilde{a}\|^2} \cdot a_d \frac{\beta^2}{(1+\beta^2)^{\frac{1}{2}+1}} e^{-\frac{\beta^2}{2(1+\beta^2)} a_d^2} \right) \\
&= a \frac{\beta^2}{(1+\beta^2)^{\frac{d}{2}+1}} e^{-\frac{\beta^2}{2(1+\beta^2)} \|a\|^2}. \quad \square
\end{aligned}$$

**Lemma 19** Für die in Gleichung (1.10) auf Seite 12 definierte Differenz  $\Delta_j$  zwischen den standardisierten Daten  $Y_j$  und den ursprünglichen Daten  $X_j$  gilt:

- 1)  $\frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \|X_j\|^k \rightarrow 0 \quad \text{f.s.} \quad (k = 0, 1, 2, 3),$
- 2)  $\frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^2 \|X_j\|^k \rightarrow 0 \quad \text{f.s.} \quad (k = 0, 1, 2),$
- 3)  $\frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^3 \|X_j\|^k \rightarrow 0 \quad \text{f.s.} \quad (k = 0, 1),$
- 4)  $\frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^4 \rightarrow 0 \quad \text{f.s.}$

BEWEIS: Nach Definition ist  $\Delta_j = (S_n^{-1/2} - I_d)X_j - S_n^{-1/2}\bar{X}_n$ , und es gilt

$$\|\Delta_j\| \leq \|S_n^{-1/2} - I_d\|_S \|X_j\| + \|S_n^{-1/2}\|_S \|\bar{X}_n\|.$$

Damit folgen die Behauptungen aus der fast sicheren Konvergenz von  $\bar{X}_n$  gegen 0 und von  $S_n^{-1}$  bzw.  $S_n^{-1/2}$  gegen  $I_d$ , sowie aus der fast sicheren Konvergenz von  $\frac{1}{n} \sum_{j=1}^n \|X_j\|^l$  ( $l = 1, 2, 3, 4$ ): 1) ist klar, 2) folgt aus

$$\begin{aligned}
\frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^2 \|X_j\|^k &\leq \|S_n^{-1/2} - I_d\|_S^2 \frac{1}{n} \sum_{j=1}^n \|X_j\|^{k+2} + \|S_n^{-1/2}\|_S^2 \|\bar{X}_n\|^2 \frac{1}{n} \sum_{j=1}^n \|X_j\|^k \\
&\quad + 2 \|S_n^{-1/2}\|_S \|\bar{X}_n\| \|S_n^{-1/2} - I_d\|_S \frac{1}{n} \sum_{j=1}^n \|X_j\|^{k+1} \rightarrow 0 \quad \text{f.s.},
\end{aligned}$$

und die Beweise von 3) und 4) verlaufen analog. Man beachte, daß nur Potenzen von  $\|X_j\|$  der Ordnung  $\leq 4$  auftreten, so daß die vorausgesetzte Existenz der Momente vierter Ordnung ausreicht.  $\square$

### Beweis von Lemma 7:

Es werden nur die Aussagen (i) bewiesen, da der Beweis von (ii) jeweils analog verläuft. Mit

der in Gleichung (1.10) auf Seite 12 definierten Differenz  $\Delta_j = Y_j - X_j$  zwischen den standardisierten Daten und den ursprünglichen Daten gilt wie in den Gleichungen (1.13) und (1.14) auf Seite 14

$$\begin{aligned}\cos(t^T Y_j) &= \cos(t^T X_j) - t^T \Delta_j \sin(t^T X_j) + \varepsilon_{n,j}(t), \\ \sin(t^T Y_j) &= \sin(t^T X_j) + t^T \Delta_j \cos(t^T X_j) + \eta_{n,j}(t)\end{aligned}$$

mit  $|\varepsilon_{n,j}(t)| \leq \|t\|^2 \|\Delta_j\|^2$  und  $|\eta_{n,j}(t)| \leq \|t\|^2 \|\Delta_j\|^2$ . Im folgenden sei  $t \in \mathbb{R}^d$  fest, und  $o_{f.s.}(1)$  bezeichne einen Restterm, der nicht mehr von  $t$  abhängt und fast sicher gegen 0 strebt.

1.(i) Es ist  $\Phi_{1,n}(t) = \frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) = \Phi_{1,n}^\diamond(t) + R_{1,n}(t)$  mit

$$R_{1,n}(t) := -\frac{1}{n} \sum_{j=1}^n (t^T \Delta_j) \sin(t^T X_j) + \frac{1}{n} \sum_{j=1}^n \varepsilon_{n,j}(t)$$

und  $|R_{1,n}(t)| \leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}$  nach Lemma 19:

$$\begin{aligned}\left| \frac{1}{n} \sum_{j=1}^n (t^T \Delta_j) \sin(t^T X_j) \right| &\leq \frac{1}{n} \sum_{j=1}^n |t^T \Delta_j| \leq \|t\| \frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \rightarrow 0 \quad \text{f.s.}, \\ \left| \frac{1}{n} \sum_{j=1}^n \varepsilon_{n,j}(t) \right| &\leq \|t\|^2 \frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^2 \rightarrow 0 \quad \text{f.s.}\end{aligned}$$

2.(i) Es ist  $\Phi'_{1,n}(t) = -\frac{1}{n} \sum_{j=1}^n \sin(t^T Y_j) Y_j = \Phi'_{1,n}^\diamond(t) + R'_{1,n}(t)$  mit

$$R'_{1,n}(t) := -\frac{1}{n} \sum_{j=1}^n \sin(t^T Y_j) \Delta_j - \frac{1}{n} \sum_{j=1}^n (t^T \Delta_j) \cos(t^T X_j) X_j - \frac{1}{n} \sum_{j=1}^n \eta_{n,j}(t) X_j$$

und  $\|R'_{1,n}(t)\| \leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}$  nach Lemma 19:

$$\begin{aligned}\left\| \frac{1}{n} \sum_{j=1}^n \sin(t^T Y_j) \Delta_j \right\| &\leq \frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \rightarrow 0 \quad \text{f.s.}, \\ \left\| \frac{1}{n} \sum_{j=1}^n (t^T \Delta_j) \cos(t^T X_j) X_j \right\| &\leq \|t\| \frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \|X_j\| \rightarrow 0 \quad \text{f.s.}, \\ \left\| \frac{1}{n} \sum_{j=1}^n \eta_{n,j}(t) X_j \right\| &\leq \|t\|^2 \frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^2 \|X_j\| \rightarrow 0 \quad \text{f.s.}\end{aligned}$$

3.(i) Es ist  $\Phi''_{1,n}(t) = -\frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) Y_j Y_j^T = \Phi''_{1,n}^\diamond(t) + R''_{1,n}(t)$  mit

$$\begin{aligned}R''_{1,n}(t) &:= -\frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) X_j \Delta_j^T - \frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) \Delta_j X_j^T - \frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) \Delta_j \Delta_j^T \\ &\quad + \frac{1}{n} \sum_{j=1}^n (t^T \Delta_j) \sin(t^T X_j) X_j X_j^T - \frac{1}{n} \sum_{j=1}^n \varepsilon_{n,j}(t) X_j X_j^T\end{aligned}$$

und  $\|R'_{1,n}(t)\|_S \leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}$  nach Lemma 19:

$$\begin{aligned} \left\| \frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) X_j \Delta_j^T \right\|_S &\leq \frac{1}{n} \sum_{j=1}^n \|X_j \Delta_j^T\|_S = \frac{1}{n} \sum_{j=1}^n \|X_j\| \|\Delta_j\| \rightarrow 0 \quad \text{f.s.}, \\ \left\| \frac{1}{n} \sum_{j=1}^n \cos(t^T Y_j) \Delta_j \Delta_j^T \right\|_S &\leq \frac{1}{n} \sum_{j=1}^n \|\Delta_j \Delta_j^T\|_S = \frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^2 \rightarrow 0 \quad \text{f.s.}, \\ \left\| \frac{1}{n} \sum_{j=1}^n (t^T \Delta_j) \sin(t^T X_j) X_j X_j^T \right\|_S &\leq \|t\| \frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \|X_j\|^2 \rightarrow 0 \quad \text{f.s.}, \\ \left\| \frac{1}{n} \sum_{j=1}^n \varepsilon_{n,j}(t) X_j X_j^T \right\|_S &\leq \|t\|^2 \frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^2 \|X_j\|^2 \rightarrow 0 \quad \text{f.s.} \end{aligned}$$

Hierbei beachte man, daß nach Definition der Schur-Norm für  $x, y \in \mathbb{R}^d$  gilt:  $\|xy^T\|_S = \|x\| \|y\|$  (vgl. Anhang A).

4.(i) Es ist  $M'_{3,1,n}(t) = -\frac{1}{n^2} \sum_{j,k=1}^n \sin(t^T Y_j) (Y_k^T Y_j) (t^T Y_k) Y_k = M'_{3,1,n}{}^\diamond(t) + R'_{3,1,n}(t)$  mit

$$\begin{aligned} R'_{3,1,n}(t) &:= -\frac{1}{n^2} \sum_{j,k=1}^n \sin(t^T Y_j) \left\{ (X_j^T X_k) [(t^T X_k) \Delta_k + (t^T \Delta_k) X_k + (t^T \Delta_k) \Delta_k] \right. \\ &\quad \left. + [(X_j^T \Delta_k) + (\Delta_j^T X_k) + (\Delta_j^T \Delta_k)] \cdot \right. \\ &\quad \left. [(t^T X_k) X_k + (t^T X_k) \Delta_k + (t^T \Delta_k) X_k + (t^T \Delta_k) \Delta_k] \right\} \\ &\quad - \frac{1}{n^2} \sum_{j,k=1}^n (t^T \Delta_j) \cos(t^T X_j) (X_k^T X_j) (t^T X_k) X_k \\ &\quad - \frac{1}{n^2} \sum_{j,k=1}^n \eta_{n,j}(t) (X_k^T X_j) (t^T X_k) X_k, \end{aligned}$$

und nach Lemma 19 gilt

$$\begin{aligned} \|R'_{3,1,n}(t)\| &\leq \|t\| \left\{ \frac{1}{n} \sum_{j=1}^n \|X_j\| \frac{1}{n} \sum_{k=1}^n [3 \|X_k\|^2 \|\Delta_k\| + 3 \|X_k\| \|\Delta_k\|^2 + \|\Delta_k\|^3] \right. \\ &\quad \left. + \frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \frac{1}{n} \sum_{k=1}^n [\|X_k\|^3 + 3 \|X_k\|^2 \|\Delta_k\| + 3 \|X_k\| \|\Delta_k\|^2 + \|\Delta_k\|^3] \right\} \\ &\quad + \|t\|^2 \frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \|X_j\| \frac{1}{n} \sum_{k=1}^n \|X_k\|^3 \\ &\quad + \|t\|^3 \frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^2 \|X_j\| \frac{1}{n} \sum_{k=1}^n \|X_k\|^3 \\ &\leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}. \end{aligned}$$

5.(i) Es ist  $M''_{4,1,n}(t) = -\frac{1}{n^2} \sum_{j,k=1}^n \sin(t^T Y_j) (t^T Y_k) (Y_k^T Y_j) Y_k Y_k^T = M''_{4,1,n}{}^\diamond(t) + R''_{4,1,n}(t)$  mit

$$\begin{aligned}
R''_{4,1,n}(t) &:= -\frac{1}{n^2} \sum_{j,k=1}^n \sin(t^T Y_j) \cdot \\
&\quad \left\{ (X_j^T X_k) [(t^T X_k) X_k \Delta_k^T + (t^T X_k) \Delta_k X_k^T + (t^T \Delta_k) X_k X_k^T \right. \\
&\quad + (t^T \Delta_k) \Delta_k X_k^T + (t^T X_k) \Delta_k \Delta_k^T + (t^T \Delta_k) X_k \Delta_k^T + (t^T \Delta_k) \Delta_k \Delta_k^T] \\
&\quad + [(X_j^T \Delta_k) + (\Delta_j^T X_k) + (\Delta_j^T \Delta_k)] \cdot \\
&\quad \left. [(t^T X_k) X_k X_k^T + (t^T X_k) \Delta_k X_k^T + (t^T \Delta_k) X_k X_k^T + (t^T \Delta_k) \Delta_k X_k^T \right. \\
&\quad \left. + (t^T X_k) X_k \Delta_k^T + (t^T X_k) \Delta_k \Delta_k^T + (t^T \Delta_k) X_k \Delta_k^T + (t^T \Delta_k) \Delta_k \Delta_k^T] \right\} \\
&\quad - \frac{1}{n^2} \sum_{j,k=1}^n (t^T \Delta_j) \cos(t^T X_j) (X_j^T X_k) (t^T X_k) X_k X_k^T \\
&\quad - \frac{1}{n^2} \sum_{j,k=1}^n \eta_{n,j}(t) (X_j^T X_k) (t^T X_k) X_k X_k^T,
\end{aligned}$$

und nach Lemma 19 gilt

$$\begin{aligned}
&\|R''_{4,1,n}(t)\|_S \\
&\leq \|t\| \left\{ \frac{1}{n} \sum_{j=1}^n \|X_j\| \frac{1}{n} \sum_{k=1}^n [4 \|X_k\|^3 \|\Delta_k\| + 6 \|X_k\|^2 \|\Delta_k\|^2 + 4 \|X_k\| \|\Delta_k\|^3 + \|\Delta_k\|^4] \right. \\
&\quad \left. + \frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \frac{1}{n} \sum_{k=1}^n [\|X_k\|^4 + 4 \|X_k\|^3 \|\Delta_k\| + 6 \|X_k\|^2 \|\Delta_k\|^2 + 4 \|X_k\| \|\Delta_k\|^3 \right. \\
&\quad \left. + \|\Delta_k\|^4] \right\} \\
&\quad + \|t\|^2 \frac{1}{n} \sum_{j=1}^n \|\Delta_j\| \|X_j\| \frac{1}{n} \sum_{k=1}^n \|X_k\|^4 \\
&\quad + \|t\|^3 \frac{1}{n} \sum_{j=1}^n \|\Delta_j\|^2 \|X_j\| \frac{1}{n} \sum_{k=1}^n \|X_k\|^4 \\
&\leq o_{f.s.}(1) \cdot \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}. \quad \square
\end{aligned}$$

### Beweis von Satz 10:

I) Im folgenden wird jeweils nur ein Teil der Aussagen 1. bis 8. bewiesen, da der Nachweis des anderen analog verlauft. Desweiteren bezeichne  $o_{f.s.}(1)$  einen Restterm, der fast sicher gegen 0 strebt und nicht mehr vom Parameter  $t$  abhangt. Alle Integrale haben den gesamten  $\mathbb{R}^d$  als

Integrationsbereich.

1.(i),(ii) Nach Definition und nach Lemma 7 gilt

$$K_{\beta 1n}^1 = \iint \Phi_{1,n}(t-s) \left[ \Phi_n(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi_n(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt = K_{\beta 1n}^{1\diamond} + R_n$$

mit

$$\begin{aligned} K_{\beta 1n}^{1\diamond} &:= \iint \Phi_{1,n}^\diamond(t-s) \left[ \Phi_n^\diamond(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi_n^\diamond(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt \\ &= \frac{1}{n^3} \sum_{i,j,k=1}^n \iint \cos((t-s)^T X_i) [\cos(s^T X_j) + \sin(s^T X_j)] [\cos(t^T X_k) + \sin(t^T X_k)] \\ &\quad \varphi_\beta(s) ds \varphi_\beta(t) dt \\ &\quad - \frac{2}{n^2} \sum_{i,j=1}^n \iint \cos((t-s)^T X_i) [\cos(s^T X_j) + \sin(s^T X_j)] e^{-\frac{\|t\|^2}{2}} \varphi_\beta(s) ds \varphi_\beta(t) dt \\ &\quad + \frac{1}{n} \sum_{i=1}^n \iint \cos((t-s)^T X_i) e^{-\frac{\|s\|^2}{2}} e^{-\frac{\|t\|^2}{2}} \varphi_\beta(s) ds \varphi_\beta(t) dt \end{aligned}$$

und

$$\begin{aligned} R_n &:= \iint R_{1,n}(t-s) \left[ \Phi_n(s) - e^{-\frac{\|s\|^2}{2}} \right] \left[ \Phi_n(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(s) ds \varphi_\beta(t) dt \\ &\quad + 2 \iint \Phi_{1,n}^\diamond(t-s) \left[ \Phi_n^\diamond(s) - e^{-\frac{\|s\|^2}{2}} \right] [R_{1,n}(t) + R_{2,n}(t)] \varphi_\beta(s) ds \varphi_\beta(t) dt \\ &\quad + \iint \Phi_{1,n}^\diamond(t-s) [R_{1,n}(s) + R_{2,n}(s)] [R_{1,n}(t) + R_{2,n}(t)] \varphi_\beta(s) ds \varphi_\beta(t) dt . \end{aligned}$$

Der Restterm  $R_n$  genügt der Abschätzung

$$\begin{aligned} |R_n| &\leq o_{f.s.}(1) \cdot 9 \iint \max_{l \in \{0,1,2,3\}} \{\|t-s\|^l\} \varphi_\beta(s) ds \varphi_\beta(t) dt + o_{f.s.}(1) \cdot 6 \int \max_{l \in \{0,1,2,3\}} \{\|t\|^l\} \varphi_\beta(t) dt \\ &\quad + o_{f.s.}(1) \cdot \iint \max_{l \in \{0,1,2,3\}} \{\|s\|^l\} \max_{l \in \{0,1,2,3\}} \{\|t\|^l\} \varphi_\beta(s) ds \varphi_\beta(t) dt \quad \rightarrow 0 \quad \text{f.s.} \end{aligned}$$

Folglich gilt  $K_{\beta 1n}^1 = K_{\beta 1n}^{1\diamond} + R_n = K_{\beta 1n}^{1\diamond} + o_{f.s.}(1)$ , und die Behauptung folgt aus dem starken Gesetz großer Zahlen in Verbindung mit Theorem D.2, denn danach konvergiert  $K_{\beta 1n}^{1\diamond}$  f.s. gegen

$$\begin{aligned} \mathbb{E} &\left[ \iint \cos((t-s)^T X_1) [\cos(s^T X_2) + \sin(s^T X_2)] [\cos(t^T X_3) + \sin(t^T X_3)] \varphi_\beta(s) ds \varphi_\beta(t) dt \right] \\ &\quad - 2\mathbb{E} \left[ \iint \cos((t-s)^T X_1) [\cos(s^T X_2) + \sin(s^T X_2)] e^{-\frac{\|t\|^2}{2}} \varphi_\beta(s) ds \varphi_\beta(t) dt \right] \\ &\quad + \mathbb{E} \left[ \iint \cos((t-s)^T X_1) e^{-\frac{\|s\|^2}{2}} e^{-\frac{\|t\|^2}{2}} \varphi_\beta(s) ds \varphi_\beta(t) dt \right] \end{aligned}$$

$$\begin{aligned}
&= \iint \Phi_1(t-s)\Phi(s)\Phi(t)\varphi_\beta(s)ds\varphi_\beta(t)dt - 2 \iint \Phi_1(t-s)\Phi(s)e^{-\frac{\|t\|^2}{2}}\varphi_\beta(s)ds\varphi_\beta(t)dt \\
&\quad + \iint \Phi_1(t-s)e^{-\frac{\|s\|^2}{2}}e^{-\frac{\|t\|^2}{2}}\varphi_\beta(s)ds\varphi_\beta(t)dt \\
&= K_{\beta 1}^1.
\end{aligned}$$

Man beachte, daß in diesem Fall die Voraussetzung von Theorem D.2 trivialerweise erfüllt ist, da die Zufallsvektoren  $X_j$  nur in den beschränkten Sinus/Cosinus-Termen auftreten.

2.(i),(ii) Nach Definition und nach Lemma 7 gilt

$$K_{\beta 1n}^2 = \int \Phi_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt = K_{\beta 1n}^{2\diamond} + R_n$$

mit

$$\begin{aligned}
K_{\beta 1n}^{2\diamond} &:= \int \Phi_{1,n}^\diamond(t) \left[ \Phi_{1,n}^\diamond(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\
&= \frac{1}{n^2} \sum_{j,k=1}^n \int \cos(t^T X_j) \cos(t^T X_k) \varphi_\beta(t) dt - \frac{1}{n} \sum_{j=1}^n \int \cos(t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt
\end{aligned}$$

und

$$R_n := \int R_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + \int \Phi_{1,n}^\diamond(t) R_{1,n}(t) \varphi_\beta(t) dt.$$

Der Restterm genügt der Abschätzung

$$|R_n| \leq o_{f.s.}(1) \cdot 3 \int \max_{l \in \{0,1,2,3\}} \{\|t\|^l\} \varphi_\beta(t) dt \rightarrow 0 \quad \text{f.s.}$$

Folglich gilt  $K_{\beta 1n}^2 = K_{\beta 1n}^{2\diamond} + R_n = K_{\beta 1n}^{2\diamond} + o_{f.s.}(1)$ , und die Behauptung folgt aus dem starken Gesetz großer Zahlen in Verbindung mit Theorem D.2. Man beachte, daß im Falle von  $K_{\beta 1n}^{2\diamond}$  die Existenz von Momenten vierter Ordnung von  $X_1$  problemlos ausreicht, um die Voraussetzung von Theorem D.2 zu erfüllen. Dies gilt auch für die nachfolgenden Beweisteile 3.(i),(ii), 4.(i),(ii), 5.(i),(ii) und 7.(i),(ii).

3.(i),(ii) Nach Definition und nach Lemma 7 gilt

$$K_{\beta 1n}^3 = \int t^T \Phi'_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt = K_{\beta 1n}^{3\diamond} + R_n$$

mit

$$\begin{aligned}
K_{\beta 1n}^{3\diamond} &:= \int t^T \Phi'_{1,n}^\diamond(t) \left[ \Phi_{1,n}^\diamond(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\
&= -\frac{1}{n^2} \sum_{j,k=1}^n \int \sin(t^T X_j) (t^T X_j) \cos(t^T X_k) \varphi_\beta(t) dt
\end{aligned}$$

$$+\frac{1}{n} \sum_{j=1}^n \int \sin(t^T X_j) (t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt$$

und

$$R_n := \int t^T R'_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + \int t^T \Phi'_{1,n}(t) R_{1,n}(t) \varphi_\beta(t) dt .$$

Der Restterm genügt der Abschätzung

$$\begin{aligned} |R_n| &\leq o_{f.s.}(1) \cdot 2 \int \max_{l \in \{1,2,3,4\}} \{\|t\|^l\} \varphi_\beta(t) dt + o_{f.s.}(1) \frac{1}{n} \sum_{j=1}^n \|X_j\| \int \max_{l \in \{1,2,3,4\}} \{\|t\|^l\} \varphi_\beta(t) dt \\ &\rightarrow 0 \quad \text{f.s.} \end{aligned}$$

Folglich gilt  $K_{\beta 1n}^3 = K_{\beta 1n}^{3\diamond} + R_n = K_{\beta 1n}^{3\diamond} + o_{f.s.}(1)$ , und die Behauptung folgt aus dem starken Gesetz großer Zahlen in Verbindung mit Theorem D.2.

4.(i),(ii) Nach Definition und nach Lemma 7 gilt

$$K_{\beta 2n}^{4'} = \int \Phi'_{2,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt = K_{\beta 2n}^{4'\diamond} + R_n$$

mit

$$\begin{aligned} K_{\beta 2n}^{4'\diamond} &:= \int \Phi'_{2,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\ &= \frac{1}{n^2} \sum_{j,k=1}^n X_j \int \cos(t^T X_j) \cos(t^T X_k) \varphi_\beta(t) dt - \frac{1}{n} \sum_{j=1}^n X_j \int \cos(t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt \end{aligned}$$

und

$$R_n := \int R'_{2,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + \int \Phi'_{2,n}(t) R_{1,n}(t) \varphi_\beta(t) dt .$$

Der Restterm genügt nach Lemma A.4 aus Anhang A der Abschätzung

$$\begin{aligned} \|R_n\| &\leq \left( \int \|R'_{2,n}(t)\|^2 \varphi_\beta(t) dt \cdot \int \left| \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt \right)^{1/2} \\ &\quad + \left( \int \|\Phi'_{2,n}(t)\|^2 \varphi_\beta(t) dt \cdot \int R_{1,n}(t)^2 \varphi_\beta(t) dt \right)^{1/2} \\ &\leq 2 \left( o_{f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}^2 \varphi_\beta(t) dt \right)^{1/2} \\ &\quad + \left( \left( \frac{1}{n} \sum_{j=1}^n \|X_j\| \right)^2 \cdot o_{f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}^2 \varphi_\beta(t) dt \right)^{1/2} \rightarrow 0 \quad \text{f.s.} \end{aligned}$$

Folglich gilt  $K_{\beta 2n}^{4'} = K_{\beta 2n}^{4'\diamond} + R_n = K_{\beta 2n}^{4'\diamond} + o_{f.s.}(1)$ , und die Behauptung folgt aus dem starken Gesetz großer Zahlen in Verbindung mit Theorem D.2.

5.(i),(ii) Nach Definition und nach Lemma 7 gilt

$$K_{\beta 2n}^{5'} = \int t \Phi_{2,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_{\beta}(t) dt = K_{\beta 2n}^{5'\diamond} + R_n$$

mit

$$\begin{aligned} K_{\beta 2n}^{5'\diamond} &:= \int t \Phi_{2,n}^{\diamond}(t) \left[ \Phi_{1,n}^{\diamond}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_{\beta}(t) dt \\ &= \frac{1}{n^2} \sum_{j,k=1}^n \int t \sin(t^T X_j) \cos(t^T X_k) \varphi_{\beta}(t) dt - \frac{1}{n} \sum_{j=1}^n \int t \sin(t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_{\beta}(t) dt \end{aligned}$$

und

$$R_n := \int t R_{2,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_{\beta}(t) dt + \int t \Phi_{2,n}^{\diamond}(t) R_{1,n}(t) \varphi_{\beta}(t) dt .$$

Der Restterm genügt nach Lemma A.4 der Abschätzung

$$\begin{aligned} \|R_n\| &\leq \left( \int |R_{2,n}(t)|^2 \|t\|^2 \varphi_{\beta}(t) dt \cdot \int \left| \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_{\beta}(t) dt \right)^{1/2} \\ &\quad + \left( \int |\Phi_{2,n}^{\diamond}(t)|^2 \|t\|^2 \varphi_{\beta}(t) dt \cdot \int R_{1,n}(t)^2 \varphi_{\beta}(t) dt \right)^{1/2} \\ &\leq 2 \left( o_{f.s.}(1) \int \max_{l \in \{1,2,3,4\}} \{\|t\|^l\}^2 \varphi_{\beta}(t) dt \right)^{1/2} \\ &\quad + \left( \int \|t\|^2 \varphi_{\beta}(t) dt \cdot o_{f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{\|t\|^l\}^2 \varphi_{\beta}(t) dt \right)^{1/2} \rightarrow 0 \quad f.s. \end{aligned}$$

Folglich gilt  $K_{\beta 2n}^{5'} = K_{\beta 2n}^{5'\diamond} + R_n = K_{\beta 2n}^{5'\diamond} + o_{f.s.}(1)$ , und die Behauptung folgt aus dem starken Gesetz großer Zahlen in Verbindung mit Theorem D.2.

6.(i),(ii) Nach Definition und nach Lemma 7 gilt

$$K_{\beta 1n}^{6'} = \int M'_{3,1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_{\beta}(t) dt = K_{\beta 1n}^{6'\diamond} + R_n$$

mit

$$\begin{aligned} K_{\beta 1n}^{6'\diamond} &:= \int M'_{3,1,n}{}^{\diamond}(t) \left[ \Phi_{1,n}^{\diamond}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_{\beta}(t) dt \\ &= -\frac{1}{n^3} \sum_{i,j,k=1}^n X_k (X_k^T X_j) \left( X_k^T \int t \sin(t^T X_j) \cos(t^T X_i) \varphi_{\beta}(t) dt \right) \end{aligned}$$

$$+ \frac{1}{n^2} \sum_{j,k=1}^n X_k (X_k^T X_j) \left( X_k^T \int t \sin(t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt \right)$$

und

$$R_n := \int R'_{3,1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + \int M'_{3,1,n}{}^\diamond(t) R_{1,n}(t) \varphi_\beta(t) dt .$$

Der Restterm genügt nach Lemma A.4 der Abschätzung

$$\begin{aligned} \|R_n\| &\leq \left( \int \left\| R'_{3,1,n}(t) \right\|^2 \varphi_\beta(t) dt \cdot \int \left| \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt \right)^{1/2} \\ &\quad + \left( \int \left\| M'_{3,1,n}{}^\diamond(t) \right\|^2 \varphi_\beta(t) dt \cdot \int R_{1,n}(t)^2 \varphi_\beta(t) dt \right)^{1/2} \\ &\leq 2 \left( o_{f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{ \|t\|^l \}^2 \varphi_\beta(t) dt \right)^{1/2} \\ &\quad + \left( \left\{ \frac{1}{n} \sum_{j=1}^n \|X_j\| \frac{1}{n} \sum_{k=1}^n \|X_k\|^3 \right\}^2 \int \|t\|^2 \varphi_\beta(t) dt \cdot o_{f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{ \|t\|^l \}^2 \varphi_\beta(t) dt \right)^{1/2} \\ &\rightarrow 0 \quad f.s. \end{aligned}$$

Folglich gilt  $K_{\beta 1n}^{6'} = K_{\beta 1n}^{6'\diamond} + R_n = K_{\beta 1n}^{6'\diamond} + o_{f.s.}(1)$ , und die Behauptung folgt aus Theorem D.2. Man beachte, daß die Existenz der Momente vierter Ordnung ausreicht, um die Voraussetzung von Theorem D.2 zu erfüllen.

7.(i),(ii) Nach Definition und nach Lemma 7 gilt

$$K_{\beta 1n}^{7''} = \int \Phi''_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt = K_{\beta 1n}^{7''\diamond} + R_n$$

mit

$$\begin{aligned} K_{\beta 1n}^{7''\diamond} &:= \int \Phi''_{1,n}{}^\diamond(t) \left[ \Phi_{1,n}{}^\diamond(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\ &= -\frac{1}{n^2} \sum_{j,k=1}^n X_j X_j^T \int \cos(t^T X_j) \cos(t^T X_k) \varphi_\beta(t) dt \\ &\quad + \frac{1}{n} \sum_{j=1}^n X_j X_j^T \int \cos(t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt \end{aligned}$$

und

$$R_n := \int R''_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + \int \Phi''_{1,n}{}^\diamond(t) R_{1,n}(t) \varphi_\beta(t) dt .$$

Der Restterm genügt nach Lemma A.4 der Abschätzung

$$\|R_n\|_S \leq \left( \int \left\| R''_{1,n}(t) \right\|_S^2 \varphi_\beta(t) dt \cdot \int \left| \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt \right)^{1/2}$$

$$\begin{aligned}
& + \left( \int \left\| \Phi_{1,n}'' \diamond(t) \right\|_S^2 \varphi_\beta(t) dt \cdot \int R_{1,n}(t)^2 \varphi_\beta(t) dt \right)^{1/2} \\
& \leq 2 \left( \text{o.f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{ \|t\|^l \}^2 \varphi_\beta(t) dt \right)^{1/2} \\
& + \left( \left( \frac{1}{n} \sum_{j=1}^n \|X_j\|^2 \right)^2 \cdot \text{o.f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{ \|t\|^l \}^2 \varphi_\beta(t) dt \right)^{1/2} \rightarrow 0 \quad \text{f.s.}
\end{aligned}$$

Folglich gilt  $K_{\beta 1n}^{7''} = K_{\beta 1n}^{7'' \diamond} + R_n = K_{\beta 1n}^{7'' \diamond} + \text{o.f.s.}(1)$ , und die Behauptung folgt aus dem starken Gesetz großer Zahlen in Verbindung mit Theorem D.2.

8.(i),(ii) Nach Definition und nach Lemma 7 gilt

$$K_{\beta 1n}^{8''} = \int M_{4,1,n}''(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt = K_{\beta 1n}^{8'' \diamond} + R_n^\ddagger \quad (1.31)$$

mit

$$\begin{aligned}
K_{\beta 1n}^{8'' \diamond} & := \int M_{4,1,n}'' \diamond(t) \left[ \Phi_{1,n}^\diamond(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\
& = -\frac{1}{n^3} \sum_{i,j,k=1}^n X_k X_k^T (X_j^T X_k) \left( X_k^T \int t \sin(t^T X_j) \cos(t^T X_i) \varphi_\beta(t) dt \right) \\
& \quad + \frac{1}{n^2} \sum_{j,k=1}^n X_k X_k^T (X_j^T X_k) \left( X_k^T \int t \sin(t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt \right)
\end{aligned}$$

und

$$R_n^\ddagger := \int R_{4,1,n}''(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + \int M_{4,1,n}'' \diamond(t) R_{1,n}(t) \varphi_\beta(t) dt.$$

Der Restterm genügt nach Lemma A.4 der Abschätzung

$$\begin{aligned}
\|R_n^\ddagger\|_S & \leq \left( \int \left\| R_{4,1,n}''(t) \right\|_S^2 \varphi_\beta(t) dt \cdot \int \left| \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt \right)^{1/2} \\
& + \left( \int \left\| M_{4,1,n}'' \diamond(t) \right\|_S^2 \varphi_\beta(t) dt \cdot \int R_{1,n}(t)^2 \varphi_\beta(t) dt \right)^{1/2} \\
& \leq 2 \left( \text{o.f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{ \|t\|^l \}^2 \varphi_\beta(t) dt \right)^{1/2} \\
& + \left( \left\{ \frac{1}{n} \sum_{j=1}^n \|X_j\| \frac{1}{n} \sum_{k=1}^n \|X_k\|^4 \right\}^2 \int \|t\|^2 \varphi_\beta(t) dt \cdot \text{o.f.s.}(1) \int \max_{l \in \{0,1,2,3\}} \{ \|t\|^l \}^2 \varphi_\beta(t) dt \right)^{1/2} \\
& \rightarrow 0 \quad \text{f.s.} \tag{1.32}
\end{aligned}$$

Folglich gilt  $K_{\beta 1n}^{8''} = K_{\beta 1n}^{8''\diamond} + R_n^\ddagger = K_{\beta 1n}^{8''\diamond} + o_{f.s.}(1)$ , und die Behauptung folgt aus Theorem D.2. Man beachte, daß hierfür die Existenz der Momente vierter Ordnung ausreicht, denn es ist z.B.

$$\begin{aligned} & -\frac{1}{n^3} \sum_{i,j,k=1}^n X_k X_k^T (X_k^T X_j) \left( X_k^T \int t \sin(t^T X_j) \cos(t^T X_i) \varphi_\beta(t) dt \right) \\ & = - \sum_{l,m=1}^d \left( \frac{1}{n} \sum_{k=1}^n X_k X_k^T X_{k,l} X_{k,m} \right) \left( \frac{1}{n^2} \sum_{i,j=1}^n X_{j,l} \int t_m \sin(t^T X_j) \cos(t^T X_i) \varphi_\beta(t) dt \right), \end{aligned}$$

wobei  $X_{j,l}$  bzw.  $t_l$  ( $l = 1, \dots, d$ ) die  $d$  Komponenten des Vektors  $X_j$  bzw.  $t$  bezeichnen.

II) Wir zeigen exemplarisch, daß  $K_{\beta 1n}^9 (K_{\beta 1n}^{8''}) \rightarrow K_{\beta 1}^9 (K_{\beta 1}^{8''})$  f.s. gilt. Die anderen Beweisteile verlaufen analog. Nach Definition 9, Lemma 7 und Gleichung (1.31) ist

$$\begin{aligned} K_{\beta 1n}^9 (K_{\beta 1n}^{8''}) & = \int t^T K_{\beta 1n}^{8''} \Phi'_{1,n}(t) \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\ & = \int t^T K_{\beta 1n}^{8''\diamond} \Phi'_{1,n}{}^\diamond(t) \left[ \Phi_{1,n}^\diamond(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt + R_n, \end{aligned}$$

wobei der Restterm

$$\begin{aligned} R_n & := \int t^T \left\{ K_{\beta 1n}^{8''\diamond} R'_{1,n}(t) + R_n^\ddagger \Phi'_{1,n}{}^\diamond(t) + R_n^\ddagger R'_{1,n}(t) \right\} \left[ \Phi_{1,n}(t) - e^{-\frac{\|t\|^2}{2}} \right] \varphi_\beta(t) dt \\ & \quad + \int t^T K_{\beta 1n}^{8''\diamond} \Phi'_{1,n}{}^\diamond(t) R_{1,n} \varphi_\beta(t) dt \end{aligned}$$

nach Lemma 7 und Gleichung (1.32) der Abschätzung

$$\begin{aligned} |R_n| & \leq 2 \int \|t\| \left\{ \left\| K_{\beta 1n}^{8''\diamond} \right\|_S \left\| R'_{1,n}(t) \right\| + \left\| R_n^\ddagger \right\|_S \left\| \Phi'_{1,n}{}^\diamond(t) \right\| + \left\| R_n^\ddagger \right\|_S \left\| R'_{1,n}(t) \right\| \right\} \varphi_\beta(t) dt \\ & \quad + \int \|t\| \left\| K_{\beta 1n}^{8''\diamond} \right\|_S \left\| \Phi'_{1,n}{}^\diamond(t) \right\| |R_{1,n}| \varphi_\beta(t) dt \\ & \leq 2 \int \|t\| \left\{ \text{const} \frac{1}{n} \sum_{j=1}^n \|X_j\| \frac{1}{n} \sum_{k=1}^n \|X_k\|^4 \cdot o_{f.s.}(1) \max_{l \in \{0,1,2,3\}} \{\|t\|^l\} \right. \\ & \quad \left. + o_{f.s.}(1) \frac{1}{n} \sum_{j=1}^n \|X_j\| + o_{f.s.}(1) \cdot o_{f.s.}(1) \max_{l \in \{0,1,2,3\}} \{\|t\|^l\} \right\} \varphi_\beta(t) dt \\ & \quad + \int \|t\| \text{const} \frac{1}{n} \sum_{j=1}^n \|X_j\| \frac{1}{n} \sum_{k=1}^n \|X_k\|^4 \cdot \frac{1}{n} \sum_{j=1}^n \|X_j\| \cdot o_{f.s.}(1) \max_{l \in \{0,1,2,3\}} \{\|t\|^l\} \varphi_\beta(t) dt \\ & \rightarrow 0 \quad \text{f.s.} \end{aligned}$$

genügt. Der Ausdruck  $\text{const}$  bezeichne hierbei eine generische, endliche Konstante, die in verschiedenen Kontexten unterschiedliche Werte annimmt. Damit gilt

$$K_{\beta 1n}^9 (K_{\beta 1n}^{8''}) = \frac{1}{n^5} \sum_{i,j,k,l,m=1}^n (X_j^T X_k) \left( X_k^T \int t \sin(t^T X_j) \cos(t^T X_i) \varphi_\beta(t) dt \right).$$

$$\begin{aligned}
& (X_l^T X_k) \left( X_k^T \int t \sin(t^T X_l) \cos(t^T X_m) \varphi_\beta(t) dt \right) \\
& - \frac{1}{n^4} \sum_{i,j,k,l=1}^n (X_j^T X_k) \left( X_k^T \int t \sin(t^T X_j) \cos(t^T X_i) \varphi_\beta(t) dt \right) \cdot \\
& \quad (X_l^T X_k) \left( X_k^T \int t \sin(t^T X_l) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt \right) \\
& - \frac{1}{n^4} \sum_{j,k,l,m=1}^n (X_j^T X_k) \left( X_k^T \int t \sin(t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt \right) \cdot \\
& \quad (X_l^T X_k) \left( X_k^T \int t \sin(t^T X_l) \cos(t^T X_m) \varphi_\beta(t) dt \right) \\
& + \frac{1}{n^3} \sum_{j,k,l=1}^n (X_j^T X_k) \left( X_k^T \int t \sin(t^T X_j) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt \right) \cdot \\
& \quad (X_l^T X_k) \left( X_k^T \int t \sin(t^T X_l) e^{-\frac{\|t\|^2}{2}} \varphi_\beta(t) dt \right) + o_{f.s.}(1),
\end{aligned}$$

und die Behauptung folgt aus Theorem D.2. Man beachte, daß auch hierfür die Existenz der Momente vierter Ordnung ausreicht, denn es ist z.B.

$$\begin{aligned}
& \frac{1}{n^5} \sum_{i,j,k,l,m=1}^n (X_j^T X_k) \left( X_k^T \int t \sin(t^T X_j) \cos(t^T X_i) \varphi_\beta(t) dt \right) \\
& \quad (X_l^T X_k) \left( X_k^T \int t \sin(t^T X_l) \cos(t^T X_m) \varphi_\beta(t) dt \right) \\
& = \sum_{\substack{r_1, r_2, \\ r_3, r_4=1}}^d \left( \frac{1}{n} \sum_{k=1}^n X_{k,r_1} X_{k,r_2} X_{k,r_3} X_{k,r_4} \right) \left( \frac{1}{n^2} \sum_{i,j=1}^n X_{j,r_1} \int t_{r_3} \sin(t^T X_j) \cos(t^T X_i) \varphi_\beta(t) dt \right) \\
& \quad \left( \frac{1}{n^2} \sum_{l,m=1}^n X_{l,r_2} \int t_{r_4} \sin(t^T X_l) \cos(t^T X_m) \varphi_\beta(t) dt \right)
\end{aligned}$$

wobei  $X_{j,l}$  bzw.  $t_l$  ( $l = 1, \dots, d$ ) die  $d$  Komponenten des Vektors  $X_j$  bzw.  $t$  bezeichnen.  $\square$

## 2. Die Bowman–Foster Statistik

Im Kontext der Konstruktion von Anpassungstests auf multivariate Normalverteilung wurde in Kapitel 1 die auf der empirischen charakteristischen Funktion  $\Psi_n(t) = \frac{1}{n} \sum_{j=1}^n \exp\{it^T Y_j\}$ ,  $t \in \mathbb{R}^d$ , der mit dem empirischen Mittel  $\bar{X}_n := \frac{1}{n} \sum_{j=1}^n X_j$  und der empirischen Kovarianzmatrix  $S_n = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T$  standardisierten Daten

$$Y_j := S_n^{-1/2}(X_j - \bar{X}_n) \quad , \quad j = 1, \dots, n,$$

basierende BHEP–Statistik

$$T_{n,\beta} = n \int_{\mathbb{R}^d} \left| \Psi_n(t) - e^{-\frac{\|t\|^2}{2}} \right|^2 \varphi_\beta(t) dt \quad (2.1)$$

betrachtet. Man beachte an dieser Stelle noch einmal, daß die empirische Kovarianzmatrix  $S_n$  nach Eaton und Perlman (1973) fast sicher nichtsingulär ist, falls alle  $(d - 1)$ –dimensionalen Hyperebenen des  $\mathbb{R}^d$  unter der Verteilung von  $X$  das Maß 0 besitzen und der Stichprobenumfang  $n$  echt größer als die Dimension  $d$  ist. Wie in Kapitel 1 nehmen wir stets an, daß dies erfüllt ist (vgl. die Ausführungen auf Seite 6).

Um die Rolle des Gewichtsparameters  $\beta$  in der BHEP–Statistik genauer zu beleuchten, stellten Henze und Zirkler (1990) auf Seite 3599 die nachfolgenden Überlegungen an. Diese Betrachtungen sind in gleicher Weise für den fast sicheren Grenzwert  $L_\beta$  von  $\frac{1}{n} T_{n,\beta}$  durchführbar, was in Abschnitt 1.3.4 auf Seite 33 bei der genaueren Untersuchung von  $L_\beta$  angesprochen wurde.

Die Fourier–Transformierte  $\hat{g}(t) := \int \exp\{it^T x\} g(x) dx$  einer Funktion  $g \in L^2(\mathbb{R}^d, \mathfrak{B}^d, dx)$  liegt ebenfalls in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, dx)$ , und die *Formel von Plancherel* aus der Fourier–Analysis besagt, daß zwischen den beiden  $L^2$ –Normen der folgende Zusammenhang besteht:

$$\int_{\mathbb{R}^d} |\hat{g}(t)|^2 dt = (2\pi)^d \int_{\mathbb{R}^d} |g(t)|^2 dt$$

(vgl. z.B. Rudin (1987)). Setzt man die Dichte  $\varphi_\beta(t) = (2\pi\beta^2)^{-d/2} \exp\{-\frac{\|t\|^2}{2\beta^2}\}$  von  $\mathcal{N}_d(0, \beta^2 I_d)$  in (2.1) ein, erhält man die Darstellung

$$T_{n,\beta} = n \frac{1}{(2\pi\beta^2)^{d/2}} \int_{\mathbb{R}^d} \left| \Psi_n(t) e^{-\frac{\|t\|^2}{4\beta^2}} - e^{-\frac{1}{2} \left[ \frac{2\beta^2+1}{2\beta^2} \right] \|t\|^2} \right|^2 dt .$$

Nun stellt man fest, daß mit  $h := 1/(\sqrt{2}\beta)$  und  $Q_n := \frac{1}{n} \sum_{j=1}^n \delta_{Y_j}$  die Funktion  $\Psi_n(t) \exp\{-\frac{\|t\|^2}{4\beta^2}\}$  die Fourier–Transformierte der Faltung  $Q_n * \mathcal{N}_d(0, h^2 I_d)$  ist, wobei  $\delta_{\{\cdot\}}$  das

Dirac–Maß bezeichnet, und daß mit  $\frac{2\beta^2+1}{2\beta^2} = 1 + h^2$  die Funktion  $\exp\left\{-\frac{1}{2}\left[\frac{2\beta^2+1}{2\beta^2}\right]\|t\|^2\right\}$  die Fourier–Transformierte der Normalverteilung  $\mathcal{N}_d(0, (1+h^2)I_d)$  ist. Damit liefert die Formel von Plancherel

$$T_{n,\beta} = n \frac{(2\pi)^{d/2}}{\beta^d} \int_{\mathbb{R}^d} \left| \frac{1}{nh^d} \sum_{j=1}^n \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x-Y_j\|^2}{2h^2}} - \frac{1}{(2\pi(1+h^2))^{d/2}} e^{-\frac{\|x\|^2}{2(1+h^2)}} \right|^2 dx .$$

Die BHEP–Statistik besitzt also eine Darstellung als quadrierter Abstand in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, dx)$  zwischen dem auf dem  $d$ –dimensionalen *Gaußkern*

$$K(x) := \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x\|^2}{2}} \quad (2.2)$$

und den standardisierten Daten  $Y_j = S_n^{-1/2}(X_j - \bar{X}_n)$  beruhenden nichtparametrischen multivariaten *Kerndichteschätzer*

$$\hat{f}_{n,\text{st}}(x) := \frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{x-Y_j}{h}\right) = \frac{1}{nh^d} \sum_{j=1}^n \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x-Y_j\|^2}{2h^2}}$$

mit *Bandbreite*  $h$  und einem deterministischen Term. Schreibt man

$$\hat{f}_n(x) := \frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right) = \frac{1}{nh^d} \sum_{j=1}^n \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x-X_j\|^2}{2h^2}}$$

für den entsprechenden auf den ursprünglichen unabhängigen  $d$ –dimensionalen Zufallsvektoren  $X_1, \dots, X_n$  anstelle von  $Y_1, \dots, Y_n$  basierenden Kerndichteschätzer, so stellt sich dieser deterministische Term gerade als der Erwartungswert von  $\hat{f}_n(x)$  unter der Hypothese einer  $d$ –dimensionalen Standard–Normalverteilung heraus:

$$\mathbb{E}_{\mathcal{N}_d(0, I_d)} [\hat{f}_n(x)] = \mathbb{E}_{\mathcal{N}_d(0, I_d)} \left[ \frac{1}{h^d (2\pi)^{d/2}} e^{-\frac{\|x-X_1\|^2}{2h^2}} \right] = \frac{1}{(2\pi(1+h^2))^{d/2}} e^{-\frac{\|x\|^2}{2(1+h^2)}}. \quad (2.3)$$

Wählt man nun die Bandbreite  $h$  nicht fest, sondern in Abhängigkeit des Stichprobenumfangs  $n$  in einem gewissen Sinne *optimal* (vgl. die Ausführungen auf Seite 76), d.h. gemäß

$$h = h(n) := \left(\frac{4}{2d+1}\right)^{\frac{1}{d+4}} \left(\frac{1}{n}\right)^{\frac{1}{d+4}}, \quad (2.4)$$

und damit den Gewichtsparameter  $\beta$  gemäß der Vorschrift

$$\beta = \beta(n) = \frac{1}{\sqrt{2}h} = \frac{1}{\sqrt{2}} \left(\frac{2d+1}{4}\right)^{\frac{1}{d+4}} n^{\frac{1}{d+4}}, \quad (2.5)$$

so gelangt man zur *Bowman–Foster–Statistik*, die von Bowman und Foster (1993) aus dem Kontext der Kerndichteschätzung heraus zur Realisierung eines Anpassungstests auf multivariate Normalverteilung vorgeschlagen wurde:

$$T_{n,h,\text{st}} := nh^d 2^d \pi^{d/2} \int_{\mathbb{R}^d} \left( \hat{f}_{n,\text{st}}(x) - \mathbb{E}_{\mathcal{N}_d(0, I_d)} [\hat{f}_n(x)] \right)^2 dx. \quad (2.6)$$

Hinter diesem Ansatz steckt dieselbe Idee wie hinter dem BHEP–Test: Trifft die Hypothese nicht zu, wird der integrierte Abstand zwischen dem Kerndichteschätzer und dessen Erwartungswert unter  $H_0$  zu groß, und der zugehörige Test lehnt die Hypothese ab.

Damit sind wir bei der bereits in Henze (1997) vermerkten überraschenden Beobachtung angelangt, daß zwei a priori völlig unterschiedliche Konzepte zu beinahe identischen Verfahren führen: Die BHEP–Statistik basiert auf dem gewichteten  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \varphi_\beta(t)dt)$ –Abstand der empirischen charakteristischen Funktion und ihrem Erwartungswert unter  $H_0$  mit festem Gewichtsparameter  $\beta$ . Die Bowman–Foster–Statistik hingegen beruht auf dem  $L^2(\mathbb{R}^d, \mathfrak{B}^d, dx)$ –Abstand zwischen dem Kerndichteschätzer  $\hat{f}_{n,st}$  und seinem Erwartungswert unter der Hypothese, mit variabler Bandbreite  $h$ , und damit über die Koppelung  $\beta = 1/(\sqrt{2}h)$  auch variablem Gewichtsparameter  $\beta$ .

Die Asymptotik der BHEP–Statistik bei festem Gewichtsparameter  $\beta$  ist mittlerweile geklärt: Die Grenzverteilung unter  $H_0$  wurde von Baringhaus und Henze (1988), Henze und Zirkler (1990) bzw. Henze und Wagner (1997) erhalten. Henze und Wagner (1997) behandelten das Verhalten unter einer Folge von benachbarten Alternativen, die mit Geschwindigkeit  $n^{-1/2}$  gegen die Klasse der Normalverteilungen konvergiert. Die Grenzverteilung unter festen Alternativen wurde in Kapitel 1 der vorliegenden Arbeit bestimmt.

Es stellt sich nun die Frage nach dem Verhalten bei variablem Gewichtsparameter  $\beta$ , also nach der Asymptotik der Bowman–Foster–Statistik unter der Hypothese einer  $d$ –dimensionalen Normalverteilung sowie unter festen und unter lokalen Alternativen. Dies ist Gegenstand von Kapitel 2 der vorliegenden Arbeit. Die Problemstellung ist keineswegs neu und teilweise gelöst, da die Bowman–Foster–Statistik eng verwandt ist mit dem aus der Theorie der Kerndichteschätzung stammenden Bickel–Rosenblatt–Test, der in zahlreichen Veröffentlichungen untersucht wurde. Deshalb beginnen wir in Abschnitt 2.1 mit einem Überblick über den Stand der Forschung, in dem die relevante Literatur und die bisher erzielten Resultate vorgestellt werden. Dabei wird Bilanz über die bisher gelösten und die noch offenen Fragen gezogen und das weitere Vorgehen skizziert.

Es kann allerdings nicht Aufgabe der vorliegenden Arbeit sein, die umfangreiche Literatur zum Gebiet der Kerndichteschätzung umfassend zu behandeln. Zu diesem äußerst aktiven Forschungsgebiet wurden allein in den letzten 15 Jahren Myriaden von Artikeln veröffentlicht. Wir beschränken uns deshalb darauf, das im vorliegenden Kontext unmittelbar Notwendige zu schildern. Umfassendere Darstellungen dieses Gebietes findet man z.B. in den Büchern von Prakasa Rao (1983), Devroye und Györfi (1985), Devroye (1987) und Silverman (1986), denen auch Teile der allgemeinen Bemerkungen zum Thema Kerndichteschätzung entnommen sind. Der Fall der multivariaten Kerndichteschätzung ist z.B. in Prakasa Rao (1983), Kapitel 3 ausführlich beschrieben.

## 2.1. Historischer Kontext

Das Gebiet der nichtparametrischen Dichteschätzung befaßt sich mit dem Problem, anhand von unabhängigen und identisch verteilten  $d$ –dimensionalen Zufallsvektoren  $X_1, \dots, X_n$ , denen man eine zugrundeliegende Lebesgue–Dichte  $f$  unterstellt, diese unbekannte Größe zu schät-

zen. In diesem Zusammenhang haben sich besonders die Kerndichteschätzer bewährt, die auf einer Kernfunktion  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  basieren. Ein klassischer Vertreter dieser Spezies ist der von Rosenblatt (1956) und Parzen (1962) vorgeschlagene *Rosenblatt–Parzen–Schätzer* mit Bandbreite  $h$ , der definiert ist durch

$$\hat{f}_n(x) := \frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right) \quad (2.7)$$

und auf den wir uns in dieser Arbeit beschränken wollen. Andere Möglichkeiten wie z.B. Histogramm–Dichteschätzer werden hier nicht betrachtet. Die Bowman–Foster–Statistik beruht also auf einem speziellen Rosenblatt–Parzen–Schätzer, nämlich auf der Variante, die den Exponentialkern aus (2.2) verwendet. Im gesamten weiteren Verlauf gehen wir von der folgenden Voraussetzung 1 aus, die die im Kontext der Kerndichteschätzung gängigen Voraussetzungen wiedergibt (vgl. Hall (1984), S. 7).

**Voraussetzung 1** Für die Bandbreite  $h$  gelte  $h \rightarrow 0$  und  $nh^d \rightarrow \infty$  für  $n \rightarrow \infty$ . Der beschränkte, nichtnegative Kern  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  genüge den Bedingungen

$$\int_{\mathbb{R}^d} K(x) dx = 1, \quad \int_{\mathbb{R}^d} x_i K(x) dx = 0, \quad \int_{\mathbb{R}^d} x_i x_j K(x) dx = 2k \mathbf{1}_{\{i=j\}}$$

für alle  $i, j \in \{1, \dots, d\}$ , wobei die Konstante  $k$  nicht von  $i$  abhängt. Die zugrundeliegende Dichte  $f$  sei zweimal stetig differenzierbar;  $f$  und ihre partiellen Ableitungen zweiter Ordnung seien beschränkt und gleichmäßig stetig auf  $\mathbb{R}^d$ .

Ein solcher Kern  $K$  ist insbesondere eine Wahrscheinlichkeitsdichte. Nach Konstruktion erbt  $\hat{f}_n$  alle Stetigkeits– und Differenzierbarkeitseigenschaften von  $K$  und ist selbst wieder eine Dichte. Offensichtlich genügt der für uns interessante Gaußkern aus (2.2) der Voraussetzung 1. Für die nachfolgenden Betrachtungen wird nicht immer jede der dort geforderten Eigenschaften benötigt, wohl aber für die uns interessierenden Resultate. In der umfangreichen Literatur über Kerndichteschätzung wurden zahlreiche Aussagen unter teilweise sehr unterschiedlichen Voraussetzungen bewiesen. Da der Schwerpunkt der vorliegenden Arbeit nicht auf der Untersuchung von Kernschätzern als solchen liegt, beschränken wir uns auf den angedeuteten Rahmen, der zwar nicht der allgemeinste ist, aber unter den Zielvorgaben sinnvoll erscheint.

### 2.1.1. Eigenschaften von Kerndichteschätzern

Die Konsistenz des Schätzers  $\hat{f}_n$  wurde zunächst im Fall  $d = 1$  von Silverman (1978), Silverman (1980), Bertrand-Retali (1974) oder Bertrand-Retali (1978) untersucht. Eine ausführliche Betrachtung des multivariaten Falles findet man in Prakasa Rao (1983), Kapitel 3, dessen Theoreme 3.1.5, 3.1.6 und 3.1.7 Bedingungen für die punktweise und gleichmäßige Konsistenz von  $\hat{f}_n$  angeben: Gilt zusätzlich zu Voraussetzung 1 die Konvergenz  $\lim_{\|x\| \rightarrow \infty} \|x\|^d K(x) = 0$  und für alle  $\alpha > 0$  die Beziehung  $\sum_{n=1}^{\infty} \exp\{-\alpha n h^d\} < \infty$ , so folgt für alle  $x \in \mathbb{R}^d$

$$\hat{f}_n(x) \rightarrow f(x) \quad \text{f.s.}$$

Ferner gilt unter der Voraussetzung 1

$$\sup_x \left| \mathbb{E} \widehat{f}_n(x) - f(x) \right| \rightarrow 0 .$$

Sind außer der Voraussetzung 1 die Bedingungen  $nh^d/\log n \rightarrow \infty$ ,  $\sup_x \|x\|^d K(x) < \infty$ ,  $\sup_x |K(x+y) - K(x)| \leq \text{const} \|y\|$ ,  $y \in \mathbb{R}^d$  und  $\int_{\mathbb{R}^d} \|x\|^\gamma f(x) dx < \infty$  für ein  $\gamma > 0$  erfüllt, so folgt

$$\sup_x \left| \widehat{f}_n(x) - f(x) \right| \rightarrow 0 \quad \text{f.s.}$$

Ein weit verbreitetes Maß für die globale Diskrepanz zwischen dem Kerndichteschätzer  $\widehat{f}_n$  und der zugrundeliegenden Dichte  $f$  ist der *mittlere integrierte quadratische Fehler (mean integrated square error, MISE)*, der definiert ist durch

$$\text{MISE} := \int_{\mathbb{R}^d} \mathbb{E} \left[ \left( \widehat{f}_n(x) - f(x) \right)^2 \right] dx = \text{BIAS} + \text{VAR} \quad (2.8)$$

und der aus zwei Anteilen besteht. Der erste ist der deterministische Bias

$$\text{BIAS} := \int_{\mathbb{R}^d} \left( \mathbb{E} \widehat{f}_n(x) - f(x) \right)^2 dx ,$$

ein systematischer Fehler, denn es ist

$$\mathbb{E}_f \left[ \widehat{f}_n(x) \right] = \frac{1}{h^d} \int_{\mathbb{R}^d} K \left( \frac{x-t}{h} \right) f(t) dt = \int_{\mathbb{R}^d} K(u) f(x-hu) du \neq f(x) ,$$

und der zweite ist die integrierte Varianz von  $\widehat{f}_n(x)$ :

$$\text{VAR} := \int_{\mathbb{R}^d} \mathbb{E} \left[ \left( \widehat{f}_n(x) - \mathbb{E} \widehat{f}_n(x) \right)^2 \right] dx .$$

Ein Ziel ist es, die Bandbreite  $h$  so zu wählen, daß der mittlere quadratische Fehler möglichst klein wird. In diesem Zusammenhang spricht man auch von *optimaler Glättung* der Daten. Eine Taylor–Entwicklung für  $h \rightarrow 0$  der beiden Bestandteile von MISE (vgl. Silverman (1986), S. 38 ff) zeigt, daß zur Eliminierung des Bias die Bandbreite  $h$  sehr klein gewählt werden sollte, was aber die Varianz vergrößert. Zur Reduktion der Varianz hingegen wäre eine größere Bandbreite  $h$  wünschenswert, was wiederum den Bias erhöht. Variiert man diese beiden Effekte so aus, daß der führende Term von MISE in der Taylor–Entwicklung minimal wird, so ergibt sich die zunächst vom Kern  $K$ , der Dichte  $f$  und dem Stichprobenumfang  $n$  abhängige *asymptotisch optimale Bandbreite*

$$h = h(n) := \left( \frac{d \cdot \int K(x)^2 dx}{\left[ \int x_1^2 K(x) dx \right]^2 \int (\nabla^2 f(x))^2 dx} \right)^{\frac{1}{d+4}} \left( \frac{1}{n} \right)^{\frac{1}{d+4}}$$

(vgl. Silverman (1986), S. 40 und S. 85–87, bzw. Parzen (1962)). Bei zugrundeliegender Normalverteilung, d.h. mit der Dichte  $f(x) := (2\pi)^{-d/2} \exp\{-\frac{\|x\|^2}{2}\}$ , und dem Gaußkern aus Gleichung (2.2) besitzt diese Bandbreite gerade die bereits in Gleichung (2.4) angegebene Gestalt

$$h = \left(\frac{4}{2d+1}\right)^{\frac{1}{d+4}} \left(\frac{1}{n}\right)^{\frac{1}{d+4}}$$

(vgl. Silverman (1986), S. 85–87). Wenn in den folgenden Kapiteln von *optimaler Wahl der Bandbreite*  $h$  die Rede ist, meinen wir allgemein (d.h. ohne Berücksichtigung eines speziellen Kerns oder einer bestimmten Dichte) die Festlegung

$$h := \kappa_d \cdot n^{-\frac{1}{d+4}}, \quad \text{d.h.} \quad nh^{d+4} = \kappa_d, \quad (2.9)$$

mit einer von der Dimension  $d$  abhängigen Konstanten  $\kappa_d$ , die im Normalverteilungsfall und für den Gaußkern die Gestalt

$$\kappa_d = \left(\frac{4}{2d+1}\right)^{\frac{1}{d+4}}$$

annimmt.

### 2.1.2. Der Bickel–Rosenblatt–Test

Das Hauptaugenmerk unserer Betrachtungen liegt auf der Konstruktion von Anpassungstests auf der Grundlage von Kerndichteschätzern. Es seien hierzu  $f$  eine feste Lebesgue–Dichte auf  $\mathbb{R}^d$  und

$$\mathcal{F} := \{f(\cdot, \vartheta) \mid \vartheta \in \Theta\}$$

eine parametrisierte Familie  $d$ –dimensionaler Dichten. Anhand von unabhängigen und identisch verteilten Zufallsvektoren  $X_1, \dots, X_n$ , von denen man annimmt, daß ihre zugrundeliegende Verteilung eine Dichte  $g$  besitzt, soll entweder die einfache Hypothese

$$H_0 : g = f \quad \text{gegen} \quad H_1 : g \neq f$$

oder die zusammengesetzte Hypothese

$$H_0 : g \in \mathcal{F} \quad \text{gegen} \quad H_1 : g \notin \mathcal{F}$$

getestet werden. Ähnlich wie mit der empirischen Verteilungsfunktion, der empirischen charakteristischen Funktion oder anderen empirischen Größen, die dem charakterisierenden Merkmal einer Verteilung entsprechen, kann man auch anhand von Kerndichteschätzern Anpassungstests konstruieren. Der klassische *Bickel–Rosenblatt–Test* für die einfache Hypothese beruht auf dem Rosenblatt–Parzen–Schätzer aus Gleichung (2.7), und die zugehörige *Bickel–Rosenblatt–Statistik* besitzt die Gestalt

$$T_{n,h,f} := \int_{\mathbb{R}^d} \left(\hat{f}_n(x) - f(x)\right)^2 dx, \quad (2.10)$$

mit der Maßgabe, daß zu große Abweichungen zwischen Schätzer und  $H_0$ -Dichte  $f$  zur Ablehnung der Hypothese führen.

Der Erwartungswert von  $T_{n,h,f}$  ist gerade der mittlere quadratische Fehler MISE aus Gleichung (2.8), der aus dem systematischen Fehler und der integrierten Varianz von  $\hat{f}_n(x)$  besteht. Um ersteren zu eliminieren, wird die Statistik häufig nicht mit  $f(x)$  zentriert, sondern mit  $E_f[\hat{f}_n(x)]$ , was auf die Modifikation

$$T_{n,h} := \int_{\mathbb{R}^d} \left( \hat{f}_n(x) - E_f[\hat{f}_n(x)] \right)^2 dx$$

führt, deren Erwartungswert nur noch aus der integrierten Varianz von  $\hat{f}_n(x)$  besteht. Die Zerlegung

$$T_{n,h,f} = T_{n,h} + 2 \int_{\mathbb{R}^d} \left( \hat{f}_n(x) - E_f[\hat{f}_n(x)] \right) \left( E_f[\hat{f}_n(x)] - f(x) \right) dx + \int_{\mathbb{R}^d} \left( E_f[\hat{f}_n(x)] - f(x) \right)^2 dx$$

zeigt, daß der Unterschied zwischen den Grenzverteilungen von  $T_{n,h}$  und  $T_{n,h,f}$  nicht wesentlich ist: Er besteht neben dem deterministischen Bias nur aus dem gemischten Glied, dessen Asymptotik als Summe von unabhängigen und identisch verteilten Zufallsvariablen leicht mit dem Zentralen Grenzwertsatz zu behandeln ist. Da  $T_{n,h}$  den systematischen Fehler nicht enthält und da die uns interessierende Bowman–Foster–Statistik von demselben Typ ist, werden wir uns auf diese Statistik konzentrieren.

Beim Testen der zusammengesetzten Hypothese stößt man zusätzlich auf das Problem, den unbekannt Parameter  $\vartheta$  schätzen zu müssen. Ist die parametrisierte Familie  $\mathcal{F}$   $d$ -dimensionaler Dichten eine Lokations–Skalen–Familie, so existieren hierfür zwei grundsätzlich verschiedene Lösungsansätze: Man kann anstelle der ursprünglichen Daten  $X_1, \dots, X_n$  die standardisierten Daten  $Y_1, \dots, Y_n$  in die Statistik einsetzen, oder man kann in der Zentrierung der Statistik mit  $E_{f(\cdot, \vartheta)}[\hat{f}_n(x)]$  an der Stelle von  $\vartheta$  einen Schätzer  $\hat{\vartheta}_n$  verwenden.

Der erste Ansatz taucht bisher nur in Bowman (1992) bzw. Bowman und Foster (1993) auf, allerdings ohne jede theoretische Untersuchung. Er führt im Falle eines Tests auf multivariate Normalverteilung auf die in Gleichung (2.6) hergeleitete Bowman–Foster–Statistik

$$\begin{aligned} T_{n,h,\text{st}} &= nh^d 2^d \pi^{d/2} \int_{\mathbb{R}^d} \left( \hat{f}_{n,\text{st}}(x) - E_{\mathcal{N}_d(0, I_d)}[\hat{f}_n(x)] \right)^2 dx & (2.11) \\ &= nh^d 2^d \pi^{d/2} \int_{\mathbb{R}^d} \left( \frac{1}{nh^d} \sum_{j=1}^n \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x - Y_j\|^2}{2h^2}} - \frac{1}{(2\pi(1+h^2))^{d/2}} e^{-\frac{\|x\|^2}{2(1+h^2)}} \right)^2 dx \\ &= \frac{1}{n} \sum_{j,k=1}^n e^{-\frac{\|Y_j - Y_k\|^2}{4h^2}} - \frac{2^{\frac{d+2}{2}} h^d}{(2h^2 + 1)^{d/2}} \sum_{j=1}^n e^{-\frac{\|Y_j\|^2}{2(1+2h^2)}} + \frac{nh^d}{(h^2 + 1)^{d/2}}, \end{aligned}$$

wobei sich die die letzte Zeile durch Integration nach  $x$  für festes  $\omega \in \Omega$  ergibt (vgl. Gleichung (1.2) auf Seite 6). Da die Daten in die Berechnung der Bowman–Foster–Statistik nur

über  $\|Y_j\|^2$  bzw.  $\|Y_j - Y_k\|^2$  eingehen, besitzt diese die wichtige Eigenschaft der affinen Invarianz (vgl. auch Seite 7 im BHEP-Kontext).

Der zweite Ansatz fand vielfältige Beachtung in der Literatur: Liegt den Daten die Dichte  $f(x, \vartheta)$  mit festem Parameter  $\vartheta$  zugrunde, so ist

$$T_{n,h}(\vartheta) := \int_{\mathbb{R}^d} \left( \widehat{f}_n(x) - E_{f(\cdot, \vartheta)} \left[ \widehat{f}_n(x) \right] \right)^2 dx$$

die zugehörige Bias-korrigierte Bickel-Rosenblatt-Statistik. Zum Testen der zusammengesetzten Hypothese verwendet man an ihrer Stelle die Modifikation

$$T_{n,h}(\widehat{\vartheta}_n) := \int_{\mathbb{R}^d} \left( \widehat{f}_n(x) - E_{f(\cdot, \widehat{\vartheta}_n)} \left[ \widehat{f}_n(x) \right] \right)^2 dx \quad (2.12)$$

mit dem Schätzer  $\widehat{\vartheta}_n$  für den unbekanntem  $H_0$ -Parameter  $\vartheta$ . Im Falle des Tests auf multivariate Normalverteilung mit dem konsistenten Schätzer  $\widehat{\vartheta}_n = (\bar{X}_n, S_n)$  für  $\vartheta = (\mu, \Sigma)$  ergibt sich die Darstellung

$$\begin{aligned} T_{n,h}(\widehat{\vartheta}_n) &= \int_{\mathbb{R}^d} \left( \widehat{f}_n(x) - E_{f(\cdot, \widehat{\vartheta}_n)} \left[ \widehat{f}_n(x) \right] \right)^2 dx \\ &= \int_{\mathbb{R}^d} \left( \frac{1}{nh^d} \sum_{j=1}^n \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x-X_j\|^2}{2h^2}} - \frac{e^{-\frac{1}{2}\|(S_n+h^2I_d)^{-1/2}(x-\bar{X}_n)\|^2}}{(2\pi)^{d/2} |\det(S_n+h^2I_d)|^{1/2}} \right)^2 dx \\ &= \frac{1}{n^2} \sum_{i,j=1}^n \frac{e^{-\frac{1}{4h^2}\|X_i-X_j\|^2}}{(2h\sqrt{\pi})^d} - \frac{2}{n} \sum_{j=1}^n \frac{e^{-\frac{1}{2}\|(S_n+2h^2I_d)^{-1/2}(X_j-\bar{X}_n)\|^2}}{(2\pi)^{d/2} |\det(S_n+2h^2I_d)|^{1/2}} + \frac{(2\sqrt{\pi})^{-d}}{|\det(S_n+h^2I_d)|^{1/2}}, \end{aligned}$$

wobei

$$E_{\mathcal{N}_d(\mu, \Sigma)} \left[ \widehat{f}_n(x) \right] = E_{\mathcal{N}_d(\mu, \Sigma)} \left[ \frac{1}{h^d (2\pi)^{d/2}} e^{-\frac{\|x-X_1\|^2}{2h^2}} \right] = \frac{e^{-\frac{1}{2}(x-\mu)^T(\Sigma+h^2I_d)^{-1}(x-\mu)}}{(2\pi)^{d/2} |\det(\Sigma+h^2I_d)|^{1/2}}$$

analog zu Gleichung (2.3) auf Seite 73 direkt nachzurechnen ist. Anhand der ausintegrierten Darstellung der Bickel-Rosenblatt-Statistik stellt man fest, daß diese zwar translationsinvariant, im Gegensatz zur Bowman-Foster-Statistik aber nicht affin invariant, also mit einem entscheidenden Nachteil behaftet ist. Zur Asymptotik der Bickel-Rosenblatt-Statistik liegen allerdings umfangreiche theoretische Untersuchungen vor.

Die beiden vorgestellten Ansätze sind im Falle einer einfachen Hypothese ohne geschätzte Parameter äquivalent (man gleiche den Faktor  $nh^d 2^d \pi^{d/2}$  an), im Falle einer zusammengesetzten Hypothese jedoch nicht, wie man an der ausintegrierten Darstellung der beiden Statistiken sieht. Im folgenden sprechen wir daher immer von der *Bowman-Foster-Statistik*, wenn das erste Konzept mit den standardisierten Daten  $Y_1, \dots, Y_n$  gemeint ist, und von der *Bickel-Rosenblatt-Statistik*, wenn es sich um den zweiten Ansatz mit den ursprünglichen Daten  $X_1, \dots, X_n$  und dem Schätzer  $\widehat{\vartheta}_n$  handelt.

Schließlich führen wir für eine Funktion  $g \in L^2(\mathbb{R}^d, \mathfrak{B}^d, dx)$  die Bezeichnungen

$$\mu(g) := \int_{\mathbb{R}^d} K^2(x) dx \int_{\mathbb{R}^d} g(t) dt, \quad (2.13)$$

$$\sigma^2(g) := \int_{\mathbb{R}^d} (K * K)^2(x) dx \int_{\mathbb{R}^d} g^2(t) dt \quad (2.14)$$

ein. Üblicherweise ist  $g$  eine Wahrscheinlichkeitsdichte, so daß das zweite Integral in (2.13) gleich 1 ist. Betrachtet man jedoch allgemeiner anstelle des Lebesgue–Maßes ein Maß  $\nu$  auf  $\mathbb{R}^d$ , z.B.  $d\nu(x) = w(x)dx$  mit einer Gewichtsfunktion  $w$ , dann ist dies nicht mehr der Fall, so daß die Definition von  $\mu(g)$  in ihrer allgemeineren Form angegeben wurde. Ist  $f$  die wahre den Daten zugrundeliegende Dichte, so gibt  $\mu(f)$  den führenden Term des Erwartungswertes von  $T_{n,h}$  und  $\sigma^2(f)$  den führenden Term der Varianz von  $T_{n,h}$  bei einer Taylor–Entwicklung für  $h \rightarrow 0$  an, wie z.B. aus den Rechnungen in Abschnitt 2.3 ersichtlich wird. Bei dem führenden Term des Erwartungswertes von  $T_{n,h,f}$  sind entsprechende Bias–Korrekturen zu berücksichtigen.

### 2.1.3. Vorhandene Resultate zum Bickel–Rosenblatt–Test

Die Asymptotik der Statistik  $T_{n,h}$  unter der einfachen Hypothese  $H_0 : g = f$  bzw. unter einer entsprechenden zusammengesetzten Hypothese ist Gegenstand zahlreicher Veröffentlichungen und wurde mit mehreren, zum Teil sehr unterschiedlichen Methoden untersucht. Grundsätzlich handelt es sich dabei um Resultate der Form

$$\frac{nh^d T_{n,h} - \mu(f)}{\sqrt{2} h^{d/2} \sigma(f)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad (2.15)$$

die zunächst im Eindimensionalen, später in beliebiger Dimension  $d$  und unter diversen Voraussetzungen an den Kern  $K$  und die zugrundeliegende Dichte  $f$  bewiesen wurden. In einigen Fällen wurde beim Testen der einfachen Hypothese anstelle der Statistik  $nh^d T_{n,h}$  mit der Zentrierung  $\mu(f)$  die Statistik  $nh^d T_{n,h,f}$  mit einer entsprechenden Zentrierung betrachtet. Man beachte, daß im Gegensatz zur BHEP–Statistik mit festem Gewichtsparameter bei der Bickel–Rosenblatt–Statistik mit variabler Bandbreite die Skalierung  $nh^d$  und anschließende Zentrierung nötig sind, um zu einer nichttrivialen Grenzverteilung zu gelangen.

In dem grundlegenden Artikel von Bickel und Rosenblatt (1973) wurde zum ersten Mal die *Konstruktion von Komlós–Major–Tusnádý (KMT)* (vgl. Komlós, Major und Tusnádý (1975)), d.h. eine fast sichere Approximation des empirischen Prozesses durch eine Brownsche Brücke verwendet, um die Konvergenzaussage aus (2.15) zu beweisen. In den folgenden Jahren wurde diese Methodik u.a. von Rosenblatt (1976), Hall (1982), Csörgő und Horváth (1987), Csörgő und Horváth (1988) und Horváth (1990) verfeinert und erweitert, um die anfangs sehr restriktiven Voraussetzungen abzuschwächen. Csörgő, Gombay und Horváth (1991) wandten diese Technik auf den Fall zensierter Daten an. Problematisch an dem KMT–Ansatz ist zum einen, daß wesentliche Einschränkungen an die Rate, mit der die Bandbreite  $h$  gegen 0 strebt, benötigt werden; der Fall optimaler Glättung im Sinne von Seite 76 ist darin nicht enthalten. Zum anderen sind gewisse Einschränkungen an den Kern  $K$  wesentlich, z.B. ein beschränkter Träger. Folglich kann der uns interessierende Exponentialkern aus Gleichung (2.2) auf Seite 73 mit der KMT–Methode nicht behandelt werden. Mehrdimensionale Verallgemeinerungen der KMT–Konstruktion sind ebenfalls schwierig, und geschätzte Parameter im Fall einer zusammengesetzten Hypothese wurden mit dieser Methodik bisher gar nicht betrachtet.

Auf der Suche nach einer mehrdimensionalen Version der Konvergenz (2.15) verwendete Ro-

senblatt (1975) angesichts der Probleme mit dem KMT–Ansatz eine grundlegend andere Methode, die der *Poissonization*. Dabei wird der Stichprobenumfang  $n$  durch eine Poisson–verteilte Zufallsvariable mit Mittelwert  $n$  modelliert und der empirische Prozeß durch einen Poisson–Prozeß angenähert. Schließlich findet der Zentrale Grenzwertsatz von Lyapounov Anwendung. Weitere Arbeiten mit diesem Ansatz folgten, so z.B. Horváth (1991) und Beirlant und Mason (1995), die die Resultate von Rosenblatt (1975) unter verschiedenen Zielsetzungen verallgemeinerten. Diese Methode funktioniert im Mehrdimensionalen, benötigt aber wie die KMT–Konstruktion Kerne mit beschränktem Träger. Der Fall von geschätzten Parametern wurde in diesem Zusammenhang bisher nicht behandelt. Neumann (1998) betrachtete den Fall abhängiger Daten mit einem Poissonization–Ansatz.

Einen grundlegenden Fortschritt brachte der Artikel Hall (1984), der ein völlig neues Beweisprinzip einführte: Die Statistik  $T_{n,h}$  läßt sich auch als zentrierte, degenerierte  $U$ –Statistik mit variablem, vom Stichprobenumfang  $n$  abhängigen Kern schreiben. Hall (1984) leitete für derartige  $U$ –Statistiken einen Zentralen Grenzwertsatz her, indem er ihre Martingalstruktur erkannte und Brown’s *Zentralen Grenzwertsatz für Martingale* anwandte. Dieser Ansatz brachte völlig neue Perspektiven mit sich: Er funktioniert mehrdimensional genauso gut wie im Eindimensionalen, man benötigt nur minimale Voraussetzungen an die Bandbreite  $h$ , den Kern  $K$  und die zugrundeliegende Dichte  $f$ , und Schätzer für Parameter können relativ problemlos behandelt werden. In den Folgejahren wurden mehrere Arbeiten veröffentlicht, die dieses neue Konzept zur Behandlung der noch ungelösten Fragen einsetzten: Fan (1994) und Fan (1995) betrachteten zu der Statistik  $T_{n,h}$  in beliebiger Dimension  $d$  Kerne allgemeiner Ordnung  $m$  (nicht nur zweiter Ordnung wie Hall (1984)) und bestimmten eine (2.15) entsprechende Asymptotik unter einer zusammengesetzten Hypothese und unter lokalen Pitman– bzw. Sharp Peaked Alternativen für beliebige Raten der Bandbreite  $h$ . Die Konsistenz des Bickel–Rosenblatt–Tests und eine theoretische Behandlung der nötigen Bootstrap–Prozedur zur konkreten Durchführung des Tests wurden ebenfalls betrachtet. Nicht berücksichtigt wurden feste Alternativen und der Einsatz standardisierter Daten (also die Bowman–Foster–Statistik). Kim, Hong, Jeong und Yang (1997) behandelten eine zu (2.15) analoge Asymptotik unter verschiedenen Raten der Bandbreite  $h$  und für allgemeine Gewichtsfunktionen  $w(x)dx$  anstelle des Lebesgue–Maßes  $dx$ , allerdings nur im Eindimensionalen und unter der einfachen Hypothese. Feste Alternativen wurden ebenfalls behandelt, allerdings nur für  $d = 1$  und für die Statistik  $T_{n,h}$  ohne Schätzer. Liero, Läuter und Konakov (1998) schließlich betrachteten die Asymptotik vom Typ (2.15) im Eindimensionalen, für Kerne mit beschränktem Träger, für allgemeine Gewichtsfunktionen  $w(x)dx$ , unter einer zusammengesetzten Hypothese und unter lokalen Pitman– bzw. Sharp Peaked Alternativen. Sie untersuchten insbesondere die asymptotische Güte des Bickel–Rosenblatt Tests gegen beide Typen von Alternativen. Ghosh und Huang (1991) beschäftigten sich überdies mit der Wahl des Kerns  $K$ , so daß die asymptotische Güte des Tests maximal wird.

Fan und Li (1999), Fan und Ullah (1999), Takahata und Yoshihara (1987) und Neumann und Paparoditis (1998) verallgemeinerten die Methodik von Hall (1984) auf die Situation schwach abhängiger bzw. stationärer stochastischer Prozesse. Dabei wurden der mehrdimensionale Fall und die zusammengesetzte Hypothese behandelt.

Bowman (1992) und Bowman und Foster (1993) schließlich untersuchten in Simulationsstudien die auf den standardisierten Daten basierende Bowman–Foster–Statistik aus Gleichung (2.6)

bzw. (2.11) zum Testen auf multivariate Normalverteilung. Dabei wurden als einziger Kern  $K$  der Gaußkern aus Gleichung (2.2) auf Seite 73 verwendet und für  $h$  die asymptotisch optimale Bandbreite aus Gleichung (2.4). Theoretische Betrachtungen fehlen, wie bereits erwähnt, in beiden Publikationen.

### 2.1.4. Offene Probleme

Angesichts des zu Beginn von Kapitel 2 auf Seite 72 geschilderten Zusammenhangs zwischen der BHEP– und der Bowman–Foster–Statistik für einen Anpassungstest auf multivariate Normalverteilung ergab sich die Frage nach der Asymptotik der letzteren bei variablem Gewichtparameter  $\beta$  bzw. von  $n$  abhängiger Bandbreite  $h$  unter der Hypothese sowie unter festen und lokalen Alternativen. Diese Fragen bleiben auch angesichts der zahlreichen aufgeführten Publikationen unbeantwortet, da diese stets den Bickel–Rosenblatt–Test betrachten und da sich dieser im Fall der zusammengesetzten Hypothese von der Bowman–Foster–Statistik unterscheidet. Die dort erzielten Resultate lassen höchstens vermuten, wie die Ergebnisse im Bowman–Foster–Fall aussehen könnten.

Nach den Überlegungen aus Abschnitt 2.1.2 erscheint es lohnenswert, die Bowman–Foster–Statistik theoretisch zu untersuchen, denn aufgrund ihrer affinen Invarianz und damit Verteilungsfreiheit unter der Hypothese einer  $d$ –dimensionalen Standard–Normalverteilung besitzt sie einen erheblichen Vorteil gegenüber der Bickel–Rosenblatt–Statistik. Insbesondere kommt man bei Monte–Carlo–Simulationen ohne zeitaufwendiges Bootstrap–Verfahren aus. Untersuchungen wie in Fan (1995), wo der auf der Bickel–Rosenblatt–Statistik basierende parametrische Bootstrap–Test kritisch beleuchtet wird, werden im Bowman–Foster–Fall also gar nicht benötigt.

Gegenstand der nachfolgenden Abschnitte ist die Asymptotik der Bowman–Foster–Statistik unter  $H_0$  und unter festen bzw. lokalen Alternativen. Dabei formulieren wir einige Resultate für allgemeinere Lokations–Skalen–Familien als die der  $d$ –dimensionalen Normalverteilungen und für allgemeinere Kerne als den Exponentialkern aus Gleichung (2.2). Ähnlich wie im Fall der BHEP–Statistik aus Kapitel 1 zerfällt auch die Behandlung der Bowman–Foster–Statistik in zwei Teile: das „Herausrechnen“ der in den standardisierten Daten enthaltenen Schätzer und die Asymptotik der übriggebliebenen, dann auf unabhängigen und identisch verteilten Zufallsvektoren basierenden Statistik für eine einfache Hypothese. An dieser Stelle kann auf die bereits vorhandenen Resultate zurückgegriffen werden, was an geeigneter Stelle gekennzeichnet wird. Wir beginnen im nachfolgenden Abschnitt mit einer Festlegung des Rahmens, der allen weiteren Überlegungen zugrundeliegt.

## 2.2. Situation und Notationen

Die von einer  $d$ –dimensionalen Lebesgue–Dichte  $f$  mit existierenden zweiten Momenten erzeugte Lokations–Skalen–Familie ist definiert durch

$$\mathcal{F} := \left\{ f(x, \vartheta) = \frac{1}{|\det \Sigma|^{1/2}} f(\Sigma^{-1/2}(x - \mu)) \mid \vartheta = (\mu, \Sigma), \mu \in \mathbb{R}^d, \Sigma \text{ positiv definit} \right\}. \quad (2.16)$$

Für eine unbekannte  $d$ -dimensionale Verteilung mit Dichte  $g$  möchte man anhand von unabhängigen und identisch verteilten Realisierungen  $X_1, \dots, X_n$  testen, ob  $g$  der Familie  $\mathcal{F}$  angehört, also die zusammengesetzte Hypothese

$$H_0 : g \in \mathcal{F} \quad \text{gegen} \quad H_1 : g \notin \mathcal{F} \quad (2.17)$$

prüfen. Dies soll anhand der reskalierten Bowman–Foster–Statistik

$$\tilde{T}_{n,h,st} := nh^d \int_{\mathbb{R}^d} \left( \hat{f}_{n,st}(x) - e_h(x, f) \right)^2 dx \quad (2.18)$$

geschehen, wobei

$$\hat{f}_n(x) = \frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right) \quad \text{bzw.} \quad \hat{f}_{n,st}(x) = \frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{x - Y_j}{h}\right) \quad (2.19)$$

wie früher den auf den ursprünglichen Daten  $X_1, \dots, X_n$  bzw. auf den standardisierten Daten  $Y_j = S_n^{-1/2}(X_j - \bar{X}_n)$ ,  $j = 1, \dots, n$ , basierenden Rosenblatt–Parzen–Schätzer mit Kern  $K$  bezeichnet und

$$e_h(x, f) := E_f \left[ \hat{f}_n(x) \right] = \frac{1}{h^d} E_f \left[ K\left(\frac{x - X_1}{h}\right) \right] = \int_{\mathbb{R}^d} K(u) f(x - hu) du \quad (2.20)$$

den Erwartungswert des ersteren unter der Dichte  $f$ . Bei der Betrachtung des auf  $\tilde{T}_{n,h,st}$  basierenden asymptotischen Tests für die zusammengesetzte Hypothese aus (2.17) beschränken wir uns auf die Fälle, in denen die Bowman–Foster–Statistik  $\tilde{T}_{n,h,st}$  affin invariant ist, also z.B. den Fall  $d = 1$ , oder auf die Familie  $\mathcal{F}$  der  $d$ -dimensionalen Normalverteilungen und den Gaußkern  $K$  aus (2.2). Damit kann die zugrundeliegende Dichte o.B.d.A. als standardisiert angenommen werden, d.h. für einen  $d$ -dimensionalen Zufallsvektor  $X$  mit Dichte  $f$  gilt  $EX = 0$  und  $EXX^T = I_d$ . Insbesondere wird dadurch die Verteilungsfreiheit von  $\tilde{T}_{n,h,st}$  unter  $H_0$  sichergestellt, und bei der konkreten Durchführung des Tests kann auf ein aufwendiges Bootstrap–Verfahren verzichtet werden. Die im folgenden hergeleiteten Grenzwertsätze werden jedoch allgemeiner für standardisierte Dichten  $f$  und Kerne  $K$ , die gewissen Regularitätsbedingungen genügen, formuliert.

Im weiteren Verlauf ist außerdem die (mit der Bias–korrigierten Bickel–Rosenblatt–Statistik übereinstimmende) Bowman–Foster–Statistik ohne geschätzte Parameter

$$\tilde{T}_{n,h} := nh^d \int_{\mathbb{R}^d} \left( \hat{f}_n(x) - e_h(x, f) \right)^2 dx \quad (2.21)$$

von Bedeutung, die die einfache Hypothese

$$H_0 : g = f \quad \text{gegen} \quad H_1 : g \neq f$$

testet. Schließlich erinnern wir noch einmal an die Bezeichnungen

$$\mu(f) = \int_{\mathbb{R}^d} K^2(x) dx \quad \text{und} \quad \sigma^2(f) = \int_{\mathbb{R}^d} (K * K)^2(x) dx \int_{\mathbb{R}^d} f^2(t) dt \quad (2.22)$$

aus Gleichung (2.13) bzw. (2.14) auf Seite 79, die speziell bei zugrundeliegender  $d$ -dimensionaler Standard–Normalverteilung und für den Gaußkern aus (2.2) die Werte

$$\mu(f) = \frac{1}{2^d \pi^{d/2}} \quad \text{und} \quad \sigma^2(f) = \frac{1}{2^{\frac{5}{2}d} \pi^d} \quad (2.23)$$

annehmen. Integrale ohne explizit angegebenen Integrationsbereich laufen im folgenden stets über den gesamten  $\mathbb{R}^d$ ; Integrale über Vektoren bzw. Matrizen sind komponentenweise zu lesen.

An diesem Punkt stellt sich die Frage nach dem adäquaten Rahmen für das weitere Vorgehen, d.h. nach den Voraussetzungen an den Kern  $K$ , an die Dichte  $f$  und an die Bandbreite  $h$ , sowie nach den Beweismethoden, die sinnvollerweise einzusetzen sind. Angesichts der Gestalt von  $\tilde{T}_{n,h}$  bzw.  $\tilde{T}_{n,h,\text{st}}$  als  $L^2$ -Abstand zwischen dem Kerndichteschätzer  $\hat{f}_n$  bzw.  $\hat{f}_{n,\text{st}}$  und dessen Erwartungswert unter der Dichte  $f$  und angesichts der Behandlung der Asymptotik der BHEP-Statistik ebenfalls in einem  $L^2$ -Kontext (vgl. Kapitel 1) ist es naheliegend, auch die Statistik  $\tilde{T}_{n,h}$  derartigen Überlegungen zu unterziehen. Man würde also die schwache Konvergenz von  $c(n)(\hat{f}_n(\cdot) - e_h(\cdot, f))$  bzw. von  $c(n)(\hat{f}_{n,\text{st}}(\cdot) - e_h(\cdot, f))$  mit einer passenden Rate  $c(n)$  im Rahmen der in Anhang B erläuterten Theorie in dem Hilbertraum  $L^2(\mathbb{R}^d, \mathfrak{B}^d, dx)$  nachweisen und dann mit dem Abbildungssatz auf die Verteilungskonvergenz von  $\tilde{T}_{n,h}$  bzw.  $\tilde{T}_{n,h,\text{st}}$  schließen. Dies funktioniert im Kontext der Kerndichteschätzung jedoch aus strukturellen Gründen nicht, wie die folgende Bemerkung zeigt.

**Bemerkung 20** Ruymgaart (1998) betrachtete den Fall eindimensionaler unabhängiger und identisch verteilter Zufallsvariablen  $X_1, \dots, X_n$  mit einer gemeinsamen Dichte  $f \in L^2(\mathbb{R}, \mathfrak{B}, dx)$ , deren charakteristische Funktion gewissen Regularitätsbedingungen genügt. Desweiteren betrachtete er den auf dem speziellen Kern  $K(x) := \sin(x)/(\pi x)$  basierenden Rosenblatt–Parzen–Schätzer mit Bandbreite  $h = h(n)$ , mit dem Hinweis, daß sich die nachfolgenden Überlegungen auf beliebige nichtnegative Kerne  $K$  mit  $\int K(x)dx = 1$  übertragen lassen.

Die zentrierte Größe  $Z_n(\cdot) := \sqrt{n}(\hat{f}_n(\cdot) - e_h(\cdot, f))$  ist ein  $L^2(\mathbb{R}, \mathfrak{B}, dx)$ -wertiges Zufallselement mit einem Kovarianzoperator  $S_n$ . Ruymgaart (1998) zeigte, daß für jede Bandbreite der Gestalt  $h = h(n) = n^{-r}$  ( $0 < r < 1$ ) und für jedes  $g \in L^2(\mathbb{R}, \mathfrak{B}, dx)$  gilt:

$$\langle Z_n, g \rangle_{L^2} \xrightarrow{D} \mathcal{N}(0, \langle g, Sg \rangle_{L^2}), \quad n \rightarrow \infty,$$

wobei  $S : L^2(\mathbb{R}, \mathfrak{B}, dx) \rightarrow L^2(\mathbb{R}, \mathfrak{B}, dx)$  ein symmetrischer positiver, aber nicht nuklearer Operator ist, den Ruymgaart explizit angibt. Die endlichdimensionalen Verteilungen von  $Z_n$  konvergieren also gegen Normalverteilungen, und der einzige Kandidat für einen schwachen Grenzwert von  $Z_n$  in  $L^2(\mathbb{R}, \mathfrak{B}, dx)$  wäre ein Gaußprozeß mit Kovarianzoperator  $S$ . Dies kann aber nicht sein, da Kovarianzoperatoren von Gaußschen Zufallselementen in einem Hilbertraum nuklear sind (vgl. Anhang B.2.4 und B.4).

Die intuitive, elegante Behandlung der Asymptotik von  $\tilde{T}_{n,h}$  in  $L^2$  ist also leider nicht möglich, und es müssen andere Verfahren in Betracht gezogen werden. Bei der Schilderung der

vorhandenen Resultate zum Bickel–Rosenblatt–Test in Abschnitt 2.1.3 wurden unter anderem Techniken erwähnt, die auf der KMT–Konstruktion bzw. der Poissonization aufbauen. Beiden sind jedoch erhebliche Einschränkungen an den Kern (beschränkter Träger), an die Konvergenzrate der Bandbreite  $h$  und an die zugrundeliegende Dichte  $f$  gemeinsam. Ferner ist nicht klar, ob und wenn ja mit welchen Einschränkungen sich geschätzte Parameter behandeln lassen.

Deshalb arbeiten wir im folgenden mit der „simplen“ Methodik von Hall (1984), die auf Martingalmethoden bzw. auf der in Anhang D.2 zusammengefaßten Theorie der  $U$ –Statistiken mit variablem Kern beruht und die mit minimalen Voraussetzungen an die zugrundeliegende Dichte  $f$  sowie an den Kern  $K$  auskommt. Sie wird in Abschnitt 2.3 bei der Beweisskizze von Theorem 29 und in Anhang D.2 bei der Beweisskizze von Theorem D.5 erläutert. Der Vollständigkeit halber wiederholen wir noch einmal die grundlegende Voraussetzung 1 aus Abschnitt 2.1 (vgl. Hall (1984)), von der wir im gesamten weiteren Verlauf ausgehen.

**Voraussetzung 1** Für die Bandbreite  $h$  gelte  $h \rightarrow 0$  und  $nh^d \rightarrow \infty$  für  $n \rightarrow \infty$ . Der beschränkte, nichtnegative Kern  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  sei von zweiter Ordnung, d.h. er genüge den Bedingungen

$$\int K(x) dx = 1, \quad \int xK(x) dx = 0, \quad \int xx^T K(x) dx = 2kI_d$$

mit einer Konstanten  $k$ . Die zugrundeliegende Dichte  $f$  sei zweimal stetig differenzierbar;  $f$  und ihre partiellen Ableitungen zweiter Ordnung seien beschränkt und gleichmäßig stetig auf  $\mathbb{R}^d$ .

**Bemerkung 21** Aus der Voraussetzung 1 folgt die Beschränktheit des Gradienten  $\nabla f$ : Denn für eine eindimensionale zweimal stetig differenzierbare Funktion  $g : \mathbb{R} \rightarrow \mathbb{R}$  gilt nach dem Satz von Taylor mit einem Zwischenpunkt  $\xi \in (x, x+1)$

$$g'(x) = g(x+1) - g(x) - \frac{1}{2} g''(\xi).$$

Die Beschränktheit von  $g$  und  $g''$  impliziert also die Beschränktheit der ersten Ableitung  $g'$ . Wendet man diese Überlegung auf die  $d$  partiellen Ableitungen von  $f$  an, folgt die Behauptung.

**Bemerkung 22** Unter der Voraussetzung 1 besitzt die Dichte  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  die folgende Taylor–Entwicklung ersten Grades mit Restglied in Integraldarstellung (vgl. Satz C.16 auf Seite 165):

$$f(x - hy) = f(x) - h \cdot \nabla f(x)y + h^2 \cdot y^T \left( \int_0^1 (1-t) \nabla^2 f(x - thy) dt \right) y. \quad (2.24)$$

Dabei sind  $x, y$  Vektoren des  $\mathbb{R}^d$  und  $h$  ein Skalar;  $\nabla f$  bezeichnet den Gradienten von  $f$ , den wir als Zeilenvektor auffassen, und  $\nabla^2 f$  die Hesse–Matrix von  $f$ . Für das Restglied in (2.24) gilt die Abschätzung

$$\left| h^2 \cdot y^T \left( \int_0^1 (1-t) \nabla^2 f(x - thy) dt \right) y \right| \leq \|y\|^2 O(h^2), \quad (2.25)$$

wobei der Term  $O(h^2)$  gleichmäßig in  $x \in \mathbb{R}^d$  zu verstehen ist.

**Bemerkung 23** Die Glattheit der zugrundeliegenden Dichte  $f$  hat durchaus einen Einfluß auf die Asymptotik der Bowman–Foster–Statistik  $\tilde{T}_{n,h}$  bzw. der Bickel–Rosenblatt–Statistik  $T_{n,h,f}$  aus Gleichung (2.10) auf Seite 77: van Es (1997) betrachtete Dichten auf  $\mathbb{R}$  mit einer endlichen Anzahl von Sprüngen bzw. Knicken, die durch den Rosenblatt–Parzen–Schätzer mit Bandbreite  $h$  und nichtnegativem symmetrischem beschränktem Kern  $K$  mit Träger in  $[-1, 1]$  und  $\int K(x)dx = 1$  geschätzt wurden. Zunächst ergaben sich in den drei Fällen Sprünge, Knicke oder Glattheit unterschiedliche optimale Bandbreiten im Sinne der Minimierung der mittleren integrierten quadratischen Abweichung MISE (vgl. die Diskussion auf S. 76). Für Dichten mit zusätzlich beschränktem Träger leitete van Es (1997) teilweise unterschiedliche Grenzverteilungen der geeignet standardisierten Größen  $\tilde{T}_{n,h}$  bzw.  $T_{n,h,f}$  in den drei Fällen Sprünge, Knicke oder Glattheit her.

Voraussetzung 1 reicht aus, wenn die Asymptotik von  $\tilde{T}_{n,h}$  unter einer einfachen Hypothese mit zugrundeliegender Dichte  $f$  behandelt werden soll. Sobald jedoch die standardisierten Daten  $Y_j$  auftreten, d.h. sobald die Schätzer  $\bar{X}_n$  und  $S_n$  „herausgerechnet“ werden müssen, sind zusätzliche Eigenschaften vonnöten. Wir haben uns dafür entschieden, Taylor–Entwicklungen des Kerns  $K$  nach den Schätzern um den Punkt  $(0, I_d)$  zu verwenden, also von einer gewissen Glattheit auszugehen. Da uns vor allem der Gaußkern aus Gleichung (2.2) interessiert, stören wir uns nicht an Voraussetzungen wie der dreifachen stetigen Differenzierbarkeit von  $K$ . Ferner werden die Taylor–Entwicklungen nach  $\bar{X}_n$  und  $S_n$  stets um den Punkt  $(0, I_d)$  durchgeführt, d.h. unter der Annahme, daß die zugrundeliegende Dichte standardisiert ist. Im Hinblick auf die Anwendbarkeit der Ergebnisse auf andere Lokations–Skalen–Familien  $\mathcal{F}$  als die der Normalverteilungen und im Hinblick auf eine Untersuchung der Asymptotik von  $\tilde{T}_{n,h,st}$  unter festen Alternativen, die dann insbesondere die Konsistenz des auf der Bowman–Foster–Statistik basierenden Tests sicherstellt, soll die Dichte  $f$  im übrigen jedoch so allgemein wie möglich gehalten werden.

Die für die Taylor–Entwicklungen des Kerns  $K$  benötigten Eigenschaften werden zusammen mit der Existenz von Momenten vierter Ordnung der zugrundeliegenden Verteilung in Voraussetzung 2 festgehalten. Dabei bezeichnen wir mit  $\nabla K$  den Gradienten der reellwertigen Funktion  $K$ , den wir als ein Element von  $\mathcal{L}(\mathbb{R}^d; \mathbb{R})$ , d.h. als Zeilenvektor auffassen.  $\nabla^2 K$  ist die Hesse–Matrix von  $K$  und liegt in  $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R})$ . Schließlich schreiben wir  $\nabla^3 K(x) \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d, \mathbb{R}^d; \mathbb{R})$  für das Differential dritter Ordnung von  $K$ . Eine Definition dieser Räume, eine Einführung in den allgemeinen Differentialkalkül und eine ausführliche Herleitung der in der nachfolgenden Bemerkung 27 angegebenen Taylor–Entwicklungen befinden sich in Anhang C.

**Voraussetzung 2** Die Bandbreite  $h$  sei nun gemäß Gleichung (2.9) auf Seite 77 gewählt, d.h. es gilt  $nh^{d+4} = \text{const.}$  Zusätzlich zu Voraussetzung 1 sei der Kern  $K$  dreimal stetig differenzierbar mit

$$\int \|x\|^m K(x)dx < \infty, \quad \int \|x\|^m \|\nabla K(x)\| dx < \infty, \quad \int \|x\|^m \|\nabla^2 K(x)\| dx < \infty$$

( $m = 0, 1, 2, 3, 4$ ),

$$\int \|\nabla^2 K(x)\|^2 dx < \infty, \quad \int \|\nabla^3 K(x)\|^2 dx < \infty$$

und  $\|x\| K(x) \rightarrow 0$  sowie  $\|\nabla K(x)\| \rightarrow 0$  für  $\|x\| \rightarrow \infty$ . Für die zugrundeliegende Dichte  $f$  gelte zusätzlich zu Voraussetzung 1: Die Momente vierter Ordnung existieren;  $\int \|x\|^l \|\nabla f(x)\| dx < \infty$  ( $l = 0, 1, 2$ ).

**Bemerkung 24** Der uns interessierende Gaußkern aus Gleichung (2.2) genügt den Voraussetzungen 1 und 2.

**Bemerkung 25** Die Voraussetzungen  $\|x\| K(x) \rightarrow 0$  und  $\|\nabla K(x)\| \rightarrow 0$  für  $\|x\| \rightarrow \infty$  implizieren zusammen mit Stetigkeitsargumenten die Beschränktheit von  $\nabla K$ , die Eigenschaft  $K(x) \rightarrow 0$  ( $\|x\| \rightarrow \infty$ ) und

$$\int \nabla K(x) dx = 0, \quad \int \nabla^2 K(x) dx = 0. \quad (2.26)$$

Dabei beweist man die beiden Gleichungen in (2.26) ähnlich wie das folgende Lemma.

**Lemma 26** Unter der Voraussetzung 2 an den Kern  $K$  gilt:  $\int x \nabla K(x) dx = -I_d$ .

BEWEIS: Für  $r \in \{1, \dots, d\}$  gilt mit der Notation  $\tilde{x} := (x_1, \dots, x_{r-1}, x_{r+1}, \dots, x_d) \in \mathbb{R}^{d-1}$

$$\begin{aligned} \int x_r \frac{\partial}{\partial x_r} K(x) dx &= \int \left\{ \int x_r \frac{\partial}{\partial x_r} K(x) dx_r \right\} d\tilde{x} \\ &= \int \left\{ [x_r K(x)]_{-\infty}^{\infty} - \int K(x) dx_r \right\} d\tilde{x} = - \int K(x) dx = -1, \end{aligned}$$

denn nach Voraussetzung 2 ist  $\lim_{x_r \rightarrow \infty} x_r K(x) = \lim_{x_r \rightarrow -\infty} x_r K(x) = 0$ .

Für  $r, s \in \{1, \dots, d\}$ ,  $r \neq s$ , gilt

$$\int x_s \frac{\partial}{\partial x_r} K(x) dx = \int x_s \left\{ \int \frac{\partial}{\partial x_r} K(x) dx_r \right\} d\tilde{x} = \int x_s \{ [K(x)]_{-\infty}^{\infty} \} d\tilde{x} = 0,$$

denn nach Voraussetzung 2 ist  $\lim_{x_r \rightarrow \infty} K(x) = \lim_{x_r \rightarrow -\infty} K(x) = 0$ .  $\square$

Im weiteren Verlauf wird die folgende Taylor-Entwicklung des Kerns  $K$  nach den Parametern benötigt, die in allgemeinerer Form in Anhang C.3 hergeleitet wird.

**Bemerkung 27 (Taylor-Entwicklung des Kerns nach den Parametern)** Der Kern  $K$  genüge den Voraussetzungen 1 und 2. Wir betrachten die Differenz

$$K\left(\frac{x - Y_i}{h}\right) - K\left(\frac{x - X_i}{h}\right) = K_h(x, X_i, \hat{\vartheta}_n) - K_h(x, X_i, \vartheta_0) \quad (2.27)$$

des Kerns  $K$  mit standardisierten Daten  $Y_i = S_n^{-1/2}(X_i - \bar{X}_n)$  und ursprünglichen Daten  $X_i$  ( $i = 1, \dots, n$ ) als Funktion der Parameter  $\hat{\vartheta}_n := (\bar{X}_n, S_n)$  und  $\vartheta_0 := (0, I_d)$  unter Verwendung der Ergebnisse aus Anhang C.3. Dort werden mit den Bezeichnungen

$$SP_d = \{A \in \mathbb{R}^{d \times d} \mid A \text{ symmetrisch und positiv definit}\},$$

$\vartheta = (\mu, \Sigma) \in \mathbb{R}^d \times SP_d$  und

$$\text{st}(x, X_i, \vartheta) = \frac{1}{h} (x - \Sigma^{-1/2}(X_i - \mu))$$

Entwicklungen der Funktion

$$K_h : \begin{cases} \mathbb{R}^d \times SP_d & \rightarrow \mathbb{R} \\ \vartheta & \mapsto K_h(x, X_i, \vartheta) := K(\text{st}(x, X_i, \mu, \Sigma)) \end{cases}$$

in  $\vartheta$  um den Punkt  $\vartheta_0$  hergeleitet, die sich wegen Gleichung (2.27) auf das vorliegende Problem anwenden lassen. Für  $\vartheta = \hat{\vartheta}_n$  ist nach Gleichung (C.3) auf Seite 175 die Taylor–Entwicklung ersten Grades gegeben durch

$$K\left(\frac{x - Y_i}{h}\right) - K\left(\frac{x - X_i}{h}\right) = dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) + R_1(x, X_i) \quad (2.28)$$

und die zweiten Grades nach Gleichung (C.5) auf Seite 176 durch

$$\begin{aligned} K\left(\frac{x - Y_i}{h}\right) - K\left(\frac{x - X_i}{h}\right) & \quad (2.29) \\ & = dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) + \frac{1}{2} d^2 K_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) + R_2(x, X_i) \end{aligned}$$

mit den Restgliedern

$$R_1(x, X_i) = \int_0^1 (1-t) d^2 K_h(x, X_i, \vartheta_0 + t(\hat{\vartheta}_n - \vartheta_0)) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) dt, \quad (2.30)$$

$$R_2(x, X_i) = \int_0^1 \frac{1}{2} (1-t)^2 d^3 K_h(x, X_i, \vartheta_0 + t(\hat{\vartheta}_n - \vartheta_0)) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) dt. \quad (2.31)$$

Die Ableitungen in (2.28) und (2.29) sind nach den Gleichungen (C.7) und (C.8) auf Seite 176 gegeben durch

$$\begin{aligned} & dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) & (2.32) \\ & = \frac{1}{h} \nabla K\left(\frac{x - X_i}{h}\right) \bar{X}_n + \frac{1}{2h} \nabla K\left(\frac{x - X_i}{h}\right) (S_n - I_d) X_i \\ & = \frac{1}{h} \sum_{r=1}^d \nabla K\left(\frac{x - X_i}{h}\right)_r \bar{X}_{nr} + \frac{1}{2h} \sum_{r,s=1}^d \nabla K\left(\frac{x - X_i}{h}\right)_r (S_n - I_d)_{rs} X_{is} \end{aligned}$$

und

$$d^2 K_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) \quad (2.33)$$

$$\begin{aligned}
&= \frac{1}{h^2} \bar{X}_n^T \nabla^2 K \left( \frac{x - X_i}{h} \right) \bar{X}_n + \frac{1}{h^2} \bar{X}_n^T \nabla^2 K \left( \frac{x - X_i}{h} \right) (S_n - I_d) X_i \\
&\quad + \frac{1}{4h^2} X_i^T (S_n - I_d) \nabla^2 K \left( \frac{x - X_i}{h} \right) (S_n - I_d) X_i \\
&\quad - \frac{1}{h} \nabla K \left( \frac{x - X_i}{h} \right) (S_n - I_d) \bar{X}_n - \frac{3}{4h} \nabla K \left( \frac{x - X_i}{h} \right) (S_n - I_d)^2 X_i \\
&= \frac{1}{h^2} \sum_{r,s=1}^d \bar{X}_{nr} \nabla^2 K \left( \frac{x - X_i}{h} \right)_{rs} \bar{X}_{ns} + \frac{1}{h^2} \sum_{q,r,s=1}^d \bar{X}_{nq} \nabla^2 K \left( \frac{x - X_i}{h} \right)_{qr} (S_n - I_d)_{rs} X_{is} \\
&\quad + \frac{1}{4h^2} \sum_{l,q,r,s=1}^d X_{il} (S_n - I_d)_{lq} \nabla^2 K \left( \frac{x - X_i}{h} \right)_{qr} (S_n - I_d)_{rs} X_{is} \\
&\quad - \frac{1}{h} \sum_{r,s=1}^d \nabla K \left( \frac{x - X_i}{h} \right)_r (S_n - I_d)_{rs} \bar{X}_{ns} \\
&\quad - \frac{3}{4h} \sum_{q,r,s=1}^d \nabla K \left( \frac{x - X_i}{h} \right)_q (S_n - I_d)_{qr} (S_n - I_d)_{rs} X_{is} ,
\end{aligned}$$

wobei  $X_{ir}$ ,  $\bar{X}_{nr}$ ,  $\nabla K \left( \frac{x - X_i}{h} \right)_r$ ,  $(S_n - I_d)_{rs}$  und  $\nabla^2 K \left( \frac{x - X_i}{h} \right)_{rs}$  die Komponenten der Spalten- bzw. Zeilenvektoren und Matrizen bezeichnen.

**Lemma 28** Für die Restglieder  $R_1(x, X_i)$  und  $R_2(x, X_i)$  aus (2.30) bzw. (2.31) der Taylor-Entwicklungen aus Gleichung (2.28) bzw. (2.29) gelten mit den Abkürzungen  $\tilde{\vartheta}_n = (\tilde{\mu}_n, \tilde{\Sigma}_n)$ ,

$$\tilde{\vartheta}_n := \vartheta_0 + t(\hat{\vartheta}_n - \vartheta_0) , \quad \tilde{\mu}_n := t\bar{X}_n , \quad \tilde{\Sigma}_n := I_d + t(S_n - I_d)$$

die Abschätzungen

$$\begin{aligned}
&|R_1(x, X_i)| \\
&\leq \frac{1}{h^2} O_P \left( \frac{1}{n} \right) \left( \sum_{l=0}^2 \|X_i\|^l \right) \int_0^1 (1-t) \left\{ \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \right\} dt, \\
&|R_2(x, X_i)| \\
&\leq \frac{1}{h^3} O_P \left( \frac{1}{n^{3/2}} \right) \left( \sum_{l=0}^3 \|X_i\|^l \right) \\
&\quad \int_0^1 (1-t)^2 \left\{ \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \right\} dt.
\end{aligned}$$

**BEWEIS:** Zunächst werden für die in den Gleichungen (2.30) und (2.31) auftretenden Ableitungen in dem Zwischenpunkt  $\tilde{\vartheta}_n$  zu den Ungleichungen (C.10) und (C.11) auf Seite 178 analoge Abschätzungen hergeleitet. Dabei ist zu berücksichtigen, daß die dortige obere Schranke  $C$  für die Ableitungen  $\left\| \Upsilon(\tilde{\Sigma}_n) \right\|$ ,  $\left\| d\Upsilon(\tilde{\Sigma}_n) \right\|_{\mathcal{L}}$ ,  $\left\| d^2\Upsilon(\tilde{\Sigma}_n) \right\|_{\mathcal{L}}$  und  $\left\| d^3\Upsilon(\tilde{\Sigma}_n) \right\|_{\mathcal{L}}$  der Funktion  $\Upsilon : \mathfrak{J} \rightarrow SP_d$  nach (C.9) auf Seite 178 nur dann Gültigkeit besitzt, falls  $S_n$  in der kompakten

Umgebung  $\tilde{\mathfrak{J}} \subset \mathfrak{J}$  von  $I_d$  liegt. Dies gilt für fast alle  $\omega \in \Omega$  ab hinreichend großem  $n$ , genauer gesagt für  $n \geq n_0(\omega)$ . Die Ableitungen der Funktion  $\Upsilon$  sind also  $\omega$ -fast sicher beschränkt durch Konstanten, die weder von  $x$  noch von  $t$  abhängen, und wir können die Potenzen von  $C$  samt konstanter Vorfaktoren in den Ungleichungen (C.10) und (C.11) durch ein  $O_{f.s.}(1)$  ersetzen:

$$\begin{aligned} & \left| d^2 K_h(x, X_i, \tilde{\vartheta}_n) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) \right| \\ & \leq \frac{O_{f.s.}(1)}{h^2} \|\bar{X}_n\|^2 \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h^2} \|\bar{X}_n\| \|S_n - I_d\| (\|\bar{X}_n\| + \|X_i\|) \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h^2} \|S_n - I_d\|^2 (\|\bar{X}_n\| + \|X_i\|)^2 \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h} \|\bar{X}_n\| \|S_n - I_d\| \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h} \|S_n - I_d\|^2 (\|\bar{X}_n\| + \|X_i\|) \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \end{aligned}$$

und

$$\begin{aligned} & \left| d^3 K_h(x, X_i, \tilde{\vartheta}_n) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) \right| \\ & \leq \frac{O_{f.s.}(1)}{h^3} \|\bar{X}_n\|^3 \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h^3} \|\bar{X}_n\|^2 \|S_n - I_d\| (\|\bar{X}_n\| + \|X_i\|) \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h^3} \|\bar{X}_n\| \|S_n - I_d\|^2 (\|\bar{X}_n\| + \|X_i\|)^2 \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h^3} \|S_n - I_d\|^3 (\|\bar{X}_n\| + \|X_i\|)^3 \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h^2} \|\bar{X}_n\|^2 \|S_n - I_d\| \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h^2} \|\bar{X}_n\| \|S_n - I_d\|^2 (\|\bar{X}_n\| + \|X_i\|) \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h^2} \|S_n - I_d\|^3 (\|\bar{X}_n\| + \|X_i\|)^2 \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h} \|\bar{X}_n\| \|S_n - I_d\|^2 \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \\ & \quad + \frac{O_{f.s.}(1)}{h} \|S_n - I_d\|^3 (\|\bar{X}_n\| + \|X_i\|) \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\|. \end{aligned}$$

Berücksichtigt man nun die Ungleichungen  $h^{-1} \leq h^{-2} \leq h^{-3}$  und die Straffheit von  $\sqrt{n}\bar{X}_n$  sowie  $\sqrt{n}(S_n - I_d)$  aufgrund der Annahme, daß die zugrundeliegende Verteilung standardisiert ist, und schätzt man  $(\|\bar{X}_n\| + \|X_i\|)$ ,  $(\|\bar{X}_n\| + \|X_i\|)^2$  durch  $O_P(1) \sum_{l=0}^2 \|X_i\|^l$  und  $(\|\bar{X}_n\| + \|X_i\|)^3$  durch  $O_P(1) \sum_{l=0}^3 \|X_i\|^l$  ab, so ergeben sich die Abschätzungen

$$\left| d^2 K_h(x, X_i, \tilde{\vartheta}_n) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) \right|$$

$$\leq \frac{1}{h^2} O_P \left( \frac{1}{n} \right) \left( \sum_{l=0}^2 \|X_i\|^l \right) \left\{ \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \right\}$$

bzw.

$$\begin{aligned} & \left| d^3 K_h(x, X_i, \tilde{\vartheta}_n) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) \right| \\ & \leq \frac{1}{h^3} O_P \left( \frac{1}{n^{3/2}} \right) \left( \sum_{l=0}^3 \|X_i\|^l \right) \\ & \quad \cdot \left\{ \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \right\} . \end{aligned}$$

Einsetzen in (2.30) bzw. (2.31) liefert die Behauptung.  $\square$

## 2.3. $H_0$ -Konvergenz der Bowman-Foster-Statistik

Ziel dieses Abschnitts ist es, die Konvergenz der Bowman-Foster-Statistik

$$\tilde{T}_{n,h,\text{st}} = nh^d \int_{\mathbb{R}^d} \left( \hat{f}_{n,\text{st}}(x) - e_h(x, f) \right)^2 dx$$

mit geschätzten Parametern herzuleiten. Zu diesem Zweck stellt man zunächst fest, daß sich  $\tilde{T}_{n,h,\text{st}}$  gemäß

$$\tilde{T}_{n,h,\text{st}} = \tilde{T}_{n,h} + \tilde{T}_{n,1} + \tilde{T}_{n,2} \quad (2.34)$$

zerlegen läßt in die Bowman-Foster-Statistik

$$\tilde{T}_{n,h} = nh^d \int_{\mathbb{R}^d} \left( \hat{f}_n(x) - e_h(x, f) \right)^2 dx$$

ohne geschätzte Parameter und in zwei Restterme  $\tilde{T}_{n,1}$  und  $\tilde{T}_{n,2}$ , die definiert sind durch

$$\tilde{T}_{n,1} := nh^d \int \left( \hat{f}_{n,\text{st}}(x) - \hat{f}_n(x) \right)^2 dx , \quad (2.35)$$

$$\tilde{T}_{n,2} := 2nh^d \int \left( \hat{f}_{n,\text{st}}(x) - \hat{f}_n(x) \right) \left( \hat{f}_n(x) - e_h(x, f) \right) dx . \quad (2.36)$$

Die Asymptotik der Bowman-Foster-Statistik  $\tilde{T}_{n,h}$  ohne geschätzte Parameter wurde in Hall (1984) unter der Voraussetzung 1 mit grundlegend neuen Methoden gezeigt und ist in Theorem 29 festgehalten.

**Theorem 29 (Hall (1984))** Wenn  $f$  die den Daten zugrundeliegende Dichte ist, gilt unter der Voraussetzung 1

$$\frac{\tilde{T}_{n,h} - \mu(f)}{\sqrt{2} h^{d/2} \sigma(f)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) .$$

**Korollar 30** Unter der Hypothese einer  $d$ -dimensionalen Standard-Normalverteilung gilt nach (2.23) unter der Voraussetzung 1 an den Kern  $K$  und die Bandbreite  $h$

$$\frac{2^d \pi^{d/2} \tilde{T}_{n,h} - 1}{2^{1/2-d/4} h^{d/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

BEWEIS VON THEOREM 29: Die Struktur des Beweises von P. Hall läßt sich wie folgt skizzieren. Zunächst zerlegt man  $\tilde{T}_{n,h} = L_n + U_n$  in einen linearen Term

$$L_n := \frac{1}{nh^d} \sum_{i=1}^n \int \left\{ K\left(\frac{x-X_i}{h}\right) - \mathbb{E}\left[K\left(\frac{x-X_1}{h}\right)\right] \right\}^2 dx$$

und einen quadratischen Term

$$U_n := \frac{2}{nh^d} \sum_{i < j} \int \left\{ K\left(\frac{x-X_i}{h}\right) - \mathbb{E}\left[K\left(\frac{x-X_1}{h}\right)\right] \right\} \left\{ K\left(\frac{x-X_j}{h}\right) - \mathbb{E}\left[K\left(\frac{x-X_1}{h}\right)\right] \right\} dx,$$

deren Asymptotik getrennt behandelt wird.

1) BEHANDLUNG VON  $L_n$ :

Es ist

$$\mathbb{E}L_n = \mu(f) - h^d \iint K(u)K(u+v) du dv \int f(x)f(x+hv)dx, \quad \text{Var}(L_n) = O\left(\frac{1}{n}\right)$$

und folglich

$$L_n = \mu(f) - h^d \iint K(u)K(u+v) du dv \int f(x)f(x+hv)dx + O_P\left(\frac{1}{n^{1/2}}\right).$$

2) BEHANDLUNG VON  $U_n$ :

$U_n = \frac{2}{nh^d} \sum_{i < j} H_n(X_i, X_j)$  ist eine  $U$ -Statistik mit dem symmetrischen, zentrierten und degenerierten Kern

$$H_n(X_i, X_j) := \int \left\{ K\left(\frac{x-X_i}{h}\right) - \mathbb{E}K\left(\frac{x-X_1}{h}\right) \right\} \left\{ K\left(\frac{x-X_j}{h}\right) - \mathbb{E}K\left(\frac{x-X_1}{h}\right) \right\} dx,$$

der von der Bandbreite  $h = h(n)$  abhängt und daher mit dem Stichprobenumfang  $n$  variiert. Hall (1984) leitete zur Behandlung von  $U_n$  einen Zentralen Grenzwertsatz für  $U$ -Statistiken mit variablem Kern her, indem er bislang in diesem Kontext noch nicht verwendete Martingalmethoden einsetzte. Das resultierende Theorem ist als Theorem D.4 auf Seite 182 in Anhang D.2 wiedergegeben. Mit

$$G_n(x, y) := \mathbb{E}[H_n(X_1, x)H_n(X_1, y)]$$

läßt sich

$$\mathbb{E}[H_n^2(X_1, X_2)] = h^{3d} \sigma^2(f) + O(h^{4d})$$

zeigen sowie

$$\mathbb{E} [H_n^4(X_1, X_2)] = O(h^{5d}) , \quad \mathbb{E} [G_n^2(X_1, X_2)] = O(h^{7d}) ,$$

was es erlaubt, die Voraussetzung von Theorem D.4 nachzuprüfen:

$$\frac{\mathbb{E} [G_n^2(X_1, X_2)] + \frac{1}{n} \mathbb{E} [H_n^4(X_1, X_2)]}{(\mathbb{E} [H_n^2(X_1, X_2)])^2} = \frac{O(h^d) + O\left(\frac{1}{nh^d}\right)}{[\sigma^2(f) + O(h^d)]^2} \rightarrow 0 .$$

Damit folgt die Asymptotik von  $U_n$  aus Theorem D.4:

$$\frac{U_n}{\sqrt{2} h^{d/2} \sigma(f)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) .$$

3) ASYMPTOTIK VON  $\tilde{T}_{n,h}$ :

Mit einer asymptotisch standard-normalverteilten Zufallsvariablen  $Z_n$  gilt

$$\begin{aligned} \tilde{T}_{n,h} &= L_n + U_n \\ &= \mu(f) - h^d \iint K(u)K(u+v) du dv \int f(x)f(x+hv) dx + O_P\left(\frac{1}{n^{1/2}}\right) + \sqrt{2} h^{d/2} \sigma(f) Z_n, \end{aligned}$$

also

$$\frac{\tilde{T}_{n,h} - \mu(f)}{\sqrt{2} h^{d/2} \sigma(f)} = Z_n + o(1) + O_P\left(\frac{1}{\sqrt{nh^d}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) .$$

□

Die Behauptung ist nun, daß die Schätzer im  $H_0$ -Fall bei einer zugrundeliegenden standardisierten Dichte und unter der zusätzlichen Voraussetzung 2 keinen Einfluß auf die Grenzverteilung haben, d.h. daß die Asymptotik der Bowman-Foster-Statistik  $\tilde{T}_{n,h,st}$  mit geschätzten Parametern dieselbe ist wie die Asymptotik der Bowman-Foster-Statistik  $\tilde{T}_{n,h}$  ohne geschätzte Parameter.

**Theorem 31** Wenn die standardisierte Dichte  $f$  den Daten zugrundeliegt, gilt unter den Voraussetzungen 1 und 2

$$\frac{\tilde{T}_{n,h,st} - \mu(f)}{\sqrt{2} h^{d/2} \sigma(f)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) . \quad (2.37)$$

**Korollar 32** Unter der Hypothese einer  $d$ -dimensionalen Normalverteilung gilt unter den Voraussetzungen 1 und 2 an den Kern  $K$  und die Bandbreite  $h$

$$\frac{2^d \pi^{d/2} \tilde{T}_{n,h,st} - 1}{2^{1/2-d/4} h^{d/2}} = \frac{T_{n,h,st} - 1}{2^{1/2-d/4} h^{d/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) .$$

BEWEIS VON THEOREM 31: Mit der Zerlegung

$$\frac{\tilde{T}_{n,h,\text{st}} - \mu(f)}{\sqrt{2} h^{d/2} \sigma(f)} = \frac{\tilde{T}_{n,h} - \mu(f)}{\sqrt{2} h^{d/2} \sigma(f)} + \frac{\tilde{T}_{n,1}}{\sqrt{2} h^{d/2} \sigma(f)} + \frac{\tilde{T}_{n,2}}{\sqrt{2} h^{d/2} \sigma(f)}$$

gemäß Gleichung (2.34) ist angesichts von Theorem 29 zu zeigen, daß die Restterme  $\tilde{T}_{n,1}$  und  $\tilde{T}_{n,2}$  hinreichend schnell stochastisch gegen 0 streben:

$$\frac{\tilde{T}_{n,1}}{h^{d/2}} \xrightarrow{P} 0, \quad \frac{\tilde{T}_{n,2}}{h^{d/2}} \xrightarrow{P} 0.$$

1) BETRACHTUNG VON  $\tilde{T}_{n,1}/h^{d/2}$ :

Setzt man in die Definition von  $\tilde{T}_{n,1}$  die Taylor–Entwicklung ersten Grades aus Gleichung (2.28) ein, so gelangt man zu der Zerlegung

$$\begin{aligned} \frac{\tilde{T}_{n,1}}{h^{d/2}} &= nh^{d/2} \int \left( \hat{f}_{n,\text{st}}(x) - \hat{f}_n(x) \right)^2 dx \\ &= h^{d/2} \cdot \frac{1}{nh^{2d}} \sum_{i,j=1}^n \int \left[ K\left(\frac{x-Y_i}{h}\right) - K\left(\frac{x-X_i}{h}\right) \right] \left[ K\left(\frac{x-Y_j}{h}\right) - K\left(\frac{x-X_j}{h}\right) \right] dx \\ &= \tilde{T}_{n,3} + 2\tilde{T}_{n,4} + \tilde{T}_{n,5} \end{aligned}$$

mit

$$\begin{aligned} \tilde{T}_{n,3} &:= \frac{h^{d/2}}{nh^{2d}} \sum_{i,j=1}^n \int dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \cdot dK_h(x, X_j, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) dx, \\ \tilde{T}_{n,4} &:= \frac{h^{d/2}}{nh^{2d}} \sum_{i,j=1}^n \int dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \cdot R_1(x, X_j) dx, \\ \tilde{T}_{n,5} &:= \frac{h^{d/2}}{nh^{2d}} \sum_{i,j=1}^n \int R_1(x, X_i) R_1(x, X_j) dx, \end{aligned}$$

wobei  $R_1(x, X_i)$  das Restglied aus Gleichung (2.30) ist. Folglich ist zu zeigen, daß die Terme  $\tilde{T}_{n,3}$ ,  $\tilde{T}_{n,4}$  und  $\tilde{T}_{n,5}$  stochastisch gegen 0 streben. Da nach der Cauchy–Schwarz–Ungleichung die Abschätzung

$$\begin{aligned} & \left| \tilde{T}_{n,4} \right| \\ & \leq \int \left| \frac{h^{d/4}}{\sqrt{nh^d}} \sum_{i=1}^n dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \right| \left| \frac{h^{d/4}}{\sqrt{nh^d}} \sum_{j=1}^n R_1(x, X_j) \right| dx \\ & \leq \left\{ \int \left[ \frac{h^{d/4}}{\sqrt{nh^d}} \sum_{i=1}^n dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \right]^2 dx \cdot \int \left[ \frac{h^{d/4}}{\sqrt{nh^d}} \sum_{j=1}^n R_1(x, X_j) \right]^2 dx \right\}^{1/2} \end{aligned}$$

$$= \left\{ \tilde{T}_{n,3} \cdot \tilde{T}_{n,5} \right\}^{1/2}$$

für den gemischten Term  $\tilde{T}_{n,4}$  gilt, reduziert sich das Problem auf den Nachweis der Gültigkeit von  $\tilde{T}_{n,3} = o_P(1)$  und  $\tilde{T}_{n,5} = o_P(1)$ .

(i) BETRACHTUNG VON  $\tilde{T}_{n,3}$ :

Setzt man in die Definition von  $\tilde{T}_{n,3}$  die Ableitung aus Gleichung (2.32) ein, so ergibt sich

$$\begin{aligned} \tilde{T}_{n,3} &= \frac{h^{d/2}}{nh^{2d}} \sum_{i,j=1}^n \int dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \cdot dK_h(x, X_j, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) dx \\ &= \sum_{r,s=1}^d (\sqrt{n}\bar{X}_{nr}) (\sqrt{n}\bar{X}_{ns}) \frac{h^{d/2}}{n^2 h^{2d+2}} \sum_{i,j=1}^n \int \nabla K \left( \frac{x - X_i}{h} \right)_r \nabla K \left( \frac{x - X_j}{h} \right)_s dx \\ &\quad + \sum_{q,r,s=1}^d (\sqrt{n}\bar{X}_{nq}) (\sqrt{n}(S_n - I_d)_{rs}) \frac{h^{d/2}}{n^2 h^{2d+2}} \sum_{i,j=1}^n X_{js} \int \nabla K \left( \frac{x - X_i}{h} \right)_q \nabla K \left( \frac{x - X_j}{h} \right)_r dx \\ &\quad + \sum_{l,q,r,s=1}^d \frac{1}{4} (\sqrt{n}(S_n - I_d)_{lq}) (\sqrt{n}(S_n - I_d)_{rs}) \\ &\quad \quad \frac{h^{d/2}}{n^2 h^{2d+2}} \sum_{i,j=1}^n X_{iq} X_{js} \int \nabla K \left( \frac{x - X_i}{h} \right)_l \nabla K \left( \frac{x - X_j}{h} \right)_r dx \\ &= o_P(1) \end{aligned}$$

aufgrund der Straffheit von  $\sqrt{n}\bar{X}_n$  und  $\sqrt{n}(S_n - I_d)$  sowie nach Lemma 53 auf Seite 125.

(ii) BETRACHTUNG VON  $\tilde{T}_{n,5}$ :

Nach Lemma 28 auf Seite 89 gilt mit den Abkürzungen

$$\tilde{\vartheta}_n := \vartheta_0 + t(\hat{\vartheta}_n - \vartheta_0), \quad \tilde{\tilde{\vartheta}}_n := \vartheta_0 + s(\hat{\vartheta}_n - \vartheta_0)$$

die Abschätzung

$$\begin{aligned} \left| \tilde{T}_{n,5} \right| &\leq \frac{h^{d/2}}{nh^{2d}} \sum_{i,j=1}^n \int |R_1(x, X_i)| \cdot |R_1(x, X_j)| dx & (2.38) \\ &\leq O_P(1) \frac{1}{nh^{d+4}} \cdot \frac{h^{d/2}}{n^2 h^d} \sum_{i,j=1}^n \left( \sum_{l=0}^2 \|X_i\|^l \right) \left( \sum_{q=0}^2 \|X_j\|^q \right) \\ &\quad \int \left[ \int_0^1 (1-t) \left\{ \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \right\} dt \right] \\ &\quad \left[ \int_0^1 (1-s) \left\{ \left\| \nabla^2 K(\text{st}(x, X_j, \tilde{\tilde{\vartheta}}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_j, \tilde{\tilde{\vartheta}}_n)) \right\| \right\} ds \right] dx \end{aligned}$$

$$\begin{aligned}
&= O_P(1) \frac{1}{nh^{d+4}} \cdot \frac{h^{d/2}}{n^2 h^d} \sum_{i,j=1}^n \left( \sum_{l=0}^2 \|X_i\|^l \right) \left( \sum_{q=0}^2 \|X_j\|^q \right) \left\{ \right. \\
&\quad \int_0^1 \int_0^1 (1-t)(1-s) \int \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \left\| \nabla^2 K(\text{st}(x, X_j, \tilde{\vartheta}_n)) \right\| dx ds dt \\
&\quad + 2 \int_0^1 \int_0^1 (1-t)(1-s) \int \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \left\| \nabla K(\text{st}(x, X_j, \tilde{\vartheta}_n)) \right\| dx ds dt \\
&\quad \left. + \int_0^1 \int_0^1 (1-t)(1-s) \int \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \left\| \nabla K(\text{st}(x, X_j, \tilde{\vartheta}_n)) \right\| dx ds dt \right\}.
\end{aligned}$$

Da die Bandbreite  $h$  optimal, d.h. gemäß Gleichung (2.9) auf Seite 77 gewählt wurde, ist  $nh^{d+4}$  konstant und braucht im folgenden nicht mehr berücksichtigt zu werden, ebenso wie der straffe Term  $O_P(1)$  am Anfang der fünften Zeile von (2.38). Ferner verläuft die Behandlung der drei Summanden in den Zeilen 6,7 und 8 von (2.38) analog, so daß wir uns auf einen Vertreter beschränken können. Mit

$$\begin{aligned}
\tilde{\mu}_n &:= t\bar{X}_n, & \tilde{\Sigma}_n &:= I_d + t(S_n - I_d), \\
\tilde{\mu}_n &:= s\bar{X}_n, & \tilde{\Sigma}_n &:= I_d + s(S_n - I_d)
\end{aligned}$$

ist also z.B. die stochastische Konvergenz gegen 0 von

$$\begin{aligned}
&\frac{h^{d/2}}{n^2 h^d} \sum_{i,j=1}^n \left( \sum_{l=0}^2 \|X_i\|^l \right) \left( \sum_{q=0}^2 \|X_j\|^q \right) \\
&\quad \int_0^1 \int_0^1 (1-t)(1-s) \int \left\| \nabla K \left( \frac{x - \tilde{\Sigma}_n^{-1/2}(X_i - \tilde{\mu}_n)}{h} \right) \right\| \left\| \nabla K \left( \frac{x - \tilde{\Sigma}_n^{-1/2}(X_j - \tilde{\mu}_n)}{h} \right) \right\| dx ds dt
\end{aligned}$$

zu zeigen. Mit der Cauchy–Schwarz–Ungleichung und der Substitution  $y = (x - \tilde{\Sigma}_n^{-1/2}(X_i - \tilde{\mu}_n))/h$ ,  $dy = h^{-d} dx$  ergibt sich nach Voraussetzung 2 und dem starken Gesetz großer Zahlen für diesen Term die obere Schranke

$$\begin{aligned}
&h^{d/2} \left\{ \frac{1}{nh^{d/2}} \sum_{i=1}^n \left( \sum_{l=0}^2 \|X_i\|^l \right) \int_0^1 (1-t) \left[ \int \left\| \nabla K \left( \frac{x - \tilde{\Sigma}_n^{-1/2}(X_i - \tilde{\mu}_n)}{h} \right) \right\|^2 dx \right]^{1/2} dt \right\}^2 \\
&= h^{d/2} \left\{ \frac{1}{n} \sum_{i=1}^n \left( \sum_{l=0}^2 \|X_i\|^l \right) \int_0^1 (1-t) \left[ \int \|\nabla K(y)\|^2 dy \right]^{1/2} dt \right\}^2 \\
&= h^{d/2} \left\{ \frac{1}{2} \left[ \int \|\nabla K(y)\|^2 dy \right]^{1/2} \frac{1}{n} \sum_{i=1}^n [1 + \|X_i\| + \|X_i\|^2] \right\}^2 \\
&= o_P(1).
\end{aligned}$$

2) BETRACHTUNG VON  $\tilde{T}_{n,2}/h^{d/2}$ :

Setzt man in die Definition von  $\tilde{T}_{n,2}$  die Taylor-Entwicklung zweiten Grades aus Gleichung (2.29) ein, so gelangt man zu der Zerlegung

$$\begin{aligned}\frac{\tilde{T}_{n,2}}{h^{d/2}} &= 2nh^{d/2} \int \left( \hat{f}_{n,\text{st}}(x) - \hat{f}_n(x) \right) \left( \hat{f}_n(x) - e_h(x, f) \right) dx \\ &= 2 \frac{1}{h^{d/2}} \sum_{i=1}^n \int \left[ K \left( \frac{x-Y_i}{h} \right) - K \left( \frac{x-X_i}{h} \right) \right] \left[ \hat{f}_n(x) - e_h(x, f) \right] dx \\ &= 2 \tilde{T}_{n,6} + \tilde{T}_{n,7} + 2 \tilde{T}_{n,8}\end{aligned}$$

mit

$$\begin{aligned}\tilde{T}_{n,6} &:= \frac{1}{h^{d/2}} \sum_{i=1}^n \int dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \left[ \hat{f}_n(x) - e_h(x, f) \right] dx, \\ \tilde{T}_{n,7} &:= \frac{1}{h^{d/2}} \sum_{i=1}^n \int d^2 K_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) \left[ \hat{f}_n(x) - e_h(x, f) \right] dx, \\ \tilde{T}_{n,8} &:= \frac{1}{h^{d/2}} \sum_{i=1}^n \int R_2(x, X_i) \left[ \hat{f}_n(x) - e_h(x, f) \right] dx,\end{aligned}$$

und es ist zu zeigen, daß die Terme  $\tilde{T}_{n,6}$ ,  $\tilde{T}_{n,7}$  und  $\tilde{T}_{n,8}$  stochastisch gegen 0 streben.

(i) BETRACHTUNG VON  $\tilde{T}_{n,6}$ :

Setzt man in die Definition von  $\tilde{T}_{n,6}$  die Ableitung aus Gleichung (2.32) ein, so ergibt sich

$$\begin{aligned}\tilde{T}_{n,6} &= \frac{1}{nh^{(3d)/2}} \sum_{i,j=1}^n \int dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E}K \left( \frac{x-X_1}{h} \right) \right] dx \\ &= \sum_{r=1}^d (\sqrt{n} \bar{X}_{nr}) \frac{h^{1/2}}{n^{3/2} h^{(3d+3)/2}} \sum_{i,j=1}^n \int \nabla K \left( \frac{x-X_i}{h} \right)_r \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E}K \left( \frac{x-X_1}{h} \right) \right] dx \\ &\quad + \sum_{r,s=1}^d \frac{1}{2} (\sqrt{n} (S_n - I_d)_{rs}) \\ &\quad \frac{h^{1/2}}{n^{3/2} h^{(3d+3)/2}} \sum_{i,j=1}^n X_{is} \int \nabla K \left( \frac{x-X_i}{h} \right)_r \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E}K \left( \frac{x-X_1}{h} \right) \right] dx \\ &= o_P(1)\end{aligned}$$

aufgrund der Straffheit von  $\sqrt{n} \bar{X}_n$  und  $\sqrt{n} (S_n - I_d)$  sowie nach Lemma 54 auf Seite 131. Dabei wird der Erwartungswert  $\mathbb{E}K((x - X_1)/h)$  bezüglich der zugrundeliegenden Dichte  $f$  gebildet.

(ii) BETRACHTUNG VON  $\tilde{T}_{n,7}$ :

Setzt man in die Definition von  $\tilde{T}_{n,7}$  die Ableitung aus Gleichung (2.33) ein, so ergibt sich

$$\begin{aligned}
\tilde{T}_{n,7} &= \frac{1}{nh^{(3d)/2}} \sum_{i,j=1}^n \int d^2 K_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) \left[ K\left(\frac{x-X_j}{h}\right) - \mathbb{E}K\left(\frac{x-X_1}{h}\right) \right] dx \\
&= \sum_{r,s=1}^d (\sqrt{n}\bar{X}_{nr}) (\sqrt{n}\bar{X}_{ns}) \\
&\quad \frac{h^{1/2}}{n^2 h^{(3d+5)/2}} \sum_{i,j=1}^n \int \nabla^2 K\left(\frac{x-X_i}{h}\right)_{rs} \left[ K\left(\frac{x-X_j}{h}\right) - \mathbb{E}K\left(\frac{x-X_1}{h}\right) \right] dx \\
&+ \sum_{q,r,s=1}^d (\sqrt{n}\bar{X}_{nq}) (\sqrt{n}(S_n - I_d)_{rs}) \\
&\quad \frac{h^{1/2}}{n^2 h^{(3d+5)/2}} \sum_{i,j=1}^n X_{is} \int \nabla^2 K\left(\frac{x-X_i}{h}\right)_{qr} \left[ K\left(\frac{x-X_j}{h}\right) - \mathbb{E}K\left(\frac{x-X_1}{h}\right) \right] dx \\
&+ \sum_{l,q,r,s=1}^d \frac{1}{4} (\sqrt{n}(S_n - I_d)_{lq}) (\sqrt{n}(S_n - I_d)_{rs}) \\
&\quad \frac{h^{1/2}}{n^2 h^{(3d+5)/2}} \sum_{i,j=1}^n X_{il} X_{is} \int \nabla^2 K\left(\frac{x-X_i}{h}\right)_{qr} \left[ K\left(\frac{x-X_j}{h}\right) - \mathbb{E}K\left(\frac{x-X_1}{h}\right) \right] dx \\
&- \sum_{r,s=1}^d (\sqrt{n}\bar{X}_{ns}) (\sqrt{n}(S_n - I_d)_{rs}) \\
&\quad \frac{h^{1/2}}{n^2 h^{(3d+3)/2}} \sum_{i,j=1}^n \int \nabla K\left(\frac{x-X_i}{h}\right)_r \left[ K\left(\frac{x-X_j}{h}\right) - \mathbb{E}K\left(\frac{x-X_1}{h}\right) \right] dx \\
&- \sum_{q,r,s=1}^d \frac{3}{4} (\sqrt{n}(S_n - I_d)_{qr}) (\sqrt{n}(S_n - I_d)_{rs}) \\
&\quad \frac{h^{1/2}}{n^2 h^{(3d+3)/2}} \sum_{i,j=1}^n X_{is} \int \nabla K\left(\frac{x-X_i}{h}\right)_q \left[ K\left(\frac{x-X_j}{h}\right) - \mathbb{E}K\left(\frac{x-X_1}{h}\right) \right] dx \\
&= o_P(1)
\end{aligned}$$

aufgrund der Straffheit von  $\sqrt{n}\bar{X}_n$  und  $\sqrt{n}(S_n - I_d)$  sowie nach Lemma 54 auf Seite 131. Für die Terme mit der Gewichtung  $1/(n^2 h^{(3d+5)/2})$  beachte man hierbei, daß aufgrund der Wahl von  $h$  gemäß (2.9) auf Seite 77 gilt:

$$\frac{1}{n^2 h^{(3d+5)/2}} = \frac{1}{n^{3/2} h^{(3d+3)/2}} \cdot \frac{1}{n^{1/2} h} = o\left(\frac{1}{n^{3/2} h^{(3d+3)/2}}\right).$$

(iii) BETRACHTUNG VON  $\tilde{T}_{n,8}$ :

Nach der Cauchy-Schwarz-Ungleichung gilt die Abschätzung

$$\begin{aligned} \left| \tilde{T}_{n,8} \right| &\leq \int \left| \frac{1}{h^{d/2}} \sum_{i=1}^n R_2(x, X_i) \right| \cdot \left| \hat{f}_n(x) - e_h(x, f) \right| dx \\ &\leq \left\{ \frac{1}{nh^d} \int \left[ \frac{1}{h^{d/2}} \sum_{i=1}^n R_2(x, X_i) \right]^2 dx \cdot nh^d \int \left[ \hat{f}_n(x) - e_h(x, f) \right]^2 dx \right\}^{1/2} \\ &= \left\{ \frac{1}{nh^{2d}} \sum_{i,j=1}^n \int R_2(x, X_i) R_2(x, X_j) dx \cdot \tilde{T}_{n,h} \right\}^{1/2}, \end{aligned}$$

und aufgrund der Straffheit von  $\tilde{T}_{n,h}$  nach Theorem 29 auf Seite 91 reicht es, den Term

$$\frac{1}{nh^{2d}} \sum_{i,j=1}^n \int |R_2(x, X_i)| |R_2(x, X_j)| dx \leq \quad (2.39)$$

$$\begin{aligned} &O_P(1) \frac{1}{n^2 h^{d+7}} \cdot \frac{h}{n^2 h^d} \sum_{i,j=1}^n \left( \sum_{l=0}^3 \|X_i\|^l \right) \left( \sum_{q=0}^3 \|X_j\|^q \right) \\ &\int \left[ \int_0^1 (1-t)^2 \left\{ \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \right\} dt \right] \\ &\left[ \int_0^1 (1-s)^2 \left\{ \left\| \nabla^3 K(\text{st}(x, X_j, \tilde{\vartheta}_n)) \right\| + \left\| \nabla^2 K(\text{st}(x, X_j, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_j, \tilde{\vartheta}_n)) \right\| \right\} ds \right] dx \end{aligned}$$

bzw. dessen obere Schranke zu betrachten. Dabei wurden wie bei der Behandlung von  $\tilde{T}_{n,5}$  Lemma 28 auf Seite 89 angewandt und

$$\tilde{\vartheta}_n := \vartheta_0 + t(\hat{\vartheta}_n - \vartheta_0), \quad \tilde{\tilde{\vartheta}}_n := \vartheta_0 + s(\hat{\vartheta}_n - \vartheta_0)$$

gesetzt. Da die Bandbreite  $h$  optimal, d.h. gemäß Gleichung (2.9) auf Seite 77 gewählt wurde, konvergiert der Faktor  $1/(n^2 h^{d+7})$  gegen 0 und wird im folgenden nicht mehr berücksichtigt, ebenso wie der straffe Term  $O_P(1)$  am Anfang der zweiten Zeile von (2.39). Das Integral in (2.39) besteht nach Ausmultiplizieren und Zusammenfassen aus sechs verschiedenen Summanden, deren Behandlung analog verläuft, so daß wir uns auf einen Vertreter beschränken. Mit

$$\begin{aligned} \tilde{\mu}_n &:= t\bar{X}_n, & \tilde{\Sigma}_n &:= I_d + t(S_n - I_d), \\ \tilde{\tilde{\mu}}_n &:= s\bar{X}_n, & \tilde{\tilde{\Sigma}}_n &:= I_d + s(S_n - I_d) \end{aligned}$$

gilt z.B. mit der Cauchy-Schwarz-Ungleichung und der Substitution  $y = (x - \tilde{\Sigma}_n^{-1/2}(X_i - \tilde{\mu}_n))/h$ ,  $dy = h^{-d} dx$  sowie nach Voraussetzung 2 in Verbindung mit dem starken Gesetz großer Zahlen

$$\frac{h}{n^2 h^d} \sum_{i,j=1}^n \left( \sum_{l=0}^3 \|X_i\|^l \right) \left( \sum_{q=0}^3 \|X_j\|^q \right)$$

$$\begin{aligned}
& \int_0^1 \int_0^1 (1-t)^2 (1-s)^2 \int \left\| \nabla K \left( \frac{x - \tilde{\Sigma}_n^{-1/2} (X_i - \tilde{\mu}_n)}{h} \right) \right\| \left\| \nabla K \left( \frac{x - \tilde{\Sigma}_n^{-1/2} (X_j - \tilde{\mu}_n)}{h} \right) \right\| dx ds dt \\
& \leq h \left\{ \frac{1}{nh^{d/2}} \sum_{i=1}^n \left( \sum_{l=0}^3 \|X_i\|^l \right) \int_0^1 (1-t)^2 \left[ \int \left\| \nabla K \left( \frac{x - \tilde{\Sigma}_n^{-1/2} (X_i - \tilde{\mu}_n)}{h} \right) \right\|^2 dx \right]^{1/2} dt \right\}^2 \\
& = h \left\{ \frac{1}{n} \sum_{i=1}^n \left( \sum_{l=0}^3 \|X_i\|^l \right) \int_0^1 (1-t)^2 \left[ \int \|\nabla K(y)\|^2 dy \right]^{1/2} dt \right\}^2 \\
& = h \left\{ \frac{1}{3} \left[ \int \|\nabla K(y)\|^2 dy \right]^{1/2} \frac{1}{n} \sum_{i=1}^n [1 + \|X_i\| + \|X_i\|^2 + \|X_i\|^3] \right\}^2 \\
& = o_P(1).
\end{aligned}$$

□

**Bemerkung 33 (Asymptotischer Niveau- $\alpha$  Test zur Bowman–Foster–Statistik)** Anhand von Theorem 31 läßt sich unter den Voraussetzungen 1 und 2 an die Dichte  $f$ , den Kern  $K$  und die Bandbreite  $h$  ein auf der Bowman–Foster–Statistik basierender asymptotischer Niveau- $\alpha$ -Test für die zusammengesetzte Hypothese

$$H_0 : g \in \mathcal{F} \quad \text{gegen} \quad H_1 : g \notin \mathcal{F}$$

mit der von  $f$  erzeugten Lokations–Skalen–Familie

$$\mathcal{F} = \left\{ f(x, \vartheta) = \frac{1}{|\det \Sigma|^{1/2}} f(\Sigma^{-1/2}(x - \mu)) \mid \vartheta = (\mu, \Sigma), \mu \in \mathbb{R}^d, \Sigma \text{ positiv definit} \right\}$$

konstruieren. Dabei beschränken wir uns auf die Fälle, in denen  $\tilde{T}_{n,h,st}$  affin invariant ist, also z.B. die Familie  $\mathcal{F}$  der  $d$ -dimensionalen Normalverteilungen und den Gaußkern aus Gleichung (2.2), oder aber die Dimension  $d = 1$ . Aufgrund der Problemstellung besitzt der Test einen oberen Ablehnbereich. Bezeichnet  $\Phi^{-1}(1 - \alpha)$ ,  $\alpha \in (0, 1)$ , das  $(1 - \alpha)$ -Quantil der Standard–Normalverteilung, so wird nach (2.37) die Hypothese genau dann abgelehnt, wenn

$$\frac{\tilde{T}_{n,h,st} - \mu(f)}{\sqrt{2} h^{d/2} \sigma(f)} > \Phi^{-1}(1 - \alpha)$$

gilt, also wenn der Wert der Teststatistik auf gegebenen Daten den *kritischen Wert*

$$q_{h,1-\alpha} := \mu(f) + \sqrt{2} h^{d/2} \sigma(f) \Phi^{-1}(1 - \alpha) \quad (2.40)$$

des Tests überschreitet. Für den auf der Bowman–Foster–Statistik basierenden asymptotischen Test auf  $d$ -dimensionale Normalverteilung ergibt sich nach Gleichung (2.23) auf Seite 84

$$q_{h,1-\alpha} = 2^{-d} \pi^{-d/2} [1 + h^{d/2} 2^{1/2-d/4} \Phi^{-1}(1 - \alpha)].$$

Wie im BHEP–Fall kann man das asymptotische Quantil  $q_{h,1-\alpha}$  der Bowman–Foster–Statistik aus (2.40) durch ein bei festem Stichprobenumfang  $n$  anhand von Monte–Carlo–Simulationen ermitteltes empirisches Quantil  $q_{h,n,1-\alpha}$  ersetzen.

## 2.4. $H_1$ -Konvergenz der Bowman-Foster-Statistik

Bisher wurde in Kapitel 2.3 die Konvergenz der Bowman-Foster-Statistik unter einer standardisierten Dichte, die den Voraussetzungen 1 und 2 genügt, gezeigt. Damit konnte die Wohldefiniertheit des in Bemerkung 33 konstruierten, auf  $\tilde{T}_{n,h,st}$  basierenden Tests für die zusammengesetzte Hypothese nachgewiesen werden, wobei wir uns auf die Fälle, in denen  $\tilde{T}_{n,h,st}$  affin invariant ist, beschränkt haben. Nun soll analog zum BHEP-Fall das Verhalten der Bowman-Foster-Statistik unter einer festen Alternative untersucht werden, mit dem Ziel einer Gütenäherung für den Test aus Bemerkung 33, die mit der aus dem BHEP-Fall verglichen werden kann.

Der Deutlichkeit halber bezeichnen wir im folgenden die der Lokations-Skalen-Familie  $\mathcal{F}$  zugrundeliegende standardisierte  $H_0$ -Dichte mit  $f_0$  und die einer festen Alternative zugrundeliegende Dichte mit  $f_1$ . Im weiteren Verlauf gehen wir stets davon aus, daß  $f_0$ , die Bandbreite  $h$  und der Kern  $K$  die Voraussetzungen 1 und 2 erfüllen. Damit die Ergebnisse des  $H_0$ -Falles auf  $f_1$  angewandt werden können, möge auch die alternative Dichte diesen Voraussetzungen genügen. Desweiteren wählen wir  $f_1$  standardisiert, d.h. für einen  $d$ -dimensionalen Zufallsvektor  $X$  mit zugrundeliegender Dichte  $f_1$  gelte  $EX = 0$  und  $EXX^T = I_d$ .

Seien im folgenden  $X_1, \dots, X_n$  unabhängige und identisch verteilte  $d$ -dimensionale Zufallsvektoren mit zugrundeliegender Dichte  $f_1$ . Um zur Bestimmung der Asymptotik der Bowman-Foster-Statistik die bereits vorhandenen Ergebnisse anwenden zu können, ist es zweckmäßig, die Zerlegung

$$\begin{aligned} \frac{\tilde{T}_{n,h,st}}{nh^d} &= \int \left( \hat{f}_{n,st}(x) - e_h(x, f_0) \right)^2 dx \\ &= \int \left( \hat{f}_{n,st}(x) - e_h(x, f_1) \right)^2 dx + \int \left( e_h(x, f_1) - e_h(x, f_0) \right)^2 dx \\ &\quad + 2 \int \left( \hat{f}_{n,st}(x) - e_h(x, f_1) \right) \left( e_h(x, f_1) - e_h(x, f_0) \right) dx \end{aligned}$$

zu betrachten, die aus einer mit der zugrundeliegenden Dichte  $f_1$  zentrierten Statistik vom Bowman-Foster-Typ, einem konstanten Term und einem gemischten Glied besteht. Zunächst stellt man fest, daß für die mit  $f_1$  zentrierte Statistik vom Bowman-Foster-Typ nach Theorem 31 auf Seite 93 mit einer asymptotisch standard-normalverteilten Zufallsvariablen  $Z_n$  gilt:

$$\begin{aligned} \sqrt{n} \int \left( \hat{f}_{n,st}(x) - e_h(x, f_1) \right)^2 dx &= \sqrt{n} \left( \frac{\mu(f_1) + \sqrt{2} h^{d/2} \sigma(f_1) Z_n}{nh^d} \right) \\ &= \frac{\mu(f_1)}{\sqrt{nh^d}} + \frac{\sqrt{2} \sigma(f_1) Z_n}{\sqrt{nh^d}} = \frac{\mu(f_1)}{\sqrt{nh^d}} + o_P(1), \end{aligned}$$

denn nach Voraussetzung strebt  $nh^d$  gegen Unendlich. Da die Bandbreite  $h$  gemäß Gleichung (2.9) auf Seite 77 gewählt wurde, ist der Term  $\mu(f_1)/(\sqrt{nh^d})$  nur für  $d \leq 3$  vernachlässigbar; für  $d = 4$  ist er konstant, und für  $d \geq 5$  liefert er einen Beitrag zur Asymptotik. Insgesamt kann man nach Subtraktion der deterministischen Terme und Reskalieren mit  $\sqrt{n}$  auf eine nichttriviale Grenzverteilung hoffen, d.h. wir interessieren uns im folgenden für die Zufallsvariable

$$\tilde{S}_{n,h,st} := \sqrt{n} \left( \frac{\tilde{T}_{n,h,st}}{nh^d} - \frac{\mu(f_1)}{nh^d} - \int \left( e_h(x, f_1) - e_h(x, f_0) \right)^2 dx \right) \quad (2.41)$$

$$= 2\sqrt{n} \int \left( \widehat{f}_{n,\text{st}}(x) - e_h(x, f_1) \right) (e_h(x, f_1) - e_h(x, f_0)) dx + o_P(1).$$

Zur Bestimmung der Varianz der Grenzverteilung von  $\widetilde{S}_{n,h,\text{st}}$  benötigen wir den Zeilenvektor  $\mathcal{I}_1$  und den Skalar  $\mathcal{I}_2$ , die durch

$$\mathcal{I}_1 := \int \nabla f_1(x) (f_1(x) - f_0(x)) dx, \quad \mathcal{I}_2 := \int f_1(x) (f_1(x) - f_0(x)) dx \quad (2.42)$$

definiert sind, sowie die Größen

$$\begin{aligned} \mathcal{I}_{3,r,s} &:= \int \nabla f_1(x) {}_r x_s (f_1(x) - f_0(x)) dx, \\ \mathcal{I}_3(A) &:= \int \nabla f_1(x) Ax (f_1(x) - f_0(x)) dx = \sum_{r,s=1}^d a_{rs} \mathcal{I}_{3,r,s} \end{aligned} \quad (2.43)$$

mit  $r, s \in \{1, \dots, d\}$  und  $A = (a_{rs})_{r,s=1,\dots,d} \in \mathbb{R}^{d \times d}$ . Ferner seien

$$\gamma_1 := \int f_1(x) (f_1(x) - f_0(x)) dx + \frac{d}{2} \mathcal{I}_2 + \frac{1}{2} \mathcal{I}_3(I_d) = \frac{d+2}{2} \mathcal{I}_2 + \frac{1}{2} \mathcal{I}_3(I_d), \quad (2.44)$$

$$\begin{aligned} \gamma_2 &:= \int f_1(x) (f_1(x) - f_0(x))^2 dx + \int (2\mathcal{I}_1 x + \mathcal{I}_2 x^T x + \mathcal{I}_3(x x^T)) f_1(x) (f_1(x) - f_0(x)) dx \\ &\quad + \mathcal{I}_1 \mathcal{I}_1^T + \frac{1}{4} \mathcal{I}_2^2 \mathbb{E} \left[ (X_1^T X_1)^2 \right] + \frac{1}{4} \mathbb{E} \left[ \mathcal{I}_3(X_1 X_1^T)^2 \right] + \mathcal{I}_2 \mathcal{I}_1 \mathbb{E} \left[ (X_1^T X_1) X_1 \right] \\ &\quad + \mathcal{I}_1 \mathbb{E} \left[ \mathcal{I}_3(X_1 X_1^T) X_1 \right] + \frac{1}{2} \mathcal{I}_2 \mathbb{E} \left[ (X_1^T X_1) \mathcal{I}_3(X_1 X_1^T) \right], \end{aligned} \quad (2.45)$$

wobei die Erwartungswerte bezüglich der alternativen Dichte  $f_1$  gebildet werden, und

$$\widetilde{\sigma}^2(f_1, f_0) := \gamma_2 - \gamma_1^2. \quad (2.46)$$

Damit läßt sich zeigen, daß  $\widetilde{S}_{n,h,\text{st}}$  unter  $f_1$  asymptotisch normalverteilt mit Erwartungswert 0 und Varianz  $4\widetilde{\sigma}^2(f_1, f_0)$  ist.

**Theorem 34** Unter den Voraussetzungen 1 und 2 an die Bandbreite  $h$ , die standardisierte  $H_0$ -Dichte  $f_0$ , den Kern  $K$  und die standardisierte alternative Dichte  $f_1 \neq f_0$  gilt

$$\frac{\widetilde{S}_{n,h,\text{st}}}{2\widetilde{\sigma}(f_1, f_0)} = \frac{\sqrt{n}}{2\widetilde{\sigma}(f_1, f_0)} \left( \frac{\widetilde{T}_{n,h,\text{st}}}{nh^d} - \frac{\mu(f_1)}{nh^d} - \int (e_h(x, f_1) - e_h(x, f_0))^2 dx \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

BEWEIS: Anhand der Taylor-Entwicklung ersten Grades

$$\widehat{f}_{n,\text{st}}(x) = \frac{1}{nh^d} \sum_{i=1}^n \left[ K\left(\frac{x - X_i}{h}\right) + dK_h(x, X_i, \vartheta_0) \circ (\widehat{\vartheta}_n - \vartheta_0) \right] + \frac{1}{nh^d} \sum_{i=1}^n R_1(x, X_i)$$

nach Gleichung (2.28) auf Seite 88 gelangt man zu der Darstellung

$$\widetilde{S}_{n,h,\text{st}} = 2\sqrt{n} \int \left( \widehat{f}_{n,\text{st}}(x) - e_h(x, f_1) \right) (e_h(x, f_1) - e_h(x, f_0)) dx + o_P(1)$$

$$= \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{1}{h^d} \int \left\{ K \left( \frac{x-X_i}{h} \right) - \mathbb{E}_{f_1} \left[ K \left( \frac{x-X_1}{h} \right) \right] \right\} (e_h(x, f_1) - e_h(x, f_0)) dx \\ + \tilde{S}_{n,1} + \tilde{S}_{n,2} + o_P(1)$$

mit

$$\tilde{S}_{n,1} := \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{1}{h^d} \int dK_h(x, X_i, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) (e_h(x, f_1) - e_h(x, f_0)) dx, \\ \tilde{S}_{n,2} := \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{1}{h^d} \int R_1(x, X_i) (e_h(x, f_1) - e_h(x, f_0)) dx. \quad (2.47)$$

Nach Lemma 35 auf Seite 105 gilt  $\tilde{S}_{n,2} = o_P(1)$ , d.h. dieser Term ist vernachlässigbar. Setzt man in  $\tilde{S}_{n,1}$  die Ableitung aus Gleichung (2.32) auf Seite 88 ein, so erhält man unter Berücksichtigung von Lemma 56 auf Seite 138 und Gleichung (1.15) auf Seite 14

$$\tilde{S}_{n,1} = \sum_{r=1}^d \frac{2}{n} \sum_{i=1}^n \frac{1}{h^{d+1}} \int \nabla K \left( \frac{x-X_i}{h} \right)_r (e_h(x, f_1) - e_h(x, f_0)) dx (\sqrt{n} \bar{X}_{nr}) \\ + \sum_{r,s=1}^d \frac{1}{n} \sum_{i=1}^n \frac{1}{h^{d+1}} X_{is} \int \nabla K \left( \frac{x-X_i}{h} \right)_r (e_h(x, f_1) - e_h(x, f_0)) dx (\sqrt{n} (S_n - I_d)_{rs}) \\ = \sum_{r=1}^d 2(\mathcal{I}_{1r} + o_P(1)) \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ir} \\ + \sum_{r,s=1}^d (\mathcal{I}_2 \mathbf{1}_{\{r=s\}} + \mathcal{I}_{3,r,s} + o_P(1)) \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i X_i^T - I_d)_{rs} + o_P(1) \right] \\ = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ 2\mathcal{I}_1 X_i + \sum_{r,s=1}^d (\mathcal{I}_2 \mathbf{1}_{\{r=s\}} + \mathcal{I}_{3,r,s}) (X_i X_i^T - I_d)_{rs} \right] + o_P(1) \\ = \frac{2}{\sqrt{n}} \sum_{i=1}^n \left[ \mathcal{I}_1 X_i + \frac{1}{2} \mathcal{I}_2 (X_i^T X_i - d) + \frac{1}{2} \mathcal{I}_3 (X_i X_i^T - I_d) \right] + o_P(1).$$

Dabei wurde in der letzten Zeile die Identität

$$\sum_{r,s=1}^d (X_i X_i^T - I_d)_{rs} \mathbf{1}_{\{r=s\}} = \sum_{r=1}^d (X_{ir} X_{ir} - (I_d)_{rr}) = X_i^T X_i - d$$

verwendet. Insgesamt ergibt sich für  $\tilde{S}_{n,h,st}$

$$\tilde{S}_{n,h,st} = \frac{2}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{1}{h^d} \int \left\{ K \left( \frac{x-X_i}{h} \right) - \mathbb{E}_{f_1} \left[ K \left( \frac{x-X_1}{h} \right) \right] \right\} (e_h(x, f_1) - e_h(x, f_0)) dx \right. \\ \left. + \mathcal{I}_1 X_i + \frac{1}{2} \mathcal{I}_2 (X_i^T X_i - d) + \frac{1}{2} \mathcal{I}_3 (X_i X_i^T - I_d) \right\} + o_P(1)$$

$$= \frac{2}{\sqrt{n}} \sum_{i=1}^n [Y_{ni} - \mathbb{E}Y_{ni}] + o_P(1),$$

d.h. das asymptotische Verhalten von  $\tilde{S}_{n,h,st}$  wird bestimmt durch eine zentrierte Summe von unabhängigen Zufallsvariablen

$$Y_{ni} := \frac{1}{h^d} \int K\left(\frac{x-X_i}{h}\right) (e_h(x, f_1) - e_h(x, f_0)) dx + \mathcal{I}_1 X_i + \frac{1}{2} \mathcal{I}_2 X_i^T X_i + \frac{1}{2} \mathcal{I}_3 (X_i X_i^T) \quad (2.48)$$

( $i = 1, \dots, n$ ). Zur Bestimmung der Grenzverteilung von  $\tilde{S}_{n,h,st}$  bietet sich folglich der Zentrale Grenzwertsatz von Lindeberg–Feller an, d.h. mit

$$\sigma_n^{*2} := \text{Var} \left( \sum_{i=1}^n Y_{ni} \right) = n \text{Var}(Y_{n1}) = n [\tilde{\sigma}^2(f_1, f_0) + O(h^2)] \rightarrow \infty \quad (n \rightarrow \infty)$$

nach Lemma 57 auf Seite 142 ist die Lindeberg–Bedingung

$$\frac{1}{\sigma_n^{*2}} \sum_{i=1}^n \mathbb{E} [(Y_{ni} - \mathbb{E}Y_{ni})^2 \mathbf{1}\{|Y_{ni} - \mathbb{E}Y_{ni}| > \varepsilon \sigma_n^*\}] \rightarrow 0 \quad (n \rightarrow \infty) \quad \forall \varepsilon > 0$$

zu überprüfen. Sei also im folgenden  $\varepsilon > 0$  fest. Zunächst stellen wir fest, daß punktweise auf  $\Omega$  mit der Substitution  $y = (x - X_1)/h$ ,  $dy = h^{-d} dx$ , der Definition von  $\mathcal{I}_3$  und nach Lemma 55 auf Seite 136 die Abschätzung

$$\begin{aligned} |Y_{n1}| &\leq \int K(y) |e_h(X_1 + hy, f_1) - e_h(X_1 + hy, f_0)| dy + \|\mathcal{I}_1\| \|X_1\| + \frac{1}{2} \|\mathcal{I}_2\| \|X_1\|^2 \\ &\quad + \frac{1}{2} \|X_1\|^2 \int \|\nabla f_1(x)\| \|x\| |f_1(x) - f_0(x)| dx \\ &\leq \text{const} \cdot (1 + \|X_1\| + \|X_1\|^2) \end{aligned}$$

mit einer von  $n$  unabhängigen oberen Schranke gilt. Unter Berücksichtigung von  $\sigma_n^{*2} \rightarrow \infty$  folgt daraus

$$(Y_{n1} - \mathbb{E}Y_{n1})^2 \mathbf{1}\{|Y_{n1} - \mathbb{E}Y_{n1}| > \varepsilon \sigma_n^*\} \rightarrow 0 \quad (n \rightarrow \infty)$$

punktweise auf  $\Omega$ . Da außerdem aufgrund der Existenz der Momente vierter Ordnung eine integrierbare Majorante durch die von  $n$  unabhängige obere Schranke

$$|Y_{n1} - \mathbb{E}Y_{n1}|^2 \mathbf{1}\{|Y_{n1} - \mathbb{E}Y_{n1}| > \varepsilon \sigma_n^*\} \leq |Y_{n1} - \mathbb{E}Y_{n1}|^2 \leq \text{const} \sum_{l=0}^4 \|X_1\|^l$$

gegeben ist, gilt die Lindeberg–Bedingung nach dem Satz von der majorisierten Konvergenz:

$$\begin{aligned} &\frac{1}{\sigma_n^{*2}} \sum_{i=1}^n \mathbb{E} [(Y_{ni} - \mathbb{E}Y_{ni})^2 \mathbf{1}\{|Y_{ni} - \mathbb{E}Y_{ni}| > \varepsilon \sigma_n^*\}] \\ &= \frac{1}{\tilde{\sigma}^2(f_1, f_0) + O(h^2)} \mathbb{E} [(Y_{n1} - \mathbb{E}Y_{n1})^2 \mathbf{1}\{|Y_{n1} - \mathbb{E}Y_{n1}| > \varepsilon \sigma_n^*\}] \rightarrow 0. \end{aligned}$$

Damit folgt die Behauptung aus dem Zentralen Grenzwertsatz von Lindeberg-Feller:

$$\frac{\sum_{i=1}^n [Y_{ni} - \mathbb{E}Y_{ni}]}{\sqrt{n} \tilde{\sigma}(f_1, f_0)} = \frac{\sqrt{\tilde{\sigma}^2(f_1, f_0) + O(h^2)}}{\tilde{\sigma}(f_1, f_0)} \cdot \frac{\sum_{i=1}^n [Y_{ni} - \mathbb{E}Y_{ni}]}{\sigma_n^*} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

□

**Lemma 35** Es gilt  $\tilde{S}_{n,2} = o_P(1)$ .

BEWEIS: Nach der Definition (2.47) von  $\tilde{S}_{n,2}$  gilt zunächst

$$\left| \tilde{S}_{n,2} \right| \leq \frac{2}{\sqrt{nh^d}} \sum_{i=1}^n \int |R_1(x, X_i)| |e_h(x, f_1) - e_h(x, f_0)| dx.$$

Wählen wir wie früher die Abkürzungen

$$\tilde{\vartheta}_n := \vartheta_0 + t(\hat{\vartheta}_n - \vartheta_0), \quad \tilde{\mu}_n := t\bar{X}_n, \quad \tilde{\Sigma}_n := I_d + t(S_n - I_d)$$

für die in dem Restglied  $R_1(x, X_i)$  auftretenden Parameter, so ergibt sich mit der Abschätzung von  $R_1(x, X_i)$  aus Lemma 28 auf Seite 89 und der Abschätzung von  $e_h(x, f_1)$  aus Lemma 55 auf Seite 136

$$\left| \tilde{S}_{n,2} \right| \leq O(1) \frac{1}{n^{3/2} h^{d+2}} \sum_{i=1}^n \left( \sum_{l=0}^2 \|X_i\|^l \right) \int_0^1 (1-t) \int \left\{ \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\| \right\} dx dt.$$

Da die Behandlung von  $\left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\|$  und  $\left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta}_n)) \right\|$  identisch verläuft, beschränken wir uns auf einen Vertreter. Für diesen gilt mit der Substitution  $y = (x - \tilde{\Sigma}_n^{-1/2}(X_i - \tilde{\mu}_n))/h$ ,  $dy = h^{-d} dx$  nach Voraussetzung 2 und dem Starken Gesetz großer Zahlen die Abschätzung

$$\begin{aligned} & O(1) \frac{1}{n^{3/2} h^{d+2}} \sum_{i=1}^n \left( \sum_{l=0}^2 \|X_i\|^l \right) \int_0^1 (1-t) \int \left\| \nabla^2 K \left( \frac{x - \tilde{\Sigma}_n^{-1/2}(X_i - \tilde{\mu}_n)}{h} \right) \right\| dx dt \\ &= O(1) \frac{1}{\sqrt{nh^2}} \frac{1}{n} \sum_{i=1}^n \left( \sum_{l=0}^2 \|X_i\|^l \right) \int_0^1 (1-t) \int \left\| \nabla^2 K(y) \right\| dy dt = o_P(1). \end{aligned}$$

□

**Korollar 36** Die Bowman-Foster-Statistik sei affin-invariant (z.B. gelte  $d = 1$ , oder  $\mathcal{F}$  sei die Familie der  $d$ -dimensionalen Normalverteilungen und  $K$  der Gaußkern aus Gleichung (2.2)).

Unter den Voraussetzungen 1 und 2 an die Bandbreite  $h$ , die  $H_0$ -Dichte und den Kern  $K$  ist der auf der Bowman–Foster–Statistik basierende Test für die zusammengesetzte Hypothese aus Bemerkung 33 auf Seite 100 konsistent gegen jede alternative Verteilung, deren Dichte  $f_1$  den Voraussetzungen 1 und 2 genügt.

BEWEIS: Nach Theorem 34 gilt unter der alternativen Dichte  $f_1$ , die wir aufgrund der affinen Invarianz von  $\tilde{T}_{n,h,st}$  o.B.d.A. standardisiert wählen können, mit einer asymptotisch standardnormalverteilten Zufallsvariablen  $Z_n$

$$\begin{aligned} \frac{\tilde{T}_{n,h,st}}{nh^d} &= \frac{2\tilde{\sigma}(f_1, f_0)}{\sqrt{n}} Z_n + \frac{\mu(f_1)}{nh^d} + \int (e_h(x, f_1) - e_h(x, f_0))^2 dx \\ &\xrightarrow{P_{f_1}} \int (f_1(x) - f_0(x))^2 dx > 0, \end{aligned}$$

denn nach Voraussetzung strebt  $nh^d \rightarrow \infty$ , und nach Lemma 55 auf Seite 136 gilt die Beziehung  $\int (e_h(x, f_1) - e_h(x, f_0))^2 dx = \int (f_1(x) - f_0(x))^2 dx + O(h^2)$ .

Da nun nach Bemerkung 33 der auf der Bowman–Foster–Statistik basierende Test definiert ist durch

$$H_0 \text{ ablehnen} \iff \tilde{T}_{n,h,st} > \mu(f_0) + \sqrt{2} h^{d/2} \sigma(f_0) \Phi^{-1}(1 - \alpha)$$

mit dem  $(1 - \alpha)$ -Quantil der Standard–Normalverteilung  $\Phi^{-1}(1 - \alpha)$ , folgt für die Wahrscheinlichkeit, daß der Test eine feste Alternative  $f_1$  erkennt

$$\begin{aligned} P_{f_1} \left( \tilde{T}_{n,h,st} > \mu(f_0) + \sqrt{2} h^{d/2} \sigma(f_0) \Phi^{-1}(1 - \alpha) \right) \\ = P_{f_1} \left( \frac{\tilde{T}_{n,h,st}}{nh^d} > \frac{\mu(f_0)}{nh^d} + \frac{\sqrt{2} \sigma(f_0) \Phi^{-1}(1 - \alpha)}{nh^{d/2}} \right) \rightarrow 1, \end{aligned}$$

denn  $\mu(f_0)/(nh^d) + \sqrt{2}\sigma(f_0)\Phi^{-1}(1 - \alpha)/(nh^{d/2})$  strebt gegen 0.  $\square$

**Bemerkung 37** Wie in der Situation des BHEP–Tests in Abschnitt 1.3.2 beschrieben, läßt sich auch im Fall der (affin invariant vorausgesetzten) Bowman–Foster–Statistik unter den Voraussetzungen 1 und 2 anhand von Theorem 34 eine Näherung für die Güte des zugehörigen Tests angeben. Diese hat im Fall der Bowman–Foster–Statistik jedoch eine geringe praktische Relevanz, da hier die Konvergenz von  $\tilde{S}_{n,h,st}/(2\tilde{\sigma}(f_1, f_0))$  gegen die Standard–Normalverteilung nur sehr langsam vonstatten geht, wie sich in Simulationen gezeigt hat. Dennoch ist es interessant, die Gütenäherung der Bowman–Foster–Statistik mit der des BHEP–Tests aus Gleichung (1.21) auf Seite 30 zu vergleichen.

Bezeichnen  $q_{h,1-\alpha}$  das  $H_0$ -Quantil der Bowman–Foster–Statistik und  $f_1$  die feste alternative Dichte, so ist nach Theorem 34

$$P_{f_1} \left( \tilde{T}_{n,h,st} > q_{h,1-\alpha} \right)$$

$$\begin{aligned}
&= 1 - P_{f_1} \left( \frac{\sqrt{n}}{2\tilde{\sigma}(f_1, f_0)} \left[ \frac{\tilde{T}_{n,h,st}}{nh^d} - \frac{\mu(f_1)}{nh^d} - \int (e_h(x, f_1) - e_h(x, f_0))^2 dx \right] \right) \\
&\leq \frac{\sqrt{n}}{2\tilde{\sigma}(f_1, f_0)} \left[ \frac{q_{h,1-\alpha}}{nh^d} - \frac{\mu(f_1)}{nh^d} - \int (e_h(x, f_1) - e_h(x, f_0))^2 dx \right] \\
&\approx 1 - \Phi \left( \frac{q_{h,1-\alpha} - \mu(f_1) - nh^d \int (e_h(x, f_1) - e_h(x, f_0))^2 dx}{2\tilde{\sigma}(f_1, f_0)\sqrt{nh^d}} \right)
\end{aligned}$$

die gesuchte Gütenäherung, wobei  $\Phi$  die Verteilungsfunktion der Standard-Normalverteilung bezeichnet.

**Bemerkung 38** Die Aussage von Theorem 34 behält ihre Gültigkeit, wenn die Varianz  $\tilde{\sigma}^2(f_1, f_0)$  durch einen konsistenten Schätzer  $\tilde{\sigma}_n^2(f_1, f_0)$  ersetzt wird:

$$\frac{\tilde{S}_{n,h,st}}{2\tilde{\sigma}_n(f_1, f_0)} = \frac{\sqrt{n}}{2\tilde{\sigma}_n(f_1, f_0)} \left( \frac{\tilde{T}_{n,h,st}}{nh^d} - \frac{\mu(f_1)}{nh^d} - \int (e_h(x, f_1) - e_h(x, f_0))^2 dx \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Ein konsistenter Schätzer  $\tilde{\sigma}_n^2(f_1, f_0)$  für die Grenzvarianz  $\tilde{\sigma}^2(f_1, f_0)$  läßt sich konstruieren, indem die in den Gleichungen (2.44) und (2.45) auf Seite 102 aufgeführten Bestandteile von  $\tilde{\sigma}^2(f_1, f_0)$  durch auf standardisierten Daten beruhende empirische Analoga ersetzt werden. Der Konsistenznachweis erfolgt ähnlich wie in Abschnitt 1.2 im Fall der BHEP-Tests durch Herausrechnen geschätzter Parameter, den Nachweis der Vernachlässigbarkeit von Resttermen und das Anwenden des starken Gesetzes großer Zahlen sowie von Konvergenzsätzen für  $U$ -Statistiken. Da im Fall der Bowman-Foster-Statistik aufgrund der sehr langsamen Konvergenz von  $\tilde{S}_{n,h,st}$  jedoch keine praxisrelevanten statistischen Folgerungen — wie z.B. Konfidenzintervalle, eine Gütenäherung oder ein inverser Test — aus Theorem 34 zu erwarten sind, für die ein solcher Schätzer benötigt würde, verzichten wir darauf, die Einzelheiten der Konstruktion von  $\tilde{\sigma}_n^2(f_1, f_0)$  anzugeben.

**Bemerkung 39** Durch Nachvollziehen der Überlegungen zu Beginn dieses Abschnitts und von Teilen des Beweises von Theorem 34 läßt sich unter den Voraussetzungen 1 und 2 leicht die Asymptotik der Bowman-Foster-Statistik  $\tilde{T}_{n,h}$  ohne geschätzte Parameter unter einer festen alternativen Dichte  $f_1$  angeben: Wie im Fall von  $\tilde{T}_{n,h,st}$  schreibt man

$$\begin{aligned}
\frac{\tilde{T}_{n,h}}{nh^d} &= \int \left( \hat{f}_n(x) - e_h(x, f_0) \right)^2 dx \\
&= \int \left( \hat{f}_n(x) - e_h(x, f_1) \right)^2 dx + \int (e_h(x, f_1) - e_h(x, f_0))^2 dx \\
&\quad + 2 \int \left( \hat{f}_n(x) - e_h(x, f_1) \right) (e_h(x, f_1) - e_h(x, f_0)) dx
\end{aligned}$$

und stellt anhand von Theorem 29 auf Seite 91 fest, daß mit einer asymptotisch standardnormalverteilten Zufallsvariablen  $Z_n$  gilt:

$$\sqrt{n} \int \left( \widehat{f}_n(x) - e_h(x, f_1) \right)^2 dx = \frac{\mu(f_1)}{\sqrt{nh^d}} + \frac{\sqrt{2} \sigma(f_1) Z_n}{\sqrt{nh^d}} = \frac{\mu(f_1)}{\sqrt{nh^d}} + o_P(1).$$

Analog zu Gleichung (2.41) auf Seite 101 und dem Beweis von Theorem 34 gilt für die Größe

$$\begin{aligned} \widetilde{S}_{n,h} &:= \sqrt{n} \left( \frac{\widetilde{T}_{n,h}}{nh^d} - \frac{\mu(f_1)}{nh^d} - \int (e_h(x, f_1) - e_h(x, f_0))^2 dx \right) \\ &= 2\sqrt{n} \int \left( \widehat{f}_n(x) - e_h(x, f_1) \right) (e_h(x, f_1) - e_h(x, f_0)) dx + o_P(1) \\ &= \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{1}{h^d} \int \left\{ K\left(\frac{x - X_i}{h}\right) - E_{f_1} \left[ K\left(\frac{x - X_1}{h}\right) \right] \right\} (e_h(x, f_1) - e_h(x, f_0)) dx + o_P(1) \\ &\xrightarrow{\mathcal{D}} \mathcal{N}(0, 4 \widetilde{\sigma}_*^2(f_1, f_0)) \end{aligned}$$

mit

$$\widetilde{\sigma}_*^2(f_1, f_0) = \gamma_2^* - (\gamma_1^*)^2 \quad (2.49)$$

und

$$\gamma_1^* := \int f_1(x) (f_1(x) - f_0(x)) dx, \quad \gamma_2^* := \int f_1(x) (f_1(x) - f_0(x))^2 dx.$$

Die Details ergeben sich aus dem Nachvollziehen des Beweises von Theorem 34 und Lemma 57 auf Seite 142, wobei die von den Schätzern stammenden Anteile in  $\widetilde{S}_{n,h,st}$  wegzulassen sind.

**Bemerkung 40** Das Ergebnis von Bemerkung 39 ist konform mit den Resultaten von Kim et al. (1997), die im Kontext einer einfachen Hypothese und im Fall  $d = 1$  für die Bickel–Rosenblatt–Statistik

$$T_{n,h,f} = \int \left( \widehat{f}_n(x) - f_0(x) \right)^2 dx$$

aus Gleichung (2.10) auf Seite 77 mit  $\widetilde{\sigma}_*^2(f_1, f_0)$  aus (2.49) die Asymptotik

$$\frac{\sqrt{n}(T_n - ET_n)}{2\widetilde{\sigma}_*(f_1, f_0)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

unter einer festen Alternative  $f_1$  herleiteten. Man beachte in diesem Zusammenhang, daß sich die unterschiedlichen Zentrierungen bei Kim et al. (1997) und im Fall der Bowman–Foster–Statistik unter Berücksichtigung von Resttermen leicht ineinander umrechnen lassen.

## 2.5. Lokale Alternativen

Ausgangspunkt unserer Überlegungen war ein Vergleich der Bowman–Foster–Statistik mit variabler Bandbreite  $h$  und der BHEP–Statistik mit festem Glättungsparameter  $\beta$  aus Kapitel 1 im Kontext des Testens auf multivariate Normalverteilung. Mittlerweile ist das Verhalten beider Statistiken unter  $H_0$  und unter festen Alternativen geklärt, und es verbleibt die Frage nach der Asymptotik unter lokalen Pitman–Alternativen. Im Fall der BHEP–Statistik wurde diese Frage von Henze und Wagner (1997) durch den Nachweis einer nichttrivialen Grenzverteilung unter lokalen Alternativen, die mit der Geschwindigkeit  $n^{-1/2}$  gegen die Hypothese einer  $d$ –dimensionalen Normalverteilung konvergieren, beantwortet. Folglich ist zu klären, ob die Bowman–Foster–Statistik ebenfalls in der Lage ist, lokale Alternativen zu erkennen, die mit der Geschwindigkeit  $n^{-1/2}$  gegen die Hypothese (2.17) einer bestimmten Lokations–Skalen–Familie streben. Dabei beschränken wir uns bei der Untersuchung der zusammengesetzten Hypothese auf die Fälle, in denen die Bowman–Foster–Statistik affin invariant ist, also z.B. auf die Dimension  $d = 1$  oder auf die Familie der  $d$ –dimensionalen Normalverteilungen und den Gaußkern aus Gleichung (2.2). Für alle weiteren Betrachtungen legen wir die folgende Situation lokaler Pitman–Alternativen zugrunde.

Sei  $X_{n1}, \dots, X_{nn}$  ein Dreiecksschema unabhängiger und zeilenweise identisch verteilter  $d$ –dimensionaler Zufallsvektoren mit einer gemeinsamen Dichte  $f_n$ , die definiert ist durch

$$f_n(x) := f_0(x) + N_n \mathfrak{w}(x) \quad \text{mit } N_n \rightarrow 0 \quad \text{und} \quad \int \mathfrak{w}(x) dx = 0. \quad (2.50)$$

Dabei ist  $f_0$  die zugrundeliegende  $H_0$ –Dichte, und  $N_n \mathfrak{w}$  bezeichnet einen Störanteil, der mit Geschwindigkeit  $N_n$  gegen  $f_0$  konvergiert, mit einer meßbaren und integrierbaren Funktion  $\mathfrak{w} : \mathbb{R}^d \rightarrow \mathbb{R}$ . Um sicherzustellen, daß  $f_n = f_0(1 + N_n \mathfrak{w}/f_0)$  nichtnegativ ist, fordern wir die Beschränktheit des Quotienten

$$\left| \frac{\mathfrak{w}(x)}{f_0(x)} \right| \leq \text{const} \quad \text{für alle } x \in \mathbb{R}^d \quad (2.51)$$

(im Fall  $f_0(x) = \mathfrak{w}(x) = 0$  ist per definitionem  $\mathfrak{w}(x)/f_0(x) := 0$ ) und setzen im folgenden stets  $n$  als hinreichend groß voraus.

Wie im  $H_0$ – bzw.  $H_1$ –Fall werden wir bei der Herleitung von Konvergenzaussagen für die Bowman–Foster–Statistik fordern, daß  $f_n$  den in den Voraussetzungen 1 und 2 festgehaltenen Regularitätsbedingungen genügt. Damit ist gemeint, daß  $f_0$  und  $\mathfrak{w}$  dies tun, mit der Ausnahme, daß  $\mathfrak{w}$  keine Dichte ist. Aus Gründen der Übersichtlichkeit formulieren wir die resultierenden Erweiterungen der Voraussetzungen 1 und 2 auf die Situation lokaler Alternativen als eigenständigen Punkt.

**Voraussetzung 1\***: Für die Bandbreite  $h$  gelte  $h \rightarrow 0$  und  $nh^d \rightarrow \infty$  für  $n \rightarrow \infty$ . Der Kern  $K$  genüge der Voraussetzung 1, ebenso wie die  $H_0$ –Dichte  $f_0$ . Die Dichte  $f_n$  sei gegeben durch  $f_n(x) = f_0(x) + N_n \mathfrak{w}(x)$  mit  $N_n \rightarrow 0$  und einer meßbaren, integrierbaren, und zweimal stetig differenzierbaren Funktion  $\mathfrak{w} : \mathbb{R}^d \rightarrow \mathbb{R}$ , für die gelte

$$\int \mathfrak{w}(x) dx = 0 \quad \text{und} \quad \forall x \in \mathbb{R}^d : \quad \left| \frac{\mathfrak{w}(x)}{f_0(x)} \right| \leq \text{const} .$$

$\mathfrak{w}$  und ihre partiellen Ableitungen zweiter Ordnung seien beschränkt und gleichmäßig stetig auf  $\mathbb{R}^d$ . Der Stichprobenumfang  $n$  sei so groß gewählt, daß  $f_n$  nichtnegativ ist.

**Voraussetzung 2\*:** Für die Bandbreite  $h$  gelte  $nh^{d+4} = \text{const.}$  Der Kern  $K$  genüge der Voraussetzung 2, ebenso wie die  $H_0$ -Dichte  $f_0$ . Zusätzlich zu Voraussetzung 1 gelte für  $\mathfrak{w}$ :  $\int \|x\|^m |\mathfrak{w}(x)| dx < \infty$  ( $m = 1, 2, 3, 4$ );  $\int \|x\|^l \|\nabla \mathfrak{w}(x)\| dx < \infty$  ( $l = 0, 1, 2$ ).

Wie in Lemma 55 auf Seite 136 definieren wir für beschränkte Kerne  $K$  die Erweiterung von  $e_h(x, \cdot)$  aus (2.20) auf Seite 83 auf meßbare integrierbare Funktionen  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ :

$$e_h(x, g) := \int K(y)g(x - hy)dy.$$

Analog zum i.i.d. –Fall seien

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_{ni}, \quad S_n := \frac{1}{n} \sum_{i=1}^n (X_{ni} - \bar{X}_n) (X_{ni} - \bar{X}_n)^T$$

das Stichprobenmittel und die empirische Kovarianzmatrix sowie  $Y_{ni} := S_n^{-1/2}(X_{ni} - \bar{X}_n)$  die standardisierten Daten. Die Kerndichteschätzer  $\hat{f}_n$  und  $\hat{f}_{n,\text{st}}$  seien wie in Gleichung (2.19) auf Seite 83 definiert. Wie bei der Untersuchung der  $H_0$ -Konvergenz in Kapitel 2.3 ist es zweckmäßig, die Bowman–Foster–Statistik

$$\tilde{T}_{n,h,\text{st}} = nh^d \int \left( \hat{f}_{n,\text{st}}(x) - e_h(x, f_0) \right)^2 dx$$

mit geschätzten Parametern gemäß  $\tilde{T}_{n,h,\text{st}} = \tilde{T}_{n,h} + \tilde{T}_{n,1} + \tilde{T}_{n,2}$  zu zerlegen in die Bowman–Foster–Statistik

$$\tilde{T}_{n,h} = nh^d \int \left( \hat{f}_n(x) - e_h(x, f_0) \right)^2 dx$$

ohne geschätzte Parameter und in zwei Restterme  $\tilde{T}_{n,1}$  und  $\tilde{T}_{n,2}$ , die wie früher definiert sind durch

$$\begin{aligned} \tilde{T}_{n,1} &:= nh^d \int \left( \hat{f}_{n,\text{st}}(x) - \hat{f}_n(x) \right)^2 dx, \\ \tilde{T}_{n,2} &:= 2nh^d \int \left( \hat{f}_{n,\text{st}}(x) - \hat{f}_n(x) \right) \left( \hat{f}_n(x) - e_h(x, f_0) \right) dx. \end{aligned}$$

### 2.5.1. Asymptotik ohne geschätzte Parameter

Wie im  $H_0$ -Fall bestimmen wir zunächst die Asymptotik der Bowman–Foster–Statistik  $\tilde{T}_{n,h}$  ohne geschätzte Parameter. Hierzu zerlegen wir diese gemäß

$$\tilde{T}_{n,h} = \tilde{R}_{n,h} + \tilde{R}_{n,1} + \tilde{R}_{n,2}$$

in eine entsprechende mit  $f_n$  zentrierte Statistik

$$\tilde{R}_{n,h} := nh^d \int \left( \hat{f}_n(x) - e_h(x, f_n) \right)^2 dx,$$

in einen deterministischen Restterm  $\tilde{R}_{n,1}$  und eine Zufallsvariable  $\tilde{R}_{n,2}$ , die definiert sind durch

$$\begin{aligned} \tilde{R}_{n,1} &:= nh^d \int (e_h(x, f_n) - e_h(x, f_0))^2 dx, \\ \tilde{R}_{n,2} &:= 2nh^d \int \left( \hat{f}_n(x) - e_h(x, f_n) \right) (e_h(x, f_n) - e_h(x, f_0)) dx \end{aligned}$$

und deren asymptotisches Verhalten zunächst getrennt studiert wird.

**Bemerkung 41** Für die gemäß Gleichung (2.22) auf Seite 83 definierten Größen  $\mu(f_n)$  und  $\sigma^2(f_n)$  gelten

$$\mu(f_n) = \mu(f_0) \quad \text{und} \quad \sigma^2(f_n) \rightarrow \sigma^2(f_0) \quad (n \rightarrow \infty).$$

Denn  $\mu(\cdot)$  ist nach Gleichung (2.22) konstant, also ist trivialerweise  $\mu(f_n) = \mu(f_0)$ . Der Varianzterm  $\sigma^2(f_n)$  strebt gegen  $\sigma^2(f_0)$  aufgrund der Definition in Gleichung (2.22) und der Tatsache

$$\int f_n^2(x) dx = \int f_0^2(x) dx + 2N_n \int f_0(x) \mathbf{w}(x) dx + N_n^2 \int \mathbf{w}^2(x) dx \rightarrow \int f_0^2(x) dx.$$

**Satz 42** Unter der Voraussetzung 1\* gilt

$$\frac{\tilde{R}_{n,h} - \mu(f_0)}{\sqrt{2} h^{d/2} \sigma(f_0)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

BEWEIS: Der Beweis ist eine wörtliche Wiedergabe von Teilen des Artikels Hall (1984) und wurde bereits im  $H_0$ -Fall in Kapitel 2.3 als Beweis des dortigen Theorems 29 skizziert. Außerdem ist die Aussage von Satz 42 bereits als Theorem 1 in Liero et al. (1998) angegeben, allerdings im Fall  $d = 1$ , unter weit restriktiveren Voraussetzungen und ohne Beweis. Wir skizzieren daher ein weiteres Mal die Struktur des Beweises von P. Hall und weisen an geeigneter Stelle darauf hin, warum man bei den einzelnen Schritten unter lokalen Alternativen zu denselben Resultaten wie im  $H_0$ -Fall gelangt.

Wie früher zerlegt man  $\tilde{R}_{n,h} = L_n + U_n$  in einen linearen Term

$$L_n := \frac{1}{nh^d} \sum_{i=1}^n \int \left\{ K\left(\frac{x - X_{ni}}{h}\right) - \mathbb{E}\left[K\left(\frac{x - X_{n1}}{h}\right)\right] \right\}^2 dx$$

und einen quadratischen Term

$$U_n := \frac{2}{nh^d} \sum_{i < j} \int \left\{ K\left(\frac{x - X_{ni}}{h}\right) - \mathbb{E}\left[K\left(\frac{x - X_{n1}}{h}\right)\right] \right\} \left\{ K\left(\frac{x - X_{nj}}{h}\right) - \mathbb{E}\left[K\left(\frac{x - X_{n1}}{h}\right)\right] \right\} dx,$$

deren Asymptotik getrennt behandelt wird.

1) BEHANDLUNG VON  $L_n$ :

Mit denselben Rechnungen wie im  $H_0$ -Fall (Substitutionen, Beschränktheit und Dichte-eigenschaft von  $f_n$ , Beschränktheit von  $K$ ) ergibt sich analog zu Hall (1984)

$$EL_n = \mu(f_0) - h^d \iint K(u)K(u+v) du dv \int f_n(x)f_n(x+hv)dx, \quad \text{Var}(L_n) = O\left(\frac{1}{n}\right)$$

und folglich

$$L_n = \mu(f_0) - h^d \iint K(u)K(u+v) du dv \int f_n(x)f_n(x+hv)dx + O_P\left(\frac{1}{n^{1/2}}\right).$$

2) BEHANDLUNG VON  $U_n$ :

$U_n = \frac{2}{nh^d} \sum_{i < j} H_n(X_{ni}, X_{nj})$  ist eine  $U$ -Statistik mit dem von  $n$  abhängenden, symmetrischen, zentrierten und degenerierten Kern

$$H_n(X_{ni}, X_{nj}) := \int \left\{ K\left(\frac{x-X_{ni}}{h}\right) - \text{EK}\left(\frac{x-X_{n1}}{h}\right) \right\} \left\{ K\left(\frac{x-X_{nj}}{h}\right) - \text{EK}\left(\frac{x-X_{n1}}{h}\right) \right\} dx.$$

Die Behandlung von  $U_n$  erfolgt mit dem in Hall (1984) hergeleiteten und in Anhang D.2 als Theorem D.4 auf Seite 182 aufgeführten Grenzwertsatz für  $U$ -Statistiken mit variablem Kern. Genauer gesagt benötigen wir die Version dieses Satzes für Dreiecksschemata unabhängiger Zufallsvariablen, die in Theorem D.5 festgehalten und im Anschluß daran bewiesen ist. Mit

$$G_n(x, y) := \text{E} [H_n(X_{n1}, x)H_n(X_{n1}, y)]$$

läßt sich

$$\text{E} [H_n^2(X_{n1}, X_{n2})] = h^{3d} \sigma^2(f_n) + O(h^{4d})$$

zeigen sowie

$$\text{E} [H_n^4(X_{n1}, X_{n2})] = O(h^{5d}), \quad \text{E} [G_n^2(X_{n1}, X_{n2})] = O(h^{7d}).$$

Die zugehörigen umfangreichen Rechnungen verlaufen exakt wie im  $H_0$ -Fall, wobei hauptsächlich Substitutionen und die Taylor-Entwicklung von  $f_n$  nach Bemerkung 22 auf Seite 85 verwendet werden. Damit können wir angesichts von  $\sigma^2(f_n) \rightarrow \sigma^2(f_0)$  nach Bemerkung 41 die Voraussetzung von Theorem D.5 nachprüfen:

$$\frac{\text{E} [G_n^2(X_{n1}, X_{n2})] + \frac{1}{n} \text{E} [H_n^4(X_{n1}, X_{n2})]}{(\text{E} [H_n^2(X_{n1}, X_{n2})])^2} = \frac{O(h^d) + O\left(\frac{1}{nh^d}\right)}{[\sigma^2(f_n) + O(h^d)]^2} \rightarrow 0.$$

Damit folgt die Asymptotik von  $U_n$  nach Theorem D.5:

$$\frac{U_n}{\sqrt{2} h^{d/2} \sigma(f_n)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{bzw.} \quad \frac{U_n}{\sqrt{2} h^{d/2} \sigma(f_0)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

3) ASYMPTOTIK VON  $\tilde{R}_{n,h}$ :

Wegen  $\tilde{R}_{n,h} = L_n + U_n$  gilt mit einer asymptotisch standard-normalverteilten Zufallsvariablen  $Z_n$

$$\begin{aligned} \tilde{R}_{n,h} &= \mu(f_0) - h^d \iint K(u)K(u+v)dudv \int f_n(x)f_n(x+hv)dx + O_P\left(\frac{1}{n^{1/2}}\right) \\ &\quad + \sqrt{2} h^{d/2} \sigma(f_0) Z_n, \end{aligned}$$

also

$$\frac{\tilde{R}_{n,h} - \mu(f_0)}{\sqrt{2} h^{d/2} \sigma(f_0)} = Z_n + o(1) + O_P\left(\frac{1}{\sqrt{nh^d}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \quad \square$$

**Satz 43** Unter der Voraussetzung 1\* gilt

$$\tilde{R}_{n,1} = nh^d N_n^2 \left[ \int \mathfrak{w}^2(x) dx + O(h^2) \right].$$

BEWEIS: Nach Definition ist

$$\begin{aligned} e_h(x, f_n) - e_h(x, f_0) &= \int K(y) [f_n(x-hy) - f_0(x-hy)] dy \\ &= N_n \int K(y) \mathfrak{w}(x-hy) dy = N_n e_h(x, \mathfrak{w}), \end{aligned} \quad (2.52)$$

also

$$\tilde{R}_{n,1} = nh^d \int (e_h(x, f_n) - e_h(x, f_0))^2 dx = nh^d N_n^2 \int e_h(x, \mathfrak{w})^2 dx,$$

und nach Lemma 55 auf Seite 136 gilt  $\int e_h(x, \mathfrak{w})^2 dx = \int \mathfrak{w}(x)^2 dx + O(h^2)$ . □

**Satz 44** Mit der Bezeichnung

$$\sigma_R^2 := \int \mathfrak{w}(x)^2 f_0(x) dx - \left( \int \mathfrak{w}(x) f_0(x) dx \right)^2 \quad (2.53)$$

gilt unter der Voraussetzung 1\*

$$\frac{\tilde{R}_{n,2}}{2 \sigma_R \sqrt{nh^{2d} N_n^2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

BEWEIS: Mit  $e_h(x, f_n) - e_h(x, f_0) = N_n e_h(x, \mathfrak{w})$  nach Gleichung (2.52) besitzt  $\tilde{R}_{n,2}$  die Darstellung

$$\begin{aligned}\tilde{R}_{n,2} &= 2nh^d \int \left( \hat{f}_n(x) - e_h(x, f_n) \right) (e_h(x, f_n) - e_h(x, f_0)) dx \\ &= 2N_n h^d \sum_{i=1}^n \frac{1}{h^d} \int \left[ K\left(\frac{x - X_{ni}}{h}\right) - \mathbb{E}K\left(\frac{x - X_{n1}}{h}\right) \right] e_h(x, \mathfrak{w}) dx \\ &= 2N_n h^d \sum_{i=1}^n \left[ \tilde{Y}_{ni} - \mathbb{E}\tilde{Y}_{ni} \right]\end{aligned}$$

als zentrierte Summe von unabhängigen Zufallsvariablen

$$\tilde{Y}_{ni} := \frac{1}{h^d} \int K\left(\frac{x - X_{ni}}{h}\right) e_h(x, \mathfrak{w}) dx \quad (2.54)$$

( $i = 1, \dots, n$ ), für die nach Lemma 58 auf Seite 147 gilt:

$$\sigma_{R,n}^2 := \text{Var}\left(\sum_{i=1}^n \tilde{Y}_{ni}\right) = n \text{Var}\tilde{Y}_{n1} = n[\sigma_R^2 + o(1)] \rightarrow \infty \quad (n \rightarrow \infty).$$

Wie im Beweis von Theorem 34 auf Seite 102 wenden wir den Zentralen Grenzwertsatz von Lindeberg–Feller an und überprüfen die Lindeberg–Bedingung

$$\frac{1}{\sigma_{R,n}^2} \sum_{i=1}^n \mathbb{E} \left[ \left( \tilde{Y}_{ni} - \mathbb{E}\tilde{Y}_{ni} \right)^2 \mathbf{1} \left\{ \left| \tilde{Y}_{ni} - \mathbb{E}\tilde{Y}_{ni} \right| > \varepsilon \sigma_{R,n} \right\} \right] \rightarrow 0 \quad (n \rightarrow \infty) \quad \forall \varepsilon > 0.$$

Sei also im folgenden  $\varepsilon > 0$  fest. Zunächst stellen wir fest, daß punktweise auf  $\Omega$  mit der Substitution  $y = (x - X_{n1})/h$ ,  $dy = h^{-d} dx$  nach Lemma 55 auf Seite 136 die Abschätzung

$$\left| \tilde{Y}_{n1} \right| \leq \int K(y) |e_h(X_{n1} + hy, \mathfrak{w})| dy \leq \text{const}$$

mit einer von  $n$  und von  $\omega$  unabhängigen oberen Schranke gilt. Unter Berücksichtigung von  $\sigma_{R,n}^2 \rightarrow \infty$  folgt, daß die Indikatorfunktion  $\mathbf{1} \left\{ \left| \tilde{Y}_{n1} - \mathbb{E}\tilde{Y}_{n1} \right| > \varepsilon \sigma_{R,n} \right\}$  für hinreichend großes  $n$  identisch 0 ist, und damit die Gültigkeit der Lindeberg–Bedingung:

$$\begin{aligned}\frac{1}{\sigma_{R,n}^2} \sum_{i=1}^n \mathbb{E} \left[ \left( \tilde{Y}_{ni} - \mathbb{E}\tilde{Y}_{ni} \right)^2 \mathbf{1} \left\{ \left| \tilde{Y}_{ni} - \mathbb{E}\tilde{Y}_{ni} \right| > \varepsilon \sigma_{R,n} \right\} \right] \\ = \frac{1}{\sigma_R^2 + o(1)} \mathbb{E} \left[ \left( \tilde{Y}_{n1} - \mathbb{E}\tilde{Y}_{n1} \right)^2 \mathbf{1} \left\{ \left| \tilde{Y}_{n1} - \mathbb{E}\tilde{Y}_{n1} \right| > \varepsilon \sigma_{R,n} \right\} \right] \rightarrow 0.\end{aligned}$$

Also gilt die Behauptung nach dem Zentralen Grenzwertsatz von Lindeberg–Feller:

$$\frac{\sum_{i=1}^n \left[ \tilde{Y}_{ni} - \mathbb{E}\tilde{Y}_{ni} \right]}{\sqrt{n} \sigma_R} = \frac{\sqrt{\sigma_R^2 + o(1)}}{\sigma_R} \cdot \frac{\sum_{i=1}^n \left[ \tilde{Y}_{ni} - \mathbb{E}\tilde{Y}_{ni} \right]}{\sigma_{R,n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad \square$$

Aus den Sätzen 42, 43 und 44 läßt sich nun unter der Voraussetzung 1\* angesichts der Zerlegung

$$\frac{\tilde{T}_{n,h} - \mu(f_0)}{\sqrt{2} \sigma(f_0) h^{d/2}} = \frac{\tilde{R}_{n,h} - \mu(f_0)}{\sqrt{2} \sigma(f_0) h^{d/2}} + \frac{\tilde{R}_{n,1}}{\sqrt{2} \sigma(f_0) h^{d/2}} + \frac{\tilde{R}_{n,2}}{\sqrt{2} \sigma(f_0) h^{d/2}} \quad (2.55)$$

das Verhalten von  $\tilde{T}_{n,h}$  unter lokalen Alternativen  $f_n$  gemäß (2.50) bestimmen. Nach Satz 42 ist der erste Term in (2.55) für jede Lokalisierungsrate  $N_n$  asymptotisch standard-normalverteilt. Für den zweiten Term gilt nach Satz 43 beim Grenzübergang  $n \rightarrow \infty$  mit einer Konstanten  $c \in (0, \infty)$

$$\frac{\tilde{R}_{n,1}}{\sqrt{2} \sigma(f_0) h^{d/2}} = nh^{d/2} N_n^2 \frac{\int \mathfrak{w}^2(x) dx + O(h^2)}{\sqrt{2} \sigma(f_0)} \rightarrow \begin{cases} \infty & , \quad nh^{d/2} N_n^2 \rightarrow \infty , \\ c \frac{\int \mathfrak{w}^2(x) dx}{\sqrt{2} \sigma(f_0)} & , \quad nh^{d/2} N_n^2 \rightarrow c , \\ 0 & , \quad nh^{d/2} N_n^2 \rightarrow 0 , \end{cases}$$

und für den dritten Term nach Satz 44 mit einer asymptotisch standard-normalverteilten Zufallsvariablen  $Z_n$  und einer Konstanten  $c \in (0, \infty)$

$$\frac{\tilde{R}_{n,2}}{\sqrt{2} \sigma(f_0) h^{d/2}} = \sqrt{nh^d N_n^2} \sqrt{2} \frac{\sigma_R}{\sigma(f_0)} Z_n \xrightarrow{\mathcal{D}} \begin{cases} \frac{1}{2} (\delta_\infty + \delta_{-\infty}) & , \quad \sqrt{nh^d N_n^2} \rightarrow \infty , \\ c \sqrt{2} \frac{\sigma_R}{\sigma(f_0)} \mathcal{N}(0, 1) & , \quad \sqrt{nh^d N_n^2} \rightarrow c , \\ 0 & , \quad \sqrt{nh^d N_n^2} \rightarrow 0 , \end{cases}$$

wobei  $\delta_x$  das Dirac-Maß im Punkt  $x$  bezeichnet. Im folgenden wählen wir  $h$  gemäß Gleichung (2.9) auf Seite 77, also  $h = \kappa_d \cdot n^{-\frac{1}{d+4}}$  mit einer Konstanten  $\kappa_d$ , und betrachten Lokalisierungsraten der Form  $N_n = n^{-r}$  mit  $r \in (0, \frac{1}{2}]$ . Setzen wir ferner

$$\gamma_d := \frac{\kappa_d^{d/2}}{\sqrt{2} \sigma(f_0)} \int \mathfrak{w}^2(x) dx > 0 , \quad (2.56)$$

so nimmt die Asymptotik von  $\tilde{R}_{n,1}$  und  $\tilde{R}_{n,2}$  die Gestalt

$$\frac{\tilde{R}_{n,1}}{\sqrt{2} \sigma(f_0) h^{d/2}} = n^{1-\frac{d}{2(d+4)}} N_n^2 \frac{\kappa_d^{d/2} (\int \mathfrak{w}^2(x) dx + O(h^2))}{\sqrt{2} \sigma(f_0)} \rightarrow \begin{cases} \infty & , \quad 0 < r < \frac{1}{2} - \frac{d}{4(d+4)} , \\ \gamma_d & , \quad r = \frac{1}{2} - \frac{d}{4(d+4)} , \\ 0 & , \quad \frac{1}{2} - \frac{d}{4(d+4)} < r \leq \frac{1}{2} \end{cases}$$

bzw.

$$\begin{aligned} & \frac{\tilde{R}_{n,2}}{\sqrt{2} \sigma(f_0) h^{d/2}} \\ &= n^{\frac{1}{2}-\frac{d}{2(d+4)}} N_n \kappa_d^{d/2} \sqrt{2} \frac{\sigma_R}{\sigma(f_0)} Z_n \xrightarrow{\mathcal{D}} \begin{cases} \frac{1}{2} (\delta_\infty + \delta_{-\infty}) & , \quad 0 < r < \frac{1}{2} - \frac{d}{2(d+4)} , \\ \kappa_d^{d/2} \sqrt{2} \frac{\sigma_R}{\sigma(f_0)} \mathcal{N}(0, 1) & , \quad r = \frac{1}{2} - \frac{d}{2(d+4)} , \\ 0 & , \quad \frac{1}{2} - \frac{d}{2(d+4)} < r \leq \frac{1}{2} \end{cases} \end{aligned}$$

an. Angesichts von (2.55) ergibt sich daraus das folgende Resultat für die Asymptotik von  $\tilde{T}_{n,h}$  in dem für uns interessanten Bereich in der Nähe der Lokalisierungsrate  $n^{-1/2}$ .

**Theorem 45** Unter der Voraussetzung 1\* gilt mit  $\gamma_d$  aus Gleichung (2.56)

$$\frac{\tilde{T}_{n,h} - \mu(f_0)}{\sqrt{2} \sigma(f_0) h^{d/2}} \xrightarrow{\mathcal{D}} \begin{cases} \delta_\infty & , \quad \frac{1}{2} - \frac{d}{2(d+4)} < r < \frac{1}{2} - \frac{d}{4(d+4)} , \\ \mathcal{N}(\gamma_d, 1) & , \quad r = \frac{1}{2} - \frac{d}{4(d+4)} , \\ \mathcal{N}(0, 1) & , \quad \frac{1}{2} - \frac{d}{4(d+4)} < r \leq \frac{1}{2} . \end{cases} \quad (2.57)$$

Die Bowman–Foster–Statistik  $\tilde{T}_{n,h}$  ohne geschätzte Parameter besitzt also eine nichttriviale Grenzverteilung bei der Lokalisierungsrate  $n^{-\left(\frac{1}{2} - \frac{d}{4(d+4)}\right)}$  und ist nicht in der Lage, lokale Pitman–Alternativen zu erkennen, die mit der Geschwindigkeit  $n^{-r}$ ,  $\frac{1}{2} - \frac{d}{4(d+4)} < r \leq \frac{1}{2}$ , gegen die Hypothese konvergieren. Insbesondere werden die „klassischen“ benachbarten Alternativen im Sinne von Witting und Müller–Funk (1995) (vgl. Anhang E und Bemerkung 50 auf Seite 123) mit Lokalisierungsrate  $n^{-1/2}$  nicht erkannt.

**Bemerkung 46** Das Resultat von Theorem 45 für die Bowman–Foster–Statistik  $\tilde{T}_{n,h}$  ohne geschätzte Parameter ist konform mit den Ergebnissen in Fan (1994), Theorem 4.2 (ii) für den Bickel–Rosenblatt–Test. Im Eindimensionalen findet man eine ähnliche Diskussion in Liero et al. (1998). Allerdings wird dort bei der Untersuchung der zusammengesetzten Hypothese das Problem der zu schätzenden Parameter anders gelöst: Es werden keine standardisierten Daten  $Y_{nj}$  anstelle der  $X_{nj}$  eingesetzt, sondern die Bickel–Rosenblatt–Statistik mit  $E_{f(\cdot, \hat{\vartheta}_n)}[\hat{f}_n(x)]$  anstelle von  $E_{f(\cdot, \vartheta)}[\hat{f}_n(x)]$  zentriert (vgl. die Diskussion auf Seite 78).

## 2.5.2. Behandlung der geschätzten Parameter

Um das Verhalten der Bowman–Foster–Statistik  $\tilde{T}_{n,h,\text{st}}$  mit geschätzten Parametern unter lokalen Alternativen exakt zu bestimmen, müssen die entsprechenden Rechnungen des  $H_0$ –Falles, d.h. der Beweis von Theorem 31 ab Seite 94 neu aufgerollt werden. Im folgenden gehen wir deshalb davon aus, daß die Bandbreite  $h$ , der Kern  $K$ , die Dichte  $f_0$  und  $\mathfrak{w}$  nicht nur der Voraussetzung 1\*, sondern auch der Voraussetzung 2\* genügen. Desweiteren sei ab jetzt die Bowman–Foster–Statistik  $\tilde{T}_{n,h,\text{st}}$  affin invariant (z.B. gelte  $d = 1$ , oder  $\mathcal{F}$  sei die Familie der  $d$ –dimensionalen Normalverteilungen und  $K$  der Gaußkern aus (2.2)) im Hinblick auf die Untersuchung des Tests für die zusammengesetzte Hypothese aus (2.17). Damit können wir die  $H_0$ –Dichte  $f_0$  o.B.d.A. standardisiert wählen, d.h. für einen  $d$ –dimensionalen Zufallsvektor  $X$  mit Dichte  $f$  gilt  $EX = 0$  und  $EXX^T = I_d$ . Angesichts der Zerlegung

$$\frac{\tilde{T}_{n,h,\text{st}} - \mu(f_0)}{\sqrt{2} \sigma(f_0) h^{d/2}} = \frac{\tilde{T}_{n,h} - \mu(f_0)}{\sqrt{2} \sigma(f_0) h^{d/2}} + \frac{\tilde{T}_{n,1}}{\sqrt{2} \sigma(f_0) h^{d/2}} + \frac{\tilde{T}_{n,2}}{\sqrt{2} \sigma(f_0) h^{d/2}} \quad (2.58)$$

mit den Termen  $\tilde{T}_{n,1}$  und  $\tilde{T}_{n,2}$  aus den Gleichungen (2.35) bzw. (2.36) auf Seite 91 ist die Asymptotik von  $\tilde{T}_{n,1}/h^{d/2}$  und  $\tilde{T}_{n,2}/h^{d/2}$  unter der Folge  $f_n$  von alternativen Dichten zu bestimmen.

Die exakte Asymptotik dieser beiden Größen erweist sich als sehr aufwendig, da hierfür der gesamte  $H_0$ –Fall zu betrachten ist, d.h. Kapitel 2.3 sowie die ersten beiden technischen Lemmata in Abschnitt 2.6 (Lemma 53 und 54), und außerdem zwei Zusatzterme, die von den lokalen Alternativen herrühren. Deshalb beschränken wir uns auf Lokalisierungsraten  $N_n = n^{-r}$ , wobei  $r$

einem „kleinen“ Bereich links von  $1/2$  entstammt, d.h.  $1/2 - \varepsilon < r \leq 1/2$  ( $\varepsilon > 0$  hinreichend klein), und die zugehörige Argumentation verläuft auf einer heuristischen Ebene. Da nach Theorem 45 die Bowman–Foster–Statistik  $\tilde{T}_{n,h}$  ohne geschätzte Parameter lokale Alternativen mit einer derartigen Lokalisierungsrate  $N_n$  nicht mehr von der Hypothese  $H_0$  unterscheiden kann, ist zu vermuten, daß dies auch mit geschätzten Parametern so bleibt.

Den Abschluß dieses Abschnittes bildet ein rigoroser Beweis, daß die Bowman–Foster–Statistik  $\tilde{T}_{n,h,st}$  mit geschätzten Parametern lokale Alternativen, die mit der Rate  $n^{-1/2}$  gegen die Hypothese konvergieren, nicht erkennt. Dieser beruht auf einem anderen Beweisprinzip, nämlich der in Witting und Müller-Funk (1995), Kapitel 6.3, geschilderten Theorie benachbarter Alternativen (vgl. Anhang E).

### Heuristik

Zunächst beschäftigen wir uns mit der Evaluierung eines Bereiches  $1/2 - \varepsilon < r \leq 1/2$ , so daß für lokale Alternativen mit Lokalisierungsrate  $N_n = n^{-r}$  die Konvergenz  $\tilde{T}_{n,1}/h^{d/2} \xrightarrow{P} 0$  bzw.  $\tilde{T}_{n,2}/h^{d/2} \xrightarrow{P} 0$  gilt. Wie im  $H_0$ -Fall ist hierzu die stochastische Konvergenz gegen 0 der Größen

$$\begin{aligned} \tilde{T}_{n,3} &:= \frac{h^{d/2}}{nh^{2d}} \sum_{i,j=1}^n \int dK_h(x, X_{ni}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \cdot dK_h(x, X_{nj}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) dx, \\ \tilde{T}_{n,4} &:= \frac{h^{d/2}}{nh^{2d}} \sum_{i,j=1}^n \int dK_h(x, X_{ni}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) \cdot R_1(x, X_{nj}) dx, \\ \tilde{T}_{n,5} &:= \frac{h^{d/2}}{nh^{2d}} \sum_{i,j=1}^n \int R_1(x, X_{ni}) R_1(x, X_{nj}) dx, \\ \tilde{T}_{n,6} &:= \frac{1}{h^{d/2}} \sum_{i=1}^n \int dK_h(x, X_{ni}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) [\hat{f}_n(x) - e_h(x, f_n)] dx, \\ \tilde{T}_{n,7} &:= \frac{1}{h^{d/2}} \sum_{i=1}^n \int d^2 K_h(x, X_{ni}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) [\hat{f}_n(x) - e_h(x, f_n)] dx, \\ \tilde{T}_{n,8} &:= \frac{1}{h^{d/2}} \sum_{i=1}^n \int R_2(x, X_{ni}) [\hat{f}_n(x) - e_h(x, f_0)] dx, \\ \tilde{T}_{n,9} &:= \frac{1}{h^{d/2}} \sum_{i=1}^n \int dK_h(x, X_{ni}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) [e_h(x, f_n) - e_h(x, f_0)] dx, \\ \tilde{T}_{n,10} &:= \frac{1}{h^{d/2}} \sum_{i=1}^n \int d^2 K_h(x, X_{ni}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) [e_h(x, f_n) - e_h(x, f_0)] dx \end{aligned}$$

nachzuweisen. Man beachte, daß unter den lokalen Alternativen  $f_n$  die zwei zusätzlichen Terme  $\tilde{T}_{n,9}$  und  $\tilde{T}_{n,10}$  benötigt werden, da  $\tilde{T}_{n,6}$  und  $\tilde{T}_{n,7}$  eigentlich mit  $[\hat{f}_n(x) - e_h(x, f_0)]$  anstelle von  $[\hat{f}_n(x) - e_h(x, f_n)]$  zentriert waren, was durch  $\tilde{T}_{n,9}$  und  $\tilde{T}_{n,10}$  ausgeglichen wird.

Mit dem Spaltenvektor  $\widetilde{M}_1$  und der Matrix  $\widetilde{M}_2$ , die definiert sind durch

$$\widetilde{M}_1 := \int x \mathfrak{w}(x) dx, \quad \widetilde{M}_2 := \int xx^T \mathfrak{w}(x) dx$$

berechnen sich aufgrund der Standardisierung von  $f_0$  die ersten beiden Momente von  $X_{n1}$  unter der Dichte  $f_n$  zu

$$EX_{n1} = \int x f_n(x) dx = N_n \widetilde{M}_1, \quad E[X_{n1} X_{n1}^T] = \int xx^T f_n(x) dx = I_d + N_n \widetilde{M}_2,$$

und für die Kovarianzmatrix gilt

$$\text{Cov}(X_{n1}) = E[X_{n1} X_{n1}^T] - EX_{n1} EX_{n1}^T = I_d + N_n \widetilde{M}_2 - N_n^2 \widetilde{M}_1 \widetilde{M}_1^T.$$

Da nach Voraussetzung die Momente vierter Ordnung existieren, gilt für  $l = 1, 2, 3, 4$  wegen

$$E \|X_{n1}\|^l = \int \|x\|^l f_0(x) dx + N_n \int \|x\|^l \mathfrak{w}(x) dx \xrightarrow{n \rightarrow \infty} \int \|x\|^l f_0(x) dx$$

nach der Markov-Ungleichung die Beziehung

$$\frac{1}{n} \sum_{i=1}^n \|X_{ni}\|^l = O_P(1).$$

Ferner gelten für das Stichprobenmittel und die empirische Kovarianzmatrix

$$\begin{aligned} \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_{ni} = \frac{1}{n} \sum_{i=1}^n (X_{ni} - N_n \widetilde{M}_1) + N_n \widetilde{M}_1 \xrightarrow{P} 0, \\ S_n &= \frac{1}{n} \sum_{i=1}^n X_{ni} X_{ni}^T - \bar{X}_n \bar{X}_n^T = \frac{1}{n} \sum_{i=1}^n (X_{ni} X_{ni}^T - I_d - N_n \widetilde{M}_2) + I_d + N_n \widetilde{M}_2 - \bar{X}_n \bar{X}_n^T \\ &\xrightarrow{P} I_d. \end{aligned}$$

Beachtet man, daß wir nur Lokalisierungsraten mit  $N_n = n^{-r}$  mit  $1/2 - \varepsilon < r \leq 1/2$  betrachten, so folgt nach dem multivariaten zentralen Grenzwertsatz von Lindberg-Feller

$$\begin{aligned} \frac{1}{N_n} \bar{X}_n &= \frac{1}{\sqrt{n} N_n} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_{ni} - N_n \widetilde{M}_1) + \widetilde{M}_1 = O_P(1), \\ \frac{1}{N_n} (S_n - I_d) &= \frac{1}{\sqrt{n} N_n} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_{ni} X_{ni}^T - I_d - N_n \widetilde{M}_2) + \widetilde{M}_2 - N_n \left( \frac{1}{N_n} \bar{X}_n \right) \left( \frac{1}{N_n} \bar{X}_n \right)^T \\ &= O_P(1). \end{aligned}$$

Hier werden die Schwierigkeiten deutlich, die bei einer Verallgemeinerung des  $H_0$ -Falles auf lokale Alternativen  $f_n$  mit Lokalisierungsrate  $N_n$  zu Tage treten: Man kann nicht mehr von der Straffheit von  $\sqrt{n} \bar{X}_n$  bzw.  $\sqrt{n} (S_n - I_d)$  ausgehen, sondern nur noch von der Straffheit der Terme  $N_n^{-1} \bar{X}_n$  bzw.  $N_n^{-1} (S_n - I_d)$ , was für  $N_n = n^{-r}$  mit  $r < 1/2$  eine Abschwächung

darstellt. Dies beeinflußt nicht nur die eigentliche Behandlung der Terme  $\tilde{T}_{n,3}$  bis  $\tilde{T}_{n,10}$ , sondern man muß bereits die Abschätzungen der Restglieder  $R_1(x, X_{ni})$  und  $R_2(x, X_{ni})$  aus Lemma 28 auf Seite 89 revidieren: Setzt man wie früher  $\tilde{\vartheta}_n := \vartheta_0 + t(\hat{\vartheta}_n - \vartheta_0)$ , so wird man anstelle der dortigen Abschätzungen nur noch Resultate der Form

$$\begin{aligned}
& |R_1(x, X_{ni})| \\
& \leq \frac{1}{h^2} O_P(N_n^2) \left( \sum_{l=0}^2 \|X_{ni}\|^l \right) \int_0^1 (1-t) \left\{ \left\| \nabla^2 K(\text{st}(x, X_{ni}, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_{ni}, \tilde{\vartheta}_n)) \right\| \right\} dt, \\
& |R_2(x, X_{ni})| \\
& \leq \frac{1}{h^3} O_P(N_n^3) \left( \sum_{l=0}^3 \|X_{ni}\|^l \right) \\
& \quad \int_0^1 (1-t)^2 \left\{ \left\| \nabla^3 K(\text{st}(x, X_{ni}, \tilde{\vartheta}_n)) \right\| + \left\| \nabla^2 K(\text{st}(x, X_{ni}, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_{ni}, \tilde{\vartheta}_n)) \right\| \right\} dt
\end{aligned}$$

erwarten können. Damit lassen sich die Terme  $\tilde{T}_{n,3}$ ,  $\tilde{T}_{n,5}$  und  $\tilde{T}_{n,6}$  analog zum  $H_0$ -Fall zerlegen bzw. abschätzen gemäß

$$\begin{aligned}
\tilde{T}_{n,3} &= N_n^2 n h^{d/2} \cdot \\
& \left\{ \sum_{r,s=1}^d \frac{\bar{X}_{nr} \bar{X}_{ns}}{N_n} \frac{1}{n^2 h^{2d+2}} \sum_{i,j=1}^n \int \nabla K \left( \frac{x - X_{ni}}{h} \right)_r \nabla K \left( \frac{x - X_{nj}}{h} \right)_s dx \right. \\
& + \sum_{q,r,s=1}^d \frac{\bar{X}_{nq} (S_n - I_d)_{rs}}{N_n} \frac{1}{n^2 h^{2d+2}} \sum_{i,j=1}^n X_{njs} \int \nabla K \left( \frac{x - X_{ni}}{h} \right)_q \nabla K \left( \frac{x - X_{nj}}{h} \right)_r dx \\
& \left. + \sum_{l,q,r,s=1}^d \frac{1}{4} \frac{(S_n - I_d)_{lq} (S_n - I_d)_{rs}}{N_n} \frac{1}{n^2 h^{2d+2}} \sum_{i,j=1}^n X_{niq} X_{njs} \int \nabla K \left( \frac{x - X_{ni}}{h} \right)_l \nabla K \left( \frac{x - X_{nj}}{h} \right)_r dx \right\}
\end{aligned}$$

und

$$\begin{aligned}
|\tilde{T}_{n,5}| &\leq \frac{h^{d/2} n N_n^4}{h^{d+4}} \cdot O_P(1) \frac{1}{n^2 h^d} \sum_{i,j=1}^n \left( \sum_{l=0}^2 \|X_{ni}\|^l \right) \left( \sum_{q=0}^2 \|X_{nj}\|^q \right) \\
& \quad \int \left[ \int_0^1 (1-t) \left\{ \left\| \nabla^2 K(\text{st}(x, X_{ni}, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_{ni}, \tilde{\vartheta}_n)) \right\| \right\} dt \right] \\
& \quad \left[ \int_0^1 (1-s) \left\{ \left\| \nabla^2 K(\text{st}(x, X_{nj}, \tilde{\vartheta}_n)) \right\| + \left\| \nabla K(\text{st}(x, X_{nj}, \tilde{\vartheta}_n)) \right\| \right\} ds \right] dx
\end{aligned}$$

sowie

$$\begin{aligned}
\tilde{T}_{n,6} &= \sqrt{n} N_n h^{1/2} \cdot \\
& \left\{ \sum_{r=1}^d \frac{\bar{X}_{nr}}{N_n} \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i,j=1}^n \int \nabla K \left( \frac{x - X_{ni}}{h} \right)_r \left[ K \left( \frac{x - X_{nj}}{h} \right) - \text{EK} \left( \frac{x - X_{n1}}{h} \right) \right] dx \right. \\
& \left. + \sum_{r,s=1}^d \frac{1}{2} \frac{(S_n - I_d)_{rs}}{N_n} \right.
\end{aligned}$$

$$\left. \frac{1}{n^{3/2}h^{(3d+3)/2}} \sum_{i,j=1}^n X_{nis} \int \nabla K \left( \frac{x - X_{ni}}{h} \right)_r \left[ K \left( \frac{x - X_{nj}}{h} \right) - \mathbb{E}K \left( \frac{x - X_{n1}}{h} \right) \right] dx \right\}.$$

$\tilde{T}_{n,4}$  läßt sich wie im  $H_0$ -Fall mit der Cauchy–Schwarz–Ungleichung durch das Produkt der Wurzeln von  $\tilde{T}_{n,3}$  und  $\tilde{T}_{n,4}$  abschätzen, die Zerlegung von  $\tilde{T}_{n,7}$  verläuft analog zu der von  $\tilde{T}_{n,6}$  und die Abschätzung von  $\tilde{T}_{n,8}$  analog zu der von  $\tilde{T}_{n,5}$ , so daß wir unsere Betrachtungen exemplarisch auf die drei explizit aufgeführten Fälle beschränken. Ein Kommentar zu  $\tilde{T}_{n,9}$  und  $\tilde{T}_{n,10}$  erfolgt weiter unten.

Verallgemeinert man die Lemmata 53 auf Seite 125 und 54 auf Seite 131 auf den Fall lokaler Alternativen  $f_n$  und hat damit die Straffheit der Summen in geschweiften Klammern in den obigen Zerlegungen von  $\tilde{T}_{n,3}$  und  $\tilde{T}_{n,6}$  gesichert, und berücksichtigt man außerdem, daß die Rechnung für  $\tilde{T}_{n,5}$  von Seite 96 ihre Gültigkeit behält, so sind die Bedingungen

$$N_n^2 n h^{d/2} \rightarrow 0, \quad \frac{h^{d/2} n N_n^4}{h^{d+4}} \rightarrow 0, \quad \sqrt{n} N_n h^{1/2} \rightarrow 0 \quad (2.59)$$

hinreichend für die asymptotische Vernachlässigbarkeit von  $\tilde{T}_{n,3}$ ,  $\tilde{T}_{n,5}$  und  $\tilde{T}_{n,6}$ . Dabei beachte man, daß nur die zweite Bedingung in (2.59) auf einer Abschätzung, nämlich der von  $\tilde{T}_{n,5}$ , beruht, daß die erste und dritte Bedingung jedoch exakt sind. Denn weder in den Berechnungen von  $\tilde{T}_{n,3}$  und  $\tilde{T}_{n,6}$  noch in den Lemmata 53 und 54 wurde „etwas verschenkt“.

An der dritten Bedingung in (2.59) sieht man sofort, daß für die „interessanten“ Lokalisierungsraten  $N_n = n^{-r}$  der Bowman–Foster–Statistik  $\tilde{T}_{n,h}$  ohne geschätzte Parameter aus Abschnitt 2.5.1, also für

$$\begin{aligned} r &= \frac{1}{2} - \frac{d}{2(d+4)}, & \text{d.h.} & \quad \sqrt{n} N_n h^{d/2} \rightarrow \text{const} \quad \text{und} \\ r &= \frac{1}{2} - \frac{d}{4(d+4)}, & \text{d.h.} & \quad \sqrt{n} N_n h^{d/4} \rightarrow \text{const} \end{aligned}$$

mit  $\text{const} \in (0, \infty)$ , der Term  $\tilde{T}_{n,6}$  in Dimensionen  $d \geq 3$  nicht mehr kontrollierbar ist. Man kann also nur in einem viel kleineren Bereich  $1/2 - \varepsilon < r \leq 1/2$  auf seine asymptotische Vernachlässigbarkeit hoffen. Eine Möglichkeit für einen Bereich, bei dem alle oben beschriebenen Rechnungen funktionieren, ist die Wahl

$$\frac{1}{2} - \frac{1}{8(d+4)} < r \leq \frac{1}{2}, \quad \text{d.h.} \quad \sqrt{n} N_n h^{1/8} \rightarrow 0.$$

Unter dieser Voraussetzung stellt sich bei exakter Rechnung auch der Term

$$\begin{aligned} \tilde{T}_{n,9} &= \frac{N_n}{h^{d/2}} \sum_{i=1}^n \int dK_h(x, X_{ni}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0) e_h(x, \mathbf{w}) dx \\ &= n N_n^2 h^{d/2} \left\{ \sum_{r=1}^d \frac{\bar{X}_{nr}}{N_n} \frac{1}{n h^{d+1}} \sum_{i=1}^n \int \nabla K \left( \frac{x - X_{ni}}{h} \right)_r e_h(x, \mathbf{w}) dx \right. \end{aligned}$$

$$+ \frac{1}{2} \sum_{r,s=1}^d \frac{(S_n - I_d)_{rs}}{N_n} \frac{1}{nh^{d+1}} \sum_{i=1}^n X_{nis} \int \nabla K \left( \frac{x - X_{ni}}{h} \right)_r e_h(x, \mathfrak{w}) dx \Big\}$$

als asymptotisch vernachlässigbar heraus, denn  $nN_n^2 h^{d/2}$  strebt gegen 0, und die Summen in geschweiften Klammern sind straff. Gleiches gilt für

$$\tilde{T}_{n,10} = \frac{N_n}{h^{d/2}} \sum_{i=1}^n \int d^2 K_h(x, X_{ni}, \vartheta_0) \circ (\hat{\vartheta}_n - \vartheta_0, \hat{\vartheta}_n - \vartheta_0) e_h(x, \mathfrak{w}) dx.$$

Da, wie bereits diskutiert, die Rechnungen an dieser Stelle nicht im Detail durchgeführt werden sollen, formulieren wir das Resultat der vorangegangenen Betrachtungen als Vermutung.

**Vermutung 47** Unter den Voraussetzungen 1\* und 2\* und einer Lokalisierungsrate der Gestalt  $N_n = n^{-r}$  mit

$$\frac{1}{2} - \frac{1}{8(d+4)} < r \leq \frac{1}{2}$$

gilt für die Bowman–Foster–Statistik  $\tilde{T}_{n,h,st}$  mit geschätzten Parametern unter der standardisierten  $H_0$ -Dichte  $f_0$

$$\frac{\tilde{T}_{n,h,st} - \mu(f_0)}{\sqrt{2} \sigma(f_0) h^{d/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

**Bemerkung 48** In der nachfolgenden Tabelle sind die „interessanten“ Werte der Exponenten  $r$  von  $N_n = n^{-r}$ , d.h. die Werte aus Abschnitt 2.5.1, bei denen nichttriviale Grenzverteilungen auftreten, sowie die Grenze aus Vermutung 47 für einige Dimensionen  $d$  aufgeführt.

$d$	$\frac{1}{2} - \frac{d}{2(d+4)}$	$\frac{1}{2} - \frac{d}{4(d+4)}$	$\frac{1}{2} - \frac{1}{8(d+4)}$
1	$\frac{2}{5} = 0.4$	$\frac{9}{20} = 0.45$	$\frac{19}{40} = 0.475$
2	$\frac{1}{3} = 0.\bar{3}$	$\frac{5}{12} = 0.41\bar{6}$	$\frac{23}{48} = 0.4791\bar{6}$
3	$\frac{2}{7} \approx 0.28571$	$\frac{11}{28} \approx 0.39286$	$\frac{27}{56} \approx 0.48214$
4	$\frac{1}{4} = 0.25$	$\frac{3}{8} = 0.375$	$\frac{31}{64} = 0.484375$
5	$\frac{2}{9} = 0.\bar{2}$	$\frac{13}{36} = 0.36\bar{1}$	$\frac{35}{72} = 0.486\bar{1}$
10	$\frac{1}{7} \approx 0.14286$	$\frac{9}{28} \approx 0.32143$	$\frac{55}{112} \approx 0.49107$
20	$\frac{1}{12} = 0.8\bar{3}$	$\frac{7}{24} = 0.291\bar{6}$	$\frac{95}{192} \approx 0.49479$
50	$\frac{1}{27} = 0.0\bar{37}$	$\frac{29}{108} \approx 0.26852$	$\frac{215}{432} \approx 0.49769$
100	$\frac{1}{52} \approx 0.01923$	$\frac{27}{104} \approx 0.25962$	$\frac{415}{832} \approx 0.49880$
$\infty$	0	$\frac{1}{4} = 0.25$	$\frac{1}{2}$

### Exakte Behandlung der $n^{-1/2}$ -Rate

Wie bereits zu Beginn dieses Abschnittes angedeutet, folgt als letzter Punkt ein exakter Beweis, daß die Bowman–Foster–Statistik  $\tilde{T}_{n,h,\text{st}}$  mit geschätzten Parametern lokale Alternativen, die mit der Rate  $n^{-1/2}$  gegen die Hypothese konvergieren, nicht erkennt. Hierzu ist angesichts der Zerlegung (2.58) auf Seite 116 die stochastische Konvergenz gegen 0 der Restterme  $\tilde{T}_{n,1}/h^{d/2}$  und  $\tilde{T}_{n,2}/h^{d/2}$  unter der Folge  $f_n$  von lokalen Alternativen zu zeigen. Dies erfolgt durch Anwenden der in Witting und Müller-Funk (1995), Kapitel 6.3, geschilderten Theorie benachbarter Alternativen (vgl. Anhang E). An dieser Stelle sei darauf hingewiesen, daß die Definition „benachbarter Alternativen“ nach Witting und Müller-Funk (1995) (vgl. Definition E.1 in Anhang E) restriktiver ist als die in dem aktuellen Kapitel 2.5 der vorliegenden Arbeit zugrundegelegte Definition „lokaler Alternativen“ nach Gleichung (2.50). Dieser Zusammenhang wird in Bemerkung 50 näher erläutert.

**Satz 49** Unter benachbarten Alternativen  $f_n$  mit standardisierter  $H_0$ -Dichte  $f_0$  und der Lokalisierungsrates  $N_n = n^{-1/2}$  gelten unter den Voraussetzungen 1\* und 2\* für die Restterme  $\tilde{T}_{n,1}$  und  $\tilde{T}_{n,2}$  die Grenzwertaussagen

$$\frac{\tilde{T}_{n,1}}{h^{d/2}} \xrightarrow{P} 0 \quad \text{und} \quad \frac{\tilde{T}_{n,2}}{h^{d/2}} \xrightarrow{P} 0 .$$

BEWEIS: Bezeichnet  $\lambda^d$  das Lebesgue-Maß auf  $\mathbb{R}^d$ , so definieren wir für  $j = 1, \dots, n$  die Wahrscheinlichkeitsmaße  $u_{nj}$  und  $v_{nj}$  auf dem Raum  $(\mathbb{R}^d, \mathfrak{B}^d)$  durch

$$\frac{du_{nj}}{d\lambda^d} := f_0 \quad \text{und} \quad \frac{dv_{nj}}{d\lambda^d} := f_n$$

sowie die Produktmaße  $u_n$  und  $v_n$  auf dem Produktraum  $\bigotimes_{j=1}^n (\mathbb{R}^d, \mathfrak{B}^d)$  durch

$$u_n := \bigotimes_{j=1}^n u_{nj} \quad \text{und} \quad v_n := \bigotimes_{j=1}^n v_{nj} .$$

Das Wahrscheinlichkeitsmaß  $u_n$  ist gerade die Verteilung einer Folge von unabhängigen und identisch verteilten Zufallsvektoren  $X_1, \dots, X_n$  unter der Hypothese  $H_0$ , und  $v_n$  die Verteilung einer identisch verteilten Zeile  $X_{n1}, \dots, X_{nn}$  eines Dreiecksschemas unabhängiger Zufallsvektoren unter der lokalen Alternative  $f_n$ .

Aufgrund der Forderung (2.51) auf Seite 109 ist der Dichtequotient

$$L_n := \frac{dv_n}{du_n}, \quad \text{d.h.} \quad L_n(x_1, \dots, x_n) = \prod_{j=1}^n \left( 1 + \frac{1}{\sqrt{n}} \frac{\mathfrak{w}(x_j)}{f_0(x_j)} \right)$$

$(x_1, \dots, x_n \in \mathbb{R}^d)$ , wohldefiniert und endlich. Ferner gelten aufgrund der Voraussetzung (2.50) und der Beschränktheit von  $\mathfrak{w}/f_0$  für  $j = 1, \dots, n$

$$E_{u_n} \left[ \frac{\mathfrak{w}(X_{nj})}{f_0(X_{nj})} \right] = \int \frac{\mathfrak{w}(x)}{f_0(x)} f_0(x) dx = 0, \quad (2.60)$$

$$\tilde{\kappa}^2 := E_{u_n} \left[ \frac{\mathfrak{w}(X_{n_j})^2}{f_0(X_{n_j})^2} \right] = \int \frac{\mathfrak{w}(x)^2}{f_0(x)^2} f_0(x) dx < \infty. \quad (2.61)$$

Mit der Taylor-Entwicklung  $\log(1+t) = t - \frac{1}{2}t^2 + \frac{1}{3} \frac{1}{(1+\xi)^3} t^3$ ,  $|t| < 1$ ,  $\xi$  zwischen 0 und  $t$ , ergibt sich für den Logarithmus des Dichtequotienten

$$\begin{aligned} \log L_n(X_{n_1}, \dots, X_{n_n}) &= \sum_{j=1}^n \log \left( 1 + \frac{1}{\sqrt{n}} \frac{\mathfrak{w}(X_{n_j})}{f_0(X_{n_j})} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{\mathfrak{w}(X_{n_j})}{f_0(X_{n_j})} - \frac{1}{2n} \sum_{j=1}^n \frac{\mathfrak{w}(X_{n_j})^2}{f_0(X_{n_j})^2} + \frac{1}{3n^{3/2}} \sum_{j=1}^n \frac{1}{(1+\xi_{n_j})^3} \frac{\mathfrak{w}(X_{n_j})^3}{f_0(X_{n_j})^3} \end{aligned}$$

mit  $\xi_{n_j}(\omega)$  zwischen 0 und  $\frac{1}{\sqrt{n}} \frac{\mathfrak{w}(X_{n_j}(\omega))}{f_0(X_{n_j}(\omega))}$ . Aufgrund der Beschränktheit von  $\mathfrak{w}/f_0$  ist der Restterm von der Größenordnung  $o_{u_n}(1)$ , und angesichts von (2.60) und (2.61) folgt in Verbindung mit dem Zentralen Grenzwertsatz von Lindeberg-Lévy und dem Gesetz großer Zahlen

$$\begin{aligned} \log L_n(X_{n_1}, \dots, X_{n_n}) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{\mathfrak{w}(X_{n_j})}{f_0(X_{n_j})} - \frac{1}{2n} \sum_{j=1}^n \frac{\mathfrak{w}(X_{n_j})^2}{f_0(X_{n_j})^2} + o_{u_n}(1) \\ &\xrightarrow{\mathcal{D}} \mathcal{N}\left(-\frac{1}{2}\tilde{\kappa}^2, \tilde{\kappa}^2\right) \quad \text{unter } u_n. \end{aligned}$$

Daraus ergibt sich nach dem ersten Lemma von Le Cam (zitiert in Satz E.5 auf Seite 187 in Anhang E) die wechselseitige Benachbarkeit von  $u_n$  und  $v_n$ .

Nach dem Beweis von Theorem 31 auf Seite 93 im  $H_0$ -Fall gelten für die Restterme  $\tilde{T}_{n,1}$  und  $\tilde{T}_{n,2}$  die stochastischen Konvergenzaussagen

$$\frac{\tilde{T}_{n,1}}{h^{d/2}} \xrightarrow{u_n} 0 \quad \text{und} \quad \frac{\tilde{T}_{n,2}}{h^{d/2}} \xrightarrow{u_n} 0,$$

und aufgrund der wechselseitigen Benachbarkeit der Wahrscheinlichkeitsmaße  $u_n$  und  $v_n$  erhalten wir nach Satz E.2 aus Anhang E die Behauptung:

$$\frac{\tilde{T}_{n,1}}{h^{d/2}} \xrightarrow{v_n} 0 \quad \text{und} \quad \frac{\tilde{T}_{n,2}}{h^{d/2}} \xrightarrow{v_n} 0.$$

□

**Bemerkung 50** Es bezeichne  $\lambda^d$  das Lebesgue-Maß auf  $\mathbb{R}^d$ ,  $f_n(x) = f_0(x) + N_n \mathfrak{w}(x)$  die Dichte der lokalen Alternative und

$$u_n := \bigotimes_{j=1}^n u_{n_j}, \quad v_n := \bigotimes_{j=1}^n v_{n_j} \quad \text{mit} \quad \frac{du_{n_j}}{d\lambda^d} := f_0, \quad \frac{dv_{n_j}}{d\lambda^d} := f_n$$

die zu den Dichten  $f_0$  und  $f_n$  gehörigen Produktmaße. Unter der Voraussetzung 1\* ist  $v_n$  genau dann benachbart zu  $u_n$  im Sinne von Witting und Müller-Funk (1995) (vgl. Definition E.1), wenn für die Lokalisierungsrate gilt:  $N_n = O(n^{-1/2})$ .

Für Raten  $N_n = n^{-r}$  mit  $1/2 - \varepsilon < r < 1/2$  ist die Theorie benachbarter Alternativen aus Witting und Müller-Funk (1995) also nicht anwendbar, d.h. es ergibt keinen Sinn, eine Verallgemeinerung des Beweises von Satz 49 auf diese Situation anzustreben.

**BEWEIS:** Für den Hellinger–Abstand der Wahrscheinlichkeitsmaße  $u_{nj}$  und  $v_{nj}$  gilt nach Gleichung (E.1) auf Seite 185 und der Entwicklung  $\sqrt{1+t} = 1 + \frac{1}{2}t - \frac{1}{8}t^2 + \frac{1}{16} \frac{1}{(1+\xi)^{5/2}} t^3$ ,  $|t| < 1$ ,  $\xi$  zwischen 0 und  $t$ ,

$$\begin{aligned} H^2(u_{nj}, v_{nj}) &= 1 - \int \sqrt{f_n(x)f_0(x)} dx = 1 - \int \sqrt{1 + N_n \frac{\mathfrak{w}(x)}{f_0(x)}} f_0(x) dx \\ &= 1 - \int f_0(x) dx - \frac{1}{2} N_n \int \mathfrak{w}(x) dx + \frac{1}{8} N_n^2 \int \frac{\mathfrak{w}(x)^2}{f_0(x)^2} f_0(x) dx \\ &\quad - \frac{1}{16} N_n^3 \int \frac{1}{(1+\xi)^{5/2}} \frac{\mathfrak{w}(x)^3}{f_0(x)^3} f_0(x) dx \end{aligned}$$

mit  $\xi$  zwischen 0 und  $N_n \frac{\mathfrak{w}(x)}{f_0(x)}$ . Aufgrund der Beschränktheit von  $\mathfrak{w}/f_0$  ist der Restterm von der Größenordnung  $o(N_n^2)$ , und mit  $\tilde{\kappa}^2$  aus Gleichung (2.61) ergibt sich anhand der Voraussetzungen

$$H^2(u_{nj}, v_{nj}) = \frac{1}{8} \tilde{\kappa}^2 N_n^2 + o(N_n^2),$$

also

$$\sum_{j=1}^n H^2(u_{nj}, v_{nj}) = \frac{\tilde{\kappa}^2}{8} n N_n^2 + o(n N_n^2).$$

Im Fall  $N_n = n^{-1/2}$  haben wir bereits im Beweis von Satz 49 gesehen, daß  $u_n$  und  $v_n$  wechselseitig benachbart sind. Im Fall  $N_n = o(n^{-1/2})$  gilt  $\sum_{j=1}^n H^2(u_{nj}, v_{nj}) = o(1)$ , und die wechselseitige Benachbarkeit von  $u_n$  und  $v_n$  folgt nach Satz E.3 aus Anhang E. Für  $N_n = n^{-r}$  mit  $r < 1/2$  ist  $\sum_{j=1}^n H^2(u_{nj}, v_{nj})$  nicht beschränkt, und nach dem Satz von Oosterhoff–van Zwet (Satz E.4 in Anhang E) kann  $v_n$  nicht zu  $u_n$  benachbart sein.  $\square$

Angesichts der Zerlegung (2.58) auf Seite 116 erhalten wir aus Satz 49 das folgende Resultat für die Bowman–Foster–Statistik  $\tilde{T}_{n,h,st}$  mit geschätzten Parametern.

**Theorem 51** Unter benachbarten Alternativen  $f_n$  mit standardisierter  $H_0$ –Dichte  $f_0$  und der Lokalisierungsrate  $N_n = n^{-1/2}$  gilt unter den Voraussetzungen 1\* und 2\* die Konvergenzaussage

$$\frac{\tilde{T}_{n,h,st} - \mu(f_0)}{\sqrt{2} \sigma(f_0) h^{d/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Die Bowman–Foster–Statistik mit geschätzten Parametern weist also unter lokalen Alternativen  $f_n$  mit Lokalisierungsrate  $N_n = n^{-1/2}$  exakt dasselbe asymptotische Verhalten auf wie unter der Hypothese  $H_0$ , d.h. sie ist nicht in der Lage, derartige Alternativen von  $H_0$  zu unterscheiden. Nach Bemerkung E.10 auf Seite 188 besitzt der auf der Bowman–Foster–Statistik basierende asymptotische Test aus Bemerkung 33 auf Seite 100 die Pitman–Effizienz 0 gegenüber jedem asymptotischen Test, der unter der Folge  $(f_n)$  gegen eine im Vergleich zu seiner  $H_0$ –Asymptotik verschobene Normalverteilung konvergiert. Dies trifft im Fall des Testens auf multivariate Normalverteilung nach Henze und Wagner (1997) insbesondere auf den BHEP–Test aus Kapitel 1 zu. Damit sind wir bei dem überraschenden Resultat angelangt, daß der BHEP–Test bei variablem Glättungsparameter unter benachbarten Alternativen eine grundsätzlich andere Asymptotik aufweist als bei festem Glättungsparameter.

**Bemerkung 52** Das Resultat von Theorem 51 ist vorwiegend von theoretischem Interesse. In einem konkreten Vergleich (z.B. einer Simulationsstudie) schneidet der auf der Bowman–Foster–Statistik basierende Test auf multivariate Normalverteilung bei festem Stichprobenumfang  $n$  ebenso gut ab wie der BHEP–Test aus Kapitel 1. Denn der einzige Unterschied zwischen den beiden besteht darin, daß der Gewichtsparameter  $\beta$  der BHEP–Statistik fest gewählt ist, während er bei der Bowman–Foster–Statistik mit dem Stichprobenumfang  $n$  gegen  $\infty$  strebt. Da dies jedoch äußerst langsam geschieht ( $\beta = \text{const} \cdot n^{\frac{1}{d+4}}$  gemäß Gleichung (2.9) auf Seite 77), ist bei üblichen Stichprobengrößen kein Unterschied zum BHEP–Test zu erkennen. Zum Beispiel ist für  $n < 10\,000$  der Glättungsparameter  $\beta := n^{\frac{1}{d+4}}$  kleiner als 6.5 und liegt damit in einem Bereich, wo der BHEP–Test noch ein sehr stabiles Verhalten aufweist. Zum Vergleich sei an die Ergebnisse aus Abschnitt 1.4.7 erinnert, wo darauf hingewiesen wurde, daß der BHEP–Test besonders gut für  $\beta \in [0.8, 3.5]$  funktioniert.

## 2.6. Technische Lemmata

**Lemma 53** Unter den Voraussetzungen 1 und 2 an die Bandbreite  $h$ , den Kern  $K$  und die den Daten  $X_1, \dots, X_n$  zugrundeliegende Dichte  $f$  gilt für jedes feste  $l, q, r, s \in \{1, \dots, d\}$  und  $\eta, \zeta \in \{0, 1\}$

$$\frac{1}{n^2 h^{2d+2}} \sum_{i,j=1}^n X_{il}^\eta X_{jq}^\zeta \int \nabla K \left( \frac{x - X_i}{h} \right)_r \nabla K \left( \frac{x - X_j}{h} \right)_s dx = O_P(1).$$

BEWEIS: Es seien  $l, q, r, s \in \{1, \dots, d\}$  und  $\eta, \zeta \in \{0, 1\}$  fest. Die zu untersuchende Größe läßt sich gemäß

$$\begin{aligned} & \frac{1}{n^2 h^{2d+2}} \sum_{i,j=1}^n X_{il}^\eta X_{jq}^\zeta \int \nabla K \left( \frac{x - X_i}{h} \right)_r \nabla K \left( \frac{x - X_j}{h} \right)_s dx \\ &= \frac{1}{n^2 h^{2d+2}} \sum_{i=1}^n X_{il}^\eta X_{iq}^\zeta \int \nabla K \left( \frac{x - X_i}{h} \right)_r \nabla K \left( \frac{x - X_i}{h} \right)_s dx \end{aligned}$$

$$+ \frac{1}{n^2 h^{2d+2}} \sum_{i \neq j} X_{il}^\eta X_{jq}^\zeta \int \nabla K \left( \frac{x - X_i}{h} \right)_r \nabla K \left( \frac{x - X_j}{h} \right)_s dx$$

zerlegen in einen linearen und einen quadratischen Term. Für den linearen Term gilt mittels der Substitution  $y = (x - X_i)/h$ ,  $dy = h^{-d} dx$ , nach dem starken Gesetz großer Zahlen und nach Voraussetzung 2

$$\begin{aligned} & \frac{1}{n^2 h^{2d+2}} \sum_{i=1}^n X_{il}^\eta X_{iq}^\zeta \int \nabla K \left( \frac{x - X_i}{h} \right)_r \nabla K \left( \frac{x - X_i}{h} \right)_s dx \\ &= \frac{h^2}{n h^{d+4}} \cdot \left( \frac{1}{n} \sum_{i=1}^n X_{il}^\eta X_{iq}^\zeta \right) \left( \int \nabla K(y)_r \nabla K(y)_s dy \right) = o_P(1). \end{aligned}$$

Folglich ist nur der quadratische Term

$$\begin{aligned} U_n &:= \frac{1}{n^2 h^{2d+2}} \sum_{i \neq j} X_{il}^\eta X_{jq}^\zeta \int \nabla K \left( \frac{x - X_i}{h} \right)_r \nabla K \left( \frac{x - X_j}{h} \right)_s dx \\ &= \frac{1}{2} \cdot \frac{1}{n^2 h^{(d+3)/2}} \sum_{i \neq j} H_n(X_i, X_j), \end{aligned}$$

also eine  $U$ -Statistik mit dem symmetrischen, vom Stichprobenumfang  $n$  abhängenden Kern

$$\begin{aligned} H_n(X_i, X_j) &:= \frac{1}{h^{(3d+1)/2}} \left\{ X_{il}^\eta X_{jq}^\zeta \int \nabla K \left( \frac{x - X_i}{h} \right)_r \nabla K \left( \frac{x - X_j}{h} \right)_s dx \right. \\ &\quad \left. + X_{jl}^\eta X_{iq}^\zeta \int \nabla K \left( \frac{x - X_j}{h} \right)_r \nabla K \left( \frac{x - X_i}{h} \right)_s dx \right\} \end{aligned}$$

weiter zu untersuchen. Dies geschieht durch Reduzieren der quadratischen  $U$ -Statistik  $U_n$  auf eine Summe von unabhängigen und identisch verteilten Zufallsgrößen mit Hilfe der Hájek-Projektion, genauer gesagt durch Anwenden von Lemma D.3 auf Seite 181. Hierzu stellen wir fest, daß mit der Cauchy–Schwarz–Ungleichung, der Substitution  $y = (x - X_1)/h$ ,  $dy = h^{-d} dx$  und nach Voraussetzung 2 ( $n h^{d+4} = \text{const}$ ,  $\mathbb{E} \|X_1\|^4 < \infty$  und  $\int \|\nabla K(x)\|^2 dx < \infty$ ) gilt:

$$\begin{aligned} & \mathbb{E} [H_n^2(X_1, X_2)] \\ &= \frac{1}{h^{3d+1}} \mathbb{E} \left[ \left\{ X_{1l}^\eta X_{2q}^\zeta \int \nabla K \left( \frac{x - X_1}{h} \right)_r \nabla K \left( \frac{x - X_2}{h} \right)_s dx \right. \right. \\ &\quad \left. \left. + X_{2l}^\eta X_{1q}^\zeta \int \nabla K \left( \frac{x - X_2}{h} \right)_r \nabla K \left( \frac{x - X_1}{h} \right)_s dx \right\}^2 \right] \\ &\leq \frac{1}{h^{3d+1}} \mathbb{E} \left[ \left\{ X_{1l}^\eta X_{2q}^\zeta \left( \int \nabla K \left( \frac{x - X_1}{h} \right)_r^2 dx \right)^{1/2} \left( \int \nabla K \left( \frac{x - X_2}{h} \right)_s^2 dx \right)^{1/2} \right. \right. \\ &\quad \left. \left. + X_{2l}^\eta X_{1q}^\zeta \left( \int \nabla K \left( \frac{x - X_2}{h} \right)_r^2 dx \right)^{1/2} \left( \int \nabla K \left( \frac{x - X_1}{h} \right)_s^2 dx \right)^{1/2} \right\}^2 \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{h^{d+1}} \mathbb{E} \left[ \left( X_{1l}^\eta X_{2q}^\zeta + X_{2l}^\eta X_{1q}^\zeta \right)^2 \right] \left( \int \nabla K(y)_r^2 dy \right) \left( \int \nabla K(y)_s^2 dy \right) \\
&= o(n) .
\end{aligned}$$

Dies impliziert nach Lemma D.3 mit den Bezeichnungen

$$\theta_n := \mathbb{E} [H_n(X_1, X_2)] , \quad r_n(X_1) := \mathbb{E} [H_n(X_1, X_2) | X_1]$$

die folgende asymptotische Äquivalenz zwischen der Hájek–Projektion

$$\widehat{U}_n := \theta_n + \frac{2}{n} \sum_{i=1}^n (r_n(X_i) - \theta_n)$$

und der auf dem Kern  $H_n$  basierenden  $U$ –Statistik:

$$\frac{1}{n(n-1)} \sum_{i \neq j} H_n(X_i, X_j) = \widehat{U}_n + o_P \left( \frac{1}{\sqrt{n}} \right) .$$

Daraus ergibt sich für  $U_n$  zunächst

$$\begin{aligned}
U_n &= \frac{1}{2} \frac{n-1}{n} \frac{1}{h^{(d+3)/2}} \widehat{U}_n + \frac{1}{2} \frac{n-1}{n} \frac{1}{h^{(d+3)/2}} o_P \left( \frac{1}{\sqrt{n}} \right) \\
&= \frac{1}{2} \frac{n-1}{n} \frac{1}{h^{(d+3)/2}} \theta_n + \frac{n-1}{n} \frac{1}{h^{(d+3)/2}} \frac{1}{n} \sum_{i=1}^n (r_n(X_i) - \theta_n) + o_P \left( \frac{1}{\sqrt{nh^{d+3}}} \right) .
\end{aligned}$$

Im weiteren Verlauf des Beweises werden wir zeigen, daß  $\theta_n$  und  $\mathbb{E} [r_n(X_1)^2]$  das asymptotische Verhalten

$$\theta_n = O(h^{(d+3)/2}) , \quad \mathbb{E} [r_n(X_1)^2] = O(h^{d-1}) \quad (2.62)$$

besitzen, was angesichts von

$$\begin{aligned}
\mathbb{E} \left[ \frac{1}{h^{(d+3)/2}} \frac{1}{n} \sum_{i=1}^n (r_n(X_i) - \theta_n) \right] &= 0 , \\
\text{Var} \left( \frac{1}{h^{(d+3)/2}} \frac{1}{n} \sum_{i=1}^n (r_n(X_i) - \theta_n) \right) &= \frac{1}{h^{d+3}n} \{ \mathbb{E} [r_n(X_1)^2] - \mathbb{E} [r_n(X_1)]^2 \} \\
&\leq \frac{1}{h^{d+3}n} O(h^{d-1}) = O \left( \frac{1}{h^4 n} \right) = o(1)
\end{aligned}$$

das gewünschte Resultat liefert:

$$U_n = \frac{1}{2} \frac{n-1}{n} \frac{1}{h^{(d+3)/2}} O(h^{(d+3)/2}) + \frac{n-1}{n} o_P(1) + o_P \left( \frac{1}{\sqrt{nh^{d+3}}} \right) = O_P(1) .$$

Es verbleibt also, die beiden Aussagen in (2.62) zu zeigen.

(i) BEWEIS VON  $\theta_n = O(h^{(d+3)/2})$ :

Mit den Substitutionen  $u = (x - z)/h$ ,  $du = h^{-d}dz$  und  $v = (x - y)/h$ ,  $dv = h^{-d}dy$  ist

$$\begin{aligned}\theta_n &= \frac{2}{h^{(3d+1)/2}} \mathbb{E} \left[ X_{1l}^\eta X_{2q}^\zeta \int \nabla K \left( \frac{x - X_1}{h} \right)_r \nabla K \left( \frac{x - X_2}{h} \right)_s dx \right] \\ &= \frac{2}{h^{(3d+1)/2}} \iiint z_l^\eta y_q^\zeta \nabla K \left( \frac{x - z}{h} \right)_r \nabla K \left( \frac{x - y}{h} \right)_s f(z) f(y) dx dz dy \\ &= 2 h^{(d-1)/2} \iiint (x - hu)_l^\eta (x - hv)_q^\zeta \nabla K(u)_r \nabla K(v)_s f(x - hu) f(x - hv) dx dudv.\end{aligned}$$

Setzt man nun für  $f(x - hu)$  die Taylor–Entwicklung

$$f(x - hu) = f(x) - h \nabla f(x) u + h^2 u^T \left( \int_0^1 (1 - t) \nabla^2 f(x - thu) dt \right) u$$

aus Bemerkung 22 auf Seite 85 ein, so gilt nach Gleichung (2.25) für das Integral über das Restglied die Abschätzung

$$\begin{aligned}h^{(d-1)/2} O(h^2) \iiint |(x - hu)_l^\eta| |(x - hv)_q^\zeta| |\nabla K(u)_r| |\nabla K(v)_s| \|u\|^2 f(x - hv) dx dudv \\ = O(h^{(d+3)/2}) \iint \|u\|^2 |\nabla K(u)_r| |\nabla K(v)_s| \int |(y + h(v - u))_l^\eta| |y_q^\zeta| f(y) dy dudv \\ = O(h^{(d+3)/2}).\end{aligned}$$

Folglich besitzt  $\theta_n$  die Gestalt

$$\begin{aligned}\theta_n &= 2 h^{(d-1)/2} \iiint (x - hu)_l^\eta (x - hv)_q^\zeta \nabla K(u)_r \nabla K(v)_s f(x) f(x - hv) dx dudv \\ &\quad - 2 h^{(d+1)/2} \iiint (x - hu)_l^\eta (x - hv)_q^\zeta \nabla K(u)_r \nabla K(v)_s \nabla f(x) u f(x - hv) dx dudv \\ &\quad + O(h^{(d+3)/2}).\end{aligned}$$

Setzt man in die beiden Integrale eine analoge Taylor–Entwicklung von  $f(x - hv)$  ein, so gelten für die Restglieder die Abschätzungen

$$\begin{aligned}h^{(d-1)/2} O(h^2) \iint |\nabla K(u)_r| \|v\|^2 |\nabla K(v)_s| \int |(x - hu)_l^\eta| |(x - hv)_q^\zeta| f(x) dx dudv \\ = O(h^{(d+3)/2})\end{aligned}$$

und

$$\begin{aligned}h^{(d+1)/2} O(h^2) \iint \|u\| |\nabla K(u)_r| \|v\|^2 |\nabla K(v)_s| \int |(x - hu)_l^\eta| |(x - hv)_q^\zeta| \|\nabla f(x)\| dx dudv \\ = O(h^{(d+5)/2}).\end{aligned}$$

Damit folgt insgesamt für  $\theta_n$

$$\theta_n = 2 h^{(d-1)/2} \iiint (x - hu)_l^\eta (x - hv)_q^\zeta \nabla K(u)_r \nabla K(v)_s f(x)^2 dx dudv$$

$$\begin{aligned}
& -2 h^{(d+1)/2} \iiint (x-hu)_l^\eta (x-hv)_q^\zeta \nabla K(u)_r \nabla K(v)_s f(x) \nabla f(x) v dx du dv \\
& -2 h^{(d+1)/2} \iiint (x-hu)_l^\eta (x-hv)_q^\zeta \nabla K(u)_r \nabla K(v)_s \nabla f(x) u f(x) dx du dv \\
& +2 h^{(d+3)/2} \iiint (x-hu)_l^\eta (x-hv)_q^\zeta \nabla K(u)_r \nabla K(v)_s \nabla f(x) u \nabla f(x) v dx du dv \\
& + O(h^{(d+3)/2}) .
\end{aligned}$$

Um zu der behaupteten Größenordnung  $O(h^{(d+3)/2})$  zu gelangen, muß noch gezeigt werden, daß die ersten drei Integrale in der Darstellung von  $\theta_n$  von der Ordnung  $O(h^2)$  bzw.  $O(h)$  sind. Dies geschieht durch eine Fallunterscheidung nach den Werten von  $\eta$  und  $\zeta$ , wobei nur der umfangreichste Fall  $\eta = \zeta = 1$  aufgeführt wird. Die anderen Fälle  $\eta = \zeta = 0$  bzw.  $\eta = 1, \zeta = 0$  bzw.  $\eta = 0, \zeta = 1$  verlaufen mit derselben Argumentation etwas einfacher.

Ausmultiplizieren von  $(x-hu)_l(x-hv)_q$  und Berücksichtigen von  $\int \nabla K(u)_r du = 0$  nach Bemerkung 25 auf Seite 87 liefert

$$\begin{aligned}
\theta_n &= 2 h^{(d-1)/2} \left\{ \int \nabla K(u)_r du \int \nabla K(v)_s dv \int x_l x_q f(x)^2 dx \right. \\
&\quad - h \int \nabla K(v)_s dv \iint u_l x_q \nabla K(u)_r f(x)^2 dx du \\
&\quad - h \int \nabla K(u)_r du \iint x_l v_q \nabla K(v)_s f(x)^2 dx dv \\
&\quad \left. + h^2 \iiint u_l v_q \nabla K(u)_r \nabla K(v)_s f(x)^2 dx du dv \right\} \\
&-2 h^{(d+1)/2} \left\{ \int \nabla K(u)_r du \iint x_l x_q \nabla K(v)_s f(x) \nabla f(x) v dx dv \right. \\
&\quad + h \iiint [-x_q u_l - x_l v_q + h u_l v_q] \nabla K(u)_r \nabla K(v)_s f(x) \nabla f(x) v dx du dv \\
&\quad + \int \nabla K(v)_s dv \iint x_l x_q \nabla K(u)_r \nabla f(x) u f(x) dx du \\
&\quad \left. + h \iiint [-x_q u_l - x_l v_q + h u_l v_q] \nabla K(u)_r \nabla K(v)_s \nabla f(x) u f(x) dx du dv \right\} \\
&\quad + O(h^{(d+3)/2}) \\
&= O(h^{(d+3)/2}) .
\end{aligned}$$

(ii) BEWEIS VON  $E[r_n(X_1)^2] = O(h^{d-1})$ :

Wegen

$$\begin{aligned}
r_n(x) &= E[H_n(x, X_2)] \\
&= \frac{1}{h^{(3d+1)/2}} \iint \left[ x_l^\eta z_q^\zeta \nabla K\left(\frac{y-x}{h}\right)_r \nabla K\left(\frac{y-z}{h}\right)_s \right.
\end{aligned}$$

$$+ z_l^\eta x_q^\zeta \nabla K \left( \frac{y-z}{h} \right)_r \nabla K \left( \frac{y-x}{h} \right)_s \Big] f(z) dy dz$$

ist

$$\begin{aligned} r_n(X_1)^2 &= \frac{1}{h^{3d+1}} \iiint \iiint \left[ X_{1l}^\eta z_q^\zeta \nabla K \left( \frac{y-X_1}{h} \right)_r \nabla K \left( \frac{y-z}{h} \right)_s + z_l^\eta X_{1q}^\zeta \nabla K \left( \frac{y-z}{h} \right)_r \nabla K \left( \frac{y-X_1}{h} \right)_s \right] \\ &\quad \left[ X_{1l}^\eta u_q^\zeta \nabla K \left( \frac{w-X_1}{h} \right)_r \nabla K \left( \frac{w-u}{h} \right)_s + u_l^\eta X_{1q}^\zeta \nabla K \left( \frac{w-u}{h} \right)_r \nabla K \left( \frac{w-X_1}{h} \right)_s \right] \\ &\quad f(z) f(u) dy dz dw du . \end{aligned}$$

Durch Ausmultiplizieren der beiden Klammern stellt man fest, daß zur Bestimmung der Größenordnung von  $E[r_n(X_1)^2]$  nur Terme der Form

$$\frac{1}{h^{3d+1}} X_{1c}^\eta X_{1k}^\zeta \iiint \iiint z_l^\alpha u_m^\delta \nabla K \left( \frac{y-X_1}{h} \right)_p \nabla K \left( \frac{y-z}{h} \right)_q \nabla K \left( \frac{w-X_1}{h} \right)_r \nabla K \left( \frac{w-u}{h} \right)_s f(z) f(u) dy dz dw du$$

mit  $\eta, \zeta, \alpha, \delta \in \{0, 1\}$  und  $c, k, l, m, p, q, r, s \in \{1, \dots, d\}$  auftreten. Mittels der Substitutionen  $v = (y-x)/h$ ,  $dv = h^{-d} dy$ ,  $t = (w-x)/h$ ,  $dt = h^{-d} dw$ ,  $\xi = (x-z)/h$ ,  $d\xi = h^{-d} dz$ ,  $\varrho = (x-u)/h$ ,  $d\varrho = h^{-d} du$  gilt für den Erwartungswert eines solchen Terms

$$\begin{aligned} &\frac{1}{h^{3d+1}} \iiint \iiint \iiint x_{1c}^\eta x_{1k}^\zeta z_l^\alpha u_m^\delta \nabla K \left( \frac{y-x}{h} \right)_p \nabla K \left( \frac{y-z}{h} \right)_q \nabla K \left( \frac{w-x}{h} \right)_r \nabla K \left( \frac{w-u}{h} \right)_s \\ &\quad f(z) f(u) f(x) dx dy dz dw du \\ &= \frac{1}{h^{d+1}} \iiint \iiint \iiint x_{1c}^\eta x_{1k}^\zeta z_l^\alpha u_m^\delta \nabla K(v)_p \nabla K \left( v + \frac{x-z}{h} \right)_q \nabla K(t)_r \nabla K \left( t + \frac{x-u}{h} \right)_s \\ &\quad f(z) f(u) f(x) dx dv dz dt du \\ &= h^{d-1} \iiint \iiint \iiint x_{1c}^\eta x_{1k}^\zeta (x-h\xi)_l^\alpha (x-h\varrho)_m^\delta \nabla K(v)_p \nabla K(v+\xi)_q \nabla K(t)_r \nabla K(t+\varrho)_s \\ &\quad f(x-h\xi) f(x-h\varrho) f(x) dx dv d\xi dt d\varrho \\ &= O(h^{d-1}) , \end{aligned}$$

denn wegen  $|f(x-h\xi)f(x-h\varrho)| \leq \text{const}$  und der Existenz der Momente vierter Ordnung von  $f$  läßt sich das letzte Integral abschätzen durch

$$\begin{aligned} &\text{const} \cdot \iiint \iiint \left| \nabla K(v)_p \right| \left| \nabla K(v+\xi)_q \right| \left| \nabla K(t)_r \right| \left| \nabla K(t+\varrho)_s \right| \\ &\quad \cdot \left\{ \int \left| x_{1c}^\eta x_{1k}^\zeta (x-h\xi)_l^\alpha (x-h\varrho)_m^\delta \right| f(x) dx \right\} dv d\xi dt d\varrho < \infty . \end{aligned}$$

□

**Lemma 54** Unter den Voraussetzungen 1 und 2 an die Bandbreite  $h$ , den Kern  $K$  und die den Daten  $X_1, \dots, X_n$  zugrundeliegende Dichte  $f$  gilt für jedes feste  $l, q, r, s \in \{1, \dots, d\}$  und  $\eta, \zeta \in \{0, 1\}$

$$\begin{aligned} & \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i,j=1}^n X_{ir}^\eta \int \nabla K \left( \frac{x-X_i}{h} \right)_s \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E} K \left( \frac{x-X_1}{h} \right) \right] dx = O_P(1), \\ & \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i,j=1}^n X_{il}^\eta X_{iq}^\zeta \int \nabla^2 K \left( \frac{x-X_i}{h} \right)_{rs} \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E} K \left( \frac{x-X_1}{h} \right) \right] dx = O_P(1). \end{aligned}$$

BEWEIS: Es seien  $l, q, r, s \in \{1, \dots, d\}$  und  $\eta, \zeta \in \{0, 1\}$  fest. Wir führen den Beweis für die zweite Behauptung aus; der Beweis für die erste Aussage des Lemmas ergibt sich unmittelbar durch Ersetzen von  $\nabla^2 K(\cdot)_{rs}$  durch  $\nabla K(\cdot)_s$  und mit der speziellen Wahl  $\zeta = 0$ .

Die zu untersuchende Größe läßt sich gemäß

$$\begin{aligned} & \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i,j=1}^n X_{il}^\eta X_{iq}^\zeta \int \nabla^2 K \left( \frac{x-X_i}{h} \right)_{rs} \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E} K \left( \frac{x-X_1}{h} \right) \right] dx \\ &= \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i=1}^n X_{il}^\eta X_{iq}^\zeta \int \nabla^2 K \left( \frac{x-X_i}{h} \right)_{rs} \left[ K \left( \frac{x-X_i}{h} \right) - \mathbb{E} K \left( \frac{x-X_1}{h} \right) \right] dx \\ & \quad + \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i \neq j} X_{il}^\eta X_{iq}^\zeta \int \nabla^2 K \left( \frac{x-X_i}{h} \right)_{rs} \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E} K \left( \frac{x-X_1}{h} \right) \right] dx \end{aligned}$$

zerlegen in einen linearen und einen quadratischen Term. Der lineare Term läßt sich mittels der Substitution  $y = (x - X_i)/h$ ,  $dy = h^{-d} dx$ , nach dem starken Gesetz großer Zahlen und nach Voraussetzung 2 abschätzen durch

$$\begin{aligned} & \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i=1}^n |X_{il}^\eta| |X_{iq}^\zeta| \int \left| \nabla^2 K \left( \frac{x-X_i}{h} \right)_{rs} \right| \left| K \left( \frac{x-X_i}{h} \right) - \mathbb{E} K \left( \frac{x-X_1}{h} \right) \right| dx \\ & \leq 2 \sup_{x \in \mathbb{R}^d} K(x) \cdot \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i=1}^n |X_{il}^\eta| |X_{iq}^\zeta| \int \left| \nabla^2 K \left( \frac{x-X_i}{h} \right)_{rs} \right| dx \\ & = 2 \sup_{x \in \mathbb{R}^d} K(x) \cdot \frac{h^{1/2}}{\sqrt{nh^{d+4}}} \cdot \frac{1}{n} \sum_{i=1}^n |X_{il}^\eta| |X_{iq}^\zeta| \cdot \int |\nabla^2 K(y)_{rs}| dy \\ & = o_P(1). \end{aligned}$$

Folglich ist nur der quadratische Term

$$\begin{aligned} \tilde{U}_n &:= \frac{1}{n^{3/2} h^{(3d+3)/2}} \sum_{i \neq j} X_{il}^\eta X_{iq}^\zeta \int \nabla^2 K \left( \frac{x-X_i}{h} \right)_{rs} \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E} K \left( \frac{x-X_1}{h} \right) \right] dx \\ &= \frac{1}{2} \cdot \frac{1}{n^{3/2}} \sum_{i \neq j} \tilde{H}_n(X_i, X_j), \end{aligned}$$

also eine  $U$ -Statistik mit dem symmetrischen, vom Stichprobenumfang  $n$  abhängenden Kern

$$\begin{aligned} \tilde{H}_n(X_i, X_j) &:= \frac{1}{h^{(3d+3)/2}} \left\{ X_{il}^\eta X_{iq}^\zeta \int \nabla^2 K \left( \frac{x-X_i}{h} \right)_{rs} \left[ K \left( \frac{x-X_j}{h} \right) - \mathbb{E}K \left( \frac{x-X_1}{h} \right) \right] dx \right. \\ &\quad \left. + X_{jl}^\eta X_{jq}^\zeta \int \nabla^2 K \left( \frac{x-X_j}{h} \right)_{rs} \left[ K \left( \frac{x-X_i}{h} \right) - \mathbb{E}K \left( \frac{x-X_1}{h} \right) \right] dx \right\} \end{aligned}$$

weiter zu untersuchen. Dies geschieht durch Reduzieren der quadratischen  $U$ -Statistik  $\tilde{U}_n$  auf eine Summe von unabhängigen und identisch verteilten Zufallsgrößen mit Hilfe der Hájek-Projektion, genauer gesagt durch Anwenden von Lemma D.3 auf Seite 181. Hierzu stellen wir fest, daß mit der Substitution  $y = (x - X_1)/h$ ,  $dy = h^{-d}dx$  und nach Voraussetzung 2 die Abschätzungen

$$\begin{aligned} &\left| X_{1l}^\eta X_{1q}^\zeta \int \nabla^2 K \left( \frac{x-X_1}{h} \right)_{rs} \left[ K \left( \frac{x-X_2}{h} \right) - \mathbb{E}K \left( \frac{x-X_1}{h} \right) \right] dx \right| \\ &\leq \left| X_{1l}^\eta X_{1q}^\zeta \right| 2 \sup_{x \in \mathbb{R}^d} K(x) \int \left| \nabla^2 K \left( \frac{x-X_1}{h} \right)_{rs} \right| dx \\ &= h^d \left| X_{1l}^\eta X_{1q}^\zeta \right| 2 \sup_{x \in \mathbb{R}^d} K(x) \int |\nabla^2 K(y)_{rs}| dy \leq \text{const} \cdot h^d \left| X_{1l}^\eta X_{1q}^\zeta \right| \end{aligned}$$

und

$$\begin{aligned} \mathbb{E} \left[ \tilde{H}_n(X_1, X_2)^2 \right] &\leq \frac{1}{h^{3d+3}} \mathbb{E} \left[ \left\{ \text{const} \cdot h^d \left( \left| X_{1l}^\eta X_{1q}^\zeta \right| + \left| X_{2l}^\eta X_{2q}^\zeta \right| \right) \right\}^2 \right] \\ &= \frac{\text{const}}{h^{d+3}} \mathbb{E} \left[ \left( \left| X_{1l}^\eta X_{1q}^\zeta \right| + \left| X_{2l}^\eta X_{2q}^\zeta \right| \right)^2 \right] \\ &= o(n) \end{aligned}$$

gelten. Dies impliziert nach Lemma D.3 mit den Bezeichnungen

$$\tilde{\theta}_n := \mathbb{E} \left[ \tilde{H}_n(X_1, X_2) \right] = 0, \quad \tilde{r}_n(X_1) := \mathbb{E} \left[ \tilde{H}_n(X_1, X_2) \mid X_1 \right]$$

die folgende asymptotische Äquivalenz zwischen der Hájek-Projektion

$$\hat{U}_n := \tilde{\theta}_n + \frac{2}{n} \sum_{i=1}^n \left( \tilde{r}_n(X_i) - \tilde{\theta}_n \right) = \frac{2}{n} \sum_{i=1}^n \tilde{r}_n(X_i)$$

und der auf dem Kern  $\tilde{H}_n$  basierenden  $U$ -Statistik:

$$\frac{1}{n(n-1)} \sum_{i \neq j} \tilde{H}_n(X_i, X_j) = \hat{U}_n + o_P \left( \frac{1}{\sqrt{n}} \right).$$

Daraus ergibt sich für  $\tilde{U}_n$  zunächst

$$\tilde{U}_n = \frac{1}{2} \frac{n-1}{\sqrt{n}} \hat{U}_n + \frac{1}{2} \frac{n-1}{\sqrt{n}} o_P \left( \frac{1}{\sqrt{n}} \right) = \frac{n-1}{n} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{r}_n(X_i) + o_P(1).$$

Offensichtlich gilt  $E[\tilde{r}_n(X_i)] = \tilde{\theta}_n = 0$ , und im weiteren Verlauf des Beweises werden wir zeigen, daß  $\text{Var}(\tilde{r}_n(X_1)) = O(1)$  gilt. Dies liefert angesichts von

$$E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{r}_n(X_i)\right] = 0, \quad \text{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{r}_n(X_i)\right) = \text{Var}(\tilde{r}_n(X_1)) = O(1)$$

und damit  $n^{-1/2}\sum_{i=1}^n\tilde{r}_n(X_i) = O_P(1)$  nach der Tchebychev-Ungleichung das gewünschte Resultat:

$$\tilde{U}_n = \frac{n-1}{n}O_P(1) + o_P(1) = O_P(1).$$

Es verbleibt also, den Term  $\tilde{r}_n(X_1)$  näher zu untersuchen.

(i) BETRACHTUNG VON  $\tilde{r}_n(X_1)$ : UMFORMULIERUNG DES PROBLEMS:

Wegen

$$\begin{aligned}\tilde{r}_n(x) &= E\left[\tilde{H}_n(x, X_2)\right] \\ &= \frac{1}{h^{(3d+3)/2}} \int E\left[X_{2l}^\eta X_{2q}^\zeta \nabla^2 K\left(\frac{y-X_2}{h}\right)_{rs}\right] \left[K\left(\frac{y-x}{h}\right) - EK\left(\frac{y-X_1}{h}\right)\right] dy \\ &= \frac{1}{h^{(3d+3)/2}} \iint z_l^\eta z_q^\zeta \nabla^2 K\left(\frac{y-z}{h}\right)_{rs} \left[K\left(\frac{y-x}{h}\right) - EK\left(\frac{y-X_1}{h}\right)\right] f(z) dz dy\end{aligned}$$

und folglich

$$\tilde{r}_n(X_1) = \tilde{r}_n^*(X_1) - E[\tilde{r}_n^*(X_1)]$$

mit

$$\tilde{r}_n^*(X_1) := \frac{1}{h^{(3d+3)/2}} \iint z_l^\eta z_q^\zeta \nabla^2 K\left(\frac{y-z}{h}\right)_{rs} K\left(\frac{y-X_1}{h}\right) f(z) dz dy$$

ist nur der Term  $\tilde{r}_n^*(X_1)$  näher zu betrachten. Aus den nachfolgend gezeigten Tatsachen  $E[\tilde{r}_n^*(X_1)] = O(1)$  und  $E[\tilde{r}_n^*(X_1)^2] = O(1)$  folgt dann die gewünschte Größenordnung  $\text{Var}(\tilde{r}_n(X_1)) = E[\tilde{r}_n^*(X_1)^2] - (E[\tilde{r}_n^*(X_1)])^2 = O(1)$ .

(ii) BEWEIS VON  $E[\tilde{r}_n^*(X_1)] = O(1)$ :

Mit den Substitutionen  $u = (y-x)/h$ ,  $du = h^{-d}dy$  und  $w = (x-z)/h$ ,  $dw = h^{-d}dx$  ergibt sich zunächst

$$\begin{aligned}E[\tilde{r}_n^*(X_1)] &= \frac{1}{h^{(3d+3)/2}} \iiint z_l^\eta z_q^\zeta \nabla^2 K\left(\frac{y-z}{h}\right)_{rs} K\left(\frac{y-x}{h}\right) f(z)f(x) dz dy dx \\ &= \frac{1}{h^{(d+3)/2}} \iiint z_l^\eta z_q^\zeta \nabla^2 K\left(u + \frac{x-z}{h}\right)_{rs} K(u) f(z)f(x) dz dudx\end{aligned}$$

$$= h^{(d-3)/2} \iiint z_l^\eta z_q^\zeta \nabla^2 K(u+w)_{rs} K(u) f(z) f(z+hw) dz dudw .$$

Setzt man nun für  $f(z+hw)$  die Taylor–Entwicklung aus Bemerkung 22 auf Seite 85 ein, so gilt nach Gleichung (2.25) für das Integral über das Restglied die Abschätzung

$$h^{(d-3)/2} O(h^2) \iint |\nabla^2 K(u+w)_{rs}| |K(u)| \|w\|^2 \int |z_l^\eta z_q^\zeta| f(z) dz dudw = O(h^{(d+1)/2}) .$$

Damit ergibt sich unter Berücksichtigung von  $\int \nabla^2 K(u+w)_{rs} dw = 0$  nach Bemerkung 25 auf Seite 87 für  $E[\tilde{r}_n^*(X_1)]$  die Darstellung

$$\begin{aligned} E[\tilde{r}_n^*(X_1)] &= h^{(d-3)/2} \int K(u) \left[ \int \nabla^2 K(u+w)_{rs} dw \right] \int z_l^\eta z_q^\zeta f(z)^2 dz du \\ &\quad + h^{(d-1)/2} \iint \nabla^2 K(u+w)_{rs} K(u) \left[ \int z_l^\eta z_q^\zeta f(z) \nabla f(z) dz \right] w dudw + O(h^{(d+1)/2}) \\ &= O(h^{(d-1)/2}) , \end{aligned}$$

also  $E[\tilde{r}_n^*(X_1)] = O(1)$ .

(iii) BEWEIS VON  $E[\tilde{r}_n^*(X_1)^2] = O(1)$ :

Mit den Substitutionen  $v = (y-z)/h$ ,  $dv = h^{-d} dz$ ,  $t = (y-x)/h$ ,  $dt = h^{-d} dx$ ,  $\xi = (w-u)/h$ ,  $d\xi = h^{-d} du$ ,  $\varrho = (w-y)/h$ ,  $d\varrho = h^{-d} dw$  gilt

$$\begin{aligned} E[\tilde{r}_n^*(X_1)^2] &= \frac{1}{h^{3d+3}} \iiint \iiint z_l^\eta z_q^\zeta u_l^\eta u_q^\zeta \nabla^2 K\left(\frac{y-z}{h}\right)_{rs} \nabla^2 K\left(\frac{w-u}{h}\right)_{rs} K\left(\frac{y-x}{h}\right) K\left(\frac{w-x}{h}\right) \\ &\quad f(z) f(u) f(x) dz dy dx dudw \\ &= \frac{1}{h^{d+3}} \iiint \iiint (y-hv)_l^\eta (y-hv)_q^\zeta u_l^\eta u_q^\zeta \nabla^2 K(v)_{rs} \nabla^2 K\left(\frac{w-u}{h}\right)_{rs} K(t) K\left(t + \frac{w-y}{h}\right) \\ &\quad f(y-hv) f(u) f(y-hv) dv dy dt dudw \\ &= h^{d-3} \iiint \iiint (y-hv)_l^\eta (y-hv)_q^\zeta (y+h(\varrho-\xi))_l^\eta (y+h(\varrho-\xi))_q^\zeta \\ &\quad \nabla^2 K(v)_{rs} \nabla^2 K(\xi)_{rs} K(t) K(t+\varrho) f(y-hv) f(y+h(\varrho-\xi)) f(y-hv) dv dy dt d\xi d\varrho . \end{aligned}$$

Setzt man nun für  $f(y-hv)$  die Taylor–Entwicklung aus Bemerkung 22 auf Seite 85 ein, so gilt nach Gleichung (2.25) und nach den Voraussetzungen 1 und 2 für das Integral über das Restglied die Abschätzung

$$\begin{aligned} h^{d-3} O(h^2) \iiint \iiint \|v\|^2 |\nabla^2 K(v)_{rs}| |\nabla^2 K(\xi)_{rs}| K(t) K(t+\varrho) \\ \left[ \int |(y-hv)_l^\eta| |(y-hv)_q^\zeta| |(y+h(\varrho-\xi))_l^\eta| |(y+h(\varrho-\xi))_q^\zeta| f(y-hv) dy \right] dv dt d\xi d\varrho \end{aligned}$$

$$\begin{aligned}
&= O(h^{d-1}) \iiint \int \|v\|^2 |\nabla^2 K(v)_{rs}| |\nabla^2 K(\xi)_{rs}| K(t) K(t+\varrho) \\
&\quad \left[ \int (z+h(t-v))_i^\eta |(z+h(t-v))_q^\zeta| |(z+h(t+\varrho-\xi))_i^\eta| |(z+h(t+\varrho-\xi))_q^\zeta| f(z) dz \right] dv dt d\xi d\varrho \\
&= O(h^{d-1}) .
\end{aligned}$$

Daraus folgt für  $E[\tilde{r}_n^*(X_1)^2]$  die Entwicklung

$$\begin{aligned}
&E[\tilde{r}_n^*(X_1)^2] \\
&= h^{d-3} \iiint \int \int \int (y-hv)_i^\eta (y-hv)_q^\zeta (y+h(\varrho-\xi))_i^\eta (y+h(\varrho-\xi))_q^\zeta \\
&\quad \nabla^2 K(v)_{rs} \nabla^2 K(\xi)_{rs} K(t) K(t+\varrho) f(y) f(y+h(\varrho-\xi)) f(y-h\varrho) dv dy dt d\xi d\varrho \\
&\quad - h^{d-2} \iiint \int \int \int (y-hv)_i^\eta (y-hv)_q^\zeta (y+h(\varrho-\xi))_i^\eta (y+h(\varrho-\xi))_q^\zeta \\
&\quad \nabla^2 K(v)_{rs} \nabla^2 K(\xi)_{rs} K(t) K(t+\varrho) \nabla f(y) v f(y+h(\varrho-\xi)) f(y-h\varrho) dv dy dt d\xi d\varrho \\
&\quad + O(1) .
\end{aligned}$$

Um zu der Behauptung  $E[\tilde{r}_n^*(X_1)^2] = O(1)$  zu gelangen, muß noch gezeigt werden, daß das erste Integral auf der rechten Seite von der Größenordnung  $O(h^2)$  und das zweite Integral von der Größenordnung  $O(h)$  ist. Dies geschieht durch eine Fallunterscheidung nach den Werten von  $\eta$  und  $\zeta$ , wobei nur der umfangreichste Fall  $\eta = \zeta = 1$  aufgeführt wird. Die anderen Fälle  $\eta = \zeta = 0$  bzw.  $\eta = 1, \zeta = 0$  bzw.  $\eta = 0, \zeta = 1$  verlaufen mit derselben Argumentation etwas einfacher. Setzt man

$$\begin{aligned}
&(y-hv)_i (y-hv)_q (y+h(\varrho-\xi))_i (y+h(\varrho-\xi))_q \\
&= y_i^2 y_q^2 + h \{ -y_i^2 y_q v_q - y_i y_q^2 v_i + y_i^2 y_q (\varrho-\xi)_q + y_i y_q^2 (\varrho-\xi)_i \} \\
&\quad + h^2 \{ y_i y_q (\varrho-\xi-v)_i (\varrho-\xi-v)_q - y_q^2 v_i (\varrho-\xi)_i - y_i^2 v_q (\varrho-\xi)_q \} \\
&\quad + h^3 \{ v_i v_q [y_i (\varrho-\xi)_q + y_q (\varrho-\xi)_i] - [y_q v_i + y_i v_q] (\varrho-\xi)_i (\varrho-\xi)_q \} \\
&\quad + h^4 v_i v_q (\varrho-\xi)_i (\varrho-\xi)_q
\end{aligned}$$

in die beiden Integrale ein, so erhält man wegen  $\int \nabla^2 K(v)_{rs} dv = 0$  nach Bemerkung 25 auf Seite 87

$$\begin{aligned}
&E[\tilde{r}_n^*(X_1)^2] \\
&= h^{d-3} \int \nabla^2 K(v)_{rs} dv \iiint \int \int y_i^2 y_q^2 \nabla^2 K(\xi)_{rs} K(t) K(t+\varrho) f(y) f(y+h(\varrho-\xi)) f(y-h\varrho) \\
&\quad dy dt d\xi d\varrho \\
&\quad - h^{d-2} \iiint \int \int \int [y_i^2 y_q v_q + y_i y_q^2 v_i] \nabla^2 K(v)_{rs} \nabla^2 K(\xi)_{rs} K(t) K(t+\varrho) f(y) f(y+h(\varrho-\xi)) \\
&\quad \quad \quad f(y-h\varrho) dv dy dt d\xi d\varrho \\
&\quad + h^{d-2} \int \nabla^2 K(v)_{rs} dv \iiint \int \int [y_i^2 y_q (\varrho-\xi)_q + y_i y_q^2 (\varrho-\xi)_i] \nabla^2 K(\xi)_{rs} K(t) K(t+\varrho) \\
&\quad \quad \quad f(y) f(y+h(\varrho-\xi)) f(y-h\varrho) dy dt d\xi d\varrho
\end{aligned}$$



$$5.) \int (e_h(x, g_1) - e_h(x, g_2))^2 dx = \int (g_1(x) - g_2(x))^2 dx + O(h^2).$$

BEWEIS: 1.) Mit der Taylor-Entwicklung aus Bemerkung 22 auf Seite 85 und nach Voraussetzung 1 gilt

$$\begin{aligned} e_h(x, g) &= \int K(y)g(x - hy)dy \\ &= g(x) \int K(y)dy - h\nabla g(x) \int yK(y)dy + O(h^2) \int \|y\|^2 K(y)dy \\ &= g(x) + O(h^2), \end{aligned}$$

wobei der Term  $O(h^2)$  nach Gleichung (2.25) gleichmäßig in  $x \in \mathbb{R}^d$  gilt.

2.) Mit der Substitution  $u = (x - z)/h$ ,  $du = h^{-d}dz$  und der Taylor-Entwicklung aus Bemerkung 22 auf Seite 85 gilt

$$\begin{aligned} e_h(x + hy, g) &= \frac{1}{h^d} \int K\left(\frac{x - z}{h} + y\right) g(z) dz = \int K(u + y) g(x - hu) du \\ &= g(x) \int K(u + y) du - h\nabla g(x) \int uK(u + y) du + O(h^2) \int \|u\|^2 K(u + y) du. \end{aligned}$$

Wegen

$$\int K(u + y) du = \int K(x) dx = 1, \quad (2.63)$$

$$\int uK(u + y) du = \int xK(x) dx - y \int K(x) dx = -y \quad (2.64)$$

nach Voraussetzung 1 und

$$\begin{aligned} \int \|u\|^2 K(u + y) du &= \int \|x - y\|^2 K(x) dx \\ &\leq \int \|x\|^2 K(x) dx + 2\|y\| \int \|x\| K(x) dx + \|y\|^2 \int K(x) dx \end{aligned}$$

folgt

$$e_h(x + hy, g) = g(x) + h\nabla g(x)y + O(h^2) \sum_{l=0}^2 \|y\|^l.$$

3.) Klar nach 1.)

4.) Mit den Taylor-Entwicklungen aus Bemerkung 22 auf Seite 85 ergibt sich

$$\int e_h(x, g)^2 dx$$

$$\begin{aligned}
&= \iiint K(y)K(z)g(x-hy)g(x-hz)dzdydx \\
&= \int K(y)dy \iint K(z)g(x)g(x-hz)dzdx - h \iint K(z)\nabla g(x) \left[ \int yK(y)dy \right] g(x-hz)dzdx \\
&\quad + O(h^2) \iiint \|y\|^2 K(y)K(z)g(x-hz)dzdydx \\
&= \iint K(z)g(x)g(x-hz)dzdx + O(h^2) \\
&= \int K(z)dz \int g(x)^2 dx - h \int g(x)\nabla g(x) \left[ \int zK(z)dz \right] dx + O(h^2) \iint \|z\|^2 K(z)g(x)dzdx \\
&= \int g(x)^2 dx + O(h^2) .
\end{aligned}$$

5.) Wegen

$$e_h(x, g_1) - e_h(x, g_2) = \int K(y)(g_1 - g_2)(x - hy)dy = e_h(x, g_1 - g_2)$$

ergibt sich die Behauptung aus 4.). □

**Lemma 56** Mit  $\mathcal{I}_1$  und  $\mathcal{I}_2$  aus Gleichung (2.42) auf Seite 102 sowie  $\mathcal{I}_{3,r,s}$  aus Gleichung (2.43) gelten für  $r, s \in \{1, \dots, d\}$  unter den Voraussetzungen 1 und 2 an  $K, f_0$  und die den Daten  $X_1, \dots, X_n$  zugrundeliegende Dichte  $f_1$

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \frac{1}{h^{d+1}} \int \nabla K \left( \frac{x - X_i}{h} \right)_r (e_h(x, f_1) - e_h(x, f_0)) dx &= \mathcal{I}_{1r} + o_P(1), \\
\frac{1}{n} \sum_{i=1}^n \frac{1}{h^{d+1}} X_{is} \int \nabla K \left( \frac{x - X_i}{h} \right)_r (e_h(x, f_1) - e_h(x, f_0)) dx &= \mathcal{I}_2 \mathbf{1}_{\{r=s\}} + \mathcal{I}_{3,r,s} + o_P(1).
\end{aligned}$$

**BEWEIS:** Für beide Behauptungen reicht es zu zeigen, daß der Erwartungswert der linken Seite jeweils gegen die Konstante auf der rechten Seite der obigen Gleichungen strebt und daß die Varianz der linken Seite mit wachsendem Stichprobenumfang  $n$  verschwindet. Da die Rechnungen in den beiden Fällen über weite Strecken parallel verlaufen, betrachten wir allgemeiner den generischen Term

$$\tilde{S}_{n,3,i} := \frac{1}{h^{d+1}} X_{is}^\eta \int \nabla K \left( \frac{x - X_i}{h} \right)_r (e_h(x, f_1) - e_h(x, f_0)) dx$$

mit  $r, s \in \{1, \dots, d\}$  sowie  $\eta \in \{0, 1\}$  und unterscheiden nur bei Bedarf die Fälle 1 ( $\eta = 0$ ) und 2 ( $\eta = 1$ ).

(i) BERECHNUNG DER ERWARTUNGSWERTE:

Mit der Substitution  $y = (x - z)/h$ ,  $dy = h^{-d}dz$  ist

$$\begin{aligned} \mathbb{E}\tilde{S}_{n,3,i} &= \frac{1}{h^{d+1}} \iint z_s^\eta \nabla K \left( \frac{x-z}{h} \right)_r (e_h(x, f_1) - e_h(x, f_0)) f_1(z) dx dz \\ &= \frac{1}{h} \iint (x - hy)_s^\eta \nabla K(y)_r (e_h(x, f_1) - e_h(x, f_0)) f_1(x - hy) dx dy \\ &= \frac{1}{h} \iint (x - hy)_s^\eta \nabla K(y)_r (f_1(x) - f_0(x)) f_1(x - hy) dx dy + O(h) , \end{aligned}$$

denn nach Lemma 55 und den Voraussetzungen 1 und 2 gilt für den Restterm die Abschätzung

$$\frac{1}{h} O(h^2) \int |\nabla K(y)_r| \int |(x - hy)_s^\eta| f_1(x - hy) dx dy = O(h) .$$

Setzt man nun für  $f_1(x - hy)$  die Taylor-Entwicklung aus Bemerkung 22 auf Seite 85 ein, so erhält man

$$\begin{aligned} \mathbb{E}\tilde{S}_{n,3,i} &= \frac{1}{h} \iint (x - hy)_s^\eta \nabla K(y)_r (f_1(x) - f_0(x)) f_1(x) dx dy \\ &\quad - \iint (x - hy)_s^\eta \nabla K(y)_r (f_1(x) - f_0(x)) \nabla f_1(x)_y dx dy + O(h) , \end{aligned}$$

denn für den Restterm gilt nach Gleichung (2.25) die Abschätzung

$$\frac{1}{h} O(h^2) \int \|y\|^2 |\nabla K(y)_r| \int |(x - hy)_s^\eta| |f_1(x) - f_0(x)| dx dy = O(h) .$$

FALL 1:  $\eta = 0$ :

Nach Bemerkung 25 und Lemma 26 auf Seite 87 ist

$$\begin{aligned} \mathbb{E}\tilde{S}_{n,3,i} &= \frac{1}{h} \int \nabla K(y)_r dy \int (f_1(x) - f_0(x)) f_1(x) dx \\ &\quad - \int (f_1(x) - f_0(x)) \nabla f_1(x)_y dx \int y \nabla K(y)_r dy + O(h) \\ &= \int \nabla f_1(x)_r (f_1(x) - f_0(x)) dx + O(h) = \mathcal{I}_{1r} + O(h) . \end{aligned}$$

FALL 2:  $\eta = 1$ :

Mit derselben Argumentation ergibt sich

$$\begin{aligned} \mathbb{E}\tilde{S}_{n,3,i} &= \frac{1}{h} \int \nabla K(y)_r dy \int x_s (f_1(x) - f_0(x)) f_1(x) dx \\ &\quad - \int y_s \nabla K(y)_r dy \int (f_1(x) - f_0(x)) f_1(x) dx \\ &\quad - \int x_s (f_1(x) - f_0(x)) \nabla f_1(x)_y dx \int y \nabla K(y)_r dy + O(h) \\ &= \mathbf{1}_{\{r=s\}} \int (f_1(x) - f_0(x)) f_1(x) dx + \int \nabla f_1(x)_r x_s (f_1(x) - f_0(x)) dx + O(h) \end{aligned}$$

$$= \mathcal{I}_2 \mathbf{1}_{\{r=s\}} + \mathcal{I}_{3,r,s} + O(h) .$$

(ii) BERECHNUNG DER ZWEITEN MOMENTE:

Mit den Substitutionen  $u = (x - z)/h$ ,  $du = h^{-d}dz$  und  $v = (y - x)/h$ ,  $dv = h^{-d}dy$  ist

$$\begin{aligned} \mathbb{E} \left[ \tilde{S}_{n,3,i}^2 \right] &= \frac{1}{h^{2d+2}} \iiint z_s^{2\eta} \nabla K \left( \frac{x-z}{h} \right)_r \nabla K \left( \frac{y-z}{h} \right)_r (e_h(x, f_1) - e_h(x, f_0)) \\ &\quad (e_h(y, f_1) - e_h(y, f_0)) f_1(z) dx dy dz \\ &= \frac{1}{h^{d+2}} \iiint (x - hu)_s^{2\eta} \nabla K(u)_r \nabla K \left( u + \frac{y-x}{h} \right)_r (e_h(x, f_1) - e_h(x, f_0)) \\ &\quad (e_h(y, f_1) - e_h(y, f_0)) f_1(x - hu) dx dy du \\ &= \frac{1}{h^2} \iiint (x - hu)_s^{2\eta} \nabla K(u)_r \nabla K(u+v)_r (e_h(x, f_1) - e_h(x, f_0)) \\ &\quad (e_h(x+ hv, f_1) - e_h(x+ hv, f_0)) f_1(x - hu) dx dv du \\ &= \frac{1}{h^2} \iiint (x - hu)_s^{2\eta} \nabla K(u)_r \nabla K(u+v)_r (f_1(x) - f_0(x)) \\ &\quad (e_h(x+ hv, f_1) - e_h(x+ hv, f_0)) f_1(x - hu) dx dv du + O(1) , \end{aligned}$$

denn nach Lemma 55 und den Voraussetzungen 1 und 2 gilt für den Restterm die Abschätzung

$$\begin{aligned} O(1) \iiint |(x - hu)_s^{2\eta}| |\nabla K(u)_r| |\nabla K(u+v)_r| |e_h(x+ hv, f_1) - e_h(x+ hv, f_0)| f_1(x - hu) dx dv du \\ \leq O(1) \iiint |\nabla K(u+v)_r| |\nabla K(u)_r| \int |(x - hu)_s^{2\eta}| f_1(x - hu) dx dudv = O(1) . \end{aligned}$$

Setzt man nun für  $f_1(x - hu)$  die Taylor–Entwicklung aus Bemerkung 22 auf Seite 85 ein, so erhält man

$$\begin{aligned} \mathbb{E} \left[ \tilde{S}_{n,3,i}^2 \right] &= \frac{1}{h^2} \iiint (x - hu)_s^{2\eta} \nabla K(u)_r \nabla K(u+v)_r (f_1(x) - f_0(x)) \\ &\quad (e_h(x+ hv, f_1) - e_h(x+ hv, f_0)) f_1(x) dx dv du \\ &\quad - \frac{1}{h} \iiint (x - hu)_s^{2\eta} \nabla K(u)_r \nabla K(u+v)_r (f_1(x) - f_0(x)) \\ &\quad (e_h(x+ hv, f_1) - e_h(x+ hv, f_0)) \nabla f_1(x) u dx dv du + O(1) , \end{aligned}$$

denn für den Restterm gilt nach Gleichung (2.25) die Abschätzung

$$\begin{aligned} O(1) \iiint |(x - hu)_s^{2\eta}| |\nabla K(u)_r| |\nabla K(u+v)_r| |f_1(x) - f_0(x)| \\ |e_h(x+ hv, f_1) - e_h(x+ hv, f_0)| \|u\|^2 dx dv du \end{aligned}$$

$$\begin{aligned} &\leq O(1) \iint |\nabla K(u+v)_r| \|u\|^2 |\nabla K(u)_r| \int |(x-hu)_s^{2\eta}| |f_1(x) - f_0(x)| dx dv \\ &= O(1). \end{aligned}$$

Setzt man nun für  $e_h(x+hv, f_1) - e_h(x+hv, f_0)$  die Darstellung aus Lemma 55 ein, ergibt sich wegen

$$\begin{aligned} \int \nabla K(u+v)_r dv &= \int \nabla K(x)_r dx = 0, \\ \int v \nabla K(u+v)_r dv &= \int x \nabla K(x)_r dx - u \int \nabla K(x)_r dx = \int x \nabla K(x)_r dx \end{aligned}$$

nach Lemma 26 die Darstellung

$$\begin{aligned} &\mathbb{E} \left[ \tilde{S}_{n,3,i}^2 \right] \\ &= \frac{1}{h^2} \iint \left[ \int \nabla K(u+v)_r dv \right] (x-hu)_s^{2\eta} \nabla K(u)_r (f_1(x) - f_0(x))^2 f_1(x) dx du \\ &\quad + \frac{1}{h} \iint (x-hu)_s^{2\eta} \nabla K(u)_r (f_1(x) - f_0(x)) (\nabla f_1(x) - \nabla f_0(x)) \left[ \int v \nabla K(u+v)_r dv \right] f_1(x) dx du \\ &\quad - \frac{1}{h} \iint \left[ \int \nabla K(u+v)_r dv \right] (x-hu)_s^{2\eta} \nabla K(u)_r (f_1(x) - f_0(x))^2 \nabla f_1(x) u dx du \\ &\quad - \iint \int (x-hu)_s^{2\eta} \nabla K(u)_r \nabla K(u+v)_r (f_1(x) - f_0(x)) (\nabla f_1(x) - \nabla f_0(x)) v \nabla f_1(x) u dx dv du \\ &\quad + O(1) \\ &= -\frac{1}{h} \iint (x-hu)_s^{2\eta} \nabla K(u)_r (f_1(x) - f_0(x)) (\nabla f_1(x)_r - \nabla f_0(x)_r) f_1(x) dx du + O(1), \end{aligned}$$

denn nach Lemma 55 gelten für die Restterme die Abschätzungen

$$\begin{aligned} &\frac{1}{h^2} O(h^2) \iint \int |(x-hu)_s^{2\eta}| |\nabla K(u)_r| |\nabla K(u+v)_r| |f_1(x) - f_0(x)| \left( \sum_{l=0}^2 \|v\|^l \right) f_1(x) dx dv du \\ &= O(1) \end{aligned}$$

und

$$\begin{aligned} &O(h) \iint \int |(x-hu)_s^{2\eta}| |\nabla K(u)_r| |\nabla K(u+v)_r| |f_1(x) - f_0(x)| \left( \sum_{l=0}^2 \|v\|^l \right) \|\nabla f_1(x)\| \|u\| dx dv du \\ &= O(1). \end{aligned}$$

FALL 1:  $\eta = 0$ :

Wegen  $\int \nabla K(u)_r du = 0$  nach Bemerkung 25 auf Seite 87 gilt

$$\mathbb{E} \left[ \tilde{S}_{n,3,i}^2 \right] = -\frac{1}{h} \int \nabla K(u)_r du \int (f_1(x) - f_0(x)) (\nabla f_1(x)_r - \nabla f_0(x)_r) f_1(x) dx + O(1)$$

$$= O(1).$$

FALL 2:  $\eta = 1$ :

Wegen  $\int \nabla K(u)_r du = 0$ ,  $(x - hu)_s^2 = x_s^2 - 2hx_s u_s + h^2 u_s^2$  und der Integrierbarkeit aller Terme gilt

$$\begin{aligned} \mathbb{E} \left[ \tilde{S}_{n,3,i}^2 \right] &= -\frac{1}{h} \int \nabla K(u)_r du \int x_s^2 (f_1(x) - f_0(x)) (\nabla f_1(x)_r - \nabla f_0(x)_r) f_1(x) dx + O(1) \\ &= O(1). \end{aligned}$$

Insbesondere gilt in beiden Fällen

$$\text{Var} \left( \tilde{S}_{n,3,i} \right) = \mathbb{E} \left[ \tilde{S}_{n,3,i}^2 \right] - \left( \mathbb{E} \left[ \tilde{S}_{n,3,i} \right] \right)^2 = O(1) - O(1)^2 = O(1).$$

(iii) ZUSAMMENFÜGEN DER ERGEBNISSE:

Nach dem bisher Gezeigten gilt

$$\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n \tilde{S}_{n,3,i} \right] = \mathbb{E} \left[ \tilde{S}_{n,3,1} \right], \quad \text{Var} \left( \frac{1}{n} \sum_{i=1}^n \tilde{S}_{n,3,i} \right) = \frac{\text{Var} \left( \tilde{S}_{n,3,1} \right)}{n} = \frac{O(1)}{n} \rightarrow 0,$$

und aus  $\mathbb{E} \tilde{S}_{n,3,i} = \mathcal{I}_{1r} + O(h)$  bzw.  $\mathbb{E} \tilde{S}_{n,3,i} = \mathcal{I}_2 \mathbf{1}_{\{r=s\}} + \mathcal{I}_{3,r,s} + O(h)$  folgt die Behauptung.  $\square$

**Lemma 57** Mit den in den Gleichungen (2.44), (2.45) und (2.46) auf Seite 102 definierten Größen gelten für die Zufallsvariablen  $Y_{n1}$  aus Gleichung (2.48) auf Seite 104 unter den Voraussetzungen 1 und 2 an  $K$ ,  $f_0$  und die den Daten  $X_1, \dots, X_n$  zugrundeliegende Dichte  $f_1$

- 1.)  $\mathbb{E} Y_{n1} = \gamma_1 + O(h^2)$ ,
- 2.)  $\mathbb{E} Y_{n1}^2 = \gamma_2 + O(h^2)$ ,
- 3.)  $\text{Var} (Y_{n1}) = \tilde{\sigma}^2(f_1, f_0) + O(h^2)$ .

BEWEIS: Zunächst sei an die Definition

$$Y_{n1} = \frac{1}{h^d} \int K \left( \frac{x - X_1}{h} \right) (e_h(x, f_1) - e_h(x, f_0)) dx + \mathcal{I}_1 X_1 + \frac{1}{2} \mathcal{I}_2 X_1^T X_1 + \frac{1}{2} \mathcal{I}_3 (X_1 X_1^T)$$

erinnert.

1.) Mit der Substitution  $u = (x - z)/h$ ,  $du = h^{-d} dz$  gilt

$$\mathbb{E} Y_{n1} = \frac{1}{h^d} \iint K \left( \frac{x - z}{h} \right) (e_h(x, f_1) - e_h(x, f_0)) f_1(z) dx dz + \frac{d}{2} \mathcal{I}_2 + \frac{1}{2} \mathcal{I}_3 (I_d)$$

$$\begin{aligned}
&= \iint K(u) (e_h(x, f_1) - e_h(x, f_0)) f_1(x - hu) dx du + \frac{d}{2} \mathcal{I}_2 + \frac{1}{2} \mathcal{I}_3(I_d) \\
&= \iint K(u) (f_1(x) - f_0(x)) f_1(x - hu) dx du + \frac{d}{2} \mathcal{I}_2 + \frac{1}{2} \mathcal{I}_3(I_d) + O(h^2),
\end{aligned}$$

denn nach Lemma 55 gilt für den Restterm des Integrals die Abschätzung

$$O(h^2) \iint K(u) f_1(x - hu) dx du = O(h^2).$$

Die Taylor-Entwicklung von  $f_1(x - hu)$  aus Bemerkung 22 auf Seite 85 liefert mit Voraussetzung 1

$$\begin{aligned}
EY_{n1} &= \int K(u) du \int (f_1(x) - f_0(x)) f_1(x) dx \\
&\quad - h \int (f_1(x) - f_0(x)) \nabla f_1(x) dx \int u K(u) du + \frac{d}{2} \mathcal{I}_2 + \frac{1}{2} \mathcal{I}_3(I_d) + O(h^2) \\
&= \int f_1(x) (f_1(x) - f_0(x)) dx + \frac{d}{2} \mathcal{I}_2 + \frac{1}{2} \mathcal{I}_3(I_d) + O(h^2) \\
&= \gamma_1 + O(h^2),
\end{aligned}$$

denn nach Gleichung (2.25) gilt für den Restterm die Abschätzung

$$O(h^2) \int \|u\|^2 K(u) du \int |f_1(x) - f_0(x)| dx = O(h^2).$$

2.) Aufgrund von  $E[Y_{n1}^2] = \tilde{S}_{n,4} + 2\tilde{S}_{n,5} + \tilde{S}_{n,6}$  mit

$$\begin{aligned}
\tilde{S}_{n,4} &:= \frac{1}{h^{2d}} E \left[ \left( \int K \left( \frac{x - X_1}{h} \right) (e_h(x, f_1) - e_h(x, f_0)) dx \right)^2 \right] \\
\tilde{S}_{n,5} &:= \frac{1}{h^d} E \left[ \left( \mathcal{I}_1 X_1 + \frac{1}{2} \mathcal{I}_2 X_1^T X_1 + \frac{1}{2} \mathcal{I}_3 (X_1 X_1^T) \right) \right. \\
&\quad \left. \cdot \left( \int K \left( \frac{x - X_1}{h} \right) (e_h(x, f_1) - e_h(x, f_0)) dx \right) \right] \\
\tilde{S}_{n,6} &:= E \left[ \left( \mathcal{I}_1 X_1 + \frac{1}{2} \mathcal{I}_2 X_1^T X_1 + \frac{1}{2} \mathcal{I}_3 (X_1 X_1^T) \right)^2 \right]
\end{aligned}$$

folgt die Behauptung durch Behandlung der drei Terme  $\tilde{S}_{n,4}$ ,  $\tilde{S}_{n,5}$  und  $\tilde{S}_{n,6}$ .

(i) BEHANDLUNG VON  $\tilde{S}_{n,4}$ :

Mit den Substitutionen  $u = (x - z)/h$ ,  $du = h^{-d} dz$  und  $v = (y - x)/h$ ,  $dv = h^{-d} dy$  gilt

$$\tilde{S}_{n,4} = \frac{1}{h^{2d}} \iiint K \left( \frac{x - z}{h} \right) K \left( \frac{y - z}{h} \right) (e_h(x, f_1) - e_h(x, f_0))$$

$$\begin{aligned}
& (e_h(y, f_1) - e_h(y, f_0)) f_1(z) dz dx dy \\
= & \frac{1}{h^d} \iiint K(u) K\left(u + \frac{y-x}{h}\right) (e_h(x, f_1) - e_h(x, f_0)) \\
& (e_h(y, f_1) - e_h(y, f_0)) f_1(x-hu) du dx dy \\
= & \iiint K(u) K(u+v) (e_h(x, f_1) - e_h(x, f_0)) \\
& (e_h(x+hv, f_1) - e_h(x+hv, f_0)) f_1(x-hu) du dx dv \\
= & \iiint K(u) K(u+v) (f_1(x) - f_0(x)) \\
& (e_h(x+hv, f_1) - e_h(x+hv, f_0)) f_1(x-hu) du dx dv \\
& + O(h^2) ,
\end{aligned}$$

denn nach Lemma 55 gilt für den Restterm die Abschätzung

$$O(h^2) \iint K(u) K(u+v) \int f_1(x-hu) dx du dv = O(h^2) .$$

Eine weitere Anwendung von Lemma 55 liefert unter Berücksichtigung der Gleichungen (2.63) und (2.64) auf Seite 137

$$\begin{aligned}
\tilde{S}_{n,4} &= \iint K(u) \left[ \int K(u+v) dv \right] (f_1(x) - f_0(x))^2 f_1(x-hu) du dx \\
&+ h \iint K(u) (f_1(x) - f_0(x)) (\nabla f_1(x) - \nabla f_0(x)) \left[ \int v K(u+v) dv \right] f_1(x-hu) du dx \\
&+ O(h^2) \\
&= \iint K(u) (f_1(x) - f_0(x))^2 f_1(x-hu) du dx \\
&- h \iint K(u) (f_1(x) - f_0(x)) (\nabla f_1(x) - \nabla f_0(x)) u f_1(x-hu) du dx + O(h^2) ,
\end{aligned}$$

denn auch hier gilt für den Restterm die Abschätzung

$$O(h^2) \iint K(u) K(u+v) \left( \sum_{l=0}^2 \|v\|^l \right) \int |f_1(x) - f_0(x)| f_1(x-hu) dx du dv = O(h^2) .$$

Schließlich liefert eine Taylor-Entwicklung von  $f_1(x-hu)$  gemäß Bemerkung 22 auf Seite 85

$$\begin{aligned}
\tilde{S}_{n,4} &= \int K(u) du \int (f_1(x) - f_0(x))^2 f_1(x) dx \\
&- h \int (f_1(x) - f_0(x))^2 \nabla f_1(x) \left[ \int u K(u) du \right] dx \\
&- h \int (f_1(x) - f_0(x)) (\nabla f_1(x) - \nabla f_0(x)) \left[ \int u K(u) du \right] f_1(x) dx \\
&+ h^2 \iint K(u) (f_1(x) - f_0(x)) (\nabla f_1(x) - \nabla f_0(x)) u \nabla f_1(x) u du dx + O(h^2)
\end{aligned}$$

$$= \int f_1(x) (f_1(x) - f_0(x))^2 dx + O(h^2) ,$$

denn für die jeweiligen Restterme gelten die Abschätzungen

$$\begin{aligned} O(h^2) \iint K(u) \|u\|^2 (f_1(x) - f_0(x))^2 dudx &= O(h^2) , \\ O(h^3) \iint K(u) \|u\|^3 |f_1(x) - f_0(x)| \|\nabla f_1(x) - \nabla f_0(x)\| dudx &= O(h^2) . \end{aligned}$$

(ii) BEHANDLUNG VON  $\tilde{S}_{n,5}$ :

Mit der Substitution  $u = (x - z)/h$ ,  $du = h^{-d}dz$  gilt

$$\begin{aligned} \tilde{S}_{n,5} &= \frac{1}{h^d} \iint \left( \mathcal{I}_1 z + \frac{1}{2} \mathcal{I}_2 z^T z + \frac{1}{2} \mathcal{I}_3 (zz^T) \right) K\left(\frac{x-z}{h}\right) (e_h(x, f_1) - e_h(x, f_0)) f_1(z) dx dz \\ &= \iint \left( \mathcal{I}_1(x-hu) + \frac{1}{2} \mathcal{I}_2(x-hu)^T(x-hu) + \frac{1}{2} \mathcal{I}_3\left((x-hu)(x-hu)^T\right) \right) K(u) \\ &\quad (e_h(x, f_1) - e_h(x, f_0)) f_1(x-hu) dx du \\ &= \iint \left( \mathcal{I}_1(x-hu) + \frac{1}{2} \mathcal{I}_2(x-hu)^T(x-hu) + \frac{1}{2} \mathcal{I}_3\left((x-hu)(x-hu)^T\right) \right) K(u) \\ &\quad (f_1(x) - f_0(x)) f_1(x-hu) dx du + O(h^2) , \end{aligned}$$

denn für den Restterm gilt die Abschätzung

$$\begin{aligned} O(h^2) \int K(u) \\ \int \left( \|\mathcal{I}_1\| \|x-hu\| + \frac{1}{2} |\mathcal{I}_2| \|x-hu\|^2 + \frac{1}{2} \left| \mathcal{I}_3\left((x-hu)(x-hu)^T\right) \right| \right) f_1(x-hu) dx du \\ = O(h^2) . \end{aligned}$$

Berücksichtigt man nun die Identität

$$\begin{aligned} \mathcal{I}_1(x-hu) + \frac{1}{2} \mathcal{I}_2(x-hu)^T(x-hu) + \frac{1}{2} \mathcal{I}_3\left((x-hu)(x-hu)^T\right) \\ = \mathcal{I}_1 x + \frac{1}{2} \mathcal{I}_2 x^T x + \frac{1}{2} \mathcal{I}_3 (xx^T) - h \left\{ \mathcal{I}_1 u + \mathcal{I}_2 u^T x + \frac{1}{2} \mathcal{I}_3 (ux^T + xu^T) \right\} \\ + h^2 \frac{1}{2} \left\{ \mathcal{I}_2 u^T u + \mathcal{I}_3 (uu^T) \right\} , \end{aligned}$$

so ergibt sich weiter

$$\begin{aligned} \tilde{S}_{n,5} &= \iint K(u) \left( \mathcal{I}_1 x + \frac{1}{2} \mathcal{I}_2 x^T x + \frac{1}{2} \mathcal{I}_3 (xx^T) \right) (f_1(x) - f_0(x)) f_1(x-hu) dx du \\ &\quad - h \iint \left( \mathcal{I}_1 u + \mathcal{I}_2 u^T x + \frac{1}{2} \mathcal{I}_3 (ux^T + xu^T) \right) K(u) (f_1(x) - f_0(x)) f_1(x-hu) dx du \\ &\quad + O(h^2) \end{aligned}$$

und mit einer Taylor–Entwicklung von  $f_1(x - hu)$  gemäß Bemerkung 22 auf Seite 85

$$\begin{aligned}
& \tilde{S}_{n,5} \\
&= \int K(u) du \int \left( \mathcal{I}_1 x + \frac{1}{2} \mathcal{I}_2 x^T x + \frac{1}{2} \mathcal{I}_3 (xx^T) \right) (f_1(x) - f_0(x)) f_1(x) dx \\
&\quad - h \int \left( \mathcal{I}_1 x + \frac{1}{2} \mathcal{I}_2 x^T x + \frac{1}{2} \mathcal{I}_3 (xx^T) \right) (f_1(x) - f_0(x)) \nabla f_1(x) \left[ \int u K(u) du \right] dx \\
&\quad - h \iint \left( \mathcal{I}_1 u + \mathcal{I}_2 u^T x + \frac{1}{2} \mathcal{I}_3 (ux^T + xu^T) \right) K(u) (f_1(x) - f_0(x)) f_1(x) dx du \\
&\quad + h^2 \iint \left( \mathcal{I}_1 u + \mathcal{I}_2 u^T x + \frac{1}{2} \mathcal{I}_3 (ux^T + xu^T) \right) K(u) (f_1(x) - f_0(x)) \nabla f_1(x) u dx du + O(h^2) \\
&= \int \left( \mathcal{I}_1 x + \frac{1}{2} \mathcal{I}_2 x^T x + \frac{1}{2} \mathcal{I}_3 (xx^T) \right) (f_1(x) - f_0(x)) f_1(x) dx \\
&\quad - h \mathcal{I}_1 \int u K(u) du \int (f_1(x) - f_0(x)) f_1(x) dx \\
&\quad - h \mathcal{I}_2 \left[ \int u K(u) du \right]^T \int x (f_1(x) - f_0(x)) f_1(x) dx \\
&\quad - \frac{h}{2} \sum_{r,s=1}^d \mathcal{I}_{3,r,s} \int u_r K(u) du \int x_s (f_1(x) - f_0(x)) f_1(x) dx \\
&\quad - \frac{h}{2} \sum_{r,s=1}^d \mathcal{I}_{3,r,s} \int u_s K(u) du \int x_r (f_1(x) - f_0(x)) f_1(x) dx + O(h^2) \\
&= \int \left( \mathcal{I}_1 x + \frac{1}{2} \mathcal{I}_2 x^T x + \frac{1}{2} \mathcal{I}_3 (xx^T) \right) (f_1(x) - f_0(x)) f_1(x) dx + O(h^2) ,
\end{aligned}$$

denn aufgrund der Abschätzungen

$$\begin{aligned}
& O(h^2) \iint \|u\|^2 K(u) \left( \|\mathcal{I}_1\| \|x\| + \frac{1}{2} |\mathcal{I}_2| \|x\|^2 + \frac{1}{2} |\mathcal{I}_3 (xx^T)| \right) |f_1(x) - f_0(x)| dx du \\
& O(h^3) \iint \|u\|^2 K(u) \left( \|\mathcal{I}_1\| \|u\| + \frac{1}{2} |\mathcal{I}_2| \|x\| \|u\| + \frac{1}{2} |\mathcal{I}_3 (ux^T + xu^T)| \right) |f_1(x) - f_0(x)| dx du
\end{aligned}$$

sind die jeweiligen Restterme von der Größenordnung  $O(h^2)$ .

(iii) BEHANDLUNG VON  $\tilde{S}_{n,6}$ :

Es ist

$$\begin{aligned}
\tilde{S}_{n,6} &= \mathbb{E} \left[ \left( \mathcal{I}_1 X_1 + \frac{1}{2} \mathcal{I}_2 X_1^T X_1 + \frac{1}{2} \mathcal{I}_3 (X_1 X_1^T) \right)^2 \right] \\
&= \mathbb{E} \left[ (\mathcal{I}_1 X_1)^2 + \frac{1}{4} \mathcal{I}_2^2 (X_1^T X_1)^2 + \frac{1}{4} \mathcal{I}_3^2 (X_1 X_1^T)^2 + \mathcal{I}_1 X_1 \mathcal{I}_2 X_1^T X_1 \right.
\end{aligned}$$

$$\begin{aligned}
& + \mathcal{I}_1 X_1 \mathcal{I}_3 (X_1 X_1^T) + \frac{1}{2} \mathcal{I}_2 X_1^T X_1 \mathcal{I}_3 (X_1 X_1^T) \Big] \\
= & \mathcal{I}_1 \mathcal{I}_1^T + \frac{1}{4} \mathcal{I}_2^2 \mathbb{E} \left[ (X_1^T X_1)^2 \right] + \frac{1}{4} \mathbb{E} \left[ \mathcal{I}_3 (X_1 X_1^T)^2 \right] + \mathcal{I}_2 \mathcal{I}_1 \mathbb{E} \left[ (X_1^T X_1) X_1 \right] \\
& + \mathcal{I}_1 \mathbb{E} \left[ \mathcal{I}_3 (X_1 X_1^T) X_1 \right] + \frac{1}{2} \mathcal{I}_2 \mathbb{E} \left[ (X_1^T X_1) \mathcal{I}_3 (X_1 X_1^T) \right] .
\end{aligned}$$

3.) Aufgrund von  $\tilde{\sigma}^2(f_1, f_0) = \gamma_2 - \gamma_1^2$  klar nach 1.) und 2.) □

**Lemma 58** Für die in Gleichung (2.54) auf Seite 114 definierten Zufallsvariablen  $\tilde{Y}_{n1}$  gelten mit  $\sigma_R^2$  aus (2.53) bei zugrundeliegenden Dichten  $f_n$  unter der Voraussetzung 1\*

1.)  $\mathbb{E} \tilde{Y}_{n1} = \int \mathfrak{w}(x) f_0(x) dx + o(1),$

2.)  $\mathbb{E} \tilde{Y}_{n1}^2 = \int \mathfrak{w}(x)^2 f_0(x) dx + o(1),$

3.)  $\text{Var}(\tilde{Y}_{n1}) = \sigma_R^2 + o(1).$

BEWEIS: 1.) Mit der Substitution  $u = (x - z)/h, du = h^{-d} dz$  gilt

$$\begin{aligned}
\mathbb{E} \tilde{Y}_{n1} &= \frac{1}{h^d} \iint K \left( \frac{x - z}{h} \right) e_h(x, \mathfrak{w}) f_n(z) dx dz \\
&= \iint K(u) e_h(x, \mathfrak{w}) f_n(x - hu) dx du \\
&= \iint K(u) \mathfrak{w}(x) f_n(x - hu) dx du + O(h^2) ,
\end{aligned}$$

denn nach Lemma 55 auf Seite 136 gilt für den Restterm des Integrals die Abschätzung

$$O(h^2) \iint K(u) f_n(x - hu) dx du = O(h^2) .$$

Die Taylor-Entwicklung von  $f_n(x - hu)$  nach Bemerkung 22 auf Seite 85 liefert mit Voraussetzung 1\*

$$\begin{aligned}
\mathbb{E} \tilde{Y}_{n1} &= \int K(u) du \int \mathfrak{w}(x) f_n(x) dx - h \int \mathfrak{w}(x) \nabla f_n(x) dx \int u K(u) du + O(h^2) \\
&= \int \mathfrak{w}(x) f_0(x) dx + N_n \int \mathfrak{w}(x)^2 dx + O(h^2) \\
&= \int \mathfrak{w}(x) f_0(x) dx + o(1) ,
\end{aligned}$$

denn nach Gleichung (2.25) gilt für den Restterm die Abschätzung

$$O(h^2) \int \|u\|^2 K(u) du \int |\mathfrak{w}(x)| dx = O(h^2) .$$

2.) Mit den Substitutionen  $u = (x - z)/h$ ,  $du = h^{-d}dz$  und  $v = (y - x)/h$ ,  $dv = h^{-d}dy$  gilt

$$\begin{aligned} \mathbb{E} \left[ \tilde{Y}_{n1}^2 \right] &= \frac{1}{h^{2d}} \iiint K \left( \frac{x-z}{h} \right) K \left( \frac{y-z}{h} \right) e_h(x, \mathfrak{w}) e_h(y, \mathfrak{w}) f_n(z) dz dx dy \\ &= \frac{1}{h^d} \iiint K(u) K \left( u + \frac{y-x}{h} \right) e_h(x, \mathfrak{w}) e_h(y, \mathfrak{w}) f_n(x-hu) du dx dy \\ &= \iiint K(u) K(u+v) e_h(x, \mathfrak{w}) e_h(x+hv, \mathfrak{w}) f_n(x-hu) du dx dv \\ &= \iiint K(u) K(u+v) \mathfrak{w}(x) e_h(x+hv, \mathfrak{w}) f_n(x-hu) du dx dv + O(h^2) , \end{aligned}$$

denn nach Lemma 55 gilt für den Restterm die Abschätzung

$$O(h^2) \iint K(u) K(u+v) \int f_n(x-hu) dx du dv = O(h^2) .$$

Eine weitere Anwendung von Lemma 55 liefert unter Berücksichtigung der Gleichungen (2.63) und (2.64) auf Seite 137

$$\begin{aligned} \mathbb{E} \left[ \tilde{Y}_{n1}^2 \right] &= \iint K(u) \left[ \int K(u+v) dv \right] \mathfrak{w}(x)^2 f_n(x-hu) du dx \\ &\quad + h \iint K(u) \mathfrak{w}(x) \nabla \mathfrak{w}(x) \left[ \int v K(u+v) dv \right] f_n(x-hu) du dx + O(h^2) \\ &= \iint K(u) \mathfrak{w}(x)^2 f_n(x-hu) du dx + O(h) , \end{aligned}$$

denn auch hier gilt für den Restterm die Abschätzung

$$O(h^2) \iint K(u) K(u+v) \left( \sum_{l=0}^2 \|v\|^l \right) \int |\mathfrak{w}(x)| f_n(x-hu) dx du dv = O(h^2) .$$

Schließlich liefert eine Taylor–Entwicklung von  $f_n(x-hu)$  gemäß Bemerkung 22 auf Seite 85

$$\begin{aligned} \mathbb{E} \left[ \tilde{Y}_{n1}^2 \right] &= \int K(u) du \int \mathfrak{w}(x)^2 f_n(x) dx - h \int \mathfrak{w}(x)^2 \nabla f_n(x) \left[ \int u K(u) du \right] dx + O(h) \\ &= \int \mathfrak{w}(x)^2 f_n(x) dx + O(h) \\ &= \int \mathfrak{w}(x)^2 f_0(x) dx + N_n \int \mathfrak{w}(x)^3 dx + O(h) \\ &= \int \mathfrak{w}(x)^2 f_0(x) dx + o(1) . \end{aligned}$$

3.) Aufgrund von  $\sigma_R^2 = \int f_0(x) \mathfrak{w}(x)^2 dx - \left( \int f_0(x) \mathfrak{w}(x) dx \right)^2$  klar nach 1.) und 2.) □

# A. Vektor– und Matrixnormen

In diesem Anhang werden einige Tatsachen über Normen zusammengefaßt und Schreibweisen erläutert, die insbesondere im Kapitel 1.2 benötigt werden. Die folgende Definition und Bemerkung kann man z.B. in Stoer (1999), S.209 ff., nachlesen.

**Definition und Bemerkung A.1** Der reelle Vektorraum  $\mathbb{R}^d$  sei mit der Euklidnorm  $\|\cdot\|$  versehen. Als Matrix–Norm auf dem Raum der  $d \times d$ –Matrizen  $\mathbb{R}^{d \times d}$  bietet sich die *Schur–Norm*  $\|\cdot\|_S$  an, d.h. für  $A = (a_{ij})_{1 \leq i, j \leq d}$  gilt

$$\|A\|_S = \left( \sum_{i,j=1}^d a_{ij}^2 \right)^{1/2}.$$

Die Schur–Norm ist mit der Euklidnorm *verträglich*, d.h. für jeden Vektor  $x \in \mathbb{R}^d$  und jede  $d \times d$ –Matrix  $A$  gilt  $\|Ax\| \leq \|A\|_S \|x\|$ .

Die Schur–Norm ist außerdem *submultiplikativ*, d.h. für beliebige  $d \times d$ –Matrizen  $A, B$  gilt  $\|AB\|_S \leq \|A\|_S \|B\|_S$ .

**Bemerkung A.2** Für beliebige Vektoren  $x, y \in \mathbb{R}^d$  gilt  $\|xy^T\|_S = \|x\| \|y\|$ .

BEWEIS: Einsetzen der Definitionen. □

**Definition A.3** Es seien  $\mu$  ein Maß auf einem Meßraum  $(\mathcal{X}, \mathcal{A})$  und  $h : \mathcal{X} \rightarrow \mathbb{R}$  eine reellwertige Funktion.  $g : \mathcal{X} \rightarrow \mathbb{R}^d$ ,  $g := (g_1, \dots, g_d)^T$  bezeichne eine vektorwertige Funktion mit den Komponenten  $g_j : \mathcal{X} \rightarrow \mathbb{R}$  ( $j = 1, \dots, d$ ) und  $A : \mathcal{X} \rightarrow \mathbb{R}^{d \times d}$ ,  $A := (a_{ij})_{i,j=1, \dots, d}$ , eine Matrix mit den Komponenten  $a_{ij} : \mathcal{X} \rightarrow \mathbb{R}$ . Wir schreiben

$$\int gh \, d\mu := \begin{pmatrix} \int g_1 h \, d\mu \\ \vdots \\ \int g_d h \, d\mu \end{pmatrix}$$

für das komponentenweise definierte Integral, sofern die Integrale  $\int g_j h \, d\mu$  ( $j = 1, \dots, d$ ) alle existieren. In gleicher Weise ist

$$\int Ah \, d\mu := \left( \int a_{ij} h \, d\mu \right)_{i,j=1, \dots, d}$$

zu verstehen.

**Lemma A.4** Es seien  $\mu$  ein Maß auf  $(\mathcal{X}, \mathcal{A})$  und  $h : \mathcal{X} \rightarrow \mathbb{R}$  eine reellwertige Funktion mit  $\int h^2 d\mu < \infty$ .

1. Ist  $g : \mathcal{X} \rightarrow \mathbb{R}^d$  eine vektorwertige Funktion mit  $g := (g_1, \dots, g_d)^T$  und  $\int g_j^2 d\mu < \infty$  für jedes  $j = 1, \dots, d$ , so gilt für das komponentenweise definierte Integral  $\int gh d\mu$  das folgende Analogon zur Cauchy–Schwarz–Ungleichung:

$$\left\| \int gh d\mu \right\| \leq \left( \int \|g\|^2 d\mu \int h^2 d\mu \right)^{1/2}.$$

2. Ist  $A : \mathcal{X} \rightarrow \mathbb{R}^{d \times d}$  mit  $A := (a_{ij})_{i,j=1,\dots,d}$  eine Matrix mit in  $x \in \mathcal{X}$  variierenden Komponenten und  $\int a_{ij}^2 d\mu < \infty$  für jedes  $i, j \in \{1, \dots, d\}$ , so gilt für das komponentenweise definierte Integral  $\int Ah d\mu$  das folgende Analogon zur Cauchy–Schwarz–Ungleichung:

$$\left\| \int Ah d\mu \right\|_S \leq \left( \int \|A\|_S^2 d\mu \int h^2 d\mu \right)^{1/2}.$$

BEWEIS:

1. Anwenden der klassischen Cauchy–Schwarz–Ungleichung liefert

$$\left\| \int gh d\mu \right\|^2 = \sum_{j=1}^d \left( \int g_j h d\mu \right)^2 \leq \sum_{j=1}^d \int g_j^2 d\mu \int h^2 d\mu = \int \|g\|^2 d\mu \int h^2 d\mu.$$

2. Anwenden der klassischen Cauchy–Schwarz–Ungleichung liefert

$$\left\| \int Ah d\mu \right\|_S^2 = \sum_{i,j=1}^d \left( \int a_{ij} h d\mu \right)^2 \leq \sum_{i,j=1}^d \int a_{ij}^2 d\mu \int h^2 d\mu = \int \|A\|_S^2 d\mu \int h^2 d\mu.$$

□

**Definition und Bemerkung A.5** Die zu der Euklidnorm  $\|\cdot\|$  des  $\mathbb{R}^d$  passende Matrixnorm auf dem Raum der  $d \times d$ -Matrizen ist für  $A \in \mathbb{R}^{d \times d}$  durch  $\|A\| := \sup_{\|x\| \leq 1} \|Ax\|$  definiert und gleich der *Spektralnorm*, d.h.

$$\|A\| = \left( \lambda_{\max}(A^T A) \right)^{1/2},$$

wobei  $\lambda_{\max}(A^T A)$  den größten Eigenwert der Matrix  $A^T A$  bezeichnet. Die Spektralnorm ist submultiplikativ und unter allen mit der Euklidnorm des  $\mathbb{R}^d$  verträglichen Matrixnormen die kleinste.

---

Für *symmetrische*  $d \times d$ -Matrizen  $A$  gilt

$$\|A\| = |\lambda_{\max}(A)|$$

und für *symmetrische positiv definite* Matrizen  $A$  sogar

$$\|A\| = \lambda_{\max}(A),$$

denn in diesem Fall sind alle Eigenwerte positiv. Die symmetrische positiv definite Wurzel  $A^{1/2}$  einer symmetrischen positiv definiten Matrix  $A$  (vgl. Abschnitt C.2) besitzt als Eigenwerte gerade die Wurzeln der Eigenwerte von  $A$ , woraus sich für ihre Spektralnorm

$$\|A^{1/2}\| = (\|A\|)^{1/2}$$

ergibt. Ferner besitzt die  $d$ -dimensionale Einheitsmatrix  $I_d$  die Spektralnorm  $\|I_d\| = 1$ .

**Bemerkung A.6** Im Endlichdimensionalen sind alle Normen äquivalent. Insbesondere existieren positive Konstanten  $C_1, C_2$ , so daß für jede beliebige  $d \times d$ -Matrix  $A$  gilt:

$$C_1 \|A\| \leq \|A\|_S \leq C_2 \|A\| .$$

# B. Schwache Konvergenz und Gaußprozesse in Hilberträumen

In diesem Anhang sollen grundlegende Tatsachen der Wahrscheinlichkeitstheorie in Hilberträumen (Zufallselemente, schwache Konvergenz, Gaußverteilungen und Zentrale Grenzwertsätze) zusammengestellt werden, soweit sie für die vorliegende Arbeit relevant sind. Dies geschieht nicht immer in der größtmöglichen Allgemeinheit. Die Grundlagen der im folgenden zitierten Tatsachen gehen auf die Dissertation von Mourier (1954) und den bahnbrechenden Artikel von Prokhorov (1956) zurück. Als unmittelbare Quellen seien Heuser (1975), Achieser und Glasmann (1975), Prokhorov (1956), Vakhania (1981), Parthasarathy (1967) und Ledoux und Talagrand (1991) genannt. Die Struktur der nachfolgenden Ausführungen lehnt sich an Chen und White (1992) an.

Bis zum Ende des Kapitels sei die folgende Situation zugrunde gelegt:  $\mathbb{H}$  bezeichne einen reellen separablen Hilbertraum mit dem Skalarprodukt  $\langle \cdot, \cdot \rangle$ , der kanonischen Norm  $\|x\| := \sqrt{\langle x, x \rangle}$  und einer fest gewählten Orthonormalbasis (ONB)  $\{e_j\}_{j \geq 1}$ . Bezeichnet  $\mathbb{H}' := \{f : \mathbb{H} \rightarrow \mathbb{R} \mid f \text{ linear, stetig}\}$  den Dualraum von  $\mathbb{H}$ , so gilt nach dem Darstellungssatz von Fréchet–Riesz:  $\mathbb{H}' = \{f(\cdot) = \langle \cdot, h \rangle \mid h \in \mathbb{H}\}$ . Ferner sei  $\mathbb{H}$  mit der Borel– $\sigma$ –Algebra  $\mathfrak{B}$  versehen.

## B.1. Operatoren auf Hilberträumen

Ein linearer Operator  $S : \mathbb{H} \rightarrow \mathbb{H}$  heißt *symmetrisch*, wenn für alle  $g, h \in \mathbb{H}$  gilt:  $\langle Sg, h \rangle = \langle g, Sh \rangle$ . Nach dem Satz von Hellinger–Toeplitz ist jeder symmetrische Operator eines Hilbertraumes stetig (vgl. Heuser (1975), Satz 68.6).

Ein symmetrischer Operator  $S : \mathbb{H} \rightarrow \mathbb{H}$  heißt *positiv*, wenn für alle  $h \in \mathbb{H}$  gilt:  $\langle Sh, h \rangle \geq 0$ .

Ein linearer Operator  $S : \mathbb{H} \rightarrow \mathbb{H}$  heißt *kompakt* oder *vollstetig*, wenn für jede beschränkte Folge  $(h_n)_{n \geq 1}$  in  $\mathbb{H}$  das Bild  $(Sh_n)_{n \geq 1}$  eine konvergente Teilfolge besitzt. Ein kompakter Operator ist stetig.

Die zu einem linearen stetigen Operator  $S : \mathbb{H} \rightarrow \mathbb{H}$  *adjungierte Abbildung*  $S^* : \mathbb{H} \rightarrow \mathbb{H}$  ist definiert durch die Beziehung  $\langle Sg, h \rangle = \langle g, S^*h \rangle$  für alle  $g, h \in \mathbb{H}$ .  $S^*$  ist ebenfalls eine stetige lineare Abbildung.

Eine Zahl  $\lambda$  heißt *Eigenwert* eines linearen Operators  $S : \mathbb{H} \rightarrow \mathbb{H}$ , wenn ein  $h \in \mathbb{H}$ ,  $h \neq 0$  existiert, so daß  $Sh = \lambda h$  gilt.

Ist  $S : \mathbb{H} \rightarrow \mathbb{H}$  ein kompakter Operator, so ist  $S^*S$  ein positiver kompakter Operator, und

seine von Null verschiedenen Eigenwerte sind positiv. Es bezeichne  $\mu_1^2 \geq \mu_2^2 \geq \mu_3^2 \geq \dots$  die vollständige Folge dieser Eigenwerte ( $\mu_k > 0$  für alle  $k$ ), vgl. Achieser und Glasman (1975), S.178.

**Definition B.1 (Achieser und Glasman (1975), S.178)** Ein kompakter Operator  $S$  heißt *nuklear*, wenn gilt:  $\sum_{k=1}^{\infty} \mu_k < \infty$ .

**Satz B.2 (Achieser und Glasman (1975), S.178–181)** Für einen nuklearen Operator  $S : \mathbb{H} \rightarrow \mathbb{H}$  konvergiert die Summe  $\sum_{k=1}^{\infty} \langle S f_k, f_k \rangle$  absolut für jede ONB  $\{f_k\}_{k \geq 1}$  von  $\mathbb{H}$ , und der Reihenwert hängt nicht von der gewählten ONB ab. Damit ist die *Spur* des Operators definiert als

$$\operatorname{tr}(S) := \sum_{k=1}^{\infty} \langle S f_k, f_k \rangle .$$

Konvergiert umgekehrt für einen stetigen Operator  $S : \mathbb{H} \rightarrow \mathbb{H}$  die Reihe  $\sum_{k=1}^{\infty} \langle S f_k, f_k \rangle$  absolut für eine ONB  $\{f_k\}_{k \geq 1}$  von  $\mathbb{H}$ , so ist  $S$  nuklear.

**Definition B.3 (Vakhania (1981), S. 55/56, bzw. Prokhorov (1956))** Ein symmetrischer, positiver, nuklearer Operator  $S : \mathbb{H} \rightarrow \mathbb{H}$  heißt *S-Operator*, und für die Menge aller S-Operatoren eines Hilbertraumes schreibt man

$$S := \{S : \mathbb{H} \rightarrow \mathbb{H} \mid S \text{ symmetrisch, positiv, nuklear} \} .$$

## B.2. Maße und Zufallselemente auf Hilberträumen

### B.2.1. Erwartungswert und Kovarianzoperator

Ähnlich wie im  $\mathbb{R}^d$  definiert man zu einem  $\mathbb{H}$ -wertigen Zufallselement  $X$  den Erwartungswert und den Kovarianzoperator. Den Erwartungswert bezeichnen wir wieder mit  $EX$ . Bei der Definition des Kovarianzoperators von  $X$  beschränken wir uns auf den Fall  $E \|X\|^2 < \infty$ , was für unsere Zwecke ausreicht.

**Definition und Satz B.4 (Vakhania (1981), S. 55; Parthasarathy (1967), S. 154–155, S. 168)** Sei  $X$  ein  $\mathbb{H}$ -wertiges Zufallselement.

- (i) Gilt  $E \|X\| < \infty$ , so ist der *Erwartungswert*  $EX$  von  $X$  definiert als das Element  $a \in \mathbb{H}$ , das charakterisiert ist durch

$$\forall h \in \mathbb{H} : \langle h, a \rangle = E \langle h, X \rangle . \quad (\text{B.1})$$

$X$  heißt *zentriert*, falls  $a = 0$  gilt.

- (ii) Gilt  $E \|X\|^2 < \infty$ , so ist der *Kovarianzoperator*  $S_X$  von  $X$  der symmetrische positive Operator  $S : \mathbb{H} \rightarrow \mathbb{H}$ , der charakterisiert ist durch

$$\forall h \in \mathbb{H} : \langle Sh, h \rangle = E [\langle h, X - a \rangle^2]. \quad (\text{B.2})$$

**Bemerkung B.5** Für ein  $\mathbb{H}$ -wertiges Zufallselement  $X$  mit  $E \|X\|^2 < \infty$  gilt:

- (i) Der Kovarianzoperator  $S_X$  von  $X$  ist ein  $S$ -Operator.  
(ii)  $E \|X\|^2 = \sum_{k=1}^{\infty} E [\langle X, e_k \rangle^2]$ .  
(iii) Ist  $X$  zusätzlich zentriert, so folgt

$$\text{tr}(S_X) = \sum_{k=1}^{\infty} \langle S_X e_k, e_k \rangle = \sum_{k=1}^{\infty} E [\langle e_k, X \rangle^2] = E \|X\|^2.$$

## B.2.2. Straffheit und Verteilungskonvergenz

Eine Folge  $(\mu_n)_{n \geq 1}$  von Wahrscheinlichkeitsmaßen auf einem Hilbertraum  $(\mathbb{H}, \mathfrak{B})$  heißt *straff*, wenn zu jedem  $\varepsilon > 0$  ein Kompaktum  $K \subset \mathbb{H}$  existiert mit  $\mu_n(K) \geq 1 - \varepsilon$  für alle  $n \geq 1$ . Eine Folge  $(X_n)_{n \geq 1}$  von  $\mathbb{H}$ -wertigen Zufallselementen heißt *straff*, wenn die Folge  $(P^{X_n})_{n \geq 1}$  ihrer Verteilungen straff ist in  $(\mathbb{H}, \mathfrak{B})$ . Für Wahrscheinlichkeitsmaße auf einem Hilbertraum gilt das folgende Kriterium für Straffheit:

**Lemma B.6 (Vakhania (1981), Theorem 3.1.2)** Eine Folge  $(\mu_n)_{n \geq 1}$  von Wahrscheinlichkeitsmaßen auf  $\mathbb{H}$  ist straff, wenn gilt:

- (i)  $\sup_{n \geq 1} \sum_{k=1}^{\infty} \int_{\mathbb{H}} \langle x, e_k \rangle^2 \mu_n(dx) \leq C < \infty$ ,  
(ii)  $\lim_{m \rightarrow \infty} \sup_{n \geq 1} \sum_{k=m}^{\infty} \int_{\mathbb{H}} \langle x, e_k \rangle^2 \mu_n(dx) = 0$ .

In der Formulierung für Zufallselemente lautet die Aussage von Lemma B.6 wie folgt (zur Formulierung von (i) vgl. Bemerkung B.5):

**Lemma B.7** Die Folge  $(X_n)_{n \geq 1}$  von  $\mathbb{H}$ -wertigen Zufallselementen ist straff, wenn gilt:

- (i)  $\sup_{n \geq 1} E [\|X_n\|^2] \leq C < \infty$ ,  
(ii)  $\lim_{m \rightarrow \infty} \sup_{n \geq 1} \sum_{k=m}^{\infty} E [\langle X_n, e_k \rangle^2] = 0$ .

Die schwache Konvergenz von Wahrscheinlichkeitsmaßen auf  $(\mathbb{H}, \mathfrak{B})$  bzw. die Verteilungskonvergenz von  $\mathbb{H}$ -wertigen Zufallselementen ist wie in metrischen Räumen üblich erklärt:

**Definition B.8 (Ledoux und Talagrand (1991), S. 39–41)** (i) Für Wahrscheinlichkeitsmaße  $(\mu_n)_{n \geq 1}$  und  $\mu$  auf  $(\mathbb{H}, \mathfrak{B})$  gilt:

$$\begin{aligned} \mu_n \xrightarrow{\mathcal{D}} \mu &: \iff \mu_n \text{ konvergiert schwach gegen } \mu \\ &: \iff \int f(x) \mu_n(dx) \xrightarrow{n \rightarrow \infty} \int f(x) \mu(dx) \text{ für alle stetigen, beschränkten } f : \mathbb{H} \rightarrow \mathbb{R}. \end{aligned}$$

(ii) Für  $\mathbb{H}$ -wertige Zufallselemente  $(X_n)_{n \geq 1}$  und  $X$  gilt:

$$\begin{aligned} X_n \xrightarrow{\mathcal{D}} X &: \iff X_n \text{ konvergiert nach Verteilung gegen } X \\ &: \iff \mathbb{E}f(X_n) \xrightarrow{n \rightarrow \infty} \mathbb{E}f(X) \text{ für alle stetigen, beschränkten } f : \mathbb{H} \rightarrow \mathbb{R}. \end{aligned}$$

Der Zusammenhang zwischen Straffheit und Verteilungskonvergenz auf einem Hilbertraum ist durch das folgende Lemma gegeben.

**Lemma B.9 (Ledoux und Talagrand (1991), S. 41)** Für  $\mathbb{H}$ -wertige Zufallselemente  $(X_n)_{n \geq 1}$  und  $X$  gilt die folgende Charakterisierung der Verteilungskonvergenz:

$$X_n \xrightarrow{\mathcal{D}} X \iff \begin{aligned} &(i) \quad \forall h \in \mathbb{H} : \langle X_n, h \rangle \xrightarrow{\mathcal{D}} \langle X, h \rangle \text{ in } \mathbb{R}, \\ &(ii) \quad (X_n)_{n \geq 1} \text{ ist straff.} \end{aligned}$$

### B.2.3. Das charakteristische Funktional

**Definition B.10 (Ledoux und Talagrand (1991), S. 39)** (i) Das *charakteristische Funktional*  $\chi : \mathbb{H} \rightarrow \mathbb{C}$  eines Wahrscheinlichkeitsmaßes  $\mu$  auf  $(\mathbb{H}, \mathfrak{B})$  ist definiert durch

$$\chi(h) := \chi(h; \mu) := \int_{\mathbb{H}} e^{i\langle x, h \rangle} \mu(dx) \quad , \quad h \in \mathbb{H}.$$

(ii) Das *charakteristische Funktional*  $\chi : \mathbb{H} \rightarrow \mathbb{C}$  eines  $\mathbb{H}$ -wertigen Zufallselementes  $X$  ist definiert durch

$$\chi(h) := \chi(h; X) := \mathbb{E} [e^{i\langle X, h \rangle}] = \int_{\mathbb{H}} e^{i\langle x, h \rangle} P^X(dx) \quad , \quad h \in \mathbb{H}.$$

**Lemma B.11 (Prokhorov (1956), Lemma 1.6)** Seien  $(\mu_n)_{n \geq 1}$  eine Folge von Wahrscheinlichkeitsmaßen auf einem Hilbertraum  $(\mathbb{H}, \mathfrak{B})$  und  $\chi_n(h) := \chi(h; \mu_n)$  die Folge der zugehörigen charakteristischen Funktionale. Unter den Voraussetzungen

(i)  $(\mu_n)_{n \geq 1}$  ist straff,

(ii)  $\chi_n(h) \rightarrow \chi(h)$  punktweise in  $\mathbb{H}$  bei  $n \rightarrow \infty$  für eine Funktion  $\chi : \mathbb{H} \rightarrow \mathbb{C}$

gilt:  $\mu_n$  konvergiert schwach gegen ein Wahrscheinlichkeitsmaß  $\mu$  auf  $(\mathbb{H}, \mathfrak{B})$ , und es ist  $\chi(h) = \chi(h; \mu)$ .

### B.2.4. Gaußsche Zufallselemente auf Hilberträumen

**Definition und Bemerkung B.12 (Ledoux und Talagrand (1991), S. 54–56)** Ein  $\mathbb{H}$ -wertiges Zufallselement  $X$  heißt *Gaußsches Zufallselement in  $\mathbb{H}$* , wenn  $\langle X, h \rangle$  für jedes  $h \in \mathbb{H}$  eine reelle Gaußsche Zufallsvariable ist.

Der Erwartungswert  $EX$  und der Kovarianzoperator  $S_X$  eines Gaußschen Zufallselementes  $X$  in  $\mathbb{H}$  existieren. Wie im Reellen schreibt man  $\mathcal{N}(EX, S_X)$  für die zugehörige Verteilung.

**Theorem B.13 (Vakhania (1981), S. 55–56)** (i) Das charakteristische Funktional einer Gaußverteilung auf  $\mathbb{H}$  besitzt die Gestalt

$$\chi(h) = e^{i\langle h, a \rangle - \frac{1}{2}\langle Sh, h \rangle}, \quad h \in \mathbb{H} \quad (\text{B.3})$$

mit einem  $a \in \mathbb{H}$  und einem  $S$ -Operator  $S$ . Umgekehrt definiert (B.3) für jedes  $a \in \mathbb{H}$  und jeden  $S$ -Operator  $S$  das charakteristische Funktional einer eindeutig bestimmten Gaußverteilung auf  $\mathbb{H}$ .

(ii)  $a$  und  $S$  sind Erwartungswert und Kovarianzoperator der zugehörigen Gaußverteilung.

(iii) Die Menge  $\mathcal{S}$  der  $S$ -Operatoren auf  $\mathbb{H}$  stimmt mit der Menge der Kovarianzoperatoren von Gaußverteilungen auf  $\mathbb{H}$  überein.

## B.3. Zentrale Grenzwertsätze auf Hilberträumen

**Definition B.14 (Ledoux und Talagrand (1991), S. 272–273)** Seien  $X$  ein  $\mathbb{H}$ -wertiges Zufallselement und  $(X_i)_{i \geq 1}$  eine Folge unabhängiger Kopien von  $X$ . Dann definiert man:

$X$  genügt dem Zentralen Grenzwertsatz (ZGWS) in  $\mathbb{H}$

$$: \iff \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \text{ konvergiert schwach in } \mathbb{H}.$$

**Bemerkung B.15 (Ledoux und Talagrand (1991), S. 273–274)** Wenn ein  $\mathbb{H}$ -wertiges Zufallselement  $X$  dem Zentralen Grenzwertsatz genügt, so folgt:

(i)  $EX = 0$  und  $E\|X\|^2 < \infty$ .

(ii)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow{\mathcal{D}} G$ , wobei  $G$  ein zentriertes Gaußsches Zufallselement in  $\mathbb{H}$  ist mit derselben Kovarianzstruktur wie  $X$ .

Umgekehrt genügt ein  $\mathbb{H}$ -wertiges Zufallselement  $X$  mit  $EX = 0$  und  $E\|X\|^2 < \infty$  dem Zentralen Grenzwertsatz. Separable Hilberträume werden sogar durch diese Eigenschaft charakterisiert, wie das folgende Theorem zeigt. Der dort auftretende separable Banachraum  $B$  sei mit der ebenfalls als  $\|\cdot\|$  notierten Norm sowie mit der Borel- $\sigma$ -Algebra  $\mathfrak{B}$  versehen. Verteilungskonvergenz in  $B$  sei wie üblich, d.h. analog zu Definition B.8 erklärt, und das Erfüllen des Zentralen Grenzwertsatzes analog zu Definition B.14.

**Theorem B.16 (Ledoux und Talagrand (1991), Corollary 10.9)** Ein separabler Banachraum  $B$  ist genau dann isomorph zu einem Hilbertraum, wenn für jedes  $B$ -wertige,  $\mathfrak{B}$ -meßbare Zufallselement  $X$  gilt:

$$\left\{ \begin{array}{l} \mathbb{E}X = 0 \\ \mathbb{E} \|X\|^2 < \infty \end{array} \right\} \iff X \text{ genügt dem ZGWS in } B.$$

Lemma B.7, Lemma B.11 und Theorem B.13 implizieren den folgenden Zentralen Grenzwertsatz für Dreiecksschemata in Hilberträumen. Der Beweis ist in Chen und White (1992) ausgeführt; er besteht im wesentlichen aus dem Nachweis der Existenz des Grenzoperators  $S \in \mathcal{S}$ .

**Theorem B.17 (Chen und White (1992), Corollary 3.1)** Sei  $\{X_{nj}\}_{\substack{1 \leq j \leq r_n \\ n \geq 1}}$  ( $n \in \mathbb{N}$  und  $r_n \rightarrow \infty$  für  $n \rightarrow \infty$ ) ein Dreiecksschema von zentrierten,  $\mathbb{H}$ -wertigen Zufallselementen mit  $\mathbb{E} \|X_{nj}\|^2 < \infty$  für alle  $n, j \in \mathbb{N}$ . Unter den Voraussetzungen

(i)  $\langle \sum_{j=1}^{r_n} X_{nj}, h \rangle \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(h))$  in  $\mathbb{R}$  für alle  $0 \neq h \in \mathbb{H}$ , wobei  $\sigma^2(h) \in (0, \infty)$  von  $h$  abhängt,

(ii)  $\sup_{n \geq 1} \mathbb{E} \left[ \left\| \sum_{j=1}^{r_n} X_{nj} \right\|^2 \right] \leq C < \infty$ ,

(iii)  $\lim_{m \rightarrow \infty} \sup_{n \geq 1} \sum_{k=m}^{\infty} \mathbb{E} \left[ \langle \sum_{j=1}^{r_n} X_{nj}, e_k \rangle^2 \right] = 0$

folgen die Existenz eines  $S \in \mathcal{S}$  mit  $\langle Sh, h \rangle = \sigma^2(h)$  für alle  $0 \neq h \in \mathbb{H}$  und die Verteilungskonvergenz

$$\sum_{j=1}^{r_n} X_{nj} \xrightarrow{\mathcal{D}} \mathcal{N}(0, S) \quad \text{in } \mathbb{H}.$$

## B.4. Gaußprozesse und Gaußsche Zufallselemente in $L^2(\mathbb{R}^d, \mathfrak{B}^d, d\mu)$

Abschließend betrachten wir speziell den reellen separablen Hilbertraum  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$  mit dem Skalarprodukt  $\langle g, h \rangle_{L^2} = \int_{\mathbb{R}^d} g(t)h(t)d\mu(t)$ ,  $g, h \in L^2$ , wobei der  $\mathbb{R}^d$  mit der Borel- $\sigma$ -Algebra  $\mathfrak{B}^d$  versehen sei und  $\mu$  ein  $\sigma$ -endliches Maß auf  $(\mathbb{R}^d, \mathfrak{B}^d)$  bezeichne. In dieser Situation soll geklärt werden, wie ein Gaußprozeß auf  $\mathbb{R}^d$  mit Pfaden in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$  und Gaußsche Zufallselemente mit Werten in dem Hilbertraum  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$  im Sinne der Definitionen aus Abschnitt B.2.4 miteinander zusammenhängen. Die folgenden Überlegungen sind z.B. in Ibragimov und Rozanov (1978), S. 1–13, Gihman und Skorohod (1974) oder Cramér und Leadbetter (1967) zu finden. Alle im folgenden auftretenden Zufallsvariablen und -elemente seien auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  definiert.

Ein *Gaußprozeß*  $Z$  auf dem  $\mathbb{R}^d$  ist eine Familie  $(Z(t))_{t \in \mathbb{R}^d}$  von Zufallsvariablen  $Z(\cdot, t) : \Omega \rightarrow \mathbb{R}$ , deren endlichdimensionale Verteilungen  $(Z(t_1), \dots, Z(t_k))$  für alle  $k \in \mathbb{N}$ ,  $t_1, \dots, t_k \in \mathbb{R}^d$

multivariat normalverteilt sind.  $Z$  ist charakterisiert durch seine Erwartungswertfunktion  $m(t)$ ,  $t \in \mathbb{R}^d$ , und seine Kovarianzfunktion  $B(s, t)$ ,  $s, t \in \mathbb{R}^d$ , mit

$$m(t) := \mathbb{E} [Z(t)] , \quad B(s, t) := \mathbb{E} [(Z(s) - m(s))(Z(t) - m(t))] .$$

Gelten

$$\lim_{s \rightarrow t} [m(s) - m(t)] = 0 \quad \text{und} \quad \lim_{s \rightarrow t} [B(s, s) - 2B(s, t) + B(t, t)] = 0 ,$$

so ist  $Z$  stetig im quadratischen Mittel, d.h. es gilt  $\lim_{s \rightarrow t} \mathbb{E} [(Z(s) - Z(t))^2] = 0$  für jedes  $t \in \mathbb{R}^d$ . Nach Theorem 1 aus Gihman und Skorohod (1974), S. 171, existiert ein  $\mathcal{A} \otimes \mathfrak{B}^d$ -meßbarer stochastischer Prozeß  $\tilde{Z} : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}$ , der äquivalent zu  $Z$  ist, d.h. es gilt  $P(\{\omega \in \Omega : Z(t, \omega) = \tilde{Z}(t, \omega)\}) = 1$  für jedes  $t \in \mathbb{R}^d$ . Ohne Beschränkung der Allgemeinheit nehmen wir im folgenden  $Z$  selbst als meßbar an.

Unter den Bedingungen

$$\int_{\mathbb{R}^d} m^2(t) d\mu(t) < \infty \quad \text{und} \quad \int_{\mathbb{R}^d} B(t, t) d\mu(t) < \infty$$

folgt  $\int_{\mathbb{R}^d} \mathbb{E} [Z(t)^2] d\mu(t) < \infty$ , und nach dem Satz von Fubini liegen fast alle Pfade  $Z(\omega, \cdot)$  in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$ . Für diejenigen  $\omega$  aus einer Nullmenge, für die das nicht gilt, können wir die Pfade  $Z(\omega, \cdot)$  umdefinieren und gehen im folgenden o.B.d.A. von einem meßbaren Prozeß mit Pfaden in  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$  aus.

Dieser Prozeß kann auch als Gaußsches Zufallselement in dem separablen Hilbertraum  $L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$  im Sinne von Abschnitt B.2.4 aufgefaßt werden. Dabei ist die Meßbarkeit klar, und

$$\langle Z, h \rangle_{L^2} = \int_{\mathbb{R}^d} Z(t) h(t) d\mu(t) \tag{B.4}$$

ist für jedes  $h \in L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$  eine reelle normalverteilte Zufallsvariable. Die nötige Argumentation ist z.B. in Ibragimov und Rozanov (1978) auf Seite 12 ausgeführt. Zunächst zeigt man für stetige Funktionen  $h$ , daß das Integral in (B.4) als Grenzwert im quadratischen Mittel definiert werden kann und damit in dem  $L^2$ -Abschluß des Gaußprozesses  $Z$  liegt, also selbst wieder normalverteilt ist. Allgemeine Funktionen  $h \in L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$  müssen zusätzlich in geeigneter Weise durch stetige Funktionen angenähert werden.

Im Lichte dieser Erkenntnisse ergibt sich durch Anwendung des Satzes von Fubini der folgende einfache Zusammenhang zwischen dem Erwartungswert und der Erwartungswertfunktion von  $Z$ : Es gilt für jedes  $h \in L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu)$

$$\mathbb{E} [\langle Z, h \rangle_{L^2}] = \mathbb{E} \left[ \int_{\mathbb{R}^d} Z(t) h(t) d\mu(t) \right] = \int_{\mathbb{R}^d} \mathbb{E} [Z(t)] h(t) d\mu(t) = \langle h, m \rangle_{L^2} ,$$

d.h. der Erwartungswert der Hilbertraum-wertigen Zufallselementes  $Z$  nach (B.1) ist  $\mu$ -fast sicher gleich der Erwartungswertfunktion  $m$ . Für zentriertes  $Z$  ergibt sich ferner der Zusammenhang

$$\mathbb{E} [\langle Z, h \rangle_{L^2}^2] = \mathbb{E} \left[ \int_{\mathbb{R}^d} Z(s) h(s) d\mu(s) \int_{\mathbb{R}^d} Z(t) h(t) d\mu(t) \right]$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} B(s, t) h(s) h(t) d\mu(s) d\mu(t) .$$

Der Kovarianzoperator nach (B.2) ist also durch die Kovarianzfunktion  $B$  eindeutig bestimmt.

# C. Hilfsmittel aus dem Differentialkalkül

## C.1. Der allgemeine Differentialkalkül

Gegenstand dieses Abschnitts ist eine Zusammenfassung der hier benötigten Aussagen des allgemeinen Differentialkalküls in Banachräumen, wie man ihn z.B. in Cartan (1977), S. 28–108, oder Avez (1983), S. 11–46, findet.

Im folgenden seien  $(E, \|\cdot\|_E)$ ,  $(E_1, \|\cdot\|_{E_1})$ ,  $\dots$ ,  $(E_m, \|\cdot\|_{E_m})$ ,  $(F, \|\cdot\|_F)$ ,  $(F_1, \|\cdot\|_{F_1})$ ,  $(F_2, \|\cdot\|_{F_2})$  und  $(G, \|\cdot\|_G)$  Banachräume.  $U$  bezeichne eine offene Teilmenge von  $E$ . Der Produktraum  $\tilde{E} := E_1 \times \dots \times E_m$  sei mit der Supremumsnorm  $\|(x_1, \dots, x_m)\|_{\tilde{E}} := \sup_{i=1, \dots, m} \|x_i\|_{E_i}$  versehen. Der Raum

$$\mathfrak{L}(E; F) := \{f : E \rightarrow F \mid f \text{ linear, stetig}\}$$

ist mit der üblichen Norm  $\|f\|_{\mathfrak{L}(E; F)} := \sup_{x \neq 0} \frac{\|f(x)\|_F}{\|x\|_E} = \sup_{\|x\|_E \leq 1} \|f(x)\|_F$  ein Banachraum, da  $F$  ein Banachraum ist. Entsprechend ist

$$\mathfrak{L}(E_1, \dots, E_m; F) := \{f : E_1 \times \dots \times E_m \rightarrow F \mid f \text{ } m\text{-fach linear und stetig}\}$$

mit der Norm

$$\|f\|_{\mathfrak{L}(E_1, \dots, E_m; F)} = \sup_{x_i \neq 0} \frac{\|f(x_1, \dots, x_m)\|_F}{\|x_1\|_{E_1} \cdots \|x_m\|_{E_m}} = \sup_{\|x_1\|_{E_1} \leq 1, \dots, \|x_m\|_{E_m} \leq 1} \|f(x_1, \dots, x_m)\|_F$$

versehen, und für alle  $(x_1, \dots, x_m) \in E_1 \times \dots \times E_m$  gilt die Abschätzung

$$\|f(x_1, \dots, x_m)\|_F \leq \|f\|_{\mathfrak{L}(E_1, \dots, E_m; F)} \cdot \|x_1\|_{E_1} \cdots \|x_m\|_{E_m}. \quad (\text{C.1})$$

Ferner gilt für jedes  $l \in \{1, \dots, m\}$  die Normisomorphie

$$\mathfrak{L}(E_1, \dots, E_m; F) \cong \mathfrak{L}(E_1, \dots, E_l; \mathfrak{L}(E_{l+1}, \dots, E_m; F)).$$

**Definition und Bemerkung C.1** Es seien  $U$  eine offene Teilmenge von  $E$  und  $a \in U$ .  $f : U \rightarrow F$  heißt *differenzierbar im Punkt  $a$* , falls eine lineare stetige Abbildung  $L : E \rightarrow F$  existiert mit:

$$\lim_{x \rightarrow a} \frac{\|f(x) - f(a) - L(x - a)\|_F}{\|x - a\|_E} = 0.$$

In diesem Fall schreibt man für die Abbildung  $L$  auch  $df(a)$  und nennt sie das *Differential* (*Fréchet-Ableitung*) von  $f$  im Punkt  $a$ . Die Funktion  $f$  heißt *differenzierbar auf  $U$* , wenn sie in jedem Punkt  $a \in U$  differenzierbar ist.  $f$  heißt *stetig differenzierbar auf  $U$*  ( $f \in C^1(U)$ ), wenn in jedem Punkt  $a \in U$  die Funktion  $f$  differenzierbar und die Funktion

$$df : \begin{cases} U & \rightarrow \mathfrak{L}(E; F) \\ a & \mapsto df(a) \end{cases}$$

stetig ist. Durch  $df(a) \circ v := df(a)(v)$  ist das Bild von  $v \in E$  unter der Abbildung  $df(a) \in \mathfrak{L}(E; F)$  oder die *Richtungsableitung von  $f$  im Punkt  $a \in E$  in Richtung  $v$*  erklärt.

**Bemerkung C.2 (Linearität der Ableitung)** Seien  $U$  eine offene Teilmenge von  $E$ ,  $f, g : U \rightarrow F$  differenzierbar im Punkt  $a \in U$  und  $\lambda, \mu \in \mathbb{R}$ . Dann ist  $\lambda f + \mu g$  differenzierbar in  $a$ , und es gilt

$$d(\lambda f + \mu g)(a) = \lambda df(a) + \mu dg(a).$$

**Definition und Bemerkung C.3** Eine Funktion  $f : U \rightarrow F$  auf einer offenen Teilmenge  $U$  von  $E$  heißt *zweifach differenzierbar in einem Punkt  $a \in U$* , wenn  $f$  in einer Umgebung von  $a$  und  $df$  im Punkt  $a$  differenzierbar ist.  $f$  heißt *zweifach differenzierbar auf  $U$* , wenn  $f$  in jedem Punkt  $a \in U$  zweifach differenzierbar ist. Im Falle der zweifachen Differenzierbarkeit der Funktion  $f$  auf  $U$  schreibt man

$$d^2 f := d(df) : U \rightarrow \mathfrak{L}(E; \mathfrak{L}(E; F)) \cong \mathfrak{L}(E, E; F)$$

und bemerkt, daß für  $a \in U$  die Funktion  $d^2 f(a) \in \mathfrak{L}(E; \mathfrak{L}(E; F)) \cong \mathfrak{L}(E, E; F)$  linear und stetig ist. Analog zur ersten Ableitung von  $f$  bezeichnet man mit

$$d^2 f(a) \circ (h_1, h_2) := d^2 f(a)(h_1) \circ h_2$$

den Funktionswert von  $d^2 f(a) \in \mathfrak{L}(E, E; F)$  im Punkt  $(h_1, h_2) \in E \times E$ . Die Funktion  $f$  heißt *zweimal stetig differenzierbar auf  $U$*  ( $f \in C^2(U)$ ), wenn  $f$  und  $df$  dort differenzierbar und  $d^2 f$  dort stetig ist.

Die Definition der Differenzierbarkeit einer Funktion  $f$  läßt sich wie in dem Schritt von der ersten zur zweiten Ableitung per Induktion auf  $n$ -fache Differenzierbarkeit verallgemeinern. Falls  $f$  im Punkt  $a \in U$   $n$ -fach differenzierbar ist, setzt man

$$d^n f(a) := d(d^{n-1} f)(a) \in \mathfrak{L}^n(E; F) := \mathfrak{L}(E; \underbrace{\mathfrak{L}(E, \dots, E; F)}_{n-1}) \cong \mathfrak{L}(\underbrace{E, \dots, E}_n; F)$$

und  $d^n f(a) \circ (h_1, \dots, h_n)$  für den Funktionswert im Punkt  $(h_1, \dots, h_n) \in E \times \dots \times E$ . Die Klasse  $C^n(U)$  umfaßt die differenzierbaren Funktionen  $f$ , deren Ableitungen bis zur Ordnung  $n - 1$  ebenfalls differenzierbar auf  $U$  sind und für die außerdem  $d^n f$  dort stetig ist. Ferner ist  $C^\infty(U) := \bigcap_{n \in \mathbb{N}} C^n(U)$ .

**Satz C.4 (Satz von Schwarz)** Wenn  $f : U \subset E \rightarrow F$   $n$ -mal differenzierbar im Punkt  $a \in U$  ist, so ist  $d^n f(a) \in \mathfrak{L}^n(E; F)$  eine  $n$ -lineare symmetrische Abbildung, d.h. für alle  $h_1, \dots, h_n \in E$  und für jede Permutation  $\tau$  der Zahlen  $\{1, \dots, n\}$  gilt

$$d^n f(a) \circ (h_1, \dots, h_n) = d^n f(a) \circ (h_{\tau(1)}, \dots, h_{\tau(n)}) .$$

Wie im Reellen lassen sich auch partielle Ableitungen einer Funktion  $f$  erklären.

**Definition und Bemerkung C.5 (Cartan (1977), S. 38, 67)** Es seien  $U$  eine offene Teilmenge von  $E = E_1 \times \dots \times E_m$  und  $f : U \rightarrow F$  eine stetige Abbildung. Für jedes  $a = (a_1, \dots, a_m) \in U$  betrachten wir die durch

$$\lambda_i : \begin{cases} E_i & \rightarrow E \\ x_i & \mapsto (a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_m) \end{cases} , \quad i = 1, \dots, m ,$$

definierte  $i$ -te Injektion. Die Verkettung  $f \circ \lambda_i$  ist definiert auf der offenen Menge  $(\lambda_i)^{-1}(U) \subset E_i$ , die  $a_i \in E_i$  enthält.

Ist  $f$  differenzierbar im Punkt  $a$ , so ist jedes  $f \circ \lambda_i$  differenzierbar im Punkt  $a_i$ . Die Ableitung  $d_i f(a)$  dieser Funktion heißt  $i$ -te partielle Ableitung von  $f$  im Punkt  $a$  und liegt in  $\mathfrak{L}(E_i; F)$ . Ferner gilt

$$df(a) \circ (h_1, \dots, h_m) = \sum_{i=1}^m d_i f(a) \circ h_i$$

für  $h_i \in E_i$  ( $i = 1, \dots, m$ ). Ist  $f$  zweimal differenzierbar in  $a$ , so gilt für  $h_i, k_i \in E_i$

$$\begin{aligned} (d^2 f(a) \circ (k_1, \dots, k_m)) \circ (h_1, \dots, h_m) &= \sum_{i=1}^m (d_i(df)(a) \circ k_i) \circ (h_1, \dots, h_m) \\ &= \sum_{i,j=1}^m (d_{ij}^2 f(a) \circ k_i) \circ h_j , \end{aligned}$$

wobei  $d_{ij}^2 f(a)$  den Wert von  $d_i(d_j f)$  im Punkt  $a$  bezeichnet. Dieser ist ein Element aus  $\mathfrak{L}(E_i; \mathfrak{L}(E_j; F))$ , und es gilt  $d_{ij}^2 f(a) = d_{ji}^2 f(a)$  für alle  $i$  und  $j$ . Im Fall der dreifachen Differenzierbarkeit von  $f$  in  $a$  gilt analog für  $h_i, k_i, \xi_i \in E_i$

$$((d^3 f(a) \circ (\xi_1, \dots, \xi_m)) \circ (k_1, \dots, k_m)) \circ (h_1, \dots, h_m) = \sum_{i,j,l=1}^m ((d_{ijl}^3 f(a) \circ \xi_i) \circ k_j) \circ h_l$$

mit  $d_{ijl}^3 f(a) = d_{\tau(i)\tau(j)\tau(l)}^3 f(a)$  für alle Permutationen  $\tau$  der drei Indizes.

Im folgenden werden einige Rechenregeln zusammengefaßt.

**Satz C.6 (Konstanten; Cartan (1977), Prop. 2.4.1)** Eine konstante Funktion  $f : U \rightarrow F$  auf einer offenen Teilmenge  $U$  von  $E$  ist differenzierbar auf  $U$  mit  $df(a) = 0$  für jedes  $a \in U$ .

**Satz C.7 (Lineare Abbildungen; Cartan (1977), Prop. 2.4.2)** Eine lineare stetige Funktion  $f : U \rightarrow F$  auf einer offenen Teilmenge  $U$  von  $E$  ist in  $C^\infty(U)$  mit  $df(a) = f$  für jedes  $a \in U$  und  $d^n f(a) = 0$  für  $n \geq 2$  und  $a \in U$ .

**Satz C.8 (Bilineare Abbildungen; Cartan (1977), Prop. 2.4.3)** Eine bilineare stetige Funktion  $b : E_1 \times E_2 \rightarrow F$  liegt in  $C^\infty(E_1 \times E_2)$  mit

$$\begin{aligned} db(a_1, a_2) \circ (h_1, h_2) &= b(h_1, a_2) + b(a_1, h_2), \\ d^2 b(a_1, a_2) \circ ((h_1, h_2), (k_1, k_2)) &= b(h_1, k_2) + b(k_1, h_2), \end{aligned}$$

$a_1, h_1, k_1 \in E_1, a_2, h_2, k_2 \in E_2$ , und  $d^n b = 0$  für  $n > 2$ .

**Satz C.9 (Multilineare Abbildungen; Cartan (1977), S. 34)** Eine multilineare stetige Funktion  $f : E_1 \times \dots \times E_n \rightarrow F$  liegt in  $C^\infty(E_1 \times \dots \times E_n)$  mit

$$\begin{aligned} df(a_1, \dots, a_n) \circ (h_1, \dots, h_n) &= f(h_1, a_2, \dots, a_n) + f(a_1, h_2, a_3, \dots, a_n) \\ &\quad + \dots + f(a_1, \dots, a_{n-1}, h_n). \end{aligned}$$

**Satz C.10 (Kettenregel; Cartan (1977), Th. 2.2.1, Fraenkel (1978), S. 161)** Seien  $f : U \rightarrow F$  und  $g : V \rightarrow G$  stetige Funktionen auf einer offenen Teilmenge  $U$  von  $E$  bzw. einer offenen Teilmenge  $V$  von  $F$ . Ferner seien  $f$  differenzierbar in  $a \in U$  ( $n$ -mal differenzierbar in  $a$ ;  $\in C^n(U)$ ) und  $g$  differenzierbar in  $f(a) \in V$  ( $n$ -mal differenzierbar in  $f(a)$ ;  $\in C^n(V)$ ).

Dann ist die Verkettung  $g \circ f : f^{-1}(V) \cap U \rightarrow G$  differenzierbar in  $a$  ( $n$ -mal differenzierbar in  $a$ ;  $\in C^n(f^{-1}(V) \cap U)$ ), und es gilt, sofern die folgenden Ableitungen definiert sind:

$$d(g \circ f)(a) \circ h_1 = dg(f(a)) \circ df(a) \circ h_1,$$

$$\begin{aligned} d^2(g \circ f)(a) \circ (h_1, h_2) \\ = dg(f(a)) \circ d^2 f(a) \circ (h_1, h_2) + d^2 g(f(a)) \circ (df(a) \circ h_1, df(a) \circ h_2), \end{aligned}$$

$$\begin{aligned} d^n(g \circ f)(a) \circ (h_1, \dots, h_n) \\ = \sum_{l=1}^n \sum_{\substack{r_1 + \dots + r_l = n \\ r_j \geq 1 \forall j=1, \dots, l}} \frac{1}{l! r_1! \dots r_l!} \sum_{\tau \in \text{Per}\{1, \dots, n\}} \\ d^l g(f(a)) \circ (d^{r_1} f(a) \circ (h_{\tau(1)}, \dots, h_{\tau(r_1)}), \dots, d^{r_l} f(a) \circ (h_{\tau(n-r_l+1)}, \dots, h_{\tau(n)})). \end{aligned}$$

**Satz C.11 (Formel von Leibniz; Avez (1983), S. 15 und 37)** Es seien  $U$  eine offene Teilmenge von  $E$ ,  $a \in U$ ,  $f : U \rightarrow F_1$  differenzierbar in  $a$  ( $\in C^1(U)$ ,  $\in C^n(U)$ ) und  $g : U \rightarrow F_2$  differenzierbar in  $a$  ( $\in C^1(U)$ ,  $\in C^n(U)$ ).  $b : F_1 \times F_2 \rightarrow G$  sei bilinear und stetig. Dann ist

$$p : \begin{cases} U & \rightarrow G \\ x & \mapsto b(f(x), g(x)) \end{cases}$$

differenzierbar in  $a$  ( $\in C^1(U)$ ,  $\in C^n(U)$ ) mit

$$\begin{aligned} dp(a) \circ h_1 &= b(df(a) \circ h_1, g(a)) + b(f(a), dg(a) \circ h_1) \\ d^2 p(a) \circ (h_1, h_2) &= b(f(a), d^2 g(a) \circ (h_1, h_2)) + b(df(a) \circ h_1, dg(a) \circ h_2) + b(df(a) \circ h_2, dg(a) \circ h_1) \\ &\quad + b(d^2 f(a) \circ (h_1, h_2), g(a)) \\ d^n p(a) \circ (h_1, \dots, h_n) &= \sum_{q=0}^n \sum b(d^q f(a) \circ (h_{i_1}, \dots, h_{i_q}), d^{n-q} g(a) \circ (h_{j_1}, \dots, h_{j_{n-q}})), \end{aligned}$$

wobei sich die zweite Summe in der letzten Zeile über die  $\binom{n}{q}$  möglichen Einteilungen der Indizes  $\{1, \dots, n\}$  in zwei Teilmengen  $i_1 < \dots < i_q$  und  $j_1 < \dots < j_{n-q}$  erstreckt.

**Satz C.12 (Inverse eines Isomorphismus, Cartan (1977), Th. 2.4.4 und S. 74)** Für zwei Banachräume  $E$  und  $F$  ist

$$\text{Isom}(E, F) := \{u : E \rightarrow F \mid u \text{ linear, stetig, bijektiv}\}$$

eine offene Teilmenge von  $\mathcal{L}(E; F)$ , und die Abbildung

$$\varphi : \begin{cases} \text{Isom}(E, F) & \rightarrow \text{Isom}(F, E) \subset \mathcal{L}(F; E) \\ u & \mapsto u^{-1} \end{cases}$$

liegt in  $C^\infty(\text{Isom}(E, F))$ . Die Ableitung  $d\varphi(u)$  liegt in  $\mathcal{L}(\mathcal{L}(E; F), \mathcal{L}(F; E))$  und ist gegeben durch

$$d\varphi(u) \circ h = -u^{-1} \circ h \circ u^{-1}, \quad h \in \mathcal{L}(E; F).$$

Die höheren Ableitungen von  $\varphi$  sind mit  $h_i \in \mathcal{L}(E; F)$  gegeben durch

$$\begin{aligned} d^2 \varphi(u) \circ (h_1, h_2) &= u^{-1} \circ h_1 \circ u^{-1} \circ h_2 \circ u^{-1} + u^{-1} \circ h_2 \circ u^{-1} \circ h_1 \circ u^{-1}, \\ d^n \varphi(u) \circ (h_1, \dots, h_n) &= (-1)^n \sum_{\tau \in \text{Per}\{1, \dots, n\}} u^{-1} \circ h_{\tau(1)} \circ u^{-1} \circ \dots \circ u^{-1} \circ h_{\tau(n)} \circ u^{-1}. \end{aligned}$$

Wie im Reellen stellt sich die Frage nach den Glattheitseigenschaften der Inversen einer Abbildung bzw. nach lokaler Umkehrbarkeit.

**Definition C.13** Eine Funktion  $f : U \rightarrow V$  von einer offenen Menge  $U \subset E$  in eine offene Menge  $V \subset F$  heißt  $C^1$ -Diffeomorphismus, wenn  $f : U \rightarrow V$  bijektiv und von der Klasse  $C^1$  ist (als Abbildung von  $U$  in  $F$  betrachtet) und wenn zusätzlich die Umkehrabbildung  $f^{-1} : V \rightarrow U$  von der Klasse  $C^1$  ist (als Abbildung von  $V$  in  $E$  betrachtet).

**Satz C.14 (Cartan (1977), Th. 5.4.4)** Es seien  $U$  eine offene Teilmenge eines Banachraums  $E$ ,  $V$  eine offene Teilmenge eines Banachraums  $F$  und  $f : U \rightarrow V$  ein  $C^1$ -Diffeomorphismus. Wenn  $f$  in  $C^n$  liegt, so liegt auch  $f^{-1}$  in  $C^n$ .  $f$  heißt dann  $C^n$ -Diffeomorphismus.

**Satz C.15 (Umkehrabbildung; Cartan (1977), Th. 4.2.1, Cor. 5.4.5, Avez (1983), S. 27, 30)** Es seien  $E, F$  Banachräume und  $U$  eine offene Teilmenge von  $E$ . Ferner sei  $f : U \rightarrow F$  eine stetig differenzierbare Funktion ( $f \in C^n$ ) mit  $df(a) \in \text{Isom}(E, F)$ ,  $a \in U$ . Dann existieren offene Umgebungen  $I$  von  $a$  ( $I \subset U$ ) und  $J$  von  $f(a)$ , so daß  $f : I \rightarrow J$  ein  $C^1$ -Diffeomorphismus ( $C^n$ -Diffeomorphismus) ist. Desweiteren ist  $df(x) \in \text{Isom}(E, F)$  für alle  $x \in I$ , und es gilt

$$df^{-1}(f(x)) = [df(x)]^{-1}$$

für jedes  $x \in I$ . Anders ausgedrückt: Für jedes  $y \in J$  gilt

$$df^{-1}(y) = [df(f^{-1}(y))]^{-1} = \varphi \circ df \circ f^{-1}(y)$$

mit der Abbildung  $\varphi$  aus Satz C.12.

**Satz C.16 (Taylor mit Integralrest; Avez (1983), S. 42)** Es seien  $E$  und  $F$  Banachräume,  $U$  eine offene Teilmenge von  $E$  und  $f : U \rightarrow F$  eine Funktion der Klasse  $C^{n+1}(U)$ . Wenn das Segment  $[a, a+h]$  in  $U$  liegt, so gilt die Entwicklung

$$f(a+h) - f(a) = df(a) \circ h + \frac{1}{2} d^2 f(a) \circ (h, h) + \dots + \frac{1}{n!} d^n f(a) \circ \underbrace{(h, \dots, h)}_{n \text{ mal}} + R_n,$$

wobei das Restglied  $R_n$  die folgende Integraldarstellung besitzt:

$$R_n := \int_0^1 \frac{1}{n!} (1-t)^n d^{n+1} f(a+th) \circ \underbrace{(h, \dots, h)}_{n+1 \text{ mal}} dt. \quad (\text{C.2})$$

(zur Definition des Integrals vgl. Avez (1983)).

## C.2. Differentiation spezieller Funktionen

Der in Abschnitt C.1 vorgestellte Differentialkalkül soll nun auf den für uns interessanten Spezialfall angewandt werden.

Hierzu seien  $M(d \times d, \mathbb{R})$  der reelle Vektorraum der reellen  $d \times d$ -Matrizen, versehen mit der Spektralnorm (vgl. Definition A.5 in Anhang A), und  $\text{Symm}$  der Untervektorraum der reellen symmetrischen  $d \times d$ -Matrizen. Da bei symmetrischen Matrizen z.B. das untere Dreieck redundant ist, ist  $\text{Symm}$  isomorph zum  $\mathbb{R}^{d(d+1)/2}$ . Ferner bezeichne

$$SP_d := \{A \in \text{Symm} \mid A \text{ positiv definit}\}$$

die Menge der symmetrischen positiv definiten Matrizen.  $SP_d$  ist eine offene Teilmenge von  $Symm$ .

Nach Brieskorn (1985), S. 470–476 und S. 486–488, kann man zu einer symmetrischen positiv definiten Matrix  $B$  eine eindeutige Quadratwurzel  $B^{1/2}$  definieren, die wiederum symmetrisch und positiv definit ist. Die Abbildungen

$$\text{sq} : \begin{cases} SP_d & \rightarrow & SP_d \\ A & \mapsto & A^2 \end{cases} , \quad \text{sqr} : \begin{cases} SP_d & \rightarrow & SP_d \\ A & \mapsto & A^{1/2} \end{cases}$$

sind auf der offenen Teilmenge  $SP_d$  des Vektorraums  $Symm$  definiert, bijektiv und Umkehrabbildungen zueinander (vgl. Brieskorn (1985), S. 486–488). Insbesondere liegt  $I_d$  in  $SP_d$ , und man kann die Ableitungen von  $\text{sq}$ ,  $\text{sqr}$  und Verkettungen dieser Funktionen in einer Umgebung von  $I_d$  betrachten. Um diese zu berechnen, wenden wir den Differentialkalkül aus Abschnitt C.1 auf  $SP_d \subset Symm$  an.

### C.2.1. Die Funktion $\text{sq}$

Die Abbildung  $\text{sq}$  läßt sich schreiben in der Form  $\text{sq}(A) = A^2 = b(\text{id}(A), \text{id}(A))$  als Verkettung der bilinearen Funktion

$$b : \begin{cases} \mathbb{R}^{d \times d} \times \mathbb{R}^{d \times d} & \rightarrow & \mathbb{R}^{d \times d} \\ (A, B) & \mapsto & A \cdot B \end{cases}$$

mit der Identität  $\text{id}$  auf  $SP_d$ . Für diese gilt nach Satz C.7  $d(\text{id})(A) = \text{id}$  und  $d^2(\text{id})(A) = 0$  für jede  $d \times d$ -Matrix  $A$ . Nach Satz C.11 ist  $\text{sq} \in C^\infty(SP_d)$ , und der Wert der Ableitung  $d(\text{sq})(A) \in \mathcal{L}(Symm; Symm)$  berechnet sich zu

$$d(\text{sq})(A) \circ C = b(\text{id}(C), \text{id}(A)) + b(\text{id}(A), \text{id}(C)) = C \cdot A + A \cdot C \quad , \quad C \in Symm .$$

Insbesondere ergibt sich

$$d(\text{sq})(I_d) \circ C = 2C \quad , \quad [d(\text{sq})(I_d)]^{-1} \circ C = \frac{1}{2}C \quad , \quad C \in Symm .$$

Die zweite Ableitung  $d^2(\text{sq})(A) \in \mathcal{L}(Symm, Symm; Symm)$  berechnet sich ebenfalls nach Satz C.11 zu

$$\begin{aligned} d^2(\text{sq})(A) \circ (C, D) &= b(\text{id}(A), d^2(\text{id})(A) \circ (C, D)) + b(d(\text{id})(A) \circ C, d(\text{id})(A) \circ D) \\ &\quad + b(d(\text{id})(A) \circ D, d(\text{id})(A) \circ C) + b(d^2(\text{id})(A) \circ (C, D), \text{id}(A)) \\ &= C \cdot D + D \cdot C \quad , \end{aligned}$$

$C, D \in Symm$ . Insbesondere ergibt sich

$$d^2(\text{sq})(I_d) \circ (C, D) = C \cdot D + D \cdot C \quad , \quad C, D \in Symm .$$

### C.2.2. Die Funktion $\text{sqr}$

Die Funktion  $\text{sq}$  ist auf der offenen Teilmenge  $SP_d$  des Banachraums  $\text{Symm}$  beliebig oft stetig differenzierbar, und es gilt  $d(\text{sq})(I_d) = 2id \in \text{Isom}(\text{Symm}, \text{Symm})$ . Nach Satz C.15 existieren offene Umgebungen  $\mathfrak{J}, \mathfrak{I} \subset SP_d$  von  $I_d$ , so daß

$$\text{sq} : \begin{cases} \mathfrak{J} & \rightarrow \mathfrak{I} \\ A & \mapsto A^2 \end{cases}$$

ein  $C^n$ -Diffeomorphismus ist für beliebig großes  $n$ . Für die Ableitung der Umkehrfunktion  $\text{sqr} = (\text{sq})^{-1} : \mathfrak{I} \rightarrow \mathfrak{J}$  in einem Punkt  $A \in \mathfrak{I}$  gilt  $d(\text{sqr})(A) \in \mathcal{L}(\text{Symm}; \text{Symm})$  mit

$$d(\text{sqr})(A) = [d(\text{sq})(A^{1/2})]^{-1}$$

und insbesondere

$$d(\text{sqr})(I_d) \circ C = [d(\text{sq})(I_d)]^{-1} \circ C = \frac{1}{2} C, \quad C \in \text{Symm}.$$

Mit

$$\varphi : \begin{cases} \text{Isom}(\text{Symm}, \text{Symm}) & \rightarrow \text{Isom}(\text{Symm}, \text{Symm}) \\ u & \mapsto u^{-1} \end{cases}$$

ergibt sich die weitere Darstellung

$$d(\text{sqr})(A) = \varphi \circ d(\text{sq}) \circ \text{sqr}(A),$$

die es mit Hilfe von Satz C.10 (Kettenregel) erlaubt, die zweite Ableitung zu berechnen:

$$\begin{aligned} d^2(\text{sqr})(A) &= d\varphi(d(\text{sq})(\text{sqr}(A))) \circ d^2(\text{sq})(\text{sqr}(A)) \circ d(\text{sqr})(A) \\ &= d\varphi(d(\text{sq})(A^{1/2})) \circ d^2(\text{sq})(A^{1/2}) \circ d(\text{sqr})(A) \in \mathcal{L}(\text{Symm}; \mathcal{L}(\text{Symm}; \text{Symm})). \end{aligned}$$

Mit einem Argument  $B \in \text{Symm}$  ergibt sich nach Satz C.12

$$\begin{aligned} d^2(\text{sqr})(A) \circ B &= d\varphi(d(\text{sq})(A^{1/2})) \circ [d^2(\text{sq})(A^{1/2}) \circ d(\text{sqr})(A) \circ B] \\ &= -[d(\text{sq})(A^{1/2})]^{-1} \circ [d^2(\text{sq})(A^{1/2}) \circ d(\text{sqr})(A) \circ B] \circ [d(\text{sq})(A^{1/2})]^{-1} \\ &= -d(\text{sqr})(A) \circ [d^2(\text{sq})(A^{1/2}) \circ d(\text{sqr})(A) \circ B] \circ d(\text{sqr})(A) \end{aligned}$$

und schließlich mit  $B, C \in \text{Symm}$

$$\begin{aligned} d^2(\text{sqr})(A) \circ (B, C) &= -d(\text{sqr})(A) \circ d^2(\text{sq})(A^{1/2}) \circ (d(\text{sqr})(A) \circ B, d(\text{sqr})(A) \circ C) \\ &= -d(\text{sqr})(A) \circ [(d(\text{sqr})(A) \circ B) \cdot (d(\text{sqr})(A) \circ C) + (d(\text{sqr})(A) \circ C) \cdot (d(\text{sqr})(A) \circ B)]. \end{aligned}$$

An der Stelle  $A = I_d$  besitzt die zweite Ableitung folglich den Wert

$$d^2(\text{sqr})(I_d) \circ (B, C) = -d(\text{sqr})(I_d) \circ \left[ \frac{1}{2} B \frac{1}{2} C + \frac{1}{2} C \frac{1}{2} B \right] = -\frac{1}{8} [B \cdot C + C \cdot B].$$

### C.2.3. Die Funktion $\Upsilon$

Die Abbildung

$$\varphi : \begin{cases} SP_d & \rightarrow SP_d \\ A & \mapsto A^{-1} \end{cases}$$

ist nach Satz C.12 beliebig oft stetig differenzierbar, ebenso wie die Funktion  $\text{sqr} : \mathfrak{J} \rightarrow \mathfrak{J}$  aus Abschnitt C.2.2 auf einer Umgebung  $\mathfrak{J} \subset SP_d$  von  $I_d$ . Damit ist nach Satz C.10 auch die Verkettung  $\Upsilon := \varphi \circ \text{sqr}$ , d.h.

$$\Upsilon : \begin{cases} \mathfrak{J} & \xrightarrow{\text{sqr}} \mathfrak{J} & \xrightarrow{\varphi} SP_d \\ A & \mapsto A^{1/2} & \mapsto A^{-1/2} \end{cases}$$

beliebig oft stetig differenzierbar. Für ihre Ableitung  $d\Upsilon(A) \in \mathfrak{L}(\text{Symm}; \text{Symm})$  gilt nach den Sätzen C.10 und C.12

$$d\Upsilon(A) \circ B = d\varphi(A^{1/2}) \circ d(\text{sqr})(A) \circ B = -A^{-1/2} \cdot (d(\text{sqr})(A) \circ B) \cdot A^{-1/2}$$

und speziell

$$d\Upsilon(I_d) \circ B = -d(\text{sqr})(I_d) \circ B = -\frac{1}{2} B.$$

Die zweite Ableitung  $d^2\Upsilon(A) \in \mathfrak{L}(\text{Symm}, \text{Symm}; \text{Symm})$  berechnet sich ebenfalls nach den Sätzen C.10 und C.12 zu

$$\begin{aligned} d^2\Upsilon(A) \circ (B, C) &= d\varphi(\text{sqr}(A)) \circ (d^2(\text{sqr})(A) \circ (B, C)) \\ &\quad + d^2\varphi(\text{sqr}(A)) \circ (d(\text{sqr})(A) \circ B, d(\text{sqr})(A) \circ C) \\ &= -A^{-1/2} \cdot (d^2(\text{sqr})(A) \circ (B, C)) \cdot A^{-1/2} \\ &\quad + A^{-1/2} \cdot (d(\text{sqr})(A) \circ B) \cdot A^{-1/2} \cdot (d(\text{sqr})(A) \circ C) \cdot A^{-1/2} \\ &\quad + A^{-1/2} \cdot (d(\text{sqr})(A) \circ C) \cdot A^{-1/2} \cdot (d(\text{sqr})(A) \circ B) \cdot A^{-1/2} \end{aligned}$$

und speziell

$$\begin{aligned} d^2\Upsilon(I_d) \circ (B, C) &= -d^2(\text{sqr})(I_d) \circ (B, C) \\ &\quad + (d(\text{sqr})(I_d) \circ B) \cdot (d(\text{sqr})(I_d) \circ C) + (d(\text{sqr})(I_d) \circ C) \cdot (d(\text{sqr})(I_d) \circ B) \\ &= \frac{1}{8} [B \cdot C + C \cdot B] + \frac{1}{2} B \frac{1}{2} C + \frac{1}{2} C \frac{1}{2} B \\ &= \frac{3}{8} [B \cdot C + C \cdot B]. \end{aligned}$$

### C.2.4. Verkettungen der bisher betrachteten Funktionen

Für feste Vektoren  $v, w \in \mathbb{R}^d$  sind die Abbildungen

$$g : \begin{cases} \mathbb{R}^{d \times d} & \rightarrow \mathbb{R}^d \\ A & \mapsto Av + w \end{cases}, \quad \tilde{g} : \begin{cases} \mathbb{R}^{d \times d} & \rightarrow \mathbb{R}^d \\ A & \mapsto Av \end{cases}$$

affin bzw. linear und damit beliebig oft stetig differenzierbar. Damit ist für jedes  $A \in \mathbb{R}^{d \times d}$  die Ableitung von  $g$  durch  $dg(A) = \tilde{g}$  gegeben, d.h. es gilt  $dg(A) \circ C = C v$  für  $C \in \mathbb{R}^{d \times d}$ .

Die Verkettung  $\tilde{h} := g \circ \Upsilon$ , die durch

$$\tilde{h} : \begin{cases} \tilde{\mathcal{J}} & \rightarrow & \mathbb{R}^d \\ \Sigma & \mapsto & \Sigma^{-1/2} v + w \end{cases}$$

auf der offenen Umgebung  $\tilde{\mathcal{J}} \subset SP_d$  von  $I_d$  erklärt ist, ist damit ebenfalls beliebig oft stetig differenzierbar und besitzt die Ableitung

$$d\tilde{h}(\Sigma) = dg(\Upsilon(\Sigma)) \circ d\Upsilon(\Sigma) \in \mathfrak{L}(\text{Symm}, \mathbb{R}^d).$$

Angewandt auf ein Argument  $C \in \text{Symm}$  ergibt sich

$$d\tilde{h}(\Sigma) \circ C = dg(\Sigma^{-1/2}) \circ d\Upsilon(\Sigma) \circ C = (d\Upsilon(\Sigma) \circ C) v$$

und speziell

$$d\tilde{h}(I_d) \circ C = -\frac{1}{2} C v.$$

### C.3. Taylor–Entwicklung der Funktion $K_h$

Gegenstand der folgenden Überlegungen ist es, eine Taylorentwicklung des Kerndichteschätzers der standardisierten Daten aus Abschnitt 2.2 (vgl. Bemerkung 27 auf Seite 87) in den Parametern herzuleiten, mit dem Ziel, die mit der Standardisierung der Daten einhergehenden Störparameter herauszurechnen. Dabei werden partielle Ableitungen verwendet, die anhand des in den Abschnitten C.1 und C.2 zusammengestellten Differentialkalküls berechnet werden. Hierfür benötigen wir zunächst einige Notationen.

Wir betrachten für festes  $x, X_i \in \mathbb{R}^d$  und  $h > 0$  die *Standardisierung*

$$\text{st}(x, X_i, \mu, \Sigma) := \frac{1}{h} (x - \Sigma^{-1/2} (X_i - \mu))$$

als Funktion in  $\mu \in \mathbb{R}^d$  und  $\Sigma \in SP_d$ , d.h. in der Form

$$\text{st} : \begin{cases} \mathbb{R}^d \times SP_d & \rightarrow & \mathbb{R}^d \\ (\mu, \Sigma) & \mapsto & \frac{1}{h} (x - \Sigma^{-1/2} (X_i - \mu)) \end{cases}.$$

Die partiellen Ableitungen nach  $\mu$  und  $\Sigma$  im Sinne von Definition C.5 aus Abschnitt C.1 bezeichnen wir im folgenden mit  $d_\mu$  bzw.  $d_\Sigma$ .

Anhand der Darstellung

$$\text{st}(x, X_i, \mu, \Sigma) = \frac{1}{h} \Sigma^{-1/2} \mu + \frac{1}{h} (x - \Sigma^{-1/2} X_i)$$

sieht man, daß  $\text{st}$  eine affine Funktion in  $\mu$  ist. Damit ist nach Satz C.7 die Ableitung  $d_\mu \text{st}(x, X_i, \mu, \Sigma) \in \mathcal{L}(\mathbb{R}^d; \mathbb{R})$  an einer Stelle  $\xi \in \mathbb{R}^d$  gegeben durch

$$d_\mu \text{st}(x, X_i, \mu, \Sigma) \circ \xi = \frac{1}{h} \Sigma^{-1/2} \xi,$$

und für die zweite Ableitung gilt  $d_{\mu\mu} \text{st}(x, X_i, \mu, \Sigma) \circ (\xi_1, \xi_2) = 0$  für alle  $(\xi_1, \xi_2) \in \mathbb{R}^d \times \mathbb{R}^d$ .

Um  $\text{st}$  als Funktion von  $\Sigma$  in einer Umgebung von  $I_d$  zu untersuchen, ist die Darstellung

$$\text{st}(x, X_i, \mu, \Sigma) = \Sigma^{-1/2} \left( \frac{1}{h} (\mu - X_i) \right) + \frac{1}{h} x$$

adäquat, die zeigt, daß  $\text{st}$  vom Typ  $\tilde{h} = g \circ \Upsilon$  aus Abschnitt C.2.4 ist. Damit ist die Ableitung  $d_\Sigma \text{st}(x, X_i, \mu, \Sigma) \in \mathcal{L}(\text{Symm}; \mathbb{R}^d)$  an einer Stelle  $C \in \text{Symm}$  gegeben durch

$$d_\Sigma \text{st}(x, X_i, \mu, \Sigma) \circ C = (d\Upsilon(\Sigma) \circ C) \frac{1}{h} (\mu - X_i).$$

Es sei nun  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  eine dreimal stetig differenzierbare Funktion. Wir interessieren uns für die Verkettung  $K \circ \text{st}$  als Funktion in  $(\mu, \Sigma)$  bei festem  $x, X_i \in \mathbb{R}^d$  und  $h \in \mathbb{R}$ , d.h. für die Abbildung

$$K_h : \begin{cases} \mathbb{R}^d \times \text{SP}_d & \rightarrow & \mathbb{R} \\ (\mu, \Sigma) & \mapsto & K_h(x, X_i, \mu, \Sigma) := K(\text{st}(x, X_i, \mu, \Sigma)) \end{cases}.$$

Im den folgenden Abschnitten werden die partiellen Ableitungen erster, zweiter und dritter Ordnung von  $K_h$  nach  $\mu$  bzw.  $\Sigma$  berechnet, mit dem Ziel, zu der oben erwähnten Taylor-Entwicklung zu gelangen. Dabei gehen wir stets davon aus, daß die Matrix  $\Sigma$  in einer hinreichend kleinen Umgebung der Einheitsmatrix  $I_d$  liegt.

### Ableitungen erster Ordnung

Die erste Ableitung von  $K_h$  nach  $\mu$  in einem Punkt  $(\mu, \Sigma)$  liegt in  $\mathcal{L}(\mathbb{R}^d; \mathbb{R})$ , und ihr Wert in einem Vektor  $\xi \in \mathbb{R}^d$  berechnet sich nach der Kettenregel zu

$$\begin{aligned} d_\mu K_h(x, X_i, \mu, \Sigma) \circ \xi &= \nabla K(\text{st}(x, X_i, \mu, \Sigma)) \circ d_\mu \text{st}(x, X_i, \mu, \Sigma) \circ \xi \\ &= \frac{1}{h} \nabla K(\text{st}(x, X_i, \mu, \Sigma)) \Sigma^{-1/2} \xi, \end{aligned}$$

wobei wir  $\nabla K(\cdot)$  als ein Element von  $\mathcal{L}(\mathbb{R}^d; \mathbb{R})$ , d.h. als Zeilenvektor auffassen. Speziell ergibt sich

$$d_\mu K_h(x, X_i, 0, I_d) \circ \xi = \frac{1}{h} \nabla K \left( \frac{x - X_i}{h} \right) \xi.$$

Die erste Ableitung von  $K_h$  nach  $\Sigma$  in einem Punkt  $(\mu, \Sigma)$  liegt in  $\mathcal{L}(\text{Symm}; \mathbb{R})$ , und ihr Wert in einem Argument  $C \in \text{Symm}$  berechnet sich nach der Kettenregel zu

$$d_\Sigma K_h(x, X_i, \mu, \Sigma) \circ C = \nabla K(\text{st}(x, X_i, \mu, \Sigma)) \circ d_\Sigma \text{st}(x, X_i, \mu, \Sigma) \circ C$$

$$= \frac{1}{h} \nabla K(\text{st}(x, X_i, \mu, \Sigma)) \circ (d\Upsilon(\Sigma) \circ C) (\mu - X_i) .$$

Speziell ergibt sich

$$d_{\Sigma} K_h(x, X_i, 0, I_d) \circ C = \frac{1}{2h} \nabla K \left( \frac{x - X_i}{h} \right) C X_i .$$

### Ableitungen zweiter Ordnung

Die Ableitung von  $K_h$  zweimal nach  $\mu$  in einem Punkt  $(\mu, \Sigma)$  liegt in  $\mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R})$ , und ihr Wert in einer Richtung  $(\xi_1, \xi_2) \in \mathbb{R}^d \times \mathbb{R}^d$  berechnet sich nach der Kettenregel (zweite Ableitung in Satz C.10) zu

$$\begin{aligned} d_{\mu\mu}^2 K_h(x, X_i, \mu, \Sigma) \circ (\xi_1, \xi_2) &= \nabla K(\text{st}(x, X_i, \mu, \Sigma)) \circ d_{\mu\mu}^2 \text{st}(x, X_i, \mu, \Sigma) \circ (\xi_1, \xi_2) \\ &\quad + \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \circ (d_{\mu} \text{st}(x, X_i, \mu, \Sigma) \circ \xi_1, d_{\mu} \text{st}(x, X_i, \mu, \Sigma) \circ \xi_2) \\ &= \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \circ \left( \frac{1}{h} \Sigma^{-1/2} \xi_1, \frac{1}{h} \Sigma^{-1/2} \xi_2 \right) \\ &= \frac{1}{h^2} \xi_1^T \Sigma^{-1/2} \cdot \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \cdot \Sigma^{-1/2} \xi_2 . \end{aligned}$$

Dabei bezeichnet  $\nabla^2 K$  die Hesse–Matrix der reellwertigen Funktion  $K$ . Speziell ergibt sich

$$d_{\mu\mu}^2 K_h(x, X_i, 0, I_d) \circ (\xi_1, \xi_2) = \frac{1}{h^2} \xi_1^T \nabla^2 K \left( \frac{x - X_i}{h} \right) \xi_2 .$$

Die gemischte Ableitung  $d_{\Sigma\mu}^2 K_h(x, X_i, \mu, \Sigma)$  berechnen wir anhand der ersten Ableitung von  $K_h$  nach  $\mu$

$$d_{\mu} K_h(x, X_i, \mu, \Sigma) = \frac{1}{h} \nabla K(\text{st}(x, X_i, \mu, \Sigma)) \Sigma^{-1/2} \in \mathfrak{L}(\mathbb{R}^d; \mathbb{R}) ,$$

die wir nun als Funktion in  $\Sigma \in SP_d$  betrachten. Um ihre Ableitung nach  $\Sigma$  zu berechnen, schreiben wir  $d_{\mu} K_h(x, X_i, \mu, \Sigma) = b(\nabla K(\text{st}(x, X_i, \mu, \Sigma)), \Upsilon(\Sigma))$  mit der bilinearen Funktion

$$b : \begin{cases} \mathfrak{L}(\mathbb{R}^d; \mathbb{R}) \times \text{Symm} & \rightarrow \mathfrak{L}(\mathbb{R}^d; \mathbb{R}) \\ (x^T, A) & \mapsto \frac{1}{h} x^T A \end{cases} .$$

Die Ableitung von  $d_{\mu} K_h(x, X_i, \mu, \Sigma)$  nach  $\Sigma$  liegt in  $\mathfrak{L}(\text{Symm}; \mathfrak{L}(\mathbb{R}^d; \mathbb{R}))$  und berechnet sich nach den Sätzen C.10 und C.11 für ein Argument  $C \in \text{Symm}$  zu

$$\begin{aligned} d_{\Sigma\mu}^2 K_h(x, X_i, \mu, \Sigma) \circ C &= b(\nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \circ d_{\Sigma} \text{st}(x, X_i, \mu, \Sigma) \circ C, \Upsilon(\Sigma)) \\ &\quad + b(\nabla K(\text{st}(x, X_i, \mu, \Sigma)), d\Upsilon(\Sigma) \circ C) \\ &= \frac{1}{h} \left[ \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \circ (d\Upsilon(\Sigma) \circ C) \frac{1}{h} (\mu - X_i) \right]^T \Upsilon(\Sigma) \end{aligned}$$

$$+\frac{1}{h}\nabla K(\text{st}(x, X_i, \mu, \Sigma)) (d\Upsilon(\Sigma) \circ C) .$$

Folglich gilt für  $C \in \text{Symm}$  und  $\xi \in \mathbb{R}^d$

$$\begin{aligned} (d_{\Sigma\mu}^2 K_h(x, X_i, \mu, \Sigma) \circ C) \circ \xi &= \frac{1}{h^2}(\mu - X_i)^T [d\Upsilon(\Sigma) \circ C] \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \Sigma^{-1/2} \xi \\ &\quad + \frac{1}{h} \nabla K(\text{st}(x, X_i, \mu, \Sigma)) [d\Upsilon(\Sigma) \circ C] \xi \end{aligned}$$

und speziell

$$(d_{\Sigma\mu}^2 K_h(x, X_i, 0, I_d) \circ C) \circ \xi = \frac{1}{2h^2} X_i^T C \nabla^2 K \left( \frac{x - X_i}{h} \right) \xi - \frac{1}{2h} \nabla K \left( \frac{x - X_i}{h} \right) C \xi .$$

Die zweifache Ableitung von  $K_h$  nach  $\Sigma$  berechnet sich anhand der Darstellung  $d_{\Sigma} K_h(x, X_i, \mu, \Sigma) = b(\nabla K(\text{st}(x, X_i, \mu, \Sigma)), d\Upsilon(\Sigma)) \in \mathfrak{L}(\text{Symm}; \mathbb{R})$  mit der bilinearen Funktion

$$b : \begin{cases} \mathfrak{L}(\mathbb{R}^d; \mathbb{R}) \times \mathfrak{L}(\text{Symm}; \text{Symm}) & \rightarrow \mathfrak{L}(\text{Symm}; \mathbb{R}) \\ (x^T, L) & \mapsto x^T (L \circ (\cdot)) \frac{1}{h} (\mu - X_i) \end{cases}$$

ebenfalls nach den Sätzen C.10 und C.11 für ein Argument  $C \in \text{Symm}$  zu

$$\begin{aligned} &d_{\Sigma\Sigma}^2 K_h(x, X_i, \mu, \Sigma) \circ C \\ &= b(\nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \circ d_{\Sigma} \text{st}(x, X_i, \mu, \Sigma) \circ C, d\Upsilon(\Sigma)) \\ &\quad + b(\nabla K(\text{st}(x, X_i, \mu, \Sigma)), d^2 \Upsilon(\Sigma) \circ C) \\ &= \left[ \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \circ (d\Upsilon(\Sigma) \circ C) \frac{1}{h} (\mu - X_i) \right]^T [d\Upsilon(\Sigma) \circ (\cdot)] \frac{1}{h} (\mu - X_i) \\ &\quad + \nabla K(\text{st}(x, X_i, \mu, \Sigma)) [d^2 \Upsilon(\Sigma) \circ (C, \cdot)] \frac{1}{h} (\mu - X_i) . \end{aligned}$$

Damit ergibt sich für  $C, D \in \text{Symm}$

$$\begin{aligned} &d_{\Sigma\Sigma}^2 K_h(x, X_i, \mu, \Sigma) \circ (C, D) \\ &= \frac{1}{h^2} (\mu - X_i)^T [d\Upsilon(\Sigma) \circ C] \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) [d\Upsilon(\Sigma) \circ D] (\mu - X_i) \\ &\quad + \frac{1}{h} \nabla K(\text{st}(x, X_i, \mu, \Sigma)) [d^2 \Upsilon(\Sigma) \circ (C, D)] (\mu - X_i) \end{aligned}$$

und speziell

$$\begin{aligned} &d_{\Sigma\Sigma}^2 K_h(x, X_i, 0, I_d) \circ (C, D) \\ &= \frac{1}{4h^2} X_i^T C \nabla^2 K \left( \frac{x - X_i}{h} \right) D X_i - \frac{3}{8h} \nabla K \left( \frac{x - X_i}{h} \right) (CD + DC) X_i . \end{aligned}$$

### Ableitungen dritter Ordnung

Die dreifache Ableitung von  $K_h$  nach  $\mu$  berechnet sich anhand der Darstellung  $d_{\mu\mu}^2 K_h(x, X_i, \mu, \Sigma) = L(\nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)))$  mit der linearen Funktion

$$L : \begin{cases} \mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}) & \rightarrow \mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}) \\ A & \mapsto \frac{1}{h^2} \Sigma^{-1/2} A \Sigma^{-1/2} \end{cases}$$

unter Beachtung von  $dL(A) = L$  leicht nach der Kettenregel. An dieser Stelle sei darauf hingewiesen, daß  $\nabla^3 K(x)$  das Differential dritter Ordnung von  $K$  bezeichnet und daß  $d_{\mu\mu\mu}^3 K_h(x, X_i, \mu, \Sigma)$  und  $\nabla^3 K(x)$  beide in  $\mathfrak{L}(\mathbb{R}^d; \mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}))$  liegen. Für  $\xi \in \mathbb{R}^d$  gilt mit der Kettenregel

$$\begin{aligned} d_{\mu\mu\mu}^3 K_h(x, X_i, \mu, \Sigma) \circ \xi &= dL(\nabla^2 K(\text{st}(x, X_i, \mu, \Sigma))) \circ \nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ \frac{1}{h} \Sigma^{-1/2} \xi \\ &= \frac{1}{h^3} \Sigma^{-1/2} [\nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ \Sigma^{-1/2} \xi] \Sigma^{-1/2} \end{aligned}$$

und damit für  $\xi_1, \xi_2, \xi_3 \in \mathbb{R}^d$

$$d_{\mu\mu\mu}^3 K_h(x, X_i, \mu, \Sigma) \circ (\xi_1, \xi_2, \xi_3) = \frac{1}{h^3} \xi_2^T \Sigma^{-1/2} [\nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ \Sigma^{-1/2} \xi_1] \Sigma^{-1/2} \xi_3.$$

Die Ableitung  $d_{\Sigma\mu\mu}^3 K_h(x, X_i, \mu, \Sigma)$  liegt in  $\mathfrak{L}(\text{Symm}; \mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}))$  und läßt sich z.B. durch Betrachten von  $d_{\mu\mu}^2 K_h(x, X_i, \mu, \Sigma)$  als Funktion von  $\Sigma$  berechnen. Dies geschieht anhand der Darstellung  $d_{\mu\mu}^2 K_h(x, X_i, \mu, \Sigma) = L(\Upsilon(\Sigma), \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), \Upsilon(\Sigma))$  mit der trilinearen Funktion

$$L : \begin{cases} \text{Symm} \times \mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}) \times \text{Symm} & \rightarrow \mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}) \\ (C, D, E) & \mapsto \frac{1}{h^2} CDE \end{cases}.$$

Damit folgt nach Satz C.9 für eine symmetrische Matrix  $B$

$$\begin{aligned} d_{\Sigma\mu\mu}^3 K_h(x, X_i, \mu, \Sigma) \circ B &= L(d\Upsilon(\Sigma) \circ B, \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), \Upsilon(\Sigma)) \\ &\quad + L\left(\Upsilon(\Sigma), \nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ [d\Upsilon(\Sigma) \circ B] \frac{1}{h} (\mu - X_i), \Upsilon(\Sigma)\right) \\ &\quad + L(\Upsilon(\Sigma), \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), d\Upsilon(\Sigma) \circ B) \\ &= \frac{1}{h^2} [d\Upsilon(\Sigma) \circ B] \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \Upsilon(\Sigma) + \frac{1}{h^2} \Upsilon(\Sigma) \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) [d\Upsilon(\Sigma) \circ B] \\ &\quad + \frac{1}{h^3} \Upsilon(\Sigma) \{ \nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ [d\Upsilon(\Sigma) \circ B] (\mu - X_i) \} \Upsilon(\Sigma) \end{aligned}$$

und für  $\xi_1, \xi_2 \in \mathbb{R}^d$

$$\begin{aligned} (d_{\Sigma\mu\mu}^3 K_h(x, X_i, \mu, \Sigma) \circ B) \circ (\xi_1, \xi_2) &= \frac{1}{h^2} \xi_1^T [d\Upsilon(\Sigma) \circ B] \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \Sigma^{-1/2} \xi_2 \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{h^2} \xi_1^T \Sigma^{-1/2} \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) [d\Upsilon(\Sigma) \circ B] \xi_2 \\
& + \frac{1}{h^3} \xi_1^T \Sigma^{-1/2} \left\{ \nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ [d\Upsilon(\Sigma) \circ B] (\mu - X_i) \right\} \Sigma^{-1/2} \xi_2 .
\end{aligned}$$

Die Ableitung  $d_{\mu\Sigma\Sigma}^3 K_h(x, X_i, \mu, \Sigma)$  liegt in  $\mathfrak{L}(\mathbb{R}^d; \mathfrak{L}(\text{Symm}, \text{Symm}; \mathbb{R}))$  und läßt sich z.B. durch Betrachten von  $d_{\Sigma\Sigma}^2 K_h(x, X_i, \mu, \Sigma)$  als Funktion von  $\mu$  berechnen. Dies geschieht anhand der Darstellung

$$\begin{aligned}
d_{\Sigma\Sigma}^2 K_h(x, X_i, \mu, \Sigma) \\
= L_1(\mu - X_i, \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), \mu - X_i) + L_2(\nabla K(\text{st}(x, X_i, \mu, \Sigma)), \mu - X_i)
\end{aligned}$$

mit der trilinearen Funktion

$$L_1 : \begin{cases} \mathfrak{L}(\mathbb{R}^d, \mathbb{R}) \times \mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}) \times \mathbb{R}^d & \rightarrow \mathfrak{L}(\text{Symm}, \text{Symm}; \mathbb{R}) \\ (x^T, A, y) & \mapsto \frac{1}{h^2} x^T [d\Upsilon(\Sigma) \circ (\cdot)] A [d\Upsilon(\Sigma) \circ (\cdot)] y \end{cases}$$

und der bilinearen Funktion

$$L_2 : \begin{cases} \mathfrak{L}(\mathbb{R}^d; \mathbb{R}) \times \mathbb{R}^d & \rightarrow \mathfrak{L}(\text{Symm}, \text{Symm}; \mathbb{R}) \\ (x^T, y) & \mapsto \frac{1}{h} x^T [d^2 \Upsilon(\Sigma) \circ (\cdot, \cdot)] y \end{cases} .$$

Damit folgt nach Satz C.9 für einen Vektor  $\xi \in \mathbb{R}^d$  unter Beachtung von  $d(\text{id} - X_i)(\mu) = \text{id}$  für alle  $\mu$

$$\begin{aligned}
d_{\mu\Sigma\Sigma}^3 K_h(x, X_i, \mu, \Sigma) \circ \xi \\
= L_1(\xi, \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), \mu - X_i) + L_1(\mu - X_i, \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), \xi) \\
+ L_1\left(\mu - X_i, \nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ \frac{1}{h} \Sigma^{-1/2} \xi, \mu - X_i\right) \\
+ L_2\left(\nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \circ \frac{1}{h} \Sigma^{-1/2} \xi, \mu - X_i\right) + L_2(\nabla K(\text{st}(x, X_i, \mu, \Sigma)), \xi)
\end{aligned}$$

sowie für  $B, C \in \text{Symm}$

$$\begin{aligned}
& (d_{\mu\Sigma\Sigma}^3 K_h(x, X_i, \mu, \Sigma) \circ \xi) \circ (B, C) \\
& = \frac{1}{h^3} (\mu - X_i)^T [d\Upsilon(\Sigma) \circ B] \left\{ \nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ \Sigma^{-1/2} \xi \right\} [d\Upsilon(\Sigma) \circ C] (\mu - X_i) \\
& + \frac{1}{h^2} \xi^T [d\Upsilon(\Sigma) \circ B] \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) [d\Upsilon(\Sigma) \circ C] (\mu - X_i) \\
& + \frac{1}{h^2} (\mu - X_i)^T [d\Upsilon(\Sigma) \circ B] \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) [d\Upsilon(\Sigma) \circ C] \xi \\
& + \frac{1}{h^2} \xi^T \Sigma^{-1/2} \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) [d^2 \Upsilon(\Sigma) \circ (B, C)] (\mu - X_i) \\
& + \frac{1}{h} \nabla K(\text{st}(x, X_i, \mu, \Sigma)) [d^2 \Upsilon(\Sigma) \circ (B, C)] \xi .
\end{aligned}$$

Die Ableitung  $d_{\Sigma\Sigma\Sigma}^3 K_h(x, X_i, \mu, \Sigma)$  liegt in  $\mathfrak{L}(\text{Symm}; \mathfrak{L}(\text{Symm}, \text{Symm}; \mathbb{R}))$  und lässt sich durch Ableiten von  $d_{\Sigma\Sigma}^2 K_h(x, X_i, \mu, \Sigma)$  nach  $\Sigma$  berechnen. Dies geschieht anhand der Darstellung

$$d_{\Sigma\Sigma}^2 K_h(x, X_i, \mu, \Sigma) = L_1(d\Upsilon(\Sigma) \circ (\cdot), \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), d\Upsilon(\Sigma) \circ (\cdot)) \\ + L_2(\nabla K(\text{st}(x, X_i, \mu, \Sigma)), d^2\Upsilon(\Sigma) \circ (\cdot, \cdot))$$

mit der trilinearen Funktion

$$L_1 : \begin{cases} \mathfrak{L}(\text{Symm}; \text{Symm}) \times \mathfrak{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}) \times \mathfrak{L}(\text{Symm}; \text{Symm}) & \rightarrow \mathfrak{L}(\text{Symm}, \text{Symm}; \mathbb{R}) \\ (P_1, A, P_2) & \mapsto \frac{1}{h^2}(\mu - X_i)^T P_1(\cdot) A P_2(\cdot) (\mu - X_i) \end{cases}$$

und der bilinearen Funktion

$$L_2 : \begin{cases} \mathfrak{L}(\mathbb{R}^d; \mathbb{R}) \times \mathfrak{L}(\text{Symm}, \text{Symm}; \text{Symm}) & \rightarrow \mathfrak{L}(\text{Symm}, \text{Symm}; \mathbb{R}) \\ (x^T, P_3) & \mapsto \frac{1}{h} x^T P_3(\cdot, \cdot) (\mu - X_i) \end{cases} .$$

Damit folgt nach Satz C.9 für  $B, C, D \in \text{Symm}$

$$d_{\Sigma\Sigma\Sigma}^3 K_h(x, X_i, \mu, \Sigma) \circ (B, C, D) \\ = L_1(d^2\Upsilon(\Sigma) \circ (B, C), \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), d\Upsilon(\Sigma) \circ D) \\ + L_1\left(d\Upsilon(\Sigma) \circ B, \nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ [d\Upsilon(\Sigma) \circ C] \frac{1}{h}(\mu - X_i), d\Upsilon(\Sigma) \circ D\right) \\ + L_1(d\Upsilon(\Sigma) \circ B, \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)), d^2\Upsilon(\Sigma) \circ (C, D)) \\ + L_2\left(\nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) \circ [d\Upsilon(\Sigma) \circ B] \frac{1}{h}(\mu - X_i), d^2\Upsilon(\Sigma) \circ (C, D)\right) \\ + L_2(\nabla K(\text{st}(x, X_i, \mu, \Sigma)), d^3\Upsilon(\Sigma) \circ (B, C, D)) \\ = \frac{1}{h^3}(\mu - X_i)^T [d\Upsilon(\Sigma) \circ B] \{ \nabla^3 K(\text{st}(x, X_i, \mu, \Sigma)) \circ [d\Upsilon(\Sigma) \circ C] (\mu - X_i) \} [d\Upsilon(\Sigma) \circ D] (\mu - X_i) \\ + \frac{1}{h^2}(\mu - X_i)^T [d^2\Upsilon(\Sigma) \circ (B, C)] \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) [d\Upsilon(\Sigma) \circ D] (\mu - X_i) \\ + \frac{2}{h^2}(\mu - X_i)^T [d\Upsilon(\Sigma) \circ B] \nabla^2 K(\text{st}(x, X_i, \mu, \Sigma)) [d^2\Upsilon(\Sigma) \circ (C, D)] (\mu - X_i) \\ + \frac{1}{h} \nabla K(\text{st}(x, X_i, \mu, \Sigma)) \{ d^3\Upsilon(\Sigma) \circ (B, C, D) \} (\mu - X_i) .$$

### Taylor–Entwicklung der Funktion $K_h$

Nun stehen alle Hilfsmittel bereit, um bei festem  $x, X_i \in \mathbb{R}^d$  eine Taylor–Entwicklung ersten bzw. zweiten Grades der Funktion

$$K_h : \begin{cases} \mathbb{R}^d \times \text{SP}_d & \rightarrow \mathbb{R} \\ \vartheta := (\mu, \Sigma) & \mapsto K_h(x, X_i, \vartheta) := K_h(x, X_i, \mu, \Sigma) = K(\text{st}(x, X_i, \mu, \Sigma)) \end{cases}$$

in  $\vartheta = (\mu, \Sigma)$  um den Punkt  $\vartheta_0 := (0, I_d)$  anzugeben. Mit  $E_1 \times E_2 := \mathbb{R}^d \times \text{SP}_d$  und  $F := \mathbb{R}$  ist die Taylor–Entwicklung ersten Grades nach Satz C.16 gegeben durch

$$K_h(x, X_i, \vartheta) - K_h(x, X_i, \vartheta_0) = dK_h(x, X_i, \vartheta_0) \circ (\vartheta - \vartheta_0) + R_1 \quad (\text{C.3})$$

mit dem Restglied

$$R_1 = \int_0^1 (1-t) d^2 K_h(x, X_i, \vartheta_0 + t(\vartheta - \vartheta_0)) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0) dt \quad (\text{C.4})$$

und die zweiten Grades durch

$$\begin{aligned} K_h(x, X_i, \vartheta) - K_h(x, X_i, \vartheta_0) \\ = dK_h(x, X_i, \vartheta_0) \circ (\vartheta - \vartheta_0) + \frac{1}{2} d^2 K_h(x, X_i, \vartheta_0) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0) + R_2 \end{aligned} \quad (\text{C.5})$$

mit dem Restglied

$$R_2 = \int_0^1 \frac{1}{2} (1-t)^2 d^3 K_h(x, X_i, \vartheta_0 + t(\vartheta - \vartheta_0)) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0, \vartheta - \vartheta_0) dt. \quad (\text{C.6})$$

Dabei sind die Ableitungen in (C.3) und (C.5) nach Bemerkung C.5 und den bisherigen Ausführungen in diesem Abschnitt gegeben durch

$$\begin{aligned} dK_h(x, X_i, \vartheta_0) \circ (\vartheta - \vartheta_0) &= d_\mu K_h(x, X_i, \vartheta_0) \circ \mu + d_\Sigma K_h(x, X_i, \vartheta_0) \circ (\Sigma - I_d) \quad (\text{C.7}) \\ &= \frac{1}{h} \nabla K \left( \frac{x - X_i}{h} \right) \mu + \frac{1}{2h} \nabla K \left( \frac{x - X_i}{h} \right) (\Sigma - I_d) X_i \end{aligned}$$

und durch

$$\begin{aligned} d^2 K_h(x, X_i, \vartheta_0) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0) & \quad (\text{C.8}) \\ &= d_{\mu\mu}^2 K_h(x, X_i, \vartheta_0) \circ (\mu, \mu) + 2 d_{\Sigma\mu}^2 K_h(x, X_i, \vartheta_0) \circ (\Sigma - I_d, \mu) \\ & \quad + d_{\Sigma\Sigma}^2 K_h(x, X_i, \vartheta_0) \circ (\Sigma - I_d, \Sigma - I_d) \\ &= \frac{1}{h^2} \mu^T \nabla^2 K \left( \frac{x - X_i}{h} \right) \mu \\ & \quad + \frac{1}{h^2} X_i^T (\Sigma - I_d) \nabla^2 K \left( \frac{x - X_i}{h} \right) \mu - \frac{1}{h} \nabla K \left( \frac{x - X_i}{h} \right) (\Sigma - I_d) \mu \\ & \quad + \frac{1}{4h^2} X_i^T (\Sigma - I_d) \nabla^2 K \left( \frac{x - X_i}{h} \right) (\Sigma - I_d) X_i - \frac{3}{4h} \nabla K \left( \frac{x - X_i}{h} \right) (\Sigma - I_d)^2 X_i. \end{aligned}$$

### Abschätzungen der Restglieder

Desweiteren werden Abschätzungen der Restglieder  $R_1$  bzw.  $R_2$  aus den Entwicklungen (C.3) bzw. (C.5) benötigt. Hierzu stellt man zunächst fest, daß die in (C.4) bzw. (C.6) auftretenden Ableitungen bei festem  $x, X_i \in \mathbb{R}^d$  und  $t \in (0, 1)$  unter Verwendung der Abkürzungen  $\tilde{\vartheta} = (\tilde{\mu}, \tilde{\Sigma})$  mit

$$\tilde{\vartheta} := \vartheta_0 + t(\vartheta - \vartheta_0), \quad \tilde{\mu} := t\mu, \quad \tilde{\Sigma} := I_d + t(\Sigma - I_d)$$

gegeben sind durch

$$d^2 K_h(x, X_i, \tilde{\vartheta}) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0)$$

$$\begin{aligned}
&= d_{\mu\mu}^2 K_h(x, X_i, \tilde{\vartheta}) \circ (\mu, \mu) + 2 d_{\Sigma\mu}^2 K_h(x, X_i, \tilde{\vartheta}) \circ (\Sigma - I_d, \mu) \\
&\quad + d_{\Sigma\Sigma}^2 K_h(x, X_i, \tilde{\vartheta}) \circ (\Sigma - I_d, \Sigma - I_d) \\
&= \frac{1}{h^2} \mu^T \tilde{\Sigma}^{-1/2} \cdot \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \cdot \tilde{\Sigma}^{-1/2} \mu \\
&\quad + \frac{2}{h^2} (\tilde{\mu} - X_i)^T \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \tilde{\Sigma}^{-1/2} \mu \\
&\quad + \frac{1}{h^2} (\tilde{\mu} - X_i)^T \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] (\tilde{\mu} - X_i) \\
&\quad + \frac{2}{h} \nabla K(\text{st}(x, X_i, \tilde{\vartheta})) \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \mu \\
&\quad + \frac{1}{h} \nabla K(\text{st}(x, X_i, \tilde{\vartheta})) \left[ d^2\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d, \Sigma - I_d) \right] (\tilde{\mu} - X_i)
\end{aligned}$$

und durch

$$\begin{aligned}
&d^3 K_h(x, X_i, \tilde{\vartheta}) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0, \vartheta - \vartheta_0) \\
&= d_{\mu\mu\mu}^3 K_h(x, X_i, \tilde{\vartheta}) \circ (\mu, \mu, \mu) + 3 d_{\Sigma\mu\mu}^3 K_h(x, X_i, \tilde{\vartheta}) \circ (\Sigma - I_d, \mu, \mu) \\
&\quad + 3 d_{\mu\Sigma\Sigma}^3 K_h(x, X_i, \tilde{\vartheta}) \circ (\mu, \Sigma - I_d, \Sigma - I_d) + d_{\Sigma\Sigma\Sigma}^3 K_h(x, X_i, \tilde{\vartheta}) \circ (\Sigma - I_d, \Sigma - I_d, \Sigma - I_d) \\
&= \frac{1}{h^3} \mu^T \tilde{\Sigma}^{-1/2} \left[ \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta})) \circ \tilde{\Sigma}^{-1/2} \mu \right] \tilde{\Sigma}^{-1/2} \mu \\
&\quad + \frac{3}{h^3} \mu^T \tilde{\Sigma}^{-1/2} \left\{ \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta})) \circ \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] (\tilde{\mu} - X_i) \right\} \tilde{\Sigma}^{-1/2} \mu \\
&\quad + \frac{3}{h^3} (\tilde{\mu} - X_i)^T \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \left\{ \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta})) \circ \tilde{\Sigma}^{-1/2} \mu \right\} \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] (\tilde{\mu} - X_i) \\
&\quad + \frac{1}{h^3} (\tilde{\mu} - X_i)^T \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \left\{ \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta})) \circ \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] (\tilde{\mu} - X_i) \right\} \\
&\quad \quad \quad \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] (\tilde{\mu} - X_i) \\
&\quad + \frac{6}{h^2} \mu^T \tilde{\Sigma}^{-1/2} \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \mu \\
&\quad + \frac{3}{h^2} \mu^T \tilde{\Sigma}^{-1/2} \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \left[ d^2\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d, \Sigma - I_d) \right] (\tilde{\mu} - X_i) \\
&\quad + \frac{6}{h^2} (\tilde{\mu} - X_i)^T \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \mu \\
&\quad + \frac{3}{h^2} (\tilde{\mu} - X_i)^T \left[ d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d) \right] \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \left[ d^2\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d, \Sigma - I_d) \right] (\tilde{\mu} - X_i) \\
&\quad + \frac{3}{h} \nabla K(\text{st}(x, X_i, \tilde{\vartheta})) \left[ d^2\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d, \Sigma - I_d) \right] \mu \\
&\quad + \frac{1}{h} \nabla K(\text{st}(x, X_i, \tilde{\vartheta})) \left\{ d^3\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d, \Sigma - I_d, \Sigma - I_d) \right\} (\tilde{\mu} - X_i) .
\end{aligned}$$

Wir versehen den  $\mathbb{R}^d$  mit der Euklidnorm  $\|\cdot\|$  und den Vektorraum  $\text{Symm}$  der symmetrischen  $d \times d$ -Matrizen mit der dazu passenden Spektralnorm  $\|\cdot\|$  (vgl. Anhang A). Nach Definition gilt zunächst

$$\|\tilde{\mu} - X_i\| \leq \|\tilde{\mu}\| + \|X_i\| \leq \|\mu\| + \|X_i\| .$$

Berücksichtigt man  $d\Upsilon(\tilde{\Sigma}) \in \mathcal{L}(\text{Symm}; \text{Symm})$  und schreibt abkürzend

$$\|d\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}} := \|d\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}(\text{Symm}; \text{Symm})}$$

für die zugehörige Operatornorm (analog  $\|d^2\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}}$  sowie  $\|d^3\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}}$ ), so gelten die Abschätzungen (vgl. (C.1) auf Seite 160)

$$\begin{aligned} \|d\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d)\| &\leq \|d\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}} \|\Sigma - I_d\|, \\ \|d^2\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d, \Sigma - I_d)\| &\leq \|d^2\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}} \|\Sigma - I_d\|^2, \\ \|d^3\Upsilon(\tilde{\Sigma}) \circ (\Sigma - I_d, \Sigma - I_d, \Sigma - I_d)\| &\leq \|d^3\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}} \|\Sigma - I_d\|^3. \end{aligned}$$

Nach Abschnitt C.2.3 ist die Funktion  $\Upsilon : \tilde{\mathfrak{J}} \rightarrow SP_d$  beliebig oft stetig differenzierbar in der offenen Umgebung  $\tilde{\mathfrak{J}}$  der Einheitsmatrix  $I_d$ . Damit existieren eine kompakte Umgebung  $\tilde{\mathfrak{J}} \subset \mathfrak{J}$  und eine Konstante  $C$  mit

$$\|\Upsilon(\Sigma)\| \leq C, \quad \|d\Upsilon(\Sigma)\|_{\mathcal{L}} \leq C, \quad \|d^2\Upsilon(\Sigma)\|_{\mathcal{L}} \leq C, \quad \|d^3\Upsilon(\Sigma)\|_{\mathcal{L}} \leq C$$

für jedes  $\Sigma \in \tilde{\mathfrak{J}}$ . Falls also unsere Matrix  $\Sigma$  in  $\tilde{\mathfrak{J}}$  liegt, folgen wegen  $t \in (0, 1)$  und

$$\|\tilde{\Sigma} - I_d\| = |t| \|\Sigma - I_d\| \leq \|\Sigma - I_d\|$$

die nicht mehr von  $t$  abhängenden Abschätzungen

$$\|\Upsilon(\tilde{\Sigma})\| \leq C, \quad \|d\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}} \leq C, \quad \|d^2\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}} \leq C, \quad \|d^3\Upsilon(\tilde{\Sigma})\|_{\mathcal{L}} \leq C. \quad (\text{C.9})$$

Damit erhalten wir für die zweite Ableitung von  $K_h$

$$\begin{aligned} &\left| d^2 K_h(x, X_i, \tilde{\vartheta}) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0) \right| \quad (\text{C.10}) \\ &\leq \frac{C^2}{h^2} \|\mu\|^2 \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\ &\quad + \frac{2C^2}{h^2} \|\mu\| \|\Sigma - I_d\| (\|\mu\| + \|X_i\|) \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\ &\quad + \frac{C^2}{h^2} \|\Sigma - I_d\|^2 (\|\mu\| + \|X_i\|)^2 \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\ &\quad + \frac{2C}{h} \|\mu\| \|\Sigma - I_d\| \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\ &\quad + \frac{C}{h} \|\Sigma - I_d\|^2 (\|\mu\| + \|X_i\|) \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta})) \right\|. \end{aligned}$$

Berücksichtigt man ferner, daß das Differential dritter Ordnung  $\nabla^3 K(x)$  in  $\mathcal{L}(\mathbb{R}^d; \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}))$  liegt, und schreibt man  $\|\nabla^3 K(x)\|$  für die zugehörige Operatornorm, so gilt für jeden Vektor  $y \in \mathbb{R}^d$  die Ungleichung  $\|\nabla^3 K(x) \circ y\| \leq \|\nabla^3 K(x)\| \|y\|$ , und es folgt eine entsprechende Abschätzung für die dritte Ableitung von  $K_h$ :

$$\left| d^3 K_h(x, X_i, \tilde{\vartheta}) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0, \vartheta - \vartheta_0) \right| \quad (\text{C.11})$$

$$\begin{aligned}
&\leq \frac{C^3}{h^3} \|\mu\|^3 \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| + \frac{3C^3}{h^3} \|\mu\|^2 \|\Sigma - I_d\| (\|\mu\| + \|X_i\|) \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\
&\quad + \frac{3C^3}{h^3} \|\mu\| \|\Sigma - I_d\|^2 (\|\mu\| + \|X_i\|)^2 \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\
&\quad + \frac{C^3}{h^3} \|\Sigma - I_d\|^3 (\|\mu\| + \|X_i\|)^3 \left\| \nabla^3 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\
&\quad + \frac{6C^2}{h^2} \|\mu\|^2 \|\Sigma - I_d\| \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\
&\quad + \frac{9C^2}{h^2} \|\mu\| \|\Sigma - I_d\|^2 (\|\mu\| + \|X_i\|) \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\
&\quad + \frac{3C^2}{h^2} \|\Sigma - I_d\|^3 (\|\mu\| + \|X_i\|)^2 \left\| \nabla^2 K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\
&\quad + \frac{3C}{h} \|\mu\| \|\Sigma - I_d\|^2 \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| \\
&\quad + \frac{C}{h} \|\Sigma - I_d\|^3 (\|\mu\| + \|X_i\|) \left\| \nabla K(\text{st}(x, X_i, \tilde{\vartheta})) \right\| .
\end{aligned}$$

Einsetzen dieser oberen Schranken in

$$\begin{aligned}
|R_1| &\leq \int_0^1 (1-t) \left| d^2 K_h(x, X_i, \vartheta_0 + t(\vartheta - \vartheta_0)) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0) \right| dt , \\
|R_2| &\leq \int_0^1 \frac{1}{2} (1-t)^2 \left| d^3 K_h(x, X_i, \vartheta_0 + t(\vartheta - \vartheta_0)) \circ (\vartheta - \vartheta_0, \vartheta - \vartheta_0, \vartheta - \vartheta_0) \right| dt
\end{aligned}$$

liefert die gesuchte Abschätzung für die Restglieder.

## D. $U$ -Statistiken

In der vorliegenden Arbeit werden einige Hilfsmittel aus der Theorie der  $U$ -Statistiken benötigt, die in diesem Anhang zusammengefaßt sind. Einzelheiten über  $U$ -Statistiken sind z.B. in Serfling (1980) zu finden. Dieses Buch beschränkt sich jedoch wie die meisten älteren Werke auf die Diskussion von  $U$ -Statistiken mit festem, d.h. nicht vom Stichprobenumfang  $n$  abhängendem Kern, während für uns vor allem  $U$ -Statistiken mit variablem, vom Stichprobenumfang  $n$  abhängendem Kern von Interesse sind. Eines der wenigen Bücher, die auf letztere eingehen, ist Borovskikh (1996).

### D.1. Fast sichere Konvergenz von $U$ - bzw. $V$ -Statistiken

Es seien  $(\mathfrak{X}, \mathfrak{F})$  ein Meßraum und  $X_1, \dots, X_n$  unabhängige und identisch verteilte Zufallselemente mit Werten in  $\mathfrak{X}$ . Ferner seien  $H : \mathfrak{X}^m \rightarrow \mathbb{R}^d$  ein meßbarer symmetrischer Kern vom Grad  $m$ ,

$$\theta := \mathbb{E}[H(X_1, \dots, X_m)]$$

der zu schätzende Parameter,

$$U_n := \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} H(X_{i_1}, \dots, X_{i_m})$$

die zum Kern  $H$  gehörende  $U$ -Statistik vom Grad  $m$  und

$$V_n := \frac{1}{n^m} \sum_{i_1=1}^n \dots \sum_{i_m=1}^n H(X_{i_1}, \dots, X_{i_m})$$

die zum Kern  $H$  gehörende  $V$ -Statistik vom Grad  $m$ .

Auf die Annahme, daß der Kern  $H$  in seinen  $m$  Argumenten symmetrisch ist, kann verzichtet werden: Besitzt  $H$  diese Eigenschaft nicht, so ersetze man ihn in der Definition von  $U_n$  bzw.  $V_n$  durch  $\frac{1}{m!} \sum H(X_{i_1}, \dots, X_{i_m})$  (Summation über alle  $m!$  Permutationen der Indizes  $(1, \dots, m)$ ), ohne daß der Wert der  $U$ - bzw.  $V$ -Statistik sich ändert.

#### Theorem D.1 (Borovskikh (1996), Theorem 3.1.1)

Unter der Voraussetzung  $\mathbb{E} \|H(X_1, \dots, X_m)\| < \infty$  gilt  $U_n \rightarrow \theta$  ( $n \rightarrow \infty$ ) fast sicher.

**Theorem D.2 (Borovskikh (1996), Theorem 3.3.2)**

Unter der Voraussetzung  $\max\{E \|H(X_{i_1}, \dots, X_{i_m})\| : 1 \leq i_1, \dots, i_m \leq m\} < \infty$  gilt  $V_n \rightarrow \theta$  ( $n \rightarrow \infty$ ) fast sicher.

**D.2.  $U$ -Statistiken mit variablem Kern**

Seien  $X_1, X_2, \dots, X_n$  unabhängige, identisch verteilte  $d$ -dimensionale Zufallsvektoren und sei  $H_n : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  ein Kern, der vom Stichprobenumfang  $n$  abhängt und der o.B.d.A. als symmetrisch angenommen werden kann, d.h.  $H_n(x, y) = H_n(y, x)$ . (Ist dies nicht der Fall, so ersetze man in der nachfolgenden Definition  $H_n(x, y)$  durch  $\frac{1}{2}(H_n(x, y) + H_n(y, x))$ .) Dann heißt

$$U_n := \frac{1}{\binom{n}{2}} \sum_{i < j} H_n(X_i, X_j) = \frac{1}{n(n-1)} \sum_{i \neq j} H_n(X_i, X_j)$$

$U$ -Statistik vom Grad 2 mit variablem Kern. Wie in der klassischen Theorie der  $U$ -Statistiken ist  $U_n$  ein erwartungstreuer Schätzer für

$$\theta_n := E[H_n(X_1, X_2)];$$

$H_n$  heißt *zentriert*, falls  $\theta_n = 0$  für alle  $n \in \mathbb{N}$  gilt. Mit der Notation

$$r_n(x) := E[H_n(X_1, X_2) | X_1 = x] = E[H_n(x, X_2)]$$

ist die *Hájek-Projektion*  $\hat{U}_n$  gegeben durch

$$\hat{U}_n := \theta_n + \frac{2}{n} \sum_{i=1}^n [r_n(X_i) - \theta_n].$$

Der Kern  $H_n$  heißt *degeneriert*, falls für alle  $n \in \mathbb{N}$  die Beziehung  $r_n(X_1) = 0$  fast sicher gilt.

**Lemma D.3 (Powell et al. (1989), Lemma 3.1)** Unter der Voraussetzung  $E[H_n(X_1, X_2)^2] = o(n)$  folgt  $\sqrt{n}(U_n - \hat{U}_n) = o_P(1)$ , oder anders ausgedrückt:  $U_n = \hat{U}_n + o_P(n^{-1/2})$ .

BEWEIS: Durch direkte Rechnung zeigt man, daß  $n \cdot E\left[\left(U_n - \hat{U}_n\right)^2\right] = o(1)$  gilt. Dies ist gleichbedeutend mit  $\sqrt{n}(U_n - \hat{U}_n) \rightarrow 0$  in  $L^2(\Omega, \mathcal{A}, P)$  und impliziert  $\sqrt{n}(U_n - \hat{U}_n) \xrightarrow{P} 0$  (vgl. Powell et al. (1989)).  $\square$

Ein wichtiges Hilfsmittel beim Nachweis der Konvergenz der Bowman-Foster-Statistik ist das folgende Theorem zur Konvergenz von  $U$ -Statistiken mit variablem Kern, das in der grundlegenden Arbeit Hall (1984) bewiesen wurde. Hierzu sei

$$G_n(x, y) := E[H_n(X_1, x)H_n(X_1, y)] \quad , \quad x, y \in \mathbb{R}^d.$$

**Theorem D.4 (Hall (1984), Theorem 1)** Für symmetrische, zentrierte, degenerierte und quadratisch integrierbare Kerne  $H_n$  ( $n \in \mathbb{N}$ ) folgt unter der Voraussetzung

$$\frac{\mathbb{E}[G_n^2(X_1, X_2)] + \frac{1}{n}\mathbb{E}[H_n^4(X_1, X_2)]}{(\mathbb{E}[H_n^2(X_1, X_2)])^2} \rightarrow 0 \quad (n \rightarrow \infty),$$

die Verteilungskonvergenz

$$\frac{\sum_{1 \leq i < j \leq n} H_n(X_i, X_j)}{\sqrt{\frac{1}{2}n^2 \mathbb{E}[H_n^2(X_1, X_2)]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Zur Behandlung der Bowman–Foster–Statistik unter einer Folge von benachbarten Alternativen benötigen wir die Verallgemeinerung von Theorem D.4 auf Dreiecksschemata von Zufallsvektoren, d.h. auf die folgende Situation:

Für  $n \in \mathbb{N}$  seien  $X_{n1}, X_{n2}, \dots, X_{nn}$  unabhängige, identisch verteilte  $d$ -dimensionale Zufallsvektoren, die zusätzlich alle unabhängig seien. Definieren wir analog zum Fall von i.i.d. Zufallsvektoren

$$G_n(x, y) := \mathbb{E}[H_n(X_{n1}, x)H_n(X_{n1}, y)] \quad , \quad x, y \in \mathbb{R}^d,$$

so gilt

**Theorem D.5** Für symmetrische, zentrierte, degenerierte und quadratisch integrierbare Kerne  $H_n$  ( $n \in \mathbb{N}$ ) folgt unter der Voraussetzung

$$\frac{\mathbb{E}[G_n^2(X_{n1}, X_{n2})] + \frac{1}{n}\mathbb{E}[H_n^4(X_{n1}, X_{n2})]}{(\mathbb{E}[H_n^2(X_{n1}, X_{n2})])^2} \rightarrow 0 \quad (n \rightarrow \infty),$$

die Verteilungskonvergenz

$$\frac{\sum_{1 \leq i < j \leq n} H_n(X_{ni}, X_{nj})}{\sqrt{\frac{1}{2}n^2 \mathbb{E}[H_n^2(X_{n1}, X_{n2})]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

**BEWEIS:** Wir überzeugen uns kurz davon, daß der Beweis aus Hall (1984) seine Gültigkeit auch unter Dreiecksschemata von Zufallsvektoren behält. Eine diesbezügliche Bemerkung findet sich bereits in Liero et al. (1998), allerdings ohne nähere Ausführungen.

Setzt man

$$Y_{nj} := \sum_{k=1}^{j-1} H_n(X_{nj}, X_{nk}) \quad , \quad 2 \leq j \leq n,$$

und

$$S_{ni} := \sum_{j=2}^i Y_{nj} = \sum_{j=2}^i \sum_{k=1}^{j-1} H_n(X_{nj}, X_{nk}),$$

so ist der zu untersuchende Term  $\sum_{1 \leq i < j \leq n} H_n(X_{ni}, X_{nj})$  gerade gleich  $S_{nn}$ . Setzt man ferner

$$\mathfrak{F}_{ni} := \sigma(X_{n1}, \dots, X_{ni}),$$

so ist  $\{S_{ni}, \mathfrak{F}_{ni}\}_{\substack{1 \leq i \leq n \\ n \geq 1}}$  ein zentriertes, quadratisch integrierbares Martingale Array, und wir können analog zum Originalbeweis von P.Hall Corollar 3.1 aus Hall und Heyde (1980) anwenden.

**Theorem D.6 (Hall und Heyde (1980), Corollary 3.1)**  $\{\tilde{S}_{ni}, \tilde{\mathfrak{F}}_{ni}\}_{\substack{1 \leq i \leq k_n \\ n \geq 1}}$  sei ein zentriertes, quadratisch integrierbares Martingale Array mit  $k_n \rightarrow \infty$  ( $n \rightarrow \infty$ ). Ferner sei  $\tilde{X}_{ni} := \tilde{S}_{ni} - \tilde{S}_{n,i-1}$  ( $1 \leq i \leq k_n$ ) mit  $\tilde{S}_{n0} := 0$ , und  $\eta$  bezeichne eine fast sicher endliche Zufallsvariable. Unter den Bedingungen

$$(i) \quad \forall \varepsilon > 0 : \sum_{i=1}^{k_n} \mathbb{E} \left[ \tilde{X}_{ni}^2 \mathbf{1}\{|\tilde{X}_{ni}| > \varepsilon\} \right] \rightarrow 0$$

$$(ii) \quad \tilde{V}_{nk_n}^2 := \sum_{i=1}^{k_n} \mathbb{E} \left[ \tilde{X}_{ni}^2 \middle| \tilde{\mathfrak{F}}_{n,i-1} \right] \xrightarrow{P} \eta^2.$$

$$(iii) \quad \forall 1 \leq i \leq k_n, n \geq 1 : \tilde{\mathfrak{F}}_{ni} \subseteq \tilde{\mathfrak{F}}_{n+1,i}$$

gilt  $\tilde{S}_{nk_n} = \sum_{i=1}^{k_n} \tilde{X}_{ni} \xrightarrow{\mathcal{D}} Z$ , wobei die Zufallsvariable  $Z$  durch die charakteristische Funktion  $\mathbb{E} [e^{itZ}] = \mathbb{E} \left[ e^{-\frac{1}{2}\eta^2 t^2} \right]$  gegeben ist.

Falls  $\eta^2$  konstant ist, kann in Theorem D.6 auf Bedingung (iii) verzichtet werden (vgl. Hall und Heyde (1980), S. 59).

Setzen wir

$$s_n^2 := \mathbb{E} [S_{nn}^2] , \quad V_n^2 := \sum_{i=2}^n \mathbb{E} [Y_{ni}^2 \middle| \mathfrak{F}_{n,i-1}]$$

und wenden Theorem D.6 mit  $k_n := n$ ,  $\eta^2 := 1$  auf  $\frac{1}{s_n} S_{nn}$  an, so folgt bei Nachprüfen der Bedingungen

$$(i) \quad \forall \varepsilon > 0 : \frac{1}{s_n^2} \sum_{i=2}^n \mathbb{E} [Y_{ni}^2 \mathbf{1}\{|Y_{ni}| > \varepsilon s_n\}] \rightarrow 0 \quad (n \rightarrow \infty)$$

$$(ii) \quad \frac{1}{s_n^2} V_n^2 \xrightarrow{P} 1$$

die Konvergenzaussage  $\frac{1}{s_n} S_{nn} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ .

Wie in Hall (1984) berechnet man

$$s_n^2 = \frac{n(n-1)}{2} \mathbb{E} [H_n^2(X_{n1}, X_{n2})]$$

und

$$\sum_{i=2}^n \mathbb{E} [Y_{ni}^4] \leq \text{const} \cdot n^3 \cdot \mathbb{E} [H_n^4(X_{n1}, X_{n2})] ,$$

woraus (i) folgt:

$$\begin{aligned} \frac{1}{s_n^2} \sum_{i=2}^n \mathbb{E} [Y_{ni}^2 \mathbf{1}\{|Y_{ni}| > \varepsilon s_n\}] &\leq \frac{1}{\varepsilon^2 s_n^4} \sum_{i=2}^n \mathbb{E} [Y_{ni}^4] \\ &\leq \frac{1}{\varepsilon^2} \cdot \text{const} \cdot \frac{\frac{1}{n} \mathbb{E} [H_n^4(X_{n1}, X_{n2})]}{(\mathbb{E} [H_n^2(X_{n1}, X_{n2})])^2} \rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

nach Voraussetzung. Ebenfalls wie in Hall (1984) berechnet man

$$\frac{\mathbb{E} [(V_n^2 - s_n^2)^2]}{s_n^4} \leq \text{const} \cdot \frac{\mathbb{E} [G_n^2(X_{n1}, X_{n2})] + \frac{1}{n} \mathbb{E} [H_n^4(X_{n1}, X_{n2})]}{(\mathbb{E} [H_n^2(X_{n1}, X_{n2})])^2} \rightarrow 0 \quad (n \rightarrow \infty)$$

nach Voraussetzung, was  $\mathbb{E} \left[ \left( \frac{V_n^2}{s_n^2} - 1 \right)^2 \right] \rightarrow 0$  impliziert und damit  $\frac{V_n^2}{s_n^2} \xrightarrow{P} 1$ . □

# E. Einige Aussagen der asymptotischen Statistik

## E.1. Benachbarte Alternativen

Im folgenden werden einige Aussagen zur Benachbarkeit von Verteilungsfolgen zitiert. Sie sind alle in Kapitel 6.3 in Witting und Müller-Funk (1995) zu finden.

Sei  $(\Omega, \mathcal{A})$  ein Meßraum. Der *Totalvariationsabstand* zweier endlicher Maße  $\mu$  und  $\nu$  auf  $\Omega$  ist definiert durch

$$\|\mu - \nu\| := \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)| .$$

Der *Hellinger-Abstand* zweier Wahrscheinlichkeitsmaße  $P, Q$  auf  $\Omega$ , die von einem Maß  $\nu$  auf  $\Omega$  dominiert werden ( $P \ll \nu, Q \ll \nu$ ), ist gegeben durch

$$H(P, Q) := \frac{1}{\sqrt{2}} \left( \int_{\Omega} \left( \sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right)^2 d\nu \right)^{1/2} ,$$

und es gilt:

$$H^2(P, Q) = 1 - \int_{\Omega} \sqrt{\frac{dP}{d\nu}} \sqrt{\frac{dQ}{d\nu}} d\nu. \quad (\text{E.1})$$

Der Hellinger-Abstand hängt nicht von dem dominierenden Maß  $\nu$  ab.

Im folgenden seien  $(\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)}) := (\times_{j=1}^n \mathfrak{X}_j, \otimes_{j=1}^n \mathfrak{F}_j)$  das  $n$ -fache Produkt von Meßräumen  $(\mathfrak{X}_j, \mathfrak{F}_j)$  und  $\mathfrak{M}^1(\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)})$  der Raum aller Wahrscheinlichkeitsmaße auf  $(\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)})$ .  $(\mathbb{R}^l, \mathfrak{B}^l)$  sei der  $l$ -dimensionale reelle Raum, versehen mit der Borel- $\sigma$ -Algebra.

**Definition E.1 (Witting und Müller-Funk (1995), Definition 6.106)** Für  $n \in \mathbb{N}$  seien  $u_n, v_n \in \mathfrak{M}^1(\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)})$  beliebige Wahrscheinlichkeitsmaße. Dann definiert man:

a)  $(v_n)$  heißt zu  $(u_n)$  *benachbart*, kurz:  $(v_n) \triangleleft (u_n)$ , genau dann, wenn gilt:

$$\forall (B_n) \in (\mathfrak{F}_{(n)}) : u_n(B_n) \rightarrow 0 \quad \Rightarrow \quad v_n(B_n) \rightarrow 0;$$

$(v_n)$  und  $(u_n)$  heißen *wechselseitig benachbart*, kurz:  $(v_n) \triangleleft\triangleright (u_n)$ , genau dann, wenn gilt:

$$(v_n) \triangleleft (u_n) \quad \text{und} \quad (u_n) \triangleleft (v_n);$$

b)  $(v_n)$  und  $(u_n)$  heißen *nicht trennend*, wenn ein  $0 < c < 1$  existiert mit

$$\forall (B_n) \in (\mathfrak{F}_{(n)}) : u_n(B_n) \rightarrow 0 \quad \Rightarrow \quad \limsup_{n \rightarrow \infty} v_n(B_n) \leq c;$$

c)  $(v_n)$  und  $(u_n)$  heißen *vollständig trennend*, kurz:  $(v_n) \triangle (u_n)$ , genau dann, wenn gilt:

$$\exists (B_n) \in (\mathfrak{F}_{(n)}) : u_n(B_n) \rightarrow 0 \quad \text{und} \quad v_n(B_n) \rightarrow 1.$$

**Satz E.2 (Witting und Müller-Funk (1995), Hilfssatz 6.110)** Für  $n \in \mathbb{N}$  seien  $u_n, v_n \in \mathfrak{M}^1(\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)})$  beliebige Wahrscheinlichkeitsmaße, und  $(\mathfrak{T}, \mathfrak{D})$  sei ein polnischer Raum.

a) Sind  $T_n : (\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)}) \rightarrow (\mathbb{R}^l, \mathfrak{B}^l)$ ,  $n \in \mathbb{N}$ , beliebige Statistiken, so gilt

$$(v_n) \triangleleft (u_n) \iff \left[ T_n \xrightarrow{u_n} 0 \quad \Rightarrow \quad T_n \xrightarrow{v_n} 0 \right].$$

b) Sind  $T_n : (\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)}) \rightarrow (\mathfrak{T}, \mathfrak{D})$ ,  $n \in \mathbb{N}$ , beliebige Statistiken, so gilt

$$(v_n) \triangleleft (u_n) \quad \Rightarrow \quad (v_n^{T_n}) \triangleleft (u_n^{T_n}).$$

c) Für jede Teilfolge  $(n') \subset \mathbb{N}$  gilt

$$(v_n) \triangleleft (u_n) \quad \Rightarrow \quad (v_{n'}) \triangleleft (u_{n'}).$$

**Satz E.3 (Witting und Müller-Funk (1995), Hilfssatz 6.119)** Für  $n \in \mathbb{N}$  und jedes  $j = 1, \dots, n$  seien  $u_{nj}, v_{nj} \in \mathfrak{M}^1(\mathfrak{X}_{nj}, \mathfrak{F}_{nj})$  sowie  $u_n := \otimes_{j=1}^n u_{nj}$  und  $v_n := \otimes_{j=1}^n v_{nj}$ . Dann gilt:

$$a) \sum_{j=1}^n H^2(u_{nj}, v_{nj}) = o(1) \quad \Rightarrow \quad (v_n) \triangleleft (u_n);$$

$$b) (v_n), (u_n) \text{ nicht trennend} \quad \Rightarrow \quad \begin{cases} \sum_{j=1}^n H^2(u_{nj}, v_{nj}) = O(1), \\ \sum_{j=1}^n \|u_{nj} - v_{nj}\|^2 = O(1); \end{cases}$$

$$c) (v_n) \triangleleft (u_n) \quad \Rightarrow \quad \left( \sum_{j=1}^n u_{nj}(B_{nj}) \rightarrow 0 \quad \Rightarrow \quad \sum_{j=1}^n v_{nj}(B_{nj}) \rightarrow 0 \right).$$

**Satz E.4 (Witting und Müller-Funk (1995), Satz 6.120 (Oosterhoff-van Zwet))** Für  $n \in \mathbb{N}$  und jedes  $j = 1, \dots, n$  seien  $u_{nj}, v_{nj} \in \mathfrak{M}^1(\mathfrak{X}_{nj}, \mathfrak{F}_{nj})$  sowie  $u_n := \otimes_{j=1}^n u_{nj}$  und  $v_n := \otimes_{j=1}^n v_{nj}$ .

Dann ist  $(v_n) \triangleleft (u_n)$  genau dann erfüllt, wenn mit dem Dichtequotienten  $L_{nj} := dv_{nj}/du_{nj}$ ,  $j = 1, \dots, n$ , gilt:

$$\begin{aligned} \sum_{j=1}^n H^2(u_{nj}, v_{nj}) &= O(1), \\ \sum_{j=1}^n v_{nj}(L_{nj} > c_n) &= o(1) \quad \forall (c_n) \subset \mathbb{R} : c_n \rightarrow \infty. \end{aligned}$$

**Satz E.5 (Witting und Müller-Funk (1995), Korollar 6.124 (1. Le Cam Lemma))** Für  $n \in \mathbb{N}$  seien  $u_n, v_n \in \mathfrak{M}^1(\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)})$  und  $L_n := dv_n/du_n$ . Dann ist

$$\mathfrak{L}_{u_n}(\log L_n) \xrightarrow{\mathfrak{L}} \mathfrak{N}(-\kappa^2/2, \kappa^2)$$

hinreichend für die wechselseitige Benachbarkeit  $(v_n) \triangleleft (u_n)$  und die Gültigkeit von

$$\mathfrak{L}_{v_n}(\log L_n) \xrightarrow{\mathfrak{L}} \mathfrak{N}(\kappa^2/2, \kappa^2).$$

## E.2. Asymptotische relative Effizienz

Im folgenden bezeichnen  $\Phi$  die Verteilungsfunktion der Standard-Normalverteilung und  $u_\alpha$  deren  $\alpha$ -Quantil.

**Definition E.6 (Witting und Müller-Funk (1995), Definition 6.92 und 6.93)** Gegeben seien für  $n \in \mathbb{N}$  Zufallsgrößen  $X_{(n)}$  mit Werten in Räumen  $(\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)})$ , deren Verteilungen eindeutig durch einen Parameter  $\vartheta \in \Theta$  festgelegt sind, sowie ein Testproblem mit Hypothesen  $\mathbf{H}$  bzw.  $\mathbf{K}$ . Dann versteht man unter *asymptotischen Hypothesen* oder *Hypothesen des asymptotischen Testproblems* Mengen  $\widehat{\mathbf{H}}$  und  $\widehat{\mathbf{K}}$  von Verteilungsfolgen aus  $\mathbf{H}$  bzw.  $\mathbf{K}$ , also Mengen

$$\begin{aligned} \widehat{\mathbf{H}} &\subset \{(u_n) : u_n = \mathfrak{L}_{\vartheta_n}(X_{(n)}), \vartheta_n \in \mathbf{H} \quad \forall n \in \mathbb{N}\}, \\ \widehat{\mathbf{K}} &\subset \{(v_n) : v_n = \mathfrak{L}_{\vartheta_n}(X_{(n)}), \vartheta_n \in \mathbf{K} \quad \forall n \in \mathbb{N}\}. \end{aligned}$$

$\widehat{\mathbf{H}}$  bzw.  $\widehat{\mathbf{K}}$  heißen *einfach*, wenn sie einelementig sind, andernfalls *zusammengesetzt*. Unter einem *asymptotischen Test zum Niveau  $\alpha$  für  $\widehat{\mathbf{H}}$*  versteht man einen asymptotischen Test  $(\varphi_n)$  mit

$$\limsup_{n \rightarrow \infty} E_{u_n} \varphi_n \leq \alpha \quad \forall (u_n) \in \widehat{\mathbf{H}}.$$

**Definition E.7 (Witting und Müller-Funk (1995), Definition 6.102)**  $(u_n)$  und  $(v_n)$  seien zwei einfache asymptotische Hypothesen sowie  $(\varphi_{1n})$  und  $(\varphi_{2n})$  zwei asymptotische Tests zum Niveau  $\alpha$  für  $(u_n)$  gegen  $(v_n)$  derart, daß gilt

$$E_{v_n} \varphi_{in} \rightarrow 1 - \Phi(u_\alpha - \delta_i) \quad \exists \delta_i \in (0, \infty), \quad i = 1, 2. \quad (\text{E.2})$$

Dann ist die *asymptotische relative Effizienz* oder *Pitman-Effizienz* von  $(\varphi_{2n})$  bezüglich  $(\varphi_{1n})$  unter der Folge  $(v_n)$  definiert durch

$$\text{ARE}(\varphi_{2n} : \varphi_{1n} | v_n) := \delta_2^2 / \delta_1^2. \quad (\text{E.3})$$

Gilt (E.2) für alle  $(v_n) \in \widehat{\mathbf{K}}$ , so heißt (E.3) *asymptotische relative Effizienz* oder *Pitman-Effizienz* von  $(\varphi_{2n})$  bzgl.  $(\varphi_{1n})$  unter  $\widehat{\mathbf{K}}$ .

**Bemerkung E.8 (Witting und Müller-Funk (1995), S. 285)** Gilt in (E.2)  $\delta_i = 0$  für (höchstens) ein  $i = 1, 2$ , so ist die Definition der Pitman-Effizienz auch noch sinnvoll, wenn man in üblicher Weise  $\text{ARE}(\varphi_{2n} : \varphi_{1n} | v_n) = \infty$  bzw.  $\text{ARE}(\varphi_{2n} : \varphi_{1n} | v_n) = 0$  setzt.

**Bemerkung E.9 (Witting und Müller-Funk (1995), S. 329)** Sind  $T_n$ ,  $n \in \mathbb{N}$  reellwertige Statistiken, für die mit  $\nu, c \in \mathbb{R}$  und  $\sigma^2 > 0$  gilt

$$\mathfrak{L}_{u_n}(T_n) \xrightarrow{\mathfrak{L}} \mathfrak{N}(\nu, \sigma^2), \quad \mathfrak{L}_{v_n}(T_n) \xrightarrow{\mathfrak{L}} \mathfrak{N}(\nu + c, \sigma^2),$$

so ergeben sich für die Tests  $\varphi_n := \mathbf{1}\{T_n > \nu + \sigma u_\alpha\}$  das Limesniveau

$$E_{u_n} \varphi_n \rightarrow 1 - \Phi(u_\alpha) = \alpha$$

und die Limeschärfe

$$E_{v_n} \varphi_n \rightarrow 1 - \Phi\left(u_\alpha + \frac{c}{\sigma}\right).$$

**Bemerkung E.10** Sind  $T_n$  und  $\tilde{T}_n$  ( $n \in \mathbb{N}$ ) reellwertige Statistiken mit  $\nu, \tilde{\nu}, \tilde{c} \in \mathbb{R}$  und  $\sigma^2, \tilde{\sigma}^2 > 0$  sowie

$$\begin{aligned} \mathfrak{L}_{u_n}(T_n) &\xrightarrow{\mathfrak{L}} \mathfrak{N}(\nu, \sigma^2), & \mathfrak{L}_{v_n}(T_n) &\xrightarrow{\mathfrak{L}} \mathfrak{N}(\nu, \sigma^2), \\ \mathfrak{L}_{u_n}(\tilde{T}_n) &\xrightarrow{\mathfrak{L}} \mathfrak{N}(\tilde{\nu}, \tilde{\sigma}^2), & \mathfrak{L}_{v_n}(\tilde{T}_n) &\xrightarrow{\mathfrak{L}} \mathfrak{N}(\tilde{\nu} + \tilde{c}, \tilde{\sigma}^2), \end{aligned}$$

so hat der Test  $\varphi_n := \mathbf{1}\{T_n > \nu + \sigma u_\alpha\}$  gegenüber dem Test  $\tilde{\varphi}_n := \mathbf{1}\{\tilde{T}_n > \tilde{\nu} + \tilde{\sigma} u_\alpha\}$  nach Definition E.7 und Bemerkung E.9 die Pitman-Effizienz

$$\text{ARE}(\varphi_n : \tilde{\varphi}_n | v_n) = 0.$$

# Symbolverzeichnis

## Allgemeines

i.i.d., o.B.d.A.	S. 4	$\Phi', \Phi'_1, \Phi'_2, \Phi'', \Phi''_1, \Phi''_2$	S. 10 ff.
$\mathbf{1}$ , const	S. 4	$\tilde{\Phi}, \tilde{\Phi}', \tilde{\Phi}''$	S. 10
$(\Omega, \mathcal{A}, P), (\mathbb{R}^d, \mathfrak{B}^d, \lambda^d)$	S. 4	$\Phi_n, \Phi_{1,n}, \Phi_{2,n}$	S. 9
$P^X$	S. 4	$\Phi'_n, \Phi'_{1,n}, \Phi'_{2,n}, \Phi''_n, \Phi''_{1,n}, \Phi''_{2,n}$	S. 21
$\mathcal{N}_d$	S. 5	$\Phi_n^\diamond, \Phi_{1,n}^\diamond, \Phi_{2,n}^\diamond, \Phi'_{1,n}, \Phi'_{2,n}, \Phi''_{1,n}, \Phi''_{2,n}$	S. 22
$\Phi^{-1}, \Phi$	S. 28,30	$\Phi_{2,n}^{\diamond\diamond}$	
$\mathcal{U}$ , Laplace, Exp	S. 42	$\hat{m}_3, \hat{m}_4$	S. 41
Po	S. 39	$M'_{3,1}, M'_{3,2}, M''_{4,1}, M''_{4,2}$	S. 11
$\delta_{\{\cdot\}}$	S. 72,115	$M'_{3,1,n}, M'_{3,2,n}, M''_{4,1,n}, M''_{4,2,n}$	S. 21
$P_\vartheta, H_0, H_1$	S. 1	$M_{3,1,n}^\diamond, M_{3,2,n}^\diamond, M_{4,1,n}^{\diamond\diamond}, M_{4,2,n}^{\diamond\diamond}$	S. 22
$\xrightarrow{\mathcal{D}}, \xrightarrow{P}, \xrightarrow{\text{f.s.}}$	S. 4	$K_{\beta 1}^1, K_{\beta 2}^1, K_{\beta 1}^2, K_{\beta 2}^2, K_{\beta 1}^3, K_{\beta 2}^3$	S. 18 ff.
$O_P(1), o_P(1), O_{\text{f.s.}}(1), o_{\text{f.s.}}(1)$	S. 4	$K_{\beta 1}^{4'}, K_{\beta 2}^{4'}, K_{\beta 1}^{5'}, K_{\beta 2}^{5'}, K_{\beta 1}^{6'}, K_{\beta 2}^{6'}$	
$O(1), o(1)$		$K_{\beta 1}^{7''}, K_{\beta 2}^{7''}, K_{\beta 1}^{8''}, K_{\beta 2}^{8''}, K_{\beta 1}^9, K_{\beta 2}^9$	
$\mathbb{R}^d, \ \cdot\ $	S. 4	$K_{\beta 1n}^1, K_{\beta 2n}^1, K_{\beta 1n}^2, K_{\beta 2n}^2, K_{\beta 1n}^3$	S. 24
$\mathbb{R}^{d \times d}, \ \cdot\ , \ \cdot\ _S, I_d$	S. 4	$K_{\beta 2n}^3, K_{\beta 1n}^{4'}, K_{\beta 2n}^{4'}, K_{\beta 1n}^{5'}, K_{\beta 2n}^{5'}$	
$(\cdot)^T$	S. 4	$K_{\beta 1n}^{6'}, K_{\beta 2n}^{6'}, K_{\beta 1n}^{7''}, K_{\beta 2n}^{7''}, K_{\beta 1n}^{8''}$	
		$K_{\beta 2n}^{8''}, K_{\beta 1n}^9, K_{\beta 2n}^9$	
		$K_{\beta 1n}^{1\diamond}, K_{\beta 2n}^{1\diamond}, K_{\beta 1n}^{2\diamond}, K_{\beta 2n}^{2\diamond}, K_{\beta 1n}^{3\diamond}$	S. 64–71
		$K_{\beta 2n}^{3\diamond}, K_{\beta 1n}^{4\diamond}, K_{\beta 2n}^{4\diamond}, K_{\beta 1n}^{5\diamond}, K_{\beta 2n}^{5\diamond}$	
		$K_{\beta 1n}^{6\diamond}, K_{\beta 2n}^{6\diamond}, K_{\beta 1n}^{7\diamond}, K_{\beta 2n}^{7\diamond}, K_{\beta 1n}^{8\diamond}$	
		$K_{\beta 2n}^{8\diamond}, K_{\beta 1n}^{9\diamond}, K_{\beta 2n}^{9\diamond}$	
		$I_1, I_2, I_3, I'_4, I'_5$	S. 26,57
		$I_{\beta,n,1-\alpha}$	S. 28
		$\tilde{I}_{a,n,1-\alpha}$	S. 38
		$q_{\beta,1-\alpha}, q_{\beta,n,1-\alpha}, q_{\beta,J,1-\alpha}, q_{\beta,P,1-\alpha}$	S. 29
		$\tilde{q}_{a,n,1-\alpha}$	S. 38
		$\beta_{1,d}, b_{1,d}, \tilde{\beta}_{1,d}, \tilde{b}_{1,d}$	S. 35
		$\Delta, H_\Delta, K_\Delta, \mathbf{d}_\beta(\cdot, \cdot)$	S. 30–31
		$\Delta_j$	S. 12
		$\varepsilon_{n,\cdot}, \eta_{n,\cdot}$	S. 14
		$A_n, \tilde{A}, B_n, \tilde{B}, C_n, D_n, \tilde{D}_n$	S. 15–18
		$R_{1,n}, R_{2,n}, R'_{1,n}, R'_{2,n}, R''_{1,n}$	S. 23
		$R''_{2,n}, R'_{3,1,n}, R'_{3,2,n}, R''_{4,1,n}, R''_{4,2,n}$	
		$a, \tilde{a}, a_d, b, \tilde{b}, b_d, t, \tilde{t}, t_d$	S. 57
		$\varphi_{\beta,d}, \varphi_{\beta,d-1}, \varphi_{\beta,1}$	S. 57
<b>BHEP-Test</b>			
$F$	S. 29 ff.		
$f$	S. 33		
$X_j, Y_j, \bar{X}_n, S_n$	S. 5–6		
$\beta, \varphi_\beta, T_{n,\beta}, D_{n,\beta}$	S. 6		
$L_\beta, S_{n,\beta}$	S. 8		
$S_\beta, \tau_\beta^2$	S. 18		
$\tau_{\beta n}^2$	S. 25		
$Z_n, Z$	S. 9 ff.		
$\tilde{Z}_n, Z_n^*, Z_j^{**}$	S. 12 ff.		
$k$	S. 10		
$\sigma^2, \tau^2$	S. 33		
$a, \tilde{T}_{n,a}, \tilde{S}_{n,a}, \tilde{S}_a, \tilde{L}_a, \sigma^2, \tilde{Z}_n, \tilde{Z}$	S. 37–38		
$\tilde{k}, \tilde{\tau}_a^2, \tilde{\tau}_{an}^2$			
$\vartheta, \hat{\vartheta}_n, \tilde{T}_n, \tilde{T}_{n,a}, \tilde{L}_a$	S. 39		
$\Psi, \Psi_n$	S. 5		
$\Psi, \Psi', \Psi_n$	S. 36–38		
$\Phi, \Phi_1, \Phi_2$	S. 8 ff.		

**Bowman–Foster–Statistik**

$h$	S. 73
$\kappa_d$	S. 77
$K$	S. 73,75
$k$	S. 75,85
$\mathcal{F}, f$	S. 77,82
$f_0, f_1$	S. 101
$f_n, \mathbf{w}, N_n$	S. 109
$\hat{f}_{n,\text{st}}, \hat{f}_n$	S. 73,83
$\mu(f), \sigma^2(f)$	S. 79,83
$e_h(x, f)$	S. 83
$\Theta, \vartheta, \hat{\vartheta}_n$	S. 77–78
$\hat{\vartheta}_n, \vartheta_0, \tilde{\vartheta}_n, \tilde{\mu}_n, \tilde{\Sigma}_n$	S. 87–89
$T_{n,h,\text{st}}$	S. 73
$\tilde{T}_{n,h,\text{st}}$	S. 83
$T_{n,h,f}, T_{n,h}$	S. 77 ff.
$\tilde{T}_{n,h}$	S. 83
$\tilde{T}_{n,1}, \tilde{T}_{n,2}$	S. 91
$\tilde{T}_{n,3}, \tilde{T}_{n,4}, \tilde{T}_{n,5}$	S. 94
$\tilde{T}_{n,6}, \tilde{T}_{n,7}, \tilde{T}_{n,8}$	S. 97
$\tilde{T}_{n,9}, \tilde{T}_{n,10}$	S. 117
$\tilde{R}_{n,h}, \tilde{R}_{n,1}, \tilde{R}_{n,2}$	S. 111
$\tilde{S}_{n,h,\text{st}}$	S. 101
$\tilde{S}_{n,1}, \tilde{S}_{n,2}$	S. 103
$\tilde{S}_{n,3,i}$	S. 138
$\tilde{S}_{n,4}, \tilde{S}_{n,5}, \tilde{S}_{n,6}$	S. 143
$q_{h,1-\alpha}, q_{h,n,1-\alpha}$	S. 100
$\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_{3,r,s}$	S. 102
$\gamma_1, \gamma_2$	S. 102
$\tilde{\sigma}^2(f_1, f_0)$	S. 102
$\sigma_n^{*2}$	S. 104
$\tilde{\sigma}_n^2(f_1, f_0)$	S. 107
$\sigma_R^2$	S. 113
$\sigma_{R,n}^2$	S. 114
$\tilde{S}_{n,h}, \tilde{\sigma}_*^2(f_1, f_0), \gamma_1^*, \gamma_2^*$	S. 108
$Y_{ni}$	S. 104
$\tilde{Y}_{ni}$	S. 114
$L_n, U_n, H_n, G_n$	S. 92 ff.
$L_n, U_n, H_n, G_n$	S. 111 ff.
$U_n, H_n, \hat{U}_n, \theta_n, r_n$	S. 126 ff.
$\tilde{U}_n, \tilde{H}_n, \tilde{\hat{U}}_n, \tilde{\theta}_n, \tilde{r}_n, \tilde{r}_n^*$	S. 131 ff.
$\gamma_d$	S. 115

$\tilde{M}_1, \tilde{M}_2$	S. 118
$\tilde{\kappa}^2$	S. 123
$\eta, \zeta$	S. 125;131

**Anhang Hilberträume**

$H, H', \langle \cdot, \cdot \rangle, \ \cdot\ $	S. 152
$S^*$	S. 152
$\mathcal{S}$	S. 153
$S_X$	S. 154
$\text{tr}(\cdot)$	S. 153
$\chi$	S. 155
$L^2(\mathbb{R}^d, \mathfrak{B}^d, \mu), \langle \cdot, \cdot \rangle_{L^2}$	S. 157
$m, B$	S. 158

**Anhang Differentialkalkül**

$\mathcal{L}(E; F), \mathcal{L}(E_1, \dots, E_m; F)$	S. 160
$df, C^n$	S. 161
Isom	S. 164
$\varphi$	S. 164
$M(d \times d, \mathbb{R}), \text{Symm}, SP_d$	S. 165
$\mathfrak{J}, \mathfrak{J}$	S. 167
$id, \text{sq}, \text{sqr}$	S. 166,167
$\Upsilon, \hbar$	S. 168,169
$\text{st}, K_h$	S. 169,170
$\nabla, \nabla^2, \nabla^3$	S. 170 ff.
$R_1, R_2$	S. 176 ff.
$\tilde{\vartheta}, \tilde{\mu}, \tilde{\Sigma}$	S. 176

**Anhang U–Statistiken**

$H, \theta, U_n, V_n$	S. 180
$H_n, U_n, \theta_n, r_n, \hat{U}_n, G_n$	S. 181
$S, \mathfrak{F}$	S. 182,183

**Anhang Asymptotische Statistik**

$\ \cdot\ , H(\cdot, \cdot)$	S. 185
$\triangleleft, \triangleleft, \triangle$	S. 185 ff.
$(\mathfrak{X}_j, \mathfrak{F}_j), (\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)})$	S. 185
$\mathfrak{M}^1(\mathfrak{X}_{(n)}, \mathfrak{F}_{(n)})$	S. 185
$L_{nj}, L_n$	S. 186,187
$H, \hat{H}, K, \hat{K}$	S. 187
ARE	S. 187

# Literaturverzeichnis

- Achieser, N. und Glasmann, I. (1975). *Theorie der linearen Operatoren im Hilbert-Raum* (6th ed.). Berlin: Akademie-Verlag.
- Araujo, A. und Giné, E. (1980). *The central limit theorem for real and Banach valued random variables*. New York: Wiley.
- Avez, A. (1983). *Calcul différentiel*. Paris: Masson.
- Baringhaus, L., Gürtler, N. und Henze, N. (2000). Weighted integral test statistics and components of smooth tests of fit. *Austral. New Zeal. J. Statist.*, 42(2), 179–192.
- Baringhaus, L. und Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, 35, 339–348.
- Beirlant, J. und Mason, D. (1995). On the asymptotic normality of  $L_P$ -norms of empirical functionals. *Mathematical Methods of Statistics*, 4(1), 1–19.
- Bertrand-Retali, M. (1974). Convergence uniforme stochastique d'un estimateur d'une densité de probabilité dans  $R^s$ . *C. R. Acad. Sci. Paris, Ser. A*, 278, 1449–1452.
- Bertrand-Retali, M. (1978). Convergence uniforme d'un estimateur de la densité par la méthode du noyau. *Rev. Roumaine Math. Pures Appl.*, 23, 361–385.
- Bickel, P. und Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Stat.*, 1(6), 1071–1095.
- Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.
- Borovskikh, Y. (1996). *U-Statistics in Banach Spaces*. Utrecht: VSP.
- Bowman, A. (1992). Density based tests for goodness-of-fit. *J. Statist. Comput. Simul.*, 40, 1–13.
- Bowman, A. und Foster, P. (1993). Adaptive smoothing and density-based tests of multivariate normality. *J. Amer. Statist. Assoc.*, 88(422), 529–537.
- Brieskorn, E. (1985). *Lineare Algebra und Analytische Geometrie II*. Braunschweig, Wiesbaden: Vieweg.
- Brown, L., Hwang, J. und Munk, A. (1997). An unbiased test for the bioequivalence problem. *Ann. Stat.*, 25(6), 2345–2367.

- Cartan, H. (1977). *Cours de calcul différentiel*. Paris: Hermann.
- Chen, X. und White, H. (1992). *Central limit and functional central limit theorems for Hilbert space-valued dependent processes* (Department of Economics Discussion Paper 92-35). San Diego: University of California.
- Cramér, H. und Leadbetter, M. (1967). *Stationary and related stochastic processes*. New York: Wiley.
- Csörgő, M., Gombay, E. und Horváth, L. (1991). Central limit theorems for  $L_p$  distances of kernel estimators of densities under random censorship. *Ann. Stat.*, 19(4), 1813–1831.
- Csörgő, M. und Horváth, L. (1987). Asymptotics for  $L_p$ -norms of kernel estimators of densities. *Computational Statistics & Data Analysis*, 6(3), 241–250.
- Csörgő, M. und Horváth, L. (1988). Central limit theorems for  $L_p$ -norms of density estimators. *Probability Theory and Related Fields*, 80(2), 269–291.
- Csörgő, S. (1989). Consistency of some tests for multivariate normality. *Metrika*, 36, 107–116.
- Czado, C. und Munk, A. (1998). Assessing the similarity of distributions – finite sample performance of the empirical Mallows distance. *J. Statist. Comput. Simul.*, 60, 319–346.
- Dette, H. und Munk, A. (1998). Validation of linear regression models. *Ann. Stat.*, 26(2), 778–800.
- Devroye, L. (1987). *A course in density estimation*. Boston/Basel/Stuttgart: Birkhäuser.
- Devroye, L. und Györfi, L. (1985). *Nonparametric density estimation: The  $L_1$ -view*. New York: Wiley.
- De Wet, T. und Randles, R. (1987). On the effect of substituting parameter estimators in limiting  $\chi^2$ ,  $U$  and  $V$  statistics. *Ann. Stat.*, 15, 398–412.
- Eaton, M. und Perlman, M. (1973). The non-singularity of generalized sample covariance matrices. *Ann. Stat.*, 1(4), 710–717.
- Epps, T. und Pulley, L. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, 70(3), 723–726.
- Es, B. van. (1997). A note on the integrated squared error of a kernel density estimator in non-smooth cases. *Statist. & Prob. Lett.*, 35, 241–250.
- Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory*, 10(2), 316–356.
- Fan, Y. (1995). Bootstrapping a consistent nonparametric goodness-of-fit test. *Econometric Reviews*, 14(3), 367–382.

- Fan, Y. und Li, Q. (1999). Central limit theorem for degenerate  $U$ -statistics of absolutely regular processes with applications to model specification testing. *Nonparametric Statistics*, 10, 245–271.
- Fan, Y. und Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *Nonparametric Statistics*, 11, 337–360.
- Feller, W. (1966). *An introduction to probability theory and its applications* (Vol. II). New York: Wiley.
- Fraenkel, L. (1978). Formulae for high derivatives of composite functions. *Math. Proc. Camb. Phil. Soc.*, 83, 159–165.
- Ghosh, B. und Huang, W.-M. (1991). The power and optimal kernel of the Bickel–Rosenblatt test for goodness of fit. *Ann. Stat.*, 19(2), 999–1009.
- Gihman, I. und Skorohod, A. (1974). *The theory of stochastic processes I*. Springer.
- Gradshteyn, I. und Ryzhik, I. (1980). *Table of integrals, series, and products*. New York: Academic Press.
- Gürtler, N. und Henze, N. (1999). *Recent and classical goodness-of-fit tests for the Poisson distribution* (Preprint 99/03). Germany: University of Karlsruhe. (Submitted to: J. Statist. Plann. Inf.)
- Hall, P. (1982). Limit theorems for stochastic measures of the accuracy of density estimators. *Stochastic Processes and their Applications*, 13, 11–25.
- Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multiv. Anal.*, 14, 1–16.
- Hall, P. und Heyde, C. (1980). *Martingale limit theory and its applications*. New York: Academic Press.
- Henze, N. (1990). An approximation to the limit distribution of the Epps-Pulley test statistics for normality. *Metrika*, 37, 7–18.
- Henze, N. (1993). A new flexible class of omnibus tests for exponentiality. *Commun. Statist. – Th. Meth.*, 22, 115–133.
- Henze, N. (1997). Extreme smoothing and testing for multivariate normality. *Statist. & Prob. Lett.*, 35, 203–213.
- Henze, N. und Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. *J. Multiv. Anal.*, 62(1), 1–23.
- Henze, N. und Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Commun. Statist. – Th. Meth.*, 19(10), 3595–3617.
- Heuser, H. (1975). *Funktionalanalysis* (1st ed.). Stuttgart: Teubner.

- Horváth, L. (1990). Asymptotics for  $L_P$ -norms of Fourier series density estimators. *Constructive Approximation*, 6(4), 375–397.
- Horváth, L. (1991). On  $L_P$ -norms of multivariate density estimators. *Ann. Stat.*, 19(4), 1933–1949.
- Ibragimov, I. und Rozanov, Y. (1978). *Gaussian random processes*. New York: Springer.
- Johnson, N. und Kotz, S. (1970). *Continuous univariate distributions* (Vol. 1, 1nd ed.). Boston: Houghton Mifflin Comp.
- Kim, C., Hong, C., Jeong, M. und Yang, M. (1997). Goodness-of-fit test for density estimation. *Commun. Statist. – Th. Meth.*, 26(11), 2725–2741.
- Komlós, J., Major, P. und Tusnády, G. (1975). An approximation of partial sums of independent rv's, and the sample df I. *Z. W.theorie Verw. Geb.*, 32, 111–131.
- Ledoux, M. und Talagrand, M. (1991). *Probability in Banach spaces. Isoperimetry and processes*. Berlin: Springer.
- Liero, H., Läuter, H. und Konakov, V. (1998). Nonparametric versus parametric goodness of fit. *Statistics*, 31, 115–149.
- Mardia, K. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530.
- Móri, T., Rohatgi, V. und Székely, G. (1993). On multivariate skewness and kurtosis. *Theoret. Probab. Appl.*, 38, 547–551.
- Mourier, E. (1954). *Éléments aléatoires dans un espace de Banach*. Unpublished doctoral dissertation, Paris.
- Munk, A. und Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *J. Roy. Statist. Soc. B*, 60(1), 223–241.
- Neumann, M. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Stat.*, 26(5), 2014–2048.
- Neumann, M. und Paparoditis, E. (1998). *A nonparametric test for the stationary density* (Tech. Rep. 58/98). Berlin: SFB 373, Humboldt Univ.
- Parthasarathy, K. (1967). *Probability measures on metric spaces*. New York-London: Academic Press.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33, 1065–1076.
- Powell, J., Stock, J. und Stoker, T. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57(6), 1403–1430.

- Prakasa Rao, B. (1983). *Nonparametric functional estimation*. Orlando, Florida: Academic Press.
- Prokhorov, Y. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.*, 1(2), 157–214.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, 27, 832–837.
- Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Stat.*, 3(1), 1–14.
- Rosenblatt, M. (1976). On the maximal deviation of  $k$ -dimensional density estimates. *Ann. Prob.*, 4(6), 1009–1015.
- Rudin, W. (1987). *Real and complex analysis* (3rd ed.). New York: Mc Graw Hill.
- Rueda, R., Pérez-Abreu, V. und O'Reilly, F. (1991). Goodness of fit for the Poisson distribution based on the probability generating function. *Commun. Statist. – Th. Meth.*, 20(10), 3093–3110.
- Ruymgaart, F. (1998). A note on weak convergence of density estimators in Hilbert spaces. *Statistics*, 30, 331–343.
- Serfling, R. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.
- Silverman, B. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Stat.*, 6(1), 177–184. (Addendum: *Ann. Stat.* 8(5), 1175–1176, 1980)
- Silverman, B. (1980). Addendum to weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Stat.*, 8(5), 1175–1176.
- Silverman, B. (1986). *Density estimation for statistics and data analysis*. London, New York: Chapman and Hall.
- Stoer, J. (1999). *Numerische Mathematik I* (8th ed.). Berlin, Heidelberg: Springer.
- Takahata, H. und Yoshihara, K. (1987). Central limit theorems for integrated square error of nonparametric density estimators based on absolutely regular random sequences. *Yokohama Mathematical Journal*, 35, 95–111.
- Vakhania, N. (1981). *Probability distributions on linear spaces*. New York: North-Holland.
- Wentzell, A. (1981). *A course in the theory of stochastic processes*. Mc Graw Hill.
- Witting, H. und Müller-Funk, U. (1995). *Mathematische Statistik II*. Stuttgart: Teubner.



# Lebenslauf

## Persönliche Daten:

Name: Nora Gürtler  
Geboren am: 05.11.1970 in Mannheim  
Eltern: Manfred und Felizitas Gürtler

## Schulbildung:

1977 – 1981 Grundschole in Mannheim  
1981 – 1990 Geschwister-Scholl-Gymnasium in Mannheim  
08.05.1990 Abiturprüfung

## Studium:

Okt. 1990 – Sept. 1992 Grundstudium in Mathematik und Informatik an der Universität Karlsruhe  
Sept. 1992 – Aug. 1993 Magistère de Mathématiques Pures an der ENS Lyon, Licence de Mathématiques Pures an der Université Claude Bernard/Lyon I  
Sept. 1993 – Aug. 1994 Magistère de Mathématiques Pures at ENS Lyon, Maîtrise de Mathématiques Pures at Université Claude Bernard/Lyon I  
Sept. 1994 – Jan. 1997 Hauptstudium in Mathematik an der Universität Karlsruhe

## Tätigkeiten während des Studiums:

Okt. 1994 – Sept. 1996 Wissenschaftliche Hilfskraft am Mathematischen Institut II der Universität Karlsruhe  
Okt. 1994 – Apr. 1995 Wissenschaftliche Hilfskraft am Institut für Mathematische Stochastik der Universität Karlsruhe  
Okt. 1995 – Sept. 1996 Wissenschaftliche Hilfskraft am Mathematischen Institut I der Universität Karlsruhe

## Berufstätigkeit:

seit Feb. 1997 Wissenschaftliche Angestellte am Institut für Mathematische Stochastik der Universität Karlsruhe