

# Klassische und neue statistische Anpassungstests

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik der  
Universität Karlsruhe  
genehmigte

DISSERTATION

von

Dipl.-Math. Bernhard Klar  
aus Riedlingen

Tag der mündlichen Prüfung:

24. Juni 1998

Referent:

Prof. Dr. N. Henze

Korreferent:

apl. Prof. Dr. L. Baringhaus



# Inhaltsverzeichnis

<b>Einleitung</b>	<b>1</b>
<b>1 Glatte Anpassungstests</b>	<b>9</b>
1.1 Tests mit diagnostischen Eigenschaften . . . . .	9
1.2 Glatte Anpassungstests in Exponentialfamilien . . . . .	18
1.3 Die mehrdimensionale Normalverteilung . . . . .	21
1.3.1 Multivariate Schiefe . . . . .	26
1.3.2 Multivariate Wölbung . . . . .	35
1.3.3 Komponenten höherer Ordnung . . . . .	45
1.4 Verallgemeinerungen von Schiefe und Wölbung . . . . .	52
1.5 Die bivariate Poissonverteilung . . . . .	65
1.6 Diagnostische Tests und Resampling-Verfahren . . . . .	70
1.6.1 Das Bootstrapverfahren . . . . .	72
1.6.2 Ein modifiziertes Jackknife-Verfahren . . . . .	77
<b>2 EIVF-Tests</b>	<b>83</b>
2.1 EIVF-Tests für diskrete Verteilungen . . . . .	86
2.1.1 Definition und grundlegende Eigenschaften . . . . .	86
2.1.2 Die Güte des Tests unter benachbarten Alternativen . . . . .	99
2.1.3 Beispiele und Simulationsergebnisse . . . . .	104
2.2 EIVF-Tests für stetige Verteilungen . . . . .	113
2.2.1 Ein Test auf Exponentialverteilung . . . . .	114
2.2.2 Ein Test auf Normalverteilung . . . . .	124

<b>A Ergänzungen</b>	<b>135</b>
A.1 Orthogonale Polynome . . . . .	135
A.2 Elliptisch-symmetrische Verteilungen . . . . .	138
A.3 Schranken für Wahrscheinlichkeiten . . . . .	141
<b>Literaturverzeichnis</b>	<b>145</b>

# Einleitung

In der induktiven Statistik werden beobachtete Daten  $x_1, \dots, x_n$  als Realisierungen von Zufallsvariablen  $X_1, \dots, X_n$  aufgefaßt. Um Schlußfolgerungen aus den Beobachtungen zu ziehen, die über eine bloße Beschreibung der Daten hinausgehen, müssen Annahmen über den zugrundeliegenden Zufallsmechanismus gemacht werden. Es ist folglich ein stochastisches Modell zu erstellen, durch das die (gemeinsame) Verteilung von  $X_1, \dots, X_n$  näher beschrieben wird. Dabei ist es intuitiv naheliegend, daß im allgemeinen um so weitreichendere Schlüsse aus den Daten gezogen werden können, je genauer das Modell die Verteilung von  $X_1, \dots, X_n$  beschreibt.

Aus wahrscheinlichkeitstheoretischer Sicht wäre es deshalb wünschenswert, die zugrundeliegende Verteilung vollständig zu spezifizieren. Handelt es sich um univariate Beobachtungen, so wäre ein mögliches Modell:  $X_1, \dots, X_n$  sind unabhängig und identisch verteilt; die gemeinsame Verteilung ist eine Normalverteilung mit Erwartungswert  $\mu = 0$  und Varianz  $\sigma^2 = 1$ . Man schreibt dafür

$$(X_j)_{j=1, \dots, n} \stackrel{uiv}{\sim} \mathcal{N}(0, 1).$$

Unter diesen Annahmen können dann alle interessierenden Größen mit wahrscheinlichkeitstheoretischen Hilfsmitteln bestimmt werden; statistische Methoden sind überflüssig.

Die Kehrseite dieser Vorgehensweise ist offensichtlich: Je genauer man ein Modell festlegt, desto größer ist die Gefahr, daß die Daten nicht mit dem Modell vereinbar sind. Folgerungen aus einem Modell, das mit den Be-

obachtungen nicht verträglich ist, können zumindest für die Praxis wertlos sein.

Die Möglichkeit eines Modellfehlers läßt sich ausschließen, indem man nur völlig unzweifelhafte Annahmen in das Modell einfließen läßt, etwa die Annahme von  $\mathbb{N}_0$ -wertigen Zufallsvariablen, wenn die Beobachtungen aus Zähldaten bestehen. Mit einem solchen Modell (das diesen Namen kaum noch verdient) wird man jedoch nicht viel anfangen können.

Der Ausweg, den die Statistik aus diesem Dilemma findet, ist ein Kompromiß: man legt einen Modell*rahmen* fest, der irgendwo zwischen den beiden oben beschriebenen Extremen liegt, und überprüft mit statistischen Methoden, ob dieser Rahmen mit den Daten verträglich ist. Wenn dies bejaht werden kann, werden aus den Daten und dem Modell Schlußfolgerungen gezogen, die über die reine Datenbeschreibung hinausgehen. Der zweite Schritt, das Überprüfen der Verträglichkeit von Modell und Beobachtungen mit Hilfe statistischer Anpassungstests, ist der Gegenstand der vorliegenden Arbeit. Dabei nehmen wir an, daß immer die Modellannahme

$X_1, \dots, X_n$  sind stochastisch unabhängig und identisch verteilt

erfüllt ist. Diese Voraussetzung ist in der Praxis bei kontrollierten Experimenten oft gegeben; in anderen Fällen ist sie durch geeignete Auswahl der Daten zumindest approximativ erfüllbar.

In der Literatur wird unter dem Begriff *Anpassungstest* oft folgende spezielle Situation verstanden (so z.B. in Kendall and Stuart (1979), Kap. 30): den Daten liegt als Grundannahme ein parametrisches Modell zugrunde. Ist  $\vartheta$  ein endlichdimensionaler Vektor und  $\{P_\vartheta : \vartheta \in \Theta\}$  eine parametrische Verteilungsfamilie, so dient ein statistischer Anpassungstest dazu, die Hypothese

$$\mathcal{H}_0 : P \in \{P_\vartheta : \vartheta \in \Theta\}$$

zu überprüfen. Dabei bezeichnet  $P$  die den Daten zugrundeliegende unbekannte Wahrscheinlichkeitsverteilung. Das wichtigste Beispiel für eine parametrische Verteilungsfamilie ist die Normalverteilungsklasse

$$\{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

Im Unterschied zum obigen Fall einer Normalverteilung mit bekannten Parametern hat man es also nicht mit einer einfachen, sondern mit einer zusammengesetzten Hypothese zu tun.

Liegt die Situation einer zusammengesetzten Hypothese vor, so muß erst ein geeigneter Schätzer  $\hat{\vartheta}_n$  für  $\vartheta$  definiert werden; danach benötigt man ein Vergleichskriterium, das in geeigneter Weise überprüft, ob die vorliegenden Daten aus der Verteilung  $P_{\hat{\vartheta}_n}$  erzeugt sein können.

Sind die Daten univariat, so ist die Verteilung  $P_{\vartheta}$  durch die zugehörige Verteilungsfunktion  $F(\cdot; \vartheta)$  bestimmt. In diesem Fall kann z.B. mit Hilfe der Kolmogoroff-Smirnoff-Statistik

$$K_n = \sqrt{n} \sup_{-\infty < t < \infty} |F_n(t) - F(t; \hat{\vartheta}_n)|$$

ein geeigneter Test konstruiert werden. Hier ist  $F_n(\cdot)$  die durch

$$F_n(t) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}\{X_j \leq t\}, \quad t \in \mathbb{R},$$

definierte empirische Verteilungsfunktion.

Im Falle der Normalverteilung kann man als Schätzer für  $\mu$  und  $\sigma^2$  den Stichprobenmittelwert  $\bar{X}_n$  und die empirische Varianz  $\hat{\sigma}_n^2$  verwenden;  $K_n$  nimmt dann die Form

$$K_n = \sqrt{n} \sup_{-\infty < t < \infty} \left| F_n(t) - \Phi\left(\frac{t - \bar{X}_n}{\hat{\sigma}_n}\right) \right|$$

an, wobei  $\Phi(\cdot)$  die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Eine wünschenswerte Eigenschaft, die die Kolmogoroff-Teststatistik besitzt, ist die *globale Konsistenz*: Jede feste Alternative zur hypothetischen Verteilungsklasse wird mit gegen 1 gehender Wahrscheinlichkeit erkannt, wenn der Stichprobenumfang gegen Unendlich strebt. Solche Tests werden *omnibus*-Tests genannt. Weitergehende theoretische Aussagen über das Güteverhalten sind nur schwer zu machen. Da die Alternative nichtparametrisch ist, existiert kein Test, der die größte Güte gegenüber jeder Alternative besitzt. In eingeschränkteren Situationen wie dem Testen einer einfachen

Hypothese gegen Lagealternativen läßt sich sogar zeigen, daß zu einem bestimmten Test meist eine Verteilungsfamilie existiert, für die der Test gewisse lokale asymptotische Optimalitätseigenschaften besitzt (Nikitin (1995), Kap. 6). Um so wichtiger ist es, das Güteverhalten eines Tests gegenüber einer möglichst großen Zahl von Alternativen durch Simulationsstudien bei in der Praxis auftretenden Stichprobenumfängen zu untersuchen.

Als Minimalforderung an einen Anpassungstest bleibt die globale Konsistenz bestehen. In den letzten Jahren hat sich aber gezeigt, daß auch solche Tests ihre Güte im Raum aller Alternativen sehr ungleichmäßig verteilen (siehe etwa Janssen (1995)). Es existieren nur wenige Richtungen von Abweichungen von der hypothetischen Verteilung, gegen die ein solcher Test eine hohe Güte besitzt. Insofern verhalten sich auch omnibus-Tests zumindest bei kleinen und mittleren Stichprobenumfängen ähnlich wie Tests, die von vornherein nur zur Aufdeckung bestimmter Alternativen konstruiert wurden.

Aus diesem Grunde werden in der Praxis häufig Tests verwendet, die zwar nicht global konsistent sind, aber dafür andere positive Eigenschaften besitzen. Bei Anwendern sind insbesondere sogenannte *diagnostische Tests* beliebt, denen die Eigenschaft zugeschrieben wird, nicht nur eine Hypothese ablehnen zu können, sondern gleichzeitig die Art der Abweichung zu diagnostizieren. Ein Beispiel hierfür ist der Dispersionsindex-Test, der einer der am häufigsten verwendeten Tests auf Poissonverteilung ist. Bei der Poissonverteilung ist der Quotient aus Varianz und Erwartungswert immer gleich 1. Ist nun der als Dispersionsindex bezeichnete Quotient aus empirischer Varianz und arithmetischem Mittel nicht „hinreichend nahe“ bei 1, so kann die Hypothese der Poissonverteilung (auf einem bestimmten Signifikanzniveau) abgelehnt werden. Darüber hinaus wird diagnostiziert, daß das hypothetische Modell deshalb verworfen werden muß, weil Erwartungswert und Varianz der zugrundeliegenden wahren Verteilung nicht übereinstimmen.

Daß die konventionelle Anwendung des Dispersionsindex-Tests nicht zu dieser Diagnose berechtigt, wurde in Henze und Klar (1996) gezeigt. Der Grundfehler liegt darin, daß Aussagen über nichtparametrische Verteilungs-



größen mittels einer bestimmten parametrischen Verteilungsklasse beurteilt werden. In Henze (1997a) und Henze und Klar (1996) wurde auch gezeigt, wie man diesen und ähnliche Tests modifizieren muß, um Verfahren zu erhalten, die zumindest bei großen Stichprobenumfängen diagnostische Eigenschaften besitzen. Die richtige Vorgehensweise ist, eine nichtparametrische Hypothese zu formulieren, die die gewünschten Beziehungen beinhaltet. Im Falle des Dispersionsindex-Tests wäre die Hypothese die Gleichheit von Erwartungswert und Varianz. Wird diese nichtparametrische Hypothese auf einem gewissen Signifikanzniveau abgelehnt, so kann man folgern, daß den Daten eine Verteilung zugrundeliegt, deren Erwartungswert von der Varianz verschieden ist.

Gemäß den angesprochenen beiden unterschiedlichen Problemstellungen besitzt die vorliegende Arbeit zwei Schwerpunkte. Im ersten Kapitel werden Tests untersucht, die als Anpassungstests im engeren Sinne verwendet werden können, die sich aber auch in natürlicher Weise zu Tests mit diagnostischen Eigenschaften erweitern lassen. Diese sogenannten glatten Anpassungstests (engl. „smooth tests of fit“) wurden schon 1937 von Jerzy Neyman eingeführt und erfreuen sich insbesondere in letzter Zeit wachsender Beliebtheit.

Das zweite Kapitel behandelt dagegen neue, auf der integrierten Verteilungsfunktion basierende Tests, die in den oben skizzierten engeren Rahmen passen. Neben der Annahme von unabhängig und identisch verteilten Daten setzen sie univariate Beobachtungen voraus; diese können vom diskreten oder stetigen Typ sein.

Im ersten Abschnitt von Kapitel 1 werden glatte Anpassungstests zum Testen der Hypothese  $P \in \{P_\vartheta : \vartheta \in \Theta\}$  eingeführt, die auf Teststatistiken der Form

$$\hat{\Psi}_{n,k}^2 = \sum_{j=s+1}^{s+k} \hat{U}_{n,j}^2 \quad \text{mit} \quad \hat{U}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_j(X_i; \hat{\vartheta}_n) \quad (0.1)$$

beruhen. Hierbei ist  $\{h_0(\cdot; \vartheta) \equiv 1, h_1(\cdot; \vartheta), h_2(\cdot; \vartheta), \dots\}$  ein bezüglich  $P_\vartheta$

orthonormales Polynomsystem, es gelten also die Relationen

$$\int h_j(\cdot; \vartheta) h_l(\cdot; \vartheta) dP_\vartheta = \delta_{j,l} \quad (0 \leq j, l \leq s+k). \quad (0.2)$$

Dabei sind die Polynome  $h_j(\cdot; \vartheta)$  vom Grade  $j$  ( $j \geq 0$ ), weshalb die Terme  $\hat{U}_{n,j}$  als Komponenten  $j$ -ten Grades bezeichnet werden. Setzt man in (0.2)  $l = 0$ , so ergibt sich für  $j \geq 1$

$$\int h_j(\cdot; \vartheta) dP_\vartheta = E_\vartheta[h_j(X_1; \vartheta)] = 0,$$

wodurch eine unter  $P_\vartheta$  gültige Relation zwischen den ersten  $j$  Momenten von  $X_1$  beschrieben wird.

Im Falle der Poissonverteilung mit Parameter  $\vartheta$  bilden die Poisson-Charlier-Polynome das Orthonormalsystem; es gilt insbesondere

$$h_1(x, \vartheta) = \frac{x - \vartheta}{\sqrt{\vartheta}}, \quad h_2(x, \vartheta) = \frac{(x - \vartheta)^2 - x}{\sqrt{2} \vartheta}.$$

Die Gleichung  $E_\vartheta[h_j(X_1; \vartheta)] = 0$  sagt für  $j = 1$  aus, daß der Erwartungswert der Poissonverteilung gerade gleich  $\vartheta$  ist, während sie für  $j = 2$  die Gleichheit von Erwartungswert und Varianz beschreibt. Da die erste Komponente  $\hat{U}_{n,1}$  wegen

$$\sum_{i=1}^n h_1(X_i, \hat{\vartheta}_n) = \frac{1}{\sqrt{\hat{\vartheta}_n}} \left( \sum_{i=1}^n X_i - n\bar{X}_n \right) = 0$$

verschwindet, setzt man in (0.1)  $s = 1$ . Weiter gilt hier

$$\hat{\Psi}_{n,2}^2 = \frac{1}{2n} (D_n - n)^2,$$

wobei  $D_n$  den oben beschriebenen Dispersionsindex bezeichnet. Der auf der ersten nichtverschwindenden Komponente  $\hat{U}_{n,2}$  beruhende glatte Anpassungstest auf Poissonverteilung und der Dispersionsindex-Test sind folglich äquivalent.

Nach der Darstellung der allgemeinen Verteilungstheorie der glatten Anpassungstests wird im ersten Abschnitt weiter gezeigt, wie diese modifiziert werden müssen, um Tests mit diagnostischen Eigenschaften zu erhalten.

Verteilungsklassen, in denen diese Modifikation besonders einfach zu bewerkstelligen ist, sind Thema des zweiten Abschnitts; dabei handelt es sich um spezielle *Exponentialfamilien*.

Der dritten Abschnitt behandelt glatte Anpassungstests für die multivariate Normalverteilung. Für diese mit weitem Abstand wichtigste Verteilung im Mehrdimensionalen wurden in der Literatur eine Vielzahl von Verteilungsgrößen und Teststatistiken wie etwa multivariate Schiefe und Wölbung eingeführt, die im Zusammenhang mit Komponenten der glatten Anpassungstests stehen und die mit Hilfe der Resultate des ersten Abschnitts einheitlich behandelt werden können. Verteilungsaussagen für die multivariate Normalverteilung können in vielen Fällen auf die nichtparametrische Klasse der elliptisch-symmetrischen Verteilungen verallgemeinert werden.

In einem Unterabschnitt werden einfache Ausdrücke für höhere Komponenten der multivariaten Normalverteilung hergeleitet, die die Anwendung der Tests in der Praxis überhaupt erst ermöglichen. In der Literatur sind solche Ausdrücke nur für auf den beiden ersten nichtverschwindenden Komponenten beruhende Tests zu finden.

Da die allgemeinen Ergebnisse im ersten Abschnitt teilweise ohne die Voraussetzung auskommen, daß die Polynome  $h_j$  ein Orthogonalsystem bilden, können auch Testgrößen behandelt werden, die eine Darstellung wie in (0.1) besitzen, wobei die  $h_j$  nicht orthogonal sind. Dieser Umstand wird in Abschnitt 1.4 benutzt, um die asymptotische Verteilung von Verallgemeinerungen der multivariaten Schiefe und Wölbung für die Klasse der elliptisch-symmetrischen Verteilungen zu bestimmen.

Diagnostische Tests für die bivariate Poissonverteilung sind Gegenstand des fünften Abschnitts, wobei auch hier die Schwierigkeit besteht, daß kein orthogonales Polynomsystem mehr definiert werden kann.

Werden glatte Anpassungstests als diagnostische Tests verwendet, so ist es außer im Falle der im zweiten Abschnitt behandelten univariaten Exponentialfamilien aufwendig, die asymptotische Verteilung der Teststatistiken konkret zu berechnen. Kapitel 1 schließt daher mit einem Abschnitt über die Bestimmung kritischer Werte für die betrachteten Tests mit Hilfe

von Resampling-Verfahren; dabei werden aus der ursprünglichen Stichprobe weitere Stichproben erzeugt, mit deren Hilfe die interessierenden Größen approximativ bestimmt werden können.

Das zweite Kapitel behandelt Anpassungstests, welche auf der integrierten Verteilungsfunktion basieren. Solche Verfahren wurden unseres Wissens nach bisher nicht betrachtet. Für eine auf  $\mathbb{R}_+$  konzentrierte Zufallsvariable  $X$  mit Verteilungsfunktion  $F_X(\cdot)$  kann die *integrierte Verteilungsfunktion*  $\Psi_X$  durch

$$\Psi_X(t) = \int_t^\infty (1 - F_X(x)) dx \quad (t \geq 0)$$

definiert werden. Es werden neue Teststatistiken vorgeschlagen, die auf der Differenz zwischen der integrierten Verteilungsfunktion und ihrem empirischen Gegenstück aufbauen.

Im ersten Abschnitt wird dabei der Fall diskreter Zufallsvariablen untersucht. Neben der asymptotischen Verteilung der Testgröße unter der Hypothese wird auch das Verhalten unter lokalen Alternativen besprochen. Es folgt ein längerer Unterabschnitt mit Simulationsergebnissen und einem Anwendungsbeispiel.

Im zweiten Abschnitt werden auf der integrierten Verteilungsfunktion basierende Anpassungstests für die beiden wichtigsten stetigen Verteilungen, die Exponential- und die Normalverteilung, vorgestellt. Auch für diese Tests wird das Güteverhalten durch Simulationsstudien illustriert.

Ich möchte mich an dieser Stelle bei Herrn Prof. Dr. N. Henze für seine Unterstützung während der Anfertigung dieser Arbeit sehr herzlich bedanken. Er hat ihren Werdegang mit zahlreichen wertvollen Hinweisen und anregenden Diskussionen begleitet.

Ebenso gilt mein Dank Herrn Prof. Dr. L. Baringhaus für die freundliche Übernahme des Korreferats.

# Kapitel 1

## Glatte Anpassungstests

### 1.1 Glatte Anpassungstests mit diagnostischen Eigenschaften

Im folgenden bezeichne  $\{P_\vartheta : \vartheta \in \Theta\}$  eine parametrische Klasse von Wahrscheinlichkeitsmaßen über den Borelmengen von  $\mathbb{R}$ , wobei  $\Theta$  eine offene Teilmenge des  $\mathbb{R}^s$  sei. Der Träger von  $P_\vartheta$  sei nicht von  $\vartheta$  abhängig; weiter seien die Verteilungen  $P_\vartheta$  unterscheidbar, d.h. die Abbildung  $\vartheta \rightarrow P_\vartheta$  sei injektiv.

Ist  $F_\vartheta : \mathbb{R} \rightarrow [0, 1]$  die zu  $P_\vartheta$  gehörende Verteilungsfunktion, so gelte

$$\int_{-\infty}^{\infty} |x|^n dF_\vartheta(x) < \infty \quad (n = 1, 2, \dots),$$

d.h. die Momente von  $F_\vartheta$  mögen existieren. Mit  $\mu_n$  wird das  $n$ -te Moment bezeichnet, es ist also

$$\mu_n = \int_{-\infty}^{\infty} x^n dF_\vartheta(x) \quad (n = 1, 2, \dots).$$

Besitzt  $F_\vartheta$  außerdem unendlich viele Wachstumspunkte (siehe Definition (A.2) in Anhang A.1), so existiert nach Satz A.1.1 genau ein bezüglich  $P_\vartheta$  orthonormales Polynomsystem  $\{h_0(\cdot; \vartheta) \equiv 1, h_1(\cdot; \vartheta), h_2(\cdot; \vartheta), \dots\}$ .

$X, X_1, \dots, X_n$  seien unabhängige und identisch verteilte Zufallsvariablen, deren (unbekannte) Verteilung mit  $P$  bezeichnet werden soll. Um die zusammengesetzte Hypothese

$$H_0 : P \in \{P_\vartheta : \vartheta \in \Theta\}$$

zu testen, können die von Neyman (1937) eingeführten „Glatte Anpassungstests“ (engl. „smooth tests of fit“) verwendet werden. Die Teststatistik des zu  $\{P_\vartheta\}$  gehörenden „smooth test“ hat die Form

$$\hat{\Psi}_{n,k}^2 = \sum_{j=k_0+1}^{k_0+k} \hat{U}_{n,j}^2, \quad (1.1)$$

wobei

$$\hat{U}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_j(X_i; \hat{\vartheta}_n)$$

ist; der untere Summationsindex  $k_0$  hängt vom verwendeten Schätzverfahren ab, wobei immer  $0 \leq k_0 \leq s$  gilt.  $\hat{\vartheta}_n$  ist ein Schätzer für  $\vartheta$ , der unter  $H_0$  die folgende Regularitätsvoraussetzung erfüllen soll:

**(R1)** Es existiert eine meßbare Funktion  $l : \mathbb{R} \times \Theta \rightarrow \mathbb{R}^s$ , so daß

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n l(X_i; \vartheta) + o_{P_\vartheta}(1),$$

wobei  $E_\vartheta[l(X; \vartheta)] = 0$  und  $E_\vartheta[l^2(X; \vartheta)] < \infty$ .

Der folgende Satz gibt die gemeinsame asymptotische Verteilung des Zufallsvektors  $\hat{U}_n = (\hat{U}_{n,k_0+1}, \dots, \hat{U}_{n,k_0+k})'$  unter  $H_0$  an. Dabei bedeutet  $h(x; \vartheta) = (h_{k_0+1}(x; \vartheta), \dots, h_{k_0+k}(x; \vartheta))'$ . Weiter bezeichne  $\nabla_\vartheta h(x; \vartheta)$  die  $(k \times s)$ -Matrix mit Elementen  $\partial h_i(x; \vartheta) / \partial \vartheta_j$ , wobei mit  $\vartheta_1, \dots, \vartheta_s$  die Komponenten von  $\vartheta$  bezeichnet werden.

### 1.1.1 Satz

- a) Die in  $h_j(x; \vartheta) = \sum_{i=0}^j a_{ji}(\vartheta) x^i$ ,  $k_0 + 1 \leq j \leq k_0 + k$ , auftretenden Koeffizienten  $a_{ji}(\vartheta)$  seien stetig differenzierbar in  $\vartheta$ .  $\hat{\vartheta}_n$  erfülle (R1).

Dann besitzt  $\hat{U}_n$  unter  $P_\vartheta$  asymptotisch eine  $k$ -dimensionale Normalverteilung  $\mathcal{N}(0, \Sigma)$  mit Kovarianzmatrix

$$\Sigma = E_\vartheta [v(X; \vartheta) v(X; \vartheta)'];$$

dabei ist  $v(x; \vartheta) = h(x; \vartheta) + E_\vartheta [\nabla_\vartheta h(X; \vartheta)] l(x; \vartheta)$ .

- b) Zusätzlich zu den Voraussetzungen aus Teil a) besitze  $P_\vartheta$  eine Radon-Nikodym-Dichte  $f(\cdot; \vartheta)$  bezüglich eines  $\sigma$ -endlichen Maßes  $\mu$ . Für festes  $x \in \mathbb{R}$  sei  $f(x; \vartheta)$  stetig differenzierbar bezüglich  $\vartheta$ . Weiter sei für  $k_0 + 1 \leq j \leq k_0 + k$  das Integral über  $h_j(\cdot; \vartheta) f(\cdot; \vartheta)$  bezüglich  $\mu$  unter dem Integral differenzierbar, d.h. Integration und Differentiation seien vertauschbar. Ist  $C_\vartheta$  die  $(k \times s)$ -Matrix mit Elementen

$$c_{ij} = E_\vartheta \left[ h_i(X; \vartheta) \frac{\partial \log f(X; \vartheta)}{\partial \vartheta_j} \right],$$

so gilt

$$E_\vartheta [\nabla_\vartheta h(X; \vartheta)] = -C_\vartheta.$$

- c) Ist  $\hat{\vartheta}_n$  der Maximum-Likelihood-Schätzer für  $\vartheta$ , so gilt unter den Voraussetzungen von Teil b)

$$\Sigma = I_k - C_\vartheta [I(\vartheta)]^{-1} C_\vartheta'$$

mit der  $k$ -dimensionalen Einheitsmatrix  $I_k$  und der Fisher-Informationsmatrix  $I(\vartheta)$ , deren Elemente durch

$$I_{jk}(\vartheta) = E_\vartheta \left[ \left( \frac{\partial}{\partial \vartheta_j} \log f(X; \vartheta) \right) \times \left( \frac{\partial}{\partial \vartheta_k} \log f(X; \vartheta) \right) \right]$$

gegeben sind.

BEWEIS: a) Im folgenden sei  $U_{n,j}(\gamma) = (1/\sqrt{n}) \sum_{i=1}^n h_j(X_i; \gamma)$ ; insbesondere gilt also  $U_{n,j}(\hat{\vartheta}_n) = \hat{U}_{n,j}$ . Da nach Voraussetzung  $h_j(x; \vartheta)$  stetig differenzierbar ist, folgt mit dem Mittelwertsatz

$$U_{n,j}(\hat{\vartheta}_n) = U_{n,j}(\vartheta) + (\hat{\vartheta}_n - \vartheta)' [\nabla_\gamma U_{n,j}(\gamma)|_{\gamma=\vartheta^*}],$$

wobei  $\vartheta^*$  zwischen  $\vartheta$  und  $\hat{\vartheta}_n$  liegt. Mit Slutsky's Lemma und unter Beachtung von (R1) folgt

$$\begin{aligned} U_{n,j}(\hat{\vartheta}_n) &= U_{n,j}(\vartheta) + \sqrt{n}(\hat{\vartheta}_n - \vartheta)' E[\nabla_\gamma h_j(X; \gamma)|_{\gamma=\vartheta}] + o_{P_\vartheta}(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [h_j(X_i; \vartheta) + l(X_i; \vartheta)' E[\nabla_\gamma h_j(X; \gamma)|_{\gamma=\vartheta}]] + o_{P_\vartheta}(1). \end{aligned}$$

Beachtet man noch, daß  $h_j(X; \vartheta)$  und  $l(X; \vartheta)$  unter  $P_\vartheta$  zentrierte Größen sind, so erhält man die Aussage a) aus dem zentralen Grenzwertsatz.

b) Unter den angegebenen Voraussetzungen gilt die Gleichungskette

$$\begin{aligned} 0 &= \frac{\partial}{\partial \vartheta_j} [E_\vartheta(h_i(X; \vartheta))] = \frac{\partial}{\partial \vartheta_j} \int h_i(x; \vartheta) f(x, \vartheta) \mu(dx) \\ &= \int \frac{\partial(h_i(x; \vartheta)f(x, \vartheta))}{\partial \vartheta_j} \mu(dx) \\ &= \int \frac{\partial h_i(x; \vartheta)}{\partial \vartheta_j} f(x, \vartheta) \mu(dx) + \int h_i(x; \vartheta) \frac{\partial \log f(x; \vartheta)}{\partial \vartheta_j} f(x, \vartheta) \mu(dx) \\ &= E_\vartheta \left[ \frac{\partial h_i(X; \vartheta)}{\partial \vartheta_j} \right] + E_\vartheta \left[ h_i(X; \vartheta) \frac{\partial \log f(X; \vartheta)}{\partial \vartheta_j} \right], \end{aligned}$$

woraus die Behauptung folgt.

c) Erfüllt ein Maximum-Likelihood-Schätzer (R1), so hat die Funktion  $l(\cdot; \vartheta)$  die Gestalt  $l(x; \vartheta) = [I(\vartheta)]^{-1} \cdot \nabla_\vartheta \log f(x; \vartheta)$ . Setzt man dies und die Matrix  $C_\vartheta$  aus Teil b) in die Kovarianzmatrix  $\Sigma$  ein, so ergibt sich die Behauptung, wenn man noch beachtet, daß  $E_\vartheta[h(X; \vartheta)h(X; \vartheta)'] = I_k$  wegen der Orthonormalität der  $h_j$  gilt. ■

### Bemerkungen:

1. Teil a) des Satzes ist die multivariate Version der Verteilungsaussage für eine einzelne Komponente in Henze und Klar (1996). Teile a) und b) gelten auch ohne die Voraussetzung, daß die  $h_j(\cdot, \vartheta)$  ein OGS bilden, wie man am Beweis sieht.
2. Auch wenn die zugrundeliegende Verteilung nur endlich viele Wachstumspunkte besitzt, kann eine Teststatistik wie in (1.1) definiert wer-



den. Besteht die Menge der Wachstumspunkte etwa aus  $N + 1$  Punkten, so muß man nur darauf achten, daß in (1.1)  $k_0 + k \leq N$  ist (vgl. Bem. 3 nach Satz A.1.1). Satz 1.1.1 bleibt dann ohne Änderung gültig.

3. In der Literatur werden die glatten Anpassungstests als sogenannte „score tests“ eingeführt. Diese Tests sind asymptotisch optimale Tests gegenüber bestimmten parametrischen Alternativen, ähnlich wie der Likelihood-Quotienten- oder der Wald-Test (siehe Rayner und Best (1989), Kapitel 3). Diese Optimalitätseigenschaft gilt nur bei Verwendung des Maximum-Likelihood-Schätzers; nach der allgemeinen Verteilungstheorie für diese Tests ergibt sich dann gerade die Verteilung aus Teil c) als asymptotische Verteilung.

Allerdings können solche Familien von parametrischen Alternativen in vielen Fällen, etwa beim Test auf Normalverteilung, gar nicht definiert werden (siehe Mardia und Kent (1991), S.356, und Kallenberg et al. (1997), S.45), so daß diese Herleitung i. allg. nicht sinnvoll ist.

Die gemeinsame asymptotische Normalverteilung der Komponenten  $\hat{U}_{n,j}$  kann entartet sein. Insbesondere können Komponenten aufgrund des Schätzverfahrens identisch Null sein. Der nächste Hilfssatz zeigt, daß dies bei Momentenschätzern immer der Fall ist. Dabei verwenden wir die Bezeichnung  $\mu_j(\vartheta) = E_{\vartheta}[X^j]$ ,  $j \geq 1$ ;  $m_j = (1/n) \sum_{i=1}^n X_i^j$  ist das  $j$ -te empirische Moment. Der Momentenschätzer  $\bar{\vartheta}_n$  wird dann durch die Gleichungen  $\mu_j(\bar{\vartheta}_n) = m_j$  für  $j = 1, \dots, s$  definiert.

**1.1.2 Lemma** Sei  $\bar{\vartheta}_n$  der Momentenschätzer für  $\vartheta$ . Gilt  $\bar{\vartheta}_n \in \Theta$ , so ist

$$\hat{U}_{n,j} = 0 \quad \text{für } j = 1, \dots, s.$$

BEWEIS: Sei  $h_r(x; \vartheta) = \sum_{j=0}^r a_j(\vartheta) x^j$ . Aus  $E_{\vartheta}[h_r(X; \vartheta)] = 0$  für  $r \geq 1$  folgt

$$\sum_{j=0}^r a_j(\vartheta) \mu_j(\vartheta) = 0, \quad \vartheta \in \Theta. \quad (1.2)$$

Verwendet man als Schätzer für  $\vartheta$  den Momentenschätzer  $\bar{\vartheta}_n$  und gilt  $\bar{\vartheta}_n \in \Theta$ , so folgt für  $r = 1, \dots, s$  aus (1.2)

$$\begin{aligned} \frac{1}{\sqrt{n}} \hat{U}_{n,r} &= \frac{1}{n} \sum_{i=1}^n h_r(X_i; \bar{\vartheta}_n) \\ &= \sum_{j=0}^r a_j(\bar{\vartheta}_n) m_j \\ &= \sum_{j=0}^r a_j(\bar{\vartheta}_n) \mu_j(\bar{\vartheta}_n) = 0. \quad \blacksquare \end{aligned}$$

Verwendet man den „smooth test“ lediglich als Anpassungstest für das parametrische Modell  $P_\vartheta$ , so können anstelle des (nichtparametrischen) Momentenschätzers auch andere Schätzer verwendet werden. Dann untersucht beispielsweise die erste Komponente  $\hat{U}_{n,1} = \sqrt{n}(\bar{X}_n - \mu_1(\hat{\vartheta}_n))$ , ob durch  $\mu_1(\hat{\vartheta}_n)$  der Erwartungswert „vernünftig“ geschätzt wird, was im allgemeinen nicht der Fall ist, wenn  $P \notin \{P_\vartheta\}$  gilt. Ist  $\hat{U}_{n,1}$  zu groß oder zu klein, so spricht dies gegen das parametrische Modell. Will man den Test jedoch als „diagnostischen“ Test verwenden, so ist der Momentenschätzer für  $\vartheta$  zu verwenden, wie im folgenden näher erläutert wird.

Die  $s$  Gleichungen  $E_\vartheta[h_j(X, \vartheta)] = 0$ ,  $j = 1, \dots, s$ , legen die Momente  $\mu_1, \dots, \mu_s$  in Abhängigkeit von  $\vartheta_1, \dots, \vartheta_s$  fest. Wir nehmen im folgenden an, daß sich umgekehrt auch  $\vartheta$  eindeutig als Funktion der Momente  $\mu_1, \dots, \mu_s$  ausdrücken läßt. Auf der Menge

$$\mathcal{P} := \left\{ P : \int |x|^j P(dx) < \infty, j = 1, \dots, s, \vartheta = \vartheta(\mu_1, \dots, \mu_s) \in \Theta \right\}$$

definieren wir ein Funktional  $\delta : \mathcal{P} \rightarrow \Theta$  durch  $\delta(P) = \vartheta \quad \forall \vartheta \in \Theta$ . Für  $P \in \mathcal{P}$  gilt dann insbesondere

$$E_P[h_j(X, \delta(P))] = 0, \quad j = 1, \dots, s. \quad (1.3)$$

Damit nun  $\hat{U}_n$  für ein  $P \notin \{P_\vartheta\}$  in Verteilung konvergiert, ist es erforderlich, daß der Schätzer  $\hat{\vartheta}_n$  für  $\vartheta = \delta(P)$  die Regularitätsvoraussetzung

(R1) erfüllt. Dies wird für den Momentenschätzer i. allg. der Fall sein, nicht jedoch für einen speziell auf das parametrische Modell  $P_\vartheta$  zugeschnittenen Schätzer. Wir nehmen deshalb im folgenden an, daß als Schätzer für  $\vartheta$  der Momentenschätzer  $\bar{\vartheta}_n = \bar{\vartheta}_n(\mu_1, \dots, \mu_s)$  verwendet wird, so daß nach Lemma 1.1.2  $\hat{U}_{n,j} = 0$ ,  $j = 1, \dots, s$ , gilt.  $\bar{\vartheta}_n$  erfülle (R1). Der glatte Anpassungstest  $\hat{\Psi}_{n,k}^2$  besteht dann aus den Komponenten  $\hat{U}_{n,s+1}, \dots, \hat{U}_{n,s+k}$ , es gilt also  $k_0 = s$ . Im weiteren sei

$$T_j(P) := \int h_j(\cdot; \delta(P)) dP, \quad (1.4)$$

und wir definieren eine Familie von Verteilungen durch

$$\mathcal{P}_0 := \{P \in \mathcal{P} : T_{s+1}(P) = \dots = T_{s+k}(P) = 0, E_P[X^{2(s+k)}] < \infty\}.$$

$\mathcal{P}_0$  enthält offensichtlich  $\{P_\vartheta\}$ ; allen Verteilungen in  $\mathcal{P}_0$  ist gemeinsam, daß ihre Momente der Ordnung  $s+1$  bis  $s+k$  die gleichen Relationen erfüllen wie unter dem parametrischen Modell  $P_\vartheta$ . Die Verteilungsaussage in Satz 1.1.1 a) überträgt sich nun mit Beweis auf die gesamte Familie  $\mathcal{P}_0$ , wobei immer  $\vartheta = \delta(P)$  zu setzen ist (vgl. Abschnitt 3 in Henze und Klar (1996)).

Ein *gerichteter* glatter Anpassungstest hat zum Ziel, bei Ablehnung der Hypothese die Diagnose „es liegen Abweichungen in den Momenten  $\mu_{s+1}, \dots, \mu_{s+k}$  vor“ geben zu können; die eigentlich zu testende Hypothese ist also  $\tilde{H}_0 : P \in \mathcal{P}_0$  gegen  $\tilde{H}_1 : P \notin \mathcal{P}_0$ . Dieses Ziel kann nicht mit Hilfe der Teststatistik  $\hat{\Psi}_{n,k}^2$  aus (1.1) erreicht werden, da ihre asymptotische Verteilung vom zugrundeliegenden  $P \in \mathcal{P}_0$  abhängt. Um eine auf  $\mathcal{P}_0$  asymptotisch verteilungsfreie Teststatistik zu erhalten, muß  $\hat{\Psi}_{n,k}^2$  geeignet skaliert werden. Wie der nächste Satz zeigt, ist dies in den für die Anwendung der glatten Anpassungstests besonders wichtigen Fällen, in denen  $\hat{\Psi}_{n,k}^2$  unter  $H_0$   $\chi^2$ -verteilt ist, einfach zu erreichen. Darunter fallen insbesondere alle in Rayner und Best (1989) behandelten Fälle.

### 1.1.3 Satz

- a) Die Testgröße  $\hat{\Psi}_{n,k}^2$  sei unter der parametrischen Hypothese  $H_0$  asymptotisch  $\chi^2$ -verteilt mit  $k$  Freiheitsgraden. Weiter sei  $P \in \mathcal{P}_0$  eine Verteilung, deren Träger aus mindestens  $s+k+1$  Elementen besteht.

Dann ist  $\Sigma_P = E_P[h(X; \vartheta) h(X; \vartheta)']$  nichtsingulär, und  $\hat{U}'_n \Sigma_P^{-1} \hat{U}_n$  ist unter  $P$  ebenfalls asymptotisch  $\chi^2$ -verteilt mit  $k$  Freiheitsgraden.

b) Gilt  $n \geq s + k + 1$  und ist die zu  $P$  gehörende Verteilungsfunktion stetig, so ist die durch

$$\hat{H}_n = \frac{1}{n} \sum_{i=1}^n [h(X_i; \hat{\vartheta}_n) h(X_i; \hat{\vartheta}_n)']$$

definierte  $(k \times k)$ -Matrix mit Wahrscheinlichkeit 1 nichtsingulär. Die Testgröße

$$\tilde{\Psi}_{n,k}^2 := \hat{U}'_n \hat{H}_n^{-1} \hat{U}_n$$

ist unter  $P$  asymptotisch  $\chi^2$ -verteilt mit  $k$  Freiheitsgraden.

BEWEIS:

a) Nach Voraussetzung gilt unter  $H_0$  die Beziehung  $\Sigma = I_k$ . Wegen 1.1.1 a) ist dann  $E_\vartheta[\nabla_\vartheta h(X; \vartheta)]$  die Nullmatrix. Nun ist  $\partial h_r(x; \vartheta)/\partial \vartheta_j$  ein Polynom in  $x$  vom Grad  $r$ :

$$\frac{\partial h_r(x; \vartheta)}{\partial \vartheta_j} = \sum_{i=0}^r \frac{\partial a_i(\vartheta)}{\partial \vartheta_j} x^i = \sum_{i=0}^r c_i(\vartheta) h_i(x; \vartheta).$$

Wegen  $E_\vartheta[h_0(X; \vartheta)] = 1$ ,  $E_\vartheta[h_i(X; \vartheta)] = 0$ ,  $i \geq 1$ , folgt  $c_0(\vartheta) = 0$  für alle  $\vartheta \in \Theta$ , woraus für  $P \in \mathcal{P}_0$  wegen (1.3) und der Definition von  $\mathcal{P}_0$  wiederum  $E_P[\nabla_\vartheta h(X; \vartheta)|_{\vartheta=\delta(P)}] = 0$  folgt.

Die Kovarianzmatrix unter  $P$  ist also  $\Sigma_P = E_P[h(X; \vartheta) h(X; \vartheta)']$ . Wegen der Voraussetzung an den Träger von  $P$  gilt für  $c_{s+1}, \dots, c_{s+k}, \gamma \in \mathbb{R}$

$$P \left( \sum_{j=s+1}^{s+k} c_j h_j(X; \vartheta) = \gamma \right) = P \left( \sum_{j=0}^{s+k} c'_j X^j = \gamma \right) < 1,$$

die Matrix  $\Sigma_P$  ist also nichtsingulär, woraus mit bekannten Sätzen Behauptung folgt.

- b) Die Matrix  $\hat{H}_n$  ist die empirische Version von  $\Sigma_P$ . Die Aussage über die Verteilung von  $\tilde{\Psi}_{n,k}^2$  folgt also aus Teil a), wenn  $\hat{H}_n$  (mit Wahrscheinlichkeit 1) nichtsingulär ist.

Dazu sei  $n \geq s + k + 1$  und  $X_1, \dots, X_{s+k+1}$  paarweise verschieden. Ergänzt man die  $(k \times s + k + 1)$ -Matrix mit Zeilen

$$(h_{s+j}(X_1; \hat{\vartheta}_n), \dots, h_{s+j}(X_{s+k+1}; \hat{\vartheta}_n)) \quad (j = 1, \dots, k)$$

um  $s + 1$  Zeilen  $(X_1^j, \dots, X_{s+k+1}^j)$  ( $j = 0, \dots, s$ ), so läßt sich die resultierende Matrix durch elementare Zeilenoperationen zu einer (nichtsingulären) Vandermonde-Matrix mit  $s + k + 1$  Zeilen umformen. Somit hat die ursprüngliche Matrix den Rang  $k$ . Hieraus folgt, daß  $\hat{H}_n$  regulär ist. Dieses Ergebnis folgt auch aus Teil a), wenn man dort den Spezialfall einer Gleichverteilung auf  $\{X_1, \dots, X_{s+k+1}\}$  betrachtet. Da unter der Voraussetzung die  $X_i$  fast sicher paarweise verschieden sind, folgt Teil b) des Satzes. ■

### Bemerkungen:

1. Die Bedingung an den Träger der Verteilung in Teil a) des Satzes ist keine Einschränkung. Eine Verteilung mit  $(r + 1)$ -elementigem Träger ist durch  $r$  Momente eindeutig festgelegt. Ein glatter Anpassungstest mit  $r$  Komponenten ist dann ein konsistenter Test, es gilt also  $\mathcal{P}_0 = \{P_\vartheta\}$ .
2. Ist  $P$  eine diskrete Verteilung, so ist die Wahrscheinlichkeit, daß  $\hat{H}_n$  nichtsingulär ist, kleiner als eins. Da die Elemente von  $\hat{H}_n$  fast sicher gegen die Elemente von  $\Sigma_P$  konvergieren und die Determinante von  $\hat{H}_n$  stetig von diesen Elementen abhängt, existiert fast sicher ein  $n_0 \in \mathbb{N}$ , so daß  $\hat{H}_n$  nichtsingulär ist für alle  $n \geq n_0$ . Die Verteilungsaussage in b) bleibt also auch für diskrete Verteilungen, die die Voraussetzung von Teil a) erfüllen, richtig.

## 1.2 Glatte Anpassungstests in Exponentialfamilien

In diesem Abschnitt soll eine große Klasse von stetigen und diskreten parametrischen Verteilungsfamilien betrachtet werden, für die  $\hat{\Psi}_{n,k}^2$  wie in Satz 1.1.3 gefordert  $\chi^2$ -verteilt ist. Sei dazu  $P_\vartheta$  eine  $s$ -parametrische Exponentialfamilie mit Dichte

$$p_\vartheta(x) = C(\vartheta) \exp \left[ \sum_{j=1}^s \zeta_j(\vartheta) t_j(x) \right] h(x)$$

bezüglich eines  $\sigma$ -endlichen dominierenden Maßes  $\mu$ , wobei  $1, \zeta_1(\vartheta), \dots, \zeta_s(\vartheta)$  und (mit Wahrscheinlichkeit 1)  $1, t_1(x), \dots, t_s(x)$  linear unabhängige Funktionen sind.

Wir machen die zusätzliche Annahme, daß die  $t_j$  Polynome vom Grade kleiner oder gleich  $s$  sind. Ohne Einschränkung kann man dann annehmen, daß  $t_j(x) = x^j$  für  $1 \leq j \leq s$  gilt.

Zu dieser Klasse von Verteilungen gehören alle von Rayner und Best (1989) betrachteten Verteilungen: im stetigen Fall also die Normal- und die Exponentialverteilung (oder allgemeiner die Gammaverteilungen mit bekanntem Formparameter); im diskreten Fall die Poisson- und die Binomialverteilung sowie die geometrische Verteilung.

Ist der Parameter  $\vartheta$  eindimensional, so spricht man von einer *linearen Exponentialfamilie*. Besonders gut untersucht ist diese Klasse im diskreten Fall, wo die zugehörigen Verteilungen auch als „modified power series distributions“, kurz MPSD, bezeichnet werden, da ihre Zähldichten die Gestalt

$$p_\vartheta(x) = \frac{a_x [u(\vartheta)]^x}{\eta(\vartheta)}, \quad a_x \geq 0,$$

mit einer positiven (und i. allg. als differenzierbar vorausgesetzten) Funktion  $u(\vartheta)$  und der Potenzreihe  $\eta(\vartheta) = \sum_x a_x [u(\vartheta)]^x$  haben. Neben den oben aufgeführten diskreten Verteilungen enthält diese Klasse z.B. die negative Binomial- und die logarithmische Verteilung sowie die sogenannten Lagrange-Verteilungen; auch gehört jede abgeschnittene („truncated“)

MSPD wie etwa die positive Poissonverteilung wiederum zu dieser Klasse. Näheres zu dieser wichtigen Verteilungsklasse wie allgemeine Formeln für die Momente findet man in Johnson, Kotz und Kemp (1992).

Für Exponentialfamilien haben die Maximum-Likelihood-Gleichungen allgemein die Form

$$E_{\vartheta}[t_j(X)] = \frac{1}{n} \sum_{i=1}^n t_j(X_i), \quad j = 1, \dots, s,$$

(siehe z.B. Lehmann (1983), S. 439); für  $t_j(x) = x^j$  ergibt sich also  $E_{\vartheta}[X^j] = m_j$ , d.h. Maximum-Likelihood- und Momentenschätzer stimmen überein. Nach Lemma 1.1.2 verschwinden also die ersten  $s$  Komponenten. Wegen

$$\nabla_{\vartheta}[\log p_{\vartheta}(x)] = \nabla_{\vartheta}[\log(C(\vartheta))] + \sum_{j=1}^s t_j(x) \nabla_{\vartheta}[\zeta_j(\vartheta)] \quad (1.5)$$

ist  $\nabla_{\vartheta}[\log p_{\vartheta}(x)]$  in diesem Fall ein Polynom in  $x$  vom Grade  $s$ ; somit gilt

$$E_{\vartheta}[h_k(X, \vartheta) \nabla_{\vartheta} \log p_{\vartheta}(X)] = 0, \quad k > s, \quad (1.6)$$

wenn mit  $h_k(\cdot, \vartheta)$  wieder die orthogonalen Polynome vom Grad  $k$  bezeichnet werden. Mit Blick auf 1.1.1 c) erhält man das folgende Resultat.

### 1.2.1 Satz

Sei  $\{P_{\vartheta}\}$  eine  $s$ -parametrische Exponentialfamilie, die die Voraussetzungen von 1.1.1 c) erfüllt. Die Funktionen  $t_j(x)$  seien Polynome vom Grad kleiner oder gleich  $s$ . Wird als Schätzer für  $\vartheta$  der Maximum-Likelihood-Schätzer verwendet, so ist die Teststatistik  $\hat{\Psi}_n^2 = \sum_{j=s+1}^{s+k} \hat{U}_{n,j}^2$  asymptotisch  $\chi^2$ -verteilt mit  $k$  Freiheitsgraden.

Für die beschriebene Klasse von Verteilungen kann der zugehörige glatte Anpassungstest nach Satz 1.1.3 leicht so modifiziert werden, daß er (zumindest asymptotisch) diagnostische Eigenschaften besitzt; weichen die  $k$  untersuchten Momente stark von den Momenten der hypothetischen Verteilung ab, so lehnt der modifizierte Test die Nullhypothese bei „hinreichend

großem“ Stichprobenumfang ab. Dabei läßt sich nicht allgemein sagen, was „hinreichend groß“ bedeutet, da die Konvergenzgeschwindigkeit gegen die Grenzverteilung stark von der zugrundeliegenden Verteilung abhängt.

Für Verteilungen, die nicht zur oben beschriebenen Klasse gehören, ist die notwendige Modifikation schwieriger, wie man an Satz 1.1.1 sieht. Beispiele sind die von Boulerice und Ducharme (1995) erwähnte logistische Verteilung, die Laplace-, die Extremwert- und die Gammaverteilung mit unbekanntem Formparameter.

### 1.2.2 Beispiel

Als konkretes Beispiel soll hier der auf den beiden ersten nichtverschwindenden Komponenten aufbauende glatte Anpassungstest für die Normalverteilung betrachtet werden. Offensichtlich erfüllt die Normalverteilung die Voraussetzungen aus Satz 1.2.1. Die orthonormalen Polynome bezüglich der Standardnormalverteilung sind die normierten Hermite-Polynome; es ist

$$\begin{aligned} H_1(x) &= x, & H_2(x) &= (x^2 - 1)/\sqrt{2}, \\ H_3(x) &= (x^3 - 3x)/\sqrt{6}, & H_4(x) &= (x^4 - 6x^2 + 3)/\sqrt{24}. \end{aligned}$$

Für die zweiparametrische Normalverteilungsfamilie  $\mathcal{N}(\mu, \sigma^2)$  bilden dann  $h_j(x, \mu, \sigma) = H_j((x - \mu)/\sigma)$ ,  $j \geq 1$ , ein Orthonormalsystem. Mit den Elementen der Matrix  $\hat{H}_n$  aus Satz 1.1.3

$$\hat{H}_{jk} = \frac{1}{n} \sum_{i=1}^n h_j(X_i, \bar{x}_n, \hat{\sigma}_n) h_k(X_i, \bar{x}_n, \hat{\sigma}_n), \quad j, k \in \{3, 4\},$$

erhält man die innerhalb der zugehörigen Klasse

$$\mathcal{P}_0 = \{P \in \mathcal{P} : T_3(P) = T_4(P) = 0, E_P[X^8] < \infty\}$$

asymptotisch  $\chi_2^2$ -verteilte Testgröße

$$\tilde{\Psi}_{n,2}^2 = \frac{\hat{U}_{n,3}^2 \hat{H}_{44} - 2 \hat{U}_{n,3} \hat{U}_{n,4} \hat{H}_{34} + \hat{U}_{n,4}^2 \hat{H}_{33}}{\hat{H}_{33} \hat{H}_{44} - \hat{H}_{34}^2}.$$

Der auf dieser Testgröße aufbauende Test ist ein diagnostischer Test; Abweichungen in der Schiefe und/oder der Wölbung der vorliegenden Verteilung



gegenüber der Normalverteilung führen zumindest bei großem Stichprobenumfang zur Ablehnung der nichtparametrischen Hypothese  $\tilde{H}_0 : P \in \mathcal{P}_0$  : ist  $\alpha \in (0, 1)$  und bezeichnet  $\chi_{2,1-\alpha}^2$  das  $(1 - \alpha)$ -Quantil der  $\chi_2^2$ -Verteilung, so gilt

$$\lim_{n \rightarrow \infty} P\left(\tilde{\Psi}_{n,2}^2 > \chi_{2,1-\alpha}^2\right) = 1.$$

### 1.3 Glatte Anpassungstests für die mehrdimensionale Normalverteilung

In diesem Abschnitt seien  $X_1, \dots, X_n$  unabhängige Kopien des  $d$ -dimensionalen Zufallsvektors  $X$  mit Erwartungswert  $\mu$  und nichtsingulärer Kovarianzmatrix  $T$ . Weiter sei  $\tilde{X} = T^{-1/2}(X - \mu)$  der normierte Zufallsvektor mit  $E[\tilde{X}] = 0$  und  $E[\tilde{X}\tilde{X}'] = I_d$ , wobei  $I_d$  die  $d$ -dimensionale Einheitsmatrix bezeichnet.  $Y$  und  $\tilde{Y}$  seien unabhängige Kopien von  $X$  bzw.  $\tilde{X}$ . Weiter seien  $\tilde{X} = (\xi_1, \dots, \xi_d)'$  und  $\tilde{Y} = (\eta_1, \dots, \eta_d)'$ .

Die im folgenden vorgestellten glatten Anpassungstests verwenden immer die empirisch standardisierten Größen  $Z_j = S_n^{-1/2}(X_j - \bar{X}_n)$ ,  $j = 1, \dots, n$ , wobei

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j, \quad S_n = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)'$$

den Mittelwert bzw. die empirische Kovarianzmatrix von  $X_1, \dots, X_n$  bezeichnen. Dabei sei  $S_n$  mit Wahrscheinlichkeit 1 nichtsingulär. Eine hinreichende Bedingung dafür ist, daß  $X_1$  eine  $\lambda^d$ -Dichte besitzt und  $n \geq d + 1$  ist (Eaton und Perlman(1973)). Die einzelnen Komponenten von  $Z_j$  werden mit  $Z_{rj}$ ,  $r = 1, \dots, d$ , bezeichnet. Getestet werden soll die Hypothese

$H_0 : X$  besitzt eine  $d$ -dimensionale nichtausgeartete Normalverteilung.

Dazu wird ein System von orthonormalen Polynomen wie folgt definiert.  $H_k$  seien die normierten Hermite-Polynome aus Beispiel 1.2.2; insbesondere

ist also  $H_k$  ein Polynom  $k$ -ten Grades. Nun werden multivariate Polynome durch

$$L_{k_1, \dots, k_d}(\tilde{X}) = H_{k_1}(\xi_1) \cdots H_{k_d}(\xi_d), \quad k_1, \dots, k_d \in \mathbb{N}_0,$$

definiert; die durch eine (beliebige) Anordnung entstehende Folge bezeichnen wir mit  $\{L_r\}$ . Wegen

$$E[L_r(\tilde{X})L_s(\tilde{X})] = \prod_{i=1}^d E[H_{k_i}(\xi_i)H_{m_i}(\xi_i)] = \prod_{i=1}^d \delta_{k_i m_i} =: \delta_{rs}$$

ist die Folge orthonormal. Der Grad eines Polynoms ist durch die Summe  $k = k_1 + \dots + k_d$  gegeben. Üblicherweise werden die Polynome nach dem Grad geordnet; insbesondere ist dann  $L_0 = 1$ . Glatte Anpassungstests basieren auf der Teststatistik

$$\hat{\Psi}_{n,k}^2 = \sum_r \hat{V}_{n,r}^2, \quad \text{mit } \hat{V}_{n,r} = \frac{1}{\sqrt{n}} \sum_{j=1}^n L_r(Z_j), \quad (1.7)$$

wobei nur diejenigen  $L_r$  in der Summe auftreten, deren Grad positiv und kleiner oder gleich  $k$  ist. Summiert man in (1.7) nur über die  $L_r$ , die genau den Grad  $k$  besitzen, so erhält man die  $k$ -te Komponente  $\hat{U}_{n,k}^2$  des glatten Anpassungstests.

Man kann sich leicht überlegen, daß alle zu Polynomen vom Grade 1 oder 2 gehörenden  $\hat{V}_{n,r}$  gleich Null sind; besitzt  $L_r$  z.B. den Grad 1, so hat  $\hat{V}_{n,r}$  die Gestalt

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n H_1(Z_{rj}) = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_{rj}, \quad r \in \{1, \dots, d\};$$

da die Größen  $Z_j$  standardisiert sind, ist die Summe gerade Null. Wie in Lemma 1.1.2 ist hier entscheidend, daß als Schätzer für die unbekannt Parameter die Momentenschätzer verwendet werden, die im vorliegenden Fall mit dem Maximum-Likelihood-Schätzer übereinstimmen. Es gilt also  $\hat{U}_{n,1}^2 = \hat{U}_{n,2}^2 = 0$ . Somit läßt sich  $\hat{\Psi}_{n,k}^2$  in der einprägsamen Form

$$\hat{\Psi}_{n,k}^2 = \hat{U}_{n,3}^2 + \dots + \hat{U}_{n,k}^2$$

schreiben.

Für die Bestimmung der asymptotischen Verteilung der Komponente vom Grad  $k$  unter der Hypothese muß bekannt sein, aus wie vielen Polynomen  $L_r$  sie gebildet wird. Diese Anzahl ist gleich der Zahl der Abbildungen  $f$  von  $\{1, \dots, d\}$  nach  $\mathbb{N}_0$  mit der Eigenschaft  $\sum_{j=1}^d f(j) = k$ ; diese Anzahl ist bekanntlich  $\binom{k+d-1}{k}$ . Somit gibt es

$$\binom{d}{1} + \binom{d+1}{2} + \dots + \binom{k+d-1}{k} = \binom{k+d}{k} - 1$$

Polynome vom Grade kleiner oder gleich  $k$  (ohne  $L_0$ ). Da die ersten beiden Komponenten verschwinden, setzt sich der glatte Anpassungstest  $\hat{\Psi}_{n,k}^2$  also aus  $\binom{k+d}{k} - \binom{2+d}{2}$  Polynomen zusammen.

Wie sehen die multivariaten Polynome vom Grad  $k$  genauer aus? Sie sind Produkte aus  $d$  univariaten Polynomen, wobei etwa  $m$  Polynome einen von Null verschiedenen Grad  $k_j$  haben. Hält man  $m$  Variablen fest, so gibt es  $\binom{k-1}{k-m}$  Polynome  $L_r$ , für die für die entsprechenden  $m$  Indizes  $k_j > 0$  gilt (und  $k_j = 0$  sonst). Da es  $\binom{d}{m}$  Möglichkeiten gibt,  $m$  Variablen auszuwählen, ergibt sich als Gesamtzahl der Polynome  $L_r$  vom Grade  $k$

$$\sum_{m=1}^d \binom{k-1}{k-m} \binom{d}{m} = \binom{k+d-1}{k},$$

also wieder das Ergebnis von oben.

Um das asymptotische Verhalten der glatten Anpassungstests unter  $H_0$  zu bestimmen, kann man wie in Satz 1.1.1 vorgehen. Als Parameter können neben dem Erwartungswertvektor  $\mu$  etwa die Elemente von  $T^{-1}$  verwendet werden. In den folgenden Abschnitten zeigt sich, daß jede Komponente aus Produkten der Form  $Z_i' Z_j$  aufgebaut ist. Da die  $Z_i$  standardisierte Größen sind, sieht man an diesen Darstellungen der Komponenten, daß es sich um affin-invariante Größen handelt; es gilt also

$$\hat{U}_{n,k}^2(X_1, \dots, X_n) = \hat{U}_{n,k}^2(b + BX_1, \dots, b + BX_n)$$

für jede nichtsinguläre  $(d \times d)$ -Matrix  $B$  und jeden Vektor  $b \in \mathbb{R}^d$ . Da mit  $\hat{U}_{n,k}^2$  auch die Testgrößen  $\hat{\Psi}_{n,k}^2$  affin-invariant sind, soll im folgenden

o.B.d.A. immer  $\mu = 0$  und  $T = I_d$  angenommen werden. Man beachte, daß die Eigenschaft der Affin-Invarianz i. allg. nicht erhalten bleibt, wenn eine Teststatistik nur auf einer Teilmenge aller zur Bildung einer Komponente benötigten Polynome aufbaut.

Unter der Annahme  $\mu = 0$  und  $T = I_d$  folgt wegen

$$\begin{aligned}\sqrt{n}(\bar{X}_n - \mu) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i + o_P(1), \\ \sqrt{n}(S_n^{-1} - I_d) &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i X_i' - I_d) + o_P(1)\end{aligned}$$

(siehe etwa Baringhaus und Henze (1992a)), daß die Regularitätsvoraussetzung **(R1)** erfüllt ist. Als gemeinsame asymptotische Verteilung von  $\hat{V}_{n,1}, \dots, \hat{V}_{n,r}$  ergibt sich folglich wieder eine Normalverteilung  $\mathcal{N}(0, \Sigma)$  mit einer gewissen Kovarianzmatrix  $\Sigma$ . Nun läßt sich durch direkte Rechnung zeigen (Rayner und Best (1989), S.101ff.), daß die in Satz 1.1.1 b) auftretenden Größen  $c_{ij}$  gleich Null sind, so daß gilt:

$$\Sigma = E[(L_1, \dots, L_r)(L_1, \dots, L_r)'] = I_r.$$

Man kann dies auch ähnlich wie in Abschnitt 1.2 zeigen: Die  $d$ -dimensionale Normalverteilung ist eine  $s$ -parametrische Exponentialfamilie mit  $s = d(d+3)/2$ . Die Funktionen  $t_j$ ,  $1 \leq j \leq d(d+3)/2$  sind lineare oder quadratische Polynome; für  $d = 2$  gilt etwa  $t_1(x) = x_1$ ,  $t_2(x) = x_1^2$ ,  $t_3(x) = x_1 x_2$ ,  $t_4(x) = x_2$ ,  $t_5(x) = x_2^2$ , ( $x = (x_1, x_2)'$ ). Wie oben erwähnt, verschwinden die zu den Orthogonalpolynomen ersten bzw. zweiten Grades gehörende erste bzw. zweite Komponente. Mit derselben Argumentation wie in Abschnitt 1.2 (insbesondere bleibt (1.5) gültig) folgt (1.6). Dies bedeutet aber gerade, daß die Größen  $c_{ij}$  verschwinden.

Somit ist die asymptotische Verteilung einer Komponente  $\hat{U}_{n,k}^2$  bzw. von  $\hat{\Psi}_{n,k}^2$  unter der Hypothese der multivariaten Normalverteilung eine  $\chi^2$ -Verteilung, deren Freiheitsgrad gleich der Anzahl der Polynome ist, aus denen  $\hat{U}_{n,k}^2$  bzw.  $\hat{\Psi}_{n,k}^2$  aufgebaut sind. Beispielsweise ist die asymptotische Verteilung von  $\hat{U}_{n,3}^2$  eine  $\chi^2$ -Verteilung mit  $\binom{d+2}{3}$  Freiheitsgraden; diejenige

von  $\hat{\Psi}_{n,4}^2$  ist eine  $\chi^2$ -Verteilung mit  $\binom{d+2}{3} + \binom{d+3}{4} = d(d+1)(d+2)(d+7)/24$  Freiheitsgraden.

Nun betrachten wir beliebige Verteilungen, wobei die auftretenden Momente immer existieren mögen. Im folgenden sei  $\tau = (\tau_1, \dots, \tau_r)' = (E[L_1], \dots, E[L_r])'$ . Wie in Abschnitt 1.1 läßt sich zeigen, daß

$$(\hat{V}_{n,1}, \dots, \hat{V}_{n,r})' - \tau \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma) \quad (1.8)$$

mit einer gewissen Kovarianzmatrix  $\Sigma$  gilt. Dabei bezeichnet  $\xrightarrow{\mathcal{D}}$  die Verteilungskonvergenz. Definiert man

$$v_i(x, \vartheta) = L_i(x) + E[\nabla_{\vartheta} L_i(\tilde{X})] l(x, \vartheta), \quad i = 1, \dots, r,$$

(man beachte, daß  $L_i(\tilde{X})$  durch die Standardisierung von  $\vartheta$  abhängt), so ergibt sich für  $\Sigma$

$$\Sigma = E \left[ \left( v_1(\tilde{X}, \vartheta), \dots, v_r(\tilde{X}, \vartheta) \right) \left( v_1(\tilde{X}, \vartheta), \dots, v_r(\tilde{X}, \vartheta) \right)' \right] - \tau \tau'. \quad (1.9)$$

Ähnlich wie im eindimensionalen Fall definieren wir eine Menge  $\mathcal{P}_0^k$  von Wahrscheinlichkeitsverteilungen auf (den Borelmengen des)  $\mathbb{R}^d$  durch

$$\mathcal{P}_0^k := \{P : E_P[L_j(\tilde{X})] = 0 \text{ für jedes Polynom } L_j \text{ vom Grade kleiner oder gleich } k, E_P\|X\|^{2k} < \infty\}; \quad (1.10)$$

$\mathcal{P}_0^k$  enthält also diejenigen Verteilungen, für welche der Erwartungswert jedes multivariaten Polynoms  $L_j$  vom maximalen Grad  $k$  verschwindet. Ist  $L_j$  ein derartiges Polynom, so vereinfacht sich die Kovarianzmatrix in (1.9) für Verteilungen aus  $\mathcal{P}_0^k$  zu

$$\Sigma = E \left[ \left( L_1(\tilde{X}), \dots, L_r(\tilde{X}) \right) \left( L_1(\tilde{X}), \dots, L_r(\tilde{X}) \right)' \right]. \quad (1.11)$$

Dies ergibt sich analog zum univariaten Fall (siehe den Beweis von Satz 1.1.3), wenn man beachtet, daß sich die Ableitung von  $L_j$  nach  $\vartheta$  wieder mit Hilfe von Polynomen höchstens  $k$ -ter Ordnung ausdrücken läßt.

Um die Verteilung einer Komponente zu bestimmen, unterscheiden wir die beiden Fälle  $\tau = 0$  und  $\tau \neq 0$ . Im Fall  $\tau = 0$  gilt bekanntlich

$$\hat{U}_{n,k}^2 \xrightarrow{\mathcal{D}} \sum_{j=1}^r N_j^2, \quad (1.12)$$

wobei  $(N_1, \dots, N_r)' \sim \mathcal{N}(0, \Sigma)$ . Die Grenzverteilung läßt sich auch als gewichtete Summe von unabhängigen  $\chi_1^2$ -Zufallsvariablen mit nicht-negativen Gewichten darstellen; die Gewichte sind dabei gerade die Eigenwerte der Matrix  $\Sigma$ . Beispiele werden im nächsten Abschnitt angegeben.

Im Falle  $\tau \neq 0$  gilt dagegen (siehe etwa Serfling (1980), Corollary 3.3)

$$\sqrt{n} \left( \frac{\hat{U}_{n,k}^2}{n} - \tau' \tau \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, 4 \sum_{i,j=1}^r \sigma_{ij} \tau_i \tau_j \right), \quad (1.13)$$

mit den Elementen  $\sigma_{ij}$  der Kovarianzmatrix  $\Sigma$ .

Um in den folgenden Abschnitten die einzelnen Komponenten genauer zu untersuchen, werden die Hermite-Polynome bis zur Ordnung 7 benötigt. Es gilt

$$\begin{aligned} H_3(x) &= (x^3 - 3x)/\sqrt{3!}, & H_4(x) &= (x^4 - 6x^2 + 3)/\sqrt{4!}, \\ H_5(x) &= (x^5 - 10x^3 + 15x)/\sqrt{5!}, & H_6(x) &= (x^6 - 15x^4 + 45x^2 - 15)/\sqrt{6!}, \\ H_7(x) &= (x^7 - 21x^5 + 105x^3 - 105x)/\sqrt{7!}. \end{aligned}$$

### 1.3.1 Multivariate Schiefe

Als affin-invariantes Maß für die multivariate Schiefe wurde von Mardia (1970) der Ausdruck

$$b_{1,d} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (Z_i' Z_j)^3$$

eingeführt. In Komponentenschreibweise ( $Z_i' = (z_{1i}, z_{2i}, \dots, z_{di})$ ) ergibt sich

$$b_{1,d} = \frac{1}{n^2} \left\{ \sum_{r=1}^d \left( \sum_i z_{ri}^3 \right)^2 + 3 \sum_{r \neq s} \left( \sum_i (z_{ri}^2 z_{si}) \right)^2 \right\}$$

$$+ 6 \sum_{1 \leq r < s < t \leq d} \left( \sum_i (z_{ri} z_{si} z_{ti}) \right)^2 \Bigg\}. \quad (1.14)$$

Die entsprechende Verteilungsgröße („population parameter“) ist

$$\begin{aligned} \beta_{1,d} &= E[(\tilde{X}'\tilde{Y})^3] \\ &= \sum_{r=1}^d (E[\xi_r^3])^2 + 3 \sum_{r \neq s} (E[\xi_r^2 \xi_s])^2 + 6 \sum_{1 \leq r < s < t \leq d} (E[\xi_r \xi_s \xi_t])^2. \end{aligned} \quad (1.15)$$

Die Komponente  $\hat{U}_{n,3}^2$  setzt sich aus allen multivariaten Polynomen der Ordnung 3, also aus  $H_3(y_j)$ ,  $H_2(y_j)H_1(y_k)$  und  $H_1(y_j)H_1(y_k)H_1(y_l)$ , zusammen, wobei die Indizes paarweise verschieden sind. Man erhält

$$\begin{aligned} \hat{U}_{n,3}^2 &= \frac{1}{n} \left\{ \sum_r \frac{1}{6} \left( \sum_i (z_{ri}^3 - 3z_{ri}) \right)^2 + \sum_{r \neq s} \frac{1}{2} \left( \sum_i (z_{ri}^2 - 1)z_{si} \right)^2 \right. \\ &\quad \left. + \sum_{r < s < t} \left( \sum_i (z_{ri} z_{si} z_{ti}) \right)^2 \right\} \end{aligned}$$

und somit durch Vergleich mit (1.14) die Identität

$$\frac{n}{6} b_{1,d} = \hat{U}_{n,3}^2 \quad (1.16)$$

(Koziol (1987)). Wie oben erläutert ist  $\hat{U}_{n,3}^2$  asymptotisch  $\chi^2$ -verteilt mit  $\binom{d+2}{3}$  Freiheitsgraden, wenn die zugrundeliegende Verteilung eine multivariate Normalverteilung ist.

### 1.3.1 Beispiel

Als konkretes Beispiel soll hier die asymptotische Verteilung der dritten Komponente im Falle einer zugrundeliegenden elliptisch-symmetrischen Verteilung bestimmt werden. Eine Zusammenfassung einiger wichtiger Eigenschaften dieser Verteilungen findet man in Anhang A.2.

Die elliptischen Verteilungen liegen in der Klasse  $\mathcal{P}_0^3$  aller Verteilungen, für die  $\beta_{1,d} = 0$  ist (siehe (A.4) in Anhang A.2). Die Kovarianzmatrix  $\Sigma$

berechnet sich also nach Gleichung (1.11). Die dritte Komponente wird aus  $L_r(\tilde{X}) = H_3(\xi_r)$ ,  $L_{rs}(\tilde{X}) = H_2(\xi_r)H_1(\xi_s)$  ( $r \neq s$ ) und aus  $L_{rst}(\tilde{X}) = H_1(\xi_r)H_1(\xi_s)H_1(\xi_t)$  ( $r < s < t$ ) gebildet ( $r, s, t \in \{1, \dots, d\}$ ).

Sei im folgenden  $\mu_{s_1, \dots, s_d} = E\left[\prod_{i=1}^d X_i^{s_i}\right]$ , wobei auftretende Nullen nicht notiert werden, da die Anordnung der  $s_i$  nach Satz A.2.3 keine Rolle spielt. Es gilt also z.B.  $E[X_2^2 X_3^4] = \mu_{0240\dots 0} = \mu_{42}$ . Damit erhält man die Kovarianzterme

$$\begin{aligned} \sigma^1 &= E[L_r^2(\tilde{X})] = \frac{1}{6}E[(\xi_1^3 - 3\xi_1)^2] = \frac{1}{6}(\mu_6 - 6\mu_4 + 9), \\ \sigma^{12} &= E[L_r(\tilde{X})L_{sr}(\tilde{X})] = \frac{1}{2\sqrt{3}}(\mu_{42} - 3\mu_{22} - \mu_4 + 3), \\ \sigma^2 &= E[L_{rs}^2(\tilde{X})] = \frac{1}{2}E[(\xi_1^2 - 1)^2 \xi_2^2] = \frac{1}{2}(\mu_{42} - 2\mu_{22} + 1), \\ \sigma^{22} &= E[L_{rs}(\tilde{X})L_{ts}(\tilde{X})] = \frac{1}{2}(\mu_{222} - 2\mu_{22} + 1), \\ \sigma^3 &= E[L_{rst}^2(\tilde{X})] = \mu_{222}. \end{aligned}$$

Die übrigen Kovarianzterme bestehen aus Momenten, die nach Satz A.2.3 verschwinden. Ordnet man die Polynome in der Reihenfolge

$$\begin{aligned} &L_1, L_{21}, L_{31}, \dots, L_{d1}, \\ &L_2, L_{12}, L_{32}, \dots, L_{d2}, \\ &\dots\dots\dots \\ &L_d, L_{1d}, L_{2d}, \dots, L_{d-1,d}, \\ &L_{123}, L_{124}, L_{125}, \dots, L_{d-2,d-1,d}, \end{aligned}$$

so verschwinden alle Kovarianzterme zwischen Polynomen verschiedener Zeilen, d.h. die Matrix zerfällt in  $(d + 1)$  zu den einzelnen Zeilen gehörende Teilmatrizen.

Zur Vereinfachung soll im weiteren die Kovarianzmatrix der zu  $6 \hat{U}_{n,3}^2$  gehörenden Polynome betrachtet werden. Beachtet man die in Korollar A.2.4 angegebenen Relationen und setzt  $u = 6\sigma^{22} = \mu_6/5 - 2\mu_4 + 3$  und  $v = 6\sigma^2 = 3/5 \mu_6 - 2\mu_4 + 3$  und damit  $6\sigma^1 = 2u + v$  und  $6\sigma^{12} = \sqrt{3}u$ , so



erhält man die zu einer der ersten  $d$  Zeilen gehörende  $(d \times d)$ -Teilmatrix

$$\Sigma_{1,d} = \left( \begin{array}{c|cccc} 2u+v & \sqrt{3}u & \sqrt{3}u & \cdots & \sqrt{3}u \\ \hline \sqrt{3}u & v & u & \cdots & u \\ \sqrt{3}u & u & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & u \\ \sqrt{3}u & u & \cdots & u & v \end{array} \right). \quad (1.17)$$

Aus der charakteristischen Gleichung dieser Matrix ergibt sich durch einfache Umformungen die äquivalente Gleichung

$$\det \left( \begin{array}{cc|ccc} 2u+v-\lambda & (d-1)\sqrt{3}u & \sqrt{3}u & \cdots & \sqrt{3}u \\ \hline \sqrt{3}u & (d-2)u+v-\lambda & u & \cdots & u \\ 0 & 0 & v-u-\lambda & & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & v-u-\lambda \end{array} \right) = 0.$$

Zwei Eigenwerte sind also Lösungen der Gleichung

$$(2u+v-\lambda)((d-2)u+v-\lambda) - 3(d-1)u^2 = 0;$$

man erhält  $\lambda_1 = v-u = 2\mu_6/5$  und  $\lambda_2 = (d+1)u+v = (d+4)\mu_6/5 - 2(d+2)\mu_4 + 3(d+2)$ . Weiter ist  $\lambda_3 = v-u = 2\mu_6/5$  ein  $(d-2)$ -facher Eigenwert.

Die zur letzten Zeile gehörende Teilmatrix ist eine Diagonalmatrix mit Einträgen  $6\sigma^3$ ; sie besitzt folglich den  $\binom{d}{3}$ -fachen Eigenwert  $\lambda_4 = 6\mu_{222} = 2\mu_6/5$ .

Die Grenzverteilung von  $6\hat{U}_{n,3}^2$  ist nach (1.12) eine gewichtete Summe von  $\binom{d+2}{3}$  unabhängigen  $\chi_1^2$ -verteilten Zufallsvariablen. Diese lassen sich zu zwei  $\chi^2$ -verteilten Zufallsvariablen mit entsprechenden Freiheitsgraden zusammenfassen, da nur zwei verschiedene Gewichte auftreten. Insgesamt ergibt sich folgendes Resultat:

### 1.3.2 Satz (vgl. Theorem 2.2, Baringhaus und Henze (1992a))

Der Zufallsvektor  $X$  sei elliptisch-symmetrisch verteilt mit Erwartungswert  $\mu$  und nichtsingulärer Kovarianzmatrix  $T$ . Weiter gelte  $E\{(X-\mu)'T^{-1}(X-\mu)\} = d$ .

$\mu\}^3] < \infty$ , d.h.  $\mathcal{V}(X) \in \mathcal{P}_0^3$ , wobei  $\mathcal{V}(X)$  für die Verteilung von  $X$  steht. Dann gilt

$$6 \hat{U}_{n,3}^2 \xrightarrow{\mathcal{D}} \alpha_1 \chi_{\nu_1}^2 + \alpha_2 \chi_{\nu_2}^2;$$

dabei ist

$$\begin{aligned} \alpha_1 &= \frac{2}{5} \mu_6, & \nu_1 &= d(d-1) + \binom{d}{3} = \frac{d}{6}(d-1)(d+4), \\ \alpha_2 &= \frac{d+4}{5} \mu_6 - 2(d+2) \mu_4 + 3(d+2), & \nu_2 &= d, \end{aligned}$$

und  $\chi_{\nu_i}^2$  sind unabhängige  $\chi^2$ -verteilte Zufallsvariablen mit  $\nu_i$  Freiheitsgraden. ■

**Bemerkung:** Definiert man  $r_{2k} = E[(\tilde{X}'\tilde{X})^k]$  und beachtet, daß für elliptisch-symmetrische Verteilungen

$$r_4 = \frac{\mu_4}{3}d(d+2), \quad r_6 = \frac{\mu_6}{15}d(d+2)(d+4),$$

gilt (siehe auch Satz 1.4.7 unten), so erhält man

$$\alpha_1 = \frac{6r_6}{d(d+2)(d+4)}, \quad \alpha_2 = \frac{3}{d} \left( \frac{r_6}{d+2} - 2r_4 + d(d+2) \right).$$

Wegen Gleichung (1.16) entspricht Satz 1.3.2 also dem auf ganz anderem Wege bewiesenen Theorem 2.2 in Baringhaus und Henze (1992a). Dort wird als weitere Voraussetzung  $P(X = \mu) = 0$  verlangt (in diesem Fall ist  $\tilde{X}/\|\tilde{X}\|$  gleichverteilt auf der Oberfläche der Einheitskugel im  $\mathbb{R}^d$ ). Beim hier gezeigten Beweis wird von dieser Tatsache kein Gebrauch gemacht, so daß man ohne diese Voraussetzung auskommt.

### 1.3.3 Beispiel

In Móri et al. (1993) wurde als weitere multivariate Verallgemeinerung der Schiefe die affin-invariante Größe

$$\tilde{b}_{1,d} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n Z_i' Z_j \|Z_i\|^2 \|Z_j\|^2$$

$$\begin{aligned}
&= \frac{1}{n^2} \left\{ \sum_{r=1}^d \left( \sum_i z_{ri}^3 \right)^2 + 2 \sum_{r \neq s} \left( \sum_i z_{ri}^3 \sum_j z_{rj} z_{sj}^2 \right) \right. \\
&\quad \left. + \sum_{r \neq s} \left( \sum_i z_{ri}^2 z_{si} \right)^2 + \sum_{r \neq s \neq t} \left( \sum_i z_{ri} z_{si}^2 \sum_j z_{rj} z_{tj}^2 \right) \right\} \quad (1.18)
\end{aligned}$$

betrachtet. Die zugehörige Verteilungsgröße ist

$$\tilde{\beta}_{1,d} = \| E(\tilde{X} \|\tilde{X}\|^2) \|^2 = E[(\tilde{X}'\tilde{Y})(\tilde{X}'\tilde{X})(\tilde{Y}'\tilde{Y})].$$

Definiert man  $\hat{V}_r = (1/\sqrt{n}) \sum_{i=1}^n L_r(Z_i)$  und  $\hat{V}_{rs} = (1/\sqrt{n}) \sum_{i=1}^n L_{rs}(Z_i)$  mit den Polynomen  $L_r$  und  $L_{rs}$  aus Beispiel 1.3.1, so folgt aus (1.18)

$$n \tilde{b}_{1,d} = 6 \sum_r \hat{V}_r^2 + 4\sqrt{3} \sum_{r \neq s} \hat{V}_r \hat{V}_{sr} + 2 \sum_{r \neq s} \hat{V}_{rs}^2 + 2 \sum_{r \neq s \neq t} \hat{V}_{sr} \hat{V}_{tr}.$$

Definiert man weiter  $W_r = (\hat{V}_r, \hat{V}_{1r}, \dots, \hat{V}_{r-1,r}, \hat{V}_{r+1,r}, \dots, \hat{V}_{dr})'$ , so folgt die Darstellung

$$n \tilde{b}_{1,d} = W_1' A W_1 + \dots + W_d' A W_d,$$

wobei die  $(d \times d)$ -Matrix  $A$  die Einträge  $a_{11} = 6, a_{1k} = a_{k1} = 2\sqrt{3}$  ( $1 < k \leq d$ ) und  $a_{kl} = 2$ , sonst, besitzt.

Mit Hilfe dieser Darstellung kann die asymptotische Verteilung von  $n \tilde{b}_{1,d}$  einfach ermittelt werden, wenn die zugrundeliegende Verteilung wie im vorherigen Beispiel eine elliptisch-symmetrische Verteilung ist.

### 1.3.4 Satz (vgl. Theorem 2.2 in Henze (1997b))

$X$  sei elliptisch-symmetrisch verteilt mit Erwartungswert  $\mu$  und nichtsingulärer Kovarianzmatrix  $T$ . Es gelte  $E[\{(X - \mu)'T^{-1}(X - \mu)\}^3] < \infty$ . Dann gilt

$$n \tilde{b}_{1,d} \xrightarrow{\mathcal{D}} \frac{(d+2) \alpha_2}{3} \chi_d^2;$$

dabei ist  $\alpha_2 = (d+4)\mu_6/5 - 2(d+2)\mu_4 + 3(d+2)$  wie in Satz 1.3.2.

BEWEIS: In Beispiel 1.3.1 wurde gezeigt, daß  $W_r$  und  $W_s$  ( $r \neq s$ ) unkorreliert und somit asymptotisch unabhängig sind; die Kovarianzmatrix von  $W_r$  ist  $\Sigma_{1,d}/6$  mit  $\Sigma_{1,d}$  in (1.17). Wir betrachten deshalb o.E. die quadratische Form  $W_1'AW_1$ ; diese ist nach bekannten Sätzen asymptotisch verteilt wie eine gewichtete Summe aus unabhängigen  $\chi_1^2$ -Zufallsvariablen, wobei die Gewichte die Eigenwerte der  $(d \times d)$ -Matrix  $\Sigma_{1,d}A/6$  sind.

Definiert man den Vektor  $e_d \in \mathbb{R}^d$  durch  $e_d = (\sqrt{3}, 1, \dots, 1)'$ , so gilt  $A = 2e_d e_d'$ . Außerdem prüft man leicht nach, daß  $e_d$  gerade der Eigenvektor von  $\Sigma_{1,d}$  zum Eigenwert  $\alpha_2$  ist. Folglich gilt  $\Sigma_{1,d}A = 2\alpha_2 e_d e_d'$ . Also ist  $(d+2)\alpha_2/3$  einfacher Eigenwert von  $\Sigma_{1,d}A/6$  mit Eigenvektor  $e_d$ ; 0 ist  $(d-1)$ -facher Eigenwert. ■

**Bemerkung:** 1. Aus dem Beweis folgt die Darstellung

$$n\tilde{b}_{1,d} = 2 \sum_{r=1}^d (W_r' e_d)^2;$$

die Schiefe-Statistik nach Móri et al. berücksichtigt also nur die Projektionen der  $W_r$  in die Eigenrichtung  $e_d$ .

2. Der Beweis kommt ohne die in Theorem 2.2 in Henze (1997b) gemachte Voraussetzung  $P(X = \mu) = 0$  aus (vgl. die Bemerkung nach Satz 1.3.2).

3. Für eine Verteilung mit  $\tilde{\beta}_{1,d} > 0$  erhält man mit Hilfe von Korollar 3.3 in Serfling (1980)

$$\sqrt{n}(\tilde{b}_{1,d} - \tilde{\beta}_{1,d}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau'(A + \text{diag}(A))' \Sigma (A + \text{diag}(A)) \tau).$$

### 1.3.5 Beispiel

Die Schiefe-Statistik nach Mardia ist nach Satz 1.3.2 innerhalb der Klasse der elliptisch-symmetrischen Verteilungen nicht asymptotisch verteilungsfrei. Aus der Schiefe nach Móri et al. kann man zwar durch eine Modifikation eine asymptotisch verteilungsfreie Testgröße erhalten; das vorhergehende Beispiel zeigte aber, daß dies durch eine Projektion der Vektoren  $W_r$  in eine bestimmte Richtung erreicht wurde. Es liegt nun nahe, eine neue Schiefe-Statistik derart zu definieren, daß sie wie die Mardia-Schiefe im Falle der

Normalverteilung alle Polynome dritten Grades gleichermaßen berücksichtigt und gleichzeitig in der Klasse der elliptisch-symmetrischen Verteilungen asymptotisch verteilungsfrei ist. An den beiden vorherigen Beispielen kann man sehen, daß dies für die Größe

$$V = 6 W_1' \Sigma_{1,d}^{-1} W_1 + \dots + 6 W_d' \Sigma_{1,d}^{-1} W_d + \frac{6}{\alpha_1} \sum_{r \neq s \neq t} \hat{V}_{rst}$$

mit  $\hat{V}_{rst} = (1/\sqrt{n}) \sum_{i=1}^n L_{rst}(Z_i)$  der Fall ist;  $V$  ist in der Klasse der elliptischen Verteilungen asymptotisch  $\chi^2$ -verteilt mit  $\binom{d+2}{3}$  Freiheitsgraden. Mit den Bezeichnungen aus Beispiel 1.3.1 gilt

$$\Sigma_{1,d}^{-1} = \frac{1}{c} \left( \begin{array}{c|cccc} (d-2)u+v & -\sqrt{3}u & -\sqrt{3}u & \cdots & -\sqrt{3}u \\ \hline -\sqrt{3}u & du+v & -u & \cdots & -u \\ -\sqrt{3}u & -u & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -u \\ -\sqrt{3}u & -u & \cdots & -u & du+v \end{array} \right),$$

wobei  $c = (v-u)((d+1)u+v) = \alpha_1 \alpha_2$  ist. Damit folgt weiter

$$\begin{aligned} V &= \frac{6}{\alpha_1 \alpha_2} \sum_{r=1}^d W_r' \left( \alpha_2 I_d - \frac{u}{2} A \right) W_r + \frac{6}{\alpha_1} \sum_{r \neq s \neq t} \hat{V}_{rst} \\ &= \frac{6}{\alpha_1} \hat{U}_{n,3}^2 - \frac{3u}{\alpha_1 \alpha_2} n \tilde{b}_{1,d}. \end{aligned}$$

Dabei gilt  $u = (\alpha_2 - \alpha_1)/(d+2)$ . Da die in  $\alpha_1$  und  $\alpha_2$  vorkommenden Momente nicht bekannt sind, müssen sie durch die entsprechenden empirischen Momente ersetzt werden. Man erhält dann folgendes Resultat.

### 1.3.6 Satz

$X$  sei elliptisch-symmetrisch verteilt mit Erwartungswert  $\mu$  und nicht-singulärer Kovarianzmatrix  $T$ , und es gelte  $E[\{(X - \mu)' T^{-1} (X - \mu)\}^3] < \infty$ . Weiter sei  $\hat{r}_{2k} = \frac{1}{n} \sum_{i=1}^n (Z_i' Z_i)^k$  und

$$\hat{\alpha}_1 = \frac{6\hat{r}_6}{d(d+2)(d+4)}, \quad \hat{\alpha}_2 = \frac{3}{d} \left( \frac{\hat{r}_6}{d+2} - 2\hat{r}_4 + d(d+2) \right).$$

Dann ist die Testgröße

$$n\tilde{b}_{1,d} := \frac{n}{\hat{\alpha}_1} b_{1,d} - \frac{3(\hat{\alpha}_2 - \hat{\alpha}_1)}{(d+2)\hat{\alpha}_1\hat{\alpha}_2} n\tilde{b}_{1,d}$$

asymptotisch  $\chi^2$ -verteilt mit  $\binom{d+2}{3}$  Freiheitsgraden. ■

**Bemerkung:** 1. Ein auf  $n\tilde{b}_{1,d}$  basierender Test auf elliptische Symmetrie ist konsistent gegen alle Verteilungen mit  $\beta_{1,d} > 0$ .

2. Die Konvergenzgeschwindigkeit von  $n\tilde{b}_{1,d}$  gegen die Grenzverteilung hängt wiederum von den Momenten der zugrundeliegenden elliptischen Verteilung ab. Mit Hilfe von Simulationsstudien wäre zu prüfen, bei welchen Stichprobenumfängen die in Baringhaus und Henze (1992a) bzw. Henze (1997b) bei Verwendung von  $nb_{1,d}$  bzw.  $n\tilde{b}_{1,d}$  auftretenden „Anomalien“ nicht mehr vorkommen.

### 1.3.7 Beispiel (siehe auch Beispiel 1.2.2)

Als letztes Beispiel betrachten wir univariate Verteilungen mit  $\beta_{1,1} > 0$ . Es ist

$$E \left[ \left. \frac{\partial H_3(\tilde{X}, \mu, \sigma)}{\partial \mu} \right|_{\mu=0, \sigma=1} \right] = 0$$

und

$$E \left[ \left. \frac{\partial H_3(\tilde{X}, \mu, \sigma)}{\partial \sigma} \right|_{\mu=0, \sigma=1} \right] = \frac{-3}{\sqrt{6}} \mu_3.$$

Weiter ist

$$\sqrt{n}(\bar{X}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$$

und

$$\sqrt{n}(\hat{\sigma}_n - \sigma) = \sqrt{n} \frac{\hat{\sigma}_n^2 - \sigma^2}{\hat{\sigma}_n + \sigma} = \frac{1}{2\sigma\sqrt{n}} \sum_{i=1}^n ((X_i - \mu)^2 - \sigma^2) + o_P(1),$$

d.h. (R1) ist erfüllt mit  $l_1(x, \mu, \sigma) = x - \mu$  und  $l_2(x, \mu, \sigma) = ((x - \mu)^2 - \sigma^2)/(2\sigma)$ . Somit erhält man

$$\begin{aligned} 6 \sigma_{11} &= E \left[ \left( \tilde{X}^3 - 3\tilde{X} - 3\mu_3 \frac{\tilde{X}^2 - 1}{2} \right)^2 \right] - \mu_3^2 \\ &= \mu_6 + 9 + \frac{9}{4} \mu_3^2 (\mu_4 - 1) - 6\mu_4 - 3\mu_3\mu_5 + 11\mu_3^2. \end{aligned}$$

Nach (1.13) ist  $\sqrt{n}(b_{1,1} - \beta_{1,1})$  folglich asymptotisch  $\mathcal{N}(0, 4\mu_3^2 6\sigma_{11})$ -verteilt (vgl. Baringhaus und Henze (1992a), Example 3.3). Nach einer Normierung folgt das bekannte Resultat von Gastwirth und Owens (1977). ■

### 1.3.2 Multivariate Wölbung

Als affin-invariantes Maß für die multivariate Wölbung wurde von Mardia (1970) der Ausdruck

$$\begin{aligned} b_{2,d} &= \frac{1}{n} \sum_{i=1}^n (Z_i' Z_i)^2 \\ &= \frac{1}{n} \left\{ \sum_{r=1}^d \left( \sum_i z_{ri}^4 \right)^2 + \sum_{r \neq s} \left( \sum_i (z_{ri}^2 z_{si}^2) \right)^2 \right\} \end{aligned}$$

vorgeschlagen. Die zugehörige Verteilungsgröße ist

$$\beta_{2,d} = E[(\tilde{X}' \tilde{X})^2].$$

Ein in Hinblick auf die Definition der multivariaten Schiefe näherliegendes Maß für die Wölbung ist

$$\begin{aligned} b_{2,d}^* &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (Z_i' Z_j)^4 \\ &= \frac{1}{n^2} \left\{ \sum_{r=1}^d \left( \sum_i z_{ri}^4 \right)^2 + 4 \sum_{r \neq s} \left( \sum_i (z_{ri}^3 z_{si}^2) \right)^2 \right\} \end{aligned}$$

$$\begin{aligned}
& +3 \sum_{r \neq s} \left( \sum_i (z_{ri}^2 z_{si}^2) \right)^2 + 6 \sum_{r \neq s \neq t} \left( \sum_i (z_{ri}^2 z_{si} z_{ti}) \right)^2 \\
& + 24 \sum_{1 \leq r < s < t < u \leq d} \left( \sum_i (z_{ri} z_{si} z_{ti} z_{ui}) \right)^2 \Big\}
\end{aligned}$$

mit der zugehörigen Verteilungsgröße

$$\beta_{2,d}^* = E[(\tilde{X}'\tilde{Y})^4].$$

Auch diese Größen stehen in enger Verbindung zu einer Komponente des glatten Anpassungstests. Betrachtet man die Polynome vierten Grades

$$H_4(y_j), H_3(y_j)H_1(y_k), H_2(y_j)H_2(y_k),$$

$$H_2(y_j)H_1(y_k)H_1(y_l), H_1(y_j)H_1(y_k)H_1(y_l)H_1(y_m)$$

und die daraus gebildete Komponente

$$\begin{aligned}
\hat{U}_{n,4}^2 &= \frac{1}{n} \left\{ \sum_r \frac{1}{24} \left( \sum_i (z_{ri}^4 - 3) \right)^2 + \sum_{r \neq s} \frac{1}{6} \left( \sum_i (z_{ri}^3 z_{si}) \right)^2 \right. \\
&+ \sum_{r \neq s} \frac{1}{8} \left( \sum_i (z_{ri}^2 z_{si}^2 - 1) \right)^2 + \frac{1}{4} \sum_{r \neq s \neq t} \left( \sum_i (z_{ri}^2 z_{si} z_{ti}) \right)^2 \\
&\left. + \sum_{r < s < t < u} \left( \sum_i (z_{ri} z_{si} z_{ti} z_{ui}) \right)^2 \right\},
\end{aligned}$$

so ergibt sich zum einen die algebraische Gleichheit (Koziol (1987))

$$\hat{U}_{n,4}^2 = \frac{n}{24} (b_{2,d}^* - 6 b_{2,d} + 3 d(d+2)).$$

Zum anderen erhält man

$$\sqrt{n} (b_{2,d} - d(d+2)) = \sum_{r=1}^d \frac{1}{\sqrt{n}} \sum_i (z_{ri}^4 - 3) + \sum_{r \neq s} \frac{1}{\sqrt{n}} \sum_i (z_{ri}^2 z_{si}^2 - 1)$$



$$= \frac{\sqrt{24}}{n} \sum_{r=1}^d \sum_i L_r(Z_i) + \frac{4}{\sqrt{n}} \sum_{r < s} \sum_i L_r^s(Z_i), \quad (1.19)$$

wobei die Polynome  $L_r(z_1, \dots, z_d) = (z_r^4 - 6z_r^2 + 3)/\sqrt{24}$  und  $L_r^s(z_1, \dots, z_d) = (z_r^2 - 1)(z_s^2 - 1)/2$  zu den oben aufgeführten „Bausteinen“ der vierten Komponente gehören (wegen  $\sum_i z_i = 0$ ,  $\sum_i z_i^2 = 1$  ergeben sich in (1.19) einige Vereinfachungen).

Die einzelnen Summanden in (1.19) sind nach Zentrierung asymptotisch normalverteilt (vgl. (1.8)), wobei sich die Kovarianzmatrix  $\hat{\Sigma}$  wegen der unterschiedlichen Gewichte in (1.19) von  $\Sigma$  unterscheidet. Somit ist die asymptotische Verteilung von  $\sqrt{n}(b_{2,d} - \beta_{2,d})$  eine Normalverteilung  $\mathcal{N}(0, e' \hat{\Sigma} e)$  mit  $e = (1, \dots, 1)'$ . Für Verteilungen mit  $\tau = 0$  gilt dabei

$$\beta_{2,d} = \sum_{j=1}^d E[\xi_j^4] + \sum_{i \neq j} E[\xi_i^2 \xi_j^2] = 3d + d(d-1) = d(d+2).$$

Für die asymptotische Verteilung von  $\hat{U}_{n,4}^2$  sind nach den Ergebnissen des vorherigen Abschnitts zwei Fälle zu unterscheiden. Für Verteilungen mit  $\tau = 0$  ist die asymptotische Verteilung von  $\hat{U}_{n,4}^2$  eine gewichtete Summe von  $\chi_1^2$ -verteilten Zufallsvariablen; unter der Hypothese der multivariaten Normalverteilung ist  $\hat{U}_{n,4}^2$  insbesondere  $\chi^2$ -verteilt mit  $\binom{d+3}{4}$  Freiheitsgraden. Im Falle  $\tau \neq 0$  ist  $(\hat{U}_{n,4}^2/\sqrt{n} - \sqrt{n}\tau)$  asymptotisch multivariat normalverteilt.

Die Grenzverteilung von  $b_{2,d}^*$  wird in Abschnitt 1.4 hergeleitet.

### 1.3.8 Beispiel

In diesem Beispiel soll die asymptotische Verteilung von  $\sqrt{n}(b_{2,d} - d(d+2))$  für den wichtigen Fall von elliptisch-symmetrischen Verteilungen, die in der Klasse  $\mathcal{P}_0^4$  (vgl. (1.10)) liegen, bestimmt werden; nach dem oben Gesagten berechnet sich die Kovarianzmatrix  $\hat{\Sigma}$  im wesentlichen nach Gleichung (1.11), wobei noch die unterschiedlichen Gewichte in (1.19) zu berücksichtigen sind. Aus den in Anhang A.2, Korollar A.2.4, angegebenen Relationen ergibt sich, daß eine elliptisch-symmetrische Verteilung genau dann zu  $\mathcal{P}_0^4$

gehört, wenn wie bei der Normalverteilung  $\mu_4 = 3$  gilt.  $\hat{\Sigma}$  besitzt die folgenden Einträge:

$$\begin{aligned}
\hat{\sigma}^{11} &= 24E[L_r^2(\tilde{X})] = E[(\xi_r^4 - 6\xi_r^2 + 3)^2] = \mu_8 - 12\mu_6 + 99, \\
\hat{\sigma}^{12} &= 24E[L_r(\tilde{X})L_s(\tilde{X})] = E[(\xi_r^4 - 6\xi_r^2 + 3)(\xi_s^4 - 6\xi_s^2 + 3)] \\
&= \mu_{44} - 12\mu_{42} + 36\mu_{22} + 6\mu_4 - 27 = 3\mu_8/35 - 12\mu_6/5 + 27, \\
\hat{\sigma}^{21} &= 16E[(L_r^s(\tilde{X}))^2] = 4(3\mu_8/35 - 4\mu_6/5 + 7), \\
\hat{\sigma}^{22} &= 16E[L_r^s(\tilde{X})L_r^t(\tilde{X})] = 4(\mu_8/35 - 8\mu_6/15 + 5), \\
\hat{\sigma}^{23} &= 16E[L_r^s(\tilde{X})L_t^u(\tilde{X})] = 4(\mu_8/105 - 4\mu_6/15 + 3), \\
\hat{\sigma}^{31} &= 4\sqrt{24}E[L_r(\tilde{X})L_r^s(\tilde{X})] = 2(\mu_8/7 - 12\mu_6/5 + 21), \\
\hat{\sigma}^{32} &= 4\sqrt{24}E[L_r(\tilde{X})L_s^t(\tilde{X})] = 2(\mu_8/35 - 4\mu_6/5 + 9),
\end{aligned}$$

wobei  $r, s, t, u \in \{1, \dots, d\}$ ,  $r < s < t < u$  ist. Für  $d = 3$  und die Anordnung  $L_1, L_2, L_3, L_1^2, L_1^3, L_2^3$  hat die Kovarianzmatrix z.B. folgendes Aussehen:

$$\begin{pmatrix}
\hat{\sigma}^{11} & \hat{\sigma}^{12} & \hat{\sigma}^{12} & \hat{\sigma}^{31} & \hat{\sigma}^{31} & \hat{\sigma}^{32} \\
\hat{\sigma}^{12} & \hat{\sigma}^{11} & \hat{\sigma}^{12} & \hat{\sigma}^{31} & \hat{\sigma}^{32} & \hat{\sigma}^{31} \\
\hat{\sigma}^{12} & \hat{\sigma}^{12} & \hat{\sigma}^{11} & \hat{\sigma}^{32} & \hat{\sigma}^{31} & \hat{\sigma}^{31} \\
\hat{\sigma}^{31} & \hat{\sigma}^{31} & \hat{\sigma}^{32} & \hat{\sigma}^{21} & \hat{\sigma}^{22} & \hat{\sigma}^{22} \\
\hat{\sigma}^{31} & \hat{\sigma}^{32} & \hat{\sigma}^{31} & \hat{\sigma}^{22} & \hat{\sigma}^{21} & \hat{\sigma}^{22} \\
\hat{\sigma}^{32} & \hat{\sigma}^{31} & \hat{\sigma}^{31} & \hat{\sigma}^{22} & \hat{\sigma}^{22} & \hat{\sigma}^{21}
\end{pmatrix}.$$

Für den allgemeinen Fall überlegt man sich, daß  $\hat{\sigma}^{11}$  in der Matrix  $d$ -mal vorkommt,  $\hat{\sigma}^{12}$   $d(d-1)$ -mal,  $\hat{\sigma}^{21}$   $\binom{d}{2}$ -mal,  $\hat{\sigma}^{22}$   $2(d-2)\binom{d}{2}$ -mal,  $\hat{\sigma}^{23}$   $\binom{d}{2}\binom{d-2}{2}$ -mal,  $\hat{\sigma}^{31}$   $4\binom{d}{2}$ -mal und schließlich  $\hat{\sigma}^{32}$   $2(d-2)\binom{d}{2}$ -mal. Summiert man alle Terme auf, so ergibt sich als Varianz der Normalverteilung, gegen die die Wölbungs-Statistik konvergiert,

$$\begin{aligned}
e'\hat{\Sigma}e &= \mu_8 \frac{d}{105}(d^3 + 12d^2 + 44d + 48) - \mu_6 \frac{4d}{15}(d^3 + 8d^2 + 20d + 16) \\
&\quad + d(3d^3 + 20d^2 + 44d + 32).
\end{aligned}$$

Ersetzt man  $\mu_{2k}$  für  $k = 2, 3$  und  $4$  gemäß Satz 1.4.7 durch  $r_{2k} = E[(X'X)^k]$ , so läßt sich die Varianz schreiben als

$$e'\hat{\Sigma}e = r_8 - 4(d+2)r_6 + d(d+2)^2(3d+8).$$

Dies ist das Resultat von Henze (1994a), Beispiel 3.3, wenn man dort  $\mu_4 = 3$  setzt. Ist die zugrundeliegende Verteilung eine multivariate Normalverteilung, so ergibt sich als asymptotische Varianz  $8d(d+2)$ , also das bekannte Resultat von Mardia (1970). ■

### 1.3.9 Beispiel

Als letztes Beispiel soll die asymptotische Verteilung von  $24\hat{U}_{n,4}^2$  berechnet werden, wenn die zugrundeliegende Verteilung wie im vorherigen Beispiel eine elliptisch-symmetrische Verteilung ist, die in  $\mathcal{P}_0^4$  liegt (für die folglich  $m_4 = 3$  gelten muß). Es müssen also die Eigenwerte der nach (1.11) gebildeten Kovarianzmatrix, multipliziert mit 24, bestimmt werden. Neben den Polynomen  $L_r$  und  $L_r^s$  wird  $\hat{U}_{n,4}^2$  aus  $L_{rs}(z_1, \dots, z_d) = (z_r^3 - 3z_r)z_s/6$  ( $r \neq s$ ),  $L_{rst}(z_1, \dots, z_d) = (z_r^2 - 1)z_s z_t/2$  ( $r \neq s, t, s < t$ ) und aus  $L_{rstu}(z_1, \dots, z_d) = z_r z_s z_t z_u$  ( $r < s < t < u$ ) gebildet. Betrachtet man diese Polynome genauer, so sieht man zunächst, daß alle Kovarianzterme zwischen Polynomen, die auch in  $b_{2,d}$  enthalten sind, und den restlichen Polynomen verschwinden, da sie nur Momente enthalten, die nach Satz A.2.3 Null sind; so gilt z.B.

$$E[(\tilde{X}^4 - 6\tilde{X}^2 + 3)(\tilde{X}^2 - 1)\tilde{Y}\tilde{Z}] = \mu_{611} - 7\mu_{411} + 9\mu_{211} - 3\mu_{11} = 0.$$

Aus dem gleichen Grund verschwinden alle Kovarianzterme zwischen  $L_{rstu}$  und allen anderen Polynomen. Die Kovarianzmatrix ist also eine Blockdiagonalmatrix und zerfällt in drei Teile.

Die zu den  $L_{rstu}$  gehörende Teilmatrix ist eine  $\binom{d}{4} \times \binom{d}{4}$ -Diagonalmatrix mit Einträgen  $24\mu_{2222} = 24\mu_8/105$ .  $\alpha_1 = 8\mu_8/35$  ist also  $\binom{d}{4}$ -facher Eigenwert.

Die zweite Teilmatrix besitzt die nichtverschwindenden Einträge

$$\begin{aligned} \sigma^{41} &= 24E[L_{rs}^2(\tilde{X})] = 4(\mu_8/7 - 6\mu_6/5 + 9), \\ \sigma^{42} &= 24E[L_{rs}(\tilde{X})L_{sr}(\tilde{X})] = 4(3\mu_8/35 - 6\mu_6/5 + 9), \\ \sigma^{43} &= 24E[L_{rst}^2(\tilde{X})] = 12(\mu_8/35 - 2\mu_6/15 + 1), \\ \sigma^{44} &= 24E[L_{rst}(\tilde{X})L_{st}(\tilde{X})] = 4\sqrt{3}(\mu_8/35 - 2\mu_6/5 + 3), \end{aligned}$$

wobei  $r, s, t \in \{1, \dots, d\}, r \neq s \neq t \neq r$  ist. Ordnet man die Polynome in der Reihenfolge

$$\begin{aligned}
 &L_{12}, L_{21}, L_{312}, L_{412}, \dots, L_{d12}, \\
 &L_{13}, L_{31}, L_{213}, L_{413}, \dots, L_{d13}, \\
 &\dots\dots\dots \\
 &L_{1d}, L_{d1}, L_{21d}, L_{31d}, \dots, L_{d-1,1,d}, \\
 &\dots\dots\dots \\
 &L_{d-1,d}, L_{d,d-1}, L_{1,d-1,d}, L_{2,d-1,d}, \dots, L_{d-2,d-1,d},
 \end{aligned}$$

so ergibt sich weiter, daß die betrachtete Teilmatrix wiederum zerfällt, und zwar in  $\binom{d}{2}$  Matrizen der Form

$$\left( \begin{array}{cc|ccc}
 \sigma^{41} & \sigma^{42} & \sigma^{44} & \dots & \sigma^{44} \\
 \sigma^{42} & \sigma^{41} & \sigma^{44} & \dots & \sigma^{44} \\
 \hline
 \sigma^{44} & \sigma^{44} & \sigma^{43} & & 0 \\
 \vdots & \vdots & & \ddots & \\
 \sigma^{44} & \sigma^{44} & 0 & & \sigma^{43}
 \end{array} \right). \quad (1.20)$$

Hierbei hat die rechte untere Diagonalmatrix die Dimension  $d - 2$ . Für die Matrix in (1.20) ist  $\alpha_2 = \sigma^{43}$  ein  $(d - 3)$ -facher Eigenwert. Die weiteren Eigenwerte sind die von  $\sigma^{43}$  verschiedenen Lösungen der Gleichung

$$((\sigma^{41} - \lambda)(\sigma^{43} - \lambda) - (\sigma^{44})^2(d - 2))^2 = (\sigma^{42}(\sigma^{43} - \lambda) - (\sigma^{44})^2(d - 2))^2.$$

Man erhält als Lösung einen weiteren Eigenwert  $\alpha_1 = 8\mu_8/35$  und die beiden von der Dimension abhängigen Eigenwerte

$$\alpha_{3,4} = \frac{22}{35}\mu_8 - \frac{28}{5}\mu_6 + 42 \pm \frac{2}{35}\sqrt{24d - 23}(\mu_8 - 14\mu_6 + 105).$$

Die Eigenwerte der dritten Teilmatrix sind schwieriger zu berechnen, da diese vollbesetzt ist; bis auf konstante Faktoren, die von der anderen Gewichtung herrühren, entspricht sie der Kovarianzmatrix von  $b_{2,d}$ . Genauer gilt: definiert man  $\sigma^{11} = \hat{\sigma}^{11}, \sigma^{12} = \hat{\sigma}^{12}, \sigma^{21} = 3\hat{\sigma}^{11}/2, \sigma^{22} = 3\hat{\sigma}^{22}/2, \sigma^{23} =$

$3\hat{\sigma}^{23}/2, \sigma^{31} = (3/2)^{1/2}\hat{\sigma}^{31}$  und  $\sigma^{32} = (3/2)^{1/2}\hat{\sigma}^{32}$ , und ersetzt man in den Kovarianzmatrizen des vorherigen Beispiels alle  $\hat{\sigma}^{ij}$  durch die entsprechenden  $\sigma^{ij}$ , so erhält man die dritte Teilmatrix, die mit  $\Sigma_3$  bezeichnet werden soll.

**1.3.10 Lemma**  $\Sigma_3$  besitzt die folgenden Eigenwerte:  $\alpha_1 = 8\mu_8/35$  ist  $\binom{d}{2}$ -facher Eigenwert (wobei  $\binom{1}{2} = 0$  gesetzt wird);

$$\alpha_5 = 4 \left( \frac{(d+6)}{35}\mu_8 - \frac{2d+8}{5}\mu_6 + 3(d+4) \right)$$

ist  $(d-1)$ -facher Eigenwert; schließlich ist

$$\alpha_6 = \frac{d^2 + 10d + 24}{35}\mu_8 - \frac{4(d^2 + 6d + 8)}{5}\mu_6 + 3(3d^2 + 14d + 16)$$

ein einfacher Eigenwert.

BEWEIS: Für  $d = 1$  ist  $\sigma^{11} = \mu_8 - 12\mu_6 + 99 = \alpha_6$ . Sei nun  $d > 1$ . Als neue Parameter verwenden wir  $u = \mu_8/35 - 4\mu_6/5 + 9$  und  $v = 4(\mu_8/35 - 2\mu_6/5 + 3)$ . Damit ergibt sich

$$\begin{aligned} \sigma^{11} &= -5u + 10v + 24, & \sigma^{12} &= 3u, & \sigma^{21} &= -6u + 6v + 24, \\ \sigma^{22} &= 2u + v, & \sigma^{23} &= 2u, & \sigma^{31} &= \sqrt{6}(u + v), & \sigma^{32} &= \sqrt{6}u. \end{aligned}$$

Die Polynome seien in der Reihenfolge

$$L_1, \dots, L_d, L_1^2, \dots, L_1^d, L_2^3, \dots, L_2^d, \dots, L_{d-1}^d$$

angeordnet. Zur besseren Vorstellung sei noch die Matrix  $\Sigma_3$  für den Fall

$d = 4$  angegeben:

$$\left( \begin{array}{cccc|cccccc} \sigma^{11} & \sigma^{12} & \sigma^{12} & \sigma^{12} & \sigma^{31} & \sigma^{31} & \sigma^{31} & \sigma^{32} & \sigma^{32} & \sigma^{32} \\ \sigma^{12} & \sigma^{11} & \sigma^{12} & \sigma^{12} & \sigma^{31} & \sigma^{32} & \sigma^{32} & \sigma^{31} & \sigma^{31} & \sigma^{32} \\ \sigma^{12} & \sigma^{12} & \sigma^{11} & \sigma^{12} & \sigma^{32} & \sigma^{31} & \sigma^{32} & \sigma^{31} & \sigma^{32} & \sigma^{31} \\ \sigma^{12} & \sigma^{12} & \sigma^{12} & \sigma^{11} & \sigma^{32} & \sigma^{32} & \sigma^{31} & \sigma^{32} & \sigma^{31} & \sigma^{31} \\ \hline \sigma^{31} & \sigma^{31} & \sigma^{32} & \sigma^{32} & \sigma^{21} & \sigma^{22} & \sigma^{22} & \sigma^{22} & \sigma^{22} & \sigma^{23} \\ \sigma^{31} & \sigma^{32} & \sigma^{31} & \sigma^{32} & \sigma^{22} & \sigma^{21} & \sigma^{22} & \sigma^{22} & \sigma^{23} & \sigma^{22} \\ \sigma^{31} & \sigma^{32} & \sigma^{32} & \sigma^{31} & \sigma^{22} & \sigma^{22} & \sigma^{21} & \sigma^{23} & \sigma^{22} & \sigma^{22} \\ \sigma^{32} & \sigma^{31} & \sigma^{31} & \sigma^{32} & \sigma^{22} & \sigma^{22} & \sigma^{23} & \sigma^{21} & \sigma^{22} & \sigma^{22} \\ \sigma^{32} & \sigma^{31} & \sigma^{32} & \sigma^{31} & \sigma^{22} & \sigma^{23} & \sigma^{22} & \sigma^{22} & \sigma^{21} & \sigma^{22} \\ \sigma^{32} & \sigma^{32} & \sigma^{31} & \sigma^{31} & \sigma^{23} & \sigma^{22} & \sigma^{22} & \sigma^{22} & \sigma^{22} & \sigma^{21} \end{array} \right)$$

Bei den folgenden Rechnungen ist zu beachten, daß für  $d = 2$  und  $d = 3$  manche Zeilen gar nicht auftreten; ist etwa von der 3. bis  $d$ -ten Zeile die Rede, so bezieht sich dies nur auf die Fälle  $d \geq 3$ .

1. Setze  $\alpha_1 = 4v - 8u + 24$ . Definiere Vektoren  $e^{ij}$  ( $1 \leq i < j \leq d$ ) mit Eintrag 1 an der  $i$ -ten und  $j$ -ten Stelle, Eintrag  $-\sqrt{6}$  an Stelle  $(ij)$  (d.h. an der Stelle, an der  $L_i^j$  steht) und Null sonst. Dann ist  $e^{ij}$  Eigenvektor von  $\Sigma_3$  zum Eigenwert  $\alpha_1$ .

Denn sei o.B.d.A.  $i = 1, j = 2$ . Bildet man das Produkt aus  $\Sigma_3$  und  $e^{12}$ , so erhält man in der 1. und 2. Zeile  $\sigma^{11} + \sigma^{12} - \sqrt{6}\sigma^{31} = 4v - 8u + 24$ , in der 3. bis  $d$ -ten Zeile  $2\sigma^{12} - \sqrt{6}\sigma^{32} = 0$ ; in Zeile (12) ergibt sich  $2\sigma^{31} - \sqrt{6}\sigma^{21} = -\sqrt{6}(4v - 8u + 24)$ , in Zeilen (13) bis  $(2d)$   $\sigma^{31} + \sigma^{32} - \sqrt{6}\sigma^{22} = 0$  und in den restlichen Zeilen  $2\sigma^{32} - \sqrt{6}\sigma^{23} = 0$ . Da die  $\binom{d}{2}$  Vektoren  $e^{ij}$  ( $i < j$ ) linear unabhängig sind, ist  $\alpha_1$  mindestens  $\binom{d}{2}$ -facher Eigenwert von  $\Sigma_3$ .

2. Setze  $\alpha_5 = (d + 8)v - 8u + 24$ . Definiere Vektoren  $e^i$  ( $1 < i \leq d$ ) mit Eintrag  $\sqrt{6}$  an 1. und  $-\sqrt{6}$  an  $i$ . Stelle, Einträge 1 an den Stellen  $(1j)$ ,  $j \neq 1, i$ , Einträgen  $-1$  an den Stellen  $(ij)$ ,  $j \neq 1, i$ , und Null sonst. Dann ist  $e^i$  Eigenvektor von  $\Sigma_3$  zum Eigenwert  $\alpha_5$ .

Denn sei o.B.d.A.  $i = 2$ . Für  $d = 4$  ist dann etwa

$$e^2 = (\sqrt{6}, -\sqrt{6}, 0, 0, 0, 1, 1, -1, -1, 0).$$

Bildet man das Produkt aus  $\Sigma_3$  und  $e^2$ , so erhält man in der 1. Zeile

$$\sqrt{6}(\sigma^{11} - \sigma^{12}) + (d-2)(\sigma^{31} - \sigma^{32}) = \sqrt{6}((d+8)v - 8u + 24),$$

in der 2. Zeile das Negative der 1. Zeile, in der 3. bis  $d$ -ten Zeile Null, da sich die positiven und negativen Terme jeweils aufheben, ebenso in Zeile (12); in den Zeilen (13) bis (1 $d$ ) ergibt sich

$$\sqrt{6}(\sigma^{31} - \sigma^{32}) + \sigma^{21} + (d-3)\sigma^{22} - (\sigma^{22} + (d-3)\sigma^{23}) = (d+8)v - 8u + 24,$$

in den Zeilen (23) bis (2 $d$ ) das Negative davon; in den restlichen Zeilen Null, da sich die positiven und negativen Terme jeweils aufheben. Da die  $d-1$  Vektoren  $e^i$  ( $1 < i \leq d$ ) linear unabhängig sind, ist  $\alpha_5$  mindestens  $(d-1)$ -facher Eigenwert von  $\Sigma_3$ .

3. Setze  $\alpha_6 = (d^2 + 2d - 8)u + (2d + 8)v + 24$ . Definiere den Vektor  $e$  mit Eintrag  $\sqrt{6}/2$  an 1. bis  $d$ -ten Stelle und den Einträgen 1 sonst. Dann ist  $e$  Eigenvektor von  $\Sigma_3$  zum Eigenwert  $\alpha_6$ . In den ersten  $d$  Zeilen erhält man nämlich

$$\sqrt{6}(\sigma^{11} + (d-1)\sigma^{12})/2 + (d-1)\sigma^{31} + \binom{d-1}{2}\sigma^{32} = \sqrt{6}\alpha_6/2;$$

in den restlichen Zeilen ergibt sich

$$\sqrt{6}(2\sigma^{31} + (d-2)\sigma^{32})/2 + \sigma^{21} + 2(d-2)\sigma^{22} + \binom{d-2}{2}\sigma^{23} = \alpha_6.$$

$\alpha_6$  ist also mindestens einfacher Eigenwert von  $\Sigma_3$ .

Da die Dimension von  $\Sigma_3$   $d + \binom{d}{2}$  ist, folgt aus 1. bis 3., daß  $\alpha_1$  genau  $\binom{d}{2}$ -facher,  $\alpha_5$  genau  $(d-1)$ -facher und  $\alpha_6$  genau einfacher Eigenwert von  $\Sigma_3$  ist. Drückt man  $\alpha_1$ ,  $\alpha_5$  und  $\alpha_6$  durch  $\mu_8$  und  $\mu_6$  aus, so folgt die Behauptung. ■

Faßt man alle Ergebnisse zusammen, so erhält man den folgenden Satz.

### 1.3.11 Satz

Der  $d$ -dimensionale Zufallsvektor  $X$  sei elliptisch-symmetrisch verteilt mit Erwartungswert  $\mu$  und nichtsingulärer Kovarianzmatrix  $T$ . Für  $\tilde{X} =$

$(\xi_1, \dots, \xi_d) = T^{-1/2}(X - \mu)$  gelte  $m_4 = E[\xi_1^4] = 3$  und  $E[(\tilde{X}'\tilde{X})^4] < \infty$ , d.h.  $\tilde{X}$  sei aus  $\mathcal{P}_0^4$ . Dann gilt

$$24 \hat{U}_{n,4}^2 \xrightarrow{\mathcal{D}} \sum_{i=1}^6 \alpha_i \chi_{\nu_i}^2 ;$$

dabei ist

$$\alpha_1 = 8\mu_8/35, \quad \nu_1 = \binom{d}{4} + 2\binom{d}{2},$$

$$\alpha_2 = 12(\mu_8/35 - 2\mu_6/15 + 1), \quad \nu_2 = (d-3)\binom{d}{2},$$

$$\alpha_{3,4} = \frac{22}{35}\mu_8 - \frac{28}{5}\mu_6 + 42 \pm \frac{2}{35}\sqrt{24d-23}(\mu_8 - 14\mu_6 + 105), \quad \nu_{3,4} = \binom{d}{2},$$

$$\alpha_5 = 4 \left( \frac{(d+6)}{35}\mu_8 - \frac{2d+8}{5}\mu_6 + 3(d+4) \right), \quad \nu_5 = d-1,$$

$$\alpha_6 = \frac{d^2 + 10d + 24}{35}\mu_8 - \frac{4(d^2 + 6d + 8)}{5}\mu_6 + 3(d^2 + 14d + 16), \quad \nu_6 = 1$$

und  $\chi_{\nu_i}^2$  sind unabhängige  $\chi^2$ -verteilte Zufallsvariablen mit  $\nu_i$  Freiheitsgraden. ■

**Bemerkung:** Unter der Hypothese der multivariaten Normalverteilung gilt insbesondere  $\alpha_i = 24$ ,  $i = 1, \dots, 6$ . Wegen  $\sum_{i=1}^6 \nu_i = \binom{d+3}{4}$  ist  $\hat{U}_{n,4}^2$  in diesem Fall  $\chi^2$ -verteilt mit  $\binom{d+3}{4}$  Freiheitsgraden.

**1.3.12 Korollar** *Unter den Bedingungen von Satz 1.3.11 gilt*

$$24 \hat{\Psi}_{n,4}^2 \xrightarrow{\mathcal{D}} \sum_{i=1}^8 \alpha_i \chi_{\nu_i}^2 ;$$

dabei sind  $\alpha_1, \dots, \alpha_6$  und  $\nu_1, \dots, \nu_6$  wie in Satz 1.3.11 und

$$\alpha_7 = 4(d+4)\mu_6/5 - 12(d+2), \quad \nu_7 = d,$$

$$\alpha_8 = 8\mu_6/5, \quad \nu_8 = d(d-1)(d+4)/6.$$



BEWEIS: Nach Definition gilt

$$24 \hat{\Psi}_{n,4}^2 = 24 (\hat{U}_{n,3}^2 + \hat{U}_{n,4}^2) = 4n b_{1,d} + 24 \hat{U}_{n,4}^2.$$

Da alle Kovarianzterme zwischen Polynomen, die zu  $\hat{U}_{n,3}$  gehören, und Polynomen, die zu  $\hat{U}_{n,4}$  gehören, verschwinden, folgt das Ergebnis aus den Sätzen 1.3.2 und 1.3.11. ■

### 1.3.3 Komponenten höherer Ordnung

Zur nächsthöheren Komponente tragen folgende multivariate Polynome  $L_r$  bei:

$$H_5(y_j), H_4(y_j)H_1(y_k), H_3(y_j)H_2(y_k), H_3(y_j)H_1(y_k)H_1(y_l),$$

$$H_2(y_j)H_2(y_k)H_1(y_l), H_2(y_j)H_1(y_k)H_1(y_l)H_1(y_m),$$

$$H_1(y_j)H_1(y_k)H_1(y_l)H_1(y_m)H_1(y_n);$$

dabei sind die Indizes jeweils paarweise verschieden und durchlaufen die Menge  $\{1, \dots, d\}$ . Dies wird sehr schnell unübersichtlich; in Rayner und Best (1989), S.103, wurde etwa die fünfte Polynomgruppe vergessen. Somit hat  $\hat{U}_{n,5}^2$  die Gestalt

$$\begin{aligned} \hat{U}_{n,5}^2 &= \frac{1}{n} \left\{ \sum_r \frac{1}{120} \left( \sum_i (z_{ri}^5 - 10z_{ri}^3) \right)^2 \right. \\ &\quad + \sum_{r \neq s} \frac{1}{24} \left( \sum_i (z_{ri}^4 - 6z_{ri}^2) z_{si} \right)^2 \\ &\quad + \sum_{r \neq s} \frac{1}{12} \left( \sum_i (z_{ri}^3 z_{si}^2 - z_{ri}^3 - 3z_{ri} z_{si}^2) \right)^2 \\ &\quad \left. + \frac{1}{2} \sum_{r \neq s \neq t} \frac{1}{6} \left( \sum_i (z_{ri}^3 z_{si} z_{ti} - 3z_{ri} z_{si} z_{ti}) \right)^2 \right\} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2} \sum_{r \neq s \neq t} \frac{1}{4} \left( \sum_i (z_{ri}^2 z_{si}^2 z_{ti} - z_{ri}^2 z_{ti} - z_{si}^2 z_{ti}) \right)^2 \\
& + \frac{1}{6} \sum_{r \neq s \neq t \neq u} \frac{1}{2} \left( \sum_i (z_{ri}^2 z_{si} z_{ti} z_{ui} - z_{si} z_{ti} z_{ui}) \right)^2 \\
& + \sum_{r < s < t < u < v} \left( \sum_i (z_{ri} z_{si} z_{ti} z_{ui} z_{vi}) \right)^2 \Big\}.
\end{aligned}$$

Führt man für die höheren Momente Bezeichnungen wie  $m_{abcd} = 1/n \sum_i z_{ri}^a z_{si}^b z_{ti}^c z_{ui}^d$  ein, so erhält man nach Ausmultiplizieren der Klammern

$$\begin{aligned}
\frac{120}{n} \hat{U}_{n,5}^2 &= \sum_r (m_5^2 - 20m_5 m_3 + 100m_3^2) \\
&+ \sum_{r \neq s} (5m_{41}^2 - 60m_{41} m_{21} + 180m_{21}^2) \\
&+ \sum_{r \neq s} (10m_{32}^2 + 10m_{30}^2 + 90m_{12}^2 - 20m_{32} m_{30} \\
&\quad - 60m_{32} m_{12} + 60m_{30} m_{12}) \\
&+ \sum_{r \neq s \neq t} (10m_{311}^2 - 60m_{311} m_{111} + 90m_{111}^2) \quad (1.21) \\
&+ \sum_{r \neq s \neq t} (15m_{221}^2 + 30m_{21}^2 - 60m_{221} m_{201} + 30m_{201} m_{021}) \\
&+ \sum_{r \neq s \neq t \neq u} (10m_{2111}^2 - 20m_{2111} m_{0111} + 10m_{1111}^2) \\
&+ \sum_{r \neq s \neq t \neq u \neq v} m_{111111}^2.
\end{aligned}$$

Man sieht, daß die Berechnung höherer Komponenten nach der Definition sehr aufwendig ist. Im folgenden sollen nun algebraischen Gleichheiten wie im Falle von Schiefe und Wölbung bestimmt werden, die erst die praktische Verwendung der Komponenten als Testgrößen ermöglichen. Wir verwenden

hierzu die Notation

$$b_{abc} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (Z'_i Z_j)^a (Z'_i Z_i)^b (Z'_j Z_j)^c.$$

In dieser Schreibweise gilt also  $6 \hat{U}_{n,3}^2/n = b_{1,d} = b_{300}$  und  $24 \hat{U}_{n,4}^2/n = b_{400} - 6 b_{020} + 3 d(d+2)$ .

### 1.3.13 Satz

Für die dritte nichtverschwindende Komponente gilt die Gleichheit

$$\frac{5!}{n} \hat{U}_{n,5}^2 = b_{500} - 20 b_{310} + 10 (d+6) b_{300} + 30 b_{111}.$$

Die dazugehörige affin-invariante Verteilungsgröße ist also

$$\begin{aligned} & E[(\tilde{X}'\tilde{Y})^5] - 20 E[(\tilde{X}'\tilde{Y})^3 |\tilde{X}|^2] \\ & + 10 (d+6) E[(\tilde{X}'\tilde{Y})^3] + 30 E[(\tilde{X}'\tilde{Y}) |\tilde{X}|^2 |\tilde{Y}|^2]. \end{aligned}$$

BEWEIS: Für die erste Größe gilt

$$\begin{aligned} b_{500} &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (Z'_i Z_j)^5 \\ &= \frac{1}{n^2} \left\{ \sum_r \left( \sum_i z_{ri}^5 \right)^2 + 5 \sum_{r \neq s} \left( \sum_i z_{ri}^4 z_{si} \right)^2 \right. \\ &\quad + 10 \sum_{r \neq s} \left( \sum_i z_{ri}^3 z_{si}^2 \right)^2 + 10 \sum_{r \neq s \neq t} \left( \sum_i z_{ri}^3 z_{si} z_{ti} \right)^2 \\ &\quad + 15 \sum_{r \neq s \neq t} \left( \sum_i z_{ri}^2 z_{si}^2 z_{ti} \right)^2 + 10 \sum_{r \neq s \neq t \neq u} \left( \sum_i z_{ri}^2 z_{si} z_{ti} z_{ui} \right)^2 \\ &\quad \left. + \sum_{r \neq s \neq t \neq u \neq v} \left( \sum_i z_{ri} z_{si} z_{ti} z_{ui} z_{vi} \right)^2 \right\} \end{aligned}$$

$$\begin{aligned}
&= \sum_r m_5^2 + \sum_{r \neq s} (5 m_{41}^2 + 10 m_{32}^2) + \sum_{r \neq s \neq t} (10 m_{311}^2 + 15 m_{221}^2) \\
&\quad + \sum_{r \neq s \neq t \neq u} 10 m_{2111}^2 + \sum_{r \neq s \neq t \neq u \neq v} m_{11111}^2.
\end{aligned}$$

Man sieht, daß die Terme den jeweils ersten Summanden in (1.21) entsprechen. Weiter gilt

$$\begin{aligned}
b_{310} &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (Z'_i Z_j)^3 (Z'_i Z_i) \\
&= \frac{1}{n^2} \sum_i \sum_j \left\{ \left( \sum_r z_{ri}^3 z_{rj}^3 + 3 \sum_{r \neq s} z_{ri}^2 z_{si} z_{rj}^2 z_{sj} \right. \right. \\
&\quad \left. \left. + \sum_{r \neq s \neq t} z_{ri} z_{si} z_{ti} z_{rj} z_{sj} z_{tj} \right) \left( \sum_u z_{ui}^2 \right) \right\} \\
&= \sum_r \left\{ \frac{1}{n} \left( \sum_i z_{ri}^5 \right) \left( \sum_j z_{rj}^3 \right) \right\} \\
&\quad + \sum_{r \neq u} \left\{ \frac{1}{n} \left( \sum_i z_{ri}^3 z_{ui}^2 \right) \left( \sum_j z_{rj}^3 \right) \right\} \\
&\quad + 3 \sum_{r \neq s} \left\{ \frac{1}{n} \left( \sum_i z_{ri}^4 z_{si} \right) \left( \sum_j z_{rj}^2 z_{sj} \right) \right\} \\
&\quad + 3 \sum_{r \neq s} \left\{ \frac{1}{n} \left( \sum_i z_{ri}^2 z_{si}^3 \right) \left( \sum_j z_{rj}^2 z_{sj} \right) \right\} \\
&\quad + 3 \sum_{r \neq s \neq u} \left\{ \frac{1}{n} \left( \sum_i z_{ri}^2 z_{si} z_{ui}^2 \right) \left( \sum_j z_{rj}^2 z_{sj} \right) \right\} \\
&\quad + 3 \sum_{r \neq s \neq t} \left\{ \frac{1}{n} \left( \sum_i z_{ri}^3 z_{si} z_{ti} \right) \left( \sum_j z_{rj} z_{sj} z_{tj} \right) \right\}
\end{aligned}$$

$$\begin{aligned}
& + \sum_{r \neq s \neq t \neq u} \left\{ \frac{1}{n} \left( \sum_i z_{ri} z_{si} z_{ti} z_{ui}^2 \right) \left( \sum_j z_{rj} z_{sj} z_{tj} \right) \right\} \\
= & \sum_r m_5 m_3 + \sum_{r \neq s} (m_{32} m_{30} + 3 m_{41} m_{21} + 3 m_{23} m_{21}) \\
& + \sum_{r \neq s \neq t} (3 m_{221} m_{201} + 3 m_{311} m_{111}) \\
& + \sum_{r \neq s \neq t \neq u} m_{2111} m_{0111}.
\end{aligned}$$

$b_{111}$  ist gerade die Schiefe nach Móri et al.; nach Beispiel 1.3.3 gilt

$$b_{111} = \sum_r m_3^2 + \sum_{r \neq s} (2 m_{30} m_{12} + m_{12}^2) + \sum_{r \neq s \neq t} m_{120} m_{102}.$$

Schließlich gilt für die Mardia-Schiefe nach Beispiel 1.3.1

$$b_{300} = \sum_r m_3^2 + 3 \sum_{r \neq s} m_{21}^2 + \sum_{r \neq s \neq t} m_{111}^2.$$

Addiert man alle Terme mit den im Satz angegebenen Vielfachheiten, so erhält man alle Terme in (1.21). ■

Die vierte nichtverschwindende Komponente ergibt sich ähnlich wie  $\hat{U}_{n,5}^2$ , indem man über alle Produkte von Hermite-Polynomen summiert, deren Gradsumme gleich sechs ist. Man erhält

$$\begin{aligned}
\frac{720}{n} \hat{U}_{n,6}^2 & = \sum_r (m_6 - 15m_4 + 30)^2 + 6 \sum_{r \neq s} (m_{51} - 10m_{31})^2 \\
& + 15 \sum_{r \neq s} (m_{42} + m_{40} - 6m_{22} + 6)^2 \\
& + 15 \sum_{r \neq s \neq t} (m_{411} - 6m_{211})^2 \\
& + 10 \sum_{r \neq s \neq t} (m_{33} - 3m_{31} - 3m_{13})^2 \\
& + 60 \sum_{r \neq s \neq t} (m_{321} - m_{301} - 3m_{121})^2 \tag{1.22}
\end{aligned}$$

$$\begin{aligned}
& +20 \sum_{r \neq s \neq t \neq u} (m_{3111} - 3m_{1111})^2 \\
& +15 \sum_{r \neq s \neq t \neq u} (m_{222} - m_{220} - m_{202} - m_{022} + 2)^2 \\
& +45 \sum_{r \neq s \neq t \neq u} (m_{2211} - m_{2011} - m_{0211})^2 \\
& +15 \sum_{r \neq s \neq t \neq u \neq v} (m_{21111} - m_{01111})^2 \\
& + \sum_{r \neq s \neq t \neq u \neq v \neq w} m_{111111}^2.
\end{aligned}$$

**1.3.14 Satz**

Für die vierte nichtverschwindende Komponente gilt die Gleichheit

$$\begin{aligned}
\frac{6!}{n} \hat{U}_{n,6}^2 &= b_{600} - 30 b_{410} + 15 (d+8) b_{400} + 90 b_{211} \\
&+ 60 b_{030} - 180 (d+4) b_{020} + 60 d (d^2 + 6d + 8).
\end{aligned}$$

Die dazugehörige affin-invariante Verteilungsgröße ist also

$$\begin{aligned}
& E[(\tilde{X}'\tilde{Y})^6] - 30 E[(\tilde{X}'\tilde{Y})^4 |\tilde{X}|^2] + 15 (d+8) E[(\tilde{X}'\tilde{Y})^4] \\
& + 90 E[(\tilde{X}'\tilde{Y})^2 |\tilde{X}|^2 |\tilde{Y}|^2] + 60 E[(\tilde{X}'\tilde{X})^3] \\
& + 180 (d+4) E[(\tilde{X}'\tilde{X})^2] + 60 d (d^2 + 6d + 8).
\end{aligned}$$

BEWEIS: Wie im Beweis von Satz 1.3.13 erhält man

$$\begin{aligned}
b_{600} &= \sum_r m_6^2 + \sum_{r \neq s} (6 m_{51}^2 + 15 m_{42}^2 + 10 m_{33}^2) \\
&+ \sum_{r \neq s \neq t} (15 m_{411}^2 + 60 m_{321}^2 + 15 m_{222}^2) \\
&+ \sum_{r \neq s \neq t \neq u} (20 m_{3111}^2 + 20 m_{3111}^2) + \sum_{r \neq s \neq t \neq u \neq v} m_{111111}^2,
\end{aligned}$$

$$\begin{aligned}
b_{410} &= \sum_r m_6 m_4 + \sum_{r \neq s} (m_{42} m_{40} + 4m_{51} m_{31} + 4m_{33} m_{31} + 6m_{42} m_{22}) \\
&+ \sum_{r \neq s \neq t} (4m_{312} m_{310} + 3m_{222} m_{220} + 6m_{411} m_{211} + 12m_{321} m_{121}) \\
&+ \sum_{r \neq s \neq t \neq u} (6m_{2211} m_{2011} + 4m_{3111} m_{1111}) \\
&+ \sum_{r \neq s \neq t \neq u \neq v} m_{21111} m_{01111}
\end{aligned}$$

und schließlich

$$\begin{aligned}
b_{211} &= \sum_r m_4^2 + \sum_{r \neq s} (m_{22}^2 + 2m_{40} m_{22} + 2m_{31}^2) \\
&+ \sum_{r \neq s \neq t} (m_{220} m_{202} + m_{211}^2 + 4m_{310} m_{112}) \\
&+ \sum_{r \neq s \neq t \neq u} m_{1120} m_{1102}.
\end{aligned}$$

Für die Wölbung nach Mardia  $b_{020}$  bzw. für die Variante  $b_{2,d}^* = b_{400}$  gilt nach Abschnitt 1.3.1

$$\begin{aligned}
b_{020} &= \sum_r m_4 + \sum_{r \neq s} m_{22} \quad \text{bzw.} \\
b_{400} &= \sum_r m_4^2 + \sum_{r \neq s} (4m_{31}^2 + 3m_{22}^2) + \sum_{r \neq s \neq t} (6m_{211}^2 + m_{1111}^2).
\end{aligned}$$

Schließlich benötigt man noch

$$b_{030} = \sum_r m_6 + 3 \sum_{r \neq s} m_{42} + \sum_{r \neq s \neq t} m_{222}.$$

Addiert man alle Terme mit den im Satz angegebenen Vielfachheiten, so erhält man alle Terme der rechten Seite in (1.22). ■

Zur Bildung der fünften nichtverschwindenden Komponente  $\hat{U}_{n,7}^2$  tragen bereits 15 verschiedene Produkte von Hermite-Polynomen bei, wobei wiederum über alle Koordinatenvariablen zu summieren ist. Ohne Beweis sei das entsprechende Ergebnis für diese Komponente angegeben.

**1.3.15 Satz**

Für die fünfte nichtverschwindende Komponente gilt die Gleichheit

$$\begin{aligned} \frac{7!}{n} \hat{U}_{n,7}^2 &= b_{700} - 42 b_{510} + 21 (d+10) b_{500} + 210 b_{311} \\ &\quad - 420 (d+10) b_{310} - 630 b_{121} \\ &\quad + 105 (d^2 + 14d + 48) b_{300} + 630 (d+6) b_{111}. \end{aligned}$$

Die dazugehörige affin-invariante Verteilungsgröße ist also

$$\begin{aligned} &E[(\tilde{X}'\tilde{Y})^7] - 42 E[(\tilde{X}'\tilde{Y})^5 |\tilde{X}|^2] + 21 (d+10) E[(\tilde{X}'\tilde{Y})^5] \\ &\quad + 210 E[(\tilde{X}'\tilde{Y})^3 |\tilde{X}|^2 |\tilde{Y}|^2] - 420 (d+8) E[(\tilde{X}'\tilde{Y})^3 |\tilde{X}|^2] \\ &\quad - 630 E[(\tilde{X}'\tilde{Y}) |\tilde{X}|^4 |\tilde{Y}|^2] + 105 (d^2 + 14d + 48) E[(\tilde{X}'\tilde{X})^3] \\ &\quad + 630 (d+6) E[(\tilde{X}'\tilde{Y}) |\tilde{X}|^2 |\tilde{Y}|^2]. \end{aligned}$$

Die Sätze 1.3.13-1.3.15 lassen erkennen, daß bei der Bildung der Komponenten immer wieder analoge Terme auftreten. Ein allgemeines Bildungsgesetz wurde allerdings nicht gefunden. Dies ist auch mehr von theoretischer Bedeutung, da in der Praxis höchstens die ersten vier oder fünf nichtverschwindenden Komponenten als Testgrößen verwendet werden.

## 1.4 Die asymptotische Verteilung von Verallgemeinerungen der multivariaten Schiefe und Wölbung

In diesem Abschnitt soll gezeigt werden, wie mit den bisher entwickelten Methoden auch andere Verteilungsgrößen behandelt werden können, die in keinem direkten Zusammenhang mit den Komponenten eines glatten Anpassungstests stehen. Unter anderem wird die asymptotische Verteilung der Testgröße  $b_{2,d}^*$  berechnet.



Grundlegend ist wieder Satz 1.1.1 a), der nicht voraussetzt, daß die verwendeten Polynome Orthogonalpolynome sind (siehe Bemerkung 1 nach Satz 1.1.1). Allerdings ist die Berechnung der Kovarianzmatrix  $\Sigma$  der Polynome, die die Testgröße bilden, schwieriger: Während in den Beispielen in Abschnitt 1.3 ausgenutzt wurde, daß sich die Kovarianzmatrix sowohl unter der parametrischen Hypothese der Normalverteilung als auch in der Verteilungsklasse  $\mathcal{P}_0^k$  nach (1.11) berechnen läßt, muß für nicht-orthogonale Polynome immer (d.h. auch unter der parametrischen Hypothese) die Kovarianzmatrix nach Gleichung (1.9) bestimmt werden. Der Grund liegt darin, daß im Beweis von Satz 1.2.1 die Ausdrücke  $E_\vartheta[\nabla_\vartheta h_k(X, \vartheta)]$  i. allg. nur verschwinden, wenn die  $h_k(\cdot, \vartheta)$  Orthogonalpolynome sind. Die in diesem Abschnitt verwendete Notation ist dieselbe wie in Abschnitt 1.3.

Im folgenden soll die asymptotische Verteilung von Größen der Form

$$b_{k00} = \frac{1}{n^2} \sum_{i,j=1}^n ((X_i - \bar{X}_n)' S_n^{-1} (X_j - \bar{X}_n))^k = \frac{1}{n^2} \sum_{i,j=1}^n (Z_i' Z_j)^k$$

mit einer festen natürlichen Zahl  $k$  ermittelt werden. Dabei wird wie oben vorausgesetzt, daß  $S_n$  mit Wahrscheinlichkeit 1 nichtsingulär ist. Die Testgröße läßt sich auch folgendermaßen schreiben:

$$b_{k00} = \sum_{\substack{k_1, \dots, k_d \geq 0 \\ k_1 + \dots + k_d = k}} \binom{k}{k_1 \dots k_d} \left( \frac{1}{n} \sum_{i=1}^n \prod_{r=1}^d z_{ri}^{k_r} \right)^2.$$

Sie besteht also aus einer Summe von

$$\sum_{\substack{k_1, \dots, k_d \geq 0 \\ k_1 + \dots + k_d = k}} \binom{k}{k_1 \dots k_d} = d^k$$

empirischen Momenten, die jeweils quadriert werden. Die zu  $b_{k00}$  gehörige Verteilungsgröße ist

$$\begin{aligned} \beta_{k00} &= E[((X - \mu)' T^{-1} (Y - \mu))^k] \\ &= \sum_{\substack{k_1, \dots, k_d \geq 0 \\ k_1 + \dots + k_d = k}} \binom{k}{k_1 \dots k_d} \left( E \left[ \prod_{r=1}^d \xi_r^{k_r} \right] \right)^2, \end{aligned}$$

wobei  $\xi_1, \dots, \xi_d$  die Komponenten des Vektors  $\tilde{X} = T^{-1/2}(X - \mu)$  sind. Da diese Größen offensichtlich affin-invariant sind, gelte hierbei wieder o.E.  $\mu = 0, T = I_d$ . Weiter setzen wir wieder  $E\|X\|^{2k} < \infty$  voraus.

Die in Abschnitt 1.3 verwendete Parametrisierung bereitet hier Schwierigkeiten, da man dann die partiellen Ableitungen von  $T^{-1/2}$  nach den Elementen von  $T^{-1}$  benötigt. Deshalb verwenden wir neben dem Erwartungswertvektor  $\mu$  die Elemente  $t_{ij}^{-1/2}$  von  $T^{-1/2}$  als Parameter. Die lineare Darstellung des Schätzers, die jetzt explizit benötigt wird, läßt sich auch in diesem Fall leicht angeben: aus

$$\sqrt{n}(S_n^{-1} - I_d) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i X_i' - I_d) + o_P(1)$$

und

$$\sqrt{n}(S_n^{-1} - I_d) = \sqrt{n}(S_n^{-1/2} - I_d)(S_n^{-1/2} + I_d)$$

erhält man

$$\sqrt{n}(S_n^{-1/2} - I_d) = -\frac{1}{2\sqrt{n}} \sum_{i=1}^n (X_i X_i' - I_d) + o_P(1).$$

Es gilt also  $l_{ij}(x, \vartheta) = -(x_i x_j - \delta_{ij})/2$ . Bei dieser Parametrisierung lassen sich die benötigten partiellen Ableitungen einfach berechnen. Unter Beachtung von

$$\left. \frac{\partial \xi_r}{\partial \mu_i} \right|_{(\mu, T) = (0, I_d)} = \left. -t_{ri}^{-1/2} \right|_{(0, I_d)} = -\delta_{ri}$$

ergibt sich im Falle  $k_i \geq 1$  für  $i = 1, \dots, d$

$$\left. \frac{\partial \xi_1^{k_1} \dots \xi_d^{k_d}}{\partial \mu_i} \right|_{(0, I_d)} = -k_i \xi_1^{k_1} \dots \xi_{i-1}^{k_{i-1}} \xi_i^{k_i-1} \xi_{i+1}^{k_{i+1}} \dots \xi_d^{k_d}. \quad (1.23)$$

Weiter erhält man mit

$$\left. \frac{\partial \xi_r}{\partial t_{ij}^{-1/2}} \right|_{(0, I_d)} = \delta_{ir} \xi_j$$

die Ableitungen

$$\left. \frac{\partial \xi_1^{k_1} \cdots \xi_d^{k_d}}{\partial t_{ij}^{-1/2}} \right|_{(0, I_d)} = k_i \xi_1^{k_1} \cdots \xi_i^{k_i-1} \cdots \xi_j^{k_j+1} \cdots \xi_d^{k_d}, \quad (1.24)$$

falls  $k_i \geq 1$  gilt ( $i, j = 1, \dots, d$ ). Damit sind alle Größen bekannt, die man benötigt, um die Kovarianzmatrix  $\Sigma$  zu berechnen. Numeriert man die „Bausteine“  $\binom{k}{k_1, \dots, k_d}^{1/2} x_1^{k_1} \cdots x_d^{k_d}$  der Testgröße in irgendeiner Weise durch und bezeichnet sie mit  $h_l(x)$ ,  $l = 1, \dots, d^k$ , definiert man  $\tau = (\tau_1, \dots, \tau_{d^k})$  durch  $\tau_l = E[h_l(\tilde{X})]$  und setzt

$$\begin{aligned} v_l(x) &= h_l(x) + \sum_{i=1}^d E \left( \frac{\partial h_l(\tilde{X})}{\partial \mu_i} \right) x_i \\ &\quad - \frac{1}{2} \sum_{i,j=1}^d E \left( \frac{\partial h_l(\tilde{X})}{\partial t_{ij}^{-1/2}} \right) (x_i x_j - \delta_{ij}), \end{aligned} \quad (1.25)$$

so berechnet sich  $\Sigma$  nach (1.9).

Für die Grenzverteilung sind wie in Abschnitt 1.3 zwei Fälle zu unterscheiden. Ist  $\tau = 0$ , so ist die asymptotische Verteilung eine gewichtete Summe von unabhängigen  $\chi_1^2$ -verteilten Zufallsvariablen:

$$n b_{k00} \xrightarrow{\mathcal{D}} \sum_{j=1}^{d^k} \lambda_j \chi_1^2(j),$$

wobei die Gewichte die Eigenwerte von  $\Sigma$  sind. Man beachte, daß dieser Fall nur für ungerades  $k$  vorkommen kann.

Ist  $\tau \neq 0$  (was bei geradem  $k$  für nicht-degenerierte Verteilungen immer der Fall ist), so folgt aus (1.13) wegen  $\tau' \tau = \beta_{k00}$

$$\sqrt{n} (b_{k00} - \beta_{k00}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4\tau' \Sigma \tau). \quad (1.26)$$

#### 1.4.1 Beispiel

In diesem Beispiel sollen die Grenzverteilungen für ungerade Werte von  $k$  für die Klasse der elliptisch-symmetrischen Verteilungen kurz angesprochen

werden. Da in diesen Fällen  $\tau = 0$  gilt, ist die asymptotische Verteilung eine gewichtete Summe von unabhängigen  $\chi_1^2$ -verteilten Zufallsvariablen. Für  $k = 3$  stimmt die Testgröße  $b_{300}$  mit der Mardia-Schiefe  $b_{1,d}$  überein (siehe Abschnitt 1.3.1); in ihr kommen Polynome der Form  $x_r^3, x_r^2 x_s$  bzw.  $x_r x_s x_t$  (mit Vorfaktoren) vor. Nach Satz A.2.3 verschwinden die meisten Summanden in (1.25); die  $v_l$  sind von der Form  $x_r^3 - 3x_r, x_r^2 x_s - x_s$  bzw.  $x_r x_s x_t$ . Dies sind gerade die  $L_r, L_{rs}$  bzw.  $L_{rst}$  aus Beispiel 1.3.1, so daß man mit denselben Überlegungen wie dort auf die genaue Form der asymptotischen Verteilung kommt. Für höhere ungerade  $k$ -Werte steigt die Zahl verschiedener Polynome schnell an, so daß eine explizite Berechnung der Eigenwerte der Kovarianzmatrix sehr schwierig sein dürfte. Dies soll nicht weiter verfolgt werden; die genaue Form der Gewichte für ungerade  $k$  wurde von Henze, Gutjahr und Folkers (1997) auf anderem Wege berechnet.

### 1.4.2 Beispiel

In Abschnitt 1.3.2 kam die Variante  $b_{2,d}^* = \frac{1}{n^2} \sum_{i,j=1}^n (Z_i' Z_j)^4$  der multivariaten Wölbung vor. Hier soll die Varianz der asymptotischen Normalverteilung bestimmt werden, falls die zugrundeliegende Verteilung zur Klasse der elliptisch-symmetrischen Verteilungen gehört.

Für beliebige Verteilungen, die den Voraussetzungen dieses Abschnitts genügen, ist  $\sqrt{n}(b_{2,d}^* - \beta_{2,d}^*) = \sqrt{n}(b_{400} - \beta_{400})$  asymptotisch normalverteilt, wobei sich die Varianz nach (1.26) berechnet. Dabei sieht man, daß Polynome  $h_l$ , für die  $\tau_l = 0$  gilt, weder in die Zentrierung noch in die Varianz der Grenzverteilung eingehen. Ist die zugrundeliegende Verteilung eine elliptisch-symmetrische Verteilung, so kann deshalb statt  $b_{400}$  die Größe

$$\tilde{b}_{400} = \frac{1}{n^2} \left\{ \sum_{r=1}^d \left( \sum_i z_{ri}^4 \right)^2 + 6 \sum_{r < s} \left( \sum_i (z_{ri}^2 z_{si}^2) \right)^2 \right\}$$

mit der zugehörigen Verteilungsgröße

$$\tilde{\beta}_{400} = \sum_{r=1}^d (E[\xi_r^4])^2 + 6 \sum_{r < s} (E[\xi_r^2 \xi_s^2])^2$$

betrachtet werden. Dabei kommen nur die Polynome  $h_r(x) = x_r^4$  ( $1 \leq r \leq d$ ) und  $h_r^s(x) = \sqrt{6}x_r^2x_s^2$  ( $1 \leq r < s \leq d$ ) vor. Nach (1.25) erhält man  $v_r(x) = x_r^4 - 2\mu_4(x_r^2 - 1)$  und  $v_r^s(x) = \sqrt{6}(x_r^2x_s^2 - \mu_{22}(x_r^2 + x_s^2 - 2))$ . Unter Beachtung von Korollar A.2.4 und mit  $\tau_1 = E[h_r(\tilde{X})] = \mu_4$ ,  $\tau_2 = E[h_r^s(\tilde{X})] = \sqrt{6}\mu_{22} = \sqrt{6}\mu_4/3$  ergeben sich die folgenden unterschiedlichen Einträge in der Kovarianzmatrix  $\Sigma$ :

$$\begin{aligned}\sigma^{11} &= E[v_r^2(\tilde{X})] - \tau_1^2 = \mu_8 + 4\mu_4(\mu_4^2 - \mu_6) - \mu_4^2, \\ \sigma^{12} &= E[v_r(\tilde{X})v_s(\tilde{X})] - \tau_1^2 = \frac{3\mu_8}{35} + 4\mu_4\left(\frac{\mu_4^2}{3} - \frac{\mu_6}{5}\right) - \mu_4^2, \\ \sigma^{21} &= E[(v_r^s(\tilde{X}))^2] - \tau_2^2 = 6\left(\frac{3\mu_8}{35} + \frac{4\mu_4}{3}\left(\frac{2\mu_4^2}{9} - \frac{\mu_6}{5}\right) - \frac{\mu_4^2}{9}\right), \\ \sigma^{22} &= E[v_r^s(\tilde{X})v_r^t(\tilde{X})] - \tau_2^2 = 6\left(\frac{\mu_8}{35} + \frac{2\mu_4}{9}\left(\mu_4^2 - \frac{4\mu_6}{5}\right) - \frac{\mu_4^2}{9}\right), \\ \sigma^{23} &= E[v_r^s(\tilde{X})v_t^u(\tilde{X})] - \tau_2^2 = 6\left(\frac{\mu_8}{105} + \frac{4\mu_4}{9}\left(\frac{\mu_4^2}{3} - \frac{\mu_6}{5}\right) - \frac{\mu_4^2}{9}\right), \\ \sigma^{31} &= E[v_r(\tilde{X})v_r^s(\tilde{X})] - \tau_1\tau_2 = \sqrt{6}\left(\frac{\mu_8}{7} + 4\mu_4\left(\frac{2\mu_4^2}{9} - \frac{\mu_6}{5}\right) - \frac{\mu_4^2}{3}\right), \\ \sigma^{32} &= E[v_r(\tilde{X})v_s^t(\tilde{X})] - \tau_1\tau_2 = \sqrt{6}\left(\frac{\mu_8}{35} + \frac{4\mu_4}{3}\left(\frac{\mu_4^2}{3} - \frac{\mu_6}{5}\right) - \frac{\mu_4^2}{3}\right),\end{aligned}$$

wobei  $r, s, t, u \in \{1, \dots, d\}$ ,  $r < s < t < u$  gilt. Damit ergibt sich

$$\begin{aligned}4\tau'\Sigma\tau &= 4\mu_4^2\left(d\sigma^{11} + d(d-1)\sigma^{12} + \frac{\sqrt{6}}{3}\left(4\binom{d}{2}\sigma^{31} + 2(d-2)\binom{d}{2}\sigma^{32}\right)\right. \\ &\quad \left.+ \left(\frac{\sqrt{6}}{3}\right)^2\left(\binom{d}{2}\sigma^{21} + 2(d-2)\binom{d}{2}\sigma^{22} + \binom{d}{2}\binom{d-2}{2}\sigma^{23}\right)\right) \\ &= 4\mu_4^2\left(\mu_8\frac{d}{105}(d^3 + 12d^2 + 44d + 48)\right. \\ &\quad \left.- \mu_6\mu_4\frac{4d}{45}(d^3 + 8d^2 + 20d + 16)\right. \\ &\quad \left.+ \mu_4^3\frac{4d}{27}(d^3 + 6d^2 + 12d + 8) - \mu_4^2\frac{d^2}{9}(d+2)^2\right).\end{aligned}$$

Ersetzt man  $\mu_{2k}$ ,  $k = 2, 3, 4$ , nach Satz 1.4.7 a) durch  $r_{2k} \prod_{j=0}^{k-1} (2j+1)/(d+2j)$ , so lässt sich die asymptotische Varianz auch kurz als

$$4\tau' \Sigma \tau = \frac{36r_4^2}{d^2(d+2)^2} \left( r_8 - r_4^2 + 4\frac{r_4}{d} \left( \frac{r_4^2}{d} - r_6 \right) \right)$$

schreiben. Dies ist das Resultat von Henze (1994b), Example 3.2. Ist die zugrundeliegende Verteilung eine multivariate Normalverteilung, so ergibt sich als Varianz  $288d(d+2)$ . ■

Um den Fall  $\beta_{k00} > 0$ ,  $k \geq 3$ , allgemein zu behandeln, ist es hilfreich, die Varianz  $\sigma^2$  der Grenzverteilung in anderer Form zu schreiben. Seien dazu  $h = (h_1, \dots, h_{d^k})$  und  $v = (v_1, \dots, v_{d^k})$  gesetzt. Weiter definieren wir

$$h_{1,k}(x) := \tau' h(x) = E[(x'X)^k]. \quad (1.27)$$

Nach (1.26) und (1.9) gilt

$$\begin{aligned} \sigma^2 &= 4\tau' E[v(\tilde{X})v(\tilde{X})'] \tau - 4\tau' (\tau \tau') \tau \\ &= E[(2\tau' v(\tilde{X})) (2\tau' v(\tilde{X}))'] - 4\beta_{k00}^2. \end{aligned} \quad (1.28)$$

In dem Skalarprodukt  $\tau' v(x)$  kommen neben  $h_{1,k}(x)$  Terme der Art  $x_i \sum_l \tau_l E[\partial h_l(\tilde{X})/\partial \mu_i]$  bzw.  $(x_i x_j - \delta_{ij}) \sum_l \tau_l E[\partial h_l(\tilde{X})/\partial t_{ij}^{-1/2}]$  vor. Für die Summen erhält man aus (1.23) und (1.24) durch Koeffizientenvergleich

$$\sum_{l=1}^{d^k} E[h_l(\tilde{X})] E \left[ \frac{\partial h_l(\tilde{X})}{\partial \mu_i} \right] = (-k) E[(\tilde{X}' \tilde{Y})^{k-1} \eta_i]$$

bzw.

$$\sum_{l=1}^{d^k} E[h_l(\tilde{X})] E \left[ \frac{\partial h_l(\tilde{X})}{\partial t_{ij}^{-1/2}} \right] = k E[\xi_i (\tilde{X}' \tilde{Y})^{k-1} \eta_j].$$

Unter Beachtung dieser Gleichungen und mit den Definitionen

$$\begin{aligned} a_k &= E \left[ (\tilde{X}' \tilde{Y})^{k-1} \tilde{Y}' \right], \\ B_k &= (b_{ij})_{1 \leq i, j \leq d} = E \left[ \tilde{X} (\tilde{X}' \tilde{Y})^{k-1} \tilde{Y}' \right], \\ u_k &= (2, -k b_{11}, -k b_{12}, \dots, -k b_{1d}, -k b_{21}, \dots, -k b_{dd}, -2k a'_k)', \\ Z_k &= (h_{1,k}(\tilde{X}) - \beta_{k00}, \xi_1^2 - 1, \xi_1 \xi_2, \dots, \xi_1 \xi_d, \xi_2 \xi_1, \dots, \xi_d^2 - 1, \tilde{X}')', \end{aligned} \quad (1.29)$$

läßt sich die asymptotische Varianz aus (1.28) als  $\sigma^2 = u'_k E[Z_k Z'_k] u_k$  schreiben. Zusammenfassend erhält man folgendes Resultat.

### 1.4.3 Satz

Für den Zufallsvektor  $X$  mit Erwartungswert  $\mu$  und nichtsingulärer Kovarianzmatrix  $T$  gelte

$$E[\{(X - \mu)' T^{-1} (X - \mu)\}^k] = E[(\tilde{X}' \tilde{X})^k] < \infty.$$

Ist die empirische Kovarianzmatrix  $S_n$  mit Wahrscheinlichkeit 1 nichtsingulär und ist  $\beta_{k00} > 0$ , so gilt mit den in (1.29) definierten  $(1 + d^2 + d)$ -dimensionalen Vektoren  $u_k$  und  $Z_k$

$$\sqrt{n} (b_{k00} - \beta_{k00}) \xrightarrow{D} \mathcal{N}(0, u'_k E[Z_k Z'_k] u_k).$$

■

**Bemerkung:** Für  $k = 3$  bzw.  $k = 4$  bzw. ungerades  $k$  ist dies die Aussage von Theorem 3.2 in [7] bzw. Theorem 2.2 in [30] bzw. Theorem 4.3 in [34]. Die dort jeweils benötigte Bedingung, daß der Träger von  $P^X$  positives Lebesgue-Maß besitzt, ist durch den Zugang über die Theorie der V-Statistiken bedingt und wird in der hier vorliegenden Fassung nicht benötigt.

### 1.4.4 Beispiel

Im univariaten Fall  $d = 1$  ergeben sich die Größen aus (1.27) und (1.29) zu

$$h_{1,k}(x) = \mu_k x^k, \quad a_k = \mu_{k-1} \mu_k, \quad B_k = \mu_k^2, \\ u_k = (2, -k\mu_k^2, -2k\mu_{k-1}\mu_k)', \quad Z_k = (\mu_k \tilde{X}^k - \mu_k^2, \tilde{X}^2 - 1, \tilde{X})'$$

und damit

$$E[Z_k Z'_k] = \begin{bmatrix} \mu_k^2 (\mu_{2k} - \mu_k^2) & \mu_k (\mu_{k+2} - \mu_k) & \mu_k \mu_{k+1} \\ \mu_k (\mu_{k+2} - \mu_k) & \mu_4 - 1 & \mu_3 \\ \mu_k \mu_{k+1} & \mu_3 & 1 \end{bmatrix}.$$

Hieraus erhält man

$$\begin{aligned} u_k' E[Z_k Z_k'] u_k &= 4\mu_k^2 \left( \mu_{2k} - k \mu_k \mu_{k+2} + \frac{k^2}{4} \mu_4 \mu_k^2 - \frac{(k-2)^2}{4} \mu_k^2 \right. \\ &\quad \left. + k^2 \mu_3 \mu_{k-1} \mu_k - 2k \mu_{k-1} \mu_{k+1} + k^2 \mu_{k-1}^2 \right). \end{aligned}$$

Im Fall  $k = 3$  ergibt sich das Resultat aus Beispiel 1.3.7. ■

#### 1.4.5 Beispiel

Hier soll in Verallgemeinerung von Beispiel 1.4.2 die Varianz der asymptotischen Verteilung von  $\sqrt{n}(b_{2k,0,0} - \beta_{2k,0,0})$  ( $k \geq 2$ ) für elliptisch-symmetrische Verteilungen bestimmt werden. In dieser Klasse besitzt  $\beta_{2k,0,0}$  eine einfache Darstellung. Für deren Beweis benötigen wir Lemma 2.5.1 in Fang und Zhang (1990):

**1.4.6 Lemma** Sind  $g(\cdot)$  eine nichtnegative meßbare Funktion auf  $\mathbb{R}$  und  $a \in \mathbb{R}^d \setminus \{0\}$ , so gilt

$$\int_{S: x'x=c^2} g(a'x) ds = \frac{2c\pi^{(d-1)/2}}{(d-1)/2} \int_{-c}^c g(\|a\|y) (c^2 - y^2)^{\frac{d-3}{2}} dy.$$

#### 1.4.7 Satz

$X = (X_1, \dots, X_d)'$  sei sphärisch-symmetrisch verteilt mit  $E[\|X\|^{2k}] < \infty$ ;  $Y$  sei eine unabhängige Kopie von  $X$ .

a) Mit  $\mu_{2k} = E[X_1^{2k}]$  gilt

$$r_{2k} = E[(X'X)^k] = \mu_{2k} \frac{d(d+2) \cdots (d+2k-2)}{1 \cdot 3 \cdot 5 \cdots (2k-1)}.$$

b) Es gilt

$$\beta_{2k,0,0} = E[(X'Y)^{2k}] = r_{2k}^2 \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{d(d+2) \cdots (d+2k-2)}$$

und

$$B_{2k} = E[X(X'Y)^{2k-1}Y'] = r_{2k}^2 \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{d^2(d+2) \cdots (d+2k-2)} I_d.$$



**Beweis:**

a) Ist  $N \sim \mathcal{N}(0, I_d)$ , so gilt bekanntlich

$$E(\|N\|^s) = \frac{, (\frac{d+s}{2}) 2^{s/2}}{, (d/2)}$$

und somit

$$E[(N'N)^k] = \frac{, (\frac{d}{2} + k) 2^k}{, (d/2)} = d(d+2) \cdots (d+2(k-1)).$$

Wegen Satz A.2.3 gilt

$$\frac{\mu_{2k}}{\mu_{2k}^N} = \frac{r_{2k}}{r_{2k}^N},$$

wobei mit  $\mu_{2k}^N$  bzw.  $r_{2k}^N$  die entsprechenden Größen für  $N$  bezeichnet werden. Wegen  $\mu_{2k}^N = 1 \cdot 3 \cdot 5 \cdots (2k-1)$  folgt die erste Behauptung.

b) Für  $d = 1$  ist nichts zu zeigen. Seien  $d > 1$  und  $U$  ein auf  $S_d$ , der Oberfläche der Einheitskugel des  $\mathbb{R}^d$ , gleichverteilter Zufallsvektor. Für die Fläche  $A$  von  $S_d$  gilt

$$A = 2 \pi^{d/2} / , (d/2).$$

Ist  $u_1 \in \mathbb{R}^d$  mit  $\|u_1\| = 1$ , so folgt mit Lemma 1.4.6

$$\begin{aligned} E[(u_1'U)^{2k}] &= \frac{1}{A} \int_{S: x'x=1} (u_1'x)^{2k} ds \\ &= \frac{1}{A} , \frac{2 \pi^{(d-1)/2}}{((d-1)/2)} \int_{-1}^1 y^{2k} (1-y^2)^{\frac{d-3}{2}} dy \\ &= \frac{, (d/2)}{\sqrt{\pi}, ((d-1)/2)} 2 \int_0^1 x^k (1-x)^{\frac{d-3}{2}} \frac{dx}{2\sqrt{x}}. \end{aligned}$$

Mit der Betafunktion

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{, (\alpha), (\beta)}{, (\alpha + \beta)}$$

folgt weiter

$$\begin{aligned}
 E[(u'_1 U)^{2k}] &= \frac{1}{\sqrt{\pi}} \frac{\Gamma(d/2)}{\Gamma((d-1)/2)} B\left(k + \frac{1}{2}, \frac{d-1}{2}\right) \\
 &= \frac{\Gamma(d/2) \Gamma(k + \frac{1}{2})}{\sqrt{\pi} \Gamma(k + \frac{d}{2})} \\
 &= \prod_{j=0}^{k-1} \frac{2j+1}{d+2j}.
 \end{aligned}$$

Ist nun  $X$  sphärisch-symmetrisch verteilt, so folgt aus der Zerlegung  $X \stackrel{D}{=} R U$ , wobei  $R \stackrel{D}{=} \|X\|$  und  $R$  und  $U$  unabhängig sind (siehe Satz A.2.2),

$$h_{1,2k}(z) = E[(z' X)^{2k}] = \|z\|^{2k} E[R^{2k}] \prod_{j=0}^{k-1} \frac{2j+1}{d+2j} \quad (1.30)$$

und daraus die erste Behauptung in Teil b) des Satzes.

Für den Beweis der letzten Aussage überlegt man sich, daß wegen Satz A.2.3  $B_{2k} = b_{11} I_d$  gelten muß. Wegen  $\text{spur}(B_{2k}) = \beta_{2k,0,0}$  folgt daraus  $b_{11} = \beta_{2k,0,0}/d$ . ■

**Bemerkung:** Aus Teil a) und b) von Satz 1.4.7 folgen die Gleichheiten

$$\beta_{2k,0,0} = \mu_{2k}^2 \frac{d(d+2) \cdots (d+2k-2)}{1 \cdot 3 \cdot 5 \cdots (2k-1)}$$

und

$$\beta_{2k,0,0} = \mu_{2k} r_{2k}.$$

Wegen Satz A.2.3 gilt für den Vektor  $a_{2k}$  aus (1.29)  $a_{2k} = 0$ . Setzt man

$$\delta = \prod_{j=0}^{k-1} (2j+1)/(d+2j), \quad (1.31)$$

so erhält man aus Satz 1.4.7 b)

$$u_k = (2, -2k \delta r_{2k}^2/d \cdot e'_1, -2k \delta r_{2k}^2/d \cdot e'_2, \dots, -2k \delta r_{2k}^2/d \cdot e'_d, 0)'$$

wobei  $e_j = (0, \dots, 1, 0, \dots, 0)'$  den  $j$ -ten Einheitsvektor im  $\mathbb{R}^d$  bezeichnet.

Mit  $L_{2k} = h_{1,2k}(\tilde{X}) - \beta_{2k,0,0}$  und

$$W_j = (\xi_j \xi_1, \dots, \xi_j \xi_{j-1}, \xi_j^2 - 1, \xi_j \xi_{j+1}, \dots, \xi_j \xi_d)' \quad (j = 1, \dots, d)$$

läßt sich der Vektor  $Z_{2k}$  aus (1.29) schreiben als

$$Z_{2k} = (L_{2k}, W'_1, \dots, W'_d, \tilde{X}')'$$

Wegen (1.30) und Satz 1.4.7 b) gilt

$$L_{2k} = \delta r_{2k} (\|\tilde{X}\|^{2k} - r_{2k}),$$

woraus  $E[L_{2k} \tilde{X}] = 0$ ,  $E[L_{2k}^2] = \delta^2 r_{2k}^2 (r_{4k} - r_{2k}^2)$  und

$$E[L_{2k} W_j] = \delta r_{2k} \left( \frac{r_{2k+2}}{d} - r_{2k} \right) e_j \quad (j = 1, \dots, d)$$

folgen. Weiter gilt  $E[\tilde{X} W'_j] = O$  ( $j = 1, \dots, d$ ), wobei  $O$  die Null-Matrix der Ordnung  $d$  bezeichnet. Definiert man  $(d \times d)$ -Matrizen

$$B_{ij} = E[W_i W'_j] = (b_{k,l}^{(i,j)})_{1 \leq k, l \leq d} \quad (i, j = 1, \dots, d)$$

und setzt  $\rho = \delta r_{2k} \left( \frac{r_{2k+2}}{d} - r_{2k} \right)$ , so nimmt die Matrix  $E[Z_{2k} Z'_{2k}]$  der Ordnung  $1 + d^2 + d$  die Form

$$E[Z_{2k} Z'_{2k}] = \begin{bmatrix} E[L_{2k}^2] & \rho e'_1 & \rho e'_2 & \cdots & \rho e'_d & 0' \\ \rho e'_1 & B_{11} & B_{12} & \cdots & B_{1d} & O \\ \rho e'_2 & B_{21} & B_{22} & \cdots & B_{2d} & O \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho e'_d & B_{d1} & B_{d2} & \cdots & B_{dd} & O \\ 0 & O & O & \cdots & O & I_d \end{bmatrix}$$

an. Beachtet man

$$b_{i,i}^{(i,i)} = \frac{3r_4}{d(d+2)} - 1, \quad i = 1, \dots, d,$$

und

$$b_{i,j}^{(i,j)} = \frac{r_4}{d(d+2)} - 1, \quad i, j = 1, \dots, d, \quad i \neq j,$$

so erhält man für  $\sigma_{2k}^2 = u'_{2k} E[Z_{2k} Z'_{2k}] u_{2k}$

$$\sigma_{2k}^2 = 4\delta^2 r_{2k}^2 \left( r_{4k} - (k-1)^2 r_{2k}^2 - \frac{2k}{d} r_{2k} r_{2k+2} + \frac{k^2}{d^2} r_4 r_{2k}^2 \right), \quad (1.32)$$

wobei  $\delta$  in (1.31) definiert ist. Wir fassen zusammen.

#### 1.4.8 Satz

Der Zufallsvektor  $X$  sei elliptisch-symmetrisch verteilt mit Erwartungswert  $\mu$  und nichtsingulärer Kovarianzmatrix  $T$ . Weiter gelte

$$E[\{(X - \mu)' T^{-1} (X - \mu)\}^k] < \infty.$$

Die empirische Kovarianzmatrix  $S_n$  sei mit Wahrscheinlichkeit 1 nichtsingulär. Dann gilt

$$\sqrt{n}(b_{2k,0,0} - \beta_{2k,0,0}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_{2k}^2),$$

wobei  $\sigma_{2k}^2$  durch (1.32) gegeben ist. ■

**Bemerkung:** Die in Henze (1994b), Korollar 3.1, für den Fall  $k = 2$  gemachte Annahme  $P(X = 0) = 0$  sowie die Bedingung, daß der Träger von  $P^X$  positives Lebesgue-Maß besitzt, werden nicht benötigt.

**1.4.9 Korollar** Besitzt der Zufallsvektor  $X$  eine nicht-degenerierte  $d$ -dimensionale Normalverteilung, so gilt

$$\sqrt{n} \left( b_{2k,0,0} - \prod_{j=0}^{k-1} (2j+1)(d+2j) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_{2k}^2)$$

mit

$$\sigma_{2k}^2 = 4 \prod_{j=0}^{k-1} (2j+1)^2 (d+2j) \left[ \prod_{j=k}^{2k-1} (d+2j) - \frac{d+2k^2}{d} \prod_{j=0}^{k-1} (d+2j) \right].$$

Insbesondere gilt für  $k = 2, 3$  bzw. 4

$$\sigma_4^2 = 288 d(d+2),$$

$$\sigma_6^2 = 4 \prod_{j=0}^2 (2j+1)^2 (d+2j) [24(3d+14)],$$

$$\sigma_8^2 = 4 \prod_{j=0}^3 (2j+1)^2 (d+2j) [96(3d^2+38d+14)]. \quad \blacksquare$$

## 1.5 Glatte Anpassungstests für die bivariate Poissonverteilung

Wie die Poisson-Verteilung im univariaten Fall ist die bivariate Poisson-Verteilung eine wichtige diskrete Verteilung im Zweidimensionalen. Eine ausführliche Beschreibung dieser Verteilung findet man in Kocherlakota und Kocherlakota (1992). Die bivariate Poisson-Verteilung besitzt drei Parameter  $\lambda_1^*$ ,  $\lambda_2^*$  und  $\lambda_3$ ; die Zähldichte ist gegeben durch

$$f(r, s; \vartheta) = e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{i=0}^{\min(r,s)} \frac{(\lambda_1^* - \lambda_3)^{r-i} (\lambda_2^* - \lambda_3)^{s-i} \lambda_3^i}{(r-i)!(s-i)!i!};$$

hier ist  $\vartheta = (\lambda_1^*, \lambda_2^*, \lambda_3)$ . Oft wird auch eine andere Parametrisierung verwendet mit  $\lambda_1 = \lambda_1^* - \lambda_3$ ,  $\lambda_2 = \lambda_2^* - \lambda_3$  und  $\lambda_3$ .

Ist der Zufallsvektor  $(X, Y)$  bivariate Poisson-verteilt, so gilt für die Randverteilungen  $X \sim \pi(\lambda_1^*)$  und  $Y \sim \pi(\lambda_2^*)$ , wobei  $\pi(\lambda)$  die (univariate) Poisson-Verteilung mit Parameter  $\lambda$  bezeichnet. Die faktoriellen Momente  $\mu_{[r,s]} = E[X^{[r]}Y^{[s]}]$  mit  $x^{[r]} = x(x-1)(x-2)\dots(x-r+1)$  sind gegeben durch

$$\mu_{[r,s]} = \lambda_1^{*r} \lambda_2^{*s} \sum_{i=0}^{\min(r,s)} \binom{r}{i} \binom{s}{i} i! \left( \frac{\lambda_3}{\lambda_1^* \lambda_2^*} \right)^i.$$

Daraus folgt insbesondere

$$E[X] = \lambda_1^*, \quad E[Y] = \lambda_2^* \quad \text{und} \quad Cov(X, Y) = \lambda_3.$$

Für die nächsthöheren nichtzentralen Momente  $\mu'_{r,s} = E[X^r Y^s]$  erhält man weiter

$$\begin{aligned}\mu'_{2,0} &= \lambda_1^{*2} + \lambda_1^*, \\ \mu'_{3,0} &= \lambda_1^{*3} + 3\lambda_1^{*2} + \lambda_1^*, \\ \mu'_{2,1} &= \lambda_1^{*2}\lambda_2^* + 2\lambda_1^*\lambda_3 + \lambda_3 + \lambda_1^*\lambda_2^*\end{aligned}$$

und entsprechende Gleichungen für  $\mu'_{0,2}, \mu'_{0,3}$  und  $\mu'_{1,2}$ . Daraus folgen für die zentralen Momente  $\mu_{r,s} = E[(X - EX)^r (Y - EY)^s]$  die Gleichungen  $\mu_{2,0} = \mu_{3,0} = \lambda_1^*$ ,  $\mu_{0,2} = \mu_{0,3} = \lambda_2^*$  und  $\mu_{2,1} = \mu_{1,2} = \lambda_3$ .

Wie in den vorangegangenen Abschnitten sollen glatte Anpassungstests konstruiert werden, die als gerichtete Tests diagnostische Eigenschaften besitzen. Nun ist es für beliebige stetige oder diskrete multivariate Verteilungen i. allg. nicht möglich, ein vollständiges System von multivariaten orthogonalen Polynomen zu konstruieren (bei der multivariaten Normalverteilung gelingt dies, da man durch eine affine Transformation alles auf den Fall von unkorrelierten und somit unabhängigen Zufallsvariablen zurückspielen kann). Stattdessen werden im Falle der bivariaten Poisson-Verteilung Polynome in zwei Variablen verwendet, mit deren Hilfe Abweichungen in bestimmten Momenten aufgedeckt werden können. Die Aussagen über die asymptotische Verteilung in (1.8) und (1.9) behalten auch in diesem Fall ihre Gültigkeit (siehe Bem. 1. nach Satz 1.1.1). Die allgemeinen Ausdrücke sollen hier nicht nochmals aufgeführt werden; als Beispiele werden zwei Tests angesprochen, die auch in Rayner und Best (1995) näher betrachtet werden.

### 1.5.1 Beispiel

In der Arbeit von Rayner und Best (1995) werden die glatten Anpassungstests wieder als "Score-Tests" gegenüber bestimmten Alternativen hergeleitet. Auch hier ergibt sich das Problem, daß die zugehörigen Alternativen überhaupt nicht existieren (es treten nichtkonvergente Reihen der Art  $\sum_j \exp(j^k)\lambda^j/j!$  auf).

Rayner und Best (1995) verwenden als Schätzer die Maximum-Likelihood-Schätzer. Als ML-Schätzer für  $\lambda_1^*$  und  $\lambda_2^*$  ergeben sich  $\hat{\lambda}_1^* = \bar{X}$  und  $\hat{\lambda}_2^* = \bar{Y}$ . Der ML-Schätzer  $\hat{\lambda}_3$  für  $\lambda_3$  unterscheidet sich dagegen von der

empirischen Kovarianz  $S_{XY} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$  (d.h. von dem zugehörigen Momentenschätzer) und muß mittels eines Iterationsverfahrens bestimmt werden (siehe Kocherlakota und Kocherlakota (1992), Abschnitt 4.7).

Aus diesem Grunde kann man einen Test auf  $h_1(x, y; \vartheta) = (x - \lambda_1^*)(y - \lambda_2^*) = (x - E[X])(y - (E[Y]))$  aufbauen. Die zugehörige zentrierte Komponente hat dann die Form

$$\begin{aligned} \hat{U}_{n,1} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( (X_i - \hat{\lambda}_1^*)(Y_i - \hat{\lambda}_2^*) - \hat{\lambda}_3 \right) \\ &= \sqrt{n}(S_{XY} - \hat{\lambda}_3) \end{aligned}$$

und ist unter der Hypothese der bivariaten Poissonverteilung asymptotisch normalverteilt mit Erwartungswert 0. Berechnet man die Varianz von  $\hat{U}_{n,1}$ , normiert damit und quadriert, so erhält man eine asymptotisch  $\chi_1^2$ -verteilte Testgröße (siehe Rayner und Best (1995), Abschnitt 3).

Der glatte Anpassungstest, der die Hypothese der bivariaten Poissonverteilung für zu große Werte von  $\hat{U}_{n,1}^2$  verwirft, ist offensichtlich kein diagnostischer Test: ein signifikantes Testergebnis bedeutet nicht notwendig, daß die Kovarianzstruktur der vorliegenden Daten mit derjenigen einer bivariaten Poisson-Verteilung (im Rahmen einer vorgegebenen Irrtumswahrscheinlichkeit) unvereinbar ist (denn  $\lambda_3$  kann jeden positiven Wert annehmen), sondern es bedeutet, daß das parametrische ML-Schätzverfahren zu einem signifikant anderem Ergebnis führt als das nichtparametrische Momentenverfahren, was auf die Unangemessenheit des hypothetischen parametrischen Modells hinweist.

Für einen diagnostischen Test, der Abweichung in höheren Momenten feststellen soll, sind dagegen die Momentenschätzer zu verwenden (siehe auch die Bemerkungen nach Lemma 1.1.2). Der Test aus Beispiel 1.5.1 ist in diesem Fall sinnlos, da die Testgröße dann immer Null ist.

Als Einwand gegen die Verwendung des Momentenschätzers von  $\lambda_3$  wird die geringe Effizienz von  $S_{XY}$  für große Werte des Korrelationskoeffizienten  $\rho$  vorgebracht (siehe Kocherlakota und Kocherlakota (1992), S.108).

Dies spielt zwar zunächst für die Testdurchführung keine Rolle; man wird aber später eventuell doch mit dem ML-Schätzer weiterarbeiten wollen. Dies ist eine prinzipielle Schwierigkeit bei diagnostischen Tests, wenn sich Momenten- und ML-Schätzer unterscheiden.

Dieses Argument geht aber an der Absicht vorbei, die man mit einem diagnostischen Test verbindet: man untersucht, ob die Daten mit einem bestimmten (einfachen) Verteilungsmodell in wichtigen Kenngrößen (nämlich den ersten Momenten) in Einklang stehen; ist dies der Fall, so verwendet man das theoretische Modell, auch wenn die „wahre“ zugrundeliegende Verteilung eine andere ist. Der Begriff der Effizienz ist dagegen nur innerhalb des parametrischen Modells sinnvoll.

### 1.5.2 Beispiel

Werden als Schätzer für die drei Parameter der bivariaten Poissonverteilung die Momentenschätzer verwendet, so kann man einen Test konstruieren, der Abweichungen in den Varianzen der Randverteilungen vom hypothetischen Modell aufdeckt. Die zugehörige Testgröße umfaßt alle nichtverschwindenden Komponenten vom Grade kleiner oder gleich zwei. Als Polynome kann man etwa  $h_1(x, y; \vartheta) = (x - \lambda_1^*)^2$  und  $h_2(x, y; \vartheta) = (y - \lambda_2^*)^2$  verwenden. Die erste zentrierte Komponente hat dann wegen  $Eh_1(X, Y; \vartheta) = Var(X) = \lambda_1^*$  die Form

$$\begin{aligned}\hat{U}_{n,1} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( (X_i - \hat{\lambda}_1^*)^2 - \hat{\lambda}_1^* \right) \\ &= \sqrt{n}(S_{XX} - \hat{\lambda}_1^*); \end{aligned}$$

analog folgt für die zweite Komponente  $\hat{U}_{n,2} = \sqrt{n}(S_{YY} - \hat{\lambda}_2^*)$ .

Es läßt sich leicht zeigen, daß wieder (R1) erfüllt ist mit  $l_1(x, y; \vartheta) = x - \lambda_1^*$ ,  $l_2(x, y; \vartheta) = y - \lambda_2^*$  und  $l_3(x, y; \vartheta) = xy - \lambda_3 - (x - \lambda_1^*)\lambda_2^* - (y - \lambda_2^*)\lambda_1^*$ . Wegen  $E[\partial h_1 / \partial \lambda_1^*] = -1$  und  $E[\partial h_1 / \partial \lambda_2^*] = E[\partial h_1 / \partial \lambda_3] = 0$  ergibt sich für das erste Element der Kovarianzmatrix

$$\begin{aligned}\sigma_{11} &= E \left[ \left( (X - \lambda_1^*)^2 - \lambda_1^* - (X - \lambda_1^*) \right)^2 \right] \\ &= \mu_{4,0} - 2\mu_{3,0} - \lambda_1^{*2} + \lambda_1^*. \end{aligned}$$



Auf diese Weise erhält man die Kovarianzmatrix

$$\Sigma = \begin{pmatrix} \mu_{4,0} - 2\mu_{3,0} - \lambda_1^{*2} + \lambda_1^* & \mu_{2,2} - \mu_{2,1} - \mu_{1,2} + \lambda_3 - \lambda_1^*\lambda_2^* \\ * & \mu_{0,4} - 2\mu_{0,3} - \lambda_2^{*2} + \lambda_2^* \end{pmatrix}.$$

Ersetzt man in  $\Sigma$  die auftretenden höheren Momente durch die entsprechenden empirischen Größen und bildet  $\hat{U}'_n \hat{\Sigma}^{-1} \hat{U}_n$ , so ergibt sich die Aussage des folgenden Satzes.

### 1.5.3 Satz

Die Testgröße  $\hat{\Psi}_n^2$ , die durch die Gleichung

$$\begin{aligned} & \frac{1}{n} \left\{ \left( \widehat{\mu}_{4,0} - 2\widehat{\mu}_{3,0} - (\hat{\lambda}_1^*)^2 + \hat{\lambda}_1^* \right) \left( \widehat{\mu}_{0,4} - 2\widehat{\mu}_{0,3} - (\hat{\lambda}_2^*)^2 + \hat{\lambda}_2^* \right) \right. \\ & \left. - \left( \widehat{\mu}_{2,2} - \widehat{\mu}_{2,1} - \widehat{\mu}_{1,2} + S_{XY} - \hat{\lambda}_1^*\hat{\lambda}_2^* \right)^2 \right\} \hat{\Psi}_n^2 \\ & = \left( \widehat{\mu}_{0,4} - 2\widehat{\mu}_{0,3} - (\hat{\lambda}_2^*)^2 + \hat{\lambda}_2^* \right) \left( S_{XX} - \hat{\lambda}_1^* \right)^2 \\ & \quad - 2 \left( \widehat{\mu}_{2,2} - \widehat{\mu}_{2,1} - \widehat{\mu}_{1,2} + S_{XY} - \hat{\lambda}_1^*\hat{\lambda}_2^* \right) \left( S_{XX} - \hat{\lambda}_1^* \right) \left( S_{YY} - \hat{\lambda}_2^* \right) \\ & \quad + \left( \widehat{\mu}_{4,0} - 2\widehat{\mu}_{3,0} - (\hat{\lambda}_1^*)^2 + \hat{\lambda}_1^* \right) \left( S_{YY} - \hat{\lambda}_2^* \right)^2 \end{aligned}$$

definiert wird, ist asymptotisch  $\chi_2^2$ -verteilt in der Klasse aller bivariaten diskreten Verteilungen, bei denen Erwartungswert und Varianz der Randverteilungen jeweils übereinstimmen.

**Bemerkung:** Verwendet man die oben angegebenen Momente der bivariaten Poissonverteilung, so vereinfacht sich die Kovarianzmatrix zu

$$\Sigma_0 = \begin{pmatrix} \mu_{4,0} - \lambda_1^{*2} - \lambda_1^* & \mu_{2,2} - \lambda_1^*\lambda_2^* - \lambda_3 \\ * & \mu_{0,4} - \lambda_2^{*2} - \lambda_2^* \end{pmatrix}.$$

Dies ist die von Rayner und Best (1995), Abschnitt 4, angegebene Kovarianzmatrix. Die entsprechende Testgröße ist wieder asymptotisch  $\chi_2^2$ -verteilt, wenn eine bivariate Poisson-Verteilung zugrunde liegt; in der oben angegebenen nichtparametrischen Verteilungsklasse ist dies nicht mehr der Fall.

Die Konstruktion von Tests, die Komponenten höherer Ordnung verwenden, verläuft entsprechend; dies soll hier nicht mehr weiter ausgeführt werden. In den nächsten Abschnitten wird gezeigt, wie man diese Tests ohne Berechnung der Kovarianzmatrix durchführen kann.

## 1.6 Anwendung der diagnostischen Tests mit Hilfe von Resampling-Verfahren

Wie in Satz 1.1.1 gezeigt wurde, sind die Komponenten des glatten Anpassungstests  $\Psi_{n,k}^2$  asymptotisch normalverteilt in der gesamten Verteilungsklasse  $\mathcal{P}_0$  der Verteilungen, für die die Erwartungswerte der ersten  $k$  Komponenten verschwinden. Dabei ist die Kovarianzmatrix oft sehr einfach gebaut, so daß sich ein asymptotisch verteilungsfreier Test konstruieren läßt, wie in Satz 1.1.3 gezeigt wurde.

Im allgemeinen ist die Kovarianzmatrix allerdings komplizierter aufgebaut und hängt vom verwendeten Schätzer ab, wie man an Teil a) von Satz 1.1.1 sieht. Dasselbe gilt auch, wenn man als Testgröße eine einzelne (höhere) Komponente verwendet; siehe dazu Henze und Klar (1996), S.69. In diesen Situationen ist es schon recht aufwendig, analog zu Satz 1.1.3 eine asymptotisch verteilungsfreie Testgröße zu konstruieren. Noch schlimmer sieht die Situation im Falle der multivariaten Normalverteilung aus: die Kovarianzmatrix läßt sich zwar theoretisch leicht angeben; wegen der Vielzahl der Polynome, aus denen die Komponenten aufgebaut sind, ist sie in der Praxis jedoch nicht zu berechnen. Es stehen lediglich die in den vorangegangenen Kapiteln angegebenen Berechnungsformeln für die Komponenten zur Verfügung.

Außerdem konvergiert die Verteilung der Testgrößen teilweise recht langsam gegen die jeweilige Grenzverteilung, so daß die Niveaueinhaltung für kleine und mittlere Stichprobenumfänge schlecht ist.

In diesem Abschnitt soll nun untersucht werden, wie mit Hilfe von verschiedenen Resampling-Verfahren die diagnostischen Tests ausgewertet wer-

den können und ob sich die angesprochenen Probleme durch diese Verfahren lösen lassen.

Allen Resampling-Verfahren ist gemeinsam, daß mit Hilfe der vorliegenden Stichprobe eine Approximation der Verteilung der interessierenden Größe  $T_n$  berechnet wird. Dabei werden aus der ursprünglichen Stichprobe weitere Stichproben nach gewissen Vorschriften erzeugt. Genauer ist das Vorgehen wie folgt:

Man zieht eine Resampling-Stichprobe  $X_1^*, \dots, X_m^*$  vom Umfang  $m$  aus der ursprünglichen Stichprobe  $X_1, \dots, X_n$ . Je nach Verfahren wird dabei mit oder ohne Zurücklegen gezogen. Daraus wird  $T_n^* = T_m(X_1^*, \dots, X_m^*)$  berechnet. Die geeignet normierte (bedingte) Verteilung von  $T_n^*$  bei gegebenen  $X_1, \dots, X_n$  ist dann in vielen Fällen eine gute Approximation für die Verteilung von  $T_n$ .

Es besteht allerdings das Problem, daß bei vielen Resampling-Verfahren auch die Verteilung von  $T_n^*$  nicht exakt bestimmt werden kann. Hier kann man sich aber mit einer Monte Carlo Simulation behelfen: man wiederholt die obige Ziehung sehr oft und berechnet jeweils  $T_n^*$ ; die empirische Verteilung der  $T_n^*$  ist dann eine gute Approximation der theoretischen Verteilung.

Die gebräuchlichsten Resampling-Verfahren sind die „klassischen“ Bootstrap- bzw. Jackknife-Verfahren. Beim Bootstrapverfahren wird  $n$ -mal mit Zurücklegen gezogen. Dies bedeutet, daß die Bootstrap-Stichprobe  $X_1^*, \dots, X_n^*$  vom Umfang  $n$  aus  $\hat{F}_n$ , der empirischen Verteilungsfunktion von  $X_1, \dots, X_n$ , gezogen wird. Beim Bootstrap-Verfahren ist im allgemeinen eine Monte Carlo Simulation erforderlich.

Beim klassischen Jackknife-Verfahren wird dagegen ohne Zurücklegen gezogen. Dabei ist  $m = n - 1$ , d.h. zur Berechnung von  $T_n^*$  werden alle bis auf einen der Werte  $X_1, \dots, X_n$  verwendet. Die (bedingte) Verteilung von  $T_n^*$  kann leicht berechnet werden: da mit gleicher Wahrscheinlichkeit  $1/n$  jeder Wert  $X_i$  weggelassen wird, ist es eine Gleichverteilung auf  $\{T_n^1, T_n^2, \dots, T_n^n\}$ , wobei  $T_n^i = T_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$  ist.

### 1.6.1 Das Bootstrapverfahren

Für  $P \in \mathcal{P} := \{P : E_P[h_j^2(X; \delta(P))] < \infty\}$  sei im folgenden  $\gamma = \gamma(P) = T_j(P) = \int h_j(\cdot; \delta(P)) dP$  (siehe (1.4)) und  $\hat{\gamma}_n = \gamma(\hat{P}_n)$ , wobei  $\hat{P}_n$  die empirische Verteilung bezeichnet. Überlegungen wie in Abschnitt 1.1 zeigen, daß  $\sqrt{n}(\hat{\gamma}_n - \gamma)$  in Verteilung gegen eine Normalverteilung  $\mathcal{N}(0, \sigma^2)$  konvergiert. Ist  $\hat{\sigma}_n^2$  ein konsistenter Schätzer von  $\sigma^2$ , so ist folglich  $Z_n = \sqrt{n}(\hat{\gamma}_n - \gamma)/\hat{\sigma}_n$  asymptotisch standardnormalverteilt.

Nun sei  $\mathcal{X}^* = (X_1^*, \dots, X_n^*)$  eine Bootstrap-Stichprobe aus  $\mathcal{X} = (X_1, \dots, X_n)$ , d.h. es gilt

$$P(X_i^* = X_j | \mathcal{X}) = \frac{1}{n}, \quad 1 \leq i, j \leq n,$$

und  $X_1^*, \dots, X_n^*$  sind bedingt unabhängig. Die Bootstrapversionen von  $\hat{\gamma}_n$  bzw.  $\hat{\sigma}_n$  werden mit  $\hat{\gamma}_n^*$  bzw.  $\hat{\sigma}_n^*$  bezeichnet. Aus der Tatsache, daß  $Z_n$  asymptotisch normalverteilt ist, folgt i. allg. nicht die (fast sichere) asymptotische Normalverteilung der entsprechenden Bootstrapgröße  $Z_n^* = \sqrt{n}(\hat{\gamma}_n^* - \hat{\gamma}_n)/\hat{\sigma}_n^*$ . Dies gilt zwar für lineare Funktionale, aber schon nicht mehr für quadratische Funktionale wie z.B. U-Statistiken (siehe Mammen (1992)). Im folgenden setzen wir deshalb voraus, daß das Funktional  $\gamma(P)$  Fréchet-differenzierbar ist (für eine Definition siehe etwa Serfling (1980), S.217). Dies erfordert eine gewisse Glattheit von  $\gamma$  in Abhängigkeit von  $P$ , die beispielsweise bei den von Henze und Klar (1996) behandelten Beispielen immer gegeben ist. Unter dieser Annahme konvergiert die bedingte Verteilung von  $Z_n^*$  gegen die asymptotische Verteilung von  $Z_n$ , hier also gegen die Standardnormalverteilung (Arcones und Giné, (1992)).

Zu testen ist nun  $\tilde{H}_0 : \gamma = 0$  gegen  $\tilde{H}_1 : \gamma \neq 0$ . Definiert man ähnlich wie früher  $\mathcal{P}_0 = \{P \in \mathcal{P} : \gamma(P) = 0\}$ , so lautet das Testproblem  $\tilde{H}_0 : P \in \mathcal{P}_0$  gegen  $\tilde{H}_1 : P \in \mathcal{P} \setminus \mathcal{P}_0$ .

Sinnvoll ist hier ein zweiseitiger Test mit symmetrischem Ablehnungsbereich, d.h. die Hypothese wird verworfen, falls  $\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n$  „zu groß“ ist. Definiert man

$$w_\alpha = \inf\{w : P(\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n \leq w) \geq 1 - \alpha\}$$

und lehnt  $\tilde{H}_0$  ab, falls  $\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n > w_\alpha$ , so erhält man einen Test von  $\tilde{H}_0$  gegen  $\tilde{H}_1$  mit der Eigenschaft  $\lim_{n \rightarrow \infty} P(\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n > w_\alpha) = \alpha$  für jedes  $P \in \mathcal{P}_0$  (wenn die zugrundeliegende Verteilung stetig ist). Bezeichnet  $z_\xi$  das  $\xi$ -Quantil der Standardnormalverteilung, so strebt  $w_\alpha$  für  $n \rightarrow \infty$  gegen  $z_{1-\alpha/2}$ . Da man  $w_\alpha$  nicht berechnen kann, ersetzt man es beim „klassischen“ Test durch  $z_{1-\alpha/2}$ ; der so entstehende Test hält dann zumindest asymptotisch das Niveau  $\alpha$  ein.

Die Idee beim Bootstrap-Test ist dagegen,  $w_\alpha$  durch die Bootstrap-Größe

$$w_\alpha^* = \inf\{w : P(|Z_n^*| \leq w | \mathcal{X}) \geq 1 - \alpha\} \quad (1.33)$$

zu ersetzen und  $H_0$  abzulehnen, falls  $\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n > w_\alpha^*$  ist.

Der nächste Satz zeigt die Gültigkeit des so definierten Bootstrap-Tests; der Test ist außerdem konsistent gegen jede nicht zur Klasse  $\mathcal{P}_0$  gehörende Alternativverteilung und erkennt benachbarte Alternativen mit einer Wahrscheinlichkeit, die größer als  $\alpha$  ist.

### 1.6.1 Satz

Unter den obigen Voraussetzungen an  $\gamma(P)$  gilt:

- a)  $\lim P(\sqrt{n}|\hat{\gamma}_n| \leq \hat{\sigma}_n w_\alpha^*) = 1 - \alpha$ ,  $n \rightarrow \infty$ ,  $\forall P \in \mathcal{P}_0$ .
- b)  $\lim P(\sqrt{n}|\hat{\gamma}_n| > \hat{\sigma}_n w_\alpha^*) = 1$ ,  $n \rightarrow \infty$ ,  $\forall P \in \mathcal{P} \setminus \mathcal{P}_0$ .
- c) Sei  $(P_n)_{n \geq 1}$  eine Folge von Verteilungen mit  $\gamma(P_n) = n^{-1/2}c$ ,  $n \geq 1$ ,  $c \neq 0$ . Dann ist

$$\pi_\infty(c) := \lim_{n \rightarrow \infty} \pi_n(c) := \lim_{n \rightarrow \infty} P_n(\sqrt{n}|\hat{\gamma}_n| > \hat{\sigma}_n w_\alpha^*) > \alpha,$$

und es gilt  $\lim_{c \rightarrow \infty} \pi_\infty(c) = 1$ .

BEWEIS: Im folgenden sei  $\gamma_0 := \gamma(P)$  der „wahre“ zugrundeliegende Parameter. Aus  $\mathcal{V}(Z_n^* | \mathcal{X}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$  (fast sicher) folgt die stochastische Konvergenz  $w_\alpha^* \xrightarrow{P} z_{1-\alpha/2}$  ( $n \rightarrow \infty$ ) und somit  $w_\alpha - w_\alpha^* = o_P(1)$ . Ist  $\gamma_0 = 0$ , so

folgt

$$\begin{aligned} P(\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n \leq w_\alpha^*) &= P(|Z_n| \leq w_\alpha^*) \\ &= P(|Z_n| \leq w_\alpha + o_P(1)) = P(|Z_n| \leq w_\alpha) + o(1) \rightarrow 1 - \alpha \quad (n \rightarrow \infty), \end{aligned}$$

also Teil a). Gilt  $\gamma_0 \neq 0$ , so folgt aus  $\sqrt{n}\hat{\gamma}_n/\hat{\sigma}_n = Z_n + \sqrt{n}\gamma_0/\hat{\sigma}_n$  die Aussage von Teil b):

$$P(\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n > w_\alpha^*) = P(|Z_n + \sqrt{n}\gamma_0| > w_\alpha) + o(1) \rightarrow 1$$

für  $n \rightarrow \infty$ . Ganz analog erhält man für  $\gamma(P_n) = cn^{-1/2}$

$$\begin{aligned} \pi_\infty(c) &= \lim_{n \rightarrow \infty} P_n(\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n > w_\alpha^*) \\ &= \lim_{n \rightarrow \infty} P_n(|Z_n + \frac{c}{\hat{\sigma}_n}| > w_\alpha^*) \\ &= P(|Z + \frac{c}{\sigma}| > z_{1-\frac{\alpha}{2}}) > \alpha; \end{aligned}$$

hierbei ist  $Z$  eine standardnormalverteilte Zufallsvariable. Aus der letzten Gleichheit folgt auch die zweite Aussage von Teil c). ■

Bei der Berechnung von  $w_\alpha^*$  in (1.33) wurde  $\hat{\gamma}_n^*$  an  $\hat{\gamma}_n$  zentriert. Alternativ könnte man stattdessen den hypothetischen Wert 0 verwenden, d.h. man berechnet  $\tilde{w}_\alpha^*$  so, daß

$$\tilde{w}_\alpha^* = \inf \{w : P(\sqrt{n}|\hat{\gamma}_n^*|/\hat{\sigma}_n^* \leq w | \mathcal{X}) \geq 1 - \alpha\} \quad (1.34)$$

erfüllt ist und lehnt  $\tilde{H}_0$  ab, falls  $\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n > \tilde{w}_\alpha^*$ .

Hall (1992), S.149, schreibt zu diesem Vorgehen: „This technique is occasionally advocated, but it has low power“. Definiert man analog wie oben die asymptotische Güte des Tests bei benachbarten Alternativen als

$$\tilde{\pi}_\infty(c) = \lim_{n \rightarrow \infty} \tilde{\pi}_n(c) = \lim_{n \rightarrow \infty} P_n(\sqrt{n}|\hat{\gamma}_n| > \hat{\sigma}_n \tilde{w}_\alpha^*),$$

so zeigt Hall weiter, daß  $\tilde{\pi}_\infty(c) \rightarrow 0$  für  $c \rightarrow \infty$  gilt. Hall schreibt: „This latter result is clearly unsatisfactory, demonstrating the superiority of bootstrapping  $\hat{\gamma}_n^* - \hat{\gamma}_n$  rather than  $\hat{\gamma}_n^*$ .“

Diese Aussage über die Güte des durch (1.34) definierten Tests klingt äußerst verdächtig. Untersucht man den Test genauer, so zeigt sich, daß er vollkommen unbrauchbar ist. Es gilt nämlich

### 1.6.2 Satz

Sei  $(P_n)_{n \geq 1}$  eine Folge von Verteilungen mit  $\gamma(P_n) = n^{-1/2}c$ ,  $n \geq 1$ ,  $c \in \mathbb{R}$ . Dann gilt

$$\tilde{\pi}_\infty(c) = \lim_{n \rightarrow \infty} P_n(\sqrt{n}|\hat{\gamma}_n| > \hat{\sigma}_n \tilde{w}_\alpha^*) = 0.$$

Insbesondere hat der Test das asymptotische Niveau 0. Die asymptotische Güte des Tests bei festen Alternativen ist ebenfalls 0.

BEWEIS: Im folgenden sei  $\gamma(P_n) = n^{-1/2}c$ , und  $Z$  sei eine standardnormalverteilte Zufallsvariable. Definiert man  $\zeta_\alpha(x)$  durch

$$P(|Z + x| \leq \zeta_\alpha(x)) = 1 - \alpha, \quad (1.35)$$

so erhält man aus der Definition von  $\tilde{w}_\alpha^*$  und unter Beachtung von

$$\sqrt{n}\hat{\gamma}_n^*/\hat{\sigma}_n^* = Z_n^* + \frac{\hat{\sigma}_n}{\hat{\sigma}_n^*}Z_n + \frac{c}{\hat{\sigma}_n^*}$$

die Beziehung

$$w_\alpha^* = \zeta_\alpha\left(Z_n + \frac{c}{\sigma}\right) + o_P(1) \quad (1.36)$$

(vgl. Hall (1992), S.149). Wegen  $\sqrt{n}|\hat{\gamma}_n - \gamma_1|/\hat{\sigma}_n = |Z_n + c/\hat{\sigma}_n|$  folgt daraus

$$\tilde{\pi}_\infty(c) = P\left(\left|Z + \frac{c}{\sigma}\right| > \zeta_\alpha\left(Z + \frac{c}{\sigma}\right)\right). \quad (1.37)$$

Wir zeigen nun

$$\zeta_\alpha(x) > |x| + z_{1-\frac{\alpha}{2}} \quad \forall x \in \mathbb{R}. \quad (1.38)$$

Für  $x \geq 0$  gilt nämlich

$$\begin{aligned} P(|Z + x| \leq x + z_{1-\frac{\alpha}{2}}) &= P(-2x - z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) \\ &= \alpha - P(Z \leq -2x - z_{1-\frac{\alpha}{2}}) < 1 - \alpha, \end{aligned}$$

woraus zusammen mit der Definition von  $\zeta_\alpha$  in (1.35)  $x + z_{1-\frac{\alpha}{2}} < \zeta_\alpha(x)$  folgt.

Ist  $x < 0$ , so folgt aus der Gleichungskette

$$1 - \alpha = P(|Z + x| \leq \zeta_\alpha(x)) = P(|-Z + x| \leq \zeta_\alpha(x)) = P(|Z + |x|| \leq \zeta_\alpha(x))$$

zuerst  $\zeta_\alpha(x) = \zeta_\alpha(|x|)$  und daraus aus dem schon Bewiesenen  $\zeta_\alpha(x) > |x| + z_{1-\frac{\alpha}{2}}$ .

Für die asymptotische Güte gegen benachbarte Alternativen erhält man nun aus (1.37) und (1.38)

$$\tilde{\pi}_\infty(c) \leq P\left(\left|Z + \frac{c}{\sigma}\right| > \left|Z + \frac{c}{\sigma}\right| + z_{1-\frac{\alpha}{2}}\right) = 0 \quad \forall c \in \mathbb{R}.$$

Die Aussage über feste Alternativen erhält man ganz analog. ■

Eine andere Modifikation des Bootstrap-Tests führt dagegen auf einen weiteren sinnvollen Test: anstelle der „studentisierten“ Größe  $\sqrt{n}\hat{\gamma}_n / \hat{\sigma}_n$  kann die Testgröße  $\sqrt{n}\hat{\gamma}_n$  verwendet werden; modifiziert man die Definition von  $w_\alpha^*$  entsprechend, so gelten für diesen Test alle Aussagen von Satz 1.6.1.

Diesen Test kann man verwenden, wenn die Kovarianz kompliziert zu berechnen ist. Der Nachteil dieses Tests liegt in der im allgemeinen geringeren Konvergenzgeschwindigkeit im Vergleich zum ursprünglichen Bootstrap-Test. Unter bestimmten Regularitätsannahmen gilt nämlich für den studentisierten Test, daß dessen tatsächliches Niveau vom theoretischen Niveau  $1 - \alpha$  um einen Fehler der Größenordnung  $O(n^{-2})$  abweicht. Beim nicht-studentisierten Test ist der Approximationsfehler dagegen wie beim klassischen mit den Normalverteilungsquantilen arbeitenden Test von der Größenordnung  $O(n^{-1})$ . Näheres dazu findet man in Hall (1992) sowie in Mammen (1992), Kap. 5.

Wie in der Einleitung erwähnt, kann man  $\mathcal{V}(Z_n^*|\mathcal{X})$  analytisch nicht berechnen, aber durch eine einfache Monte Carlo Simulation beliebig genau approximieren. Die praktische Durchführung des in (1.33) definierten Bootstrap-Tests sieht wie folgt aus; dabei wurden wegen der einfacheren Durchführung die Testgrößen quadriert.



1. Berechne die Testgröße  $T_n = n \hat{\gamma}_n^2 / \hat{\sigma}_n^2$ .
2. Ziehe Bootstrap-Stichproben  $X_{j1}^*, \dots, X_{jn}^*$ ,  $j = 1, \dots, B$ , vom Umfang  $n$  aus  $\hat{F}_n$ .
3. Berechne  $T_{j,n}^* = T_{j,n}^*(X_{j1}^*, \dots, X_{jn}^*) = n(\hat{\gamma}_n^* - \hat{\gamma}_n)^2 / (\hat{\sigma}_n^*)^2$ ,  $j = 1, \dots, B$ .
4. Die Hypothese  $\tilde{H}_0$  wird zum Niveau  $\alpha$  abgelehnt, falls  $T_n$  größer als das empirische  $(1 - \alpha)$ -Quantil  $c_{n,\alpha}^*$  von  $(T_{j,n}^*)_{1 \leq j \leq B}$  ist; bezeichnet  $[r]$  die Gaußklammer von  $r$  und  $T_{1:B}^* \leq \dots \leq T_{B:B}^*$  die Ordnungsstatistik von  $T_{1,n}^*, \dots, T_{B,n}^*$ , so ist  $c_{n,\alpha}^*$  gegeben durch

$$c_{n,\alpha}^* = \begin{cases} T_{[B(1-\alpha)]:B}^*, & \text{falls } B(1-\alpha) \text{ ganzzahlig} \\ T_{[B(1-\alpha)]+1:B}^*, & \text{sonst.} \end{cases}$$

### Bemerkungen:

1. Hat die Testgröße  $\hat{\gamma}_n$  eine stetige Verteilung, so ist  $w_\alpha$  die Lösung der Gleichung  $P(\sqrt{n}|\hat{\gamma}_n|/\hat{\sigma}_n \leq w) = 1 - \alpha$ . Für  $w_\alpha^*$  muß man stets die Definition (1.33) verwenden, da  $\mathcal{V}(Z_n^*|\mathcal{X})$  immer eine diskrete Verteilung ist.
2. Die Bootstrap-Testgrößen  $T_{j,n}^*$  approximieren die Verteilung von  $T_n$  in ganz  $\mathcal{P}_0$ . Ein Bootstrap-Test für die parametrische Hypothese  $H_0$  müßte dagegen auf einem parametrischen Bootstrapverfahren aufbauen; dabei werden die Bootstrap-Stichproben  $X_{j1}^*, \dots, X_{jn}^*$  nicht aus  $\hat{F}_n$ , sondern aus der hypothetischen Verteilung mit geschätzten Parametern  $F(\cdot, \hat{\gamma}_n)$  gezogen. Ist der Schätzer  $\hat{\gamma}_n$  konsistent und  $\gamma(P)$  differenzierbar, so liefert dieses Verfahren asymptotisch einen Niveau  $\alpha$ -Test für  $H_0$  (Arcones und Giné, (1992), Theorem 4.4).

### 1.6.2 Ein modifiziertes Jackknife-Verfahren

In diesem Abschnitt sollen verallgemeinerte Jackknife-Verfahren betrachtet werden. Bei diesen Verfahren wird aus der ursprünglichen Stichprobe

$\mathcal{X}$  eine Stichprobe vom Umfang  $r (< n)$  ohne Zurücklegen gezogen; anders ausgedrückt werden also  $d = n - r$  Werte der ursprünglichen Stichprobe gestrichen. Beim klassischen Jackknife-Verfahren ist  $r = n - 1$ , also  $d = 1$ ; es wird deshalb auch oft „delete-one Jackknife“ genannt. Dieses Verfahren wurde oft mit Erfolg zur Verringerung des Bias und zur Varianzschätzung angewandt; es ist jedoch, im Gegensatz zum Bootstrapverfahren, nicht in der Lage, die gesamte Verteilung eines Schätzers zu approximieren. Aus diesem Grunde wurden die verallgemeinerten Jackknife-Verfahren vorgeschlagen, die diesen Nachteil nicht besitzen (Wu (1986, 1990)).

Es sei zunächst dieselbe Situation wie im letzten Abschnitt gegeben, insbesondere gelte also

$$\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

Beim (verallgemeinerten) Jackknife-Verfahren wird jede Teilmenge  $\mathcal{Y}_i$ ,  $i = 1, \dots, \binom{n}{r}$  der Größe  $r$  von  $\mathcal{X}$  mit gleicher Wahrscheinlichkeit  $\binom{n}{r}^{-1}$  ausgewählt. Für jede Teilmenge  $\mathcal{Y}_i$  kann nun die Statistik  $\hat{\gamma}_{n,i}^* = \hat{\gamma}_r(\mathcal{Y}_i)$  berechnet werden. Mit der Bezeichnung

$$L_n^*(x) = P_* \left( \sqrt{\frac{nr}{d}} \frac{\hat{\gamma}_{n,i}^* - \hat{\gamma}_n}{\hat{\sigma}} \leq x \right),$$

wobei  $P_*$  bedeutet, daß die Wahrscheinlichkeit unter dem oben beschriebenen Jackknife-Mechanismus berechnet werden soll, und  $\hat{\sigma} = \hat{\sigma}(X_1, \dots, X_n)$  ein beliebiger (schwach) konsistenter Schätzer von  $\sigma$  ist, gilt folgendes Ergebnis:

### 1.6.3 Satz

Die Voraussetzungen von Satz 1.1.1 seien erfüllt. Geht  $r \rightarrow \infty$  und ist  $r/n \leq p$  mit  $0 < p < 1$ , so gilt

$$\sup_x |L_n^*(x) - \Phi(x)| \xrightarrow{P} 0.$$

BEWEIS: Unter den Voraussetzungen von Satz 1.1.1 ist  $\hat{\gamma}_n$  asymptotisch linear, wie man am Beweis sieht; d.h. die Bedingungen (3.1) und (3.2) in

Wu (1990) sind erfüllt. Die Aussage des Satzes ist dann gerade Teil (iii) von Theorem 2 in Wu (1990). ■

Aus diesem Ergebnis folgen wieder alle Aussagen aus Satz 1.6.1 (siehe auch Beran (1984), Theorem 1). Man beachte, daß für die Gültigkeit des Jackknife-Verfahrens nur die schwache Voraussetzung der asymptotischen Linearität benötigt wird, aber keinerlei Differenzierbarkeitsbedingungen. Allerdings hat das Verfahren gegenüber dem Bootstrapverfahren auch zwei schwerwiegende Nachteile. Zum einen besitzt es i. allg. keine so hohe Konvergenzordnung wie das Bootstrapverfahren, zum anderen ist wenig darüber bekannt, wie  $r = r(n)$  im Einzelfall zu wählen ist.

Das Hauptziel dieses Abschnitts, die Durchführung von diagnostischen Tests für die Komponenten des glatten Tests für die multivariate Normalverteilung, ist mit diesem Ergebnis aber noch nicht erreicht, da man nicht die einzelnen (nach Standardisierung asymptotisch normalverteilten) Polynome verwenden will. Allein die Komponenten  $\hat{U}_{n,k}^2$  stehen zur Verfügung. Hier hilft eine Modifikation des verallgemeinerten Jackknife-Verfahrens, die von Politis und Romano (1994) stammt.

Im folgenden wird die Testgröße  $T_n = \hat{U}_{n,k}^2/n$  verwendet. Ist  $\tau = 0$  (siehe Abschnitt 1.3), so konvergiert  $n T_n$  schwach gegen eine stetige Grenzverteilung, nämlich gegen eine gewichtete Summe von  $\chi^2$ -Zufallsvariablen (Existenz der Momente vorausgesetzt).  $\mathcal{P}_0$  sei die Teilmenge von Verteilungen aus  $\mathcal{P}$ , für die  $\tau = 0$  gilt. Mit  $S_{n,i}$  wird die der Statistik  $T_n$  entsprechende Jackknife-Größe bezeichnet; ist also  $\mathcal{Y}_i$ ,  $i = 1, \dots, N_n$ , wie oben eine Teilmenge der Größe  $r$  von  $\mathcal{X}$ , so gilt  $S_{n,i} = T_r(\mathcal{Y}_i)$  (hierbei ist  $N_n = \binom{n}{r}$ ). Die bedingte Verteilung  $L_n(\cdot)$  der Jackknife-Größe  $r(S_{n,i} - T_n)$  kann im Gegensatz zum Bootstrap-Verfahren direkt angegeben werden; es gilt

$$L_n(x) = N_n^{-1} \sum_{i=1}^{N_n} 1\{r(S_{n,i} - T_n) \leq x\}.$$

Weiter sei im folgenden  $c_n(1 - \alpha)$  das obere  $\alpha$ -Quantil von  $L_n$ , also

$$c_n(1 - \alpha) = \inf\{x : L_n(x) \geq 1 - \alpha\}.$$

Dann gilt folgender Satz.

#### 1.6.4 Satz

Für  $n \rightarrow \infty$  gelte  $r \rightarrow \infty$  und  $r/n \rightarrow 0$ . Dann gilt

$$a) \lim P(nT_n \leq c_n(1 - \alpha)) = 1 - \alpha, \quad n \rightarrow \infty, \quad \forall P \in \mathcal{P}_0.$$

$$b) \lim P(nT_n > c_n(1 - \alpha)) = 1, \quad n \rightarrow \infty, \quad \forall P \in \mathcal{P} \setminus \mathcal{P}_0.$$

BEWEIS: a) Da die Grenzverteilung von  $nT_n$  stetig ist, folgt Teil a) aus Politis und Romano (1994), Theorem 2.1(iii).

b) Sei  $P \in \mathcal{P} \setminus \mathcal{P}_0$ , insbesondere also  $\tau \neq 0$ . In diesem Fall konvergiert  $\sqrt{n}(T_n - \tau)$  schwach gegen eine zentrierte Normalverteilung. Definiert man

$$\tilde{L}_n(x) = N_n^{-1} \sum_{i=1}^{N_n} 1\{\sqrt{r}(S_{n,i} - T_n) \leq x\}$$

und  $\tilde{c}_n(1 - \alpha) = \inf\{x : \tilde{L}_n(x) \geq 1 - \alpha\}$ , so folgt wie in Teil a)

$$P(\sqrt{n}(T_n - \tau) \leq \tilde{c}_n(1 - \alpha)) \rightarrow 1 - \alpha \quad (n \rightarrow \infty). \quad (1.39)$$

Nun gilt

$$\begin{aligned} c_n(1 - \alpha) &= \inf\{x : L_n(x) \geq 1 - \alpha\} = \inf\{x : \tilde{L}_n(x/\sqrt{r}) \geq 1 - \alpha\} \\ &= \sqrt{r} \inf\{y : \tilde{L}_n(y) \geq 1 - \alpha\} = \sqrt{r} \tilde{c}_n(1 - \alpha). \end{aligned}$$

Da  $\tilde{c}_n(1 - \alpha)$  nach (1.39) stochastisch gegen das  $(1 - \alpha)$ -Quantil der asymptotischen Verteilung von  $\sqrt{n}(T_n - \tau)$  konvergiert, folgt für  $n \rightarrow \infty$

$$P(nT_n \leq c_n(1 - \alpha)) = P(\sqrt{n}T_n \leq \sqrt{r/n} \tilde{c}_n(1 - \alpha)) \rightarrow 0.$$

■

#### Bemerkungen:

1. Das Entscheidende am Jackknife-Verfahren von Romano und Politis (1994) ist, daß es ohne Glattheitsvoraussetzungen auskommt; dadurch ist es immer anwendbar, wenn die Teststatistik gegen eine (beliebige) Grenzverteilung konvergiert.

- 
2. Dies erfordert aber i. allg., daß das Verhältnis der Normierungskonstanten, hier also  $r$  zu  $n$  bzw.  $\sqrt{r}$  zu  $\sqrt{n}$ , gegen 0 geht (siehe etwa Example 2.1.2 in der zitierten Arbeit). Im vorliegenden Fall ist nicht auszuschließen, daß das Verfahren auch für  $r/n \rightarrow p$ ,  $0 < p < 1$  funktioniert.
  3. Da  $\binom{n}{r}$  sehr groß sein kann, ist  $L_n(x)$  eventuell nur mit sehr großem Aufwand zu berechnen. Wie beim Bootstrap-Verfahren kann man  $L_n(x)$  mit Hilfe einer Monte-Carlo-Simulation approximieren (siehe auch Abschnitt 2.2 in Politis und Romano (1994)).



## Kapitel 2

# Auf der empirischen integrierten Verteilungsfunktion basierende Tests

Anpassungstests, die auf dem empirischen Prozeß  $Z_n = \sqrt{n}(F_n(\cdot) - F(\cdot))$  aufbauen, sind neben dem Chi-Quadrat-Test die am weitesten verbreiteten Anpassungstests, insbesondere für stetige Verteilungen.

Bekannt sind vor allem der 1928 von Cramér eingeführte Cramér-von Mises-Test und der 1933 von Kolmogorov vorgeschlagene Kolmogorov-Smirnov-Test. Auch der 1952 von Anderson und Darling eingeführte Anpassungstest wird vor allem als Test auf Normalverteilung häufig verwendet.

Allen drei zugehörigen Teststatistiken ist gemeinsam, daß ihre asymptotische Verteilung unter einer einfachen Nullhypothese nicht von der zugrundeliegenden Verteilung abhängt, sofern diese stetig ist. Die Verteilungsfreiheit bleibt bei zusammengesetzten Hypothesen im Falle von Lokations-Skalen-Parameter-Familien erhalten, wenn Parameter äquivariant geschätzt

werden; handelt es sich bei den geschätzten Parametern um Formparameter, geht die Eigenschaft der Verteilungsfreiheit verloren.

Die drei Tests unterscheiden sich in ihrem Güteverhalten deutlich voneinander, wobei der Kolmogorov-Smirnov-Test meistens schlechter abschneidet. Dennoch wird er in der Praxis viel benutzt; so ist er z.B. in der NAG-Unterprogrammbibliothek der einzige Test für stetige Verteilungen neben dem Shapiro-Wilk-Test auf Normalverteilung.

Ist die zugrundeliegende Verteilung diskret, so sind die drei Testgrößen nicht asymptotisch verteilungsfrei, auch nicht im Falle einer einfachen Nullhypothese. Aus diesem Grund kann man die asymptotische Verteilung nicht analytisch berechnen, und es sind umfangreiche Tafeln simulierter Quantile für eine konventionelle Testdurchführung notwendig.

Seit allgemein Rechner verfügbar sind, sind Anpassungstests, die auf dem empirischen Prozeß aufbauen, auch bei diskreten Verteilungen wieder interessant, da die Grenzverteilung durch ein parametrisches Bootstrapverfahren approximiert werden kann; für die Kolmogorov-Smirnov- und die Cramér-von Mises-Testgröße wurde dies in Henze (1996) durchgeführt.

Im folgenden wird ein neuer Test vorgestellt, der auf der Differenz zwischen der empirischen integrierten Verteilungsfunktion und der hypothetischen integrierten Verteilungsfunktion  $\Psi(t) = \int_t^\infty \bar{F}_X d\lambda$  aufbaut. Dabei bezeichnen  $\bar{F}_X = 1 - F_X$  die survival-Funktion und  $\lambda$  das Lebesgue-Maß auf den Borelmengen des  $\mathbb{R}^1$ . Die Testgröße läßt sich zwar auch mit Hilfe des empirischen Prozesses darstellen, die Verteilungstheorie in Henze (1996) kann aber nicht verwendet werden, da die Testgröße auf dem Raum der Nullfolgen, versehen mit der Supremumsnorm, kein stetiges Funktional ist.

Wegen  $EX = \int_0^\infty P(X > t)\lambda(dt) = \Psi_X(0)$  für positive Zufallsvariablen ist die Existenz des Erwartungswertes offensichtlich notwendig und hinreichend für die Existenz der integrierten Verteilungsfunktion; erfüllt eine  $\mathbb{N}_0$ -wertige Zufallsvariable diese Bedingung, so gilt für den zugehörigen empirischen Prozeß  $Z_n = (Z_{n,k})_{k \geq 0}$

$$\sum_{k=0}^{\infty} |Z_{n,k}| = \sqrt{n} \sum_{k=0}^{\infty} |(1 - F(k)) - (1 - F_n(k))| < \infty.$$



Der geeignete Rahmen ist deshalb der Raum  $\ell_1 = L_1(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0), \delta)$  mit dem Zählmaß  $\delta$ , also der Raum der Folgen  $x = (x_k)_{k \geq 0}$  mit  $\sum_{k \geq 0} |x_k| < \infty$ , versehen mit der üblichen  $\ell_1$ -Norm;  $Z_n$  wird dann als Zufallselement auf  $\ell_1$  aufgefaßt.

Dies wird im nächsten Abschnitt genauer ausgeführt. Dort wird unter schwachen Voraussetzungen die Verteilungskonvergenz der Teststatistik gezeigt; ein weiteres Resultat betrifft die Konsistenz des Tests.

Der darauffolgende Abschnitt behandelt das Verhalten des Tests unter benachbarten Alternativen. Dort wird gezeigt, daß die Güte des Tests unter Alternativen, die mit der Geschwindigkeit  $n^{-1/2}$  gegen die hypothetische Verteilung konvergieren, zwischen dem nominalen Niveau des Tests und 1 liegt.

Entscheidend für die Verwendung eines neuen Tests sind besondere Eigenschaften, die die bisher verwendeten Tests nicht haben, etwa eine hohe Güte gegenüber bestimmten Alternativen. Der Aufbau der vorliegenden Testgröße, bei der die Verteilungsenden stark betont werden, läßt erwarten, daß der neue Test empfindlich gegenüber Differenzen zwischen dem Verteilungsende der hypothetischen Verteilung und dem Verteilungsende der Alternativverteilung ist. Diese Vermutung wird in Abschnitt 2.1 durch mehrere Simulationsstudien untermauert. Dabei zeigt sich auch, daß der Test generell im Vergleich zu anderen Tests für diskrete Verteilungen ausgezeichnete Güteeigenschaften besitzt.

Im zweiten Abschnitt werden Anpassungstests, die auf der integrierten Verteilungsfunktion basieren, für die beiden wichtigsten stetigen Modelle, die Exponential- und die Normalverteilung, vorgestellt.

Für die Exponentialverteilung mit Parameter  $\vartheta$  werden Testgrößen der Form

$$T_{n,a}^* = n \int_0^\infty (\Psi_n(t) - \Psi_\vartheta(t))^2 \exp(-a \vartheta t) dt,$$

untersucht, wobei  $\Psi_n$  die empirische integrierte Verteilungsfunktion bezeichnet; durch den Parameter  $a$  ( $a \geq 0$ ) in der Gewichtsfunktion läßt sich das Güteverhalten der zugehörigen Tests beeinflussen.

Als Test auf Normalverteilung wird eine ähnliche Integralstatistik betrachtet, nachdem die Definition der integrierten Verteilungsfunktion geeignet angepaßt worden ist.

Für beide stetige Verteilungsfamilien werden mit Hilfe von Simulationsstudien empirische Quantile der Verteilungen der Teststatistiken ermittelt sowie die Güte der Tests gegenüber einigen Alternativen untersucht.

## 2.1 Ein auf der empirischen integrierten Verteilungsfunktion basierender Anpassungstest für diskrete Verteilungen

### 2.1.1 Definition und grundlegende Eigenschaften

Die im folgenden vorgestellte Testgröße beruht auf der integrierten Verteilungsfunktion, die in der dynamischen Optimierung und bei der Untersuchung von stochastischen Ordnungen eine gewisse Rolle spielt (vgl. Hinderer (1980) und Müller (1996) sowie die in diesen Arbeiten zitierte Literatur). Sie ist auch in der Risikotheorie unter dem Namen „stop-loss transform“ bekannt. Die Definitionen unterscheiden sich zum Teil etwas; für positive Zufallsvariablen ist die folgende Definition am geeignetsten.

#### 2.1.1 Definition

- a) Es sei  $X$  eine reelle Zufallsvariable, deren Verteilung auf  $\mathbb{R}_+$  konzentriert sei, und für die  $EX < \infty$  gelte. Dann ist die integrierte Verteilungsfunktion  $\Psi_X : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  definiert durch

$$\Psi_X(t) := E(X - t)^+ = \int_t^\infty \bar{F}_X d\lambda.$$

Für  $\mathbb{N}_0$ -wertige Zufallsvariablen mit Zähldichte  $f(\cdot)$  und endlichem Erwartungswert gilt insbesondere

$$\Psi_X(t) = \sum_{k=[t]+1}^{\infty} (k - t) f(k) = ([t] + 1 - t) \bar{F}_X([t]) + \sum_{k=[t]+1}^{\infty} \bar{F}_X(k),$$

wobei  $[t]$  die größte ganze Zahl kleiner gleich  $t$  bezeichnet.

b)  $X_1, X_2, \dots, X_n$  seien unabhängige und identisch verteilte Zufallsvariablen auf  $\mathbb{R}_+$ . Dann ist die empirische integrierte Verteilungsfunktion (EIVF) gegeben durch

$$\Psi_n(t) = \int_t^\infty \bar{F}_n d\lambda = \frac{1}{n} \sum_{i=1}^n (X_i - t) \mathbf{1}\{X_i > t\}.$$

In Hinderer (1980) wird  $\int_0^t F d\lambda$  als integrierte Verteilungsfunktion bezeichnet; die hier benutzte Definition hat den Vorteil, daß  $\Psi_X$  für positive Zufallsvariablen mit endlichem Erwartungswert beschränkt ist. Offensichtlich gilt  $\Psi_X(0) = EX$  und  $\Psi_n(0) = \bar{X}_n$ .

Ist  $X_{(i-1)} < t \leq X_{(i)}$  ( $1 \leq i \leq n$ ), wobei  $0 = X_{(0)} \leq X_{(1)} \leq \dots \leq X_{(n)}$  die geordnete Stichprobe von  $X_1, \dots, X_n$  bezeichnet, so läßt sich die empirische integrierte Verteilungsfunktion auch schreiben als

$$\Psi_n(t) = \frac{1}{n} \sum_{j=i+1}^n (X_{(j)} - X_{(j-1)})(n - j + 1) + \frac{n - i + 1}{n} (X_{(i)} - t);$$

insbesondere gilt

$$\Psi_n(X_{(i)}) = \frac{1}{n} \sum_{j=i+1}^n (X_{(j)} - X_{(j-1)})(n - j + 1), \quad i = 1, \dots, n.$$

In dieser Form sieht man den Zusammenhang zu der in der Zuverlässigkeitstheorie oft verwendeten „total time on test“, definiert durch

$$t_n \left( \frac{i}{n} \right) = \frac{1}{n} \sum_{j=1}^i (X_{(j)} - X_{(j-1)})(n - j + 1), \quad i = 1, \dots, n;$$

es gilt

$$\Psi_n(X_{(i)}) = \bar{X}_n - t_n(i/n).$$

Zwischen der theoretischen „total time on test transform“  $t(p) = \int_0^{F^{-1}(p)} \bar{F}(s) ds$  und der integrierten Verteilungsfunktion besteht die entsprechende

Beziehung  $t(p) + \Psi(F^{-1}(p)) = \mu$  ( $0 < p < 1$ ). Eine Übersicht über die Zusammenhänge zwischen diesen beiden und weiteren verwandten Größen wie der Lorenz-Kurve findet man in Chandra und Singpurwalla (1982) und in Heilmann (1985). In Barlow und Campo (1975) werden Eigenschaften der „total time on test transform“ zusammengefaßt und verschiedene auf dieser Größe basierende Anpassungstests auf Exponentialverteilung gegen bestimmte Klassen von Alternativen beschrieben.

Das folgende Lemma gibt den Inhalt von Theorem 2.2 in Müller (1996) wieder; insbesondere zeigt es, daß eine Verteilung durch die integrierte Verteilungsfunktion eindeutig festgelegt ist.

**2.1.2 Lemma** a) *Ist  $X$  eine positive Zufallsvariable mit endlichem Erwartungswert, so gilt für die integrierte Verteilungsfunktion:  $\Psi_X$  ist monoton fallend und konvex; die rechtsseitige Ableitung  $D^+\Psi_X$  existiert mit  $-1 \leq D^+\Psi_X \leq 0$ ; es gilt  $\lim_{t \rightarrow \infty} \Psi_X(t) = 0$ .*

b) *Zu jeder Funktion  $\Psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  mit den Eigenschaften aus a) gibt es genau eine Verteilung auf  $\mathbb{R}_+$ , deren integrierte Verteilungsfunktion  $\Psi$  ist. Die zugehörige Verteilungsfunktion ist dann  $F(t) = D^+\Psi(t) + 1$ .*

Im folgenden sei  $X, X_1, X_2, \dots, X_n, \dots$  eine Folge von unabhängigen und identisch verteilten Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , die Werte in  $\mathbb{N}_0$  annehmen;  $F(\cdot)$  bezeichne wie üblich die Verteilungsfunktion von  $X$ . Weiter sei

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

eine parametrische Familie diskreter Verteilungen mit offenem Parameterbereich  $\Theta \subset \mathbb{R}^s$ , wobei  $E_\vartheta(X) < \infty \forall \vartheta \in \Theta$  gelte. Getestet werden soll die Hypothese  $\mathcal{H}_0 : F(\cdot) \in \mathcal{F}_\Theta$  gegen die allgemeine Alternative  $\mathcal{H}_1 : F(\cdot) \notin \mathcal{F}_\Theta$ . Dazu wird im folgenden die empirische integrierte Verteilungsfunktion  $\Psi_n(\cdot)$  mit der geschätzten integrierten Verteilungsfunktion

$$\hat{\Psi}(t) = \int_t^\infty \bar{F}(x, \hat{\vartheta}_n) \lambda(dx)$$

verglichen, wobei  $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \dots, X_n)$  ein geeigneter Schätzer für den Parametervektor  $\vartheta$  ist; insbesondere gelte  $\hat{\vartheta}_n \in \Theta$ . Als Teststatistik verwenden wir die Größe

$$T_n = \sup_{t \geq 0} \sqrt{n} |\Psi_n(t) - \hat{\Psi}(t)|, \quad (2.1)$$

die analog zur Kolmogorov-Smirnov-Teststatistik für den empirischen Prozeß aufgebaut ist. Da sowohl  $\Psi_n$  als auch  $\hat{\Psi}$  auf dem Intervall  $(k, k+1]$ ,  $k \in \mathbb{N}_0$ , linear sind, gilt mit  $M := \max_{1 \leq i \leq n} X_i$

$$T_n = \sup_{k \in \mathbb{N}_0} \sqrt{n} |\Psi_n(k) - \hat{\Psi}(k)| = \max_{0 \leq k \leq M} \sqrt{n} |\Psi_n(k) - \hat{\Psi}(k)|. \quad (2.2)$$

$T_n$  läßt sich als Funktional des geschätzten (diskreten) empirischen Prozesses  $\mathcal{Z}_n = (Z_{n,k})_{k \geq 0}$  mit  $Z_{n,k} = \sqrt{n}(F_n(k) - F(k, \hat{\vartheta}_n))$  darstellen: wegen

$$\Psi_n(k) - \hat{\Psi}(k) = \sum_{j=k}^{\infty} (F(j, \hat{\vartheta}_n) - F_n(j))$$

folgt

$$T_n = h(\mathcal{Z}_n) = \sup_{k \geq 0} \left| \sum_{j \geq k} Z_{n,j} \right|, \quad (2.3)$$

wobei  $h(\cdot)$  für eine Folge  $x = (x_k)_{k \geq 0}$  durch  $h(x) = \sup_{k \geq 0} \left| \sum_{j \geq k} x_j \right|$  definiert ist. Allerdings ist  $h(\cdot)$  auf dem Raum  $c_0$  der Nullfolgen, ausgestattet mit der Supremumsnorm, i. allg. nicht definiert. Die in Henze (1996) erzielten Ergebnisse sind deshalb nicht direkt anwendbar. Der geeignete Rahmen ist, wie in der Einleitung erwähnt, der Banachraum  $\ell_1$  mit der Norm  $\|x\|_1 = \sum_{k \geq 0} |x_k|$ . Als Funktional auf  $\ell_1$  ist  $h(\cdot)$  nämlich stetig, so daß aus der Verteilungskonvergenz des empirischen Prozesses die Konvergenz der Verteilung von  $T_n$  gegen eine Grenzverteilung geschlossen werden kann (s. Korollar 2.1.7).

**2.1.3 Lemma** *Das Funktional  $h : \ell_1 \rightarrow \mathbb{R}$  ist stetig bzgl.  $\|\cdot\|_1$ .*

BEWEIS: Sei  $\|x - y\|_1 < \delta$ ,  $x, y \in \ell_1$ . Dann folgt für  $k \in \mathbb{N}_0$

$$\left| \left| \sum_{j \geq k} x_j \right| - \left| \sum_{j \geq k} y_j \right| \right| \leq \left| \sum_{j \geq k} x_j - \sum_{j \geq k} y_j \right| \leq \sum_{j \geq k} |x_j - y_j| < \delta$$

und daraus  $|h(x) - h(y)| \leq \delta$ . ■

**2.1.4 Lemma** Sei  $\mathcal{B}$  die  $\sigma$ -Algebra der Borelmengen auf  $l_p$ ,  $1 \leq p < \infty$ , und  $\mathcal{M}$  die Projektions- $\sigma$ -Algebra, d.h. die kleinste  $\sigma$ -Algebra auf  $l_p$ , so daß die Projektionen  $x \rightarrow \pi_k x := x_k$  ( $x \in l_p, k \geq 0$ ) meßbar sind. Dann gilt  $\mathcal{B} = \mathcal{M}$ .

BEWEIS: 1.) Da der Raum  $l_p$  separabel ist, wird  $\mathcal{B}$  erzeugt von den offenen Kugeln  $K(x, \varepsilon) = \{y \in l_p : \|x - y\|_p < \varepsilon\}$  ( $x \in l_p, \varepsilon > 0$ ). Sei  $I_1 = \{(\varepsilon_k)_{k \geq 0} : \varepsilon_k \geq 0 \forall k, \sum_{k \geq 0} \varepsilon_k^p < \varepsilon^p\}$  und  $I_2 = \{(\varepsilon_k) \in I_1 : \varepsilon_k \in \mathbb{Q}_{\geq 0} \forall k \in \mathbb{N}_0\}$ . Mit  $B_k = [x_k - \varepsilon_k, x_k + \varepsilon_k] \in \mathcal{B}_1$  folgt

$$K(x, \varepsilon) = \bigcup_{I_1} \bigcap_{k \geq 0} \pi_k^{-1}(B_k) = \bigcup_{I_2} \bigcap_{k \geq 0} \pi_k^{-1}(B_k)$$

und somit  $\mathcal{B} \subset \mathcal{M}$ .

2.) Die Projektionen  $\pi_k : l_p \rightarrow \mathbb{R}$  ( $k \in \mathbb{N}_0$ ) sind offensichtlich stetig, es gilt also  $\pi_k^{-1}(\mathcal{B}_1) \subset \mathcal{B}$  und folglich  $\mathcal{M} \subset \mathcal{B}$ . ■

$\mathcal{Z}_n$  ist offenbar eine  $(\mathcal{A}, \mathcal{M})$ -meßbare Abbildung von  $\Omega$  nach  $\ell_1$ . Nach dem vorstehenden Hilfssatz ist  $\mathcal{Z}_n$  also auch  $(\mathcal{A}, \mathcal{B})$ -meßbar, d.h. die Verteilung  $Q_n = P \circ \mathcal{Z}_n^{-1}$  ist ein Borelmaß auf  $\ell_1$ . Somit läßt sich die gesamte Theorie der Verteilungskonvergenz auf metrischen Räumen einsetzen (s. etwa Billingsley, (1968)).

**2.1.5 Lemma** Eine Folge  $\{Q_n : n \geq 1\}$  von Wahrscheinlichkeitsmaßen auf  $(l_p, \mathcal{B})$  ist genau dann straff, wenn die beiden folgenden Bedingungen erfüllt sind:

(i) Für jedes positive  $\delta$  und jedes  $l \in \mathbb{N}_0$  existiert eine endliche positive Zahl  $K$ , so daß

$$Q_n(\{x \in l_p : |x_l| \leq K\}) \geq 1 - \delta, \quad n \in \mathbb{N}.$$

(ii) Für jedes positive  $\delta$  und  $\eta$  existiert ein  $l \in \mathbb{N}_0$ , so daß

$$Q_n(\{x \in l_p : \sum_{k \geq l} |x_k|^p \leq \eta\}) \geq 1 - \delta, \quad n \in \mathbb{N}.$$

BEWEIS: Für den Beweis benötigt man das folgende leicht zu zeigende Kriterium für relative Kompaktheit in  $l_p$ : Eine Teilmenge  $M$  des  $l_p$ ,  $1 \leq p < \infty$ , ist genau dann relativ kompakt, wenn  $M$  beschränkt ist und

$$\lim_{n \rightarrow \infty} \sup_{x \in M} \sum_{j \geq n} |x_j|^p = 0$$

gilt. Damit kann der Beweis analog wie der Beweis von Lemma 2.1 in Henze (1996) geführt werden, wenn man dort  $A_j = \{x \in l_p : \sum_{k \geq l(j)} |x_k|^p \leq 1/j\}$  setzt. ■

Nach diesen Vorbereitungen soll nun die Verteilungskonvergenz des empirischen Prozesses in  $\ell_1$  gezeigt werden. Dabei betrachten wir die allgemeinere Situation eines Dreiecksschemas  $X_{n1}, X_{n2}, \dots, X_{nn}$  von zeilenweise unabhängigen und identisch verteilten Zufallsvariablen mit Verteilungsfunktion  $F(\cdot, \vartheta_n)$ , wobei

$$\vartheta_n \in \Theta, \quad n \geq 1, \quad \lim_{n \rightarrow \infty} \vartheta_n = \vartheta, \quad \vartheta \in \Theta.$$

Diese Verallgemeinerung ist nötig, um die Konvergenz des später beschriebenen parametrischen Bootstrap-Verfahrens zu erhalten.

Die folgenden drei Regularitätsbedingungen sind hinreichend für die schwache Konvergenz des empirischen Prozesses in  $\ell_1$ . Die bekannten Familien diskreter Verteilungen mit endlichem Erwartungswert erfüllen diese Bedingungen, so etwa die Poissonverteilung, die geometrische sowie die logarithmische Verteilung, die im nächsten Unterabschnitt als hypothetische Modelle verwendet werden.

(R1) Es gilt die Darstellung

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n l(X_{nj}, \vartheta_n) + \epsilon_n,$$

wobei  $\epsilon_n = o_{P_{\vartheta_n}}(1)$  ( $n \rightarrow \infty$ ) gilt und  $l : \mathbb{N}_0 \times \Theta \rightarrow \mathbb{R}^s$  eine meßbare Funktion mit folgenden Eigenschaften ist:

- (i)  $E_{\vartheta} l(X, \vartheta) = 0$ ,  $\vartheta \in \Theta$ .
- (ii) Für festes  $k \in \mathbb{N}_0$  ist  $l(k, \cdot)$  eine stetige Funktion von  $\vartheta$ .
- (iii)  $D(\vartheta) = E[l(X, \vartheta)l(X, \vartheta)']$  ist eine endliche und nicht-negativ definite Matrix, die stetig von  $\vartheta$  abhängt.

(R2) Es existiert ein  $n_0 \in \mathbb{N}$ , so daß gilt:

$$\lim_{l \rightarrow \infty} \sup_{n \geq n_0} \sum_{k \geq l} \sqrt{1 - F(k, \vartheta_n)} = 0.$$

(R3) Für festes  $k \geq 0$  existiert der Gradient  $\nabla_{\vartheta} F(k, \vartheta)$  von  $F(k, \vartheta)$  und ist stetig in Abhängigkeit von  $\vartheta$ . In einer Umgebung  $\mathcal{U}(\vartheta)$  von  $\vartheta = (\theta_1, \dots, \theta_s)'$  gilt

$$\lim_{l \rightarrow \infty} \sup_{\vartheta^* \in \mathcal{U}(\vartheta)} \sum_{k \geq l} \left| \frac{\partial F(k, \vartheta^*)}{\partial \theta_j} \right| = 0, \quad j = 1, \dots, s.$$

### 2.1.6 Satz

Sind die Bedingungen (R1) bis (R3) erfüllt, so gilt

$$\mathcal{Z}_n \xrightarrow{\mathcal{D}} \mathcal{W}.$$

Hierbei ist  $\mathcal{W} = (W_k)_{k \geq 0}$  ein Gaußprozeß in  $\ell_1$ , für den  $E_{\vartheta}[W_k] = 0$  und

$$\begin{aligned} C_{\vartheta}(k, m) &= F(\min(k, m), \vartheta) - F(k, \vartheta)F(m, \vartheta) \\ &\quad - J(k, \vartheta)' \nabla_{\vartheta} F(m, \vartheta) - J(m, \vartheta)' \nabla_{\vartheta} F(k, \vartheta) \\ &\quad + \nabla_{\vartheta} F(k, \vartheta)' D(\vartheta) \nabla_{\vartheta} F(m, \vartheta) \end{aligned} \quad (2.4)$$



mit

$$J(k, \vartheta) = E_{\vartheta}[l(X, \vartheta) \mathbf{1}\{X \leq k\}] = \sum_{j=0}^k l(j, \vartheta) f(j, \vartheta)$$

für  $k, m \geq 0, \vartheta \in \Theta$ , gilt.

BEWEIS: Unter (R1) und (R3) läßt sich die Konvergenz der endlichdimensionalen Verteilungen wie in Henze (1996) zeigen, woraus auch die Aussagen über den Erwartungswert und die Kovarianzmatrix folgen. Zu zeigen ist nur noch die Straffheit der Folge  $\{\mathcal{Z}_n : n \geq 1\}$  unter  $P_{\vartheta_n}$ , also die Bedingungen von Lemma 2.1.5. Aus der Konvergenz der eindimensionalen Verteilungen folgt, daß die Folge  $(Z_{n,l})_{n \geq 1}$  für festes  $l \in \mathbb{N}_0$  straff ist, d.h. Bedingung (i) in Lemma 2.1.5 ist erfüllt. Somit ist nur noch Bedingung (ii) für den Fall  $p = 1$  nachzuweisen, also die Existenz von Zahlen  $l$  und  $n_0$  aus  $\mathbb{N}$ , so daß gilt:

$$P_{\vartheta_n} \left( \sum_{k \geq l} |Z_{n,k}| > \eta \right) \leq \delta, \quad n \geq n_0.$$

Wegen  $Z_{n,k} = T_{n,k} + V_{n,k}$  mit  $T_{n,k} = \sqrt{n}(F_n(k) - F(k, \vartheta_n))$  und  $V_{n,k} = \sqrt{n}(F(k, \vartheta_n) - F(k, \hat{\vartheta}_n))$  ist dies gleichbedeutend mit der Existenz von Zahlen  $l$  und  $n_0$  aus  $\mathbb{N}$ , so daß

$$P_{\vartheta_n} \left( \sum_{k \geq l} |T_{n,k}| > \frac{\eta}{2} \right) \leq \frac{\delta}{2}, \quad n \geq n_0, \quad (2.5)$$

und

$$P_{\vartheta_n} \left( \sum_{k \geq l} |V_{n,k}| > \frac{\eta}{2} \right) \leq \frac{\delta}{2}, \quad n \geq n_0, \quad (2.6)$$

gilt. Dabei folgt (2.5) mit Hilfe der Markov- und der Jensen-Ungleichung aus der Ungleichungskette

$$P_{\vartheta_n} \left( \sum_{k \geq l} |T_{n,k}| > \frac{\eta}{2} \right) \leq \frac{2}{\eta} E_{\vartheta_n} \left[ \sum_{k \geq l} |T_{n,k}| \right] = \frac{2}{\eta} \sum_{k \geq l} E_{\vartheta_n} [|T_{n,k}|]$$

$$\leq \frac{2}{\eta} \sum_{k \geq l} \sqrt{E_{\vartheta_n} [ |T_{n,k}^2| ]} \leq \frac{2}{\eta} \sum_{k \geq l} \sqrt{1 - F(k, \vartheta_n)} \rightarrow 0 \quad (l \rightarrow \infty),$$

und zwar gleichmäßig für alle  $n \geq n_0$  nach (R2). Dabei ergibt sich die letzte Ungleichung, wenn man beachtet, daß  $n F_n(k)$  binomialverteilt mit Parametern  $n$  und  $F(k, \vartheta_n)$  ist.

Zum Beweis von (2.6) benötigen wir folgendes bekannte Ergebnis: Ist  $z$  ein Vektor im  $\mathbb{R}^s$ , versehen mit der  $\ell_1$ -Norm, und  $A = (a_{kj})$  eine  $r \times s$ -Matrix, so ist die Spaltensummennorm

$$\|A\|_1 = \max_{1 \leq j \leq s} \sum_{k=1}^r |a_{kj}|$$

die zugehörige Abbildungsnorm; insbesondere gilt  $\|Az\|_1 \leq \|A\|_1 \|z\|_1$ , die Spaltensummennorm ist also eine mit der  $\ell_1$ -Norm verträgliche Matrixnorm. Aus dem Mittelwertsatz für vektorwertige Funktionen (Heuser (1986), S.278) ergibt sich nun für alle  $m \in \mathbb{N}$  und  $\hat{\vartheta}_n, \vartheta_n \in \mathcal{U}(\vartheta)$  die Abschätzung

$$\begin{aligned} \sum_{k=l}^{l+m} |V_{n,k}| &= \sum_{k=l}^{l+m} |\sqrt{n}(F(k, \vartheta_n) - F(k, \hat{\vartheta}_n))| \\ &\leq \|\sqrt{n}(\hat{\vartheta}_n - \vartheta_n)\|_1 \max_{\vartheta^* \in S} \max_{1 \leq j \leq s} \sum_{k=l}^{l+m} \left| \frac{\partial F(k, \vartheta^*)}{\partial \theta_j} \right| \\ &\leq \|\sqrt{n}(\hat{\vartheta}_n - \vartheta_n)\|_1 a_l, \end{aligned}$$

wobei  $S$  die Verbindungsstrecke zwischen  $\vartheta_n$  und  $\hat{\vartheta}_n$  und

$$a_l = \sup_{\vartheta^* \in \mathcal{U}(\vartheta)} \max_{1 \leq j \leq s} \sum_{k=l}^{\infty} \left| \frac{\partial F(k, \vartheta^*)}{\partial \theta_j} \right|$$

ist. Für eine hinreichend kleine Umgebung  $\mathcal{U}(\vartheta)$  von  $\vartheta$  folgt aus Bedingung (R3)  $\lim_{l \rightarrow \infty} a_l = 0$ . Damit läßt sich (2.6) wie die entsprechende Bedingung im Beweis von Theorem 3.1 in Henze (1996) zeigen. ■

Im weiteren wird immer vorausgesetzt, daß die hypothetische Verteilung  $F(\cdot, \vartheta)$  die Bedingungen (R1) bis (R3) erfüllt. Außerdem sei die Verteilung

nicht degeneriert, also kein Einpunktmaß. Da das Funktional  $h(\cdot)$  stetig ist (s. Lemma 2.1.3), erhält man aus Satz 2.1.6 und dem Abbildungssatz

**2.1.7 Korollar** *Unter  $P_{\vartheta_n}$ ,  $\vartheta_n \in \Theta$ , mit  $\lim_{n \rightarrow \infty} \vartheta_n = \vartheta \in \Theta$ , gilt*

$$T_n \xrightarrow{\mathcal{D}} h(\mathcal{W}) = \sup_{k \geq 0} \left| \sum_{j \geq k} W_j \right|.$$

Die Verteilung  $H_{\vartheta}(t) = P_{\vartheta}(h(\mathcal{W}) \leq t)$  ist nur von akademischen Interesse, da sie vom unbekanntem Parameter  $\vartheta$  abhängt. Aus diesem Grunde wird die hypothetische Verteilung  $H_{n,\vartheta}(t) = P_{\vartheta}(T_n \leq t)$  von  $T_n$  durch das folgende Monte Carlo-Verfahren approximiert.

Zu gegebenen  $X_1, \dots, X_n$  werden zuerst  $\hat{\vartheta}_n$  und  $T_n(X_1, \dots, X_n)$  berechnet. Dann wird  $T_{j,n}^* = T_n(X_{j1}^*, \dots, X_{jn}^*)$ ,  $1 \leq j \leq k_n$ , berechnet, wobei, bedingt nach  $X_1, \dots, X_n$ ,  $X_{j1}^*, \dots, X_{jn}^*$ ,  $1 \leq j \leq k_n$ , unabhängige und identisch verteilte Zufallsvariablen mit Verteilungsfunktion  $F(\cdot, \hat{\vartheta}_n)$  sind. Bezeichnet  $c_{n,\alpha}^*$  das  $(1 - \alpha)$ -Quantil von  $H_{n,k_n}^*$ , wobei

$$H_{n,k_n}^*(t) = \frac{1}{k_n} \sum_{j=1}^{k_n} \mathbf{1}\{T_{j,n}^* \leq t\}$$

die empirische Verteilungsfunktion von  $T_{1,n}^*, \dots, T_{k_n,n}^*$  ist, so wird die Hypothese  $H_0$  zum Niveau  $\alpha$  abgelehnt, falls  $T_n$  größer als  $c_{n,\alpha}^*$  ist.

Im weiteren soll gezeigt werden, daß dieser auf  $T_n$  basierende *Bootstrap-Test* asymptotisch das Niveau  $\alpha$  einhält und gegenüber Alternativverteilungen mit endlichem Erwartungswert unter schwachen Bedingungen konsistent ist. Zu diesem Zweck werden die folgenden Hilfssätze benötigt.

**2.1.8 Lemma**  $Q = P \circ \mathcal{W}^{-1}$  bezeichne die Verteilung von  $\mathcal{W}$ . Die Verteilung  $Qh^{-1}$  von  $h(\mathcal{W})$  ist auf  $(0, \infty)$  konzentriert und absolut-stetig bezüglich des Lebesgue-Maßes; die Dichte  $dQh^{-1}/d\lambda$  ist auf  $(0, \infty)$  streng positiv. Insbesondere ist die Verteilungsfunktion  $H_{\vartheta}$  stetig und streng monoton wachsend auf  $[0, \infty)$ .

BEWEIS: Das Funktional  $h$  ist eine Norm auf  $\ell_1$ ; insbesondere ist  $h$  konvex.  $N_Q$  bezeichne den linearen Träger des Maßes  $Q$  (Definition siehe Davydov und Lifshits (1985), S. 2797 unten), der für Gauß-Maße mit dem topologischen Träger übereinstimmt (Davydov und Lifshits (1985), Theorem 5.1.3). Da  $Q$  zentriert ist, gilt  $0 \in N_Q$  (siehe auch Davydov und Lifshits (1985), S. 2814 oben) und somit

$$\inf_{x \in N_Q} h(x) = 0.$$

Aus Theorem 7.1 in Davydov und Lifshits (1985) folgt, daß die Verteilung  $Qh^{-1}$  von  $h(\mathcal{W})$  auf  $[0, \infty)$  konzentriert und absolut-stetig auf  $(0, \infty)$  bezüglich des Lebesgue-Maßes ist; die Dichte  $dQh^{-1}/d\lambda$  ist positiv auf  $(0, \infty)$ .

Da  $h$  eine Norm ist, gilt  $h(\mathcal{W}) = 0$  genau dann, wenn  $\mathcal{W} = 0$  ist. Nach Voraussetzung ist die Verteilung  $F(\cdot, \vartheta)$  nicht degeneriert und somit  $Q$  nicht das Einpunktmaß in 0, woraus  $P(\mathcal{W} = 0) = P(h(\mathcal{W}) = 0) = 0$  folgt.  $Qh^{-1}$  ist also auf  $(0, \infty)$  konzentriert. Somit ist alles bewiesen. ■

**Bemerkung:** Der Raum  $\ell_1$ , versehen mit der Norm  $\|\cdot\|_h = h(\cdot)$ , ist nicht vollständig. Denn sei für  $n \in \mathbb{N}$   $x^{(n)} = (x_k^{(n)})_{k \geq 1}$  mit  $x_k^{(n)} = (-1)^k/k$  für  $1 \leq k \leq n$  und  $x_k^{(n)} = 0$  für  $k > n$ . Dann ist  $(x^{(n)})$  eine Cauchyfolge in  $(\ell_1, \|\cdot\|_h)$ , aber das Grenzelement  $x = ((-1)^k/k)_{k \geq 1}$  liegt nicht in  $\ell_1$ .

Dagegen ist der Raum aller Folgen mit konvergenter Reihe  $\sum_{k \geq 0} x_k$ , versehen mit der Norm  $\|\cdot\|_h$ , ein Banachraum.

**2.1.9 Lemma**  $X_1, X_2, \dots$  liege eine Verteilung mit endlichem Erwartungswert zugrunde; weiter konvergiere  $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \dots, X_n)$  stochastisch gegen  $\vartheta \in \Theta$ .  $\|\cdot\|_\infty$  bezeichne die Supremumsnorm auf  $\mathbb{R}$ . Dann gilt:

$$a) \|H_{n, k_n}^* - H_\vartheta\|_\infty \xrightarrow{P} 0 \quad \text{für } n, k_n \rightarrow \infty.$$

$$b) c_{n, \alpha}^* \xrightarrow{P} H_\vartheta^{-1}(1 - \alpha) \quad \text{für } n, k_n \rightarrow \infty.$$

BEWEIS: Da  $H_\vartheta$  nach Lemma 2.1.8 stetig ist, ist die Aussage von Korollar 2.1.7 nach dem Satz von Pólya gleichbedeutend mit  $\lim_{n \rightarrow \infty} \|H_{n, \vartheta_n} - H_\vartheta\|_\infty = 0$  für jede Folge  $\vartheta_n \in \Theta$ , die gegen  $\vartheta$  konvergiert. Hieraus folgt, da nach Voraussetzung  $\hat{\vartheta}_n$  stochastisch gegen  $\vartheta$  konvergiert,

$$\|H_{n, \hat{\vartheta}_n} - H_\vartheta\|_\infty \xrightarrow{P} 0 \quad \text{für } n \rightarrow \infty.$$

Da andererseits  $\|H_{n, k_n}^* - H_{n, \hat{\vartheta}_n}\|_\infty$  fast sicher gegen Null konvergiert, folgt Teil a) mit der Dreiecksungleichung. b) folgt aus a), da nach Lemma 2.1.8  $H_\vartheta$  stetig ist und auf dem Intervall  $[0, \infty)$  streng monoton wächst. ■

### 2.1.10 Satz

a) Für jedes  $\vartheta \in \Theta$  gilt

$$\lim P_\vartheta(T_n > c_{n, \alpha}^*) = \alpha \quad \text{für } n, k_n \rightarrow \infty.$$

b)  $X_1, X_2, \dots$  sei eine Folge von unabhängigen und identisch verteilten  $\mathbb{N}_0$ -wertigen Zufallsvariablen mit Verteilungsfunktion  $F$ , für die

$$\rho := \inf_{\vartheta \in \Theta} \sup_{t \in \mathbb{N}_0} |F(t) - F(t, \vartheta)| > 0$$

gelte. Existiert der Erwartungswert von  $X_1$  und konvergiert  $\hat{\vartheta}_n$  stochastisch gegen ein  $\vartheta \in \Theta$ , so ist der Test konsistent, d.h. es gilt

$$\lim P(T_n > c_{n, \alpha}^*) = 1 \quad \text{für } n, k_n \rightarrow \infty.$$

BEWEIS: a) Folgt aus Lemma 2.1.9 b), da die Voraussetzungen des Lemmas insbesondere für die hypothetische Verteilung erfüllt sind.

b) Es gilt

$$\begin{aligned} T_n &= \sup_{k \geq 0} \left| \sum_{j \geq k} \sqrt{n} (F_n(j) - F(j, \hat{\vartheta}_n)) \right| & (2.7) \\ &\geq \sqrt{n} \sup_{k \geq 0} \left| \left| \sum_{j \geq k} (F_n(j) - F(j)) \right| - \left| \sum_{j \geq k} (F(j) - F(j, \hat{\vartheta}_n)) \right| \right|. \end{aligned}$$

Nun konvergiert  $1 - F_n(j)$  für jedes  $j$  fast sicher gegen  $1 - F(j)$ , und es gilt

$$\lim_{n \rightarrow \infty} \sum_{j=0}^{\infty} (1 - F_n(j)) = \lim_{n \rightarrow \infty} \bar{X}_n = EX = \sum_{j=0}^{\infty} (1 - F(j)) \quad f.s.$$

Aus einer verallgemeinerten Version des Satzes von Scheffé (Shorak und Wellner (1986), S. 862) folgt

$$\sup_{k \geq 0} \left| \sum_{j \geq k} (F_n(j) - F(j)) \right| = \sup_{k \geq 0} \left| \sum_{j \geq k} (1 - F(j)) - (1 - F_n(j)) \right| \rightarrow 0 \quad f.s.$$

Andererseits liefert die Voraussetzung  $\rho > 0$  die Ungleichung

$$\liminf_{n \rightarrow \infty} \sup_{k \geq 0} \left| \sum_{j \geq k} (F(j) - F(j, \hat{\vartheta}_n)) \right| \geq \frac{\rho}{2} \quad f.s.$$

Mit Blick auf (2.7) ergibt sich  $\lim_{n \rightarrow \infty} T_n = \infty \quad f.s.$ , und hieraus folgt zusammen mit Lemma 2.1.9 b) die Behauptung. ■

**Bemerkungen:** 1. Die Bedingungen an die Alternativverteilung können in speziellen Fällen abgeschwächt werden. So kann man zeigen, daß der Test auf Poissonverteilung auch gegen Verteilungen mit  $EX_1 = \infty$  (und folglich  $\hat{\vartheta}_n \rightarrow \infty$ ) konsistent ist (vgl. Henze (1996), S.91).

2. Lemma 2.1.9 und Satz 2.1.10a) bleiben für jedes stetige Funktional auf  $\ell_1$  gültig, für das die Aussagen von Lemma 2.1.8 gelten. Ein Beispiel ist die  $\ell_1$ -Norm selbst; die entsprechende Testgröße ist

$$\tilde{T}_n = \sum_{k \geq 0} |Z_{n,k}|.$$

Ein weiteres Beispiel ist die modifizierte Cramér-von Mises-Statistik

$$W_{mod}^2 = \sum_{k \geq 0} Z_{n,k}^2,$$

die Choulakian et al. (1994) bzw. Spinelli und Stephens (1997) als Testgröße für diskrete Verteilungen mit einer endlichen Klassenzahl bzw. für die Poissonverteilung bei Vorliegen einer Einteilung in endlich viele Klassen verwenden.  $W_{mod}^2$  gewichtet offensichtlich Abweichungen in den Verteilungsenden stärker als die gewöhnliche Cramér-von Mises-Statistik. Für beide Testgrößen bleibt auch Satz 2.1.10b) gültig, wie man am Beweis sieht.

### 2.1.2 Die Güte des Tests unter benachbarten Alternativen

In Satz 2.1.10 wurde gezeigt, daß der auf der integrierten Verteilungsfunktion beruhende Test asymptotisch das Niveau  $\alpha$  einhält und konsistent ist, d.h. Alternativverteilungen mit gegen eins gehender Wahrscheinlichkeit erkennt. Um theoretische Aussagen über das Güteverhalten des Tests zu erhalten, die über die globale Konsistenz des Tests hinausgehen, kann man die Güte des Tests gegenüber benachbarten Alternativen untersuchen. Damit bezeichnet man eine Folge von Alternativverteilungen, deren Verteilungsfunktionen mit einer bestimmten Geschwindigkeit gegen die hypothetische Verteilungsfunktion konvergieren. Die Konvergenzgeschwindigkeit wird dabei so gewählt, daß die asymptotische Güte zwischen  $\alpha$  und 1 liegt. Gegen Folgen von Alternativverteilungen mit niedrigerer Konvergenzrate ist der Test dann konsistent, solche mit höherer Konvergenzrate werden von dem Test asymptotisch nur mit Wahrscheinlichkeit  $\alpha$  entdeckt.

Sei im folgenden

$$\mathcal{F}_\Theta = \{F(\cdot, \vartheta) : \vartheta \in \Theta\} \subset \{\tilde{F}(\cdot, \vartheta, \beta) : \vartheta \in \Theta, \beta \in B \subset \mathbb{R}^q\}$$

mit  $\tilde{F}(\cdot, \vartheta, 0) = F(\cdot, \vartheta)$ . Für  $n \in \mathbb{N}$  seien  $X_{n1}, \dots, X_{nn}$  unabhängige und identisch verteilte Zufallsvariablen. Eine Folge von benachbarten Alternativen zu der Hypothese  $\mathcal{H}_{n0} : X_{n1} \sim F(\cdot, \vartheta)$ ,  $\vartheta \in \Theta$ , ist dann gegeben durch

$$\mathcal{H}_{n1} : X_{n1} \sim \tilde{F}\left(\cdot, \vartheta, \frac{\gamma}{\sqrt{n}}\right) \quad \left(\frac{\gamma}{\sqrt{n}} \in B, n = 1, 2, \dots\right).$$

Aus dem gleichen Grund wie im vorhergehenden Abschnitt betrachten wir allgemeiner ein Dreiecksschema  $X_{n1}, X_{n2}, \dots, X_{nn}$  von zeilenweise unabhängigen und identisch verteilten Zufallsvariablen mit Verteilungsfunktion  $\tilde{F}(\cdot, \vartheta_n, \beta_n)$ , wobei

$$\lim_{n \rightarrow \infty} \vartheta_n = \vartheta, \quad \beta_n = n^{-1/2} \gamma \quad (\vartheta_n \in \Theta, n^{-1/2} \gamma \in B \text{ für } n \geq 1, \vartheta \in \Theta)$$

ist. Die folgenden Regularitätsbedingungen sind hinreichend für die schwache Konvergenz des empirischen Prozesses in  $\ell_1$  unter  $\mathcal{H}_{1n}$ ; dabei deutet der Index  $n$  immer an, daß Größen unter  $\mathcal{H}_{1n}$  berechnet werden.

(R1') Unter  $\mathcal{H}_{1n}$  gilt

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \tilde{l}(X_{nj}, \vartheta_n, \beta_n) + A\gamma + \epsilon_n,$$

wobei  $A$  eine endliche  $(s \times q)$ -Matrix ist und  $\epsilon_n = o_{P_n}(1)$  ( $n \rightarrow \infty$ ) gilt.  $\tilde{l} : \mathbb{N}_0 \times \Theta \times B \rightarrow \mathbb{R}^s$  ist eine meßbare Funktion mit folgenden Eigenschaften:

- (i)  $E_n[\tilde{l}(X, \vartheta_n, \beta_n)] = 0$ ,  $n \geq 1$ .
- (ii)  $D_n(\vartheta_n, \beta_n) = E_n[\tilde{l}(X, \vartheta_n, \beta_n)\tilde{l}(X, \vartheta_n, \beta_n)']$  ist eine endliche nicht-negativ definite Matrix, die gegen eine endliche nicht-negativ definite Matrix  $D(\vartheta)$  konvergiert.

(R2') Es existiert ein  $n_0 \in \mathbb{N}$ , so daß

$$\lim_{l \rightarrow \infty} \sup_{n \geq n_0} \sum_{k \geq l} \sqrt{1 - \tilde{F}(k, \vartheta_n, \beta_n)} = 0.$$

(R3') Es gilt Bedingung (R3) von S. 92. Für festes  $k \geq 0$  existiert der Gradient  $\nabla_{\beta} \tilde{F}(k, \vartheta, \beta)$  und ist stetig in Abhängigkeit von  $\vartheta$  und  $\beta$ . In einer Umgebung  $\mathcal{U}$  von  $(\vartheta, 0)$  gilt

$$\lim_{l \rightarrow \infty} \sup_{(\vartheta^*, \beta^*) \in \mathcal{U}} \sum_{k \geq l} \left| \frac{\partial \tilde{F}(k, \vartheta^*, \beta^*)}{\partial \beta_j} \right| = 0, \quad j = 1, \dots, q.$$

**Bemerkung:** Als Motivation für (R1') wird in der Literatur oft angemerkt, daß Maximum-Likelihood-Schätzer (im erweiterten Modell  $\tilde{F}(\cdot, \vartheta, \beta)$ !) unter recht allgemeinen Bedingungen eine solche Darstellung besitzen (siehe z.B. Durbin (1973a), S.287). Dieses Argument erscheint nicht sehr zwingend. Da die Hypothese  $X_{n1} \sim F(\cdot, \vartheta)$  getestet wird, hängt der verwendete Schätzer in der Praxis nicht von der Alternative ab; der ML-Schätzer für  $\vartheta$  unter dem Modell  $\tilde{F}(\cdot, \vartheta, \beta)$  ist also gar nicht verfügbar.



Der Grund für die Darstellung (R1') ist vielmehr folgender: Erfüllt ein Schätzer unter der Hypothese die Voraussetzung (R1) mit einem Vektor  $l(\cdot, \vartheta)$ , so gilt für diesen Schätzer meist auch unter  $\mathcal{H}_{n1}$ :

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n l(X_{nj}, \vartheta_n) + \epsilon_n, \quad \epsilon_n = o_{P_n}(1). \quad (2.8)$$

Dies ist insbesondere dann der Fall, wenn die Produktverteilung  $\otimes_{j=1}^n \tilde{F}(\cdot, \vartheta_n, \beta_n)$  (im nichtparametrischen Sinne) benachbart ist zu  $\otimes_{j=1}^n F(\cdot, \vartheta)$  (siehe etwa Witting und Müller-Funk (1995), Definition 6.106). Eine hinreichende Bedingung hierfür ist nach dem 2. Le Cam Lemma die  $\mathbf{L}_2$ -Differenzierbarkeit von  $\tilde{F}(\cdot, \vartheta, \beta)$  (Witting und Müller-Funk (1995), Satz 6.130 bzw. Korollar 6.131). Es ist aber nicht sinnvoll, diese Bedingung generell an  $\tilde{F}$  zu stellen, da die Gültigkeit von (2.8) oft einfacher direkt gezeigt werden kann.

Allerdings ist  $l(X, \vartheta_n)$  unter  $\mathcal{H}_{n1}$  nicht mehr zentriert. Gilt unter der Alternative die Entwicklung

$$E_{\vartheta_n, \beta_n} [l(X, \vartheta_n)] = \beta_n' \nabla_{\beta} E_{\vartheta, \beta} [l(X, \vartheta)]|_{\beta=0} + o_{P_n}(1),$$

so folgt aus (2.8) die Darstellung (R1'), wenn man

$$\tilde{l}(x, \vartheta_n, \beta_n) = l(x, \vartheta_n) - E_{\vartheta_n, \beta_n} [l(X, \vartheta_n)] \quad (2.9)$$

und

$$A = \nabla_{\beta} E_{\vartheta, \beta} [l(X, \vartheta)]|_{\beta=0} \quad (2.10)$$

setzt. Aus (2.9) wird ersichtlich, daß  $\tilde{l}$  von  $\beta_n$  abhängen wird. Bedingung (i) in Burke et al. (1978), S. 803, die eine von  $\beta_n$  unabhängige zentrierte Funktion  $l$  fordert, ist i. allg. nicht erfüllbar.

### 2.1.11 Satz

*Sind die Bedingungen (R1') bis (R3') erfüllt, so gilt für den empirischen Prozeß*

$$\mathcal{Z}_n \xrightarrow{\mathcal{D}} \tilde{\mathcal{W}}.$$

Hierbei ist  $\tilde{W} = (\tilde{W}_k)_{k \geq 0}$  ein Gaußprozeß in  $\ell_1$ , für den

$$E[\tilde{W}_k] = \gamma' \nabla_{\beta} \tilde{F}(k, \vartheta, \beta)|_{\beta=0} - \nabla_{\vartheta} F(k, \vartheta)' A \gamma$$

gilt und dessen Kovarianz durch (2.4) gegeben ist, wenn man dort  $l(X, \vartheta)$  durch  $\tilde{l}(X, \vartheta, 0)$  ersetzt.

BEWEIS: Zerlegt man den empirischen Prozeß gemäß

$$\begin{aligned} Z_{n,k} &= \sqrt{n} \left( F_n(k) - F(k, \hat{\vartheta}_n) \right) \\ &= \sqrt{n} \left( F_n(k) - \tilde{F}(k, \vartheta_n, \beta_n) \right) + \sqrt{n} \left( \tilde{F}(k, \vartheta_n, \beta_n) - F(k, \vartheta_n) \right) \\ &\quad + \sqrt{n} \left( F(k, \vartheta_n) - F(k, \hat{\vartheta}_n) \right) \\ &= Z_{n,k}^{(1)} + Z_{n,k}^{(2)} + Z_{n,k}^{(3)}, \end{aligned}$$

so erhält man mittels Taylorentwicklungen und unter Beachtung von (R1') und (R3')

$$Z_{n,k}^{(2)} = \gamma' \nabla_{\beta} \tilde{F}(k, \vartheta_n, \beta)|_{\beta=0} + o(1)$$

und

$$Z_{n,k}^{(3)} = -\nabla_{\vartheta} F(k, \vartheta_n)' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n \tilde{l}(X_{nj}, \vartheta_n, \beta_n) + A\gamma \right) + o_{P_n}(1)$$

und somit

$$\begin{aligned} Z_{n,k} &= \frac{1}{\sqrt{n}} \sum_{j=1}^n Y_{nj}(k) + \gamma' \nabla_{\beta} \tilde{F}(k, \vartheta_n, \beta)|_{\beta=0} \\ &\quad - \nabla_{\vartheta} F(k, \vartheta_n)' A \gamma + o_{P_n}(1), \end{aligned}$$

wobei

$$Y_{nj}(k) = \mathbf{1}\{X_{nj} \leq k\} - \tilde{F}(k, \vartheta_n, \beta_n) - \nabla_{\vartheta} F(k, \vartheta_n)' \tilde{l}(X_{nj}, \vartheta_n, \beta_n)$$

ist. Aus dieser Darstellung und unter Beachtung von (R1')(i),(ii) folgen mit Hilfe des Zentralen Grenzwertsatzes von Lindeberg-Feller die Konvergenz

der endlichdimensionalen Verteilungen und die Aussagen über Erwartungswert und Kovarianz.

Die Straffheit der Folgen  $\{(Z_{n,k}^{(j)})_{k \geq 1} : n \geq 1\}$  unter  $P_n$  folgt für  $j = 1$  wie in Satz 2.1.6 aus (R2'); für  $j = 2$  und  $j = 3$  läßt sich die Straffheit ähnlich wie für  $(V_{n,k})$  in Satz 2.1.6 mit (R3') zeigen. ■

**Bemerkung:** Ist (R1') mit  $\tilde{l}$  aus (2.9) und  $A$  aus (2.10) erfüllt, so besagt Satz 2.1.11, daß die asymptotische Verteilung des geschätzten diskreten empirischen Prozesses unter  $\mathcal{H}_{n1}$  diejenige von  $(W_k + \delta_k)_{k \geq 0}$  ist, wobei  $(W_k)_{k \geq 0}$  der Gaußprozeß aus Satz 2.1.6 ist und  $(\delta_k)_{k \geq 0}$  durch

$$\delta_k = \gamma' \nabla_{\beta} \tilde{F}(k, \vartheta, \beta)|_{\beta=0} - \nabla_{\vartheta} F(k, \vartheta)' \nabla_{\beta} E_{\vartheta, \beta} [l(X, \vartheta)]|_{\beta=0} \gamma \quad (2.11)$$

gegeben ist.

Die asymptotische Güte des EIVF-Tests zum Niveau  $\alpha$  für das Testproblem  $\mathcal{H}_{n0}$  gegen  $\mathcal{H}_{n1}$  ist in diesem Fall durch

$$P(h((W_k + \delta_k)_{k \geq 0}) \geq c_{\alpha}) = P\left(\sup_{k \geq 0} \left| \sum_{j \geq k} W_k + \delta_k \right| \geq c_{\alpha}\right)$$

mit  $c_{\alpha} = H_{\vartheta}^{-1}(1 - \alpha)$  gegeben und liegt zwischen  $\alpha$  und 1.

### 2.1.12 Beispiel

Um das Verhalten des empirischen Prozesses unter einer benachbarten Alternative zu illustrieren, betrachten wir die Mischung aus einer Poissonverteilung und einer Einpunktverteilung in Null (engl. „Zero-modified Poisson distribution“). Die Zähldichte ist gegeben durch

$$\begin{aligned} P(X = 0) &= \beta + (1 - \beta) e^{-\vartheta}, \\ P(X = j) &= (1 - \beta) \frac{e^{-\vartheta} \vartheta^j}{j!}, \quad j = 1, 2, \dots, \end{aligned}$$

wobei

$$\vartheta \in \Theta = (0, \infty) \quad \text{und} \quad \beta \in B = \left( \frac{-e^{-\vartheta}}{1 - e^{-\vartheta}}, 1 \right)$$

gilt. Für  $\beta = 0$  erhält man die Familie der Poissonverteilungen mit Parameter  $\vartheta$ . Für den ML-Schätzer unter der Hypothese der Poissonverteilung  $\hat{\vartheta}_n = \bar{X}_n$  gilt offensichtlich auch unter  $\mathcal{H}_{n1}$  die Darstellung (2.8). Wegen

$$E_{\vartheta_n, \beta_n} [l(X, \vartheta_n)] = E_{\vartheta_n, \beta_n}(X) - \vartheta_n = -\vartheta_n \beta_n$$

erhält man die Darstellung (R1'), wenn man

$$\tilde{l}(x, \vartheta_n, \beta_n) = x - (1 - \beta_n) \vartheta_n \quad \text{und} \quad A = -\vartheta$$

setzt. Ähnlich wie in Beispiel 2.1.13 (s.u.) läßt sich zeigen, daß auch (R2') und (R3') erfüllt sind. Nach Satz 2.1.11 konvergiert der empirische Prozeß in Verteilung gegen  $(W_k + \delta_k)_{k \geq 0}$ , wobei die Folge  $(\delta_k)$  nach (2.11) durch

$$\delta_k = \gamma \sum_{j \geq k+1} \frac{e^{-\vartheta} \vartheta^j}{j!} - \vartheta \gamma \frac{e^{-\vartheta} \vartheta^k}{k!}$$

gegeben ist.

### 2.1.3 Beispiele und Simulationsergebnisse

In diesem Unterabschnitt soll anhand einiger Simulationsstudien versucht werden, Aussagen über die Güte der auf den Testgrößen  $T_n$ ,  $\tilde{T}_n$  und  $W_{mod}^2$  basierenden Tests bei festem Stichprobenumfang  $n$  zu machen. Als hypothetische Verteilungen werden die (positive) Poissonverteilung, die geometrische und die logarithmische Verteilung verwendet; für diese Verteilungen gibt es bereits Simulationsstudien in der Literatur, so daß die Güte verschiedener Testgrößen gegenüber bestimmten Alternativverteilungen verglichen werden kann.

Da ein bestimmter Test i. allg. nicht „gleichmäßig“ besser als ein anderer Test ist, sondern immer nur in bestimmten „Richtungen“ eine höhere Güte besitzt, ist es schwierig, die Ergebnisse solcher Simulationsstudien zu verallgemeinern. In speziellen Fällen zeigt sich aber oft, daß ein Test Alternativen mit bestimmten allgemeinen Merkmalen wie etwa „stark besetzte Verteilungsenden“ besonders gut (oder schlecht) erkennt.

Zur numerischen Berechnung der Testgrößen ist anzumerken, daß wegen (2.2) und  $\Psi_{\hat{\vartheta}_n}(0) = E_{\hat{\vartheta}_n}(X)$

$$T_n = \sqrt{n} \sup_{1 \leq k \leq M} \left| \bar{X}_n - E_{\hat{\vartheta}_n}(X) + \sum_{j=0}^{k-1} (F_n(j) - F(j, \hat{\vartheta}_n)) \right| \quad (2.12)$$

mit  $M = \max_{1 \leq j \leq n} X_j$  gilt. In den folgenden Beispielen ist  $\hat{\vartheta}_n$  stets der Momentenschätzer für  $EX$ , so daß  $E_{\hat{\vartheta}_n}(X) = \bar{X}_n$  gilt. Ähnlich folgt für die zweite Testgröße

$$\tilde{T}_n = \sum_{k=0}^M |Z_{n,k}| + \sqrt{n} E_{\hat{\vartheta}_n}(X) - \sqrt{n} \sum_{k=0}^M (1 - F(k, \hat{\vartheta}_n)).$$

Bei  $W_{mod}^2$  muß die unendliche Summe dagegen an geeigneter Stelle abgeschnitten werden.

### 2.1.13 Beispiel

Als erstes Beispiel betrachten wir den Test auf Poissonverteilung. Hierbei ist  $\Theta = \{\vartheta \in \mathbb{R} : \vartheta > 0\}$  und  $f(j, \vartheta) = e^{-\vartheta} \vartheta^j / j!$ ,  $j \geq 0$ . Der Schätzer  $\hat{\vartheta}_n = \bar{X}_n$  erfüllt (R1); wegen  $\partial F(k, \vartheta) / \partial \vartheta = -e^{-\vartheta} \vartheta^k / k!$  sieht man sofort, daß auch (R3) erfüllt ist. Nach Satz A.3.1a) in Anhang A.3 gilt für  $k+1 \geq \vartheta$

$$1 - F(k, \vartheta) < \frac{k+2}{k+2-\vartheta} f(k+1, \vartheta)$$

und somit für genügend großes  $l \in \mathbb{N}$  mit der Bezeichnung  $\vartheta_* = \inf_{n \geq n_0} \{\vartheta_n\}$

$$\sup_{n \geq n_0} \sum_{k \geq l} \sqrt{1 - F(k, \vartheta_n)} \leq 2 \sum_{k \geq l} \sqrt{f(k+1, \vartheta_*)}.$$

Folglich ist auch (R2) erfüllt.

Tabelle 2.1 zeigt die Güte des Tests auf Poissonverteilung bei mittleren und großen Stichprobenumfängen ( $n = 50$  bzw.  $n = 200$ ). Die Einträge in dieser und den folgenden Tabellen geben die relative Ablehnhäufigkeit in Prozent an. Dabei wurde unter Alternativen zur nächsten ganzen Zahl gerundet; unter der Hypothese ist jeweils eine Nachkommastelle mit angegeben. Die gleichen Alternativverteilungen und Stichprobenumfänge wurden

auch von Epps (1995), Tabelle VI, verwendet. Dabei ist zu beachten, daß einige Parameterwerte bei Epps nicht stimmen können und deshalb abgeändert wurden: bei den Neyman-Typ-A-Verteilungen  $NA(\phi_1, \phi_2)$  wurden die Parameter vertauscht; bei der Mischung aus einer Binomial- und einer Poissonverteilung  $PB(\phi_1, m, \phi_2)$  wurde statt  $\phi_1 = 0.1$  (dem Anteil der Poissonverteilung)  $\phi_1 = 0.9$  verwendet.

Als weitere Alternativen kommen in Tabelle 2.1 neben Standardverteilungen noch Mischungen aus Poisson- und logarithmischer Verteilung  $P \circ L$ , aus Poisson- und Binomialverteilung  $PB$  und aus Poisson- und negativer Binomialverteilung  $PNB$  vor. Die Verteilung  $P0(\beta, \vartheta)$  ist die in Bsp. 2.1.12 betrachtete Mischung aus einer Poissonverteilung mit Parameter  $\vartheta$  und einer Einpunktverteilung in 0. Genaue Definitionen der Verteilungen findet man in Epps (1995), S.1467.

Zu den von Epps betrachteten Tests ( $\chi^2$ -, Dispersions-, Smooth-Test sowie ein neuer, auf der empirischen erzeugenden Funktion basierender Test) ist generell zu bemerken, daß sie sämtlich nicht konsistent sind. Sie sind aber unter  $\mathcal{H}_0$  asymptotisch verteilungsfrei, so daß kritische Werte aus Tabellen abgelesen werden können.

Als Ergebnis läßt sich folgendes feststellen: Die drei auf  $T_n$ ,  $\tilde{T}_n$  bzw.  $W_{mod}^2$  basierenden Tests verhalten sich ähnlich, wobei der erste meistens besser ist. Dies zeigte sich auch bei den weiteren Simulationsstudien, so daß dort die Ergebnisse für  $\tilde{T}_n$  bzw.  $W_{mod}^2$  nicht mehr aufgeführt werden. Der auf  $T_n$  basierende Test ist in den meisten Fällen nicht wesentlich schlechter als der beste von Epps aufgeführte Test; insbesondere bei dem Stichprobenumfang  $n = 200$  ist er häufig der Test mit der größten Güte.

Tabelle 2.2 zeigt die Güte des EIVF-Tests auf Poissonverteilung bei kleinen Stichprobenumfängen ( $n = 20$  bzw.  $n = 31$ ). Dieselben Alternativverteilungen und Stichprobenumfänge wurden auch von Nakamura und Pérez-Abreu (1993), Tabelle 3, sowie von Epps (1995), Tabelle VI, verwendet. Neben Standardverteilungen werden die Zetaverteilung  $Z(\cdot)$  (siehe Johnson et al. (1992), S. 465) und eine Mischung aus zwei Poissonverteilungen verwendet.

Testgröße Verteilung	$T_n$		$\tilde{T}_n$		$W_{mod}^2$	
	$n = 50$	$n = 200$	$n = 50$	$n = 200$	$n = 50$	$n = 200$
$P(3)$	10.0	9.5	9.7	9.8	10.0	9.6
$P(7)$	10.1	10.1	9.7	10.1	9.5	10.2
$Bin(10, .2)$	25	66	20	57	20	54
$Bin(10, .5)$	94	100	89	100	87	100
$NB(5, .71)$	47	93	44	90	40	88
$NB(10, .5)$	93	100	92	100	88	100
$G(.3)$	44	89	43	88	41	87
$G(.9)$	100	100	100	100	100	100
$NA(5, .2)$	60	98	58	98	56	98
$NA(2, .5)$	23	53	22	49	20	47
$P \circ L(.5, .5)$	27	60	26	57	23	51
$P \circ L(.8, .2)$	14	21	14	20	13	17
$P0(.1, 3)$	42	90	44	90	46	94
$P0(.3, 1)$	41	91	42	91	41	91
$PB(.9, 2, .9)$	18	40	18	40	18	43
$PB(.9, 20, .9)$	17	36	16	36	17	38
$PNB(.1, 2, .5)$	86	100	84	100	80	100

Tabelle 2.1: Niveaueinhaltung und empirische Güte beim Test auf Poissonverteilung mit unbekanntem  $\vartheta$ ,  $\alpha = 0.10$ , 1000 Wiederholungen

Bei den von Nakamura und Pérez-Abreu verwendeten Testgrößen handelt es sich um den Dispersions- und den Smooth-Test, einen weiteren auf der erzeugenden Funktion aufbauenden Test, der von Kocherlakota und Kocherlakota (1986) vorgeschlagen wurde, und einen neuen Test, der speziell auf die Poissonverteilung zugeschnitten ist. Der letzte Test ist im Gegensatz zu den ersten drei konsistent; er ist nicht verteilungsfrei, doch wurde die ursprüngliche Testgröße so transformiert, daß die kritischen Werte nur noch wenig vom zugrundeliegenden Parameter  $\vartheta$  abhängen.

Die Ergebnisse von Tabelle 2.2 sind nur schwer mit den anderen genannten Simulationsstudien zu vergleichen. Dies liegt daran, daß die mei-

Verteilung	Niveau:	$n = 20$		$n = 31$	
		0.10	0.05	0.10	0.05
$P(3)$		10.7	4.0	10.7	3.1
$P(7)$		8.6	6.1	10.1	4.9
$Bin(6, 0.5)$		57	37	76	56
$Bin(3, 0.4)$		39	23	52	32
$NB(10, 0.7)$		27	17	36	28
$NB(1, 0.4)$		74	62	90	82
$U(0..4)$		14	8	17	8
$U(0..8)$		65	53	83	73
$L(0.6)$		56	44	70	57
$L(0.8)$		89	84	98	95
$Z(0.8)$		99	98	100	100
$\frac{1}{2}(P(2) + P(6))$		73	61	90	80

Tabelle 2.2: Niveaueinhaltung und empirische Güte beim Test auf Poissonverteilung mit unbekanntem  $\vartheta$ , 1000 Wiederholungen

sten Tests bei kleinen Stichprobenumfängen konservative Tests sind, d.h. ihr tatsächliches Niveau liegt unter dem angestrebten Niveau  $\alpha$ . Aus diesem Grunde werden die kritischen Werte empirisch so ermittelt, daß das Niveau genau eingehalten wird. Diese Anpassung steigert auch die Güte der Tests; in der praktischen Anwendung wird man davon aber gerade bei asymptotisch verteilungsfreien Tests keinen Gebrauch machen, da diese Anpassungen wieder Tabellen in Abhängigkeit von  $n$  und  $\vartheta$  erfordern.

Der Test von Nakamura und Pérez-Abreu ist dagegen für kleine Stichprobenumfänge stark antikonservativ; für  $n = 20$  und  $\alpha = 0.05$  beträgt sein tatsächliches Niveau etwa 0.08 (siehe Tabelle 2 in Nakamura und Pérez-Abreu (1993)). Verwendet man für diesen Test ein parametrisches Bootstrapverfahren, so hält der Test auch für  $n = 20$  das nominale Niveau gut ein. Die Güte sinkt dementsprechend deutlich, bei manchen Alternativverteilungen auf etwa zwei Drittel der in Tabelle 3 in Nakamura und Pérez-Abreu angegebenen Werte.



Hier läßt sich als Fazit für den vorliegenden Test nur sagen, daß er für Verteilungen mit stark besetzten Enden wie die  $L(.8)$ - und die  $Z(.8)$ -Verteilung sehr stark ist, dagegen die Gleichverteilung auf  $0, \dots, 4$ , für die wie bei der Poissonverteilung  $E(X) = Var(X)$  gilt, praktisch nicht erkennt. Das Beispiel der Zetaverteilung  $Z(.8)$  bestätigt die obige Aussage, daß der Test auch gegen Verteilungen, für die der Erwartungswert nicht existiert, konsistent ist.

Da bei sehr kleinen Stichprobenumfängen Verteilungen, deren Erwartungswert etwa gleich der Varianz ist, praktisch von keinem Test erkannt werden, ist in solchen Fällen der auf dem Dispersionsindex beruhende Test wegen seiner Einfachheit den anderen Tests vorzuziehen.

### 2.1.14 Beispiel

Als weitere Beispiele betrachten wir Tests auf geometrische Verteilung, logarithmische Verteilung bzw. positive Poissonverteilung.

Im Falle der geometrischen Verteilung ist  $\Theta = \{\vartheta \in \mathbb{R} : 0 < \vartheta < 1\}$  und  $f(j, \vartheta) = (1 - \vartheta)^j \vartheta$ ,  $j \geq 0$ . Der Maximum-Likelihood-Schätzer  $\hat{\vartheta}_n = (1 + \bar{X}_n)^{-1}$  erfüllt (R1). Wegen  $1 - F(k, \vartheta) = (1 - \vartheta)^{k+1}$  und  $\partial F(k, \vartheta) / \partial \vartheta = -(k+1)(1 - \vartheta)^k$  sieht man, daß auch (R2) und (R3) gelten.

Bei der logarithmischen Verteilung ist  $\Theta = \{\vartheta \in \mathbb{R} : 0 < \vartheta < 1\}$  und  $f(j, \vartheta) = a(\vartheta) \vartheta^j / j$ ,  $j \geq 1$ , mit  $a(\vartheta) = -(\ln(1 - \vartheta))^{-1}$ .

Der (eindeutige) Maximum-Likelihood-Schätzer  $\hat{\vartheta}_n$  ist Lösung der Gleichung  $\bar{X}_n = -\vartheta ((1 - \vartheta) \ln(1 - \vartheta))^{-1}$  und erfüllt (R1). Nach Satz A.3.1c) in Anhang A.3 gilt

$$1 - F(k, \vartheta) < \frac{a(\vartheta) \vartheta^{k+1}}{(k+1)(1 - \vartheta)}, \quad (2.13)$$

woraus (R2) folgt. Weiter gilt

$$\begin{aligned} \left| \frac{\partial F(k, \vartheta)}{\partial \vartheta} \right| &= \left| \frac{\partial(1 - F(k, \vartheta))}{\partial \vartheta} \right| \\ &= \left| \frac{a'(\vartheta)}{a(\vartheta)} \sum_{j>k} a(\vartheta) \frac{\vartheta^j}{j} + a(\vartheta) \sum_{j>k} \vartheta^{j-1} \right| \end{aligned}$$

$$< \frac{-a'(\vartheta)}{a(\vartheta)}(1 - F(k, \vartheta)) + \frac{a(\vartheta) \vartheta^k}{1 - \vartheta},$$

woran mit Blick auf (2.13) erkennbar ist, daß auch (R3) erfüllt ist.

Die positive Poissonverteilung erhält man aus der Poissonverteilung, indem  $f(0, \vartheta) = 0$  gesetzt und geeignet skaliert wird. Es ergibt sich  $f(j, \vartheta) = e^{-\vartheta} \vartheta^j / (j! \cdot (1 - e^{-\vartheta}))$ ,  $j \geq 1$ . Der Maximum-Likelihood-Schätzer  $\hat{\vartheta}_n$  ist Lösung der Gleichung  $\bar{X}_n = \vartheta(1 - e^{-\vartheta})^{-1}$  und erfüllt (R1). Da (R2) und (R3) für die Poissonverteilung erfüllt sind, gilt dies auch für die positive Poissonverteilung.

Tabelle 2.3 zeigt die Güte des EIVF-Tests bei mittleren und großen Stichprobenumfängen ( $n = 50$  bzw.  $n = 200$ ) bei den 3 oben aufgeführten hypothetischen Verteilungen. Diese Tabelle kann mit Tabelle VII in Epps (1995) verglichen werden. Dort wird neben dem neuen Test von Epps der  $\chi^2$ -Test und für die geometrische Verteilung auch der entsprechende glatte Anpassungstest verwendet.

Bei allen drei hypothetischen Verteilungen verhält sich der auf  $T_n$  basierende Test ähnlich wie der Test von Epps, welcher in dessen Studie der beste Test war; insbesondere für  $n = 50$  ist die Güte des Tests bei den vorliegenden Alternativen allerdings meist etwas geringer als bei dem auf der erzeugenden Funktion basierenden Test von Epps.

### 2.1.15 Beispiel

Abschließend soll an einem Beispiel die Durchführung des Tests auf Poissonverteilung gezeigt werden. Beobachtet wurde an  $n = 243$  Handelstagen, wie oft im Zeitraum zwischen 13.00 und 13.30 Uhr Aktien der Firma American Home Products gehandelt wurden (Epps (1995), S. 1477). An 33 Tagen fand kein Handel statt, an 55 Tagen wurde (genau) einmal gehandelt, usw. Der Vektor der absoluten Häufigkeiten findet sich in der viertletzten Zeile der nachstehend abgedruckten Implementation des Tests als S-Plus-Programm. Die Funktion `eidfstat` berechnet die EIVF-Test-Statistik nach Formel (2.12) (mit  $\bar{X}_n - E_{\hat{\vartheta}_n}(X) = 0$ ). In der Funktion `eidftest` wird der approximative p-Wert mit Hilfe des parametrischen Bootstrap-Verfahrens

$H_0$ : Geometrische Verteilung			$H_0$ : Logarithmische Verteilung		
Verteilung	$n = 50$	$n = 200$	Verteilung	$n = 50$	$n = 200$
$G(0.5)$	3.9	5.0	$L(.5)$	4.3	5.1
$NA(2, .5)$	13	56	$P^+(1.05)$	45	99
$NA(4, .5)$	69	100	$P^+(1.2)$	55	100
$NA(5, .2)$	41	98	$P^+(1.3)$	62	100
$NA(10, .2)$	92	100	$G^+ (.4)$	23	90
$P(.3)$	13	52	$G^+ (.33)$	35	97
$P(.5)$	26	90	$G^+ (.25)$	43	100
$P(.7)$	43	100	$Z(0.7)$	72	100
$P \circ L(.5, .7)$	11	36	$Z(1.0)$	40	100
$P \circ L(.5, .5)$	19	68	$Z(1.3)$	66	100
$P \circ L(.5, .3)$	22	83	$Z(2.0)$	40	89
$P \circ L(.3, .5)$	10	39			
$P0(.2, 1.)$	24	89	$H_0$ : Positive Poissonverteilung		
$P0(.2, 2.)$	59	100	Verteilung	$n = 50$	$n = 200$
$P0(.5, 1.)$	4	9	$P^+(3)$	4.4	4.2
			$Bin^+(10, .2)$	46	99
			$L(.3)$	16	47
			$L(.5)$	42	92
			$L(.7)$	87	100
			$G^+ (.4)$	84	100
			$G^+ (.33)$	95	100
			$NA(.5, 2.0)$	100	100
			$Z(1.0)$	100	100
			$Z(2.0)$	74	100

Tabelle 2.3: Niveaueinhaltung und empirische Güte bei verschiedenen hypothetischen Verteilungen,  $\alpha = 0.05$ , 1000 Wiederholungen

berechnet. Ein Programmlauf mit  $m = 1000$  Bootstrap-Wiederholungen ergab den p-Wert 0. Die beobachteten Daten sind folglich nicht mit der Poissonverteilung vereinbar. Dagegen ist die Unangemessenheit des Poissonmodells bei einem visuellen Vergleich der relativen Häufigkeiten mit den entsprechenden theoretischen Wahrscheinlichkeiten keineswegs offensichtlich.

```
eidfstat <- function(n, freq, cdf)  #computes eidf statistic
{
  edf <- cumsum(freq)/n  #empirical distribution function
  zsum <- cumsum(edf-cdf)
  tt <- sqrt(n)*max(abs(zsum))  #test statistic
  return(tt)
}
```

```
eidftest <- function(freq, m)  #parametric bootstrap
{
  n <- sum(freq)  #number of observations
  xmax <- length(freq) - 1  #largest observation
  xbar <- 0:xmax %*% freq / n  #sample mean
  cdf <- ppois(0:xmax, xbar)  #cumulative distr. function
  tn <- eidfstat(n, freq, cdf)
  count <- 0
  for (i in 1:m) {
    yp <- rpois( n, xbar)  #Poisson sample of size n
    ybar <- mean(yp)
    freq <- tabulate(yp+1)  #absolute frequencies
    cdf <- ppois(0:max(yp), max(ybar,10^-9))
    tn2 <- eidfstat(n, freq, cdf)
    if (tn2 >= tn) count <- count + 1 }
  return(count/m)
}
```

```
# data: number x of trades in American Home Products,  
# x=0,1,...,12; n=243; freq: absolute frequencies  
freq <- c(33,55,68,38,20,11,8,7,2,0,0,0,1)  
m <- 1000 #number of Bootstrap replications  
pval <- eidftest(freq, m)  
print(pval) #p-value
```

Zusammenfassend läßt sich über den hier eingeführten Test sagen: es ist ein Test mit hoher Güte. Der Test ist konsistent, was zumindest bei Vorliegen eines großen Stichprobenumfangs eine wünschenswerte Eigenschaft ist. Er ist zum Testen aller Hypothesen, die die Regularitätsvoraussetzungen von Satz 2.1.6 erfüllen, einsetzbar, im Gegensatz zu vielen Tests, die auf eine bestimmte Hypothese zugeschnitten sind. Die Realisierung des Tests mit Hilfe eines Computers ist nicht sonderlich aufwendig, wie der obige Programmausdruck zeigt. Dabei ist es einfach, ein Programm, das den Test für eine bestimmte Nullhypothese durchführt, so zu modifizieren, daß eine andere Hypothese getestet wird. Man benötigt lediglich Unterprogramme zur Berechnung der Zähldichte und der Schätzer. Die Erzeugung von Zufallszahlen kann dann durch Tabellenverfahren, die für alle diskreten Verteilungen verwendbar sind, durchgeführt werden. Solche Verfahren sind auch in Standard-Programmbibliotheken wie der NAG- oder der IMSL-Bibliothek implementiert.

## 2.2 Auf der empirischen integrierten Verteilungsfunktion basierende Tests für stetige Verteilungen

In diesem Abschnitt sollen auf der integrierten Verteilungsfunktion basierende Tests für die beiden wichtigsten stetigen Verteilungen, die Exponential- und die Normalverteilung, vorgestellt werden. Für die meisten anderen stetigen Verteilungen läßt sich kein geschlossener Ausdruck für die integrier-

te Verteilungsfunktion angeben, so daß auch entsprechende Testgrößen nur numerisch berechenbar sind. Es wird aber an geeigneter Stelle angedeutet, wie die erzielten theoretischen Resultate auf andere Verteilungen übertragen werden können.

### 2.2.1 Ein Test auf Exponentialverteilung

Da die Exponentialverteilung mit Parameter  $\vartheta > 0$  die Verteilungsfunktion  $F(t, \vartheta) = 1 - \exp(-\vartheta t)$ ,  $t \geq 0$ , besitzt, gilt

$$\Psi_{\vartheta}(t) = \frac{e^{-\vartheta t}}{\vartheta}, \quad t \geq 0.$$

Durch direkte Rechnung ergibt sich mit  $\Psi_n$  aus Def. 2.1.1b)

$$\begin{aligned} \int_0^{\infty} (\Psi_n(t) - \Psi_{\vartheta}(t))^2 dt &= \frac{5}{2\vartheta^3} - \frac{2}{n\vartheta^2} \sum_{i=1}^n \left( \frac{e^{-\vartheta X_i}}{\vartheta} + X_i \right) \\ &\quad - \frac{1}{3n^2} \sum_{i=1}^n (n-i-1)X_i^3 + \frac{1}{n^2} \sum_{i<j} X_i^2 X_j, \end{aligned} \quad (2.14)$$

wobei  $X_{(i)}$ ,  $1 \leq i \leq n$ , die  $i$ -te Ordnungsstatistik bezeichnet. Wird der Parameter  $\vartheta$  durch den Maximum-Likelihood-Schätzer  $\hat{\vartheta}_n = 1/\bar{X}_n$  geschätzt, so kann als Testgröße

$$T_n^* = \hat{\vartheta}_n^3 \int_0^{\infty} \left( \sqrt{n} \left( \Psi_n(t) - \Psi_{\hat{\vartheta}_n}(t) \right) \right)^2 dt$$

verwendet werden. Mit  $Y_i = \hat{\vartheta}_n X_i = X_i/\bar{X}_n$  und der Substitution  $u = \hat{\vartheta}_n t$  erhält man

$$T_n^* = \int_0^{\infty} (Z_n^*(u))^2 du$$

mit

$$Z_n^*(u) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{(Y_j - u)^+ - e^{-u}\}, \quad 0 \leq u < \infty; \quad (2.15)$$

aus (2.14) folgt die alternative Darstellung

$$\frac{T_n^*}{n} = \frac{1}{2} - \frac{2}{n} \sum_{i=1}^n e^{-Y_i} - \frac{1}{3n^2} \sum_{i=1}^n (n-i-1)Y_{(i)}^3 + \frac{1}{n^2} \sum_{i < j} Y_{(i)}^2 Y_{(j)}.$$

Da  $T_n^*$  eine Funktion von  $Y_1, \dots, Y_n$  ist, ist  $T_n^*$  unter  $H_0$  sowohl für endliches  $n$  als auch asymptotisch verteilungsfrei. Aus diesem Grunde wird im folgenden o.E. immer  $\vartheta = 1$  angenommen.

Für den Nachweis der schwachen Konvergenz von  $T_n^*$  ist es naheliegend, im Hilbertraum  $L_2 = L_2(\mathbb{R}_+, \mathcal{B}_+, \lambda)$  der quadrat-integrierbaren Funktionen auf  $\mathbb{R}_+$  zu arbeiten, da in diesem Fall  $T_n^*$  ein stetiges Funktional des Prozesses  $Z_n^*$  ist (das Skalarprodukt bzw. die zugehörige Norm werden mit  $(\cdot, \cdot)$  bzw. mit  $\|\cdot\|$  bezeichnet). Im Rahmen von Hilberträumen führte Neuhäus (1974) den Nachweis der Verteilungskonvergenz der Cramér-von Mises-Testgröße; gegenüber anderen Funktionenräumen hat dieses Vorgehen den Vorteil, daß für Summen von unabhängig und identisch verteilten Hilbertraumwertigen Zufallsvariablen  $X_1, X_2, \dots$  der zentrale Grenzwertsatz (genau dann) gilt, wenn die Varianz von  $\|X_1\|$  endlich ist; es sind also keine zusätzlichen Bedingungen wie etwa die „metrische Entropie-Bedingung“ im Raum der stetigen Funktionen nachzuweisen.

Für den Schätzer  $\hat{\vartheta}_n$  gilt die Darstellung

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = \frac{1}{\sqrt{n}} \sum_{j=1}^n l(X_j, \vartheta) + r_n \quad (2.16)$$

mit  $l(x, \vartheta) = -(\vartheta^2 x - \vartheta)$  und  $r_n = o_P(1)$ . Wir betrachten nun die Funktion

$$\begin{aligned} g(t, x) &= (x-t)^+ - \Psi_1(t) - l(x, 1) \left. \frac{\partial \Psi_\vartheta(t)}{\partial \vartheta} \right|_{\vartheta=1} \\ &= (x-t)^+ - (x-t+xt)e^{-t}. \end{aligned} \quad (2.17)$$

$g(\cdot, \cdot)$  erfüllt die folgenden Eigenschaften:

$$\begin{aligned} g(\cdot, x) &\in L_2 \quad \forall x \in \mathbb{R}_{\geq 0}, \\ E[g(t, X)] &= \int_0^\infty g(t, x) e^{-x} dx = 0 \quad \forall t \in \mathbb{R}_{\geq 0}, \\ E[\|g(\cdot, X)\|^2] &= E\left[\int_0^\infty g^2(t, X) dt\right] < \infty. \end{aligned} \quad (2.18)$$

Wir definieren eine Zufallsvariable mit Werten in  $L_2$  durch

$$W_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n g(\cdot, X_j).$$

$W_n$  ist eine Summe von unabhängigen und identisch verteilten Zufallselementen. Offensichtlich gilt  $E[(W_n, f)] = 0 \forall f \in L_2$  und  $Var(\|W_n\|) < \infty$ . Aus dem zentralen Grenzwertsatz in Hilberträumen folgt die Existenz eines Gauß'schen Zufallselements  $W$  in  $L_2$  mit

$$W_n \xrightarrow{\mathcal{D}} W \quad (2.19)$$

(siehe etwa Araujo und Giné (1980), Abschnitt 3.7). Entwickelt man  $\Psi_{\hat{\vartheta}_n}$  für festes  $t$  in eine Taylorreihe um  $\vartheta = 1$  und beachtet die Darstellung (2.16), so ergibt sich die folgende Gleichungskette:

$$\begin{aligned} & \sqrt{n} \left( \Psi_n(t) - \Psi_{\hat{\vartheta}_n}(t) \right) \\ &= \sqrt{n} \left( \Psi_n(t) - \Psi_1(t) \right) + \sqrt{n} \left( \Psi_1(t) - \Psi_{\hat{\vartheta}_n}(t) \right) \\ &= \sqrt{n} \left( \Psi_n(t) - \Psi_1(t) \right) - \sqrt{n} \left( \hat{\vartheta}_n - \vartheta \right) \frac{\partial \Psi_{\vartheta}(t)}{\partial \vartheta} \Big|_{\vartheta=\vartheta_n^*} \\ &= \sqrt{n} \left( \Psi_n(t) - \Psi_1(t) \right) - \frac{\partial \Psi_{\vartheta}(t)}{\partial \vartheta} \Big|_{\vartheta=1} \frac{1}{\sqrt{n}} \sum_{j=1}^n l(X_j, \vartheta) \\ & \quad + \sqrt{n} \left( \hat{\vartheta}_n - \vartheta \right) \left( \frac{\partial \Psi_{\vartheta}(t)}{\partial \vartheta} \Big|_{\vartheta=1} - \frac{\partial \Psi_{\vartheta}(t)}{\partial \vartheta} \Big|_{\vartheta=\vartheta_n^*} \right) - r_n \frac{\partial \Psi_{\vartheta}(t)}{\partial \vartheta} \Big|_{\vartheta=1}, \end{aligned}$$

wobei  $\vartheta_n^*$  einen Punkt zwischen  $\hat{\vartheta}_n$  und  $\vartheta$  bezeichnet. Es gilt also

$$\sqrt{n} \left( \Psi_n(t) - \Psi_{\hat{\vartheta}_n}(t) \right) = W_n(t) - R_n(t) \quad (2.20)$$

mit

$$R_n(t) = r_n \frac{\partial \Psi_{\vartheta}(t)}{\partial \vartheta} \Big|_{\vartheta=1} + \sqrt{n} \left( \hat{\vartheta}_n - \vartheta \right) \left( \frac{\partial \Psi_{\vartheta}(t)}{\partial \vartheta} \Big|_{\vartheta=\vartheta_n^*} - \frac{\partial \Psi_{\vartheta}(t)}{\partial \vartheta} \Big|_{\vartheta=1} \right).$$

Dabei ist es nicht schwer zu sehen, daß  $\|R_n\| \xrightarrow{P} 0$  gilt. Mit (2.19) und (2.20) erhält man aus Theorem 4.1 in Billingsley (1968) und dem Abbildungssatz das folgende Ergebnis:



**2.2.1 Satz**

Unter der Hypothese gilt

$$Z_n^* \xrightarrow{\mathcal{D}} W$$

in  $L_2$ ; dabei ist  $W$  ein zentrierter Gaußprozeß mit der Kovarianzfunktion

$$\begin{aligned} k(s, t) &= \text{Cov}(W(s), W(t)) = E[g(s, X_1) g(t, X_1)] \\ &= (t - s + 2) e^{-t} - (s + t + s t + 2) e^{-(s+t)}, \quad s \leq t. \end{aligned}$$

Weiter gilt

$$T_n^* = \|Z_n^*\|^2 \xrightarrow{\mathcal{D}} \|W\|^2.$$

**Bemerkung:** Der Beweis von Satz 2.2.1 zeigt, wie die Aussagen auf andere Verteilungen auf  $\mathbb{R}_+$  verallgemeinert werden können. Der Schätzer  $\hat{\vartheta}_n$  für den Parameter(-vektor)  $\vartheta$  muß eine Darstellung wie in (2.16) mit den üblichen Eigenschaften besitzen (siehe etwa (R1) in Abschnitt 2.1). Die Funktion  $g$  muß die obigen drei Eigenschaften (2.18) erfüllen. Schließlich muß  $\Psi_\vartheta$  differenzierbar nach  $\vartheta$  sein; die Ableitung muß gewissen Glattheitsforderungen genügen, so daß die Norm von  $R_n$  in (2.20) stochastisch gegen Null konvergiert. Dies soll nicht im Detail ausgeführt werden, da sich in den meisten Fällen keine explizite Formel für die Testgröße angeben läßt und somit der Test für solche Verteilungen weniger interessant ist.

Sind  $\lambda_j$  bzw.  $\varphi_j$ ,  $j \geq 1$ , die Eigenwerte bzw. Eigenfunktionen des zum Integralkern  $k(\cdot, \cdot)$  gehörenden Integraloperators, gilt also

$$\int_0^\infty k(s, t) \varphi_j(t) dt = \lambda_j \varphi_j(s), \quad 0 \leq s < \infty,$$

so ist die Verteilung von  $\|W\|^2$  bekanntlich identisch mit der Verteilung von  $\sum_{j \geq 1} \lambda_j N_j^2$ , wobei  $N_1, N_2, \dots$  unabhängige standardnormalverteilte Zufallsvariablen sind. Als wichtige Kenngrößen der Verteilung von  $\|W\|^2$  erhält man

$$E(\|W\|^2) = \int k(t, t) dt = \frac{1}{4},$$

$n$	$1 - \alpha$				
	0.5	0.9	0.95	0.975	0.99
10	0.115	0.374	0.494	0.668	1.127
20	0.124	0.452	0.623	0.863	1.397
50	0.135	0.525	0.734	0.990	1.466
100	0.141	0.557	0.775	1.028	1.438
200	0.146	0.573	0.797	1.043	1.405
500	0.150	0.584	0.805	1.040	1.376
1000	0.152	0.587	0.808	1.040	1.363
2000	0.152	0.587	0.805	1.037	1.360

Tabelle 2.4: Empirische kritische Werte von  $T_n^*$  zum Niveau  $\alpha$ 

$$\text{Var}(\|W\|^2) = 2 \iint k^2(s, t) ds dt = \frac{17}{216}.$$

Tabelle 2.4 zeigt die empirisch ermittelten kritischen Werte für verschiedene Werte des Testniveaus  $\alpha$  in Abhängigkeit vom Stichprobenumfang  $n$ . Dabei ist zu erkennen, daß  $T_n^*$  recht langsam gegen die Grenzverteilung konvergiert. Die Einträge in Tabelle 2.4 sind jeweils die 20%-gestutzten Mittel von 100 Monte-Carlo-Schätzungen, die ihrerseits auf jeweils 10000 Wiederholungen beruhen; dabei wurde immer  $\vartheta = 1$  gewählt.

Ist  $\alpha \in (0, 1)$  der Fehler 1. Art, so bezeichne  $z_n(\alpha)$  das  $(1 - \alpha)$ -Quantil von  $T_n^*$  unter  $H_0$ . Bezüglich der Konsistenz des Tests, der die Hypothese der Exponentialverteilung ablehnt, falls  $T_n^* > z_n(\alpha)$  ist, gilt das folgende Resultat.

### 2.2.2 Satz

*Der auf  $T_n^*$  basierende Test ist konsistent gegen jede Alternativverteilung mit positivem endlichen Erwartungswert.*

BEWEIS:  $\Psi_A$  bzw.  $\mu_A$  bezeichnen die integrierte Verteilungsfunktion bzw.

den Erwartungswert der Alternativverteilung. Weiter sei  $\vartheta = 1/\mu_A$ . Es ist

$$\|\Psi_n - \Psi_{\hat{\vartheta}_n}\| \geq \left| \|\Psi_n - \Psi_\vartheta\| - \|\Psi_\vartheta - \Psi_{\hat{\vartheta}_n}\| \right|. \quad (2.21)$$

Nach Lemma 2.1.2 existiert ein  $t > 0$  mit  $|\Psi_\vartheta(t) - \Psi_A(t)| \geq \delta > 0$ . Wegen  $\Psi_n(t) \rightarrow \Psi_A(t)$  *P*-f.s. folgt

$$\lim_{n \rightarrow \infty} |\Psi_n(t) - \Psi_\vartheta(t)| \geq \delta \quad P\text{-f.s.}$$

und somit

$$\liminf_{n \rightarrow \infty} \int_0^\infty (\Psi_n(t) - \Psi_\vartheta(t))^2 dt > 0 \quad P\text{-f.s.}$$

Andererseits gilt

$$\lim_{n \rightarrow \infty} \|\Psi_\vartheta - \Psi_{\hat{\vartheta}_n}\| = 0 \quad P\text{-f.s.},$$

so daß sich mit (2.21)  $\lim_{n \rightarrow \infty} \|\Psi_n - \Psi_{\hat{\vartheta}_n}\|^2 > 0$  *P*-f.s. ergibt. Hieraus folgt unter der Alternative

$$\lim_{n \rightarrow \infty} T_n^* = \infty \quad P\text{-f.s.}$$

und somit  $\lim_{n \rightarrow \infty} P(T_n^* \leq z_n(\alpha)) = 0$ . ■

Schnellere Konvergenz der Verteilung der Teststatistik gegen die Grenzverteilung ergibt sich, wenn man ein gewichtetes Integral verwendet. Eine naheliegende Gewichtsfunktion ist  $\exp(-a\vartheta t)$  mit  $a \geq 0$ . Wir verwenden als modifizierte Testgröße

$$T_{n,a}^* = (a\hat{\vartheta}_n)^3 \int_0^\infty \left( \sqrt{n} \left( \Psi_n(t) - \Psi_{\hat{\vartheta}_n}(t) \right) \right)^2 \exp(-a\hat{\vartheta}_n t) dt,$$

woraus man wie oben

$$T_{n,a}^* = a^3 \int_0^\infty (Z_n^*(t))^2 e^{-at} dt$$

mit dem Prozeß  $Z_n^*$  aus (2.15) erhält.

**Bemerkung:** Die Anwendung eines Satzes über Laplace-Transformierte (siehe Widder (1959), S. 182, oder auch Baringhaus et al. (1998), S.5) ergibt

$$\lim_{a \rightarrow \infty} a^3 \int_0^{\infty} (Z_n^*(t))^2 \exp(-at) dt = 2n.$$

Die Testgröße  $T_{n,a}^*$  konvergiert also für festes  $n$  bei  $a \rightarrow \infty$  gegen eine Konstante.

**2.2.3 Lemma** Die Testgröße  $T_{n,a}^*$  besitzt folgende Darstellung:

$$\begin{aligned} \frac{T_{n,a}^*}{n} &= \frac{2(3a+2)}{(2+a)(1+a)^2} - \frac{2a^3}{n} \sum_{i=1}^n \frac{e^{-(1+a)Y_i}}{(1+a)^2} \\ &\quad - \frac{2}{n^2} \sum_{i=1}^n e^{-aY_i} + \frac{2}{n^2} \sum_{i < j} (a(Y_{(j)} - Y_{(i)}) - 2) e^{-aY_{(i)}}. \end{aligned}$$

BEWEIS: Aus der Definition

$$T_{n,a}^* = na^3 \int_0^{\infty} \left( \frac{1}{n} \sum_{j=1}^n (Y_j - t)^+ - e^{-t} \right)^2 e^{-at} dt$$

erhält man durch direkte Rechnung

$$\begin{aligned} \frac{T_{n,a}^*}{n} &= \frac{a^3}{2+a} + \frac{2a^3}{(1+a)^2} - \frac{2a^3}{n} \sum_{i=1}^n \frac{e^{-(1+a)Y_i} + (1+a)Y_i}{(1+a)^2} \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n (a^2 Y_i^2 - 2aY_i + 2 - 2e^{-aY_i}) \\ &\quad + \frac{2}{n^2} \sum_{i < j} (a^2 Y_{(i)} Y_{(j)} - a(Y_{(i)} + Y_{(j)}) + 2 + (a(Y_{(j)} - Y_{(i)}) - 2) e^{-aY_{(i)}}). \end{aligned}$$

Faßt man die zweite Summe und die in  $i$  und  $j$  symmetrischen Terme der dritten Summe zusammen und beachtet, daß  $1/n \sum_{i=1}^n Y_i = 1$  gilt, so folgt die obige Darstellung. ■

Der Nachweis der Verteilungskonvergenz

$$T_{n,a}^* \xrightarrow{D} a^3 \int_0^\infty W^2(t) e^{-at} dt$$

erfordert gegenüber den bisherigen Überlegungen gewisse Modifikationen; da ein gewichtetes Integral auch im nächsten Abschnitt als Testgröße verwendet wird, sollen die notwendigen Änderungen hier nicht ausgeführt werden. Tabelle 2.5 zeigt die empirisch ermittelten  $(1 - \alpha)$ -Quantile der Verteilung von  $T_{n,a}^*$  für  $a = 1, 5, 10, 20$  und verschiedene Werte von  $\alpha$  in Abhängigkeit vom Stichprobenumfang  $n$ .

Die Quantile konvergieren deutlich schneller gegen die Quantile der Grenzverteilung als dies bei  $T_n^*$  der Fall ist. Die Einträge in Tabelle 2.5 wurden auf gleiche Weise wie in Tabelle 2.4 ermittelt.

Es ist klar, daß die Güte des Tests gegenüber einer bestimmten Alternativverteilung von der Wahl der Gewichtsfunktion abhängt. Um dies genauer zu untersuchen, wurde eine Gütestudie durchgeführt.

Folgende Alternativen wurden dabei verwendet: Gamma-, Weibull- bzw. Lognormalverteilung mit Skalenparameter 1 und Formparameter  $\alpha$  ( $(, (\alpha)$ ),  $W(\alpha)$  bzw.  $LN(\alpha)$ ), Gleichverteilung auf dem Intervall  $(0,1)$  ( $U(0,1)$ ), Half-normal ( $HN$ ), Half-Cauchy ( $HC$ ),  $\chi_1^2$ -Verteilung.

Neben dem auf  $T_n^*$  beruhenden Test wurden die Tests mit Gewichten  $a = 1, 5, 10$  und  $a = 20$  verwendet.

Aus den ersten fünf Spalten von Tabelle 2.6 wird ersichtlich, daß die Güte teilweise sehr stark von  $a$  abhängt (dabei ist  $\varphi_a$  bzw.  $\varphi_0$  der auf  $T_{n,a}^*$  bzw.  $T_n^*$  beruhende Test, s.u.). Zu jedem der fünf verwendeten Tests existieren bestimmte Alternativen, bei denen dieser Test die höchste Güte besitzt. Aus diesem Grunde wurden weitere Testgrößen gebildet, die bessere omnibus-Tests ergeben sollten: eine Maximum-Testgröße wie in Baringhaus und Henze (1992c) ergab keinen omnibus-Test und ist außerdem aufwendig zu berechnen; die Ergebnisse sind nicht aufgeführt.

Weiter wurden Kombinationen aus zwei oder mehr Testgrößen  $T_{n,a_j}^*$  ( $j = 1, \dots, k$ ) gebildet;  $H_0$  wird abgelehnt, wenn mindestens einer der auf  $T_{n,a_j}^*$  beruhenden Tests die Hypothese ablehnt.

	$n$	$1 - \alpha$				
		0.5	0.9	0.95	0.975	0.99
a=1	10	0.021	0.082	0.109	0.136	0.178
	20	0.021	0.086	0.118	0.150	0.196
	50	0.021	0.089	0.122	0.157	0.205
	100	0.021	0.089	0.124	0.160	0.207
	200	0.021	0.090	0.125	0.161	0.210
a=5	10	0.070	0.302	0.408	0.511	0.650
	20	0.068	0.303	0.416	0.531	0.690
	50	0.067	0.304	0.422	0.543	0.708
	100	0.066	0.304	0.423	0.547	0.715
	200	0.066	0.304	0.426	0.549	0.718
a=10	10	0.060	0.248	0.321	0.394	0.570
	20	0.058	0.256	0.349	0.444	0.585
	50	0.057	0.262	0.364	0.469	0.615
	100	0.057	0.263	0.366	0.472	0.619
	200	0.056	0.264	0.368	0.477	0.630
a=20	10	0.042	0.125	0.197	0.301	0.462
	20	0.039	0.165	0.217	0.294	0.444
	50	0.038	0.174	0.241	0.312	0.421
	100	0.038	0.178	0.248	0.322	0.427
	200	0.038	0.179	0.250	0.326	0.429

Tabelle 2.5: Empirische kritische Werte von  $T_{n,a}^*$  zum Niveau  $\alpha$  für verschiedene Werte von  $a$ .

Formal läßt sich der Test folgendermaßen beschreiben:  $\varphi_{a,\alpha} = 1\{T_{n,a}^* > z_{n,a}(\alpha)\}$  bzw.  $\varphi_{0,\alpha} = 1\{T_n^* > z_n(\alpha)\}$  sei der auf  $T_{n,a}^*$  bzw.  $T_n^*$  beruhende Test zum Niveau  $\alpha$ ; die Tests liefern den Testausgang 1 (ablehnen) bzw. 0 (nicht ablehnen). Dann definiert man den kombinierten Test zum Niveau  $\alpha$  durch

$$\varphi_{a_1, \dots, a_k, \alpha} := \max_{j=1, \dots, k} \varphi_{a_j, \alpha^*},$$

wobei  $\alpha/k \leq \alpha^* \leq \alpha$  gilt.  $\alpha^*$  ist durch die Forderung  $E[\varphi_{a_1, \dots, a_k, \alpha}] = \alpha$

Verteilung	$\varphi_0$	$\varphi_1$	$\varphi_5$	$\varphi_{10}$	$\varphi_{20}$	$\varphi_{1,10}$	$\varphi_{0,20}$	$\varphi_{0,5,20}$
, (0.4)	66	72	84	89	92	88 (87)	91	90
, (0.6)	30	31	41	47	53	46 (44)	52	50
, (0.8)	11	10	12	15	18	15 (14)	18	17
, (1.0)	5.0	5.0	5.0	5.0	5.0	5.0 (4.2)	5.0	5.0
, (1.5)	10	18	22	18	8	16 (14)	2	11
, (2.0)	28	45	54	47	24	44 (40)	9	35
, (2.4)	46	66	76	69	40	67 (62)	19	57
, (3.0)	70	87	93	89	63	88 (85)	38	82
<i>W</i> (0.4)	96	98	99	100	100	100 (99)	100	100
<i>W</i> (0.6)	67	70	77	80	81	81 (79)	83	82
<i>W</i> (0.8)	23	22	24	26	29	28 (26)	32	31
<i>W</i> (1.4)	21	34	36	27	12	30 (26)	6	20
<i>W</i> (1.6)	45	62	63	49	25	56 (51)	17	43
<i>W</i> (2.0)	86	94	93	83	53	92 (89)	56	83
<i>LN</i> (0.6)	54	76	95	96	83	92 (89)	28	85
<i>LN</i> (0.8)	15	23	42	47	28	35 (31)	6	27
<i>LN</i> (1.0)	19	15	11	10	5	15 (13)	15	16
<i>LN</i> (1.2)	38	33	20	14	7	30 (28)	34	33
<i>LN</i> (1.5)	68	66	59	53	44	65 (64)	66	66
<i>U</i> (0, 1)	68	73	45	25	10	62 (58)	35	33
<i>HN</i>	14	22	18	11	5	16 (14)	3	9
<i>HC</i>	72	69	59	52	46	67 (66)	70	69
$\chi_1^2$	46	50	63	70	75	69 (67)	74	73

Tabelle 2.6: Empirische Güte verschiedener Tests auf Exponentialverteilung,  $\alpha = 0.05$ ,  $n = 20$ , 100000 Wiederholungen

eindeutig festgelegt; in der Praxis müssen die entsprechenden Quantile durch einen Suchalgorithmus empirisch bestimmt werden. Verwendet man die zu  $\alpha/k$  gehörenden Quantile, so erhält man einen konservativen Test.

Als erste Kombination wurde  $\varphi_{1,10,\alpha}$  untersucht. Für  $n = 20$  und  $\alpha = 0.1$  erhält man die Quantile 0.108 für  $T_{n,1}^*$  und 0.321 für  $T_{n,10}^*$ ; für  $\alpha = 0.05$  erhält man 0.141 und 0.421. Da die Quantile nahe bei den  $(1 - \alpha/2)$ -

Quantilen der ursprünglichen Testgrößen liegen, wurden auch Tests mit diesen durchgeführt; die Resultat sind in der sechsten Spalte von Tabelle 2.6 in Klammer angegeben (dabei wurde der Index  $\alpha$  bei den Tests immer weggelassen). Weiter sind die Ergebnisse für die Kombination  $\varphi_{0,20,\alpha}$  aufgeführt; die Quantile sind 0.601 bzw. 0.211 bei  $\alpha = 0.1$  und 0.845 bzw. 0.285 bei  $\alpha = 0.05$ . In der letzten Spalte findet man die Resultate für die Kombination  $\varphi_{0,5,20,\alpha}$ ; die entsprechenden Quantile sind 0.646, 0.434 bzw. 0.224 ( $\alpha = 0.1$ ) und 0.937, 0.566 bzw. 0.321 ( $\alpha = 0.05$ ).

Das Ergebnis ist ganz eindeutig:  $\varphi_{1,10,\alpha}$  ist ein echter omnibus-Test und den anderen aufgeführten Tests vorzuziehen, wenn nichts über die Alternative bekannt ist.

Gegenüber  $\varphi_{0,5,20,\alpha}$  hat dieser Test den weiteren Vorteil, daß die  $(1 - \alpha/2)$ -Quantile mit geringem Güteverlust verwendbar sind.

Die Ergebnisse in Tabelle 2.6 können mit Teilen von Tabelle 4 (bzw. Tabelle 2) in Baringhaus und Henze (1991) (bzw. (1992c)) verglichen werden. Dabei zeigt sich, daß der neue Test bei vielen Alternativen eine vergleichsweise hohe Güte besitzt. Allerdings gibt es bei der Vielzahl guter Tests auf Exponentialverteilung keinen, der allen anderen „in den meisten Fällen“ überlegen ist. Eine Wertung der Simulationsergebnisse soll daher dem Leser überlassen bleiben.

Die Güte des Tests ließe sich sicherlich weiter verbessern, wenn es gelänge, die Gewichtsfunktion in geeigneter Weise adaptiv zu wählen. Dies soll Gegenstand zukünftiger Untersuchungen sein.

### 2.2.2 Ein Test auf Normalverteilung

Für Zufallsvariablen auf  $\mathbb{R}$  ist die folgende Definition einer integrierten Verteilungsfunktion üblich. Dabei bezeichne  $y^-$  den Negativteil von  $y$ , es gelte also  $y^- = -\min\{y, 0\}$ .



**2.2.4 Definition**

- a) Es sei  $X$  eine reelle Zufallsvariable mit endlichem Erwartungswert. Dann ist die integrierte Verteilungsfunktion  $\psi_X : \mathbb{R} \rightarrow \mathbb{R}$  definiert durch

$$\psi_X(t) := E(X - t)^- = \int_{-\infty}^t F_X d\lambda.$$

- b)  $X_1, X_2, \dots, X_n$  seien unabhängige und identisch verteilte Zufallsvariablen auf  $\mathbb{R}$ . Dann ist die empirische integrierte Verteilungsfunktion gegeben durch

$$\psi_n(t) = \int_{-\infty}^t F_n d\lambda = -\frac{1}{n} \sum_{i=1}^n (X_i - t) \mathbf{1}\{X_i < t\}.$$

Ähnlich wie in Lemma 2.1.2 gelten für  $\psi_X$  die folgenden Eigenschaften (Müller (1998)):

**2.2.5 Lemma** a) Ist  $X$  eine reelle Zufallsvariable mit endlichem Erwartungswert, so gilt für die integrierte Verteilungsfunktion:

- (i)  $\lim_{t \rightarrow -\infty} \psi_X(t) = 0$ .
- (ii)  $\psi_X$  ist monoton wachsend und konvex.
- (iii)  $\lim_{t \rightarrow \infty} (\psi_X(t) - t) = -EX$ .

b) Zu jeder Funktion  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  mit den Eigenschaften aus a) gibt es genau eine Verteilung auf  $\mathbb{R}$ , deren integrierte Verteilungsfunktion  $\psi$  ist. Die zugehörige Verteilungsfunktion ist dann  $F(t) = D^+ \psi(t)$ .

Im folgenden soll die integrierte Verteilungsfunktion zum Testen der Hypothese  $\mathcal{H}_0 : \mathcal{V}(X) \in \mathcal{N} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$  verwendet werden.

Die Verteilungsfunktion bzw. die Dichte der Normalverteilung  $\mathcal{N}(\mu, \sigma^2)$  bezeichnen wir wie üblich mit  $\Phi_{\mu, \sigma^2}$  bzw.  $\varphi_{\mu, \sigma^2}$ . Weiter sei  $\Phi = \Phi_{0,1}$  und

$\varphi = \varphi_{0,1}$ . Gilt  $X \sim \mathcal{N}(\mu, \sigma^2)$ , so schreiben wir statt  $\psi_X$  auch  $\psi_{\mu, \sigma^2}$ . Durch Differentiation überzeugt man sich leicht von der Darstellung

$$\psi_{\mu, \sigma^2}(x) = (x - \mu) \Phi_{\mu, \sigma^2}(x) + \sigma^2 \varphi_{\mu, \sigma^2}(x), \quad x \in \mathbb{R}. \quad (2.22)$$

Aus Lemma 2.2.5a)(iii) folgt

$$\lim_{t \rightarrow \infty} (\psi_n(t) - \psi_{\mu, \sigma^2}(t)) = \mu - \bar{X}_n,$$

so daß ein Integralkriterium ohne Gewichtungsfunktion wie bei der Exponentialverteilung nicht verwendet werden kann. Eine geeignete Testgröße ist

$$\tilde{T}_n = \frac{n}{\hat{\sigma}_n^2} \int_{-\infty}^{\infty} \left( \psi_n(t) - \psi_{\bar{X}_n, \hat{\sigma}_n^2}(t) \right)^2 \varphi_{\bar{X}_n, \hat{\sigma}_n^2}(t) dt.$$

Setzt man  $Y_i = (X_i - \bar{X}_n)/\hat{\sigma}_n$ ,  $i = 1, \dots, n$ , mit  $\hat{\sigma}_n^2 = n^{-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2$ , führt die Substitution  $u = (t - \bar{X}_n)/\hat{\sigma}_n$  durch und beachtet (2.22), so erhält man

$$\begin{aligned} \tilde{T}_n &= n \int_{-\infty}^{\infty} \left( \frac{1}{n} \sum_{j=1}^n (Y_j - u)^- - (u \Phi(u) + \varphi(u)) \right)^2 \varphi(u) du \\ &= \int_{-\infty}^{\infty} \tilde{Z}_n^2(u) \varphi(u) du, \end{aligned} \quad (2.23)$$

wobei  $\tilde{Z}_n$  durch

$$\tilde{Z}_n(u) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{ (Y_j - u)^- - (u \Phi(u) + \varphi(u)) \}, \quad u \in \mathbb{R},$$

definiert wird.

**2.2.6 Lemma** Die Testgröße  $\tilde{T}_n$  besitzt folgende explizite Darstellung:

$$\begin{aligned} \tilde{T}_n &= \frac{n}{3} + \frac{n\sqrt{3}}{2\pi} \\ &\quad - \sum_{i=1}^n \left( 1 - \Phi^2(Y_i) - \frac{2}{\sqrt{\pi}} Y_i (1 - \Phi(\sqrt{2} Y_i)) + 2\varphi^2(Y_i) \right) \\ &\quad + \frac{1}{n} \sum_{i,j=1}^n \left( (1 - \Phi(Y_i \vee Y_j)) (1 + Y_i Y_j) - (Y_i \wedge Y_j) \varphi(Y_i \vee Y_j) \right), \end{aligned}$$

wobei  $x \vee y$  für  $\max(x, y)$  und  $x \wedge y$  für  $\min(x, y)$  steht.

BEWEIS: Aus (2.23) folgt

$$\begin{aligned}
 \frac{\tilde{T}_n}{n} &= \frac{1}{n^2} \sum_{i,j=1}^n \int_{\max(Y_i, Y_j)}^{\infty} (u - Y_i) (u - Y_j) \varphi(u) du \\
 &\quad - \frac{2}{n} \sum_{j=1}^n \int_{Y_j}^{\infty} (u \Phi(u) + \varphi(u)) (u - Y_j) \varphi(u) du \\
 &\quad + \int_{-\infty}^{\infty} (u \Phi(u) + \varphi(u))^2 \varphi(u) du \\
 &=: \frac{1}{n^2} \sum_{i,j=1}^n I_1(Y_i, Y_j) - \frac{2}{n} \sum_{j=1}^n I_2(Y_j) + I_3. \tag{2.24}
 \end{aligned}$$

Wegen

$$\int x \varphi(x) dx = -\varphi(x), \quad \int x^2 \varphi(x) dx = -x \varphi(x) + \Phi(x),$$

(wobei hier  $\int f(x) dx$  eine Stammfunktion von  $f$  bezeichnet), gilt für das erste Integral im Falle  $Y_i < Y_j$

$$I_1(Y_i, Y_j) = 1 - \Phi(Y_j) + (1 - \Phi(Y_j)) Y_i Y_j - Y_i \varphi(Y_j).$$

Weiter gilt

$$\begin{aligned}
 \int \varphi^2(x) dx &= \frac{1}{2\sqrt{\pi}} \Phi(\sqrt{2}x), \\
 \int x \varphi^2(x) dx &= -\frac{1}{2} \varphi^2(x), \\
 \int x \Phi(x) \varphi(x) dx &= \frac{1}{2\sqrt{\pi}} \Phi(\sqrt{2}x) - \varphi(x) \Phi(x), \\
 \int x^2 \Phi(x) \varphi(x) dx &= \frac{1}{2} \Phi^2(x) - x \varphi(x) \Phi(x) - \frac{1}{2} \varphi^2(x),
 \end{aligned}$$

und somit

$$I_2(Y_j) = \frac{1}{2} (1 - \Phi^2(Y_j)) - \frac{1}{\sqrt{\pi}} Y_j (1 - \Phi(\sqrt{2}Y_j)) + \varphi^2(Y_j).$$

Schließlich ergibt sich mit Hilfe von

$$\begin{aligned} \int \varphi^3(x) dx &= \frac{\sqrt{3}}{6\pi} \Phi(\sqrt{3}x), \\ \int x \Phi(x) \varphi^2(x) dx &= -\frac{1}{2} \varphi^2(x) \Phi(x) + \frac{\sqrt{3}}{12\pi} \Phi(\sqrt{3}x), \\ \int x^2 \Phi^2(x) \varphi(x) dx \\ &= \frac{1}{3} \Phi^3(x) - x \varphi(x) \Phi^2(x) - \varphi^2(x) \Phi(x) + \frac{\sqrt{3}}{6\pi} \Phi(\sqrt{3}x) \end{aligned}$$

für das letzte Integral das Ergebnis  $I_3 = 1/3 + \sqrt{3}/(2\pi)$ .

Setzt man alle Resultate in (2.24) ein, so folgt die Behauptung. ■

Als Funktion von  $Y_1, \dots, Y_n$  ist  $\tilde{T}_n$  unter  $\mathcal{H}_0$  sowohl finit als auch asymptotisch verteilungsfrei, weshalb im folgenden o.E. immer  $\mu = 0$  und  $\sigma^2 = 1$  angenommen wird.

Seien  $\vartheta = (\mu, \sigma^2)$  und  $P_\vartheta$  das zur Verteilungsfunktion  $\Phi_{\mu, \sigma^2}$  gehörende Maß auf  $(\mathbb{R}, \mathcal{B})$ . Ähnlich wie im vorherigen Abschnitt verwenden wir den Hilbertraum  $\tilde{L}_2 = L_2(\mathbb{R}, \mathcal{B}, P_\vartheta)$  der bzgl.  $P_\vartheta$  quadrat-integrierbaren Funktionen auf  $\mathbb{R}$ , um die schwache Konvergenz von  $\tilde{T}_n$  nachzuweisen. Bezeichnen  $(\cdot, \cdot)$  und  $\|\cdot\|$  das Skalarprodukt und die Norm in  $\tilde{L}_2$ , so gilt  $\tilde{T}_n = \|\tilde{Z}_n\|^2$ ; insbesondere ist  $\tilde{T}_n$  ein stetiges Funktional von  $\tilde{Z}_n$ .

Für den Schätzer  $\hat{\vartheta}_n = (\bar{X}_n, \hat{\sigma}_n^2)$  von  $\vartheta$  gilt die Darstellung (2.16) mit  $l(x, \vartheta) = (x - \mu, (x - \mu)^2 - \sigma^2)$ . Die zu  $g$  in (2.17) analoge Funktion

$$\tilde{g}(t, x) = (x - t)^- - \Psi_{(0,1)}(t) - l(x, (0, 1)) \nabla_{\vartheta} \Psi_{\vartheta}(t)|_{\vartheta=(0,1)}$$

besitzt wegen

$$\frac{\partial \psi_{\vartheta}(t)}{\partial \mu} = -\Phi_{\mu, \sigma^2}(t), \quad \frac{\partial \psi_{\vartheta}(t)}{\partial \sigma^2} = \frac{1}{2} \varphi_{\mu, \sigma^2}(t)$$

die Gestalt

$$\tilde{g}(t, x) = (x - t)^- + (x - t) \Phi(t) - \frac{1}{2}(x^2 + 1) \varphi(t).$$

$\tilde{g}(\cdot, \cdot)$  erfüllt die folgenden Eigenschaften:

$$\begin{aligned}\tilde{g}(\cdot, x) &\in \tilde{L}_2 \quad \forall x \in \mathbb{R}, \\ E[\tilde{g}(t, X)] &= \int_{-\infty}^{\infty} \tilde{g}(t, x) \varphi(x) dx = 0 \quad \forall t \in \mathbb{R}, \\ E[\|\tilde{g}(\cdot, X)\|^2] &= E\left[\int_{-\infty}^{\infty} \tilde{g}^2(t, X) \varphi(t) dt\right] < \infty.\end{aligned}$$

### 2.2.7 Satz

a) Sei  $\tilde{W}$  der zentrierte Gaußprozeß mit Kovarianzfunktion

$$\begin{aligned}\tilde{k}(s, t) &= E[\tilde{g}(s, X_1) \tilde{g}(t, X_1)] \\ &= \Phi(s)(1 + st) + t\varphi(s) - 3\varphi(s)\varphi(t)/2 \\ &\quad - t\Phi(t)\varphi(s) - s\Phi(s)\varphi(t) - \Phi(s)\Phi(t)(1 + st)\end{aligned}$$

für  $s \leq t$ . Dann gilt unter der Hypothese  $\mathcal{H}_0$

$$\tilde{Z}_n \xrightarrow{\mathcal{D}} \tilde{W}$$

in  $\tilde{L}_2$ .

b) Unter  $\mathcal{H}_0$  gilt

$$\tilde{T}_n \xrightarrow{\mathcal{D}} \|\tilde{W}\|^2.$$

BEWEIS: a) Definiert man eine Zufallsvariable mit Werten in  $\tilde{L}_2$  durch

$$\tilde{W}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n \tilde{g}(\cdot, X_j)$$

und geht im weiteren analog wie im letzten Abschnitt vor (insbesondere sind die dort auftretenden partiellen Ableitungen durch Gradienten zu ersetzen), so erhält man die Konvergenzaussage von Teil a). Die Kovarianzfunktion ergibt sich durch direkte Rechnung.

$n$	$1 - \alpha$				
	0.5	0.9	0.95	0.975	0.99
10	.0086	.0233	.0300	.0371	.0468
15	.0085	.0239	.0310	.0384	.0485
20	.0085	.0244	.0316	.0392	.0494
30	.0085	.0247	.0321	.0397	.0502
50	.0085	.0249	.0324	.0401	.0507
100	.0084	.0250	.0325	.0403	.0507
200	.0084	.0251	.0327	.0405	.0512
500	.0084	.0252	.0328	.0406	.0509
1000	.0084	.0252	.0328	.0406	.0511

Tabelle 2.7: Empirische kritische Werte von  $\tilde{T}_n$  zum Niveau  $\alpha$ 

b) Eine Taylorentwicklung von  $\varphi_{\vartheta_n}(t)$  an der Stelle  $\vartheta$  zeigt, daß

$$\tilde{T}_n - \|\tilde{Z}_n\|^2 \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

gilt. Daraus folgt die Behauptung mit Hilfe des Lemmas von Slutsky, wenn man beachtet, daß nach Teil a) und dem Abbildungssatz  $\|\tilde{Z}_n\|^2 \xrightarrow{D} \|\tilde{W}\|^2$  gilt. ■

Tabelle 2.7 zeigt die empirisch ermittelten kritischen Werte für verschiedene Werte von  $\alpha$  in Abhängigkeit vom Stichprobenumfang  $n$ . Die Einträge in Tabelle 2.7 sind jeweils die 20%-gestutzten Mittel von 100 Monte-Carlo-Schätzungen, die ihrerseits auf jeweils 10000 Wiederholungen beruhen; dabei wurde immer  $\vartheta = (0, 1)$  gewählt.

$\tilde{z}_n(\alpha)$  sei das  $(1 - \alpha)$ -Quantil von  $\tilde{T}_n$  unter  $H_0$ . Mit Hilfe von Lemma 2.2.5 läßt sich wie im letzten Abschnitt zeigen, daß der Test, der die Hypothese der Normalverteilung ablehnt, falls  $\tilde{T}_n > \tilde{z}_n(\alpha)$  ist, gegen jede Alternativverteilung mit positiver endlicher Varianz konsistent ist.

Um die Güte des auf  $\tilde{T}_n$  beruhenden Tests für endliche Stichprobenumfänge beurteilen zu können, wurde eine Simulationsstudie mit zahlreichen Alternativen zur Normalverteilung durchgeführt. Die Zufallszahlen wurden

Testgröße $b$	$\tilde{T}_n$			$W$		
	0.0	0.5	1.0	0.0	0.5	1.0
$a = 2.0$	5.0	5.0	5.0	5.1	5.1	5.0
$a = 1.8$	38.9	41.1	47.8	37.5	39.7	47.0
$a = 1.6$	69.1	72.6	82.1	65.8	69.6	81.3
$a = 1.4$	88.7	90.9	96.6	85.7	88.4	96.4
$a = 1.2$	97.3	98.1	99.6	95.9	97.1	99.6
$a = 1.0$	99.7	99.8	100.0	99.4	99.6	100.0

Tabelle 2.8: Empirische Güte bei der Alternative „stable distributions“, Stichprobenumfang  $n = 50$ , Niveau  $\alpha = 0.05$ , 100000 Wiederholungen

dabei mit Hilfe von Routinen aus der IMSL-Unterprogramm-Bibliothek erzeugt. Neben den Ergebnissen für  $\tilde{T}_n$  stehen zum Vergleich die Resultate des mit Hilfe der ISML-Routine DSPWLK durchgeführten Shapiro-Wilk-Tests  $W$ .

Tabelle 2.8 zeigt die Ergebnisse für die Familie der stabilen Verteilungen („stable distributions“)  $S(a, b)$  mit charakteristischer Funktion

$$\varphi(t; a, b) = \begin{cases} \exp(-|t|^a \exp(-ib\pi(1 - |1 - a| \operatorname{sign}(t)/2))) & : a \neq 1 \\ \exp(-|t| (1 + 2ib \log |t| \operatorname{sign}(t)/\pi)) & : a = 1 \end{cases}$$

( $-1 \leq b \leq 1, 0 \leq a \leq 2$ ).  $a$  bzw.  $b$  werden als Wölbungs- bzw. Schiefeparameter bezeichnet. Für  $b = 0$  sind die Verteilungen symmetrisch; insbesondere ist  $S(2, 0)$  die Normalverteilung  $\mathcal{N}(0, 2)$  und  $S(1, 0)$  die Cauchyverteilung. Je kleiner der Parameter  $a$  ist, desto stärker sind die Enden der Verteilung besetzt; die Wölbung der Verteilung ist größer als bei der Normalverteilung („leptokurtic distributions“). Bei beiden Tests wächst die Güte, wenn die Wölbung im Vergleich zur Normalverteilung zunimmt (dh. wenn der Parameter  $a$  abnimmt) bzw. wenn  $b$  und damit die Schiefe zunimmt. Die zweite Beobachtung deckt sich allerdings nicht mit den Simulationsergebnissen für  $W$  in Baringhaus et al. (1989), Tabelle IV; dort hatte eine Veränderung von  $b$  keinen Einfluß auf die Güte aller dort verwendeten Tests. Der Anteil der

Ablehnungen ist bei  $\tilde{T}_n$  durchgehend etwas höher als für den Shapiro-Wilk-Test.

Tabelle 2.9 zeigt die Güte der beiden Tests bei symmetrischen, Tabelle 2.10 bei schiefen Verteilungen. Alle Alternativen wurden in einer Simulationsstudie von Pearson et al. (1977) verwendet und werden in dieser Arbeit genauer beschrieben. Dabei bezeichnen  $S_B$  bzw.  $S_U$  Verteilungen aus dem Johnson-System; SC (scale contaminated) bzw. LC (location contaminated) sind Mischungen zweier Normalverteilungen mit gleichen Erwartungswerten, aber unterschiedlichen Varianzen, bzw. mit unterschiedlichen Erwartungswerten, aber gleichen Varianzen. Die Verteilungen sind nach der Größe der Wölbung  $\beta_2$  geordnet. Für die ersten sieben Verteilungen in Tabelle 2.8 bzw. für die ersten fünf Verteilungen in Tabelle 2.9 gilt  $\beta_2 < 3$  ("platykurtic distributions"), für die restlichen ist  $\beta_2 > 3$ . Die Ergebnisse für den Shapiro-Wilk-Test unterscheiden sich zum Teil erheblich von den Resultaten in Pearson et al. (1977); dies liegt wohl daran, daß die Zahlen in der dortigen Studie nur auf 200 Wiederholungen beruhen. Bei Verteilungen mit  $\beta_2 < 3$  schneidet der Shapiro-Wilk-Test deutlich besser als der auf  $\tilde{T}_n$  beruhende Test ab. Bei symmetrischen Alternativen mit  $\beta_2 > 3$  ist der Anteil der Ablehnungen beim Stichprobenumfang  $n = 20$  bei beiden Tests etwa gleich groß; bei  $n = 50$  ist  $\tilde{T}_n$  oft besser als  $W$ . Bei unsymmetrischen Alternativen mit  $\beta_2 > 3$  sind die Ergebnisse beider Tests sehr ähnlich.

Insgesamt kann gesagt werden, daß der auf der empirischen integrierten Verteilungsfunktion beruhende Test auf Normalverteilung wie erwartet eine hohe Güte gegenüber Verteilungen mit stark besetzten Verteilungsenden aufweist. Die Güte gegen Alternativen mit  $\beta_2 < 3$  ist im Vergleich zum Shapiro-Wilk-Test gering; letzterer ist bei diesen Verteilungen auch in der Studie von Pearson et al. der beste Test. Eventuell könnte man durch Einführung einer Gewichtsfunktion wie im Falle der Exponentialverteilung erreichen, daß die Güte des EIVF-Test gegenüber Alternativen mit schwach besetzten Enden höher wird.



Testgröße Verteilung	$\beta_2$	$\tilde{T}_n$		W	
		$n = 20$	$n = 50$	$n = 20$	$n = 50$
$S_B(0, 0.5)$	1.63	22.4	78.3	44.7	99.2
Tukey(1.5)	1.75	12.3	51.8	25.6	93.3
Beta(1,1)	1.80	10.0	42.5	20.5	88.3
$S_B(0, 0.707)$	1.87	8.3	33.2	15.0	75.0
Tukey(0.7)	1.92	6.5	24.8	11.6	66.1
Tukey(3.0)	2.06	3.9	10.2	6.6	44.3
Beta(2,2)	2.14	3.9	9.3	5.4	26.6
$S_U(0, 3)$	3.53	7.6	9.2	7.5	8.0
$t_{10}$	4.00	9.4	13.4	9.6	11.6
Logistic	4.20	11.4	17.4	11.4	14.1
$S_U(0, 2)$	4.51	12.2	19.0	12.3	15.6
Tukey(10)	5.38	74.0	99.0	80.2	99.4
Laplace	6.00	26.8	52.1	25.8	41.2
SC(0.20,3)	7.54	36.2	67.5	36.8	60.8
SC(0.05,3)	7.65	18.2	32.6	19.1	32.7
SC(0.10,3)	8.33	27.1	50.0	28.2	47.7
SC(0.20,5)	11.22	70.1	96.7	70.1	95.2
SC(0.20,7)	12.84	84.6	99.5	84.8	99.3
SC(0.10,5)	16.45	53.6	84.3	54.5	83.4
SC(0.05,5)	19.96	34.8	62.0	35.7	62.2
SC(0.10,7)	21.49	66.5	93.1	67.2	92.9
SC(0.05,7)	31.40	44.2	74.3	45.0	74.8
$S_U(0, 1)$	36.19	42.6	76.0	42.0	68.8
$S_U(0, 0.9)$	82.08	50.5	84.5	49.8	78.6
$t_4$	$\infty$	23.6	44.2	23.6	38.6
$t_2$	$\infty$	53.0	85.9	52.5	81.3
$t_1$	$\infty$	86.7	99.7	86.3	99.4

Tabelle 2.9: Empirische Güte bei symmetrischen Verteilungen, Niveau  $\alpha = 0.05$ , 100000 Wiederholungen

Testgröße Verteilung			$\tilde{T}_n$		$W$	
	$\sqrt{\beta_1}$	$\beta_2$	$n = 20$	$n = 50$	$n = 20$	$n = 50$
$S_B(0.533, 0.5)$	0.65	2.13	56.1	97.5	72.5	100.0
Beta(3,2)	-0.29	2.36	6.5	16.2	7.4	28.8
Beta(2,1)	-0.57	2.40	24.1	67.9	30.5	90.0
$S_B(1, 2)$	0.28	2.77	6.2	10.4	6.1	11.9
$S_B(1, 1)$	0.73	2.91	28.0	71.3	30.0	83.5
LC(0.20,3)	0.68	3.09	29.0	69.3	25.8	62.8
LC(0.20,5)	1.07	3.16	88.6	100.0	62.8	100.0
LC(0.20,7)	1.25	3.20	98.7	100.0	98.7	100.0
Weibull(2)	0.63	3.25	14.5	36.1	15.0	43.7
LC(0.10,3)	0.80	4.02	26.0	56.7	24.6	51.8
$\chi_{10}^2$	0.89	4.20	23.9	56.1	24.1	59.2
LC(0.05,3)	0.68	4.35	17.9	35.1	18.2	33.3
LC(0.10,5)	1.54	5.45	75.6	97.8	76.2	98.0
$S_U(-1, 2)$	0.87	5.59	21.2	43.3	20.6	39.0
$\chi_4^2$	1.41	6.00	50.2	91.4	52.8	95.6
LC(0.10,7)	1.96	6.40	87.7	99.5	88.0	99.5
LC(0.05,5)	1.65	7.44	52.4	83.6	54.3	85.3
$\chi_2^2$	2.00	9.00	78.1	99.6	83.5	100.0
LC(0.05,7)	2.42	10.36	65.2	92.3	65.6	92.6
$\chi_1^2$	2.83	15.00	95.7	100.0	98.4	100.0
Weibull(0.5)	6.62	87.72	99.5	100.0	99.9	100.0
$S_U(1, 1)$	-5.30	93.40	73.6	98.0	72.2	96.8
Lognormal(0,1,0)	6.18	113.94	91.2	100.0	93.2	100.0

Tabelle 2.10: Empirische Güte bei unsymmetrischen Verteilungen, Niveau  $\alpha = 0.05$ , 100000 Wiederholungen

# Anhang A

## Ergänzungen

### A.1 Systeme orthogonaler Polynome bezüglich einer Verteilungsfunktion

In diesem Abschnitt sollen einige Grundlagen aus der Theorie der Orthogonalpolynome zusammengestellt werden. Im folgenden sei  $F : \mathbb{R} \rightarrow [0, 1]$  eine Verteilungsfunktion mit existierenden Momenten, d.h. es gelte

$$\int_{-\infty}^{\infty} |x|^n dF(x) < \infty \quad (n = 1, 2, \dots). \quad (\text{A.1})$$

Mit  $\mu_n$  wird wieder das  $n$ -te Moment bezeichnet. Weiter besitze  $F$  unendlich viele Wachstumspunkte, d.h. die Menge

$$\mathcal{S}(F) = \{x \mid F(x + \epsilon) - F(x - \epsilon) > 0 \quad \forall \epsilon > 0\} \quad (\text{A.2})$$

sei unendlich.

Für  $n \geq 0$  sei  $P_n$  ein Polynom vom Grad  $n$ . Die Folge  $\{P_n(\cdot)\}_{n=0}^{\infty}$  heißt Orthogonalsystem (OGS) aus Polynomen bezüglich  $F$ , falls Konstanten  $K_n \neq 0$  ( $n \geq 0$ ) existieren, so daß für  $m, n \geq 0$  die Bedingung

$$\mathcal{L}[P_m(x)P_n(x)] = \int P_m(x) P_n(x) dF(x) = K_n \delta_{mn} \quad (\text{A.3})$$

erfüllt ist.

### A.1.1 Satz

- a) Erfüllt eine Verteilungsfunktion  $F$  die Bedingungen (A.1) und (A.2), so existiert ein OGS  $\{P_n(\cdot)\}_{n=0}^{\infty}$  reeller Polynome bezüglich  $F$ .  $P_n(\cdot)$  ist bis auf einen Faktor eindeutig bestimmt.
- b) Werden die orthogonalen Polynome dadurch eindeutig festgelegt, daß sie den Höchstkoeffizienten 1 besitzen, so gilt mit den Festlegungen  $P_{-1} = 0$  und  $P_0 = 1$  die Rekursionsformel

$$P_n(x) = (x - c_n)P_{n-1}(x) - \lambda_n P_{n-2}(x), \quad n = 1, 2, \dots,$$

wobei für  $n \geq 1$

$$c_n = \frac{\mathcal{L}[xP_{n-1}^2(x)]}{\mathcal{L}[P_{n-1}^2(x)]}, \quad \lambda_{n+1} = \frac{\mathcal{L}[P_n^2(x)]}{\mathcal{L}[P_{n-1}^2(x)]},$$

gilt und  $\lambda_1$  beliebig ist.

Insbesondere gilt  $c_1 = \mu_1$ ,  $P_1(x) = x - \mu_1$ .

- c) Für die orthogonalen Polynome  $P_n(x) = \sum_{k=0}^n a_{n,k} x^k$  mit  $a_{n,n} = 1$  ( $n \geq 1$ ) gelten für  $n \in \mathbb{N}$  die Relationen

$$\begin{aligned} \mathcal{L}[P_n^2(x)] &= \sum_{k=0}^n a_{n,k} \mu_{n+k}, \\ \mathcal{L}[xP_n^2(x)] &= \sum_{k=0}^n a_{n,k} \mu_{n+k+1} + a_{n,n-1} \mathcal{L}[P_n^2(x)]. \end{aligned}$$

BEWEIS: a) Das durch  $\mathcal{L}(g) = \int g dF$  definierte Funktional ist positiv definit, d.h. es gilt  $\mathcal{L}(h) > 0$  für jedes nichtnegative Polynom  $h \neq 0$ . Damit folgt die Behauptung mit Theorem I.3.3 und dem Korollar zu Theorem I.2.2 in Chihara (1978).

b) Theorem I.4.1 und I.4.2 in Chihara (1978).

c) Mit  $\int P_n(x)x^k dF(x) = 0 \quad \forall k < n$  ergibt sich die Behauptung aus der Definition von  $\mathcal{L}(\cdot)$ . ■

**Bemerkungen:**

1. Mit obiger Rekursionsformel lassen sich die ersten orthogonalen Polynome leicht explizit angeben, wenn die Momente bekannt sind.
2. Werden die Polynome dadurch festgelegt, daß in (A.3)  $K_n = 1$  für  $n \geq 0$  gilt, so erhält man das zu  $F$  gehörende Orthonormalsystem.
3. Ist die Menge  $\mathcal{S}(F)$  in (A.2) endlich, so ist  $\mathcal{L}$  nicht mehr positiv definit. Hat  $\mathcal{S}(F)$  etwa die Mächtigkeit  $N + 1$ , so existiert eine endliche Menge von Polynomen  $\{P_n(\cdot)\}_{n=0}^N$ , die die Orthogonalitätsrelation (A.3) erfüllen. Sind die  $N + 1$  Wachstumsstellen o.B.d.A. die Punkte  $0, 1, \dots, N$ , so nimmt (A.3) die Form

$$\sum_{k=0}^N P_m(k) P_n(k) (F(k) - F(k-1)) = K_n \delta_{mn}, \quad m, n = 0, 1, \dots, N,$$

(mit  $K_n \neq 0$ ) an. Beispiele sind die diskreten Tschebyscheff- bzw. die Krawtchoukpolynome, die orthogonal bezüglich der diskreten Gleichverteilung bzw. der Binomialverteilung sind (siehe Chihara (1978), S. 161 und 162). Auch in diesem Fall gelten wieder Rekursionsformeln analog zu Satz A.1.1 b).

4. Eine Klasse von orthogonalen Polynomen, für die die rekursive Berechnung besonders einfach ist, ist die Meixner-Klasse. Diese ist dadurch charakterisiert, daß die Koeffizienten  $c_n$  linear und die Koeffizienten  $\lambda_n$  quadratisch von  $n$  abhängen. Die Meixner-Klasse besteht aus fünf verschiedenen Unterklassen von positiv-definiten polynomialen OGS, von denen vier zu bekannten Verteilungen gehören: die Hermite- bzw. (verallgemeinerten) Laguerre-Polynome, die orthogonal bezüglich der Verteilungsfunktion der Normal- bzw. der Gammaverteilung sind, sowie die Charlier- bzw. Meixner-Polynome, die zur Poisson- bzw. Negativen Binomialverteilung gehören. Näheres findet man in Chihara (1978), Kapitel V.4; siehe auch Rayner und Best (1989), Anhang 1.

## A.2 Elliptisch-symmetrische Verteilungen

Dieser Abschnitt faßt einige Grundlagen über die Klasse der sphärisch-symmetrischen bzw. elliptisch-symmetrischen Verteilungen zusammen, die an verschiedenen Stellen im Haupttext benötigt werden. Beweise findet man etwa in Fang et al. (1989).

### A.2.1 Definition

Ein  $d$ -dimensionaler Zufallsvektor  $X$  besitzt eine sphärisch-symmetrische Verteilung, falls für alle  $H \in \mathcal{O}(d)$

$$H X \stackrel{\mathcal{D}}{=} X$$

gilt, wobei  $\mathcal{O}(d)$  die Menge der orthogonalen  $(d \times d)$ -Matrizen bezeichnet.

### A.2.2 Satz

$X$  sei ein sphärisch-symmetrisch verteilter Zufallsvektor im  $\mathbb{R}^d$ .  $U^{(d)}$  sei ein auf der Oberfläche der Einheitskugel im  $\mathbb{R}^d$  gleichverteilter Zufallsvektor. Dann existieren eine Verteilungsfunktion  $F$  auf  $[0, \infty)$  und eine reelle Zufallsvariable  $R \sim F$ , so daß

$$X \stackrel{\mathcal{D}}{=} R U^{(d)}$$

gilt, wobei  $R$  und  $U^{(d)}$  stochastisch unabhängig sind und das Symbol  $\stackrel{\mathcal{D}}{=}$  die Verteilungsgleichheit bezeichnet. Dabei gilt

$$R \stackrel{\mathcal{D}}{=} \|X\|.$$



**Bemerkung:** Setzt man  $f(x) = (x'x)^{1/2}$ , so folgt die letzte Aussage sofort aus

$$\|X\| = f(X) \stackrel{\mathcal{D}}{=} f(R U^{(d)}) = R$$

und gilt offensichtlich ohne die in Fang et al. (1989), Theorem 2.3, bzw. Fang und Zhang (1990), Corollary 1 zu Theorem 2.5.3, gemachte Voraussetzung  $P(X = 0) = 0$ . Ist diese Voraussetzung erfüllt, so gilt außerdem

$$X / \|X\| \stackrel{\mathcal{D}}{=} U^{(d)}.$$

Die Schreibweise  $X \sim S_d(F)$  bedeutet im folgenden, daß  $X$  ein sphärisch-symmetrisch verteilter Zufallsvektor im  $\mathbb{R}^d$  mit Radialverteilung  $F$  wie in Satz A.2.2 ist.

### A.2.3 Satz

Sei  $X \stackrel{\mathcal{D}}{=} RU^{(d)} \sim S_d(F)$ . Für natürliche Zahlen  $s_1, \dots, s_d$  und  $s = s_1 + \dots + s_d$  berechnen sich die gemischten Momente von  $X = (X_1, \dots, X_d)'$  folgendermaßen:

$$E \left[ \prod_{i=1}^d X_i^{s_i} \right] = \begin{cases} E[R^s] \left( \frac{2}{d} \right)^{[l]} \prod_{i=1}^d \frac{(2l_i)!}{4^{l_i} (l_i)!}, & \text{falls } s_i = 2l_i, l_i \in \mathbb{N}_0, \\ & i = 1, \dots, d, s = 2l; \\ 0 & \text{falls mindestens eines} \\ & \text{der } s_i \text{ ungerade ist;} \end{cases}$$

hierbei ist  $a^{[l]} = a(a+1) \cdots (a+l-1)$ . ■

**A.2.4 Korollar** Sei  $\mu_{s_1, \dots, s_d} = E \left[ \prod_{i=1}^d X_i^{s_i} \right]$ , wobei auftretende Nullen nicht notiert werden, da die Anordnung der  $s_i$  keine Rolle spielt. Dann folgt aus Satz A.2.3:

$$\begin{aligned} \mu_4 &= 3 \mu_{22}; & \mu_6 &= 5 \mu_{42} = 15 \mu_{222}; \\ \mu_8 &= 7 \mu_{62} = \frac{35}{3} \mu_{44} = 35 \mu_{422} = 105 \mu_{2222}. \end{aligned}$$

■

**A.2.5 Definition**

Ein  $d$ -dimensionaler Zufallsvektor  $X$  besitzt eine elliptisch-symmetrische Verteilung mit Parametern  $\mu \in \mathbb{R}^d$  und  $\Delta \in \mathbb{R}^{d \times d}$ , falls ein Zufallsvektor  $Y \sim S_k(F)$  und eine  $(k \times d)$ -Matrix  $A$  vom Rang  $k$  existieren, so daß  $\Delta = A'A$  und

$$X \stackrel{\mathcal{D}}{=} \mu + A'Y$$

gilt. Mit  $R \sim F$  gilt wegen Satz A.2.2 die Verteilungsgleichheit

$$X \stackrel{\mathcal{D}}{=} \mu + R A' U^{(k)}.$$

**A.2.6 Satz**

Der Zufallsvektor  $X$  sei elliptisch-symmetrisch verteilt mit Parametern  $\mu$  und  $\Delta$ . Ist  $E[R^2] < \infty$ , so gilt

$$E[X] = \mu, \quad \text{Cov}(X) = \frac{E[R^2]}{\text{Rang}(\Delta)} \Delta.$$

■

Ist  $\Delta$  positiv definit, also die Kovarianzmatrix von  $X$  nichtsingulär, so gilt für den normierten Zufallsvektor  $\tilde{X} = [\text{Cov}(X)]^{-1/2}(X - \mu)$

$$E[\tilde{X}] = 0 \quad \text{und} \quad E[\tilde{X}\tilde{X}'] = I_d.$$

Dies bedeutet aber gerade, daß  $\tilde{X}$  eine sphärisch-symmetrische Verteilung besitzt mit

$$E[|\tilde{X}|^2] = E[R^2] = d.$$

Die Momente von  $\tilde{X}$  berechnen sich folglich nach Satz A.2.3; insbesondere gilt für die in (1.15) definierte und der multivariaten Schiefe entsprechende Verteilungsgröße  $\beta_{1,d}$  die Beziehung

$$\beta_{1,d} = E[(\tilde{X}'\tilde{Y})^3] = 0, \tag{A.4}$$

wobei  $\tilde{Y}$  eine unabhängige Kopie von  $\tilde{X}$  ist.



### A.3 Elementare Schranken für die Überlebenswahrscheinlichkeiten einiger diskreter Verteilungen

Der Anpassungstest in Abschnitt 2.1 setzt bei der hypothetischen Verteilung gewisse Regularitätseigenschaften voraus; insbesondere müssen die Enden der Verteilung „hinreichend schnell“ abklingen. Der folgende Satz gibt für die wichtigsten diskreten Verteilungen Schranken für die Überlebenswahrscheinlichkeiten an.

#### A.3.1 Satz

Im folgenden bezeichnen  $(f_k)_{k \geq 0}$  die Zähldichte und  $F(\cdot)$  die Verteilungsfunktion der  $\mathbb{N}_0$ -wertigen Zufallsvariablen  $X$ . Für die Überlebenswahrscheinlichkeiten  $1 - F(k) = \sum_{j \geq k+1} f_j$  gelten die folgenden Schranken:

a) Ist  $X$  Poisson-verteilt mit Parameter  $\vartheta > 0$ , so gilt für  $k + 1 \geq \vartheta$

$$f_{k+1} < 1 - F(k) < \frac{k+2}{k+2-\vartheta} f_{k+1}.$$

b)  $X$  sei negativ binomial-verteilt mit Parametern  $r > 0$  und  $p$  ( $0 < p < 1$ ); es ist also  $f_k = \binom{k+r-1}{k} p^r q^k$  ( $k \geq 0$ ) mit  $q = 1 - p$ .

Ist  $r > 1$ , so gilt für  $k \geq rq/p$

$$\frac{1}{p} f_{k+1} < 1 - F(k) < \frac{k+2}{k+2-(k+r+1)q} f_{k+1}.$$

Im Fall  $r < 1$  gilt für  $k \geq rq/p$

$$\frac{k+2}{k+2-(k+r+1)q} f_{k+1} < 1 - F(k) < \frac{1}{p} f_{k+1}.$$

Für  $r > 1$  (bzw.  $r < 1$ ) gilt dabei die untere (bzw. obere) Schranke für  $k \geq 0$ .

c) Ist  $X$  logarithmisch verteilt mit Parameter  $\vartheta$ ,  $0 < \vartheta < 1$ , ist also  $f_k = a(\vartheta)\vartheta^k/k$ ,  $k \geq 1$ , mit  $a(\vartheta) = -(\ln(1-\vartheta))^{-1}$ , so gilt für  $k \geq 0$

$$\left[1 - \frac{k+1}{k+2}\vartheta\right]^{-1} f_{k+1} < 1 - F(k) < (1-\vartheta)^{-1} f_{k+1}.$$

d) Ist  $X$  binomial-verteilt mit Parametern  $n \in \mathbb{N}$  und  $p$  ( $0 < p < 1$ ), ist also  $f_k = \binom{n}{k} p^k q^{n-k}$  ( $0 \leq k \leq n$ ) mit  $q = 1 - p$ , so gilt für  $np \leq k+1 \leq n$

$$f_{k+1} \leq 1 - F(k) \leq \frac{(k+2)q}{k+2 - (n+1)p} f_{k+1}.$$

BEWEIS: Mit  $S_k = 1 - F(k)$  ( $k \geq 0$ ) gilt

$$S_k = f_{k+1} + \sum_{j=k+2}^{\infty} f_j \quad (k \geq 0). \quad (\text{A.5})$$

Im Falle der Poissonverteilung gilt  $f_{k+1} = \vartheta f_k / (k+1)$  und somit

$$\begin{aligned} S_k &< f_{k+1} + \frac{\vartheta}{k+2} \sum_{j=k+2}^{\infty} f_{j-1} \\ &= f_{k+1} + \frac{\vartheta}{k+2} S_k, \end{aligned}$$

woraus für  $\vartheta/(k+2) < 1$  die Ungleichung in a) folgt.

Im Falle der logarithmischen Verteilung folgt aus  $f_{k+1} = \vartheta k f_k / (k+1)$  und (A.5) die Ungleichung

$$S_k > f_{k+1} + \frac{\vartheta(k+1)}{k+2} S_k$$

und hieraus die untere Schranke in c). Die obere Schranke erhält man analog, wenn man in (A.5)  $f_{k+1} < \vartheta f_k$  einsetzt.

Für die negative Binomialverteilung gilt die Beziehung

$$f_{k+1} = \frac{(k+r)q}{k+1} f_k.$$

Ist  $r > 1$  und somit  $(j+r)q/(j+1)$  fallend in  $j$ , so folgt wie oben die Ungleichung

$$\left(1 - \frac{(k+r+1)q}{k+2}\right) S_k < f_{k+1}$$

und daraus für  $k > rq/p$  die erste Ungleichung aus b). Für  $r > 1$  ist weiter  $f_{k+1} > qf_k$  und somit

$$S_k > f_{k+1} + q S_k$$

für alle  $k \geq 0$ . Im Fall  $r < 1$  gelten die entsprechenden Abschätzungen, wenn man jeweils „<“- und „>“-Zeichen vertauscht.

Für die Binomialverteilung gilt

$$f_{k+1} = \frac{(n-k)p}{(k+1)q} f_k,$$

woraus für  $k > (n+1)p - 2$  wie bei der Poissonverteilung die Ungleichung in d) folgt. ■

### Bemerkungen:

1. Setzt man in Glynn (1987), Proposition 1 (ii),  $m = 1$ , so erhält man die obere Schranke für die Poissonverteilung. In Johnson et al. (1992), S. 164, ist als einfache Schranke

$$1 - F(k) \leq 1 - \exp\{\vartheta/(k+1)\} \quad (k+1 \geq \vartheta)$$

angegeben. Diese ist nur für wenige Werte von  $k$ , die nahe beim Modalwert liegen, besser als die Schranke in a), in den Verteilungsenden jedoch sehr viel schlechter und für unsere Zwecke unbrauchbar. Eine weitere häufig zitierte Schranke stammt von Bohman (1963):

$$1 - F(k) \leq 1 - [(\vartheta+1)]^{-1} \int_0^k t^\vartheta e^{-t} dt.$$

Diese Schranke ist komplizierter als die Schranke in a); einige numerische Vergleiche lassen vermuten, daß sie auch wesentlich schlechter ist.

2. Die Schranken für die negative Binomialverteilung konnten wir in der Literatur nicht finden. Ist  $r = 1$  in b), liegt also eine geometrische Verteilung vor, so gilt bekanntlich  $1 - F(k) = f_{k+1}/p$ .

3. Die Schranken in c) findet man auch in Johnson et al. (1992), S. 292 (bei der dort angegebenen unteren Schranke fehlt der Exponent -1).
4. Die obere Schranke in d) wurde von Bahadur (1960) mit Hilfe einer recht komplizierten Darstellung der Verteilungsfunktion als unendliche Reihe gefunden; siehe auch Johnson et al. (1992), S. 121. Dort wird auch auf S.111 eine ähnliche, von Diaconis und Zabell (1991) stammende obere Schranke angegeben, die jedoch etwas schlechter ist. Aus ihrer Beweismethode erhalten sie die untere Schranke  $(k+1)f_{k+1}/n$ , was jedoch schlechter als die triviale Schranke in d) ist.
5. Aus den Schranken erhält man sofort Schranken für die diskrete Versagensfunktion  $h_k = f_k/S_{k-1}$ ,  $k \geq 0$ . Insbesondere konvergiert  $h_k$  bei  $k \rightarrow \infty$  für die Poissonverteilung gegen 1, für die negative Binomialverteilung gegen  $p$  und für die logarithmische Verteilung gegen  $1 - \vartheta$ .

# Literaturverzeichnis

- [1] ARAUJO, A., GINÉ, E. (1980). *The central limit theorem for real and Banach valued random variables*. J. Wiley & Sons, New York.
- [2] ARCONES, M.A., GINÉ, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the Limits of Bootstrap*. (LePage, R., Billard, L., eds.) 13-47. New York, J. Wiley & Sons.
- [3] BAHADUR, R.R. (1960): Some approximations to the binomial distribution function. *Ann. of Math. Statist.* **31**, 43-54.
- [4] BARINGHAUS, L., DANSCHKE, R., HENZE, N. (1989): Recent and classical tests for normality - a comparative study. *Commun. Statist. - Simula.* **18**, 363-379.
- [5] BARINGHAUS, L., GÜRTLER, N., HENZE, N. (1998): Weighted  $L^2$ -statistics, an Abelian theorem for Laplace transforms, and components of smooth tests of fit. Preprint.
- [6] BARINGHAUS, L., HENZE, N. (1991): A class of consistent tests for exponentiality based on the empirical Laplace transform. *Ann. Inst. Statist. Math.* **43**, 551-564.
- [7] BARINGHAUS, L., HENZE, N. (1992a): Limit distributions for Mardia's measure of multivariate skewness. *Ann. Statist.* **20**, 1889-1902.
- [8] BARINGHAUS, L., HENZE, N. (1992c): An adaptive omnibus test for exponentiality. *Commun. Statist. - Theory Meth.* **21**, 969-978.

- 
- [9] BARLOW, R.E., CAMPO, R. (1975). Total Time on Test Processes and Applications to Failure data Analysis. In *Reliability and Fault Tree Analysis*. (Barlow, R.E., Fussell, J., Singpurwalla, N.D. eds.), SIAM, 451-481.
- [10] BERAN, R. (1984): Bootstrap Methods in Statistics. *Jahresber. Deutsch. Math.-Verein.* **86**, 14-30.
- [11] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. J. Wiley & Sons, New York.
- [12] BOHMAN, H. (1963): Two inequalities for Poisson distributions. *Skandinavisk Aktuarietidskrift* **46**, 47-52.
- [13] BOULERICE, B., DUCHARME, G.R. (1995): A note on smooth tests of goodness of fit for location-scale families. *Biometrika* **82**, 437-438.
- [14] BURKE, M. D., CSÖRGŐ, M., CSÖRGŐ, S., RÉVÉSZ, P. (1978): Approximations of the empirical process when parameters are estimated. *Ann. Prob.* **7**, 790-810.
- [15] CHANDRA, M., SINGPURWALLA, N.D. (1981): Relationships between some notions which are common to reliability theory and economics. *Mathematics of Operations Research* **6**, 113-121.
- [16] CHIHARA, T.S. (1978). *An Introduction to Orthogonal Polynomials*. New York, Gordon and Breach.
- [17] CHOULAKIAN, V., LOCKHART, R.A., STEPHENS, M. A. (1994): Cramér-von Mises statistics for discrete distributions. *Canad. J. Statist.* **22**, 125-137.
- [18] DAVYDOV, Y.A., LIFSHITS, M.A. (1985): Fiberling method in some probabilistic problems. *J. Sov. Math.* **31**, 2796-2858.
- [19] DIACONIS, P., ZABELL, S. (1991): Closed Form Summation for Classical Distributions: Variations on a Theme of De Moivre. *Statistical Science* **6**, 284-302.
- [20] DURBIN, J. (1973a). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279-290.

- 
- [21] EATON, M., PERLMAN, M.D. (1973): The nonsingularity of generalized sample covariance matrices. *Ann. Statist.* **1**, 710-717.
- [22] EPPS, T.W. (1995): A test of fit for lattice distributions. *Commun. Statist.-Theor. Meth.* **24**, 1455-1479.
- [23] FANG, K.T., KOTZ, S., NG, K.W. (1989). *Symmetric multivariate and related distributions*. Chapman & Hall, London.
- [24] FANG, K.T., ZHANG, Y. (1990). *Generalized multivariate Analysis*. Springer, New York.
- [25] GASTWIRTH, J.L., OWENS, M.E.B. (1977): On classical tests of normality. *Biometrika* **64**, 135-139.
- [26] GLYNN, P.W. (1987): Upper bounds on Poisson tail distributions. *Oper. Res. Lett.* **6**, 9-14.
- [27] HALL, P. (1992). *The Bootstrap and Edgeworth Expansions*. Springer, New York.
- [28] HEILMANN, W. (1985): Transformations of Claim Distributions. *Mitteilungen der Vereinigung schweiz. Versicherungsmathematiker* **1**, 57-69.
- [29] HENZE, N. (1994a): On Mardia's kurtosis test for multivariate normality. *Commun. Statist.-Theory Meth.* **23**, 1031-1045.
- [30] HENZE, N. (1994b): The asymptotic behavior of a variant of multivariate kurtosis. *Commun. Statist.-Theory Meth.* **23**, 1047-1061.
- [31] HENZE, N. (1996): Empirical-distribution-function goodness-of-fit tests for discrete models. *Canad. J. Statist.* **24**, 81-93.
- [32] HENZE, N. (1997a): Do Components of Smooth Tests of Fit have Diagnostic properties? *Metrika* **45**, 121-130.
- [33] HENZE, N. (1997b): Limit laws for multivariate skewness in the sense of Móri, Rohatgi and Székely. *Statist. & Prob. Letters* **33**, 299-307.
- [34] HENZE, N., GUTJAHR, S., FOLKERS, M. (1997): Limit laws for generalized affine invariant skewness measures. Preprint.

- 
- [35] HENZE, N., KLAR, B. (1996): Properly rescaled components of smooth tests of fit are diagnostic. *Austral. Journ. Statist.* **38**, 61-74.
- [36] HEUSER, H. (1986). *Analysis II*. Teubner, Stuttgart.
- [37] HINDERER, K. (1980): The integrated distribution function and its applications for some inventory problems with linear demand pattern, arbitrary demand distribution and without fixed ordering cost. *Math. Operationsforsch. Statist., Ser. Optimization* **11**, 299-310.
- [38] JANSSEN, A. (1995): Principal component decomposition of non-parametric tests. *Probab. Theory Relat. Fields* **101**, 193-209.
- [39] JOHNSON, N.L., KOTZ, S., KEMP, A.W. (1992). *Univariate Discrete Distributions*. J. Wiley & Sons, New York.
- [40] KALLENBERG, W.C.M., LEDWINA, T., RAFAJLOWICZ, E. (1997): Testing bivariate independence and normality. *Sankhyā Ser. A* **59**, 42-59.
- [41] KENDALL, M.G., STUART, A. (1979). *The advanced theory of statistics vol. 2*. Fourth edition, Griffin, London.
- [42] KOCHERLAKOTA, S., KOCHERLAKOTA, K. (1986): Goodness of fit tests for discrete distributions. *Commun. Statist.-Theor. Meth.* **15**, 815-829.
- [43] KOCHERLAKOTA, S., KOCHERLAKOTA, K. (1992): *Bivariate discrete distributions*. Marcel Dekker, New York.
- [44] KOZIOL, J.A. (1987): An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives. *Metrika* **34**, 17-24.
- [45] LEHMANN, E.L. (1983). *Theory of Point Estimation*. J. Wiley & Sons, New York.
- [46] MAMMEN, E. (1992). *When does the Bootstrap work*. Lect. Notes in Statistics, 77. New York, Springer.
- [47] MARDIA, K.V. (1970): Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.



- 
- [48] MARDIA, K.V., KENT, J.T. (1991): Rao score tests for goodness of fit and independence. *Biometrika* **78**, 355-363.
- [49] MÓRI, T.F., ROHATGI, V.K., SZÉKELY, G.J. (1993): On multivariate skewness and kurtosis. *Theoret. Probab. Appl.* **38**, 547-551.
- [50] MÜLLER, A. (1996): Ordering of risks: A comparative study via stop-loss transforms. *Insurance* **17**, 215-222.
- [51] MÜLLER, A. (1998): Comparing risks with unbounded distributions. Erscheint in *J. Math. Econ.*
- [52] NAKAMURA, M., PÉREZ-ABREU, V. (1993): Use of an empirical probability generating function for testing a Poisson model. *Canad. J. Statist.* **21**, 149-156.
- [53] NEUHAUS, G. (1974). Asymptotic properties of the Cramér-von Mises statistic when parameters are estimated. In *Proc. Prague Symp. on Asymptotic Statist.* 1973 (J. Hájek, ed.). Universita Karlova Praha **2**, 257-297.
- [54] NEYMAN, J. (1937). Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift* **20**, 149-199.
- [55] NIKITIN, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge University Press.
- [56] PEARSON, E.S., D'AGOSTINO, R.B., BOWMAN, K.O. (1977): Tests for departure from normality: Comparison of powers. *Biometrika* **64**, 231-246.
- [57] POLITIS, D.N., ROMANO, J.P. (1994): Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031-2050.
- [58] RAYNER, J.C.W., BEST, D.J. (1989). *Smooth Tests of Goodness of Fit*. New York, Oxford University Press.
- [59] RAYNER, J.C.W., BEST, D.J. (1995). Smooth tests for the bivariate Poisson distribution. *Austral. J. Statist.* **37**, 233-245.
- [60] SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York, Wiley.

- [61] SHORAK, G.R., WELLNER, J.A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- [62] SPINELLI, J.J., STEPHENS, M. A. (1997): Cramér-von Mises tests of fit for the Poisson distribution. *Canad. J. Statist.* **25**, 257-268.
- [63] WIDDER, D. V. (1959). *The Laplace transform*. Princeton University Press, Princeton.
- [64] WITTING, H., MÜLLER-FUNK, U. (1995). *Mathematische Statistik II*. Teubner, Stuttgart.
- [65] WU, C.F.J. (1986): Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14**, 1261-1350.
- [66] WU, C.F.J. (1990): On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18**, 1438-1452.

# Lebenslauf

## Persönliche Daten

Name: Bernhard Klar  
Geboren am: 5. April 1967 in Riedlingen  
Eltern: Rainer und Rotraut Klar

## Schulbildung

1973-1977 Grundschole in Riedlingen  
1977-1986 Kreisgymnasium Riedlingen  
Juni 1986 Abiturprüfung

## Studium

1986-1992 Studium der Mathematik an der Universität  
Karlsruhe  
Juni 1992 Diplomprüfung in Mathematik

## Tätigkeiten während des Studiums

Okt. 1988 - Sept. 1989 Wissenschaftliche Hilfskraft am Mathematischen  
Institut I der Universität Karlsruhe  
Okt. 1989 - Sept. 1991 Wissenschaftliche Hilfskraft am Rechenzentrum  
der Universität Karlsruhe  
Sept. 1991 - Sept. 1992 Wissenschaftliche Hilfskraft am Institut für  
Mathematische Stochastik der Universität  
Karlsruhe

## Berufstätigkeit

Okt. 1992 - Dez. 1995 Wissenschaftlicher Angestellter am Institut für  
Wissenschaftliches Rechnen und Mathematische  
Modellbildung der Universität Karlsruhe  
seit Januar 1996 Wissenschaftlicher Angestellter am Institut für  
Mathematische Stochastik der Universität  
Karlsruhe