



Testen auf Normalverteilung: Der Jarque-Bera-Test

Diplomarbeit

vorgelegt von
Felix Opetz

Betreuer: PD Dr. Volkert Paulsen
Institut für Mathematische Statistik
Fachbereich 10 - Mathematik und Informatik
Westfälische Wilhelms-Universität Münster

Inhaltsverzeichnis

Abbildungsverzeichnis	iii
Tabellenverzeichnis	iv
1 Einleitung	1
2 Der Score-Test	3
2.1 Grundlagen der Schätztheorie	3
2.2 Maximum-Likelihood-Theorie	4
2.3 Grundlagen der Testtheorie	12
2.4 Herleitung des Score-Tests	15
3 Testen auf univariate Normalverteilung	22
3.1 Das Pearson-Verteilungssystem	22
3.2 Testen von Beobachtungen	24
3.3 Testen von Regressionsresiduen	29
4 Der Jarque-Bera-Test im Vergleich	31
4.1 Kritische Werte und Testalternativen	31
4.2 Powervergleich innerhalb des Pearson-Verteilungssystems	37
4.3 Powervergleich in speziellen Situationen	42
5 Testen auf multivariate Normalverteilung	49
5.1 Multivariate Schiefe- und Wölbungsmaße	50
5.2 Asymptotische Verteilung	52
5.3 Multivariate Teststatistiken	62
6 Anwendung	64
6.1 Testen der Rendite des DAX auf Normalverteilung	64
7 Zusammenfassung und Ausblick	68

A Hilfsresultate	71
B Ergebnisse der empirischen Analysen	73
C Daten-CD	77
Literaturverzeichnis	78

Abbildungsverzeichnis

4.1	Abweichungen der empirischen kritischen Werte der JB -Statistik von den kritischen Werten der χ^2_2 -Verteilung.	33
4.2	Abweichungen der empirischen kritischen Werte der JB_U -Statistik von den kritischen Werten der χ^2_2 -Verteilung.	35
4.3	Der Jarque-Bera-Test im Powervergleich innerhalb des Pearson-Verteilungssystems zum Signifikanzniveau 5%.	41
4.4	Der Jarque-Bera-Test im Powervergleich bei einigen kontaminierten Normalverteilungen zum Signifikanzniveau 5%.	45
4.5	Der Jarque-Bera-Test im Powervergleich bei einigen kontaminierten Normalverteilungen zum Signifikanzniveau 5%.	46
4.6	Der Jarque-Bera-Test im Powervergleich bei einigen kontaminierten Normalverteilungen zum Signifikanzniveau 5%.	47

Tabellenverzeichnis

4.1	Empirische kritische Werte der JB -Statistik bei 10^7 Replikationen.	32
4.2	Empirische kritische Werte der JB_U -Statistik bei 10^7 Replikationen.	34
4.3	Empirische kritische Werte der JB_{JG} -Statistik bei 10^4 Replikationen.	37
4.4	Der Jarque-Bera-Test im Powervergleich innerhalb des Pearson-Verteilungssystems zum Signifikanzniveau 5%.	39
5.1	Empirische kritische Werte der multivariaten Jarque-Bera-Statistiken MJB und \widetilde{MJB} bei 10^4 Replikationen.	63
6.1	Jarque-Bera-Werte der Renditen einiger ausgewählter DAX-Aktien und des DAX-Index.	66
B.1	Der Jarque-Bera-Test im Powervergleich bei einigen KN-Verteilungen zum Signifikanzniveau 5%.	73
B.2	Der Jarque-Bera-Test im Powervergleich bei einigen KN-Verteilungen zum Signifikanzniveau 5%.	74
B.3	Der Jarque-Bera-Test im Powervergleich bei einigen KN-Verteilungen zum Signifikanzniveau 5%.	74
B.4	Der Jarque-Bera-Test im Powervergleich innerhalb des Pearson-Verteilungssystems zum Signifikanzniveau 1%.	75
B.5	Der Jarque-Bera-Test im Powervergleich innerhalb des Pearson-Verteilungssystems zum Signifikanzniveau 10%.	76

Kapitel 1

Einleitung

Als eine der bekanntesten Wahrscheinlichkeitsverteilungen ist die Normalverteilung nicht nur in vielen mathematischen Disziplinen zu einem vieldiskutierten Objekt in statistischen wie auch wahrscheinlichkeitstheoretischen Anwendungsbereichen herangewachsen. Nach ihrer Definition durch Carl Friedrich Gauß zu Beginn des 19. Jahrhunderts war es Pierre-Simon Laplace, der den engen Zusammenhang der Normalverteilung mit anderen Wahrscheinlichkeitsverteilungen aufdeckte und somit die zentrale Stellung der Normalverteilung innerhalb der Wahrscheinlichkeitstheorie festigte. Der angesprochene Zusammenhang spiegelt sich im zentralen Grenzwertsatz wider, der in seiner gebräuchlichsten Form besagt, dass sich die Summe einer großen Zahl unabhängiger und identisch verteilter Zufallsgrößen annähernd wie eine normalverteilte Zufallsvariable verhält. Dieser Tatsache und ihren besonderen Eigenschaften, insbesondere auch ihrer einfachen Form verdankt sie es, dass sie in vielen Bereichen der Natur-, Wirtschafts- und Ingenieurwissenschaften immer wieder Verwendung findet. So reicht ihr Anwendungsspektrum von der Beschreibung zufälliger Abweichungen bei der Produktion von Industriegütern über die Beschreibung von Messfehlern physikalischer Versuchsdurchführungen bis hin zur Modellierung des Wiener-Prozesses und damit der geometrischen Brownschen Bewegung, die in heutigen Zeiten zur Nachbildung finanzmathematischer Preisprozesse benutzt wird. Ebenfalls in der Versicherungsmathematik hat die Normalverteilung Einzug erhalten und legt somit in vielen Bereichen den Grundbaustein der modernen Risikoabschätzung.

Der hohe Stellenwert der Normalverteilung führte dazu, dass in vielen mathematischen Modellen eine Normalverteilung unterstellt wird, was eine nicht immer einhaltbare Anforderung an die im Modell betrachteten Größen nach sich zieht. Als Beispiel sei die Annahme normalverteilter Regressionsfehler in linearen Regressionsmodellen genannt. Ohne das Wissen über die tatsächliche Verteilung der in diesen Modellen betrachteten Größen ist eine Anwendung somit eher spekulativer Natur, sodass an dieser Stelle die Frage nach einer Rechtfertigung für die aufgestellte Annahme aufkommt und infolgedessen geklärt

werden muss, ob das Modell so überhaupt verwendet werden darf. An diesem Punkt greift die angewandte Statistik ein und bietet eine Vielzahl an Möglichkeiten, die getroffenen Annahmen anhand statistischer Tests zu verifizieren. Dabei beschränkt sich die Vielfalt der angebotenen Tests nicht auf solche, die ausschließlich dazu geeignet sind, Aussagen über eine mögliche Normalverteilung der getesteten Größen zu treffen. Besonders die Kernideen der einzelnen Tests, die eine Abweichung von oder eine Übereinstimmung mit der Normalverteilung als glaubwürdig erscheinen lassen, sind dabei so grundlegend verschieden, dass die Frage aufkommt, ob es denn Tests gibt, die als Favoriten aus diesem Vergleich hervorgehen.

Mit dieser Frage befasst sich die vorliegende Arbeit, wobei der Schwerpunkt nicht auf einem direkten Vergleich aller Tests liegt (was bei der großen Vielfalt an Tests eine auch nur schwer lösbare Aufgabe wäre), sondern auf einem speziellen Test auf Normalverteilung, nämlich dem, der durch Carlos M. Jarque und Anil K. Bera bekannt wurde und heute sogar ihre Namen trägt: der *Jarque-Bera-Test*. Dazu wird zu Beginn dieser Arbeit in das Themengebiet der Schätz- sowie der Testtheorie eingeführt um die mathematische Grundlage zu schaffen, die notwendig ist, um die Teststatistik und ihre besonderen Eigenschaften herzuleiten. Im anschließenden Kapitel drei wird explizit auf die Herleitung der Teststatistik eingegangen und die Unterschiede aufgezeigt, die in verschiedenen Test-situationen resultieren. Nachdem der Jarque-Bera-Test sodann als Mittel zum Testen auf Normalverteilung zur Verfügung steht, wird dieser in Kapitel vier in Konkurrenz zu anderen Teststatistiken gesetzt und seine Vor- und Nachteile herausgearbeitet, die im direkten Vergleich anhand von empirischen Analysen bestätigt werden. Nachdem der Jarque-Bera-Test in seiner Form als Test auf univariate Normalverteilung analysiert wurde, wird in Kapitel fünf eine Erweiterung und zugleich Verallgemeinerung der Teststatistik angegeben, die es ermöglicht, sogar auf multivariate Normalverteilung zu testen. In Kapitel sechs schließlich wird eine Anwendung der gewonnenen Resultate auf eine finanzmathematische Fragestellung gegeben, bevor die vorliegende Arbeit in Kapitel sieben mit einer Auswertung in Form einer Zusammenfassung und einem Ausblick beendet wird.

Kapitel 2

Der Score-Test

In der mathematischen Statistik ist man daran interessiert, anhand von Stichproben, dargestellt durch Realisierungen zumeist unabhängiger und identisch verteilter (i.i.d.) Zufallsvariablen, Aussagen über die Parameter der Verteilung einer Grundgesamtheit zu treffen. Dazu werden statistische Modelle aufgestellt, auf deren Grundlage Schätz- oder Testprobleme behandelt werden können. Ziel des ersten Kapitels ist es, eine Teststatistik herzuleiten, mit deren Hilfe Fragen in Bezug auf Hypothesen über bestimmte Verteilungsparameter beantwortet werden können. Unabdingbare Voraussetzung dafür bildet die Theorie der Likelihood-Schätzer, die im Bereich der Schätzprobleme als eine der gängigsten zum Auffinden von Schätzern für Parameterfunktionen angesehen werden kann. Aus diesem Grund wird im ersten Abschnitt eine kurze Einführung in die Schätztheorie gegeben und darauf aufbauend in Abschnitt zwei die Theorie der Likelihood-Funktionen entwickelt, in der Maximum-Likelihood-Schätzer definiert und im Anschluß einige wichtige Eigenschaften von Maximum-Likelihood-Schätzern bewiesen werden. Abschnitt drei beginnt mit einer Erläuterung der Problematik in Testproblemen und motiviert, ausgehend von einfachen Hypothesen, in Abschnitt vier mit Hilfe von Likelihood-Quotienten den Übergang zu komplexeren Hypothesen. Die Vorgehensweise orientiert sich dabei an der in [CH74]. Definitionen und Notationen sind zum Teil aus [Als09], sowie aus [Hel08] übernommen.

2.1 Grundlagen der Schätztheorie

Wie der Name bereits vermuten lässt, besteht das Ziel der Schätztheorie darin, anhand von Stichproben Schätzungen über interessierende Parameter anzustellen. Dabei ist eine Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$ ein n -Tupel von Beobachtungen, die im wahrscheinlichkeitstheoretischen Zusammenhang als Realisierungen von Zufallsvariablen X_1, \dots, X_n aufgefasst werden. Das zugrunde liegende Modell lässt sich dabei wie folgt beschreiben: Ausgangspunkt bildet ein *statistisches Experiment* $\mathcal{E} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$, welches sich aus einem

Stichprobenraum \mathcal{X} , einer σ -Algebra \mathcal{A} über \mathcal{X} und einer Familie $(P_\theta)_{\theta \in \Theta}$ von Wahrscheinlichkeitsmaßen mit $\Theta \subseteq \mathbb{R}^d$ als *Parameterraum* der Wahrscheinlichkeitsfamilie zusammensetzt. Die Stichprobe stellt ein Element des Stichprobenraums dar und kann dann als Realisierung einer Zufallsvariablen $X : (\Omega, \mathcal{F}) \rightarrow (\mathcal{X}, \mathcal{A})$ und einer zugehörigen Familie von Wahrscheinlichkeitsmaßen $(Q_\theta)_{\theta \in \Theta}$ identifiziert werden, wobei $Q_\theta^X = P_\theta$ für alle $\theta \in \Theta$ gilt. Auf die Struktur von (Ω, \mathcal{F}) kommt es bei der Untersuchung der vorliegenden Stichprobe nicht an. Auf diese Weise können Beobachtungen mit Zufallsvariablen in Verbindung gebracht werden, sodass eine Stichprobe dem Zufall unterliegt, welcher über die Zufallsvariablen – genauer über die Verteilungen der Zufallsvariablen – in das eingangs beschriebene Modell einfließt. Dazu sei $(P_\theta)_{\theta \in \Theta}$ dominiert durch ein Maß μ , sodass nach dem Satz von Radon-Nikodým $f(x; \theta) = dP_\theta/d\mu$ gilt. Im stetigen Fall ist $f(x; \theta)$ die Wahrscheinlichkeitsdichte und im diskreten Fall die Wahrscheinlichkeitsfunktion zu P_θ , was im Laufe der vorliegenden Arbeit nicht mehr explizit erwähnt wird. Die Verteilung der Zufallsvariablen wird nun bis auf den Parameter(-vektor) als bekannt angenommen, sodass die Schätzungen eben jenen unbekanntem Parameter(-vektor) betreffen. Ein *Schätzer* oder eine *Schätzfunktion* für θ ist dabei eine messbare Abbildung $T : \mathcal{X} \rightarrow \Theta$, die anhand einer vorliegenden Stichprobe dem zu schätzenden Parameter einen Wert zuweist. Sinnvollerweise versucht man dabei T so zu wählen, dass ein möglichst guter Schätzer resultiert, wobei hier auf Details über die Spezifizierung der Eigenschaft „gut“ in diesem Zusammenhang verzichtet werden soll.

Drei der wohl bekanntesten Schätzmethoden bilden die Momentenmethode, die Bayes-Methode und die Theorie der Likelihood-Schätzer. Auf letztere wird im Folgenden ausführlich eingegangen. Obwohl es sich bei den betreffenden Parametern fast ausschließlich um Vektoren handelt, werden diese, sowie auftretende Matrizen, durch Fettdruck von skalaren Parametern unterschieden.

2.2 Maximum-Likelihood-Theorie

Im Folgenden wird nun davon ausgegangen, dass es sich bei der Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$ um Realisierungen von i.i.d. Zufallsvariablen X_1, \dots, X_n handelt, die gemäß einer Verteilung P_θ gezogen werden. Die Bezeichnungsweise „eine Stichprobe ziehen“ folgt dabei dem allgemeinen Sprachgebrauch und ist somit nicht wörtlich zu verstehen. Zudem sei darauf hingewiesen, dass eine Stichprobe nicht ausschließlich aus wirklichen Beobachtungen bestehen muss. In einigen Situationen wird stattdessen von Zufallsvariablen als Bestandteil der Stichproben ausgegangen, um gewisse Größen wie Erwartungswerte und Varianzen sinnvoll herleiten zu können. Eine Rechtfertigung für dieses Vorgehen liefert das Wissen über die den Beobachtungen zugrunde liegende Verteilung. Ob es sich um Zufallsvariablen

oder Beobachtungen handelt, wird dabei in Form von Klein-, beziehungsweise Großschreibung kenntlich gemacht. Begonnen wird mit einigen Definitionen, die unter den getroffenen Annahmen gelten.

Definition 2.2.1. Es seien x_1, \dots, x_n Realisierungen von i.i.d. Zufallsvariablen X_1, \dots, X_n , die gemäß einer Verteilung $P_{\boldsymbol{\theta}}$ gezogen werden. Dann heißt die Funktion

$$L(\cdot; \mathbf{x}) : \Theta \rightarrow [0, \infty)$$

mit Abbildungsvorschrift

$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta}) = f(\mathbf{x}; \boldsymbol{\theta})$$

Likelihood-Funktion zur Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$.

Likelihood-Funktionen lassen sich wie folgt motivieren: Betrachtet man eine diskret verteilte Zufallsvariable X , so gibt die Verteilung $P_{\boldsymbol{\theta}}$ von X bei gegebenem Parametervektor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ die *relativen Häufigkeiten* der einzelnen Werte aus der Grundgesamtheit an. Somit liegt es nahe, dass bei einer zufälligen Stichprobe jener Parametervektor $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_d)$ am ehesten dem wahren Parametervektor $\boldsymbol{\theta}$ entspricht, bei dem die Wahrscheinlichkeit einer Realisierung der gezogenen Stichprobe am größten ist. Die Intuition lässt sich durch Grenzübergang auf den stetigen Fall übertragen. Mathematisch bestimmt man $\hat{\boldsymbol{\theta}}$ also durch Maximieren der Likelihood-Funktion in $\boldsymbol{\theta}$.

Definition 2.2.2. Gegeben sei die Situation aus Definition 2.2.1. Dann bezeichnet $\hat{\boldsymbol{\theta}}$ den *Maximum-Likelihood-Schätzer (MLS)* für $\boldsymbol{\theta}$, falls

$$L(\hat{\boldsymbol{\theta}}; \mathbf{x}) \geq L(\boldsymbol{\theta}'; \mathbf{x})$$

für alle $\boldsymbol{\theta}' \in \Theta$ gilt. Die Maximierung erfolgt dabei komponentenweise in θ_1 bis θ_d .

Für praktische Zwecke ist es oft vorteilhaft den Logarithmus der Likelihood-Funktion zu verwenden. Aufgrund der Monotonie des Logarithmus ändert sich bei Extremwertuntersuchungen nichts. Da von dieser Eigenschaft häufig Gebrauch gemacht wird, erhält der Logarithmus der Likelihood-Funktion eine eigene Bezeichnung. Dazu wird die *Log-Likelihood-Funktion* definiert durch

$$l(\boldsymbol{\theta}; \mathbf{x}) = \log(L(\boldsymbol{\theta}; \mathbf{x})).$$

Einige Merkmale des MLS sind für die Entwicklung der weiteren Theorie wesentlich. Unter anderem besitzt $\hat{\boldsymbol{\theta}}$ eine asymptotische Normalverteilung um den wahren Parametervektor

$\boldsymbol{\theta} \in \Theta$. Diese Eigenschaft gilt uneingeschränkt, vorausgesetzt es werden sogenannte *Regularitätsbedingungen* an die Likelihood-Funktion gestellt. Zuerst jedoch werden einige Begriffe eingeführt, die im Folgenden eine zentrale Rolle spielen. Im Zusammenhang mit der Theorie der Likelihood-Funktionen erhält die Ableitung der Log-Likelihood-Funktion nach $\boldsymbol{\theta}$ eine eigene Bezeichnung.

Definition 2.2.3. Es gelten die Voraussetzungen aus Definition 2.2.1, dann heißt

$$\boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log(L(\boldsymbol{\theta}; \boldsymbol{x})) = \left(\frac{\partial}{\partial \theta_1} l(\boldsymbol{\theta}; \boldsymbol{x}), \dots, \frac{\partial}{\partial \theta_d} l(\boldsymbol{\theta}; \boldsymbol{x}) \right)^T$$

Score-Funktion oder einfach *Score* für $\boldsymbol{x} = (x_1, \dots, x_n)$.

Um die Score-Funktion von der Abhängigkeit des gesamten Beobachtungsvektors zu lösen, wird zusätzlich der Score der i -ten Beobachtung beschrieben durch

$$\boldsymbol{S}(\boldsymbol{\theta}; x_i) = \frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}; x_i),$$

sodass im Fall unabhängiger Zufallsvariablen

$$\boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log \left(\prod_{i=1}^n f(x_i; \boldsymbol{\theta}) \right) = \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta}) = \sum_{i=1}^n \boldsymbol{S}(\boldsymbol{\theta}; x_i)$$

gilt. Die so eingeführte Score-Funktion bei festen Stichproben hängt somit nur vom unbekanntem Parametervektor $\boldsymbol{\theta}$ ab. Betrachtet man jedoch Zufallsvariablen anstelle fester Beobachtungen, so besteht $\boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{X})$ selbst aus Zufallsvariablen und dem Score kann ein Erwartungswertvektor zugeordnet werden.

Definition 2.2.4. Es gelten die Annahmen aus Definition 2.2.3, dann heißt

$$\boldsymbol{I}(\boldsymbol{\theta}; \boldsymbol{X}) = \mathbb{E}_{\boldsymbol{\theta}} [\boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{X}) \boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{X})^T]$$

die *Fisher-Information* von $\boldsymbol{\theta}$ zur Stichprobe $\boldsymbol{X} = (X_1, \dots, X_n)$. Für einzelne Beobachtungen ist somit

$$\boldsymbol{I}(\boldsymbol{\theta}; X_i) = \mathbb{E}_{\boldsymbol{\theta}} [\boldsymbol{S}(\boldsymbol{\theta}; X_i) \boldsymbol{S}(\boldsymbol{\theta}; X_i)^T]$$

die Fisher-Information von $\boldsymbol{\theta}$ zur Beobachtung X_i .

Die Fisher-Information kann angesehen werden als Maß für die Güte des zu schätzenden Parameters. Je größer $\boldsymbol{I}(\boldsymbol{\theta}; \boldsymbol{X})$ ist, desto genauer lässt sich $\boldsymbol{\theta}$ schätzen. Es kann gezeigt werden, dass unter gewissen Voraussetzungen¹ für die Kovarianz eines *erwartungstreuen*

¹Vgl. [WN70], S.70.

Schätzers $\hat{\boldsymbol{\theta}}$ für $\boldsymbol{\theta}$ gilt: $\text{Cov}_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}(\mathbf{X})] \geq \mathcal{I}(\boldsymbol{\theta}; \mathbf{X})^{-1}$. Ein Schätzer, der diese untere sogenannte (*mehrdimensionale*) *Cramer-Rao-Schranke* annimmt wird als *effizient* bezeichnet. Man beachte, dass es sich bei Score-Funktionen und Fisher-Informationen in der Situation vektorwertiger Parameter $\boldsymbol{\theta}$ selbst um Vektoren beziehungsweise Matrizen handelt.

Die angekündigten *Regularitätsbedingungen (RB)* lauten wie folgt:

1. Θ ist offen mit $\dim(\Theta) < \infty$.
2. Für alle $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ mit $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$ gilt: $f(x; \boldsymbol{\theta}) \neq f(x; \boldsymbol{\theta}')$.
3. In einer Umgebung $U_{\boldsymbol{\theta}}$ des wahren Parametervektors $\boldsymbol{\theta} \in \Theta$ gilt:
 - (a) $\mathcal{S}(\boldsymbol{\theta}'; \mathbf{X})$, $\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{S}(\boldsymbol{\theta}'; \mathbf{X})$, $\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \mathcal{S}(\boldsymbol{\theta}'; \mathbf{X})$ existieren fast sicher für alle $\boldsymbol{\theta}' \in U_{\boldsymbol{\theta}}$.
 - (b) Es existiert eine Funktion $g(\mathbf{X})$ mit $\mathbb{E}[g(\mathbf{X})] < \infty$ und $n^{-1} |\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \mathcal{S}(\boldsymbol{\theta}'; \mathbf{X})| < g(\mathbf{X})$ für alle $\boldsymbol{\theta}' \in U_{\boldsymbol{\theta}}$.
4. Integration bezüglich \boldsymbol{x} und Differentiation bezüglich θ_i dürfen vertauscht werden.

Verteilungsfamilien, die den Regularitätsbedingungen genügen werden im Folgenden auch einfach als *regulär* bezeichnet.

Lemma 2.2.5. *Es sei $\mathbf{X} = (X_1, \dots, X_n)$ eine Stichprobe bestehend aus i.i.d. Zufallsvariablen X_1, \dots, X_n mit regulärer Verteilung $P_{\boldsymbol{\theta}}$. Dann gilt:*

$$\mathbb{E}_{\boldsymbol{\theta}}[\mathcal{S}_j(\boldsymbol{\theta}; X_i)] = 0 \quad (2.1)$$

für $(i, j) \in \{1, \dots, n\} \times \{1, \dots, d\}$ mit $\mathcal{S}_j(\boldsymbol{\theta}; X_i) = \frac{\partial}{\partial \theta_j} l(\boldsymbol{\theta}; X_i)$ als j -te Komponente des Score-Vektors.

Beweis. Es gilt

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}[\mathcal{S}_j(\boldsymbol{\theta}; X_i)] &= \int \frac{\partial}{\partial \theta_j} l(\boldsymbol{\theta}; x_i) f(x_i; \boldsymbol{\theta}) \mu(dx_i) \\ &= \int \frac{\frac{\partial}{\partial \theta_j} f(x_i; \boldsymbol{\theta})}{f(x_i; \boldsymbol{\theta})} f(x_i; \boldsymbol{\theta}) \mu(dx_i) \\ &= \int \frac{\partial}{\partial \theta_j} f(x_i; \boldsymbol{\theta}) \mu(dx_i) \stackrel{\text{RB 4}}{=} \frac{\partial}{\partial \theta_j} \underbrace{\int f(x_i; \boldsymbol{\theta}) \mu(dx_i)}_{= 1} = 0. \end{aligned}$$

□

Weiterhin wichtig ist folgendes

Lemma 2.2.6. *Es sei $\mathbf{X} = (X_1, \dots, X_n)$ eine Stichprobe bestehend aus i.i.d. Zufallsvariablen X_1, \dots, X_n mit regulärer Verteilung $P_{\boldsymbol{\theta}}$. Dann gilt:*

$$\mathcal{I}(\boldsymbol{\theta}; \mathbf{X}) = -\mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{S}(\boldsymbol{\theta}; \mathbf{X})^T \right]. \quad (2.2)$$

Beweis. Es gilt

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{x})^T &= \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}; \boldsymbol{x}) \right)^T = \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\frac{\partial}{\partial \boldsymbol{\theta}} f(\boldsymbol{x}; \boldsymbol{\theta})}{f(\boldsymbol{x}; \boldsymbol{\theta})} \right)^T \\ &= \frac{\left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} f(\boldsymbol{x}; \boldsymbol{\theta}) \right) f(\boldsymbol{x}; \boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}} f(\boldsymbol{x}; \boldsymbol{\theta}) \left(\frac{\partial}{\partial \boldsymbol{\theta}} f(\boldsymbol{x}; \boldsymbol{\theta}) \right)^T}{f(\boldsymbol{x}; \boldsymbol{\theta})^2} \\ &= \frac{\frac{\partial^2}{\partial \boldsymbol{\theta}^2} f(\boldsymbol{x}; \boldsymbol{\theta})}{f(\boldsymbol{x}; \boldsymbol{\theta})} - \underbrace{\frac{\frac{\partial}{\partial \boldsymbol{\theta}} f(\boldsymbol{x}; \boldsymbol{\theta})}{f(\boldsymbol{x}; \boldsymbol{\theta})}}_{= \boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{x})} \underbrace{\frac{\left(\frac{\partial}{\partial \boldsymbol{\theta}} f(\boldsymbol{x}; \boldsymbol{\theta}) \right)^T}{f(\boldsymbol{x}; \boldsymbol{\theta})}}_{= \boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{x})^T}. \end{aligned}$$

Integriert man die Gleichung bezüglich $P_{\boldsymbol{\theta}}$, so folgt wegen $dP_{\boldsymbol{\theta}} = f(\boldsymbol{x}; \boldsymbol{\theta}) d\mu$

$$\mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{X})^T \right] = \int \frac{\partial^2}{\partial \boldsymbol{\theta}^2} f(\boldsymbol{x}; \boldsymbol{\theta}) \mu(d\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{\theta}} [\boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{X}) \boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{X})^T]$$

und mit $\int \frac{\partial^2}{\partial \boldsymbol{\theta}^2} f(\boldsymbol{x}; \boldsymbol{\theta}) \mu(d\boldsymbol{x}) = \underbrace{\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \int f(\boldsymbol{x}; \boldsymbol{\theta}) \mu(d\boldsymbol{x})}_{= 1} = 0$ die Behauptung. \square

Betrachtet man reguläre Verteilungsfamilien, so lässt sich mit Lemma 2.2.5 für i.i.d. Zufallsvariablen X_1, \dots, X_n eine weitere Darstellungsmöglichkeit der Fisher-Information angeben. Es gilt

$$\boldsymbol{I}(\boldsymbol{\theta}; X_i) = \text{Cov}[\boldsymbol{S}(\boldsymbol{\theta}; X_i)].$$

Dabei bezeichnet $\text{Cov}[\boldsymbol{S}(\boldsymbol{\theta}; X_i)] = (\text{Cov}[\boldsymbol{S}_j(\boldsymbol{\theta}; X_i), \boldsymbol{S}_k(\boldsymbol{\theta}; X_i)])_{1 \leq j, k \leq d}$ die *Kovarianzmatrix* von $\boldsymbol{S}(\boldsymbol{\theta}; X_i)$. Unter Beachtung der Unabhängigkeit der X_i für $i = 1, \dots, n$ lässt sich weiter zeigen, dass $\boldsymbol{I}(\boldsymbol{\theta}; \boldsymbol{X}) = \sum_{i=1}^n \boldsymbol{I}(\boldsymbol{\theta}; X_i)$ und somit speziell für identisch verteilte Zufallsvariablen

$$\boldsymbol{I}(\boldsymbol{\theta}; \boldsymbol{X}) = n \boldsymbol{I}(\boldsymbol{\theta}; X_i) \tag{2.3}$$

gilt. Die eingangs erwähnte Eigenschaft der asymptotischen Normalverteilung des MLS soll in einem Satz festgehalten und unter Annahme der Regularitätsbedingungen bewiesen werden. Dazu wird das Konzept der multivariaten Normalverteilung benötigt, welches im Folgenden in ausreichender Kürze mit einigen dazugehörigen Resultaten vorgestellt wird. Ab jetzt wird angenommen, dass die Likelihood-Funktion der zugrunde liegenden Verteilung den Regularitätsbedingungen genüge. Als Erweiterung des Prinzips normalverteilter Zufallsvariablen (univariate Normalverteilung) werden nun Zufallsvektoren betrachtet, denen ebenfalls eine Normalverteilung (multivariate Normalverteilung) zugeordnet wird. Dabei wird auf den Fall *nichtsingulärer* Kovarianzmatrizen eingegangen. Die Definition der multivariaten Normalverteilung sowie wichtige Eigenschaften stammen aus [SHF85].

Definition 2.2.7. Es sei $\mathbf{X} = (X_1, \dots, X_p)^T$ ein p -dimensionaler Zufallsvektor mit Werten im \mathbb{R}^p . Dann besitzt \mathbf{X} eine p -variante nichtsinguläre Normalverteilung mit Erwartungswertvektor $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^T$ und positiv definiter Kovarianzmatrix $\boldsymbol{\Sigma}$, falls die Dichte von \mathbf{X} gegeben ist durch

$$f(\mathbf{x}) = |(2\pi)\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

mit $\mu_j \in \mathbb{R}$ für $j = 1, \dots, p$ und $\boldsymbol{\Sigma} > 0$. $|\boldsymbol{\Sigma}|$ bezeichnet dabei die Determinante der Matrix $\boldsymbol{\Sigma}$. Man schreibt $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Einige für den späteren Kontext wichtige Eigenschaften lauten wie folgt:

Lemma 2.2.8. Es sei $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit $\boldsymbol{\Sigma} > 0$. Dann gilt:

$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_p^2.$$

Beweis. Dies folgt direkt aus einer Zerlegung $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^T$ mit \mathbf{C} als $(p \times p)$ -Matrix und der Definition der χ^2 -Verteilung. Eine Darstellung von \mathbf{C} in vorliegender Form ist mit Hilfe der Cholesky-Zerlegung² möglich. \square

Lemma 2.2.9. Es sei $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit $\boldsymbol{\Sigma} > 0$. Weiter sei \mathbf{A} eine $(k \times p)$ -Matrix mit vollem Rang, $k \leq p$ und \mathbf{b} ein k -Vektor. Dann gilt:

$$\mathbf{A}\mathbf{X} + \mathbf{b} \sim \mathcal{N}_k(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T).$$

Beweis. Der Beweis ergibt sich aus einer Darstellung der charakteristischen Funktion von $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$. Dazu sei \mathbf{t} ein k -Vektor mit $\mathbf{t} \neq \mathbf{0}$. Dann gilt für die charakteristische Funktion von \mathbf{Y} : $\phi_{\mathbf{Y}}(\mathbf{t}) = \phi_{\mathbf{A}\mathbf{X}}(\mathbf{t})\phi_{\mathbf{b}}(\mathbf{t}) = \mathbb{E}[\exp(it^T \mathbf{A}\mathbf{X})] \exp(it^T \mathbf{b}) = \phi_{\mathbf{X}}(\mathbf{A}^T \mathbf{t}) \exp(it^T \mathbf{b})$. Nun gilt für $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit $\boldsymbol{\Sigma} > 0$: $\phi_{\mathbf{X}}(\mathbf{t}) = \exp(it^T \boldsymbol{\mu} - \frac{1}{2} \mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t})$ und somit $\phi_{\mathbf{Y}}(\mathbf{t}) = \exp(it^T \mathbf{b}) \exp(it^T \mathbf{A}\boldsymbol{\mu} - \frac{1}{2} \mathbf{t}^T \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T \mathbf{t}) = \exp(it^T (\mathbf{A}\boldsymbol{\mu} + \mathbf{b}) - \frac{1}{2} \mathbf{t}^T (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T) \mathbf{t})$. Dies ist wiederum die charakteristische Funktion eines $\mathcal{N}_k(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$ -verteilten Zufallsvektors mit $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T > 0$. \square

Lemma 2.2.10. Es sei $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit $\boldsymbol{\Sigma} > 0$, \mathbf{X}_1 sei r -Vektor und \mathbf{X}_2 sei $(p - r)$ -Vektor. Weiter sei $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T)^T$ und $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$. Dann gilt:

$$\mathbf{X}_2 | \mathbf{X}_1 \sim \mathcal{N}_{p-r}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}).$$

²Vgl. Anhang A.2.

Beweis. Es sei $\mathbf{X}_{2.1} = \mathbf{X}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1$. Dann gilt $\mathbf{X}_{2.1} \sim \mathcal{N}_{p-r}(\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$ was leicht mit der Darstellung $\mathbf{X}_{2.1} = \begin{pmatrix} -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_{p-r} \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ und Lemma 2.2.9 eingesehen werden kann. Es sei nun $\mathbf{X}_2 = \mathbf{X}_{2.1} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1$. Unter Bedingungen von \mathbf{X}_1 ist $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1$ konstant, sodass $\mathbb{E}[\mathbf{X}_2|\mathbf{X}_1] = \mathbb{E}[\mathbf{X}_{2.1}] + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1 = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1)$ und $\text{Var}[\mathbf{X}_2|\mathbf{X}_1] = \text{Var}[\mathbf{X}_{2.1}] = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. \square

Um nun die angesprochenen Eigenschaften des MLS herleiten zu können, wird mit einem Prinzip begonnen, das unabhängig von der gewählten Schätzmethode Bedeutung findet.

Definition 2.2.11. Es seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit Verteilung P_θ und $\hat{\theta}_n$ sei ein Schätzer für θ auf der Basis X_1, \dots, X_n . Dann heißt $\hat{\theta}_n$ (*schwach*) *konsistent*, wenn $\hat{\theta}_n$ in Wahrscheinlichkeit gegen θ konvergiert, das heißt $\lim_{n \rightarrow \infty} P(\|\hat{\theta}_n - \theta\| > \epsilon) = 0$ für alle $\epsilon > 0$. $\|\cdot\|$ bezeichnet dabei die euklidische Norm.

Anschaulich bedeutet die Konsistenz eines Schätzers also, dass sich der Schätzer dem zu schätzenden Parameter bei zunehmender Beobachtungsanzahl immer weiter nähert.

Lemma 2.2.12. *Es seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit Verteilung P_θ . Sind die Regularitätsbedingungen erfüllt, dann ist der MLS $\hat{\theta}$ auf der Basis X_1, \dots, X_n konsistent.*

Der Beweis des Lemmas ist überwiegend konstruktiver Natur und im folgenden Zusammenhang nicht weiter von Relevanz. Aus diesem Grund wird von genauen Ausführungen abgesehen und der interessierte Leser auf [Wal49], S. 595-601 verwiesen.

Betrachtet man den MLS $\hat{\theta}$ nicht als Funktion bezüglich einer festen Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$, sondern als Funktion in den Zufallsvariablen X_1, \dots, X_n , so kann der Schätzer selbst als Zufallsvariable interpretiert werden. Dies ermöglicht es, dem Schätzer eine Verteilung zuzuweisen, die mit Hilfe wahrscheinlichkeitstheoretischer Mittel hergeleitet werden kann.

Satz 2.2.13. *Es seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit regulärer Verteilung P_θ , $\theta = (\theta_1, \dots, \theta_d)^T$ und $\hat{\theta}$ sei MLS für θ auf der Basis X_1, \dots, X_n . Dann gilt:*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}_d(\mathbf{0}, \mathcal{I}(\theta; X_1)^{-1}).$$

\xrightarrow{D} bezeichnet dabei die Konvergenz in Verteilung.

Beweis. Mittels der Taylorformel, die aufgrund von RB 3(a) angewendet werden darf, werden die Komponenten des Score-Vektors $(\mathcal{S}_1(\hat{\theta}; \mathbf{X}), \dots, \mathcal{S}_d(\hat{\theta}; \mathbf{X}))^T = \mathcal{S}(\hat{\theta}; \mathbf{X})$ mit

$\mathcal{S}_j(\hat{\boldsymbol{\theta}}; \mathbf{X}) = \frac{\partial}{\partial \theta_j} l(\hat{\boldsymbol{\theta}}; \mathbf{X})$ für $j = 1, \dots, d$ um den wahren Parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$ entwickelt. Es gilt

$$\mathcal{S}_j(\hat{\boldsymbol{\theta}}; \mathbf{X}) = \mathcal{S}_j(\boldsymbol{\theta}; \mathbf{X}) + \sum_{k=1}^d \frac{\partial}{\partial \theta_k} \mathcal{S}_j(\boldsymbol{\theta}; \mathbf{X})(\hat{\theta}_k - \theta_k) + \frac{1}{2} \sum_{k,l=1}^d \frac{\partial^2}{\partial \theta_k \partial \theta_l} \mathcal{S}_j(\tilde{\boldsymbol{\theta}}; \mathbf{X})(\hat{\theta}_k - \theta_k)(\hat{\theta}_l - \theta_l)$$

mit $\tilde{\boldsymbol{\theta}}$ geeignet, wobei $|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}| < |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|$. Da $\mathcal{S}_j(\hat{\boldsymbol{\theta}}; \mathbf{X}) = 0$ folgt nach Division durch \sqrt{n}

$$\begin{aligned} \frac{\mathcal{S}_j(\boldsymbol{\theta}; \mathbf{X})}{\sqrt{n}} &= - \frac{\sum_{k=1}^d \frac{\partial}{\partial \theta_k} \mathcal{S}_j(\boldsymbol{\theta}; \mathbf{X})(\hat{\theta}_k - \theta_k)}{\sqrt{n}} - \frac{1}{2} \frac{\sum_{k,l=1}^d \frac{\partial^2}{\partial \theta_k \partial \theta_l} \mathcal{S}_j(\tilde{\boldsymbol{\theta}}; \mathbf{X})(\hat{\theta}_k - \theta_k)(\hat{\theta}_l - \theta_l)}{\sqrt{n}} \\ &= \sqrt{n} \left(- \frac{1}{n} \sum_{k=1}^d \frac{\partial}{\partial \theta_k} \mathcal{S}_j(\boldsymbol{\theta}; \mathbf{X})(\hat{\theta}_k - \theta_k) - \underbrace{\frac{1}{2n} \sum_{k,l=1}^d \frac{\partial^2}{\partial \theta_k \partial \theta_l} \mathcal{S}_j(\tilde{\boldsymbol{\theta}}; \mathbf{X})(\hat{\theta}_k - \theta_k)(\hat{\theta}_l - \theta_l)}_{=: Z_n} \right). \end{aligned}$$

Nach RB 3(b) existiert eine Funktion $g(\mathbf{X})$, sodass $n^{-1} \left| \frac{\partial^2}{\partial \theta_k \partial \theta_l} \mathcal{S}_j(\tilde{\boldsymbol{\theta}}; \mathbf{X}) \right| < g(\mathbf{X}) < \infty$. Da $\hat{\boldsymbol{\theta}}$ aufgrund von Lemma 2.2.12 konsistent ist, folgt $|Z_n| < |\hat{\theta}_k - \theta_k| |\hat{\theta}_l - \theta_l| g(\mathbf{X}) \xrightarrow{n \rightarrow \infty} 0$ und somit $Z_n = o_p(1)$. Dabei gilt $Z_n = o_p(f(n)) \Leftrightarrow \lim_{n \rightarrow \infty} P\left(\left|\frac{Z_n}{f(n)}\right| \leq \epsilon\right) = 1$ für alle $\epsilon > 0$ mit einer Folge $(Z_n)_{n \in \mathbb{N}_{>0}}$ von Zufallsvariablen und einer Funktion $f: \mathbb{N}_{>0} \rightarrow \mathbb{R}$. Also ist

$$\frac{\mathcal{S}_j(\boldsymbol{\theta}; \mathbf{X})}{\sqrt{n}} = \sqrt{n} \left(- \frac{1}{n} \sum_{k=1}^d \frac{\partial}{\partial \theta_k} \left(\sum_{i=1}^n \mathcal{S}_j(\boldsymbol{\theta}; X_i) \right) (\hat{\theta}_k - \theta_k) - o_p(1) \right),$$

wobei mit dem schwachen Gesetz der großen Zahlen gilt:

$$\begin{aligned} - \frac{1}{n} \sum_{k=1}^d \frac{\partial}{\partial \theta_k} \left(\sum_{i=1}^n \mathcal{S}_j(\boldsymbol{\theta}; X_i) \right) (\hat{\theta}_k - \theta_k) &\xrightarrow{n \rightarrow \infty} - \sum_{k=1}^d \frac{\partial}{\partial \theta_k} \mathbb{E}_{\boldsymbol{\theta}} [\mathcal{S}_j(\boldsymbol{\theta}; X_1)] (\hat{\theta}_k - \theta_k) \\ &\stackrel{\text{RB 4}}{=} - \mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \theta_1} \mathcal{S}_j(\boldsymbol{\theta}; X_1) \right] (\hat{\theta}_1 - \theta_1) - \dots \\ &\quad - \mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \theta_d} \mathcal{S}_j(\boldsymbol{\theta}; X_1) \right] (\hat{\theta}_d - \theta_d) \\ &\stackrel{(2.2)}{=} \mathcal{I}_{j1}(\boldsymbol{\theta}; X_1) (\hat{\theta}_1 - \theta_1) + \dots + \mathcal{I}_{jd}(\boldsymbol{\theta}; X_1) (\hat{\theta}_d - \theta_d) \\ &= (\mathcal{I}_{j1}(\boldsymbol{\theta}; X_1), \dots, \mathcal{I}_{jd}(\boldsymbol{\theta}; X_1)) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \end{aligned}$$

in Wahrscheinlichkeit. Es folgt

$$\frac{1}{\sqrt{n}} \mathcal{S}_j(\boldsymbol{\theta}; \mathbf{X}) \stackrel{a}{=} \sqrt{n} (\mathcal{I}_{j1}(\boldsymbol{\theta}; X_1), \dots, \mathcal{I}_{jd}(\boldsymbol{\theta}; X_1)) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

und folglich mit $\mathcal{I}(\boldsymbol{\theta}; X_1) = \begin{pmatrix} \mathcal{I}_{11}(\boldsymbol{\theta}; X_1) & \cdots & \mathcal{I}_{1d}(\boldsymbol{\theta}; X_1) \\ \vdots & \ddots & \vdots \\ \mathcal{I}_{d1}(\boldsymbol{\theta}; X_1) & \cdots & \mathcal{I}_{dd}(\boldsymbol{\theta}; X_1) \end{pmatrix}$

$$\begin{aligned} \frac{1}{\sqrt{n}} \mathcal{S}(\boldsymbol{\theta}; \mathbf{X}) &\stackrel{a}{=} \sqrt{n} \mathcal{I}(\boldsymbol{\theta}; X_1) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ \Leftrightarrow \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &\stackrel{a}{=} \frac{1}{\sqrt{n}} \mathcal{I}(\boldsymbol{\theta}; X_1)^{-1} \mathcal{S}(\boldsymbol{\theta}; \mathbf{X}). \end{aligned} \tag{2.4}$$

$\stackrel{a}{=}$ steht dabei für asymptotisch äquivalent, sodass für große Stichproben der linke und der rechte Term annähernd identisch sind. Da $\mathbb{E}_{\boldsymbol{\theta}}[\mathcal{S}_j(\boldsymbol{\theta}; X_i)] = 0$ für $i = 1, \dots, n$, $j = 1, \dots, d$ wegen Lemma 2.2.5 und $\mathbb{E}_{\boldsymbol{\theta}}[\mathcal{S}_j(\boldsymbol{\theta}; X_1)\mathcal{S}_k(\boldsymbol{\theta}; X_1)] = \mathcal{I}_{jk}(\boldsymbol{\theta}; X_1)$ für $j, k = 1, \dots, d$ wegen Definition 2.2.4, folgt mit dem mehrdimensionalen zentralen Grenzwertsatz³ und Lemma 2.2.9

$$\frac{1}{\sqrt{n}}\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}; X_1)^{-1}\boldsymbol{\mathcal{S}}(\boldsymbol{\theta}; \mathbf{X}) = \sqrt{n}\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}; X_1)^{-1}\left(\frac{1}{n}\sum_{i=1}^n\boldsymbol{\mathcal{S}}(\boldsymbol{\theta}; X_i)\right) \xrightarrow{D} \mathcal{N}_d\left(\mathbf{0}, \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}; X_1)^{-1}\right)$$

und somit die Behauptung. \square

Es lässt sich zeigen, dass die Aussage der asymptotischen Normalverteilung der Likelihood-Funktion auch unter schwächeren Annahmen gültig bleibt.⁴ (Diese implizieren beispielsweise dass die Forderung RB 3(a) zu restriktiv ist.) Dies soll hier jedoch nicht näher beleuchtet werden.

2.3 Grundlagen der Testtheorie

Im Folgenden wird das Ziel sein, einen Test zu entwickeln, mit dessen Hilfe Entscheidungen über die Richtigkeit einer Hypothese getroffen werden können. Grundlegend lässt sich die Schätz- von der Testtheorie dahingehend unterscheiden, dass beim Schätzen von Parametern bestimmte Werte für die unbekannt Parameter anhand von Plausibilitäten festgelegt werden, wohingegen beim Testen von Hypothesen Plausibilitäten für bereits festgelegte Werte überprüft werden. Man unterscheidet dabei zwischen einer *Nullhypothese* $H_0 : \boldsymbol{\theta} \in \Theta_0 \subseteq \Theta$ und einer *Alternativhypothese* $H_1 : \boldsymbol{\theta} \in \Theta \setminus \Theta_0$, wobei $\Theta \subseteq \mathbb{R}^d$ den Parameterraum einer bekannten Verteilungsfamilie $(P_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \Theta}$ darstellt. Auf der Grundlage von Beobachtungen, simuliert durch Realisierungen von i.i.d. Zufallsvariablen, werden Indizien für das Annehmen oder Verwerfen der Nullhypothese gesammelt. Die Entscheidungen für oder gegen die Nullhypothese werden dabei modelliert durch eine messbare Abbildung $\delta : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{D}, \mathcal{E})$ vom Raum $(\mathcal{X}, \mathcal{A})$ der möglichen Realisierungen in den Raum $(\mathcal{D}, \mathcal{E})$ der möglichen Entscheidungen. In den folgenden Testproblemen entspricht \mathcal{D} dem abgeschlossenen Einheitsintervall, das heißt $\mathcal{D} = [0, 1]$. Eine *Entscheidungsfunktion* $\phi : \mathcal{X} \rightarrow [0, 1]$ wird als *Test* bezeichnet, sodass auf der Grundlage einer Stichprobe $\mathbf{x} \in \mathcal{X}$ der Test $\phi(\mathbf{x}) = \gamma$ mit $\gamma \in [0, 1]$, eine Entscheidung mit Wahrscheinlichkeit γ für die Alternativhypothese liefert. Im Fall $\gamma \in (0, 1)$ bedeutet dies, dass die Entscheidung keineswegs eindeutig zugunsten einer der Hypothesen getroffen wird. Dies modelliert eine

³Vgl. Anhang A.1.

⁴Vgl. [LeC70], S.802-828.

Art Unentschlossenheit zwischen den Entscheidungsmöglichkeiten und wird als *Randomisieren* bezeichnet. In *nicht randomisierten* Testproblemen, das heißt $\gamma \in \{0, 1\}$, hingegen wird eine eindeutige Entscheidung getroffen, die mit $\phi(\mathbf{x}) = 1$ für die Alternativhypothese, beziehungsweise gegen die Nullhypothese und mit $\phi(\mathbf{x}) = 0$ für die Nullhypothese ausfällt. Dabei ist zu beachten, dass eine Entscheidung gegen die Alternativhypothese, das heißt also eine Annahme der Nullhypothese aufgrund einer Stichprobe nicht gleichzusetzen ist mit einem Beweis für die Richtigkeit der Nullhypothese, sondern lediglich, dass die Nullhypothese aufgrund mangelnder Beweislage nicht verworfen werden kann (vergleichbar dem juristischen Grundsatz „in dubio pro reo“). In den vorliegenden Testproblemen können somit zwei mögliche Fehler auftreten:

Fehler 1. Art \triangleq Verwerfen der Nullhypothese, obwohl diese richtig ist.

Fehler 2. Art \triangleq Annahme der Nullhypothese, obwohl diese falsch ist.

Die Auswirkungen beider Fehler differieren in aller Regel sehr stark. Da es im Allgemeinen keinen Test gibt, der die Wahrscheinlichkeiten beider Fehler simultan minimiert, die irrtümliche Entscheidung für die Alternativhypothese jedoch schwerwiegendere Auswirkungen hat als die irrtümliche Annahme der zumeist konservativen Nullhypothese, ist es von Interesse, das Risiko eines Fehlers 1. Art unter einem vorgegebenen *Signifikanzniveau* $\alpha \in [0, 1]$ zu halten und unter diesen Tests denjenigen zu wählen, der die Fehlerwahrscheinlichkeit 2. Art minimiert. α gibt dabei die maximale Wahrscheinlichkeit an, mit der eine irrtümliche Annahme der Alternativhypothese toleriert wird, was unter Verwendung der *Gütefunktion* $\beta_\phi : \boldsymbol{\theta} \mapsto \mathbb{E}_\boldsymbol{\theta}[\phi(\mathbf{X})]$ bedeutet, einen Test ϕ mit $\mathbb{E}_\boldsymbol{\theta}[\phi(\mathbf{X})] \leq \alpha$ für alle $\boldsymbol{\theta} \in \Theta_0$ zu finden. Ein Test, der diese Eigenschaft erfüllt, wird als *Test zum Niveau* α bezeichnet und die Menge all solcher Tests zum Niveau α wird als Φ_α definiert. Da es weiter von Interesse ist unter den Tests $\phi \in \Phi_\alpha$ denjenigen zu wählen, der die Fehlerwahrscheinlichkeit 2. Art minimiert – dies entspricht einer Maximierung der Gütefunktion in $\boldsymbol{\theta} \in \Theta_1$ – gilt für den *gleichmäßig besten Test* ϕ_0 zum Niveau α :

$$\mathbb{E}_\boldsymbol{\theta}[\phi_0(\mathbf{X})] = \max_{\phi \in \Phi_\alpha} \mathbb{E}_\boldsymbol{\theta}[\phi(\mathbf{X})]$$

für alle $\boldsymbol{\theta} \in \Theta_1$. Betrachtet wird im Folgenden die auf Neyman und Pearson zurückgehende Entdeckung zur Beschreibung gleichmäßig bester Tests in Testproblemen mit einfachen Hypothesen.

Satz 2.3.1. *Es seien $P_{\boldsymbol{\theta}_0}$ und $P_{\boldsymbol{\theta}_1}$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{A})$ mit Dichten $f(x; \boldsymbol{\theta}_0)$ und $f(x; \boldsymbol{\theta}_1)$ bezüglich eines dominierenden Maßes μ . Weiter sei $\alpha \in (0, 1)$ und $c \in [0, \infty)$. Dann gilt:*

Ist $\psi \in \Phi_\alpha$ mit $\psi(x) = \mathbb{1}_{\{f(x; \boldsymbol{\theta}_1) > c f(x; \boldsymbol{\theta}_0)\}}$ und c so gewählt, dass $\mathbb{E}_{\boldsymbol{\theta}_0}[\psi(X)] = \alpha$, dann ist

$\mathbb{E}_{\theta_1}[\psi(X)] = \max_{\phi \in \Phi_\alpha} \mathbb{E}_{\theta_1}[\phi(X)]$, das heißt ψ ist gleichmäßig bester Test zum Niveau α für $H_0 : \theta = \theta_0$ gegen $H_1 : \theta = \theta_1$.

Beweis. Sei $\phi \in \Phi_\alpha$ beliebig. Dann gilt:

$$\begin{aligned} f(x; \theta_1) - cf(x; \theta_0) > 0 &\Rightarrow \psi(x) = 1 \Rightarrow \psi(x) - \phi(x) \geq 0 \quad \text{und} \\ f(x; \theta_1) - cf(x; \theta_0) < 0 &\Rightarrow \psi(x) = 0 \Rightarrow \psi(x) - \phi(x) \leq 0. \end{aligned}$$

Somit ist $(\psi(x) - \phi(x))(f(x; \theta_1) - cf(x; \theta_0)) \geq 0$. Integration bezüglich μ liefert

$$\begin{aligned} 0 &\leq \int \psi(x)f(x; \theta_1)d\mu - \int \phi(x)f(x; \theta_1)d\mu \\ &\quad - c \left(\int \psi(x)f(x; \theta_0)d\mu - \int \phi(x)f(x; \theta_0)d\mu \right) \\ \Leftrightarrow 0 &\leq \mathbb{E}_{\theta_1}[\psi(X)] - \mathbb{E}_{\theta_1}[\phi(X)] - c \left(\underbrace{\mathbb{E}_{\theta_0}[\psi(X)]}_{=\alpha} - \underbrace{\mathbb{E}_{\theta_0}[\phi(X)]}_{\leq \alpha} \right). \end{aligned}$$

Also folgt $0 \leq \mathbb{E}_{\theta_1}[\psi(X)] - \mathbb{E}_{\theta_1}[\phi(X)]$ und damit die Behauptung. \square

Der Ablehnungsbereich K_α des Neyman-Pearson-Tests wird also im Falle einfacher Hypothesen über den Quotienten $f(x; \theta_1)/f(x; \theta_0)$ bestimmt. Es gilt $K_\alpha = \{x : \frac{f(x; \theta_1)}{f(x; \theta_0)} > c\}$, wobei c so zu wählen ist, dass das Signifikanzniveau α voll ausgeschöpft wird. Es sei darauf hingewiesen, dass das Neyman-Pearson-Lemma für Wahrscheinlichkeitsdichten vorgestellt wurde. Für diskrete Wahrscheinlichkeitsverteilungen muß die Testgestalt leicht verändert werden, da eine Randomisierung nötig ist, um eine Ausschöpfung des Signifikanzniveaus α zu gewährleisten.

Dehnt man die Situation einfacher Hypothesen auf den in der Realität wesentlich interessanteren Fall zweiseitiger Alternativhypothesen aus, das heißt $H_0 : \theta \in [\theta_1, \theta_2]$ gegen $H_1 : \theta \notin [\theta_1, \theta_2]$ mit $\theta_1, \theta_2 \in \Theta \subseteq \mathbb{R}$ und $\theta_1 \leq \theta_2$, so muss man feststellen, dass gleichmäßig beste Tests in der Regel nicht existieren. Um weiterhin optimale Tests zu erhalten muss das Prinzip der Unverfälschtheit eingeführt werden. Dieses besagt, dass neben den vorigen Annahmen die zusätzliche Bedingung $\mathbb{E}_\theta[\phi(\mathbf{X})] \geq \alpha$ für alle $\theta \in \Theta_1$ erfüllt sein muss. Bezeichnet man mit $\Phi_{\alpha, u}$ die Menge aller *unverfälschten Tests zum Niveau α* , so erfüllt der *gleichmäßig beste unverfälschte Test ϕ_0 zum Niveau α* folglich die Bedingungen $\phi_0 \in \Phi_{\alpha, u}$ und

$$\mathbb{E}_\theta[\phi_0(\mathbf{X})] = \max_{\phi \in \Phi_{\alpha, u}} \mathbb{E}_\theta[\phi(\mathbf{X})]$$

für alle $\theta \in \Theta_1$. Bei Vorliegen einer einparametrischen Exponentialfamilie lässt sich auch in dieser Situation eine Testgestalt ähnlich zu der in der Situation einfacher Hypothesen angeben.⁵ Dies impliziert, dass die Verteilung der Statistik, anhand welcher eine Entscheidung für oder gegen die Alternativhypothese getroffen wird, bekannt sein muss, um die

⁵Für eine detaillierte Einführung in die Testtheorie sei auf [Als09] verwiesen.

für das Einhalten der Irrtumswahrscheinlichkeit notwendigen kritischen Werte zu bestimmen. Für einige Verteilungen stehen dabei Tafelwerke zur Verfügung, die das Festlegen der kritischen Werte bei gegebenem α erleichtern.⁶ Oft ist es auch möglich, eine nicht vertafelte Verteilung mit Hilfe einer Transformation in eine vertafelte Verteilung zu überführen. Gelingt dies nicht oder ist die Verteilung der Teststatistik nicht bekannt, so ist die Bestimmung der kritischen Werte wesentlich schwieriger. Ein Ausweg aus dieser Problematik wird im Folgenden gegeben.

2.4 Herleitung des Score-Tests

Damit der Score-Test hergeleitet werden kann, muss man sich ausführlich mit der Theorie der Likelihood-Quotienten-Tests beschäftigen. Dazu ist der vorliegende Abschnitt so aufgebaut, dass zu Beginn Likelihood-Quotienten-Tests ohne nuisance-Parameter eingeführt werden, bevor im Anschluß Likelihood-Quotienten-Tests mit nuisance-Parametern untersucht werden.

Likelihood-Quotienten-Tests ohne nuisance-Parameter

Es wird mit einem Spezialfall der Situation des vorigen Abschnitts fortgefahren, nämlich einfachen Nullhypothesen bei zweiseitigen Alternativen. Dazu sei erneut $\mathbf{x} = (x_1, \dots, x_n)$ eine Stichprobe bestehend aus i.i.d. Zufallsvariablen X_1, \dots, X_n und

$$\frac{\sup_{\boldsymbol{\theta}' \in \Theta_1} L(\boldsymbol{\theta}'; \mathbf{x})}{L(\boldsymbol{\theta}_0; \mathbf{x})} = \exp\left(\frac{1}{2}\lambda\right)$$

der *Likelihood-Quotient* mit $\Theta_1 \subseteq \Theta$. Dann gilt im Fall $\Theta_1 = \Theta \setminus \boldsymbol{\theta}_0$

$$\lambda = 2\left(\sup_{\boldsymbol{\theta}' \in \Theta \setminus \boldsymbol{\theta}_0} l(\boldsymbol{\theta}'; \mathbf{x}) - l(\boldsymbol{\theta}_0; \mathbf{x})\right)$$

und mit $\Lambda = \max(0, \lambda)$ folgt

$$\Lambda = 2(l(\hat{\boldsymbol{\theta}}; \mathbf{x}) - l(\boldsymbol{\theta}_0; \mathbf{x})).$$

Λ wird dabei als *Likelihood-Quotienten-Test* bezeichnet. Obwohl für $n < \infty$ die Verteilung von Λ sowohl von n als auch von der Wahrscheinlichkeitsdichte der X_i abhängt, kann gezeigt werden, dass in regulären Problemen die asymptotische Verteilung von Λ für $n \rightarrow \infty$ ein einheitliches Ergebnis liefert. Dies ermöglicht es, approximativ für große Stichproben unabhängig von der Verteilung von Λ die kritischen Werte anhand der α -Fraktile der asymptotischen Verteilung festzumachen. Umso erfreulicher ist die Tatsache, dass es

⁶Hierzu zählen unter anderem die Standardnormalverteilung und die Chi-Quadrat-Verteilung. Einige Fraktiltabellen sind beispielsweise in [Rin08], Kapitel E1 angegeben.

sich bei der asymptotischen Verteilung um die gut vertafelte Chi-Quadrat-Verteilung mit Anzahl der Freiheitsgrade entsprechend der Dimension von $\boldsymbol{\theta}$ handelt.

Satz 2.4.1. *Es seien x_1, \dots, x_n Realisierungen von i.i.d. Zufallsvariablen X_1, \dots, X_n , die gemäß einer regulären Verteilung $P_{\boldsymbol{\theta}}$ gezogen werden. Dann gilt für die Verteilung von Λ bei einer Nullhypothese der Form $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 = (\theta_{10}, \dots, \theta_{d0})^T$ gegen $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$:*

$$\Lambda \xrightarrow{D} \chi_{\dim(\boldsymbol{\theta})}^2.$$

Beweis. Unter Annahme der Nullhypothese $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \in \Theta$ wird Λ um $\boldsymbol{\theta}_0$ in eine Taylorreihe entwickelt. Dazu sei erneut $\mathcal{S}_k(\boldsymbol{\theta}; \mathbf{X}) = \frac{\partial}{\partial \theta_k} l(\boldsymbol{\theta}; \mathbf{X})$ für $k = 1, \dots, d$. Es gilt:

$$\begin{aligned} l(\hat{\boldsymbol{\theta}}; \mathbf{X}) &= l(\boldsymbol{\theta}_0; \mathbf{X}) + \sum_{k=1}^d \mathcal{S}_k(\boldsymbol{\theta}_0; \mathbf{X})(\hat{\theta}_k - \theta_{k0}) + \frac{1}{2} \sum_{k,m=1}^d \frac{\partial}{\partial \theta_m} \mathcal{S}_k(\tilde{\boldsymbol{\theta}}; \mathbf{X})(\hat{\theta}_k - \theta_{k0})(\hat{\theta}_m - \theta_{m0}) \\ &= l(\boldsymbol{\theta}_0; \mathbf{X}) + \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{1}{2} \sum_{m=1}^d (\hat{\theta}_m - \theta_{m0}) \frac{\partial}{\partial \theta_m} \mathbf{S}(\tilde{\boldsymbol{\theta}}; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \end{aligned} \quad (2.5)$$

mit $\tilde{\boldsymbol{\theta}}$ geeignet, wobei $|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|$ und

$$\begin{aligned} \mathcal{S}_j(\hat{\boldsymbol{\theta}}; \mathbf{X}) &= \mathcal{S}_j(\boldsymbol{\theta}_0; \mathbf{X}) + \sum_{k=1}^d \frac{\partial}{\partial \theta_k} \mathcal{S}_j(\tilde{\boldsymbol{\theta}}; \mathbf{X})(\hat{\theta}_k - \theta_{k0}) \\ &= \mathcal{S}_j(\boldsymbol{\theta}_0; \mathbf{X}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{S}_j(\tilde{\boldsymbol{\theta}}; \mathbf{X}) \right)^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \end{aligned}$$

mit $\check{\boldsymbol{\theta}}$ geeignet, wobei $|\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|$. Es folgt

$$\begin{aligned} \mathbf{S}(\hat{\boldsymbol{\theta}}; \mathbf{X}) &= \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\tilde{\boldsymbol{\theta}}; \mathbf{X})^T \right)^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \Leftrightarrow \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X}) &= \mathbf{S}(\hat{\boldsymbol{\theta}}; \mathbf{X}) - \left(\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\tilde{\boldsymbol{\theta}}; \mathbf{X})^T \right)^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0). \end{aligned} \quad (2.6)$$

Somit ist

$$\begin{aligned} \Lambda &= 2(l(\hat{\boldsymbol{\theta}}; \mathbf{X}) - l(\boldsymbol{\theta}_0; \mathbf{X})) \\ &\stackrel{(2.5)}{=} 2\left(\mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{1}{2} \sum_{m=1}^d (\hat{\theta}_m - \theta_{m0}) \frac{\partial}{\partial \theta_m} \mathbf{S}(\tilde{\boldsymbol{\theta}}; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right) \\ &\stackrel{(2.6)}{=} 2\left(\underbrace{\mathbf{S}(\hat{\boldsymbol{\theta}}; \mathbf{X})^T}_{=0} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\tilde{\boldsymbol{\theta}}; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right. \\ &\quad \left. + \frac{1}{2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\tilde{\boldsymbol{\theta}}; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right) \\ &= -(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\check{\boldsymbol{\theta}}; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad \text{mit } |\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| \end{aligned}$$

$$= n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \left(-\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X})^T \right) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \underbrace{o_p(1)}_{\xrightarrow{n \rightarrow \infty} 0}. \quad (2.7)$$

In der letzten Gleichung wurde benutzt, dass $\hat{\boldsymbol{\theta}}$ und infolgedessen auch $\check{\boldsymbol{\theta}}$ konsistent sind. Somit lässt sich $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\check{\boldsymbol{\theta}}; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ in $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X})^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ und einen $o_p(1)$ -Term zerlegen. Mit dem schwachen Gesetz der großen Zahlen folgt

$$-\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X})^T \xrightarrow{n \rightarrow \infty} -\mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\boldsymbol{\theta}_0; X_1)^T \right] \stackrel{(2.2)}{=} \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0; X_1).$$

Es gilt also $\Lambda \stackrel{a}{=} n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0; X_1) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ und mit der Cholesky-Zerlegung ist eine Darstellung der Form $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0; X_1) = \mathbf{A}(\boldsymbol{\theta}_0; X_1)^T \mathbf{A}(\boldsymbol{\theta}_0; X_1)$ möglich, sodass

$$\begin{aligned} \Lambda &\stackrel{a}{=} n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \mathbf{A}(\boldsymbol{\theta}_0; X_1)^T \mathbf{A}(\boldsymbol{\theta}_0; X_1) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= (\sqrt{n} \mathbf{A}(\boldsymbol{\theta}_0; X_1) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0))^T (\sqrt{n} \mathbf{A}(\boldsymbol{\theta}_0; X_1) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)). \end{aligned}$$

Mit Satz 2.2.13 folgt $\sqrt{n} \mathbf{A}(\boldsymbol{\theta}_0; X_1) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$ und mit Lemma 2.2.8 schließlich die Behauptung. \square

Likelihood-Quotienten-Tests mit nuisance-Parametern

Die gewonnenen Resultate werden nun dazu genutzt, einen Test zu entwickeln, der die Situation einfacher Nullhypothesen auf komplexere Nullhypothesen erweitert. Genauer bedeutet dies, dass ein Signifikanztest hergeleitet wird, der im Falle eines mehrdimensionalen Parametervektors $\boldsymbol{\theta}$ die Nullhypothese $H_0 : \boldsymbol{\theta} \in \Theta_0 \subseteq \Theta$ gegen die Alternativhypothese $H_1 : \boldsymbol{\theta} \in \Theta \setminus \Theta_0$ testet. Darauf aufbauend lassen sich asymptotisch äquivalente Tests herleiten, die häufig für die Praxis vorteilhafte Eigenschaften, wie zum Beispiel leichtere Berechenbarkeit, aufweisen.

Anders als im Fall einfacher Nullhypothesen wird es im Folgenden nicht das Ziel sein, Annahmen über den gesamten Parametervektor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$ zu verifizieren, sondern es werden nur Annahmen über $r < d$ Vektorkomponenten geprüft. Dabei ist es von Vorteil, den Parametervektor in einen zu testenden und einen nicht zu testenden Teil aufzuspalten. Es sei $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T \in \mathbb{R}^d$ mit $\boldsymbol{\theta}_1^T = (\theta_{1_1}, \dots, \theta_{1_r})$ und $\boldsymbol{\theta}_2^T = (\theta_{2_1}, \dots, \theta_{2_{d-r}})$, sodass die Nullhypothese $H_0 : (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T = (\boldsymbol{\theta}_{1_0}^T, \boldsymbol{\theta}_2^T)^T$ gegen $H_1 : (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T \neq (\boldsymbol{\theta}_{1_0}^T, \boldsymbol{\theta}_2^T)^T$ lautet und wieder als Testproblem der Situation einfacher Nullhypothesen betrachtet werden kann. Es besteht jedoch der Unterschied, dass $\boldsymbol{\theta}_2$ nicht getestet wird, sondern nur als unbekannter Parameter bei der Analyse des Problems mitgeführt wird. Man bezeichnet $\boldsymbol{\theta}_2$ deswegen auch als *nuisance-* oder *Störparameter*. Da $\boldsymbol{\theta}_2$ nicht bekannt ist, bedient man sich bei der Untersuchung von H_0 eines Schätzers für $\boldsymbol{\theta}_2$, wobei man zwischen unrestringiertem und restringiertem Schätzer unterscheidet. Der restringierte Schätzer $\hat{\boldsymbol{\theta}}_{20}$ schätzt $\boldsymbol{\theta}_2$ unter

Annahme der Nullhypothese, also $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$, wohingegen der unrestringierte Schätzer $\hat{\boldsymbol{\theta}}_2$ den Störparameter ohne Annahmen an $\boldsymbol{\theta}_1$ schätzt. Die Schätzmethode ist dabei die der Maximum-Likelihood-Schätzer. Zusammenfassend gelten folgende Bezeichnungen für die Parametervektoren $\boldsymbol{\theta}_i$ mit $i \in \{1, 2\}$:

$$\begin{aligned}\boldsymbol{\theta}_i &\triangleq \text{wahrer Parametervektor,} \\ \boldsymbol{\theta}_{i0} &\triangleq \text{Parametervektor unter der Nullhypothese,} \\ \hat{\boldsymbol{\theta}}_i &\triangleq \text{unrestringierter MLS,} \\ \hat{\boldsymbol{\theta}}_{i0} &\triangleq \text{restringierter MLS.}\end{aligned}$$

Den Fall einfacher Hypothesen erweiternd wird der Likelihood-Quotient nun definiert durch

$$\frac{\sup_{\boldsymbol{\theta}' \in \Theta_1} L(\boldsymbol{\theta}'; \mathbf{x})}{\sup_{\boldsymbol{\theta}' \in \Theta_0} L(\boldsymbol{\theta}'; \mathbf{x})} = \exp\left(\frac{1}{2}\lambda\right),$$

sodass man im Fall $\Theta_1 = \Theta \setminus \Theta_0$ mit $\Lambda = \max(0, \lambda)$ einen neuen Quotienten der Form

$$\frac{\sup_{\boldsymbol{\theta}' \in \Theta} L(\boldsymbol{\theta}'; \mathbf{x})}{\sup_{\boldsymbol{\theta}' \in \Theta_0} L(\boldsymbol{\theta}'; \mathbf{x})} = \exp\left(\frac{1}{2}\Lambda\right)$$

und folglich

$$\Lambda = 2(l(\hat{\boldsymbol{\theta}}; \mathbf{x}) - l(\hat{\boldsymbol{\theta}}_0; \mathbf{x}))$$

erhält.

Satz 2.4.2. *Es seien x_1, \dots, x_n Realisierungen von i.i.d. Zufallsvariablen X_1, \dots, X_n , die gemäß einer regulären Verteilung $P_{(\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T}$ gezogen werden. Dann gilt für die Verteilung von Λ bei einer Nullhypothese der Form $H_0 : (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T = (\boldsymbol{\theta}_{10}^T, \boldsymbol{\theta}_2^T)^T$ gegen $H_1 : (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T \neq (\boldsymbol{\theta}_{10}^T, \boldsymbol{\theta}_2^T)^T$:*

$$\Lambda \xrightarrow{D} \chi_{\dim(\boldsymbol{\theta}_1)}^2.$$

Beweis. Unter Annahme der Nullhypothese gelten folgende Abkürzungen: $\boldsymbol{\theta}_0 = \begin{pmatrix} \boldsymbol{\theta}_{10} \\ \boldsymbol{\theta}_2 \end{pmatrix}$,

$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\theta}}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{pmatrix}$ und $\hat{\boldsymbol{\theta}}_0 = \begin{pmatrix} \boldsymbol{\theta}_{10} \\ \hat{\boldsymbol{\theta}}_{20} \end{pmatrix}$. Es ist

$$\begin{aligned}\Lambda &= 2(l(\hat{\boldsymbol{\theta}}; \mathbf{X}) - l(\hat{\boldsymbol{\theta}}_0; \mathbf{X})) \\ &= 2(l(\hat{\boldsymbol{\theta}}; \mathbf{X}) - l(\boldsymbol{\theta}_0; \mathbf{X})) - 2(l(\hat{\boldsymbol{\theta}}_0; \mathbf{X}) - l(\boldsymbol{\theta}_0; \mathbf{X})).\end{aligned}$$

Mit Gleichung (2.7) folgt

$$2(l(\hat{\boldsymbol{\theta}}; \mathbf{X}) - l(\boldsymbol{\theta}_0; \mathbf{X})) = n \begin{pmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10} \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{pmatrix}^T \left(-\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{S}(\boldsymbol{\theta}_0; \mathbf{X})^T \right) \begin{pmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10} \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{pmatrix} + o_p(1) \quad (2.8)$$

und

$$2(l(\hat{\boldsymbol{\theta}}_0; \mathbf{X}) - l(\boldsymbol{\theta}_0; \mathbf{X})) = n \begin{pmatrix} \mathbf{0} \\ \hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2 \end{pmatrix}^T \left(-\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X})^T \right) \begin{pmatrix} \mathbf{0} \\ \hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2 \end{pmatrix} + o_p(1). \quad (2.9)$$

Das schwache Gesetz der großen Zahlen liefert

$$-\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X})^T \xrightarrow{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{S}(\boldsymbol{\theta}_0; \mathbf{X})^T \right],$$

sodass mit Gleichung (2.2)

$$\Lambda \stackrel{a}{=} \begin{pmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10} \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{pmatrix}^T \mathcal{I}(\boldsymbol{\theta}_0; \mathbf{X}) \begin{pmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10} \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2 \end{pmatrix}^T \mathcal{I}(\boldsymbol{\theta}_0; \mathbf{X}) \begin{pmatrix} \mathbf{0} \\ \hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2 \end{pmatrix}$$

gilt. Eine Partitionierung der Fisher-Information in $\mathcal{I}(\boldsymbol{\theta}_0; \mathbf{X}) = \begin{pmatrix} \mathcal{I}_{11}(\boldsymbol{\theta}_0; \mathbf{X}) & \mathcal{I}_{12}(\boldsymbol{\theta}_0; \mathbf{X}) \\ \mathcal{I}_{21}(\boldsymbol{\theta}_0; \mathbf{X}) & \mathcal{I}_{22}(\boldsymbol{\theta}_0; \mathbf{X}) \end{pmatrix}$

mit $\mathcal{I}_{11}(\boldsymbol{\theta}_0; \mathbf{X}) \in \mathbb{R}^{r \times r}$, $\mathcal{I}_{12}(\boldsymbol{\theta}_0; \mathbf{X}) = \mathcal{I}_{21}(\boldsymbol{\theta}_0; \mathbf{X})^T \in \mathbb{R}^{r \times (d-r)}$, $\mathcal{I}_{22}(\boldsymbol{\theta}_0; \mathbf{X}) \in \mathbb{R}^{(d-r) \times (d-r)}$ bringt mit der abkürzenden Schreibweise $\mathcal{I}_{ij} = \mathcal{I}_{ij}(\boldsymbol{\theta}_0; \mathbf{X})$ für $i, j = 1, 2$ die Darstellung

$$\begin{aligned} \Lambda &\stackrel{a}{=} \begin{pmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10} \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{pmatrix}^T \begin{pmatrix} \mathcal{I}_{11}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) + \mathcal{I}_{12}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) \\ \mathcal{I}_{21}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) + \mathcal{I}_{22}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2 \end{pmatrix}^T \begin{pmatrix} \mathcal{I}_{12}(\hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2) \\ \mathcal{I}_{22}(\hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2) \end{pmatrix} \\ &= (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})^T \mathcal{I}_{11}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) + \underbrace{(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})^T \mathcal{I}_{12}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)}_{=:[1]} + \underbrace{(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)^T \mathcal{I}_{21}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})}_{=:[2]} \\ &\quad + \underbrace{(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)^T \mathcal{I}_{22}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)}_{=:[3]} - \underbrace{(\hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2)^T \mathcal{I}_{22}(\hat{\boldsymbol{\theta}}_{20} - \boldsymbol{\theta}_2)}_{=:[4]}. \end{aligned}$$

Um den Term auf der rechten Seite der Gleichung in eine geschlosseneren Form zu bringen, wird der restringierte MLS $\hat{\boldsymbol{\theta}}_{20}$ in Abhängigkeit der beiden unrestringierten MLS $\hat{\boldsymbol{\theta}}_1$ und $\hat{\boldsymbol{\theta}}_2$, sowie dem unter der Nullhypothese wahren Parameter $\boldsymbol{\theta}_{10}$ dargestellt und damit [4] berechnet. Dazu sei

$$\mathcal{I}(\boldsymbol{\theta}_0; \mathbf{X})^{-1} = \begin{pmatrix} \mathcal{I}^{11}(\boldsymbol{\theta}_0; \mathbf{X}) & \mathcal{I}^{12}(\boldsymbol{\theta}_0; \mathbf{X}) \\ \mathcal{I}^{21}(\boldsymbol{\theta}_0; \mathbf{X}) & \mathcal{I}^{22}(\boldsymbol{\theta}_0; \mathbf{X}) \end{pmatrix}$$

eine Partitionierung der Inversen der Fisher-Information an der Stelle $\boldsymbol{\theta}_0$. Abkürzend bezeichne auch hier $\mathcal{I}^{ij} = \mathcal{I}^{ij}(\boldsymbol{\theta}_0; \mathbf{X})$ für $i, j = 1, 2$. Die Beweisidee folgt [Paw01].

Nach Satz 2.2.13 gilt $\begin{pmatrix} \hat{\boldsymbol{\theta}}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{pmatrix} \xrightarrow{D} \mathcal{N}_d \left(\begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix}, \mathcal{I}(\boldsymbol{\theta}; \mathbf{X})^{-1} \right)$, sodass bei Unkenntnis von $\boldsymbol{\theta}_1$ und $\boldsymbol{\theta}_2$ der MLS für $\mathbb{E}[\hat{\boldsymbol{\theta}}_2] = \hat{\boldsymbol{\theta}}_2$ ist. Ist $\boldsymbol{\theta}_1$ allerdings nicht unbekannt – im vorliegenden Fall ist unter Annahme der Nullhypothese $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ bekannt – so hat dies Auswirkungen

auf $\mathbb{E}[\hat{\boldsymbol{\theta}}_2]$, sofern die Verteilung von $\hat{\boldsymbol{\theta}}_2$ von $\boldsymbol{\theta}_{10}$ abhängt. Nun gilt mit Lemma 2.2.10 für die bedingte multivariate Normalverteilung von $\hat{\boldsymbol{\theta}}_2$ gegeben $\hat{\boldsymbol{\theta}}_1$

$$\hat{\boldsymbol{\theta}}_2 | \hat{\boldsymbol{\theta}}_1 \stackrel{a}{\sim} \mathcal{N}_{d-r} \left(\boldsymbol{\theta}_2 + \boldsymbol{I}^{21} (\boldsymbol{I}^{11})^{-1} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}), \boldsymbol{I}^{22} - \boldsymbol{I}^{21} (\boldsymbol{I}^{11})^{-1} \boldsymbol{I}^{12} \right),$$

also speziell $\mathbb{E}[\hat{\boldsymbol{\theta}}_2 | \hat{\boldsymbol{\theta}}_1] = \boldsymbol{\theta}_2 + \boldsymbol{I}^{21} (\boldsymbol{I}^{11})^{-1} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})$. Es folgt wegen $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$, dass $\hat{\boldsymbol{\theta}}_2$ im Mittel der bedingten Erwartung entspricht. Dies bedeutet

$$\begin{aligned} \hat{\boldsymbol{\theta}}_2 &= \mathbb{E}[\hat{\boldsymbol{\theta}}_2 | \hat{\boldsymbol{\theta}}_1] = \boldsymbol{\theta}_2 + \boldsymbol{I}^{21} (\boldsymbol{I}^{11})^{-1} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) \\ \Leftrightarrow \boldsymbol{\theta}_2 &= \hat{\boldsymbol{\theta}}_2 - \boldsymbol{I}^{21} (\boldsymbol{I}^{11})^{-1} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) \end{aligned}$$

was unter Maximum-Likelihood-Schätzung zu $\hat{\boldsymbol{\theta}}_{20} = \hat{\boldsymbol{\theta}}_2 - \boldsymbol{I}^{21} (\boldsymbol{I}^{11})^{-1} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})$ und unter Berücksichtigung von⁷

$$\boldsymbol{I}^{-1} = \begin{pmatrix} \boldsymbol{J}^{-1} & -\boldsymbol{J}^{-1} \boldsymbol{I}_{12} (\boldsymbol{I}_{22})^{-1} \\ -(\boldsymbol{I}_{22})^{-1} \boldsymbol{I}_{21} \boldsymbol{J}^{-1} & \boldsymbol{I}_{22} + (\boldsymbol{I}_{22})^{-1} \boldsymbol{I}_{21} \boldsymbol{J}^{-1} \boldsymbol{I}_{12} (\boldsymbol{I}_{22})^{-1} \end{pmatrix} \quad (2.10)$$

mit $\boldsymbol{J} = \boldsymbol{I}_{11} - \boldsymbol{I}_{12} (\boldsymbol{I}_{22})^{-1} \boldsymbol{I}_{21}$ zu

$$\hat{\boldsymbol{\theta}}_{20} = \hat{\boldsymbol{\theta}}_2 + (\boldsymbol{I}_{22})^{-1} \boldsymbol{I}_{21} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) \quad (2.11)$$

führt. Subtrahiert man $\boldsymbol{\theta}_2$ auf beiden Seiten der Gleichung (2.11), so folgt

$$\begin{aligned} [4] &= \left((\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)^T + (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})^T (\boldsymbol{I}_{21})^T ((\boldsymbol{I}_{22})^{-1})^T \right) \boldsymbol{I}_{22} \left((\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + (\boldsymbol{I}_{22})^{-1} \boldsymbol{I}_{21} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) \right) \\ &= \underbrace{(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)^T \boldsymbol{I}_{22} (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)}_{=[3]} + \underbrace{(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)^T \boldsymbol{I}_{21} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})}_{=[2]} \\ &\quad + \underbrace{(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})^T \underbrace{(\boldsymbol{I}_{21})^T ((\boldsymbol{I}_{22})^{-1})^T \boldsymbol{I}_{22}}_{=\boldsymbol{I}} (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)}_{=\boldsymbol{I}_{12}}}_{=[1]} \\ &\quad + (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})^T (\boldsymbol{I}_{21})^T ((\boldsymbol{I}_{22})^{-1})^T \boldsymbol{I}_{21} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}). \end{aligned}$$

Es ergibt sich $\Lambda \stackrel{a}{=} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})^T (\boldsymbol{I}_{11} - (\boldsymbol{I}_{21})^T ((\boldsymbol{I}_{22})^{-1})^T \boldsymbol{I}_{21}) (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})$ und mit (2.10) schließlich

$$\Lambda \stackrel{a}{=} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})^T (\boldsymbol{I}^{11})^{-1} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}). \quad (2.12)$$

Mit derselben Schlußfolgerung wie im Beweis zu Satz 2.4.1 ist der Beweis vollständig. \square

Bisher wurde ausschließlich auf Likelihood-Quotienten-Tests eingegangen und deren asymptotische Verteilung abgeleitet. Da das eigentliche Ziel jedoch darin besteht den Score-Test herzuleiten, werden nun die gewonnenen Erkenntnisse genau für diesen Zweck genutzt. Da

⁷Vgl. dazu Anhang A.3.

für unabhängige und identisch verteilte Zufallsvariablen X_1, \dots, X_n wegen Gleichung (2.3) $\mathcal{I}(\boldsymbol{\theta}_0; \mathbf{X}) = n\mathcal{I}(\boldsymbol{\theta}_0; X_1)$ gilt, folgt mit Gleichung (2.4)

$$\mathcal{I}(\boldsymbol{\theta}_0; \mathbf{X})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \stackrel{a}{=} \mathcal{S}(\boldsymbol{\theta}_0; \mathbf{X}).$$

Kürzt man wieder ab und schreibt $\mathcal{I}_{ij} = \mathcal{I}_{ij}(\boldsymbol{\theta}_0; \mathbf{X})$ für $i, j = 1, 2$, dann gilt also

$$\left(\begin{array}{l} \mathcal{I}_{11}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) + \mathcal{I}_{12}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20}) \stackrel{a}{=} \mathcal{S}_1(\boldsymbol{\theta}_0; \mathbf{X}) \\ \mathcal{I}_{21}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) + \mathcal{I}_{22}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20}) \stackrel{a}{=} \mathcal{S}_2(\boldsymbol{\theta}_0; \mathbf{X}) \end{array} \right), \text{ sodass}$$

$$(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) \stackrel{a}{=} (\mathcal{I}_{11} - \mathcal{I}_{12}(\mathcal{I}_{22})^{-1}\mathcal{I}_{21})^{-1}(\mathcal{S}_1(\boldsymbol{\theta}_0; \mathbf{X}) - \mathcal{I}_{12}(\mathcal{I}_{22})^{-1}\mathcal{S}_2(\boldsymbol{\theta}_0; \mathbf{X}))$$

folgt. Einsetzen in (2.12) ergibt als asymptotisch äquivalente Statistik zu Λ mit der abkürzenden Schreibweise $\mathcal{S}_i = \mathcal{S}_i(\boldsymbol{\theta}_0; \mathbf{X})$ für $i = 1, 2$:

$$\Lambda_{\mathcal{S}} \stackrel{a}{=} (\mathcal{S}_1 - \mathcal{I}_{12}(\mathcal{I}_{22})^{-1}\mathcal{S}_2)^T (\mathcal{I}_{11} - \mathcal{I}_{12}(\mathcal{I}_{22})^{-1}\mathcal{I}_{21})^{-1} (\mathcal{S}_1 - \mathcal{I}_{12}(\mathcal{I}_{22})^{-1}\mathcal{S}_2).$$

Setzt man $\boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_{20}$, so addiert sich aufgrund der Konsistenz des MLS ein $o_p(1)$ -Term, der für $n \rightarrow \infty$ verschwindet und es ist $\mathcal{S}_2 = 0$. Aufgrund ihrer Wichtigkeit für die folgende Theorie wird die gewonnene Teststatistik in einer Definition festgehalten.

Definition 2.4.3. Es gelten die Annahmen aus Satz 2.4.2. Dann heißt die durch

$$\Lambda_{\mathcal{S}}(\hat{\boldsymbol{\theta}}_{20}) = \mathcal{S}_1(\hat{\boldsymbol{\theta}}_0; \mathbf{X})^T \mathcal{I}_{11}(\hat{\boldsymbol{\theta}}_0; \mathbf{X})^{-1} \mathcal{S}_1(\hat{\boldsymbol{\theta}}_0; \mathbf{X}) \quad (2.13)$$

definierte Statistik *Score-Test*.

Bemerkung 2.4.4. Da $\Lambda_{\mathcal{S}}$ asymptotisch äquivalent zu Λ ist, besitzt auch $\Lambda_{\mathcal{S}}$ eine asymptotische Chi-Quadrat-Verteilung mit Anzahl an Freiheitsgraden identisch der Dimension von $\boldsymbol{\theta}_1$. Welcher der vorliegenden Tests also letztlich angewendet wird, hängt in der Regel von der Berechenbarkeit der Schätzer ab. Für den Likelihood-Quotienten-Test muss der unrestringierte MLS berechnet werden, für den Score-Test hingegen ist der MLS unter der Nullhypothese ausreichend. Dies stellt in vielen Situationen einen klaren Vorteil dar.

Kapitel 3

Testen auf univariate Normalverteilung

Nachdem nun die Grundlage für die Herleitung des eigentlichen Ziels dieser Arbeit geschaffen wurde, werden in diesem Kapitel die Verteilungen von i.i.d. Beobachtungen sowie von unbeobachtbaren i.i.d. Regressionsresiduen analysiert. Das Ziel wird es sein, einen Test zu entwickeln, mit dessen Hilfe die Hypothese der Normalverteilung der zu testenden Größen bestätigt oder verworfen werden kann. Dazu wird der Score-Test genutzt. Es resultiert ein Test, dessen Vorzüge nicht nur in seiner leichten Anwendbarkeit, sondern auch in der asymptotischen Effizienz liegen. Im ersten Abschnitt wird dazu das Pearson-Verteilungssystem eingeführt und die für die folgende Theorie notwendige Verbindung mit der Normalverteilung herausgearbeitet. Die Abschnitte zwei und drei wenden die gewonnenen Resultate auf Beobachtungen beziehungsweise Regressionsresiduen an, um die speziell auf Normalverteilung testende Jarque-Bera-Teststatistik herzuleiten. Die Vorgehensweise orientiert sich an [BJ81]. Es wird das Pearson-Verteilungssystem betrachtet und anschließend der Score-Test darauf angewendet. Da die Normalverteilung ein spezielles Mitglied dieses Systems darstellt, kann so die gewünschte Hypothese innerhalb dieses Verteilungssystems getestet werden.

3.1 Das Pearson-Verteilungssystem

Ziel des vorliegenden Abschnitts ist die Darstellung der Dichte einer Normalverteilung als homogene lineare Differentialgleichung (DGL) und die daraus resultierende Verdeutlichung des Zusammenhangs mit dem Pearson-Verteilungssystem. Dazu sei

$$\frac{\partial}{\partial x} p(x) = a(x)p(x) \tag{3.1}$$

eine DGL mit stetigen Funktionen $a(x)$ und $p(x)$, die hier vorerst nicht weiter spezifiziert werden sollen. Dann bildet

$$p(x) = c \exp(A(x))$$

mit $A(x)$ als Stammfunktion von $a(x)$ und $c \in \mathbb{R}$ konstant eine Lösung der gegebenen DGL. Es sei nun X eine Zufallsvariable mit $X \sim \mathcal{N}(0, \sigma^2)$ und Dichte $f(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/(2\sigma^2))$. Dann ist eine Darstellung der Dichte $f(x)$ in Form obiger DGL mit Lösung $f(x) = c \exp(A(x))$ mit $c = (2\pi\sigma^2)^{-1/2}$ und $A(x) = -x^2/(2\sigma^2)$ durch $\frac{\partial}{\partial x} f(x) = A'(x)f(x) = (-x/\sigma^2)f(x)$ gegeben. Der Ausdruck $a(x)$ in obiger DGL (3.1) wird konkretisiert, indem ein System von Gleichungen für $a(x)$ angegeben wird, welches durch Variation der darin enthaltenen Variablen spezifiziert werden kann. Dieses auf Karl Pearson zurückgehende System aus dem Jahre 1895 beruht auf der DGL

$$\frac{\partial}{\partial x} p(x) = -\frac{c_1 + x}{c_0 + c_1 x + c_2 x^2} p(x) \quad (3.2)$$

für $x \in \mathbb{R}$ und wird als *Pearson-Verteilungssystem* bezeichnet. Es enthält alle Funktionen $p(x)$, die eine Lösung dieser DGL darstellen, wobei c_0 , c_1 und c_2 formgebende Parameter sind, die maßgeblich für die Gestalt der Funktionen verantwortlich sind. Für eine ausführlichere Auseinandersetzung mit dieser DGL sei auf [KS69], Kapitel 6, sowie [JK94] verwiesen. Damit Wahrscheinlichkeitsverteilungen als Lösungen von (3.2) resultieren, ist auf Normiertheit und Positivität, d.h. $\int_{-\infty}^{\infty} p(x) dx = 1$ und $p(x) \geq 0$ für alle $x \in \mathbb{R}$ zu achten. In der Situation $c_0 = \sigma^2$ und $c_1 = c_2 = 0$ erhält man den Spezialfall der Normalverteilung mit Erwartungswert 0 und Varianz σ^2 , wenn die Stammfunktion von $a(x)$ wie im obigen Fall als $A(x) = \int_0^x a(t) dt = -x^2/(2\sigma^2)$ gewählt ist und der konstante Faktor $c = (2\pi\sigma^2)^{-1/2}$ entspricht. Dieser Spezialfall wird im Folgenden dazu genutzt, beobachtbare Ereignisse innerhalb dieses Systems auf Normalverteilung zu testen.

Dass das Testen auf Normalverteilung nur innerhalb dieses Systems, also nur gegen Verteilungen aus diesem System vorgenommen wird, stellt insofern eine Einschränkung dar, als dass gegen gewisse Verteilungen, wie beispielsweise die *Lognormal*-Verteilung nicht getestet werden kann. Empirische Studien belegen allerdings, dass trotz dieser misslichen Situation die Jarque-Bera-Teststatistik im Vergleich mit anderen Tests auf Normalverteilung, die auch gegen Nicht-Pearson-Verteilungen testen, eine höhere Güte aufweist.¹ Desweiteren ist festzuhalten, dass ein breites Spektrum an Verteilungen (unter anderem *Beta*-, *Gamma*-, *t*- und *F*-Verteilung) durch das Pearson-System abgedeckt wird, was die Attraktivität des Jarque-Bera-Tests nicht nur anhand seiner einfachen Form begründet.

¹Vgl. [JB87].

3.2 Testen von Beobachtungen

Bevor damit begonnen wird, Beobachtungen auf Normalverteilung zu testen, wird eine Lösung der DGL (3.2) speziell für Wahrscheinlichkeitsdichten $g(y) = g(y; c_1, c_2, c_0)$ mit $y \in \mathbb{R}$ hergeleitet. Da die Gestalt von $a(y) = a(y; c_1, c_2, c_0)$ im Wesentlichen durch die Parameter c_0 , c_1 und c_2 bestimmt wird, wird für die folgende Argumentation eine neue Notation eingeführt, die sich im späteren Zusammenhang als sehr nützlich erweisen wird. Es bezeichne $\int a(y; c_1, c_2, c_0) dy$ die Stammfunktion von $a(y; c_1, c_2, c_0)$ mit konstantem Summanden null. Mit dieser Schreibweise können unabhängig von den gewählten Parametern c_0 , c_1 und c_2 diese speziellen Stammfunktionen aus der Menge aller Stammfunktionen $\int_{y_0}^y a(t; c_1, c_2, c_0) dt$ von $a(y; c_1, c_2, c_0)$ gewählt werden, ohne dass die entsprechenden Werte für y_0 angegeben werden müssen. Dann gilt für Wahrscheinlichkeitsdichten aus dem Pearson-Verteilungssystem die folgende Darstellung:

$$\begin{aligned}
 & \frac{\partial}{\partial y} g(y; c_1, c_2, c_0) = a(y; c_1, c_2, c_0) g(y; c_1, c_2, c_0) \\
 \Leftrightarrow & \log(g(y; c_1, c_2, c_0)) = \int a(y; c_1, c_2, c_0) dy \\
 \Leftrightarrow & g(y; c_1, c_2, c_0) = \exp\left(\int a(y; c_1, c_2, c_0) dy\right) \\
 \stackrel{g \text{ Dichte}}{\Leftrightarrow} & g(y; c_1, c_2, c_0) = \frac{\exp\left(\int a(y; c_1, c_2, c_0) dy\right)}{\int_{-\infty}^{\infty} \exp\left(\int a(y; c_1, c_2, c_0) dy\right) dy}. \quad (3.3)
 \end{aligned}$$

Es seien nun x_1, \dots, x_n Beobachtungen, die man sich als Realisierungen von i.i.d. Zufallsvariablen vorstelle. Um zu prüfen, ob es sich um normalverteilte Beobachtungen mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 handelt, werden die Beobachtungen durch $y_i = x_i - \mu$ für $i = 1, \dots, n$ zentriert und geprüft, ob für die zugehörigen Zufallsvariablen $\mathbb{E}[Y_i] = 0$ und $\text{Var}[Y_i] = \sigma^2$ gilt. Es bezeichne $g(y)$ die Dichte der Zufallsvariable Y_i für $i = 1, \dots, n$ und außerdem sei $g(y)$ aus dem Pearson-Verteilungssystem. Betrachtet man nun die zentrierten Beobachtungen y_1, \dots, y_n , so ist die Log-Likelihood-Funktion der Stichprobe $\mathbf{y} = (y_1, \dots, y_n)$ unter Berücksichtigung von (3.3) gegeben durch

$$l(c_1, c_2, c_0; \mathbf{y}) = -n \log\left(\int_{-\infty}^{\infty} v_1(y; c_1, c_2, c_0) dy\right) + \sum_{i=1}^n \int a(y_i; c_1, c_2, c_0) dy_i \quad (3.4)$$

mit $v_1(y; c_1, c_2, c_0) = \exp\left(\int a(y; c_1, c_2, c_0) dy\right)$.

Um nun die Nullhypothese der Normalverteilung der Y_i zu testen sei $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \theta_2)^T$ mit $\boldsymbol{\theta}_1^T = (c_1, c_2)$ und $\theta_2 = c_0$, sodass nach Abschnitt 3.1 die Nullhypothese $H_0 : \boldsymbol{\theta}^T = (0, 0, \sigma^2)$ lautet. Auf diese wird der Score-Test $\Lambda_{\mathcal{S}}$ aus (2.13) angewendet, für den der Score $\mathcal{S}_1(c_1, c_2, c_0; \mathbf{Y})$ und die Teilmatrix $\mathcal{I}_{11}(c_1, c_2, c_0; \mathbf{Y})$ der Fisher-Informationsmatrix

$\mathcal{I}(c_1, c_2, c_0; \mathbf{Y})$ an der Stelle des restringierten MLS $\hat{\boldsymbol{\theta}}_0$ benötigt wird. Als erstes werden die Ableitungen der Log-Likelihood-Funktion nach den einzelnen Komponenten des Parametervektors $\boldsymbol{\theta}$ bestimmt und dann die Parameter entsprechend der gewünschten Hypothese festgesetzt. Da in den folgenden Darstellungen innerhalb der Ableitungen der Log-Likelihood-Funktion nach den Parameterkomponenten $v_1(y; c_1, c_2, c_0)$ nicht mehr von den Differentiationsvariablen abhängt, wird abkürzend v_1 geschrieben. Zusätzlich sei $v_2(c_j) = \frac{\partial}{\partial c_j} a(y; c_1, c_2, c_0)$, wobei hier eine Abhängigkeit von den Differentiationsvariablen innerhalb der Ableitungen der Log-Likelihood-Funktion besteht. Mit dieser Notation gilt für $j \in \{0, 1, 2\}$:

$$\frac{\partial}{\partial c_j} l(\boldsymbol{\theta}; \mathbf{y}) = -nV_1^{-1} \int_{-\infty}^{\infty} v_1 V_2(c_j) dy + Z(c_j) \quad (3.5)$$

mit $V_1 = \int_{-\infty}^{\infty} v_1 dy$, $V_2(c_j) = \int v_2(c_j) dy$ und $Z(c_j) = \sum_{i=1}^n \int v_2(c_j) dy_i$. Eine weitere Vereinfachung der Form

$$\frac{\partial}{\partial c_j} l(\boldsymbol{\theta}; \mathbf{y}) = -nUV(c_j) + Z(c_j) \quad (3.6)$$

ergibt sich mit $U = V_1^{-1}$ und $V(c_j) = \int_{-\infty}^{\infty} v_1 V_2(c_j) dy$.

Lemma 3.2.1. *Unter der Nullhypothese ist $\boldsymbol{\theta}^T = (0, 0, \sigma^2)$ und somit $v_1 = \exp(-y^2/(2\sigma^2))$, $V_1 = (2\pi\sigma^2)^{1/2}$, $U = (2\pi\sigma^2)^{-1/2}$, $v_2(c_1) = -(\sigma^2 + y^2)/\sigma^4$, $v_2(c_2) = y^3/\sigma^4$, $v_2(c_0) = y/\sigma^4$, $V_2(c_1) = y/\sigma^2 - y^3/(3\sigma^4)$, $V_2(c_2) = y^4/(4\sigma^4)$, $V_2(c_0) = y^2/(2\sigma^4)$, $V(c_1) = \int_{-\infty}^{\infty} v_1 (y/\sigma^2 - y^3/(3\sigma^4)) dy$, $V(c_2) = \int_{-\infty}^{\infty} v_1 y^4/(4\sigma^4) dy$, $V(c_0) = \int_{-\infty}^{\infty} v_1 y^2/(2\sigma^4) dy$, $Z(c_1) = \sum_{i=1}^n y_i/\sigma^2 - y_i^3/(3\sigma^4)$, $Z(c_2) = \sum_{i=1}^n y_i^4/(4\sigma^4)$ und $Z(c_0) = \sum_{i=1}^n y_i^2/(2\sigma^4)$.*

Weiter sei $\mu_j = n^{-1} \sum_{i=1}^n y_i^j = n^{-1} \sum_{i=1}^n (x_i - \mu)^j$ das j -te empirische Moment von Y_i . Da der Erwartungswert μ der X_i unbekannt ist, wird als Schätzer für μ der Stichprobenmittelwert $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ verwendet, sodass $\hat{\mu}_j = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^j$ das j -te empirische Moment von Y_i unter Schätzung von μ bezeichnet.

Proposition 3.2.2. *Es sei $\mathcal{S}_1(\boldsymbol{\theta}; \mathbf{y})^T = \left(\frac{\partial}{\partial c_1} l(\boldsymbol{\theta}; \mathbf{y}), \frac{\partial}{\partial c_2} l(\boldsymbol{\theta}; \mathbf{y}) \right)$ der Score im vorliegenden Testproblem. Dann gilt:*

$$\mathcal{S}_1(\hat{\boldsymbol{\theta}}_0; \mathbf{y})^T = n \left(-\frac{\hat{\mu}_3}{3\hat{\mu}_2^2}, \frac{\hat{\mu}_4}{4\hat{\mu}_2^2} - \frac{3}{4} \right).$$

Beweis. Zuerst werden die partiellen Ableitungen von $l(\boldsymbol{\theta}; \mathbf{y})$ nach c_1 und c_2 an der Stelle $\boldsymbol{\theta}_0 = (0, 0, \sigma^2)^T$ berechnet. Man beachte dabei, dass unter der Nullhypothese Y_i normalverteilt ist und somit $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i] = 0$, $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^2] = \sigma^2$, $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^3] = 0$ und $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] = 3\sigma^4$ gilt. Es

folgt mit Gleichung (3.6) und Lemma 3.2.1

$$\begin{aligned} \frac{\partial}{\partial c_1} l(\boldsymbol{\theta}_0; \mathbf{y}) &= -n \left(\underbrace{\frac{1}{\sigma^2} \int_{-\infty}^{\infty} y_i U v_1 dy_i}_{=\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i]} - \frac{1}{3\sigma^4} \underbrace{\int_{-\infty}^{\infty} y_i^3 U v_1 dy_i}_{=\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^3]} - \frac{1}{\sigma^2} \underbrace{\frac{1}{n} \sum_{i=1}^n y_i}_{=\mu_1} + \frac{1}{3\sigma^4} \underbrace{\frac{1}{n} \sum_{i=1}^n y_i^3}_{=\mu_3} \right) \\ &= n \left(\frac{\mu_1}{\sigma^2} - \frac{\mu_3}{3\sigma^4} \right) \quad \text{und} \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial c_2} l(\boldsymbol{\theta}_0; \mathbf{y}) &= -n \left(\underbrace{\frac{1}{4\sigma^4} \int_{-\infty}^{\infty} y_i^4 U v_1 dy_i}_{=\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4]} - \frac{1}{4\sigma^4} \underbrace{\frac{1}{n} \sum_{i=1}^n y_i^4}_{=\mu_4} \right) \\ &= n \left(\frac{\mu_4}{4\sigma^4} - \frac{3}{4} \right). \end{aligned}$$

Setzt man $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_0$, so folgt wegen $\hat{\sigma}^2 = \hat{\mu}_2$ und $\hat{\mu}_1 = 0$ die Behauptung. \square

Um die Fisher-Informationsmatrix herzuleiten, werden die zweiten partiellen Ableitungen in den einzelnen Kombinationen der Parameterkomponenten bestimmt. Mit Hilfe von Gleichung (3.6) gilt für $j, k \in \{0, 1, 2\}$:

$$\frac{\partial^2}{\partial c_k \partial c_j} l(\boldsymbol{\theta}; \mathbf{y}) = -n \left(V(c_j) \frac{\partial}{\partial c_k} U + U \frac{\partial}{\partial c_k} V(c_j) \right) + \frac{\partial}{\partial c_k} Z(c_j) \quad (3.7)$$

mit $\frac{\partial}{\partial c_k} U = -U^2 V(c_k)$, $\frac{\partial}{\partial c_k} V(c_j) = \int_{-\infty}^{\infty} v_1 \left(V_2(c_k) V_2(c_j) + \frac{\partial}{\partial c_k} V_2(c_j) \right) dy$ und $\frac{\partial}{\partial c_k} Z(c_j) = \sum_{i=1}^n \frac{\partial}{\partial c_k} V_2(c_j)$.

Proposition 3.2.3. *Es sei $\mathcal{I}(\boldsymbol{\theta}; \mathbf{Y}) = -\mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\boldsymbol{\theta}; \mathbf{Y}) \right]$ mit $i, j = 1, 2$ die Fisher-Information im vorliegenden Testproblem. Dann gilt:*

$$\mathcal{I}(\hat{\boldsymbol{\theta}}_0; \mathbf{Y}) = \begin{pmatrix} \mathcal{I}_{11}(\hat{\boldsymbol{\theta}}_0; \mathbf{Y}) & \mathcal{I}_{12}(\hat{\boldsymbol{\theta}}_0; \mathbf{Y}) \\ \mathcal{I}_{21}(\hat{\boldsymbol{\theta}}_0; \mathbf{Y}) & \mathcal{I}_{22}(\hat{\boldsymbol{\theta}}_0; \mathbf{Y}) \end{pmatrix} = n \begin{pmatrix} \frac{2}{3\hat{\mu}_2} & 0 & 0 \\ 0 & 6 & \frac{3}{2\hat{\mu}_2} \\ 0 & \frac{3}{2\hat{\mu}_2} & \frac{1}{2\hat{\mu}_2^2} \end{pmatrix}.$$

Beweis. Es werden die zweiten Ableitungen von $l(\boldsymbol{\theta}; \mathbf{y})$ nach c_1 , c_2 und c_0 erneut zuerst an der Stelle $\boldsymbol{\theta}_0$ berechnet. Mit Gleichung (3.7) und Lemma 3.2.1 gilt:

$$\begin{aligned} \frac{\partial^2}{\partial c_1^2} l(\boldsymbol{\theta}_0; \mathbf{y}) &= -n \left(- \left(\frac{1}{\sigma^2} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i] - \frac{1}{3\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^3] \right)^2 + \frac{1}{\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^2] - \frac{2}{3\sigma^6} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] \right. \\ &\quad \left. + \frac{1}{9\sigma^8} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^6] + \frac{1}{\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^2] - \frac{1}{2\sigma^6} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] - \frac{1}{\sigma^4} \mu_2 + \frac{1}{2\sigma^6} \mu_4 \right) \\ &= -n \left(\frac{1}{6\sigma^2} - \frac{\mu_2}{\sigma^4} + \frac{\mu_4}{2\sigma^6} \right), \end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial c_2^2}l(\boldsymbol{\theta}_0; y) &= -n \left(- \left(\frac{1}{4\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] \right)^2 + \frac{1}{16\sigma^8} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^8] - \frac{1}{3\sigma^6} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^6] + \frac{1}{3\sigma^6} \mu_6 \right) \\ &= -n \left(1 + \frac{\mu_6}{3\sigma^6} \right),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial c_0^2}l(\boldsymbol{\theta}_0; y) &= -n \left(- \left(\frac{1}{2\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^2] \right)^2 + \frac{1}{4\sigma^8} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] - \frac{1}{\sigma^6} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^2] + \frac{1}{\sigma^6} \mu_2 \right) \\ &= -n \left(\frac{1}{2\sigma^4} \right),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial c_1 \partial c_2}l(\boldsymbol{\theta}_0; y) &= -n \left(- \frac{1}{4\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] \left(\frac{1}{\sigma^2} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i] - \frac{1}{3\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^3] \right) + \frac{1}{4\sigma^6} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^5] \right. \\ &\quad \left. - \frac{1}{12\sigma^8} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^7] - \frac{1}{3\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^3] + \frac{2}{15\sigma^6} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^5] + \frac{1}{3\sigma^4} \mu_3 - \frac{2}{15\sigma^6} \mu_5 \right) \\ &= -n \left(\frac{\mu_3}{3\sigma^4} - \frac{2\mu_5}{15\sigma^6} \right),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial c_1 \partial c_0}l(\boldsymbol{\theta}_0; y) &= -n \left(- \frac{1}{2\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^2] \left(\frac{1}{\sigma^2} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i] - \frac{1}{3\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^3] \right) - \frac{1}{6\sigma^8} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^5] \right. \\ &\quad \left. + \frac{7}{6\sigma^6} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^3] - \frac{1}{\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i] + \frac{1}{\sigma^4} \mu_1 - \frac{2}{3\sigma^6} \mu_3 \right) \\ &= -n \left(\frac{\mu_1}{\sigma^4} - \frac{2\mu_3}{3\sigma^6} \right) \quad \text{und}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial c_2 \partial c_0}l(\boldsymbol{\theta}_0; y) &= -n \left(- \frac{1}{2\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^2] \frac{1}{4\sigma^4} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] + \frac{1}{8\sigma^8} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^6] - \frac{1}{2\sigma^6} \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] + \frac{1}{2\sigma^6} \mu_4 \right) \\ &= -n \left(\frac{\mu_4}{2\sigma^6} \right).\end{aligned}$$

Mit $\mathcal{I}(\boldsymbol{\theta}; \mathbf{Y}) = -\mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\boldsymbol{\theta}; \mathbf{Y}) \right]$ und $\mathbb{E}_{\boldsymbol{\theta}_0}[\mu_j] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^j] = \mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^j]$ und $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^2] = \sigma^2$, $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^4] = 3\sigma^4$, $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^6] = 15\sigma^6$, $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^8] = 105\sigma^8$ und $\mathbb{E}_{\boldsymbol{\theta}_0}[Y_i^k] = 0$ falls k ungerade, sowie $\hat{\sigma}^2 = \hat{\mu}_2$ folgt die Behauptung. \square

Wendet man die Propositionen 3.2.2 und 3.2.3 auf die Score-Teststatistik in (2.13) an, führt dies zu

$$\Lambda_{\mathcal{S}} = n \left(\frac{1}{6} \frac{\hat{\mu}_3^2}{\hat{\mu}_2^3} + \frac{1}{24} \left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 \right)^2 \right).$$

Um die Bedeutung dieses Ausdrucks besser zu verstehen, werden zwei Momentenverhältnisse eingeführt, mit denen eine Abweichung von der Normalverteilung gemessen werden kann.

Definition 3.2.4. Es sei X eine Zufallsvariable mit $\mathbb{E}[X] = \mu$ und $\text{Var}[X] = \sigma^2$. Weiter sei $\mu_j = \mathbb{E}[(X - \mu)^j]$ das j -te zentrierte Moment von X . Dann bezeichnet

$$\beta_1 = \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\mu_2^{3/2}}$$

die *Schiefe* von X und

$$\beta_2 = \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\mu_2^2}$$

die *Wölbung* von X . Für eine Stichprobe \mathbf{X} bestehend aus i.i.d Zufallsvariablen X_1, \dots, X_n sei entsprechend $\hat{\mu}_j = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^j$ mit $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ das j -te empirische zentrierte Moment und $\hat{\mu}_2$ die Stichprobenvarianz von \mathbf{X} . Dann bildet

$$\sqrt{b_1} = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}$$

die *empirische Schiefe* und

$$b_2 = \frac{\hat{\mu}_4}{\hat{\mu}_2^2}$$

die *empirische Wölbung*.

Anschaulich misst die Schiefe einer Verteilung die Neigung nach links oder rechts, die Wölbung die Krümmung, bzw. die Steilheit einer Verteilung. Symmetrische Verteilungen besitzen demnach eine Schiefe von null, die Umkehrung dieser Aussage gilt jedoch nicht. Die Wölbung einer Normalverteilung beträgt drei. Die Ergebnisse im vorliegenden Testproblem werden nun zusammengefasst und in einer Definition wiedergegeben.

Definition 3.2.5. Es gelte die Situation des vorliegenden Testproblems. Dann definiert

$$JB = \frac{n}{6} \left((\sqrt{b_1})^2 + \frac{(b_2 - 3)^2}{4} \right) \quad (3.8)$$

die auf [JB87] zurückgehende Teststatistik. Man bezeichnet sie auch einfach als *Jarque-Bera-Test*.

Bemerkung 3.2.6. Der Jarque-Bera-Test auf Normalverteilung von Beobachtungen nutzt somit sowohl die empirische Schiefe als auch die empirische Wölbung, um ein Abweichen von der Normalverteilung festzustellen. Mit Bemerkung 2.4.4 folgt, dass JB eine asymptotische χ_2^2 -Verteilung besitzt und die Nullhypothese der Normalverteilung zum Niveau α ablehnt, falls der Wert von JB größer als das $(1-\alpha)$ -Quantil der χ_2^2 -Verteilung ist.

3.3 Testen von Regressionsresiduen

In diesem Abschnitt wird ein Schritt weitergegangen und ein Test hergeleitet, mit dem die Möglichkeit besteht, unbekannte und zudem auch unbeobachtbare Regressionsresiduen auf Normalverteilung zu testen. Dazu wird das lineare Regressionsmodell

$$\mathbf{x} = \mathbf{Y}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

mit Beobachtungsvektor $\mathbf{x} = (x_1, \dots, x_n)^T$, bekannter $(n \times d)$ -Designmatrix $\mathbf{Y} = (y_{ij})$, unbekanntem Regressionskoeffizienten β_1, \dots, β_d , zusammengefasst im Vektor $(\beta_1, \dots, \beta_d)^T = \boldsymbol{\beta}$ und i.i.d. Zufallsvariablen (*Regressionsresiduen*) $\epsilon_1, \dots, \epsilon_n$ mit $\mathbb{E}[\epsilon_i] = 0$ für alle $i = 1, \dots, n$ und $\text{Var}[\epsilon_i] = \text{Var}[\epsilon_j]$ für alle $i \neq j$ betrachtet. Es sei $g(\epsilon)$ die Dichte der Residuen ϵ_i für $i = 1, \dots, n$ und zudem sei $g(\epsilon)$ aus dem Pearson-Verteilungssystem, sodass

$$\frac{\partial}{\partial \epsilon} g(\epsilon; c_1, c_2, c_0) = -\frac{c_1 + \epsilon}{c_0 + c_1\epsilon + c_2\epsilon^2} g(\epsilon; c_1, c_2, c_0)$$

gilt. Wie im vorherigen Modell werden die j -ten empirischen Momente der Residuen ϵ_i durch $\mu_j = n^{-1} \sum_{i=1}^n \epsilon_i^j$ definiert, wobei ein entscheidender Unterschied auftritt. Da die Parameter β_i für $i = 1, \dots, d$ im vorliegenden Modell nicht bekannt sind, müssen diese geschätzt werden. Dazu sei $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{id})$. Dann gilt $\hat{\mu}_j = n^{-1} \sum_{i=1}^n \hat{\epsilon}_i^j = n^{-1} \sum_{i=1}^n (x_i - \mathbf{Y}_i \hat{\boldsymbol{\beta}})^j$, wobei als Schätzer für $\boldsymbol{\beta}$ der *Kleinste-Quadrate-Schätzer (KQS)* $\hat{\boldsymbol{\beta}}(\mathbf{x}) = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{x}$ verwendet wird.

Erneut stellt die Funktion $g(\epsilon; c_1, c_2, c_0)$ mit $c_1 = c_2 = 0$ die Dichte einer Normalverteilung dar, sodass mit $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \theta_2)$, $\boldsymbol{\theta}_1^T = (c_1, c_2)$, $\theta_2 = c_0$ die zu testende Hypothese $H_0 : \boldsymbol{\theta}_1 = (0, 0)^T$ lautet. Die Log-Likelihood-Funktion $l(c_1, c_2, c_0; \boldsymbol{\epsilon})$ ist folglich identisch zu (3.4) mit $\boldsymbol{\epsilon}$ anstelle von \mathbf{y} , sodass an der Stelle $\boldsymbol{\theta}_0$ mit dem Beweis zu Proposition 3.2.2

$$\mathcal{S}_1(\boldsymbol{\theta}_0; \boldsymbol{\epsilon})^T = n \left(\frac{\mu_1}{\sigma^2} - \frac{\mu_3}{3\sigma^4}, \frac{\mu_4}{4\sigma^4} - \frac{3}{4} \right)$$

und mit dem Beweis zu Proposition 3.2.3

$$\mathcal{I}(\boldsymbol{\theta}_0; \boldsymbol{\epsilon}) = n \begin{pmatrix} \frac{2}{3\sigma^2} & 0 & 0 \\ 0 & 6 & \frac{3}{2\sigma^2} \\ 0 & \frac{3}{2\sigma^2} & \frac{1}{2\sigma^4} \end{pmatrix}$$

folgt. Somit gilt wegen $\hat{\sigma}^2 = \hat{\mu}_2$ und mit etwas Rechenarbeit

$$\Lambda_{\mathcal{S}} = n \left(\frac{1}{6} \frac{\hat{\mu}_3^2}{\hat{\mu}_2^3} + \frac{1}{24} \left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 \right)^2 \right) + n \left(\frac{3}{2} \frac{\hat{\mu}_1^2}{\hat{\mu}_2} - \frac{\hat{\mu}_1 \hat{\mu}_3}{\hat{\mu}_2^2} \right).$$

Beachtet man, dass für die KQS-Residuen $\hat{\epsilon}_i$ das erste empirische Moment unter Schätzung des Erwartungswerts, namentlich $\hat{\mu}_1 = n^{-1} \sum_{i=1}^n \hat{\epsilon}_i = n^{-1} \left(\sum_{i=1}^n x_i - \sum_{i=1}^n \mathbf{Y}_i \hat{\boldsymbol{\beta}}(x_i) \right)$,

wegen $\hat{\beta}(x_i) = (\mathbf{Y}_i^T \mathbf{Y}_i)^{-1} \mathbf{Y}_i^T x_i = \mathbf{Y}_i^{-1} x_i$ identisch null ist, so erhält man wie in der Situation des Testens auf Normalverteilung von i.i.d. Beobachtungen

$$\Lambda_{\mathcal{S}} = n \left(\frac{1}{6} \frac{\hat{\mu}_3^2}{\hat{\mu}_2^3} + \frac{1}{24} \left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 \right)^2 \right).$$

Schreibt man in dieser Situation für die empirischen Momente unter KQ-Schätzung $\sqrt{\hat{b}_1} = \hat{\mu}_3 / \hat{\mu}_2^{3/2}$ und $\hat{b}_2 = \hat{\mu}_4 / \hat{\mu}_2^2$, so folgt

$$JB = \frac{n}{6} \left(\left(\sqrt{\hat{b}_1} \right)^2 + \frac{(\hat{b}_2 - 3)^2}{4} \right).$$

Bemerkung 3.3.1. Es besteht also auch die Möglichkeit, Regressionsresiduen mit der von Jarque und Bera vorgeschlagenen Teststatistik auf Normalverteilung zu testen. Hierzu werden ausschließlich die ersten vier empirischen Momente der KQS-Residuen $\hat{\epsilon}_i$ benötigt, für die allerdings die Regressionskoeffizienten β_i für $i = 1, \dots, d$ geschätzt werden müssen.

Kapitel 4

Der Jarque-Bera-Test im Vergleich

Obwohl der Jarque-Bera-Test (*JB*-Test) aufgrund seiner einfachen Struktur ein praktisches Werkzeug im Testen auf Normalverteilung darstellt, stellen sich bei genauerer Untersuchung einige mangelhafte Eigenschaften heraus. Auf diese soll im vorliegenden Kapitel näher eingegangen werden und zugleich werden einige Modifizierungen der Teststatistik hergeleitet. Zusätzlich wird zur Einordnung der Effizienz des *JB*-Tests ein Vergleich mit anderen – sowohl parametrischen als auch nicht-parametrischen – Tests auf Normalverteilung vorgenommen und die Ergebnisse anhand von empirischen Analysen untermauert. Begonnen wird im ersten Abschnitt mit der Herleitung empirischer kritischer Werte, die für die nachfolgenden Untersuchungen unerlässlich sind. Die Abschnitte zwei und drei gehen dann auf einen Vergleich der Power in verschiedenen Verteilungsmodellen ein.

4.1 Kritische Werte und Testalternativen

Da in vielen Gebieten der Wissenschaft der *JB*-Test ein gängiges Mittel zum Testen auf Normalverteilung sowohl von Beobachtungen, als auch von Regressionsresiduen geworden ist, gewinnt die Frage nach seiner Effizienz immer mehr an Bedeutung. Zudem werden in den meisten Anwendungen nicht seine exakten Quantile bei den Untersuchungen zur Entscheidung herangezogen, sondern wesentlich häufiger die seiner asymptotischen Verteilung. Da jedoch die Verteilung des *JB*-Tests nur sehr langsam gegen seine Grenzverteilung konvergiert, resultieren erhebliche Abweichungen der Verteilung der Teststatistik bei endlichen Stichproben von der asymptotischen Verteilung. Dies hat zur Folge, dass ungenügende Ergebnisse resultieren können, wenn die Signifikanzwerte der χ_2^2 -Verteilung dazu genutzt werden, eine Entscheidung über die Hypothese der Normalverteilung bei Vorliegen kleiner Stichproben zu treffen.

Um dennoch in sinnvoller Weise mit dem *JB*-Test arbeiten zu können, müssen für endliche und insbesondere kleine Stichproben der Größenordnung n die kritischen Werte

(kW) empirisch bestimmt werden, da die exakte Verteilung der Teststatistik für endliches n nicht angegeben werden kann. Im Laufe der Forschung rund um den JB -Test wurden so bereits in immer umfangreicheren Untersuchungen von einigen Autoren wie [Urz96], [DS96] und aktuell von [WK09] empirische kritische Werte (ekW) publiziert. Dazu wurden Monte-Carlo-Simulationen auf der Basis normalverteilter Stichproben unterschiedlicher Größenordnung durchgeführt. Um möglichst präzise Ergebnisse zu erhalten wird eine große Anzahl solcher Simulationen, sogenannte Replikationen, durchgeführt und die Werte der JB -Statistik dieser Simulationen berechnet. Die kritischen Werte zu gegebenem Signifikanzniveau α lassen sich dann anhand der JB -Werte bestimmen, das heißt zum Niveau α bildet der $(1 - \alpha) \cdot (\# \text{Replikationen})$ -größte JB -Wert den empirischen kritischen Wert. Aufgrund des Umfangs und der daraus resultierenden Genauigkeit genannter Analysen, die in eigenen Untersuchungen so nicht möglich wäre, wird an dieser Stelle auf erneute Berechnung von empirischen kritischen Werten verzichtet und auf vorhandene Datensätze zurückgegriffen.

Zum Vergleich mit den asymptotischen kritischen Werte werden die empirischen kritischen Werte für verschiedene Stichprobengrößen n und Signifikanzniveaus α angegeben, die auf der Basis von je 10^7 Replikationen von Wurtz und Katzgraber berechnet wurden und in [WK09] zu finden sind. Diese waren nach eigenem Wissensstand die genauesten zur Zeit der Veröffentlichung ihres Papers verfügbaren. Eine Übersicht der empirischen kritischen Werte gibt die Tabelle 4.1. Zur besseren Veranschaulichung der Differenzwerte $\Delta(n, \alpha) = ekW(n, \alpha) - kW(\alpha)$ für $n \in \{10, 20, 35, 50, 75, 100, 150, 200, 300, 400, 800, 1000\}$ und $\alpha \in \{0.005, 0.01, 0.05, 0.1, 0.2\}$ dient die Abbildung 4.1. Dabei wurde eine lineare Interpolation durchgeführt.

	$n = 10$	$n = 20$	$n = 35$	$n = 50$	$n = 75$	$n = 100$	$n = 150$	$n = 200$
$\alpha = 0.005$	7.300	13.471	16.414	17.281	17.305	16.959	16.257	15.638
$\alpha = 0.01$	5.703	9.718	11.736	12.392	12.586	12.491	12.185	11.882
$\alpha = 0.05$	2.525	3.795	4.593	4.976	5.278	5.430	5.598	5.676
$\alpha = 0.05$	2.525	3.795	4.593	4.976	5.278	5.430	5.598	5.676
$\alpha = 0.1$	1.623	2.347	2.881	3.183	3.486	3.673	3.904	4.033
$\alpha = 0.2$	1.124	1.562	1.916	2.128	2.346	2.487	2.656	2.756
	$n = 300$	$n = 400$	$n = 800$	$n = 1000$	$n = 1600$	$n = 2400$	$n = 10000$	$n \rightarrow \infty$
$\alpha = 0.005$	14.669	13.583	12.726	12.366	11.762	11.384	10.792	10.597
$\alpha = 0.01$	11.358	10.778	10.299	10.117	9.810	9.608	9.313	9.210
$\alpha = 0.05$	5.773	5.855	5.910	5.924	5.957	5.967	5.986	5.991
$\alpha = 0.1$	4.189	4.332	4.427	4.457	4.513	4.542	4.589	4.605
$\alpha = 0.2$	2.876	2.988	3.065	3.091	3.136	3.161	3.207	3.219

Tabelle 4.1: Empirische kritische Werte der JB -Statistik zu einigen Stichprobengrößen n und Signifikanzniveaus α bei 10^7 Replikationen. Vgl. [WK09], Tabelle 1.

Man sieht in Abbildung 4.1, dass für $\alpha \in \{0.05, 0.1, 0.2\}$ die Δ -Werte negativ, das heißt also die empirischen kritischen Werte des JB -Tests kleiner sind als die kritischen Werte der asymptotischen Verteilung und als Folge der JB -Test bei Benutzung der kritischen Werte der asymptotischen Verteilung vor allem bei kleinen Stichproben sehr konservativ ist. Die fehlende Struktur der empirischen kritischen Werte für $\alpha \in \{0.005, 0.01\}$, nämlich für $n < 20$ negative und ab $n \geq 20$ positive Δ -Werte, lässt sich nicht so leicht erklären. Dies bedeutet jedoch, dass der JB -Test bei Stichproben der Größe $n < 20$ bei Verwendung der kritischen Werte der χ_2^2 -Verteilung seltener die Nullhypothese ablehnt, als dies bei Verwendung der empirischen kritischen Werte der Fall ist und bei Stichproben der Größenordnung $n \geq 20$ bei Verwendung der asymptotischen kritischen Werte die Nullhypothese häufiger ablehnt. Dass die Δ -Werte großer Stichproben jedoch größer sind als die Δ -Werte kleiner Stichproben der Größenordnung $n \approx 20$, die nahe null liegen, zeigt deutlich, dass die Verwendung der Quantile der χ_2^2 -Verteilung zwangsläufig zu fehlerhaften Schlussfolgerungen führt.

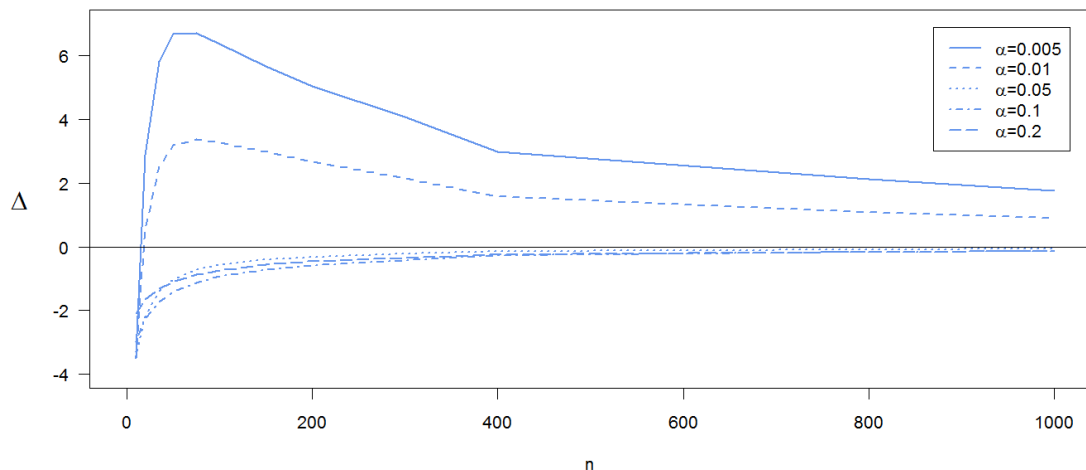


Abbildung 4.1: Abweichungen der empirischen kritischen Werte der JB -Statistik von den kritischen Werten der χ_2^2 -Verteilung in Abhängigkeit der Stichprobengröße n .

Obwohl der heute als Jarque-Bera-Test bekannte Test auf Normalverteilung erst durch Jarque und Bera, die ihn bei Ihren Untersuchungen als einen Spezialfall des Score-Tests innerhalb des Pearson-Verteilungssystems entdeckten¹, seine große Popularität erlangte, tauchte die Teststatistik im Vorfeld bereits an anderer Stelle in der Literatur auf. Bowman

¹Vgl. [JB87].

und Shenton waren es, die ihn erstmals nannten², nachdem sie herausgefunden hatten, dass sich die asymptotischen Erwartungswerte sowie die asymptotischen Varianzen der empirischen Schiefe $\sqrt{b_1}$ und der empirischen Wölbung b_2 unter Annahme der Normalverteilung als 0 und 3, bzw. $6/n$ und $24/n$ ergeben. Weiterhin zeigten sie, dass die asymptotische Kovarianz beider Größen null ist und beide Größen asymptotisch normalverteilt sind. Dies begründet in einfacher Weise die Grenzverteilung der Teststatistik, stellt vor diesem Hintergrund der JB -Test nichts weiter als die Summe zweier asymptotisch unabhängiger und quadrierter $\mathcal{N}(0, 1)$ -verteilter Zufallsvariablen dar.

	$n = 10$	$n = 20$	$n = 35$	$n = 50$	$n = 75$	$n = 100$	$n = 150$	$n = 200$
$\alpha = 0.005$	23.831	25.963	24.569	23.229	21.334	19.986	18.285	17.156
$\alpha = 0.01$	18.374	18.643	17.540	16.659	15.506	14.719	13.707	13.042
$\alpha = 0.05$	7.416	6.932	6.679	6.553	6.414	6.319	6.218	6.149
$\alpha = 0.1$	4.177	3.966	3.961	3.998	4.066	4.126	4.218	4.272
$\alpha = 0.2$	2.183	2.216	2.355	2.462	2.588	2.677	2.790	2.858
	$n = 300$	$n = 400$	$n = 800$	$n = 1000$	$n = 1600$	$n = 2400$	$n = 10000$	$n \rightarrow \infty$
$\alpha = 0.005$	15.689	14.211	13.129	12.694	11.971	11.525	10.827	10.597
$\alpha = 0.01$	12.149	11.271	10.616	10.372	9.967	9.716	9.339	9.210
$\alpha = 0.05$	6.093	6.050	6.031	6.022	6.018	6.008	5.996	5.991
$\alpha = 0.1$	4.355	4.434	4.492	4.510	4.546	4.565	4.594	4.605
$\alpha = 0.2$	2.946	3.032	3.092	3.113	3.150	3.171	3.206	3.219

Tabelle 4.2: Empirische kritische Werte der JB_U -Statistik zu einigen Stichprobengrößen n und Signifikanzniveaus α bei 10^7 Replikationen. Vgl. [WK09], Tabelle 1.

Diese Erkenntnis gibt Anlass zu einer Modifizierung der Teststatistik dahingehend, anstelle der asymptotischen Erwartungswerte und der asymptotischen Varianzen die exakten Erwartungswerte und Varianzen der Größen $\sqrt{b_1}$ und b_2 zu betrachten. Unter Verwendung der Eigenschaften der k -Statistiken in [Fis30] berechnete Urzua die Erwartungswerte und Varianzen in [Urz96] als $\mathbb{E}[\sqrt{b_1}] = 0$, $\mathbb{E}[b_2] = 3(n-1)(n+1)^{-1}$, $\text{Var}[\sqrt{b_1}] = 6(n-2)((n+1)(n+3))^{-1}$ und $\text{Var}[b_2] = 24n(n-2)(n-3)((n+1)^2(n+3)(n+5))^{-1}$, sodass eine neue, modifizierte Jarque-Bera-Teststatistik der Gestalt

$$\begin{aligned}
 JB_U &= n \left(\frac{(\sqrt{b_1})^2}{\text{Var}[\sqrt{b_1}]} + \frac{(b_2 - \mathbb{E}[b_2])^2}{\text{Var}[b_2]} \right) \\
 &= \frac{(n+1)(n+3)}{6(n-2)(n-3)} \left((n-3)(\sqrt{b_1})^2 + \frac{(n+1)(n+5)}{4n} \left(b_2 - \frac{3(n-1)}{n+1} \right)^2 \right)
 \end{aligned}$$

²Vgl. [BS75], sowie [Urz96], S.248.

resultierte. Da sich JB und JB_U asymptotisch entsprechen, besitzt auch JB_U eine χ_2^2 -Grenzverteilung und ein Vergleich der empirischen kritischen Werte der neuen Teststatistik mit denen der asymptotischen Verteilung erscheint wünschenswert. Dazu wird erneut auf die Ergebnisse in [WK09] für ausgewählte Signifikanzniveaus α zurückgegriffen. Diese sind in Tabelle 4.2 zusammengefasst. Ebenso wie im vorangehenden Fall werden die Differenzwerte $\Delta(n, \alpha)$ in Abbildung 4.2 dargestellt.

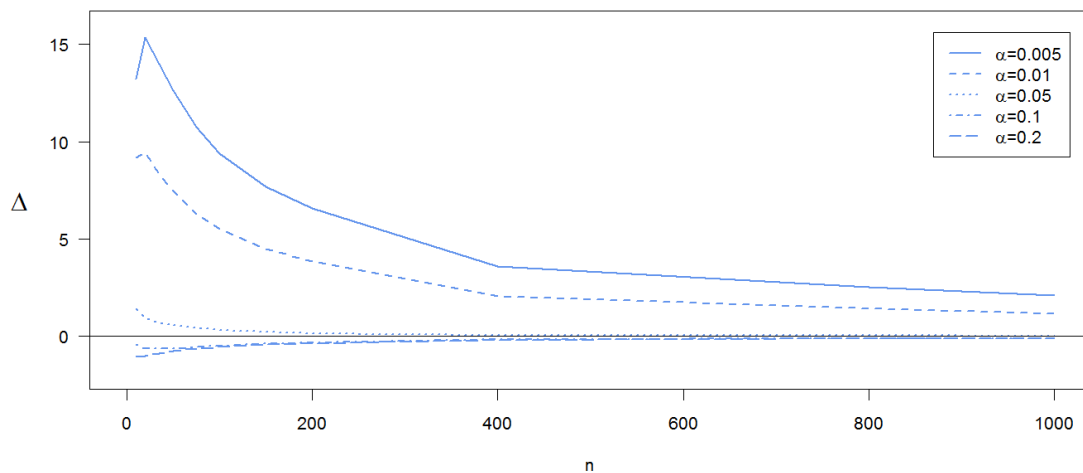


Abbildung 4.2: Abweichungen der empirischen kritischen Werte der JB_U -Statistik von den kritischen Werten der χ_2^2 -Verteilung in Abhängigkeit der Stichprobengröße n .

In Abbildung 4.2 kann man erkennen, dass für $\alpha \in \{0.005, 0.01, 0.05\}$ die Δ -Werte für Stichproben aller Größenordnungen positiv sind, wobei die Differenzwerte speziell für $\alpha = 0.05$ sehr klein sind. Daraus kann geschlossen werden, dass bei Verwendung der kritischen Werte der χ_2^2 -Verteilung besonders im Fall $\alpha \in \{0.005, 0.01\}$ die Nullhypothese bereits viel zu schnell abgelehnt wird. Die Entwicklung der Differenzwerte ist mit steigender Stichprobengröße im Bereich kleiner Stichproben wesentlich weniger von Schwankungen geprägt, als dies bei den Δ -Werten der JB -Statistik der Fall ist. Für $\alpha \in \{0.1, 0.2\}$ sind die Δ -Werte negativ, liegen jedoch etwas näher an den kritischen Werten der asymptotischen Verteilung, sodass in dieser Situation die Verwendung der kritischen Werte der χ_2^2 -Verteilung konservative Entscheidungen nach sich zieht, d.h. also die Nullhypothese erst später abgelehnt wird.

Im direkten Vergleich mit der ursprünglichen JB -Statistik zeigt sich in der Anwendung auf Regressionsresiduen, dass bei bestimmten Alternativen der neue Test eine we-

sentlich höhere Güte aufweist. Betrachtet man als Alternativhypothesen zur Hypothese der Normalverteilung beispielsweise die Studentsche t -Verteilung mit 5 Freiheitsgraden, die χ_2^2 -Verteilung, die *Laplace*-Verteilung oder die *Lognormal*-Verteilung (alle zum Erwartungswert 0 und Varianz 25), so zeigen die Ergebnisse in [Urz96], Tabelle 2, dass bei Verwendung der Signifikanzwerte der χ_2^2 -Verteilung zum Niveau $\alpha = 0.1$ die neue Statistik JB_U für Stichprobengrößen $n \in \{20, 35, 50, 100\}$ die ursprüngliche Statistik JB hinsichtlich des Fehlers 2. Art deutlich unterbietet. Bei Betrachtung der empirischen kritischen Werte anstelle der asymptotischen kritischen Werte zeigt sich ein ähnliches Bild, obwohl die Dominanz in dieser Situation nicht so eindeutig ist und bezüglich der χ_2^2 - sowie der *Lognormal*-Verteilung die JB -Statistik sogar teilweise dominiert. Zusammenfassend jedoch geben die Resultate Anlass dazu, die neue Statistik speziell in der Situation kleiner Stichproben der alten Statistik vorzuziehen.

Eine weitere Variante der Jarque-Bera-Statistik basiert auf Versionen der empirischen Schiefe und Wölbung. Da die Stichprobenwölbung b_2 aus Definition 3.2.4 nur einen verzerrten Schätzer für die theoretische Wölbung einer Normalverteilung darstellt ($\mathbb{E}[b_2 - 3] = -6(n + 1)^{-1} \neq 0$), wird diese angepasst und die auf [JG98] zurückgehenden Schiefe- und Wölbungsmaße $k_1 = \sqrt{n(n-1)}\sqrt{b_1}(n-2)^{-1}$ und $k_2 = (n-1)((n+1)(b_2-3)+6)((n-2)(n-3))^{-1}$ betrachtet, die zudem konsistent sind. Auf dieselbe Art und Weise wie im Falle des JB -Tests erhält man so eine weitere Teststatistik der Form

$$JB_{JG} = \frac{n}{6} \left(k_1^2 + \frac{k_2^2}{4} \right).$$

Da für diese Teststatistik keine empirischen kritischen Werte vorliegen, wurden diese in eigener Arbeit bestimmt. Die Vorgehensweise war dabei so, dass eine Routine in der Programmiersprache R entwickelt wurde, mit der eine Modellierung der gewünschten kritischen Werte möglich ist. Dazu wurden die JB_{JG} -Statistik, sowie die zur Beschreibung dieser Statistik notwendigen Größen implementiert und Replikationen standardnormalverteilter Stichproben verschiedener Größenordnung erzeugt. Unter Anwendung dieser Replikationen auf die JB_{JG} -Statistik konnten unter Beachtung des gewünschten Signifikanzniveaus α nach der eingangs beschriebenen Vorgehensweise, das heißt also durch Betrachtung der $(1 - \alpha) \cdot (\# \text{Replikationen})$ -größten JB_{JG} -Werte die gewünschten empirischen kritischen Werte ermittelt werden. Diese sind in der Tabelle 4.3 angegeben. Die Berechnungen beschränken sich dabei auf die Parameterwerte $\alpha = 0.05$ und $n \in \{10, 20, 50, 100, 200, 400\}$ um in dieser speziellen Situation die JB_{JG} -Statistik in die Vergleichsanalysen einbeziehen zu können. Aufgrund der geringen Computerleistung, mit der die Berechnungen durchgeführt wurden, war dies nur auf der Basis von 10^4 Replikationen möglich. Die Ergebnisse

	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
$\alpha = 0.05$	5.879	5.700	6.094	6.294	6.223	5.990

Tabelle 4.3: Empirische kritische Werte der JB_{JG} -Statistik zu einigen Stichprobengrößen n und Signifikanzniveau $\alpha = 5\%$ bei 10^4 Replikationen.

sind demzufolge wesentlich ungenauer als die in [WK09] für JB und JB_U , sollen dennoch in den Untersuchungen benutzt werden. Die Programmcodes zur Ermittlung der gewonnenen Werte sind auf der beigefügten Daten-CD im Einband am Ende der Arbeit enthalten.

4.2 Powervergleich innerhalb des Pearson-Verteilungssystems

Um zu verstehen, wie die Power der Jarque-Bera-Statistiken im Vergleich zu anderen Tests auf Normalverteilung einzuordnen ist, wird diese im vorliegenden Abschnitt innerhalb des Pearson-Verteilungssystems mit der Power anderer Tests verglichen. *Power* (auch *Teststärke* oder *Sensitivität* genannt) bezeichnet dabei die Wahrscheinlichkeit, bei Vorliegen der Alternativhypothese richtigerweise die Nullhypothese der Normalverteilung zu verwerfen. Umso höher die Power eines Tests ist, desto geringer ist also die Wahrscheinlichkeit einen Fehler 2. Art zu begehen. Dabei hängt die Power maßgeblich vom vorgegebenen Signifikanzniveau α ab, zumal dieses ja der Wahrscheinlichkeit entspricht, mit der ein Fehler 1. Art gerade noch toleriert wird. Dies motiviert also nach denjenigen Tests zu suchen, deren Power im Vergleich zu anderen Tests möglichst groß ist. Dazu wird das Testmodell in [JB87], Kapitel 5, überarbeitet und auf der Basis aktueller empirischer kritischer Werte für JB , JB_U und JB_{JG} , sowie unter Austausch und Hinzunahme einiger Tests auf Normalverteilung erweitert und präzisiert. Die Alternativen zur Normalverteilung in diesem Modell bestehen aus den folgenden Pearson-Verteilungen: $Beta(3, 2)$ -Verteilung, Student's t -Verteilung mit fünf Freiheitsgraden (5 df) und $Gamma(2, 1)$ -Verteilung. Zusätzlich wird als Nicht-Pearson-Verteilung die *Lognormal*-Verteilung untersucht. Speziell wird in der vorliegenden Situation für die Stichprobengrößen $n \in \{10, 20, 50, 100, 200, 400\}$ zum Signifikanzniveau $\alpha \in \{0.01, 0.05, 0.1\}$ die Power der Jarque-Bera-Tests mit der Power der folgenden Tests, die in der Praxis häufig angewendet werden, verglichen:

- Anderson-Darling-(Anpassungs-)Test: Der Anderson-Darling-Test ist ein Test der empirischen Verteilungsfunktion. Er basiert auf der Tatsache, dass unter Annahme der Normalverteilung eine Transformation in eine Gleichverteilung möglich ist. Dazu wird die geordnete Stichprobe $X_{(1)}, \dots, X_{(n)}$ nach der Transformation mit einem

Abstandstest auf Gleichverteilung getestet. Mit $Y_{(i)} = \hat{\sigma}^{-1}(X_{(i)} - \bar{X})$ wobei $\hat{\sigma}^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ lautet die Teststatistik

$$AD = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) (\log(\Phi(Y_{(i)})) + \log(1 - \Phi(Y_{(n-i+1)}))).$$

Φ ist die Verteilungsfunktion der $\mathcal{N}(0, 1)$ -Verteilung. Die Berechnung der p -Werte basiert dabei auf den Werten des Produkts $c \cdot AD$, wobei c ein Faktor ist, der von der Stichprobengröße sowie der zu testenden hypothetischen Verteilung abhängt.

- Lilliefors-Test: Als eine Modifizierung des Kolmogorov-Smirnov-Tests ist der Lilliefors-Test ebenso wie der Anderson-Darling-Test ein Test, der auf der empirischen Verteilungsfunktion basiert. Seine Teststatistik misst den maximalen Abstand zwischen der empirischen Verteilungsfunktion der Daten X_1, \dots, X_n und einer $\mathcal{N}(\bar{X}, \hat{\sigma}^2)$ -Verteilung, $\hat{\sigma}^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Mit $Y_i = \hat{\sigma}^{-1}(X_i - \bar{X})$ lautet die Teststatistik

$$LIL = \max\{LIL^+, LIL^-\},$$

wobei $LIL^+ = \max_{i=1, \dots, n} \{\frac{i}{n} - \Phi(Y_i)\}$ und $LIL^- = \max_{i=1, \dots, n} \{\Phi(Y_i) - \frac{i-1}{n}\}$.

- Cramér-von-Mises-Test: Der Cramér-von-Mises-Test ist ein weiterer Anpassungstest. Seine Teststatistik berechnet sich aus der aufsteigend angeordneten Stichprobe $X_{(1)}, \dots, X_{(n)}$ und besitzt mit den Bezeichnungen $Y_{(i)} = \hat{\sigma}^{-1}(X_{(i)} - \bar{X})$, $\hat{\sigma}^2 = (n-1)^{-1} \sum_{i=1}^n (X_{(i)} - \bar{X})^2$ schließlich die Gestalt

$$CVM = \frac{1}{12n} + \sum_{i=1}^n \left(\Phi(Y_{(i)}) - \frac{2i-1}{2n} \right)^2.$$

- Pearson-Chi-Quadrat-Test: Der Chi-Quadrat-Test von Pearson beruht auf einem Vergleich der Anzahl n_j an Elementen einer Stichprobe, die in k vorgegebene Klassen fallen mit der erwarteten Anzahl n_{0j} unter Annahme der Normalverteilung. Die Teststatistik lautet

$$PEA = \sum_{j=1}^k \frac{(n_j - n_{0j})^2}{n_{0j}}.$$

Aufgrund der asymptotischen χ^2 -Verteilung werden die p -Werte dabei über die χ_{k-3}^2 -Verteilung berechnet.

- Shapiro-Francia-Test: Anders als die bisherigen Tests basiert der Shapiro-Francia-Test nicht auf der empirischen Verteilungsfunktion, sondern auf der quadrierten Korrelation der geordneten Stichprobe $\mathbf{X} = (X_{(1)}, \dots, X_{(n)})$ und den geordneten

Quantilen $q_{(i)}$, $i = 1, \dots, n$, einer $\mathcal{N}(0, 1)$ -Verteilung. Dabei hängen die Quantile von der Stichprobengröße n ab. Es werden die Quantile der Werte $c_i = \frac{i-3/8}{n+1-3/4}$ für $i = 1, \dots, n$ betrachtet. Mit $\mathbf{q} = (q_{(1)}, \dots, q_{(n)})$ hat die Teststatistik somit die Form

$$SF = (\text{Cor}[\mathbf{X}, \mathbf{q}])^2.$$

Die Hypothese der Normalverteilung wird abgelehnt, falls der Wert der Statistik unterhalb des entsprechenden kritischen Wertes liegt.

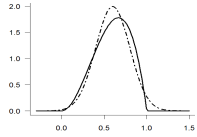
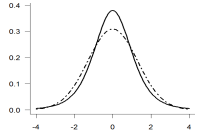
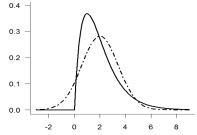
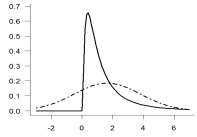
		<i>JB</i>	<i>JB_U</i>	<i>JB_{JG}</i>	<i>AD</i>	<i>LIL</i>	<i>CVM</i>	<i>PEA</i>	<i>SF</i>
 <p>Beta(3,2)</p>	$n = 10$	0.034	0.032	0.03	0.05	0.042	0.048	0.074	0.048
	$n = 20$	0.022	0.016	0.018	0.076	0.064	0.072	0.052	0.044
	$n = 50$	0.02	0.016	0.016	0.168	0.118	0.14	0.108	0.116
	$n = 100$	0.092	0.038	0.03	0.406	0.22	0.316	0.174	0.336
	$n = 200$	0.732	0.648	0.6	0.812	0.534	0.704	0.386	0.872
	$n = 400$	0.996	0.998	0.996	0.986	0.864	0.974	0.794	1
 <p>Student's t (5 df)</p>	$n = 10$	0.158	0.156	0.144	0.134	0.104	0.126	0.112	0.146
	$n = 20$	0.222	0.232	0.24	0.196	0.114	0.166	0.09	0.23
	$n = 50$	0.416	0.426	0.424	0.274	0.204	0.242	0.13	0.394
	$n = 100$	0.642	0.66	0.646	0.51	0.326	0.45	0.184	0.648
	$n = 200$	0.89	0.894	0.894	0.732	0.534	0.7	0.288	0.876
	$n = 400$	0.986	0.984	0.986	0.96	0.812	0.92	0.474	0.982
 <p>Gamma(2,1)</p>	$n = 10$	0.192	0.176	0.174	0.202	0.152	0.178	0.182	0.206
	$n = 20$	0.382	0.352	0.358	0.44	0.282	0.394	0.286	0.482
	$n = 50$	0.848	0.792	0.798	0.906	0.726	0.862	0.664	0.932
	$n = 100$	0.992	0.988	0.988	0.998	0.954	0.992	0.926	1
	$n = 200$	1	1	1	1	1	1	1	1
	$n = 400$	1	1	1	1	1	0.996	1	1
 <p>Log-Normal</p>	$n = 10$	0.5	0.432	0.43	0.584	0.464	0.558	0.53	0.594
	$n = 20$	0.832	0.81	0.818	0.892	0.798	0.874	0.82	0.904
	$n = 50$	1	1	1	1	0.996	1	0.998	1
	$n = 100$	1	1	1	1	1	0.884	1	1
	$n = 200$	1	1	1	1	1	0.158	1	1
	$n = 400$	1	1	1	1	1	0	1	1

Tabelle 4.4: Powerwerte verschiedener Tests auf Normalverteilung zum Signifikanzniveau $\alpha = 5\%$ bei einigen Stichprobengrößen n und 500 Replikationen. Die effizientesten Tests in den einzelnen Kategorien sind durch Fettdruck hervorgehoben. Die Grafiken zeigen die Dichte der entsprechenden Verteilungen (durchgezogene Linien) im Vergleich zur normierten und zentrierten Normalverteilung (gestrichelte Linien).

In der Tabelle 4.4 sind die Powerwerte der einzelnen Teststatistiken zu verschiedenen Stichprobengrößen n und dem Signifikanzniveau $\alpha = 5\%$ angegeben, die bei den durch-

geführten Simulationen im Rahmen von 500 Replikationen resultierten. Um für beliebige Stichprobengrößen zu unterschiedlichen Signifikanzniveaus die Power der einzelnen Tests ermitteln zu können, wurde erneut in der Programmiersprache R eine Routine erstellt. Dazu wurden die Jarque-Bera-Statistiken JB , JB_U und JB_{JG} implementiert und für jede der vier Verteilungen ($Beta(3, 2)$, Student's t (5 df), $Gamma(2, 1)$ und $LogNormal$) Replikationen von Stichproben der Größenordnung $n \in \{10, 20, 50, 100, 200, 400\}$ erzeugt. Mit dem Paket „nortest“ hat man Zugriff auf die bereits in R implementierten Teststatistiken AD , LIL , CVM , PEA und SF , sodass eine Implementierung dieser Teststatistiken nicht notwendig war. Die betrachteten Tests wurden dann auf die Replikationen angewendet und innerhalb der im Paket „nortest“ vorhandenen Normalverteilungstests für jede Stichprobe die Nullhypothese der Normalverteilung abgelehnt, falls der p -Wert der entsprechenden Statistik unterhalb des vorgegebenen Niveaus α lag. Die Power der einzelnen Tests innerhalb der verschiedenen Verteilungen konnte daraufhin anhand der Anzahl der abgelehnten Nullhypothesen im Verhältnis zur Gesamtzahl der durchgeführten Replikationen bestimmt werden. Die Berechnung der Power im Fall der JB -, JB_U - und JB_{JG} -Statistiken wurde über die empirischen kritischen Werte vorgenommen, die in den Tabellen 4.1 bis 4.3 angegeben sind. Hierbei wurde für jede Stichprobe die Nullhypothese der Normalverteilung abgelehnt, falls der Wert der entsprechenden Jarque-Bera-Statistik oberhalb des zugehörigen empirischen kritischen Wertes lag. Die Berechnung der Power geschah analog zur Vorgehensweise bei den vorher beschriebenen Normalitätstests, also anhand der Anzahl der abgelehnten Nullhypothesen im Verhältnis zur Gesamtzahl der durchgeführten Replikationen. Die Programmcodes der verwendeten Routine sind auf der beigefügten Daten-CD im Einband am Ende der Arbeit enthalten.

Da für den p -Wert im Falle eines rechtsseitigen Tests (also eines solchen Tests, bei dem die Nullhypothese abgelehnt wird, sobald der Wert der Teststatistik einen bestimmten kritischen Wert c überschreitet) bei Vorliegen einer Stichprobe \mathbf{x} und einer Teststatistik S mit dem Ergebnis $S(\mathbf{x}) = s$ die Beziehung $p = P(S \geq s | H_0)$ erfüllt ist und für den kritischen Wert c zum Signifikanzniveau α die Beziehung $\alpha = P(S \geq c | H_0)$ gilt, lässt sich auch eine Beziehung zwischen dem (erst nach Erhalt der Stichprobe bekannten) p -Wert und dem (vor Auswertung der Teststatistik festzulegenden) Signifikanzniveau α ausmachen. Es gilt also: $p < \alpha \Leftrightarrow s > c$. Dies liefert die Rechtfertigung dafür, dass die auf verschiedene Weise berechnete Power in den betrachteten Testsituationen miteinander verglichen werden darf.

Sicherlich können aufgrund der geringen Anzahl an Replikationen keine exakten Schlüsse über die absolute Power der einzelnen Tests gezogen werden, jedoch zeichnen sich Tendenzen ab, die eine Vermutung über die relativen Verhältnisse zulassen. Da die Untersuchungen auch in den Fällen $\alpha = 0.01$ und $\alpha = 0.1$ durchgeführt wurden, die Ergebnisse

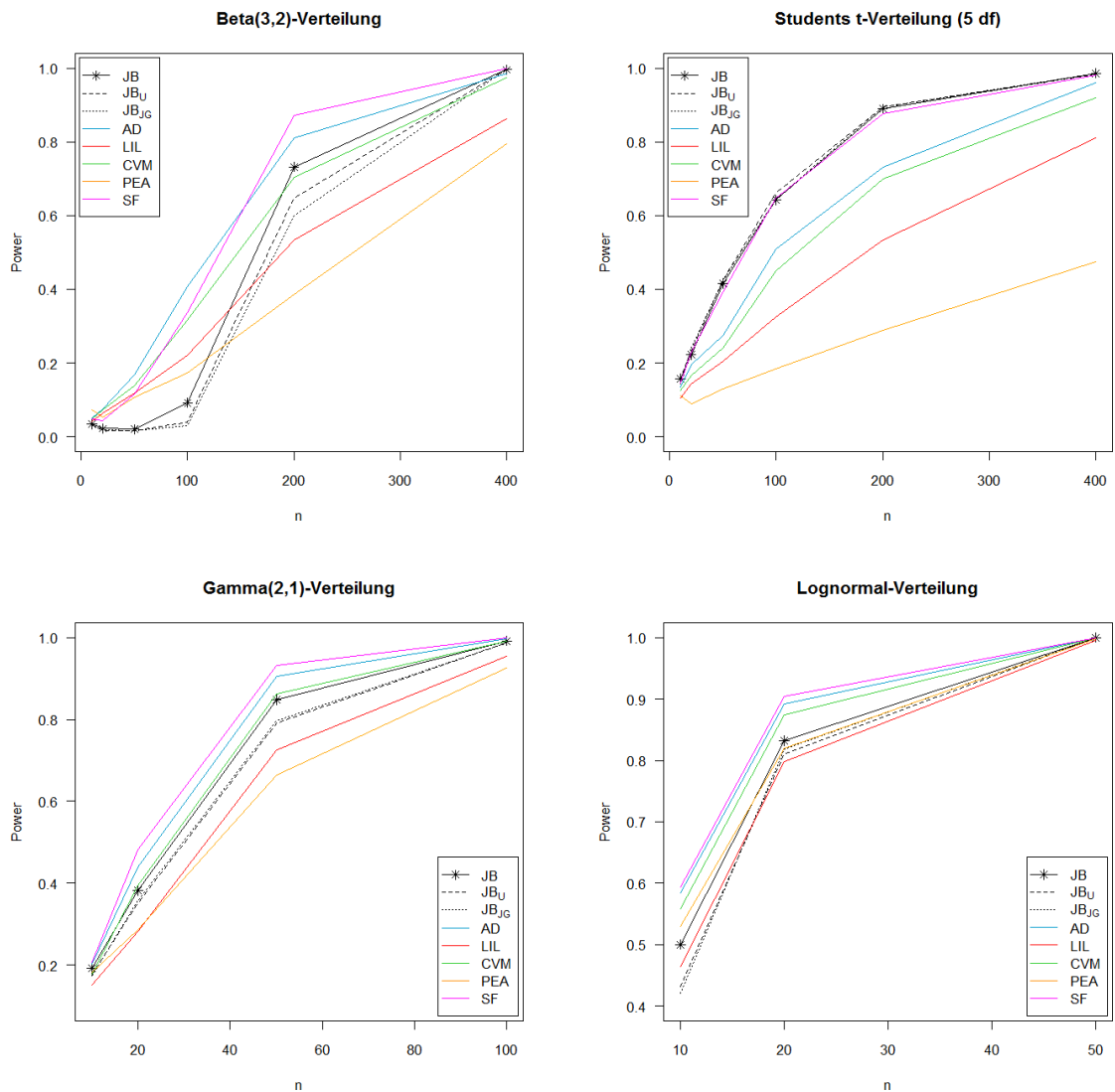


Abbildung 4.3: Power der Teststatistiken JB , JB_U , JB_{JG} , AD , LIL , CVM , PEA und SF zum Signifikanzniveau $\alpha = 5\%$ in Abhängigkeit der Stichprobengröße n . Die exakten Ergebnisse der Analysen können der Tabelle 4.4 entnommen werden.

denen aus Tabelle 4.4 jedoch recht ähnlich sind, werden diese im Anhang B.2 nachgereicht. Zur besseren Übersicht wurden die Testergebnisse aus Tabelle 4.4 mit R grafisch aufbereitet und werden in Abbildung 4.3 wiedergegeben. In den Grafiken wurde dabei eine lineare Interpolation durchgeführt. Die Ergebnisse der Stichprobengrößen $n \in \{10, 20, 50, 100, 200, 400\}$ sind auf den Powerkurven der JB -Statistik durch Sternchen gekennzeichnet, wobei für die $Gamma(2, 1)$ -Verteilung nur $n \in \{10, 20, 50, 100\}$ und für die $Lognormal$ -Verteilung nur $n \in \{10, 20, 50\}$ betrachtet wird.

Fazit: Zu erkennen ist in Abbildung 4.3, dass sowohl für die $Gamma(2, 1)$ -Verteilung (Grafik unten links) als auch für die $Lognormal$ -Verteilung (Grafik unten rechts) die Powerkurve des Shapiro-Francia-Tests (violette Kurve) bezüglich aller Stichprobengrößen n oberhalb der übrigen Powerkurven liegt und somit die anderen Teststatistiken überbietet. Hinsichtlich der $Gamma(2, 1)$ -Verteilung liegen die Jarque-Bera-Statistiken (schwarze Kurven) für alle Stichprobengrößen etwa gleichauf im Mittelfeld. Bezüglich der $Lognormal$ -Verteilung jedoch zeigt sich, dass vor allem die modifizierten Jarque-Bera-Statistiken bezüglich der Power für kleine Stichproben schlecht abschneiden. Die JB -Statistik schneidet etwas besser ab, liegt dennoch nur im unteren Mittelfeld. Im Testen gegen die $Beta(3, 2)$ -Verteilung (Grafik oben links) übertrifft speziell in der Situation kleiner bis mittelgroßer Stichproben, d.h. $n \in \{20, 50, 100\}$, die Powerkurve des Anderson-Darling-Tests (blaue Kurve) die Powerkurven der restlichen Teststatistiken. Bei großen Stichproben wiederum dominiert der Shapiro-Francia-Test. Die Jarque-Bera-Statistiken liegen hier bei Stichproben der Größenordnung $n \leq 100$ deutlich abgeschlagen am unteren Ende der Powerskala, steigen jedoch bereits bei $n \approx 200$ ins Mittelfeld auf und bieten für $n = 400$ zusammen mit dem Anderson-Darling- und Shapiro-Francia-Test sogar die höchste Power. Im Falle von Student's t -Verteilung mit fünf Freiheitsgraden (Grafik oben rechts) bieten die Jarque-Bera-Statistiken bei allen Stichprobengrößen die höchste Power. Die restlichen Teststatistiken LIL , CVM und PEA konnten bis auf den Chi-Quadrat-Test von Pearson im Falle der $Beta(3, 2)$ -Verteilung für $n = 10$ in keiner der getesteten Situationen für kleine und auch große Stichprobengrößen mit den anderen Tests mithalten.

Obwohl gegen die t -Verteilung mit fünf Freiheitsgraden die Statistiken JB_U und JB_{JG} bei fast allen Stichprobengrößen die JB -Statistik dominiert³, sind die Unterschiede doch sehr gering. Hinsichtlich der restlichen Verteilungen ist die Power der JB -Statistik bei jeder Stichprobengröße mindestens genauso groß, in den meisten Fällen sogar größer als die der Statistiken JB_U und JB_{JG} . Somit kann davon ausgegangen werden, dass weder der JB_U -Test, noch der JB_{JG} -Test in der Situation des Testens auf Normalverteilung von Beobachtungen eine überzeugende Verbesserung gegenüber dem JB -Test darstellt.

4.3 Powervergleich in speziellen Situationen

Der vorige Abschnitt hat gezeigt, dass abgesehen von der Stichprobengröße vor allem die Form der zugrunde liegenden Verteilung die Grundlage für eine Bewertung verschiedener Teststatistiken in unterschiedlichen Testsituationen darstellt. So ist es doch verwunderlich, dass sich speziell im Fall der t -Verteilung als einzige symmetrische Verteilung die Jarque-Bera-Statistiken als stärkste Tests herausgestellt haben, basieren die Jarque-Bera-

³Vgl. dazu Tabelle 4.4.

Statistiken doch auf der Schiefe und der Wölbung der zugrunde liegenden Verteilung. Die Schiefe im Falle einer symmetrischen Verteilung, also insbesondere der t -Verteilung, ist allerdings null und nur noch die Wölbung trägt ihren Teil zur Teststatistik bei. Dies motiviert die Frage nach der Gestalt, welche dazu führt, dass einige Tests in gewissen Situationen andere Tests dominieren. Um auf diese Frage näher eingehen zu können, wird im Folgenden von den Verteilungen des vorigen Abschnitts Abstand genommen und ein Modell betrachtet, mit dem nach individuellen Wünschen Verteilungsformen realisiert werden können. Dieses Modell bietet die Möglichkeit Verteilungen zu generieren, die bestimmte Eigenschaften aufweisen und zudem „nah“ an der Normalverteilung liegen. So können die Auswirkungen unterschiedlicher Abweichungen von der Normalverteilung untersucht werden. Dazu werden verschiedene Formen symmetrischer wie auch nicht symmetrischer Verteilungen mit und ohne Abweichung der für die Normalverteilung typischen Wölbung erzeugt. Zudem werden nicht nur unimodale, sondern auch bimodale Verteilungen betrachtet. In Anlehnung an das Modell von Thadewald und Büning in [TB04] zur empirischen Untersuchung des Jarque-Bera-Tests wird wie folgt vorgegangen. Es seien X_1 und X_2 unabhängige Zufallsvariablen mit $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Weiter sei $p \in [0, 1]$. Dann wird die Zufallsvariable $Z = (1 - p)X_1 + pX_2$ betrachtet, deren Verteilung als *Kontaminierte Normalverteilung* (KN -Verteilung) bezeichnet wird. Es sei o.B.d.A. $\mu_1 = 0$ und $\sigma_1^2 = 1$ sowie $\mu_2 \in \{0, 1, 2, 3\}$, $\sigma_2^2 \in \{0.5, 1, 3\}$ und $p \in \{0.1, 0.25, 0.3, 0.5, 0.75\}$. Durch Kombination dieser Parameter läßt sich bereits eine Vielzahl unterschiedlicher KN -Verteilungen erzeugen. Diese werden im Folgenden benutzt um die Power einiger Teststatistiken miteinander zu vergleichen.

Aufgrund der schlechten Resultate der Statistiken LIL , CVM und PEA in den Testsituationen des vorigen Abschnitts beschränken sich die folgenden Analysen auf die Teststatistiken JB , AD und SF . Die Vorgehensweise in der Erstellung einer geeigneten Routine zur Ermittlung der Power in den einzelnen Testsituationen entspricht dabei einer Modifizierung des Modells aus dem vorigen Abschnitt. Dazu wurde in der Programmiersprache R eine Funktion programmiert, mit der die Erzeugung kontaminiert normalverteilter Zufallsgrößen möglich ist. Diese wurde dann dazu genutzt, Replikationen von Stichproben unterschiedlicher Größenordnung zu generieren, auf welche die Teststatistiken JB , AD und SF angewendet wurden. (Die Implementierung der JB -Teststatistik wurde dabei aus dem vorigen Modell übernommen. Die Statistiken AD und SF wurden erneut aus dem Paket „nortest“ entnommen.) Im Falle des Anderson-Darling- und des Shapiro-Francia-Tests wurde für jede Replikation die Nullhypothese der Normalverteilung abgelehnt, falls der berechnete p -Wert das Signifikanzniveau unterschritten hat. Im Falle des Jarque-Bera-Tests wurde die Nullhypothese abgelehnt, falls der resultierende Wert den entsprechenden empirischen kritischen Wert aus Tabelle 4.1 überstieg. Die Anzahl der abgelehnten Nullhy-

pothesen im Verhältnis zur Gesamtzahl der durchgeführten Replikationen ergab dann die in den Tabellen B.1 bis B.3 im Anhang B.1 angegebenen Powerwerte. Dass ein Vergleich der Power, die auf unterschiedliche Weise berechnet wurde, gezogen werden darf, wurde bereits im letzten Abschnitt begründet. Die Programmcodes der verwendeten Routine sind auf der beigelegten Daten-CD im Einband am Ende der Arbeit enthalten.

Die Grafiken (A) bis (F) in den Abbildungen 4.4 bis 4.6 veranschaulichen die Ergebnisse der Simulationen, die auf der Basis von je 1000 Replikationen für die Stichprobengrößen $n \in \{20, 50, 100, 200, 400\}$ zum Signifikanzniveau $\alpha = 5\%$ durchgeführt wurden. Innerhalb der Grafiken (A) bis (F) sind in den oberen Schaubildern die Powerkurven des Jarque-Bera-, des Anderson-Darling-, und des Shapiro-Francia-Tests zu verschiedenen KN -Verteilungen dargestellt, wobei erneut eine lineare Interpolation durchgeführt wurde. Die Kombination der Parameterwerte μ_2 , σ_2^2 und p der KN -Verteilung ist oberhalb des jeweiligen Schaubilds angegeben. Zur besseren Übersicht wird mit den Sternchen auf den JB -Kurven die Lage der berechneten JB -Werte für die Stichprobengrößen $n \in \{20, 50, 100, 200, 400\}$ angedeutet. Unterhalb dieser Schaubilder sind die Dichten der zugehörigen KN -Verteilungen (durchgezogene Linien) im Vergleich zur zentrierten und standardisierten Normalverteilung (gestrichelte Linien) visualisiert. Diese bieten die Möglichkeit zu erkennen, wie stark sich eine spezielle Abweichung von der Normalverteilung auf die Power der einzelnen Teststatistiken auswirkt. Alle Grafiken wurden dabei mit R erstellt.

Fazit: Betrachtet man die Abbildungen 4.4 bis 4.6, so kann man erkennen, dass sich die absolute Power des JB -Tests hinsichtlich der verschiedenen Verteilungsformen in Abhängigkeit der Stichprobengröße deutlich unterscheidet. In den Grafiken 4.4(C), 4.4(F) und 4.6(A) ist gut zu erkennen, wie groß der Einfluss ist, den die Wölbung auf die Power des Jarque-Bera-Tests hat. Innerhalb der univariaten Verteilungen wird im Falle der (im Vergleich zur Normalverteilung) deutlich spitzigeren Verteilung (also bei einer sehr großen Wölbung) die Nullhypothese bereits bei einer Stichprobengröße von $n = 100$ fast zu 100% abgelehnt, wohingegen bei der flachen Verteilung (also einer kleinen Wölbung) selbst bei $n = 400$ die Chance nicht einmal bei 50% liegt, dass die tatsächliche Verteilung als Nicht-Normalverteilung aufgedeckt wird. Im Vergleich mit der spitzigen Verteilung schneidet die Power des JB -Tests sogar bei der bimodalen Verteilungsform deutlich schlechter ab. Hier wird bei einer Stichprobengröße von $n = 100$ nur mit etwa 20% Wahrscheinlichkeit die richtige Entscheidung gegen die Normalverteilung getroffen. Bei $n = 200$ liegt die Power jedoch schon etwa bei 0.9. Den großen Einfluss der Wölbung kann man auch erkennen, wenn man die Powerkurven in den Grafiken der Abbildung 4.5 von links nach rechts (also (A)-(C) und (D)-(F)) und die Powerkurven in den Grafiken der Abbildung 4.6 von rechts

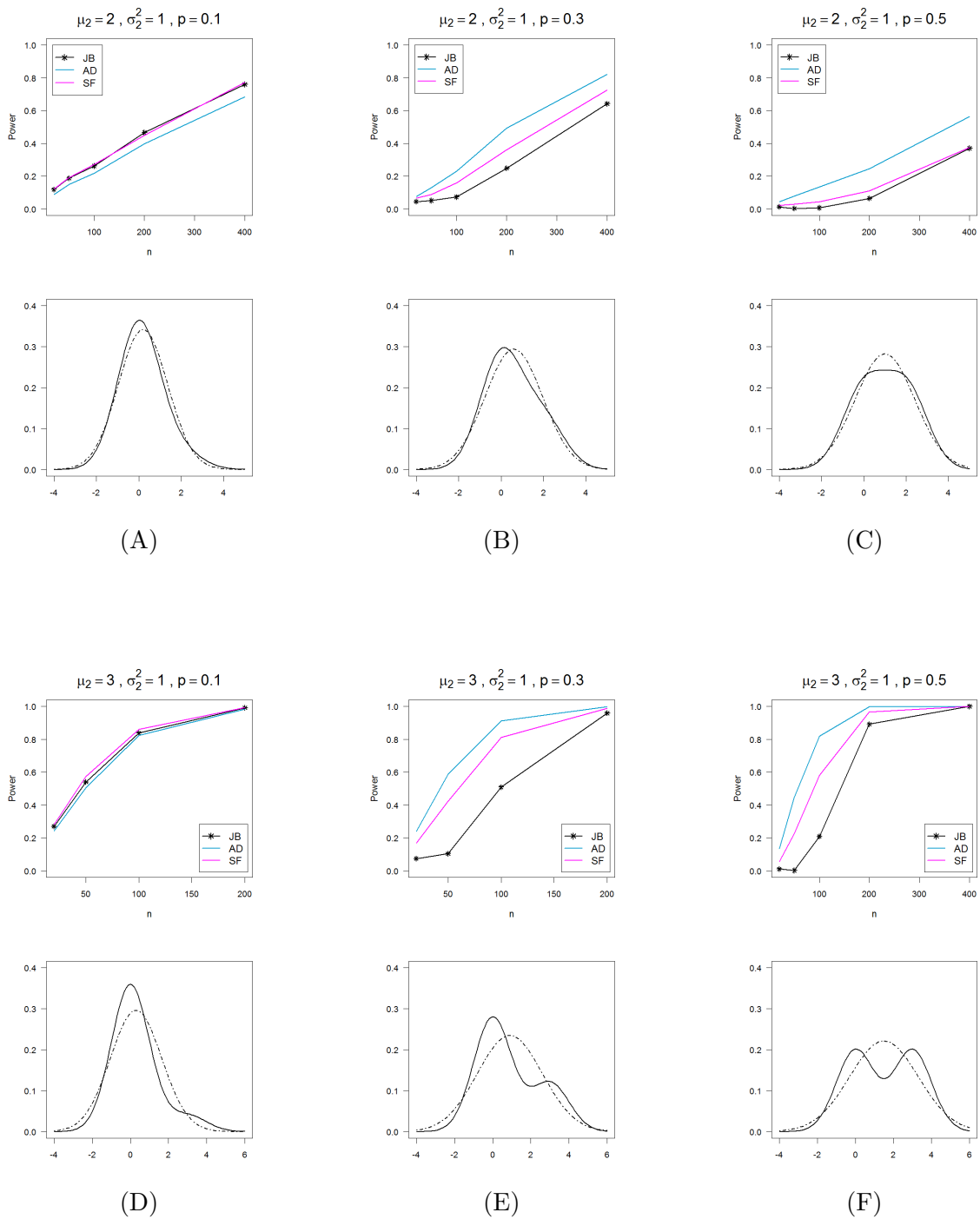


Abbildung 4.4: Power der Teststatistiken JB , AD und SF zum Signifikanzniveau $\alpha = 5\%$ in Abhängigkeit der Stichprobengröße n für verschiedene KN -Verteilungen mit Parametern $\mu_1 = 0, \mu_2 \in \{2, 3\}, \sigma_1^2 = \sigma_2^2 = 1$ und $p \in \{0.1, 0.3, 0.5\}$. Die exakten Ergebnisse der Analysen können der Tabelle B.1 im Anhang entnommen werden.

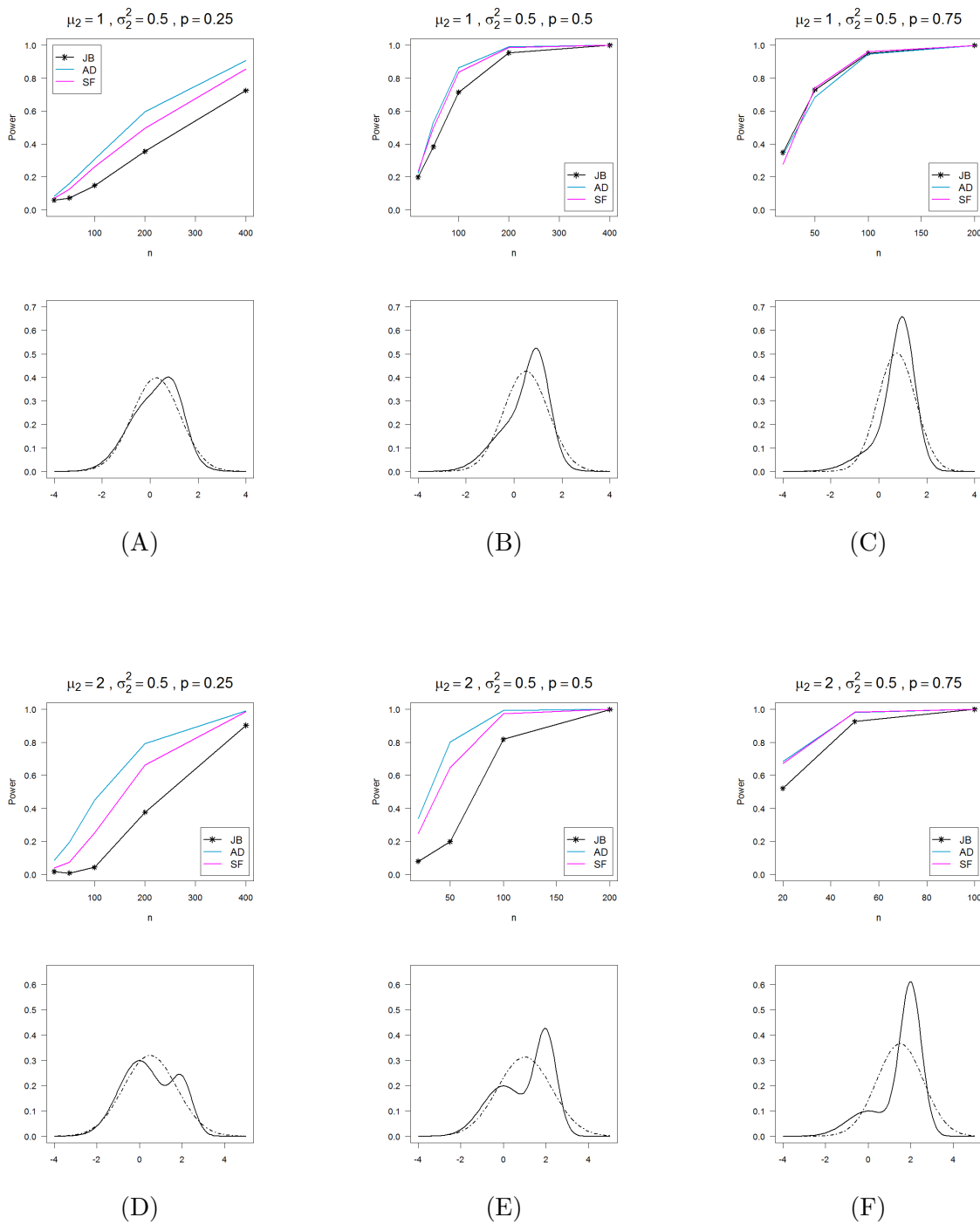


Abbildung 4.5: Power der Teststatistiken JB , AD und SF zum Signifikanzniveau $\alpha = 5\%$ in Abhängigkeit der Stichprobengröße n für verschiedene KN -Verteilungen mit Parametern $\mu_1 = 0, \mu_2 \in \{1, 2\}, \sigma_1^2 = 1, \sigma_2^2 = 0.5$ und $p \in \{0.25, 0.5, 0.75\}$. Die exakten Ergebnisse der Analysen können der Tabelle B.2 im Anhang entnommen werden.

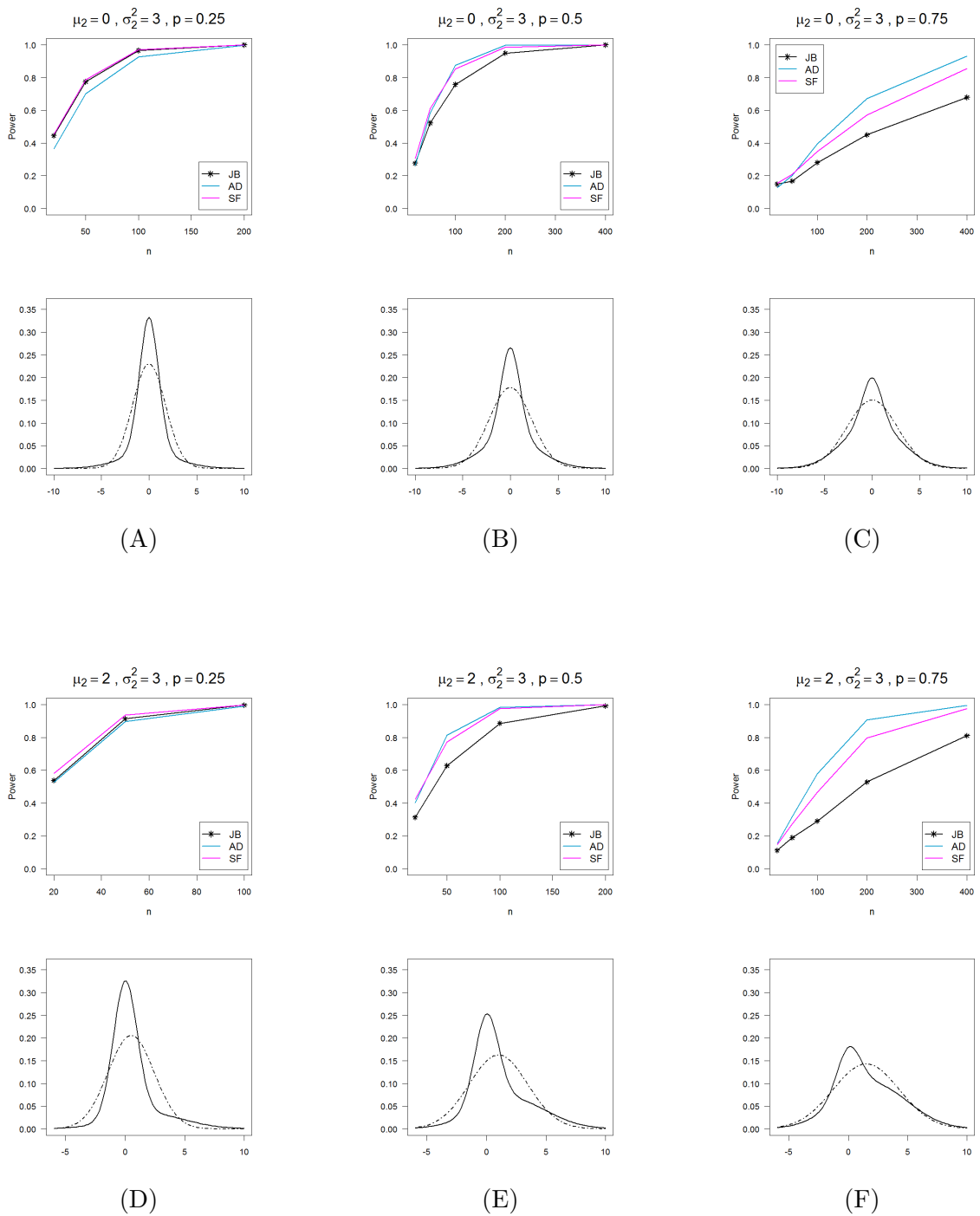


Abbildung 4.6: Power der Teststatistiken JB , AD und SF zum Signifikanzniveau $\alpha = 5\%$ in Abhängigkeit der Stichprobengröße n für verschiedene KN -Verteilungen mit Parametern $\mu_1 = 0$, $\mu_2 \in \{0, 2\}$, $\sigma_1^2 = 1$, $\sigma_2^2 = 3$ und $p \in \{0.25, 0.5, 0.75\}$. Die exakten Ergebnisse der Analysen können der Tabelle B.3 im Anhang entnommen werden.

nach links (also (C)-(A) und (F)-(D)) durchläuft. Mit steigender Wölbung der zugrunde liegenden Verteilung steigt auch die Power des Jarque-Bera-Tests erheblich. Der Einfluss der Schiefe einer Verteilung auf die Power der JB -Statistik ohne entsprechende Abweichung der Wölbung hingegen ist eher gering, wie man gut in den Grafiken 4.4(B), 4.5(A) und 4.5(D) erkennen kann. Es verwundert nicht, dass in diesen Situationen im direkten Vergleich mit dem Anderson-Darling- und dem Shapiro-Francia-Test der Jarque-Bera-Test bei allen Stichprobengrößen eine deutlich niedrigere Power aufweist. Liegt in der Grundverteilung eine Kombination von abweichender Schiefe und Wölbung vor, so spielt also die Wölbung eine weitaus wichtigere Rolle für die Power der JB -Statistik. Eine weitere Schwäche des Jarque-Bera-Tests wird bei Betrachtung der bimodalen Verteilungen in den Grafiken 4.4(E), 4.4(F) und 4.5(E) ersichtlich. Erst ab einer Stichprobengröße von $n = 200$ kann der Jarque-Bera-Test mit den anderen Teststatistiken bezüglich seiner Power konkurrieren, in 4.5(D) sogar erst ab $n = 400$.

Diese Resultate lassen erahnen, wie fehlerhaft eine Entscheidung auf Basis des Jarque-Bera-Tests sein kann, falls eine bimodale Verteilung oder eine Verteilung ohne ausreichende Wölbung zugrunde liegt und die vorliegende Stichprobe zudem keine ausreichende Größe besitzt. Ist hingegen eine starke Abweichung der Wölbung von der einer Normalverteilung zu erkennen, so stellt der Jarque-Bera-Test auch für kleine Stichproben ein hilfreiches und zudem gleichwertiges Mittel zum Testen auf Normalverteilung dar.

Kapitel 5

Testen auf multivariate Normalverteilung

Die Theorie des Testens auf univariate Normalverteilung basierte auf der Annahme einer Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$, welche als Realisierung von sowohl unabhängigen, als auch identisch verteilten Zufallsvariablen X_1, \dots, X_n angenommen wurde. Um im Folgenden nicht nur auf Zusammenhänge zwischen einzelnen Beobachtungen eingehen zu können, sondern auch auf verschiedene Merkmale dieser Beobachtungen, wird der Ansatz erweitert und in diesem Kapitel Stichproben $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ betrachten, die aus p -Zufallsvektoren $\mathbf{X}_i = (X_{1i}, \dots, X_{pi})^T$ für $i = 1, \dots, n$ bestehen. X_{ji} für $j = 1, \dots, p$ stellen dabei eindimensionale Zufallsvariablen dar. \mathbf{X}_i wird dadurch zu einer Beobachtung, die p Merkmale beinhaltet, deren Abhängigkeitsstruktur untereinander durch die $(p \times p)$ -Kovarianzmatrix $\text{Cov}[\mathbf{X}_i] = (\sigma_{jk})_{1 \leq j, k \leq p}$ mit $\sigma_{jk} = \text{Cov}[X_{ji}, X_{ki}]$ für $i = 1, \dots, n$ ausgedrückt werden kann. Dazu wird im Folgenden angenommen, dass sowohl die Erwartungswertvektoren $\mathbb{E}[\mathbf{X}_i]$, als auch die Kovarianzmatrizen $\text{Cov}[\mathbf{X}_i]$ der betrachteten Zufallsvektoren \mathbf{X}_i für $i = 1, \dots, n$ existieren. Die einzelnen Zufallsvektoren werden dabei wie im univariaten Fall als unabhängig und identisch verteilt angenommen. Eine Stichprobe \mathbf{X} bestehend aus n

Beobachtungsvektoren der Länge p kann somit als $(p \times n)$ -Matrix $\mathbf{X} = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{pi} & \cdots & X_{pn} \end{pmatrix}$

dargestellt werden. Man bezeichnet \mathbf{X} als *Daten-* oder *Beobachtungsmatrix*. Geometrisch lassen sich die Spalten der Matrix als Punkte im p -dimensionalen Raum deuten, welche der Untersuchung eines Zusammenhangs der verschiedenen Objekte (im eindimensionalen also gerade der einzelnen Beobachtungen) untereinander dienen können. Die Zeilen hingegen sind Punkte im n -dimensionalen Raum und werden benutzt um Beziehungen zwischen den verschiedenen Merkmalen aufzudecken. Statistische Fragestellungen, die dem angegebenen Modell zugrunde liegen, werden in den Bereich der multivariaten Statistik

eingeordnet. Ebenso wie im univariaten Fall besteht nun Interesse daran, anhand einer vorliegenden Stichprobe Aussagen über die Verteilung der den Beobachtungen zugrunde liegenden Grundgesamtheit zu treffen.

Im Folgenden liegt das Augenmerk speziell auf der Erweiterung des in den vorigen Kapiteln entwickelten Modells. Zwangsläufig stellt sich die Frage, ob sich das von Jarque und Bera hergeleitete Ergebnis einer Teststatistik zum Testen auf univariate Normalverteilung in ähnlicher Weise auf den multivariaten Fall übertragen lässt. Dazu sind die im eindimensionalen Fall recht anschaulichen Maße für die Schiefe und die Wölbung einer Verteilung auf die multivariaten Situationen zu übertragen. Da es sich jedoch nicht um eindeutige Fortsetzungen handelt, wie man bereits an der Vielfalt der sich mit diesem Thema auseinandersetzenen Literatur erkennen kann, bietet sich die Möglichkeit mehrerer Testerweiterungen. Die vorliegende Analyse beschränkt sich dabei auf die zwei wohl bekanntesten Fortsetzungen von Kantilal Vardichand Mardia, es wird jedoch darauf hingewiesen, dass weitere Vorschläge für multivariate Maße wie die in [Sri84] und [MA73], sowie in [Sma80] und [Son01] zur Verfügung stehen. In Abschnitt eins werden dazu die in [Mar74] angegebenen Maße für die multivariate Schiefe sowie die multivariate Wölbung eingeführt und einige wichtige Eigenschaften dieser Maße gegeben. Abschnitt zwei befasst sich mit der asymptotischen Verteilung der betrachteten Größen, bevor anschließend in Abschnitt drei einige daraus resultierende Teststatistiken untersucht werden.

5.1 Multivariate Schiefe- und Wölbungsmaße

Wie einleitend erwähnt, wird auf die Ergebnisse von Mardia zurückgegriffen und die Schiefe- und Wölbungsmaße angegeben, die in [Mar74] wie folgt definiert werden.

Definition 5.1.1. Es seien $\mathbf{X} = (X_1, \dots, X_p)^T$ und $\mathbf{Y} = (Y_1, \dots, Y_p)^T$ unabhängige und identisch verteilte p -Zufallsvektoren mit existierendem Erwartungswertvektor $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^T$ und existierender Kovarianzmatrix $\boldsymbol{\Sigma}$. Dann bezeichne in Übereinstimmung der Notation des univariaten Falls

$$\beta_{1,p} = \mathbb{E} \left[\left((\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right)^3 \right] \quad (5.1)$$

das *multivariate Schiefemaß* nach Mardia und

$$\beta_{2,p} = \mathbb{E} \left[\left((\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right)^2 \right] \quad (5.2)$$

das *multivariate Wölbungsmaß* nach Mardia.

Die Motivation für die Darstellung des Schiefemaßes in vorliegender Form lässt sich anhand von [Mar70] wie folgt begründen. Ausgehend von dem in [KS67], S.466 hergeleiteten

Ergebnis der asymptotischen Äquivalenz von $\text{Cor}[\bar{X}, S^2]$ und $(\beta_1/2)^{1/2}$ mit \bar{X} als Stichprobenmittel der univariaten Zufallsvariablen X_1, \dots, X_n , S^2 als Stichprobenvarianz und β_1 als univariatem Schiefemaß (der zugrunde liegenden Grundgesamtheit) unter gewissen Konvergenzvoraussetzungen an die zweiten Momente ($\text{Var}[\bar{X}]$, $\text{Var}[S^2]$, $\text{Cov}[\bar{X}, S^2]$ zur Ordnung n^{-1}), versuchte Mardia die asymptotische Beziehung zwischen der Korrelation oben angegebener Größen und der Schiefe auf den mehrdimensionalen Fall zu übertragen. Resultat war das Maß

$$\beta_{1,p} = \sum_{r,s,t=1}^p \sum_{r',s',t'=1}^p \sigma^{rr'} \sigma^{ss'} \sigma^{tt'} \mu_{111}^{(rst)} \mu_{111}^{r's't'}$$

mit $(\sigma^{ij})_{1 \leq i,j \leq p} = \Sigma^{-1}$ und $\mu_{r_1, \dots, r_k}^{(s_1, \dots, s_k)} = \mathbb{E}[\prod_{i=1}^k (X_{s_i} - \mu_{s_i})^{r_i}]$.

Der multivariaten Wölbung liegt eine Beziehung zum Einstichproben-Pitman-Permutationstest zugrunde, sodass

$$\beta_{2,p} = \sum_{r,s=1}^p \sum_{r',s'=1}^p \sigma^{rs} \sigma^{r's'} \mu_{1111}^{(rsr's')}$$

folgt. Daraus leitete Mardia über Bilinearformen die obigen Ergebnisse (5.1) und (5.2) ab.

Es sei nun $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ eine Stichprobe aus i.i.d. p -Zufallsvektoren mit unbekannter Verteilung. Da der Erwartungswertvektor $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}_i] = (\mathbb{E}[X_{1i}], \dots, \mathbb{E}[X_{pi}])^T$ der Zufallsvektoren \mathbf{X}_i für $i = 1, \dots, n$ unbekannt ist, wird dieser im Folgenden geschätzt. Es sei $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_p)^T$ der *Stichprobenmittelwertvektor* mit $\bar{X}_j = n^{-1} \sum_{i=1}^n X_{ji}$ für $j = 1, \dots, p$ als *Stichprobenmittel* des j -ten Merkmals. Ebenso wird Σ durch die *Stichprobenkovarianzmatrix* $\mathbf{S} = (s_{jk})_{1 \leq j,k \leq p}$ mit $s_{jk} = n^{-1} \sum_{i=1}^n (X_{ji} - \bar{X}_j)(X_{ki} - \bar{X}_k)$ geschätzt. Dann lässt sich das Stichprobenanalogon zu (5.1) und (5.2) wie folgt definieren.

Definition 5.1.2. Es gelten die getroffenen Annahmen. Dann bezeichnet

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left((\mathbf{X}_i - \bar{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}}) \right)^3 \quad (5.3)$$

die *multivariate Stichprobenschiefe* nach Mardia und

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^n \left((\mathbf{X}_i - \bar{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \right)^2 \quad (5.4)$$

die *multivariate Stichprobenwölbung* nach Mardia.

Bemerkung 5.1.3. Es seien \mathbf{X} und \mathbf{Y} unabhängige und identisch $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilte Zufallsvektoren. Dann gilt:

$$\beta_{1,p} = 0 \quad \text{und} \quad \beta_{2,p} = p(p+2).$$

Beweis. Es seien \mathbf{X} und \mathbf{Y} unabhängig und identisch $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilt. Da die Verteilung von $(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$ symmetrisch um $(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) = 0$ ist, folgt $\mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})^3] = 0$ und mit (5.1) die Behauptung der ersten Aussage. Für die zweite Aussage wird Lemma 2.2.8 benutzt. Es gilt $(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_p^2$, sodass mit (5.2) $\beta_{2,p} = \mathbb{E}[(\chi_p^2)^2] = p(p+2)$ folgt. \square

Weiter zu bemerken ist, dass die neuen Maße auch in der univariaten Situation $p = 1$ angewendet werden können. Es gilt dann $b_{1,1} = b_1$ und $b_{2,1} = b_2$.

5.2 Asymptotische Verteilung

Dieser Abschnitt befasst sich mit der asymptotischen Verteilung der von Mardia angegebenen multivariaten Maße. Diese kann dazu genutzt werden, Tests auf multivariate Normalverteilung herzuleiten, wie dies auch von Mardia getan wurde. Die asymptotische Verteilung dient dazu, die kritischen Werte der Tests zu vorgegebenem Signifikanzniveau α anhand der kritischen Werte der asymptotischen Verteilung zu bestimmen. Diese stimmen aufgrund der Asymptotik der Verteilung für wachsende Stichprobenumfänge mit steigender Genauigkeit überein. Damit ist das fehlende Wissen über die exakte Verteilung der Teststatistik nicht weiter von Relevanz, vorausgesetzt die Stichprobe ist groß genug. Dies ist allerdings in den meisten Anwendungen nicht der Fall. Für kleine Stichproben werden deshalb wie im univariaten Fall empirische Analysen durchführen und somit wenn auch nicht exakte, jedoch hinreichend genaue kritische Werte ermittelt. Auf diese auf Monte-Carlo-Simulationen basierenden Untersuchungen wird an späterer Stelle näher eingegangen. Das Ergebnis der asymptotischen Verteilung liefert der folgende

Satz 5.2.1. *Es seien $b_{1,p}$ und $b_{2,p}$ wie in (5.3) und (5.4) und $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ eine Stichprobe aus i.i.d. p -Zufallsvektoren mit $\mathbf{X}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ für $i = 1, \dots, n$. Weiter sei $B_{1,p} = \frac{n}{6} b_{1,p}$ und $B_{2,p} = \left(\frac{(n+3)(n+5)}{8p(p+2)(n-3)(n-p-1)(n-p+1)} \right)^{1/2} \left((n+1)b_{2,p} - p(p+2)(n-1) \right)$. Dann gilt:*

$$B_{1,p} \xrightarrow{D} \chi_{p(p+1)(p+2)/6}^2 \quad \text{und} \\ B_{2,p} \xrightarrow{D} \mathcal{N}(0, 1).$$

Bevor der Beweis des Satzes begonnen wird, ist die Einführung einer Verteilung notwendig, die als multivariate Verallgemeinerung der χ^2 -Verteilung angesehen werden kann.

Definition 5.2.2. Es sei $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ mit $\mathbf{X}_i \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$ für $i = 1, \dots, n$ und $\mathbf{M} = \mathbf{X}\mathbf{X}^T = \sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^T$. Dann besitzt \mathbf{M} eine *Wishart*-Verteilung mit Skalennatrix Σ und n Freiheitsgraden. Man schreibt $\mathbf{M} \sim \mathcal{W}_p(n, \Sigma)$.

Beweis des Satzes. Begonnen wird der Beweis des Satzes mit einem Lemma. Dieses liefert zusammen mit dem zentralen Grenzwertsatz die asymptotische Konvergenz von $B_{2,p}$. Danach wird die asymptotische Verteilung der multivariaten Schiefe $b_{1,p}$ bestimmt.

Lemma 5.2.3. *Es sei $b_{2,p}$ wie in (5.4). Dann gilt mit $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ und $\mathbf{X}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ für $i = 1, \dots, n$:*

$$\begin{aligned} \mathbb{E}[b_{2,p}] &= \frac{1}{n+1}p(p+2)(n-1) \quad \text{und} \\ \text{Var}[b_{2,p}] &= \frac{8p(p+2)(n-3)}{(n+1)^2(n+3)(n+5)}(n-p+1)(n-p-1). \end{aligned}$$

Beweis des Lemmas. Der Beweis basiert auf [Mar74], Kapitel 4. Es sei $(X_{1i}, \dots, X_{pi})^T = \mathbf{X}_i$ für $i = 1, \dots, n$ mit $\mathbf{X}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. Aufgrund der Invarianz von $b_{2,p}$ unter nichtsingulären Transformationen $\mathbf{X} \mapsto \mathbf{A}\mathbf{X} + \mathbf{b}$, die sich durch Einsetzen in (5.4) zeigen lässt, wird o.B.d.A. $\boldsymbol{\mu} = \mathbf{0}$ und $\Sigma = \mathbf{I}_p$ angenommen. Weiter sei

$$R_{ij} = (\mathbf{X}_i - \bar{\mathbf{X}})^T \mathbf{M}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}}) \quad (5.5)$$

mit

$$\mathbf{M} = n\mathbf{S} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T = \sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^T - n\bar{\mathbf{X}}\bar{\mathbf{X}}^T, \quad (5.6)$$

sodass

$$b_{2,p} = n \sum_{i=1}^n R_{ii}^2 \quad (5.7)$$

gilt. Die Berechnung der Momente von $b_{2,p}$ geschieht nun über die Momente von R_{ij} . Dazu wird $\mathbf{X}_{(r)} = (X_{r1}, \dots, X_{rn})^T$ als n -Vektor im Merkmal r betrachtet und $\mathbf{X}_{(r)}$ mittels einer orthogonalen Transformation zu $\boldsymbol{\zeta}_{(r)} = (\zeta_{r1}, \dots, \zeta_{rn})^T$ transformiert. Die zur Beschreibung dieser orthogonalen Abbildung benutzte $(n \times n)$ -Matrix hat die Gestalt

$$\mathbf{H} = \begin{pmatrix} d_2 & e_2 & 0 & \cdots & \cdots & 0 \\ d_3 & d_3 & e_3 & \ddots & & \vdots \\ d_4 & d_4 & d_4 & e_4 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ d_n & \cdots & \cdots & \cdots & d_n & e_n \\ n^{-1/2} & \cdots & \cdots & \cdots & \cdots & n^{-1/2} \end{pmatrix}$$

mit $d_j = (j(j-1))^{-1/2}$, $e_j = -(j-1)d_j$ und wird als verallgemeinerte *Helmert*-Matrix bezeichnet. Es sei $\zeta = (\zeta_{ri})_{1 \leq r \leq p, 1 \leq i \leq n}$ definiert durch $\zeta_{(r)} = \mathbf{H}\mathbf{X}_{(r)}$ für $r = 1, \dots, p$. Dann ist

$$\zeta^T = \mathbf{H}\mathbf{X}^T \quad (5.8)$$

und somit $\zeta\zeta^T = \mathbf{X}\mathbf{H}^T\mathbf{H}\mathbf{X}^T \stackrel{\mathbf{H} \text{ orthogonal}}{=} \mathbf{X}\mathbf{X}^T$. Es folgt

$$\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T = \mathbf{X}\mathbf{X}^T = \zeta\zeta^T = \sum_{i=1}^n \zeta_i \zeta_i^T. \quad (5.9)$$

Weiter folgt mit (5.8) und der Darstellung von \mathbf{H} , dass $\zeta_{rn} = n^{-1/2} \sum_{i=1}^n X_{ri}$, sodass $\zeta_n = n^{1/2} \bar{\mathbf{X}}$ und folglich

$$\zeta_n \zeta_n^T = n \bar{\mathbf{X}} \bar{\mathbf{X}}^T \quad (5.10)$$

gilt. Mit (5.6), (5.9) und (5.10) folgt

$$\mathbf{M} = \sum_{i=1}^n \zeta_i \zeta_i^T - \zeta_n \zeta_n^T = \sum_{i=1}^{n-1} \zeta_i \zeta_i^T. \quad (5.11)$$

Aus (5.8) kann auch eine Darstellung von \mathbf{X} in Abhängigkeit von ζ bestimmt werden. Es ist $\mathbf{X}^T = \mathbf{H}^{-1}\zeta^T$ und wegen der Orthogonalität von \mathbf{H} ist $\mathbf{X}^T = \mathbf{H}^T\zeta^T \Leftrightarrow \mathbf{X} = \zeta\mathbf{H}$. Mit der Darstellung von \mathbf{H} folgt somit

$$\begin{aligned} \mathbf{X}_i &= e_i \zeta_{i-1} + \sum_{k=i}^{n-1} d_{k+1} \zeta_k + n^{-\frac{1}{2}} \zeta_n \quad \text{für } i = 1, \dots, n-1 \\ \text{und } \mathbf{X}_n &= e_n \zeta_{n-1} + n^{-\frac{1}{2}} \zeta_n \end{aligned}$$

mit $\zeta_0 = 0$. Es ist $n^{-1/2} \zeta_n = \bar{\mathbf{X}}$. Der Kürze halber wird \mathbf{X}_i für $i = 1, \dots, n-1$ und \mathbf{X}_n als eine Gleichung der Form

$$\mathbf{X}_i - \bar{\mathbf{X}} = e_i \zeta_{i-1} + \sum_{k=i}^{n-1} d_{k+1} \zeta_k \quad \text{für } i = 1, \dots, n \quad (5.12)$$

geschrieben, wobei für $i = n$ gelte: $\sum_{k=n}^{n-1} d_{k+1} \zeta_k = 0$.

Zerlegt man \mathbf{M} derart, dass $\mathbf{M} = \mathbf{T}\mathbf{T}^T$ mit einer $(p \times p)$ -Matrix \mathbf{T} , dann folgt mit (5.5) und (5.12)

$$R_{ij} = \left(e_i \zeta_{i-1} + \sum_{k=i}^{n-1} d_{k+1} \zeta_k \right)^T \left(\mathbf{T}\mathbf{T}^T \right)^{-1} \left(e_j \zeta_{j-1} + \sum_{k=j}^{n-1} d_{k+1} \zeta_k \right)$$

$$\begin{aligned}
&= \left(e_i \boldsymbol{\zeta}_{i-1} + \sum_{k=i}^{n-1} d_{k+1} \boldsymbol{\zeta}_k \right)^T \left(\mathbf{T}^{-1} \right)^T \mathbf{T}^{-1} \left(e_j \boldsymbol{\zeta}_{j-1} + \sum_{k=j}^{n-1} d_{k+1} \boldsymbol{\zeta}_k \right) \\
&= \left(e_i \mathbf{T}^{-1} \boldsymbol{\zeta}_{i-1} + \sum_{k=i}^{n-1} d_{k+1} \mathbf{T}^{-1} \boldsymbol{\zeta}_k \right)^T \left(e_j \mathbf{T}^{-1} \boldsymbol{\zeta}_{j-1} + \sum_{k=j}^{n-1} d_{k+1} \mathbf{T}^{-1} \boldsymbol{\zeta}_k \right).
\end{aligned}$$

Definiert man $\mathbf{Z}_i = \mathbf{T}^{-1} \boldsymbol{\zeta}_i$ für $i = 1, \dots, n$, so ist

$$R_{ij} = \left(e_i \mathbf{Z}_{i-1} + \sum_{k=i}^{n-1} d_{k+1} \mathbf{Z}_k \right)^T \left(e_j \mathbf{Z}_{j-1} + \sum_{k=j}^{n-1} d_{k+1} \mathbf{Z}_k \right) \quad (5.13)$$

und die Momente von R_{ij} lassen sich anhand der Momente von \mathbf{Z}_i herleiten. Diese werden im Folgenden bestimmt. Grundlage dafür ist die Dichte der $(p \times k)$ -Matrix $\overset{(k)}{\mathbf{Z}} = (\mathbf{Z}_1, \dots, \mathbf{Z}_k)$, für deren Bestimmung auf ein Resultat von Khatri zurückgegriffen wird, auf dessen Beweis hier verzichtet werden soll. Der interessierte Leser sei auf [Kha59], S.1259, verwiesen. Das Resultat lautet wie folgt:

Hilfslemma 5.2.4. *Es seien \mathbf{Q} eine $(p \times p)$ -Matrix und \mathbf{V} eine $(p \times k)$ -Matrix. \mathbf{Q} und \mathbf{V} seien unabhängig verteilt mit $\mathbf{Q} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$ und $\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_k)$ mit $\mathbf{V}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ für $i = 1, \dots, k$. Weiter sei $\mathbf{U} = \mathbf{Q} + \mathbf{V}\mathbf{V}^T$ und \mathbf{R} eine $(p \times p)$ -Matrix, sodass $\mathbf{R}\mathbf{R}^T = \mathbf{U}$. Dann gilt:*

\mathbf{U} und $\mathbf{Z} = \mathbf{R}^{-1}\mathbf{V}$ sind unabhängig mit $\mathbf{U} \sim \mathcal{W}_p(n+k, \boldsymbol{\Sigma})$ und \mathbf{Z} hat eine Dichte der Form $f(\mathbf{Z}; n+k, k, p) = c(n+k, k, p) |\mathbf{I}_p - \mathbf{Z}\mathbf{Z}^T|^{\frac{1}{2}(n-p-1)}$ mit $c(n+k, k, p) = \pi^{-\frac{1}{2}pk} \prod_{j=1}^p \Gamma\left(\frac{n+k-j+1}{2}\right) \left(\Gamma\left(\frac{n-j+1}{2}\right)\right)^{-1}$. Γ steht in diesem Zusammenhang für die Gammafunktion.

Es sei nun $\mathbf{Q} = (\boldsymbol{\zeta}_{k+1}, \dots, \boldsymbol{\zeta}_{n-1})(\boldsymbol{\zeta}_{k+1}, \dots, \boldsymbol{\zeta}_{n-1})^T$, dann ist $\mathbf{Q} \sim \mathcal{W}_p(n-k-1, \boldsymbol{\Sigma})$. Weiter sei $\mathbf{V} = (\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_k)$, dann gilt mit (5.11) $\mathbf{M} = \mathbf{Q} + \mathbf{V}\mathbf{V}^T$. Nach Hilfslemma 5.2.4 sind \mathbf{M} und $\overset{(k)}{\mathbf{Z}} = (\mathbf{Z}_1, \dots, \mathbf{Z}_k) = \mathbf{T}^{-1}\mathbf{V}$ unabhängig und ferner besitzt $\overset{(k)}{\mathbf{Z}}$ die Dichte

$$\begin{aligned}
f(\overset{(k)}{\mathbf{Z}}; n-1, k, p) &= c(n-1, k, p) |\mathbf{I}_p - \overset{(k)}{\mathbf{Z}} \overset{(k)}{\mathbf{Z}}^T|^{\frac{1}{2}(n-k-p-2)} \\
&= \pi^{-\frac{1}{2}pk} \prod_{j=1}^p \Gamma\left(\frac{n-j}{2}\right) \left(\Gamma\left(\frac{n-k-j}{2}\right)\right)^{-1} |\mathbf{I}_p - \overset{(k)}{\mathbf{Z}} \overset{(k)}{\mathbf{Z}}^T|^{\frac{1}{2}(n-k-p-2)}.
\end{aligned}$$

Diese wird im Folgenden benutzt, um die Momente von $b_{2,p}$ zu bestimmen. Zuvor jedoch werden einige nützliche Hilfslemmata aufgestellt. Dazu seien x_i und y_i für $i = 1, \dots, n$ reelle Zahlen. Die Beweise der Hilfslemmata benutzen das Prinzip der *augmented symmetric functions*, deren Herleitung an dieser Stelle im Interesse einer verständlichen Darstellung

überflüssig erscheint und infolgedessen vernachlässigt wird. Abkürzend steht *p.v.* für die Eigenschaft *paarweise verschieden*. Die notwendigen Definitionen zum Verständnis lauten

$$\begin{aligned} (r) &= \sum_{i=1}^n x_i^r \\ s_{pq} &= \sum_{i=1}^n x_i^p y_i^q \quad \text{und} \\ [(p_1 q_1)^{\pi_1}, \dots, (p_\lambda q_\lambda)^{\pi_\lambda}] &= \sum_{\substack{i_1, \dots, i_\nu = 1 \\ i_1, \dots, i_\nu \text{ p.v.}}}^n x_{i_1}^{p_1} y_{i_1}^{q_1} \dots x_{i_\nu}^{p_\lambda} y_{i_\nu}^{q_\lambda} \end{aligned}$$

mit $\nu = \sum_{i=1}^\lambda \pi_i$. Mit Hilfe der Tabellen 1·1·2, 1·6·4 und 1·6·8 in [DKB66] können die folgenden Resultate gewonnen werden.

Hilfslemma 5.2.5. *In der vorliegenden Situation gilt:*

$$\left(\sum_{k=i-1}^{n-1} c_k Z_{j,k} \right)^2 = \sum_{k=i-1}^{n-1} c_k^2 Z_{j,k}^2 + \sum_{\substack{k_1, k_2 = i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j,k_1} c_{k_2} Z_{j,k_2}. \quad (5.14)$$

Beweis des Hilfslemmas. Dies resultiert aus einer Darstellung der Form $s_{11}^2 = [22] + [11, 11]$. \square

Hilfslemma 5.2.6. *In der vorliegenden Situation gilt:*

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{j=1}^p \sum_{k=i-1}^{n-1} c_k^2 Z_{j,k}^2 \right)^2 \right] &= \sum_{j=1}^p \left(\sum_{k=i-1}^{n-1} c_k^4 \mathbb{E}[Z_{j,k}^4] + \sum_{\substack{k_1, k_2 = i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2 \mathbb{E}[Z_{j,k_1}^2 Z_{j,k_2}^2] \right) \\ &+ \sum_{\substack{j_1, j_2 = 1 \\ j_1 \neq j_2}}^p \left(\sum_{k=i-1}^{n-1} c_k^4 \mathbb{E}[Z_{j_1,k}^2 Z_{j_2,k}^2] + \sum_{\substack{k_1, k_2 = i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2 \mathbb{E}[Z_{j_1,k_1}^2 Z_{j_2,k_2}^2] \right). \end{aligned} \quad (5.15)$$

Beweis des Hilfslemmas. (5.15) ist das Ergebnis einer Zerlegung der Form $(1)^2 = [2] + [1^2]$ und Einsetzen der Gleichungen

$$\left(\sum_{k=i-1}^{n-1} c_k^2 Z_{j,k}^2 \right)^2 = \sum_{k=i-1}^{n-1} c_k^4 Z_{j,k}^4 + \sum_{\substack{k_1, k_2 = i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 Z_{j,k_1}^2 c_{k_2}^2 Z_{j,k_2}^2 \quad (5.16)$$

und

$$\left(\sum_{k=i-1}^{n-1} c_k^2 Z_{j_1,k}^2 \right) \left(\sum_{k=i-1}^{n-1} c_k^2 Z_{j_2,k}^2 \right) = \sum_{k=i-1}^{n-1} c_k^4 Z_{j_1,k}^2 Z_{j_2,k}^2 + \sum_{\substack{k_1, k_2 = i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 Z_{j_1,k_1}^2 c_{k_2}^2 Z_{j_2,k_2}^2. \quad (5.17)$$

(5.16) und (5.17) resultieren aus der Darstellung $s_{22}^2 = [44] + [22, 22]$, wobei in (5.17) darauf zu achten ist, dass $Z_{j_1,k}$ und $Z_{j_2,k}$ wegen $j_1 \neq j_2$ in verschiedenen Merkmalen vorliegen. \square

Hilfslemma 5.2.7. *In der vorliegenden Situation gilt:*

$$\begin{aligned}
& \mathbb{E} \left[\left(\sum_{j=1}^p \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j, k_1} c_{k_2} Z_{j, k_2} \right)^2 \right] \\
&= \sum_{j=1}^p \left(2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2 \mathbb{E}[Z_{j, k_1}^2 Z_{j, k_2}^2] + 4 \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 c_{k_2} c_{k_3} \mathbb{E}[Z_{j, k_1}^2 Z_{j, k_2} Z_{j, k_3}] \right. \\
&\quad \left. + \sum_{\substack{k_1, k_2, k_3, k_4=i-1 \\ k_1, k_2, k_3, k_4 \text{ p.v.}}}^{n-1} c_{k_1} c_{k_2} c_{k_3} c_{k_4} \mathbb{E}[Z_{j, k_1} Z_{j, k_2} Z_{j, k_3} Z_{j, k_4}] \right) \\
&\quad + \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^p \left(2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2 \mathbb{E}[Z_{j_1, k_1} Z_{j_2, k_1} Z_{j_1, k_2} Z_{j_2, k_2}] \right. \\
&\quad \left. + 4 \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 c_{k_2} c_{k_3} \mathbb{E}[Z_{j_1, k_1} Z_{j_2, k_1} Z_{j_1, k_2} Z_{j_2, k_3}] \right. \\
&\quad \left. + \sum_{\substack{k_1, k_2, k_3, k_4=i-1 \\ k_1, k_2, k_3, k_4 \text{ p.v.}}}^{n-1} c_{k_1} c_{k_2} c_{k_3} c_{k_4} \mathbb{E}[Z_{j_1, k_1} Z_{j_1, k_2} Z_{j_2, k_3} Z_{j_2, k_4}] \right). \tag{5.18}
\end{aligned}$$

Beweis des Hilfslemmas. (5.18) ist das Ergebnis einer Zerlegung der Form $(1)^2 = [2] + [1^2]$ und Einsetzen der Gleichungen

$$\begin{aligned}
& \left(\sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j, k_1} c_{k_2} Z_{j, k_2} \right)^2 \\
&= 2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 Z_{j, k_1}^2 c_{k_2}^2 Z_{j, k_2}^2 + 4 \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 Z_{j, k_1}^2 c_{k_2} Z_{j, k_2} c_{k_3} Z_{j, k_3} \\
&\quad + \sum_{\substack{k_1, k_2, k_3, k_4=i-1 \\ k_1, k_2, k_3, k_4 \text{ p.v.}}}^{n-1} c_{k_1} Z_{j, k_1} c_{k_2} Z_{j, k_2} c_{k_3} Z_{j, k_3} c_{k_4} Z_{j, k_4} \tag{5.19}
\end{aligned}$$

und

$$\begin{aligned}
& \left(\sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j_1, k_1} c_{k_2} Z_{j_1, k_2} \right) \left(\sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j_2, k_1} c_{k_2} Z_{j_2, k_2} \right) \\
&= 2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 Z_{j_1, k_1} Z_{j_2, k_1} c_{k_2}^2 Z_{j_1, k_2} Z_{j_2, k_2} + 4 \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 Z_{j_1, k_1} Z_{j_2, k_1} c_{k_2} Z_{j_1, k_2} c_{k_3} Z_{j_2, k_3} \\
&\quad + \sum_{\substack{k_1, k_2, k_3, k_4=i-1 \\ k_1, k_2, k_3, k_4 \text{ p.v.}}}^{n-1} c_{k_1} Z_{j_1, k_1} c_{k_2} Z_{j_1, k_2} c_{k_3} Z_{j_2, k_3} c_{k_4} Z_{j_2, k_4}. \tag{5.20}
\end{aligned}$$

(5.19) und (5.20) ergeben sich mit Hilfe der Gleichungen $[(11)^2]^2 \stackrel{1 \cdot 6 \cdot 4}{=} (-s_{22} + s_{11}^2)^2 = s_{22}^2 - 2s_{22}s_{11}^2 + s_{11}^4 \stackrel{1 \cdot 6 \cdot 8}{=} 2[(22)^2] + 4[22, (11)^2] + [(11)^4]$. Hierbei ist darauf zu achten, dass Z in Gleichung (5.20) in zwei Merkmalen auftritt, sodass c , Z_{j_1} und Z_{j_2} drei Variablen darstellen. \square

Hilfslemma 5.2.8. *In der vorliegenden Situation gilt:*

$$\begin{aligned}
& 2\mathbb{E}\left[\left(\sum_{j=1}^p \sum_{k=i-1}^{n-1} c_k^2 Z_{j,k}^2\right) \left(\sum_{j=1}^p \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j,k_1} c_{k_2} Z_{j,k_2}\right)\right] \\
&= 2 \sum_{j=1}^p \left(2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^3 c_{k_2} \mathbb{E}[Z_{j,k_1}^3 Z_{j,k_2}] + \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 c_{k_2} c_{k_3} \mathbb{E}[Z_{j,k_1}^2 Z_{j,k_2} Z_{j,k_3}] \right) \\
&+ 2 \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^p \left(2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^3 c_{k_2} \mathbb{E}[Z_{j_1, k_1}^2 Z_{j_2, k_1} Z_{j_2, k_2}] + \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 c_{k_2} c_{k_3} \mathbb{E}[Z_{j_1, k_1}^2 Z_{j_2, k_2} Z_{j_2, k_3}] \right). \tag{5.21}
\end{aligned}$$

Beweis des Hilfslemmas. (5.21) ergibt sich durch Zerlegung der Form $(1)(1) = [2] + [1^2]$ und den Gleichungen

$$\begin{aligned}
& \left(\sum_{k=i-1}^{n-1} c_k^2 Z_{j,k}^2\right) \left(\sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j,k_1} c_{k_2} Z_{j,k_2}\right) \\
&= 2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^3 Z_{j,k_1}^3 c_{k_2} Z_{j,k_2} + \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 Z_{j,k_1}^2 c_{k_2} Z_{j,k_2} c_{k_3} Z_{j,k_3} \tag{5.22}
\end{aligned}$$

und

$$\begin{aligned}
& \left(\sum_{k=i-1}^{n-1} c_k^2 Z_{j_1, k}^2\right) \left(\sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j_2, k_1} c_{k_2} Z_{j_2, k_2}\right) \\
&= 2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^3 Z_{j_1, k_1}^2 Z_{j_2, k_1} c_{k_2} Z_{j_2, k_2} + \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 Z_{j_1, k_1}^2 c_{k_2} Z_{j_2, k_2} c_{k_3} Z_{j_2, k_3}. \tag{5.23}
\end{aligned}$$

(5.22) und (5.23) wiederum folgen als Lösungen der Gleichung $[22][(11)^2] \stackrel{1 \cdot 6 \cdot 4}{=} -s_{22}^2 + s_{22}s_{11}^2 = 2[33, 11] + [22, (11)^2]$ und wiederholter Beachtung, dass c , Z_{j_1} und Z_{j_2} unterschiedliche Variablen sind. \square

Es wird mit den Momenten von $b_{2,p}$ fortgefahren. Dazu sei $c_k = e_i \mathbb{1}_{\{k=i-1\}} + d_{k+1} \mathbb{1}_{\{k=i, \dots, n-1\}}$. Mit $\mathbb{E}[Z_{j,k}^v] = 0$ für v ungerade und der Unabhängigkeit der Z_k für $k = 1, \dots, n$ folgt

$$\begin{aligned}
\mathbb{E}[b_{2,p}] &\stackrel{(5.7)}{=} \mathbb{E}\left[n \sum_{i=1}^n R_{ii}^2\right] \\
&\stackrel{(5.13)}{=} n \sum_{i=1}^n \mathbb{E}\left[\left(\sum_{j=1}^p \left(\sum_{k=i-1}^{n-1} c_k Z_{j,k}\right)^2\right)^2\right] \\
&\stackrel{(5.14)}{=} n \sum_{i=1}^n \mathbb{E}\left[\left(\sum_{j=1}^p \sum_{k=i-1}^{n-1} c_k^2 Z_{j,k}^2 + \sum_{j=1}^p \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j,k_1} c_{k_2} Z_{j,k_2}\right)^2\right] \\
&= n \sum_{i=1}^n \left(\mathbb{E}\left[\left(\sum_{j=1}^p \sum_{k=i-1}^{n-1} c_k^2 Z_{j,k}^2\right)^2\right] + \mathbb{E}\left[\left(\sum_{j=1}^p \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j,k_1} c_{k_2} Z_{j,k_2}\right)^2\right] \right. \\
&\quad \left. + 2 \mathbb{E}\left[\left(\sum_{j=1}^p \sum_{k=i-1}^{n-1} c_k^2 Z_{j,k}^2\right) \left(\sum_{j=1}^p \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1} Z_{j,k_1} c_{k_2} Z_{j,k_2}\right)\right] \right) \\
&\stackrel{(5.15), (5.18), (5.21)}{=} n \sum_{i=1}^n \left(\sum_{j=1}^p \left(\sum_{k=i-1}^{n-1} c_k^4 \mathbb{E}[Z_{j,k}^4] + \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2 \mathbb{E}[Z_{j,k_1}^2 Z_{j,k_2}^2] \right) \right. \\
&\quad + \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^p \left(\sum_{k=i-1}^{n-1} c_k^4 \mathbb{E}[Z_{j_1,k}^2 Z_{j_2,k}^2] + \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2 \mathbb{E}[Z_{j_1,k_1}^2 Z_{j_2,k_2}^2] \right) \\
&\quad + \sum_{j=1}^p \left(2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2 \mathbb{E}[Z_{j,k_1}^2 Z_{j,k_2}^2] + 4 \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 c_{k_2} c_{k_3} \mathbb{E}[Z_{j,k_1}^2 Z_{j,k_2} Z_{j,k_3}] \right. \\
&\quad \left. + \sum_{\substack{k_1, k_2, k_3, k_4=i-1 \\ k_1, k_2, k_3, k_4 \text{ p.v.}}}^{n-1} c_{k_1} c_{k_2} c_{k_3} c_{k_4} \mathbb{E}[Z_{j,k_1} Z_{j,k_2} Z_{j,k_3} Z_{j,k_4}] \right) \\
&\quad + \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^p \left(2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2 \mathbb{E}[Z_{j_1, k_1} Z_{j_2, k_1} Z_{j_1, k_2} Z_{j_2, k_2}] \right. \\
&\quad + 4 \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 c_{k_2} c_{k_3} \mathbb{E}[Z_{j_1, k_1} Z_{j_2, k_1} Z_{j_1, k_2} Z_{j_2, k_3}] \\
&\quad \left. + \sum_{\substack{k_1, k_2, k_3, k_4=i-1 \\ k_1, k_2, k_3, k_4 \text{ p.v.}}}^{n-1} c_{k_1} c_{k_2} c_{k_3} c_{k_4} \mathbb{E}[Z_{j_1, k_1} Z_{j_1, k_2} Z_{j_2, k_3} Z_{j_2, k_4}] \right) \\
&\quad + 2 \sum_{j=1}^p \left(2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^3 c_{k_2} \mathbb{E}[Z_{j, k_1}^3 Z_{j, k_2}] + \sum_{\substack{k_1, k_2, k_3=i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 c_{k_2} c_{k_3} \mathbb{E}[Z_{j, k_1}^2 Z_{j, k_2} Z_{j, k_3}] \right) \\
&\quad + 2 \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^p \left(2 \sum_{\substack{k_1, k_2=i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^3 c_{k_2} \mathbb{E}[Z_{j_1, k_1}^2 Z_{j_2, k_1} Z_{j_2, k_2}] \right.
\end{aligned}$$

$$+ \sum_{\substack{k_1, k_2, k_3 = i-1 \\ k_1, k_2, k_3 \text{ p.v.}}}^{n-1} c_{k_1}^2 c_{k_2} c_{k_3} \mathbb{E} \left[Z_{j_1, k_1}^2 Z_{j_2, k_2} Z_{j_2, k_3} \right] \Bigg) \Bigg).$$

Ein weiteres Ergebnis von Khatri und Pillai ist hilfreich, um die Gleichung zu vereinfachen.

Hilfslemma 5.2.9. *Es sei \mathbf{M} eine $(p \times k)$ -Matrix bestehend aus Zufallsvariablen M_{ji} für $j = 1, \dots, p, i = 1, \dots, k$. \mathbf{M} besitze die Dichte $g(\mathbf{M}; f, k, p) = c(f, k, p) |\mathbf{I}_p - \mathbf{M}\mathbf{M}^T|^{-\frac{1}{2}(f-p-k-1)}$ mit $c(f, k, p) = \pi^{-\frac{1}{2}pk} \prod_{j=1}^p \Gamma\left(\frac{f-j+1}{2}\right) \left(\Gamma\left(\frac{f-k-j+1}{2}\right)\right)^{-1}$. Dann gilt:*

$$\begin{aligned} \mathbb{E}[M_{ji}^2] &= \frac{1}{f} && \text{für } j = 1, \dots, p, i = 1, \dots, k, \\ \mathbb{E}[M_{j_1 i_1}^2 M_{j_2 i_2}^2] &= \frac{3}{f(f+2)} && \text{für } j_1 = j_2, i_1 = i_2, \\ &= \frac{1}{f(f+2)} && \text{für } j_1 = j_2, i_1 \neq i_2 \text{ oder } j_1 \neq j_2, i_1 = i_2, \\ &= \frac{f+1}{(f-1)f(f+2)} && \text{für } j_1 \neq j_2, i_1 \neq i_2. \end{aligned}$$

Der Beweis des Hilfslemmas basiert auf einer Zerlegung der multivariaten Dichte von \mathbf{M} in ein Produkt univariater Beta-Verteilungen, soll hier aber nicht näher ausgeführt werden. Für Details wird auf [KP66], S.149ff. verwiesen.

Hilfslemma 5.2.10. *Es sei $c_k = e_i \mathbf{1}_{\{k=i-1\}} + d_{k+1} \mathbf{1}_{\{k=i, \dots, n-1\}}$, $d_j = (j(j-1))^{-1/2}$, $e_j = -(j-1)d_j$. Dann gilt:*

$$\sum_{k=i-1}^{n-1} c_k^2 = \frac{n-1}{n}.$$

Beweis des Hilfslemmas.

$$\sum_{k=i-1}^{n-1} c_k^2 = c_{i-1}^2 + c_i^2 + \dots + c_{n-1}^2 = e_i^2 + d_{i+1}^2 + \dots + d_n^2 \stackrel{\mathbf{H} \text{ orth.}}{=} 1 - (n^{-1/2})^2 = \frac{n-1}{n}.$$

□

Es folgt nun mit

$$\sum_{k=i-1}^{n-1} c_k^4 = \left(\sum_{k=i-1}^{n-1} c_k^2 \right)^2 - \sum_{\substack{k_1, k_2 = i-1 \\ k_1 \neq k_2}}^{n-1} c_{k_1}^2 c_{k_2}^2$$

und Hilfslemma 5.2.9 sowie mit Hilfslemma 5.2.10 unter weiterem Rechenaufwand schließlich die erste Behauptung des Lemmas 5.2.3. Die Varianz der Größe $b_{2,p}$ kann auf dieselbe Weise berechnet werden, wobei die Beziehung $\text{Var}[b_{2,p}] = \mathbb{E}[b_{2,p}^2] - \mathbb{E}[b_{2,p}]^2$ ausgenutzt werden kann. □

Die Vorgehensweise im zweiten Teil des Beweises von Satz 5.2.1, also die Bestimmung der asymptotischen Verteilung von $B_{1,p}$ orientiert sich an [Mar70], Abschnitt 2.5. Dazu sei $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ eine Stichprobe aus i.i.d. p -Zufallsvektoren mit $\mathbf{X}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ für $i = 1, \dots, n$. Aufgrund der Invarianz von $b_{1,p}$ gegenüber linearen Transformationen kann o.B.d.A. $\boldsymbol{\mu} = \mathbf{0}$ und $\boldsymbol{\Sigma} = \mathbf{I}_p$ angenommen werden. Es lässt sich $b_{1,p}$ aus (5.3) ebenfalls darstellen in der Form

$$b_{1,p} = \sum_{r,s,t=1}^p \sum_{r',s',t'=1}^p S^{rr'} S^{ss'} S^{tt'} M_{111}^{(rst)} M_{111}^{(r's't')}$$

mit $\mathbf{S}^{-1} = (S^{ij})_{1 \leq i,j \leq p}$ und $M_{r_1, \dots, r_k}^{(s_1, \dots, s_k)} = n^{-1} \sum_{i=1}^n (\prod_{j=1}^k (X_{s_j i} - \bar{X}_{s_j})^{r_j})$. Aufgrund der Konvergenz in Wahrscheinlichkeit von \mathbf{S} gegen $\boldsymbol{\Sigma} = \mathbf{I}_p$ und infolgedessen $S^{ij} \xrightarrow{n \rightarrow \infty} 0$ für $i \neq j$ folgt

$$b_{1,p} \xrightarrow{n \rightarrow \infty} \sum_{r,s,t=1}^p \left(M_{111}^{(rst)} \right)^2$$

in Wahrscheinlichkeit. Weiter ist

$$\begin{aligned} \sum_{r,s,t=1}^p \left(M_{111}^{(rst)} \right)^2 &= \sum_{\substack{r,s,t=1 \\ r=s=t}}^p \left(M_{111}^{(rst)} \right)^2 + \sum_{\substack{r,s,t=1 \\ r=s \neq t}}^p \left(M_{111}^{(rst)} \right)^2 + \sum_{\substack{r,s,t=1 \\ r=t \neq s}}^p \left(M_{111}^{(rst)} \right)^2 \\ &\quad + \sum_{\substack{r,s,t=1 \\ r \neq s=t}}^p \left(M_{111}^{(rst)} \right)^2 + \sum_{\substack{r,s,t=1 \\ r,s,t \text{ p.v.}}}^p \left(M_{111}^{(rst)} \right)^2 \\ &= \left(M_3^{(1)} \right)^2 + \dots + \left(M_3^{(p)} \right)^2 + 3 \left(M_{21}^{(12)} \right)^2 + \dots + 3 \left(M_{21}^{((p-1)p)} \right)^2 \\ &\quad + 6 \left(M_{111}^{(123)} \right)^2 + \dots + 6 \left(M_{111}^{((p-2)(p-1)p)} \right)^2. \end{aligned}$$

Nun wird der Vektor

$$\mathbf{m} = \left(M_3^{(1)}, \dots, M_3^{(p)}, M_{21}^{(12)}, \dots, M_{21}^{((p-1)p)}, M_{111}^{(123)}, \dots, M_{111}^{((p-2)(p-1)p)} \right)$$

betrachtet, der aus $p + p(p-1) + p(p-1)(p-2)/6 = p(p+1)(p+2)/6$ Einträgen besteht. Da zur Ordnung n^{-1} gilt¹

$$\begin{aligned} \mathbb{E}[M_{111}^{(rst)}] &= 0, \quad \text{Var}[M_3^{(1)}] = \frac{6}{n}, \quad \text{Var}[M_{21}^{(12)}] = \frac{2}{n}, \quad \text{Var}[M_{111}^{(123)}] = \frac{1}{n} \\ \text{und} \quad \text{Cov}[M_{111}^{(rst)}, M_{111}^{(r's't')}] &= 0 \quad \text{für} \quad (r, s, t) \neq (r', s', t'), \end{aligned}$$

folgt mit der Normalverteilung des Vektors \mathbf{m} und Lemma 2.2.8 schließlich $nb_{1,p}/6 = B_{1,p} \sim \chi_{p(p+1)(p+2)/6}^2$. Dies beendet den Beweis des Satzes 5.2.1. \square

¹Vgl. [Mar70].

Bemerkung 5.2.11. Die hier vorliegende Varianz der multivariaten Wölbung $b_{2,p}$ ist exakt. Mardia gab in [Mar70] eine weitere Darstellung für die Varianz an, die zwar eine einfacherere Struktur besitzt, allerdings nur zur Ordnung n^{-1} gültig ist. Diese lautet

$$\text{Var}[b_{2,p}] = \frac{8p(p+2)}{n}.$$

Mit dem zentralen Grenzwertsatz folgt weiter die asymptotische Standardnormalverteilung des Ausdrucks

$$\tilde{B}_{2,p} = \left(\frac{8p(p+2)}{n} \right)^{-\frac{1}{2}} (b_{2,p} - \beta_{2,p}).$$

Obwohl die asymptotischen Verteilungen von $B_{2,p}$ und $\tilde{B}_{2,p}$ übereinstimmen, sind die beiden Größen für endliches n nicht identisch. Um zu überprüfen welche der beiden Statistiken zum Testen auf Normalverteilung besser geeignet ist, müssen die kritischen Werte empirisch berechnet und die Power der beiden Statistiken miteinander verglichen werden.

5.3 Multivariate Teststatistiken

Aus den im vorigen Abschnitt gewonnenen Resultaten lassen sich mehrere Teststatistiken ableiten. Je nachdem ob die exakte oder asymptotische Varianz des multivariaten Wölbungsmaßes $b_{2,p}$ eingesetzt wird, sind dies $B_{1,p}$, $B_{2,p}$ und $\tilde{B}_{2,p}$. Mardias Test sieht dabei vor, dass die Nullhypothese der Normalverteilung der der Stichprobe zugrunde liegenden Grundgesamtheit abgelehnt wird, falls die Werte der Statistik $B_{1,p}$ oder der Statistik $B_{2,p}$ an der Stelle der Stichprobe größer als die zugehörigen kritischen Werte sind. Alternativ wähle man $\tilde{B}_{2,p}$ anstelle von $B_{2,p}$. Zu vorgegebenem Signifikanzniveau lassen sich die kritischen Werte dabei anhand der asymptotischen Verteilungen bestimmen. Diese Art des Testens auf Normalverteilung impliziert also, dass keine Normalverteilung vorliegt, falls die empirische Schiefe oder die empirische Wölbung zu stark von der theoretischen Schiefe oder der Wölbung einer Normalverteilung abweicht. Der Test beruht also genau genommen auf zwei Teststatistiken. Ein Test, welcher die beiden multivariaten Maße $b_{1,p}$ und $b_{2,p}$ in einer Teststatistik zusammenfasst, lässt sich ebenfalls ableiten.

Satz 5.3.1. *Es gelten die Bezeichnungen und Voraussetzungen des vorigen Abschnittes. Dann können*

$$MJB = B_{1,p} + (B_{2,p})^2 \quad \text{und} \quad \widetilde{MJB} = B_{1,p} + (\tilde{B}_{2,p})^2$$

als multivariate Jarque-Bera-Statistiken angesehen werden. Da $B_{2,p}$ nach Satz 5.2.1 eine asymptotische Standardnormalverteilung besitzt, haben sowohl MJB als auch \widetilde{MJB} eine asymptotische χ^2 -Verteilung mit $\frac{p(p+1)(p+2)}{6} + 1$ Freiheitsgraden.²

²Vgl. [KOS09].

Es besteht somit auch im Testen auf multivariate Normalverteilung die Möglichkeit, zu gegebenem Signifikanzniveau die kritischen Werte anhand der asymptotischen Verteilungen der Statistiken $b_{1,p}$ und $b_{2,p}$ zu bestimmen. Für kleine Stichproben jedoch resultieren wie auch im univariaten Fall nur sehr mangelhafte Ergebnisse. Deshalb hat Mardia in [Mar74] kritische Werte für seinen Test speziell im Fall $p = 2$ und für einige ausgewählte Stichprobengrößen n für jeweils $b_{1,2}$ und $b_{2,2}$ angegeben.

Da diese Berechnungen bereits für wenige Merkmale und kleine Stichprobengrößen sehr aufwendig sind, wird auf erneute Berechnungen verzichtet und nur die empirischen kritischen Werte der simultan testenden Teststatistiken MJB und \widetilde{MJB} für $p = 4$ Merkmale und ausgewählte Stichprobengrößen n sowie Signifikanzniveaus α ermittelt. Die Vorgehensweise entspricht dabei der im univariaten Fall, jedoch müssen diesmal 4-variater normalverteilter Stichproben generiert werden. Dazu wurde erneut die Programmiersprache R benutzt und unter Einbeziehung des Paketes „MASS“ Replikationen multivariat normalverteilter Stichproben unterschiedlicher Größenordnung erzeugt. Um die Statistiken MJB und \widetilde{MJB} implementieren zu können, mussten vorab der Stichprobenmittelwertvektor und die empirische Kovarianzmatrix implementiert werden. Danach konnten die multivariaten Schiefe- und Wölbungsmaße $b_{1,p}$ und $b_{2,p}$, sowie die darauf basierenden Größen $B_{1,p}$, $B_{2,p}$ und $\tilde{B}_{2,p}$ als Funktionen programmiert werden. Die kritischen Werte zu vorgegebenem Signifikanzniveau α wurden dann nach Anwendung der multivariaten Teststatistiken auf die erzeugten Stichproben anhand der $(1-\alpha) \cdot (\# \text{Replikationen})$ -größten Werte bestimmt. Die Tabelle 5.1 gibt die ermittelten Werte an. Die Programmcodes der einzelnen Funktionen finden sich auf der beiliegenden Daten-CD im Einband am Ende der Arbeit.

	$p = 4$						
	$n = 10$		$n = 50$		$n = 150$		$n \rightarrow \infty$
	MJB	\widetilde{MJB}	MJB	\widetilde{MJB}	MJB	\widetilde{MJB}	MJB, \widetilde{MJB}
$\alpha = 0.01$	29.51	23.08	43.75	39.91	42.76	40.86	38.93
$\alpha = 0.05$	23.15	20.30	32.86	31.18	34.16	33.40	32.67
$\alpha = 0.1$	20.22	18.91	28.62	27.51	29.93	29.60	29.62

Tabelle 5.1: Empirische kritische Werte der multivariaten Jarque-Bera-Statistiken zu einigen Stichprobengrößen n und Signifikanzniveaus α bei $p = 4$ Merkmalen basierend auf 10^4 Replikationen.

Bemerkung 5.3.2. Speziell im Fall $p = 1$ stimmt \widetilde{MJB} mit dem JB -Test aus (3.8) überein.

Kapitel 6

Anwendung

6.1 Testen der Rendite des DAX auf Normalverteilung

Im vorliegenden Abschnitt wird die Rendite des deutschen Aktienindex (DAX) betrachtet und diese, sowie die Rendite einiger ausgewählter DAX-Aktien auf (multivariate) Normalverteilung getestet. Dazu muß im Vorfeld geklärt werden, was genau unter dem Begriff *Rendite* verstanden wird. Als Wertzuwachs bei positiver, bzw. Wertverlust bei negativer Rendite wird die diskrete Rendite einer Aktie A über eine Zeitspanne der Länge eins, d.h. im Zeitraum $(n-1, n]$ mit $n \in \mathbb{N}$ zum Zeitpunkt n definiert durch

$$R(n) = \frac{A(n) - A(n-1)}{A(n-1)}. \quad (6.1)$$

Obwohl die betrachteten Aktien im Laufe eines Handelstages fortlaufenden Preisanpassungen unterliegen, werden nur die Schlußwerte der Handelstage betrachtet, sodass die Gleichung (6.1) in der folgenden Verwendung als Tagesrendite (R_{Tag}) interpretiert werden kann. Es gilt somit für die Tagesrendite des Handelstages t innerhalb des betrachteten Handelszeitraums $(0, T]$ mit T als letztem Handelstag

$$R_{Tag}(t) = \frac{A(t)}{A(t-1)} - 1, \quad (6.2)$$

sodass $R_{Tag}(t) \in [-1, \infty)$, da der Preis einer Aktie weder negativ werden kann, noch eine (zumindest theoretische) Beschränkung des Aktienpreises nach oben gegeben ist. Weiter werden auch die Wochen- (R_{Woche}), Monats- (R_{Monat}), Quartals- ($R_{Quartal}$) und Jahresrenditen (R_{Jahr}) betrachtet. Dabei können die Tagesrenditen jedoch nicht einfach aufsummiert werden, denn es gilt für einen Zeitraum der Länge s Tage (also $s \in \mathbb{N}$):

$$R((t, t+s]) = \frac{A(t+s) - A(t)}{A(t)} = \prod_{j=t+1}^{t+s} \left(\frac{A(j)}{A(j-1)} \right) - 1 = \prod_{j=t+1}^{t+s} (R_{Tag}(j) + 1) - 1.$$

Weiter zu beachten ist, dass an Wochenenden und Feiertagen nicht gehandelt wird. Somit entspricht T nicht der Anzahl an tatsächlichen Handelstagen innerhalb des Handelszeitraums $(0, T]$. Die Anzahl an Handelstagen ist vielmehr wesentlich kleiner. Also müssen unter Beachtung der Diskretheit des Modells gewisse Anpassungen getroffen und die Anzahl an Handelstagen möglichst gleichmäßig auf den betrachteten Handelszeitraum aufgeteilt werden. Betrachtet man einen Handelszeitraum der Größenordnung x Jahre, so entspricht dies $T = 365 \cdot x$ Tagen und in anderen Zeiteinheiten ausgedrückt $T = 52 \cdot x$ Wochen = $12 \cdot x$ Monate = $4 \cdot x$ Quartale = x Jahre. Es sei nun n die Anzahl an Handelstagen innerhalb dieser x Jahre und $k = 52 \cdot x$. Dann wird die Wochenrendite der i -ten Woche innerhalb des betrachteten Handelszeitraums $(0, T]$ als

$$R_{Woche}(i) = \prod_{j=\lfloor \frac{n(i-1)}{k} \rfloor + 1}^{\lfloor \frac{ni}{k} \rfloor} (R_{Tag}(j) + 1) - 1 \quad (6.3)$$

mit $i \in \{1, \dots, k\}$ definiert. Analog werden die Größen $R_{Monat}(i)$, $R_{Quartal}(i)$ und $R_{Jahr}(i)$ mit $k = 12 \cdot x$, $k = 4 \cdot x$ bzw. $k = x$ anstelle von $k = 52 \cdot x$ definiert.

Dieses Modell wird im Folgenden dazu benutzt, die Rendite verschiedener Laufzeiten auf Normalverteilung zu testen. Dazu wurde in der Programmiersprache R ein Programm erstellt, in dem die einzelnen Renditearten implementiert und auf historische DAX-Werte angewendet wurden. Die Analysen beschränken sich auf die DAX-Aktien der Unternehmen Allianz, Bayer, Deutsche Bank, SAP und Siemens, sowie den DAX-Index. Betrachtet wurde im Rahmen dieser Analysen der Zehnjahreszeitraum vom 2.1.2002 - 30.12.2011. Die Größe der vorliegenden Stichproben der einzelnen Renditearten ist $n \in [2552, 2580]$ für die Tagesrenditen, $n = 520$ für die Wochenrenditen, $n = 120$ für die Monatsrenditen, $n = 40$ für die Quartalsrenditen und $n = 10$ für die Jahresrenditen. Die historischen Daten stammen dabei von der Internetseite <http://finance.yahoo.com/>. Um mit den Daten arbeiten zu können wurden diese in R eingelesen und auf die (um die Dividende bereinigten) Aktienwerte zugegriffen. Unter Anwendung der in Kapitel 3 hergeleiteten JB -Statistik (die bereits in R implementiert wurde) auf die Aktienwerte konnten die in der Tabelle 6.1 aufgelisteten Werte ermittelt werden.

Zu erkennen ist in Tabelle 6.1, dass für den DAX und auch alle DAX-Aktien die Hypothese der Normalverteilung der Tages-, Wochen-, Monats- und Quartalsrenditen zu den üblichen Signifikanzniveaus $\alpha \in \{0.05, 0.1\}$ und sogar zum Niveau $\alpha = 0.005$ abgelehnt wird.¹ Ausschließlich für die Jahresrenditen des DAX, sowie die Jahresrenditen von SAP und Siemens kann die Nullhypothese der Normalverteilung zum Signifikanzniveau $\alpha = 0.05$, für die Jahresrendite von Bayer sogar nur zum Signifikanzniveau $\alpha = 0.01$ gehalten

¹Vgl. dazu Tabelle 4.1.

werden.

	$JB(R_{Tag})$	$JB(R_{Woche})$	$JB(R_{Monat})$	$JB(R_{Quartal})$	$JB(R_{Jahr})$
DAX	1877.25	376.23	34.42	25.27	2.29
Allianz	5116.19	2837.58	113.30	119.78	7.52
Bayer	53572.54	19066.24	73.92	64.76	2.64
Deutsche Bank	7054.97	901.04	5828.01	69.38	7.32
SAP	14322.55	1354.54	1246.65	92.80	2.42
Siemens	20712.39	1011.75	2010.90	25.20	2.13

Tabelle 6.1: Werte der JB -Statistik einiger ausgewählter DAX-Aktien und des DAX-Index für die Tages-, Wochen-, Monats-, Quartals- und Jahresrenditen vom 2.1.2002 - 30.12.2011.

Weiter wird überprüft ob die Jahresrenditen des DAX und die Jahresrenditen von Bayer, SAP und Siemens sogar als multivariat normalverteilt angenommen werden können. Dazu wurden die Jahresrenditen mit R in einer (4×10) -Datenmatrix \mathbf{X} zusammengefasst und auf diese die Statistiken MJB und \widetilde{MJB} aus Satz 5.3.1 angewendet, die bereits im Rahmen des vorigen Kapitels in R implementiert wurden. Zur Entscheidungsfindung werden die empirischen kritischen Werte herangezogen, die der Tabelle 5.1 entnommen werden können. Es ist $MJB(\mathbf{X}) = 29.41$ und $\widetilde{MJB}(\mathbf{X}) = 22.65$, sodass also auch hier die Annahme der multivariaten Normalverteilung der vorliegenden Jahresrenditen zum Niveau $\alpha = 0.01$ nicht verworfen werden kann. Zum Niveau $\alpha = 0.05$ jedoch kann die Nullhypothese bereits nicht mehr gehalten werden.

Fazit: Nach den Renditedefinitionen (6.2) und (6.3) muss davon ausgegangen werden, dass die Tages-, Wochen-, Monats-, und Quartalsrenditen des DAX und aller betrachteten DAX-Aktien weder multivariat, noch univariat normalverteilt sind. Für die Jahresrenditen des DAX, sowie für die Jahresrenditen von Bayer, SAP und Siemens ändert sich das Bild. Hier kann sogar eine multivariate Normalverteilung unterstellt werden. Dabei kann die Nullhypothese jedoch nur knapp zum Signifikanzniveau 1% gehalten werden. Bezüglich der Nullhypothese univariat normalverteilter Jahresrenditen kann für diese Aktienwerte keine einheitliche Entscheidung getroffen werden. Die Nullhypothesen wurden zwar nicht verworfen, allerdings mussten dazu die Signifikanzniveaus im Vorfeld unterschiedlich gewählt werden. Bei den Aktienwerten der Unternehmen Allianz und Deutsche Bank konnte die Nullhypothese zu keinem der gängigen Signifikanzniveaus $\alpha \in \{0.01, 0.05, 0.1\}$ gehalten werden. Insgesamt jedoch sollten aufgrund der geringen Stichprobengröße von lediglich

zehn Beobachtungen im Falle der Jahresrenditen, sowie aufgrund der geringen Anzahl an Replikationen, anhand derer die empirischen kritischen Werte der multivariaten Teststatistiken ermittelt wurden, die gewonnenen Erkenntnisse unter Vorbehalt betrachtet werden. Wie sich allerdings mit Sicherheit herausgestellt hat, müssen vor allem die Tagesrenditen als nicht normalverteilt angenommen werden. Um dennoch Aussagen über die den Tagesrenditen zugrunde liegende Verteilungsform treffen zu können sei beispielsweise auf [ST06] aufmerksam gemacht. In ihrem Buch geben die Autoren verschiedene Modelle zur Anpassung an die empirische Verteilung unterschiedlicher DAX-Aktien.

Kapitel 7

Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde eine Herleitung des Jarque-Bera-Tests gegeben und dabei insbesondere auf seine asymptotische Verteilung eingegangen. Es konnte gezeigt werden, dass sich der Jarque-Bera-Test als ein Spezialfall des Score-Tests innerhalb des Pearson-Verteilungssystems darstellen lässt, sodass der Jarque-Bera-Test innerhalb dieses Verteilungssystems ein Test auf Normalverteilung bildet. Anhand dieser Darstellung kann man erkennen, dass der Jarque-Bera-Test auf der Schiefe und der Wölbung der zugrunde liegenden Verteilung basiert und somit eine gewisse Abweichung dieser Größen von denen einer Normalverteilung notwendig ist, um eine Ablehnung der Hypothese der Normalverteilung herbeizuführen. Ob und in welchem Maße dies ein Vorteil des Jarque-Bera-Tests gegenüber anderen Tests auf Normalverteilung ist, wurde daraufhin in empirischen Untersuchungen analysiert. Dazu wurden, anders als es in praktischen Anwendungen häufig der Fall ist, nicht die kritischen Werte der asymptotischen Verteilung, sondern die empirischen kritischen Werte, die zum Teil selbst ermittelt wurden, benutzt. Denn obwohl sich die Grenzverteilung des Jarque-Bera-Tests als eine χ^2 -Verteilung mit zwei Freiheitsgraden herausstellte und die Verwendung der Fraktile dieser Verteilung aufgrund der guten Vertafelung wesentlich einfacher gewesen wäre, konnte gezeigt werden, dass aufgrund der langsamen Konvergenz in Verteilung die Verwendung der Grenzverteilung zur Beurteilung in Testsituationen besonders beim Vorliegen kleiner Stichproben oftmals zu falschen Ergebnissen führen wird. Nichtsdestotrotz werden die Fraktile der χ^2 -Verteilung heutzutage in vielen Anwendungen dazu benutzt, über das Verwerfen oder Einhalten der aufgestellten Nullhypothese der Normalverteilung zu entscheiden.

Um Aussagen über die Power des Jarque-Bera-Tests in Bezug zur Power anderer Tests treffen zu können, wurden mit dem Programm R Monte-Carlo-Simulationen durchgeführt. Zudem wurden zwei Modifizierungen des Jarque-Bera-Tests betrachtet, von denen die eine auf den exakten Momenten der Schiefe und der Wölbung und die andere auf unverzerrten Schätzern für die theoretische Schiefe und die Wölbung beruhte. Es stellte sich heraus,

dass in der Situation des Testens auf Normalverteilung von Beobachtungen gegen Verteilungen aus dem Pearson-Verteilungssystem die Jarque-Bera-Statistiken speziell gegen die t -Verteilung den anderen Tests überlegen waren. In den anderen Testsituationen lagen der Jarque-Bera-Test und seine Modifizierungen etwa im Mittelfeld, wobei für kleine Stichproben die Ergebnisse schlechter ausfielen als die der anderen Teststatistiken. Zusätzlich wurde auch gegen die *Lognormal*-Verteilung als Nicht-Pearson-Verteilung getestet, obwohl eine Rechtfertigung für dieses Vorgehen erst an späterer Stelle gegeben wird. In dieser Testsituation jedoch unterlag der Jarque-Bera-Test seiner Konkurrenz am deutlichsten. Im Verhältnis zu seinen Modifizierungen schnitt die ursprüngliche Jarque-Bera-Statistik abgesehen vom Testen gegen die t -Verteilung deutlich besser ab, sodass die modifizierten Jarque-Bera-Tests nicht als wirkliche Verbesserungen eingestuft werden können und infolgedessen in den anschließenden Untersuchungen nicht weiter berücksichtigt wurden.

Um nun auch gegen Nicht-Pearson-Verteilungen testen zu dürfen, wurden die Ergebnisse von Bowman und Shenton genutzt, die eine Charakterisierung des Jarque-Bera-Tests als Summe zweier standardnormalverteilter Zufallsvariablen gegeben haben. Dieses Resultat liefert gleichzeitig eine einfache Begründung für die angegebene Grenzverteilung. In ähnlicher Weise wie im vorigen Modell wurden neue Untersuchungen über die Power der Teststatistiken durchgeführt, wobei in der neuen Situation die Alternativhypothesen aus kontaminierten Normalverteilungen bestanden. Das Ergebnis war, dass sich bei einer Abweichung der Schiefe von null weniger Auswirkungen auf die Power der Jarque-Bera-Statistik zeigten, als bei einer Abweichung der für die Normalverteilung typischen Wölbung. Dies führte dazu, dass der Jarque-Bera-Test sowohl dem Anderson-Darling-, als auch dem Shapiro-Francia-Test fast immer unterlag, wobei in den Situationen, in denen nur eine starke Abweichung von der Wölbung zu beobachten war, diese Unterlegenheit teilweise ausgeglichen erschien. Besonders bimodale Verteilungsmodelle stellen eine große Herausforderung für den Jarque-Bera-Test dar, denn in diesen Situationen war die Unterlegenheit der Teststatistik für kleine Stichproben am offensichtlichsten.

Als Test auf multivariate Normalverteilung konnte der Jarque-Bera-Test auf multivariate Testsituationen erweitert werden, indem die von Kantilal Vardichand Mardia definierten multivariaten Schiefe- und Wölbungsmaße benutzt wurden. Auch hier konnte für die multivariaten Teststatistiken eine asymptotische Verteilung hergeleitet werden. In Analogie zur univariaten Situation wurden anhand von Monte-Carlo-Simulationen für den speziellen Fall von vier Merkmalen empirische kritische Werte ermittelt, mit dem Ergebnis, dass sich auch im mehrdimensionalen Fall speziell beim Vorliegen kleiner Stichproben die empirischen kritischen Werte von denen der Grenzverteilung deutlich unterscheiden. Benutzt man diese beispielsweise zur Beurteilung über eine multivariate Normalverteilung von Renditen verschiedener Laufzeiten an Aktienmärkten, so muß bei den betrachteten

DAX-Aktien davon ausgegangen werden, dass für die Jahresrenditen die Nullhypothese der multivariaten Normalverteilung zu einem deutlich niedrigeren Signifikanzniveau (5%) nicht gehalten werden kann, als dies bei Betrachtung der Fraktile der asymptotischen Verteilung der Fall ist. Hier kann selbst zu einem Signifikanzniveau von 10% die Nullhypothese nicht verworfen werden. Jedoch sollte an dieser Stelle darauf hingewiesen werden, dass die Größe der Stichprobe, anhand derer diese Ergebnisse resultieren, lediglich zehn Beobachtungen beträgt. Alle Renditen kürzerer Laufzeiten müssen generell auch bereits in der univariaten Situation zu jedem üblichen Signifikanzniveau als nicht normalverteilt angenommen werden.

Obwohl die empirischen Untersuchungen deutlich zeigen, dass die Verwendung der χ^2 -Fraktile andere Ergebnisse liefert als die Verwendung der empirischen kritischen Werte und ein Überblick darüber gewonnen werden konnte, wie die einzelnen Tests im Verhältnis zueinander einzuordnen sind, muß bedacht werden, dass diese Ergebnisse die relativen Verhältnisse lediglich andeuten. Um präzisere Aussagen treffen zu können, müssen in größerem Rahmen Simulationen durchgeführt werden.

Um eine Vorstellung darüber zu bekommen, wie die multivariaten Jarque-Bera-Statistiken im Verhältnis zu anderen Tests auf multivariate Normalverteilung einzuordnen sind, bietet es sich an, einen Powervergleich mit anderen Tests auf multivariate Normalverteilung durchzuführen. Problematisch ist jedoch die große Zahl an Rechenschritten innerhalb der Simulationen, die mit steigender Anzahl an Merkmalen bedeutend zunimmt. In der univariaten Situation könnte man versuchen den univariaten Jarque-Bera-Test dadurch zu verbessern, dass weitere Größen für die Schiefe und Wölbung als Basis einer Teststatistik betrachtet werden. Insgesamt jedoch lässt sich festhalten, dass der Jarque-Bera-Test als Test auf univariate Normalverteilung aufgrund seiner sehr einfachen Struktur durchaus ein leicht zu handhabendes und empfehlenswertes Hilfsmittel darstellt, sofern die empirischen kritischen Werte benutzt werden und die Stichprobe eine gewisse Mindestgröße besitzt.

Anhang A

Hilfsresultate

In diesem Kapitel werden einige Resultate angegeben, die im Laufe der Arbeit verwendet wurden.

A.1 Zentraler Grenzwertsatz (mehrdimensional)

Es seien $\mathbf{X}_1, \dots, \mathbf{X}_n$ i.i.d. p -Zufallsvektoren mit $\mathbb{E}[\mathbf{X}_i] = \boldsymbol{\mu}$ und $\text{Cov}[\mathbf{X}_i] = \boldsymbol{\Sigma}$ für $i = 1, \dots, n$. Weiter sei $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$. Dann gilt:

$$\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}).$$

A.2 Cholesky-Zerlegung

Eine symmetrische Matrix \mathbf{A} ist genau dann positiv definit, wenn es eine obere Dreiecks-

matrix $\mathbf{G} = \begin{pmatrix} g_{11} & \cdots & \cdots & g_{1d} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & g_{dd} \end{pmatrix}$ gibt, sodass $\mathbf{G}^T \mathbf{G} = \mathbf{A}$. Man nennt \mathbf{G} auch die

Matrix-Wurzel aus der Matrix \mathbf{A} . Die Cholesky-Zerlegung ist ein numerisches Verfahren, das eine solche Matrix \mathbf{G} berechnet.

A.3 Invertierung von Blockmatrizen

Gegeben sei die $(d \times d)$ -Matrix $\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}$ mit $\mathbf{M}_{11} \in \mathbb{R}^{r \times r}$, $\mathbf{M}_{12} \in \mathbb{R}^{r \times (d-r)}$, $\mathbf{M}_{21} \in \mathbb{R}^{(d-r) \times r}$ und $\mathbf{M}_{22} \in \mathbb{R}^{(d-r) \times (d-r)}$. Sind die Matrizen \mathbf{M} , \mathbf{M}_{11} und \mathbf{M}_{22} invertier-

bar, dann gilt mit $N = M_{11} - M_{12}M_{22}^{-1}M_{21}$:

$$M^{-1} = \begin{pmatrix} N^{-1} & -N^{-1}M_{12}M_{22}^{-1} \\ -M_{22}^{-1}M_{21}N^{-1} & M_{22}^{-1} + M_{22}^{-1}M_{21}N^{-1}M_{12}M_{22}^{-1} \end{pmatrix}.$$

Anhang B

Ergebnisse der empirischen Analysen

B.1 Numerische Ergebnisse der Analysen im Powervergleich der Teststatistiken JB , AD und SF innerhalb verschie- dener KN -Verteilungen

		$\sigma_1^2 = \sigma_2^2 = 1$								
		$p = 0.1$			$p = 0.3$			$p = 0.5$		
		JB	AD	SF	JB	AD	SF	JB	AD	SF
$\mu_2 = 2$	$n = 20$	0.119	0.088	0.122	0.045	0.075	0.065	0.013	0.044	0.023
	$n = 50$	0.188	0.149	0.191	0.051	0.13	0.088	0.003	0.079	0.029
	$n = 100$	0.262	0.218	0.271	0.073	0.231	0.16	0.007	0.135	0.043
	$n = 200$	0.466	0.396	0.448	0.248	0.492	0.361	0.064	0.244	0.109
	$n = 400$	0.76	0.683	0.772	0.642	0.821	0.725	0.37	0.563	0.375
$\mu_2 = 3$	$n = 20$	0.27	0.243	0.283	0.074	0.24	0.17	0.011	0.135	0.057
	$n = 50$	0.54	0.506	0.574	0.104	0.589	0.425	0.002	0.446	0.226
	$n = 100$	0.838	0.824	0.86	0.509	0.913	0.811	0.209	0.82	0.582
	$n = 200$	0.991	0.982	0.992	0.959	0.998	0.989	0.892	0.998	0.967
	$n = 400$	1	1	1	1	1	1	1	1	1

Tabelle B.1: Powerwerte der Teststatistiken JB , AD und SF zum Signifikanzniveau $\alpha = 5\%$ bei einigen Stichprobengrößen n und 1000 Replikationen für verschiedene KN -Verteilungen mit den Parametern $\mu_1 = 0$, $\mu_2 \in \{2, 3\}$, $\sigma_1^2 = \sigma_2^2 = 1$, und $p \in \{0.1, 0.3, 0.5\}$.

		$\sigma_2^2 = 0.5$								
		$p = 0.25$			$p = 0.5$			$p = 0.75$		
		<i>JB</i>	<i>AD</i>	<i>SF</i>	<i>JB</i>	<i>AD</i>	<i>SF</i>	<i>JB</i>	<i>AD</i>	<i>SF</i>
$\mu_2 = 1$	$n = 20$	0.059	0.084	0.07	0.198	0.224	0.235	0.348	0.335	0.378
	$n = 50$	0.072	0.16	0.126	0.382	0.529	0.495	0.729	0.683	0.741
	$n = 100$	0.148	0.308	0.263	0.714	0.862	0.837	0.952	0.946	0.96
	$n = 200$	0.355	0.596	0.495	0.954	0.99	0.985	0.999	0.998	0.999
	$n = 400$	0.725	0.908	0.856	1	1	1	1	1	1
$\mu_2 = 2$	$n = 20$	0.017	0.085	0.038	0.079	0.337	0.247	0.521	0.684	0.672
	$n = 50$	0.008	0.194	0.074	0.198	0.801	0.648	0.926	0.98	0.982
	$n = 100$	0.043	0.451	0.252	0.818	0.994	0.974	1	1	1
	$n = 200$	0.376	0.791	0.662	0.999	1	1	1	1	1
	$n = 400$	0.902	0.99	0.986	1	1	1	1	1	1

Tabelle B.2: Powerwerte der Teststatistiken *JB*, *AD* und *SF* zum Signifikanzniveau $\alpha = 5\%$ bei einigen Stichprobengrößen n und 1000 Replikationen für verschiedene *KN*-Verteilungen mit den Parametern $\mu_1 = 0$, $\mu_2 \in \{1, 2\}$, $\sigma_1^2 = 1$, $\sigma_2^2 = 0.5$ und $p \in \{0.25, 0.5, 0.75\}$.

		$\sigma_2^2 = 3$								
		$p = 0.25$			$p = 0.5$			$p = 0.75$		
		<i>JB</i>	<i>AD</i>	<i>SF</i>	<i>JB</i>	<i>AD</i>	<i>SF</i>	<i>JB</i>	<i>AD</i>	<i>SF</i>
$\mu_2 = 0$	$n = 20$	0.445	0.365	0.453	0.276	0.261	0.307	0.149	0.128	0.155
	$n = 50$	0.775	0.701	0.788	0.522	0.583	0.61	0.167	0.201	0.208
	$n = 100$	0.966	0.926	0.97	0.758	0.875	0.854	0.28	0.395	0.349
	$n = 200$	1	0.998	1	0.95	0.997	0.985	0.45	0.671	0.572
	$n = 400$	1	1	1	1	1	1	0.679	0.932	0.855
$\mu_2 = 2$	$n = 20$	0.538	0.528	0.581	0.313	0.402	0.425	0.112	0.154	0.146
	$n = 50$	0.914	0.897	0.937	0.628	0.813	0.773	0.189	0.316	0.272
	$n = 100$	0.997	0.99	0.998	0.886	0.983	0.976	0.29	0.579	0.465
	$n = 200$	1	1	1	0.992	1	1	0.529	0.906	0.796
	$n = 400$	1	1	1	0.999	1	1	0.811	0.995	0.977

Tabelle B.3: Powerwerte der Teststatistiken *JB*, *AD* und *SF* zum Signifikanzniveau $\alpha = 5\%$ bei einigen Stichprobengrößen n und 1000 Replikationen für verschiedene *KN*-Verteilungen mit den Parametern $\mu_1 = 0$, $\mu_2 \in \{0, 2\}$, $\sigma_1^2 = 1$, $\sigma_2^2 = 3$ und $p \in \{0.25, 0.5, 0.75\}$.

B.2 Numerische Ergebnisse der Analysen im Powervergleich der Teststatistiken JB , JB_U , AD , LIL , CVM , PEA und SF innerhalb des Pearson-Verteilungssystems

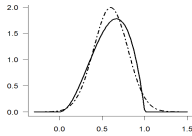
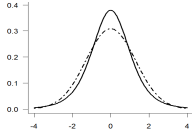
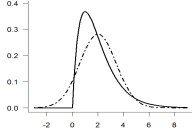
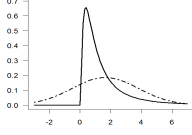
		JB	JB_U	AD	LIL	CVM	PEA	SF
<p>Beta(3,2)</p> 	$n = 10$	0.014	0.014	0.008	0.006	0.006	0.012	0.008
	$n = 20$	0.002	0.002	0.014	0.01	0.01	0.026	0.01
	$n = 50$	0	0	0.04	0.03	0.034	0.038	0.01
	$n = 100$	0	0	0.148	0.062	0.104	0.052	0.07
	$n = 200$	0.006	0.004	0.554	0.252	0.422	0.176	0.548
	$n = 400$	0.804	0.752	0.964	0.586	0.844	0.57	0.988
	<p>Students t (5df)</p> 	$n = 10$	0.054	0.056	0.05	0.048	0.048	0.034
$n = 20$		0.12	0.124	0.088	0.048	0.078	0.036	0.124
$n = 50$		0.26	0.268	0.154	0.078	0.124	0.04	0.25
$n = 100$		0.472	0.478	0.322	0.144	0.27	0.072	0.484
$n = 200$		0.784	0.786	0.568	0.302	0.498	0.11	0.772
$n = 400$		0.958	0.958	0.864	0.596	0.802	0.27	0.958
<p>Gamma(2,1)</p> 		$n = 10$	0.082	0.082	0.086	0.06	0.076	0.048
	$n = 20$	0.168	0.154	0.226	0.144	0.21	0.142	0.256
	$n = 50$	0.512	0.48	0.77	0.44	0.688	0.382	0.798
	$n = 100$	0.912	0.892	0.98	0.808	0.94	0.834	0.992
	$n = 200$	1	1	1	0.994	1	1	1
	$n = 400$	1	1	1	1	0.994	1	1
	<p>Log-Normal</p> 	$n = 10$	0.284	0.272	0.378	0.276	0.362	0.236
$n = 20$		0.622	0.59	0.812	0.624	0.78	0.604	0.826
$n = 50$		0.962	0.956	1	0.976	0.998	0.978	1
$n = 100$		0.998	0.998	1	1	0.88	0.998	1
$n = 200$		1	1	1	1	0.146	1	1
$n = 400$		1	1	1	1	0	1	1

Tabelle B.4: Powerwerte verschiedener Tests auf Normalverteilung zum Signifikanzniveau $\alpha = 1\%$ bei einigen Stichprobengrößen n und 500 Replikationen. Die effizientesten Tests in den einzelnen Kategorien sind durch Fettdruck hervorgehoben. Die Grafiken zeigen die Dichte der entsprechenden Verteilungen (durchgezogene Linien) im Vergleich zur normierten und zentrierten Normalverteilung (gestrichelte Linien).

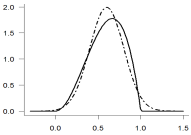
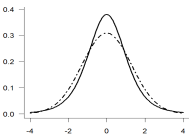
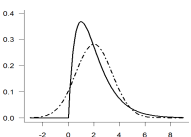
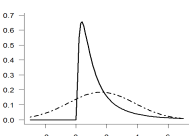
		<i>JB</i>	<i>JB_U</i>	<i>AD</i>	<i>LIL</i>	<i>CVM</i>	<i>PEA</i>	<i>SF</i>
Beta(3,2) 	$n = 10$	0.074	0.052	0.116	0.092	0.116	0.12	0.1
	$n = 20$	0.062	0.05	0.152	0.11	0.13	0.134	0.104
	$n = 50$	0.138	0.082	0.266	0.214	0.236	0.202	0.202
	$n = 100$	0.47	0.366	0.554	0.374	0.472	0.28	0.512
	$n = 200$	0.934	0.912	0.902	0.69	0.806	0.522	0.948
	$n = 400$	1	1	0.994	0.942	0.988	0.874	1
Students t (5df) 	$n = 10$	0.218	0.222	0.19	0.152	0.182	0.146	0.206
	$n = 20$	0.332	0.346	0.256	0.196	0.234	0.174	0.312
	$n = 50$	0.514	0.532	0.37	0.282	0.338	0.204	0.48
	$n = 100$	0.728	0.742	0.594	0.466	0.556	0.276	0.728
	$n = 200$	0.93	0.936	0.836	0.686	0.794	0.428	0.926
	$n = 400$	0.992	0.992	0.97	0.912	0.962	0.63	0.99
Gamma(2,1) 	$n = 10$	0.286	0.252	0.296	0.226	0.282	0.252	0.318
	$n = 20$	0.566	0.494	0.552	0.39	0.51	0.434	0.594
	$n = 50$	0.956	0.926	0.944	0.834	0.918	0.77	0.972
	$n = 100$	1	1	0.998	0.978	0.998	0.964	1
	$n = 200$	1	1	1	1	1	1	1
	$n = 400$	1	1	1	1	0.998	1	1
Log-Normal 	$n = 10$	0.628	0.566	0.664	0.562	0.644	0.596	0.686
	$n = 20$	0.912	0.894	0.94	0.87	0.918	0.888	0.96
	$n = 50$	1	1	1	0.998	1	1	1
	$n = 100$	1	1	1	1	0.89	1	1
	$n = 200$	1	1	1	1	0.158	1	1
	$n = 400$	1	1	1	1	0	1	1

Tabelle B.5: Powerwerte verschiedener Tests auf Normalverteilung zum Signifikanzniveau $\alpha = 10\%$ bei einigen Stichprobengrößen n und 500 Replikationen. Die effizientesten Tests in den einzelnen Kategorien sind durch Fettdruck hervorgehoben. Die Grafiken zeigen die Dichte der entsprechenden Verteilungen (durchgezogene Linien) im Vergleich zur normierten und zentrierten Normalverteilung (gestrichelte Linien).

Anhang C

Daten-CD

In den Kapiteln vier, fünf und sechs dieser Arbeit sind empirische Analysen durchgeführt worden. Die in diesen Analysen verwendeten Größen wurden dabei in der Programmiersprache R im Rahmen von Monte-Carlo-Simulationen generiert. Die Programmcodes zur Erzeugung dieser Größen sowie weitere für die Analysen entwickelte Programme sind auf der beigefügten Daten-CD im Einband am Ende der Arbeit enthalten. Ebenso die Aktienwerte aus Kapitel sechs, die von der Internetseite <http://finance.yahoo.com/> entnommen wurden.

Die Dateien mit den Programmcodes sind dabei den einzelnen Kapiteln und Abschnitten zugeordnet und geben namentlich an, um welche Programme es sich handelt. Auf die in R verwendeten Pakete wird in den entsprechenden Dateien zu Beginn innerhalb einer kurzen Erläuterung hingewiesen. Die genaue Vorgehensweise innerhalb der Analysen wird in den entsprechenden Kapiteln dieser Arbeit detailliert geschildert.

Literaturverzeichnis

- [Als09] GEROLD ALSMEYER: *Mathematische Statistik*. Institut für Mathematische Statistik, Fachbereich Mathematik der Westfälischen Wilhelms-Universität Münster, 3. Auflage 2009.
- [BJ81] ANIL K. BERA, CARLOS M. JARQUE: *An efficient Large-Sample Test for Normality of Observations and Regression Residuals*. Working Paper No.040, March 1981.
- [BS75] K. O. BOWMAN, L. R. SHENTON: *Omnibus Test Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2* . *Biometrika*, Vol. 62, No. 2 (Aug. 1975), pp. 243-250.
- [CH74] D.R. COX, D.V. HINKLEY: *Theoretical Statistics*. 1974; Chapman and Hall.
- [DKB66] F. N. DAVID, M. G. KENDALL, D. E. BARTON: *Symmetric function and allied tables*. Cambridge, At the University Press, 1966.
- [DS96] PARTHA DEB, MARTIN SEFTON: *The distribution of Lagrange multiplier test of normality*. *Economic Letters* 51 (1996), pp. 123-130.
- [Fis30] R. A. FISHER: *The Moments of the Distribution for Normal Samples of Measures of Departure from Normality*. *Proceedings of the Royal Society of London*, Vol. 130, No. 812 (Dec. 2, 1930), pp. 16-28.
- [Hel08] LEONHARD HELD: *Methoden der statistischen Inferenz*. Spektrum Akademischer Verlag Heidelberg, 2008.
- [JB87] CARLOS M. JARQUE, ANIL K. BERA: *A Test for Normality of Observations and Regression Residuals*. *International Statistical Review*, Vol.55, No.2 (Aug.1987), pp.167-171.
- [JG98] D. N. JOANEST, C. A. GILL: *Comparing measures of sample skewness and kurtosis*. *The Statistician* (1998), **47**, Part 1, pp. 183-189.

- [JK94] NORMAN L. JOHNSON, SAMUEL KOTZ: *Continuous univariate distributions/1*. 2nd Edition, 1994.
- [Kha59] C. G. KHATRI: *On the mutual Independence of Certain Statistics*. The Annals of Mathematical Statistics, Vol.30, No.4 (Dec.,1959), pp.1258-1262.
- [KOS09] KOIZUMI K., OKAMOTO N., SEO T.: *On Jarque-Bera tests for assessing multivariate normality*. Journal of Statistics: Advances in Theory and Applications, **1**, pp.207-220.
- [KP66] C. G. KHATRI, K. C. S. PILLAI: *On the moments of traces of two matrices in multivariate Analysis*. Annals of the Institute of Statistical Mathematics, Volume 19, Number 1 (May,1966), pp.143-156.
- [KS67] M.G. KENDALL, A. STUART: *The advanced theory of statistics*. Vol.II, 2nd edition, 1967. London:Griffin.
- [KS69] M.G. KENDALL, A. STUART: *The advanced theory of statistics*. Vol.I, 3rd edition, 1969. London:Griffin.
- [LeC70] L. LECAM: *On the assumptions used to prove asymptotic normality of Maximum Likelihood Estimates*. The Annals of Mathematical Statistics 1970, Vol.41, No.3.
- [MA73] J. F. MALKOVICH, A. A. AFIFI: *On Tests for Multivariate Normality*. Journal of the American Statistical Association, March 1973, Volume 68, Number 341, pp.176-179.
- [Mar70] K. V. MARDIA: *Measures of multivariate skewness and kurtosis with applications*. Biometrika, 57, pp.519-530, 1970.
- [Mar74] K. V. MARDIA: *Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies*. Sankhya B, 36, pp.115-128, 1974.
- [Paw01] YUDI PAWITAN: *In All Likelihood*. University Press, Oxford, 2001.
- [Rin08] HORST RINNE: *Taschenbuch der Statistik*. Verlag Harri Deutsch, Frankfurt am Main, 4. Auflage 2008.
- [SHF85] MINORU SIOTANI, TAKESI HAYAKAWA, YASONURI FUJIKOSHI: *Modern multivariate statistical analysis: a graduate course and handbook*. American Sciences Press, 1985.

- [Sma80] N. J. H. SMALL: *Marginal skewness and kurtosis in testing multivariate normality*. Applied Statistics, 22, pp.260-266.
- [Son01] K. SONG: *Rényi information, loglikelihood and an intrinsic distribution measure*. Journal of Statistical Planning and Inference, 93, pp.51-69, 2001.
- [Sri84] M. S. SRIVASTAVA: *A measure of skewness and kurtosis and a graphical method for assessing multivariate normality*. Statistics and Probability Letters, 2, pp.263-267.
- [ST06] FRIEDRICH SCHMID, MARK TREDE: *Finanzmarktstatistik*. Springer-Verlag Berlin [u.a.] 2006.
- [TB04] THORSTEN THADEWALD, HERBERT BÜNING: *Jarque-Bera Test and its Competitors for Testing Normality - A Power Comparison*. Diskussionsbeiträge des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin, Nr. 2004/09, Volkswirtschaftliche Reihe.
- [Urz96] CARLOS M. URZÚA: *On the correct use of omnibus tests for normality*. Economic Letters 53 (1996), pp. 247-251.
- [Wal49] ABRAHAM WALD: *Note on the Consistency of Maximum Likelihood Estimate*. Annals of Mathematical Statistics, Volume 20, Number 4 (1949), pp.595-601.
- [WK09] DIETHELM WUERTZ, HELMUT KATZGRABER: *Precise finite-sample quantiles of the Jarque-Bera adjusted Lagrange multiplier test*. 11. December 2009; Swiss Federal Institute of Technology, Zurich.
- [WN70] H.WITTING, G.NÖLLE: *Angewandte mathematische Statistik*. B.G. Teubner Stuttgart, 1970.

Gemäß §21 (6) der Diplomprüfungsordnung für den Studiengang Mathematik der Westfälischen Wilhelms-Universität Münster vom 15. Juli 1998 versichere ich, dass ich die vorliegende Diplomarbeit selbständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen und Hilfsmittel benutzt habe.

Ort, Datum

Unterschrift