Die χ^2-Diskrepanz und verwandte Diskrepanzen: Powerapproximationen für Anpassungstests und Konstruktion von Bootstrap-Tests

Dissertation
zur Erlangung des wirtschaftswissenschaftlichen Doktorgrades
des Fachbereichs Wirtschaftswissenschaften
der Universität Göttingen

vorgelegt
von
Achim Lewandowski
aus Uslar

Göttingen, 1993

Inhaltsverzeichnis

Abbildungsverzeichnis **vi**

Tabellenverzeichnis **viii**

1 Einleitung **1**

2 Die Verteilung der χ^2-Prüfgröße, einfache Hypothese **4**

2.1 Einleitung 4

2.2 Definitionen und verwendete allgemeine Resultate 5

2.2.1 Definitionen der Modelle und der Diskrepanzbegiffe im Falle der χ^2-Diskrepanz 5

2.2.2 Weitere Definitionen 8

2.2.3 Bereitstellung von allgemeinen Resultaten 9

2.3 Approximationen der Verteilung der χ^2-Prüfgröße, einfache Hypothese 12

2.3.1 Einleitung 12

2.3.2 Korrekt spezifiziertes Modell 13

2.3.3 Verteilung unter einer festen Alternative 13

2.3.4 „Local alternatives“ 15

2.3.5 Der Ansatz von Patnaik & Frosini 16

2.3.6 Eine verbesserte Approximation 18

2.4 Vergleich der Approximationen, einfache Hypothese 23

2.4.1 Zusammenstellung der untersuchten Approximationen . 23

2.4.2 Ein Vergleich aufgrund der Momente 23

2.4.3 Ein Vergleich aufgrund der asymptotischen Verteilung für eine feste Alternative 27

2.4.4 Ein Vergleich aufgrund der Eignung zur Approximation der Macht, Verteilung unter „local alternatives“ . . 32

2.5 Simulationen zum Vergleich der Approximationen der exakten Macht, einfache Hypothese . . . 46
2.5.1 Durchführung der Simulationen . . . 46
2.5.2 Zusammenfassung der Ergebnisse . . . 53
2.5.3 Interpretation der Simulationsergebnisse . . . 68

3 Die Verteilung der χ^2-Prüfgröße bei zusammengesetzten Hypothesen 71
3.1 Einleitung . . . 71
3.2 Diskrepanzen zur Herleitung von Minimum-Diskrepanz-Schätzern . . . 72
3.2.1 Eine allgemeine Klasse von Diskrepanzen . . . 72
3.2.2 Spezielle Klassen von Diskrepanzen . . . 74
3.3 Annahmen und asymptotische Resultate . . . 75
3.3.1 Hinreichende Bedingungen für die Ableitung der asymptotischen Resultate . . . 75
3.3.2 Asymptotische Resultate . . . 78
3.4 Approximationen der Verteilung der χ^2-Prüfgröße, zusammengesetzte Hypothese . . . 81
3.4.1 Korrekt spezifizierte Modellfamilie . . . 81
3.4.2 Verteilung unter einer festen Alternative . . . 83
3.4.3 „Local alternatives“ . . . 85
3.4.4 Eine verbesserte Approximation für die zusammengesetzte Hypothese . . . 89
3.4.5 Anpassung von Verteilungen über die Momente . . . 93
3.5 Simulationen zum Vergleich der Approximationen der Macht, zusammengesetzte Hypothese . . . 94
3.5.1 Einleitung . . . 94
3.5.2 Beispiel 1: Unabhängigkeit bei zweidimensionalen Kontingenztafeln, 2×2-Tafel . . . 96
3.5.3 Beispiel 2: Unabhängigkeit bei zweidimensionalen Kontingenztafeln, 4×4-Tafel . . . 102
3.5.4 Beispiel 3: Eine einparametrige Modellfamilie . . . 107

4 Erweiterung der Ergebnisse auf andere Diskrepanzen 115
4.1 Einleitung . . . 115
4.2 Einfache Hypothesen . . . 116
4.2.1 Verteilung unter der Hypothese . . . 116

4.2.2 Verteilung unter einer festen Alternative 116
4.2.3 Verteilung unter „local alternatives“ 118
4.2.4 Der Ansatz von Drost et al. 118
4.2.5 Abgewandelter Ansatz von Drost et al. 119
4.2.6 Anpassung von Verteilungen über die Momente 123
4.3 Zusammengesetzte Hypothesen 124
4.3.1 Verteilung unter der Hypothese 124
4.3.2 Verteilung unter einer festen Alternative 125
4.3.3 Verteilung unter „local alternatives“ 126
4.3.4 Ein verbesserter Ansatz zur Approximation der Verteilung von speziellen Prüfgrößen 127

5 Bootstrap-Tests 130
5.1 Das Bootstrap-Verfahren 130
5.2 Bootstrap-Tests für die χ^2-Diskrepanz 132
5.2.1 Vorbereitungen 132
5.2.2 Ein Test über $\Delta(\theta_0)$ 136
5.2.3 Ein alternativer Test über $\Delta(\theta_0)$ 139
5.2.4 Ein Bootstrap-Anpassungstest 142
5.3 Simulationen 148
5.3.1 Probleme bei der Durchführung der Simulationen . . . 148
5.3.2 Simulationen zur Hypothese $H_0 : \Delta(\theta_0) \leq d$ 150
5.3.3 Simulationen zum Bootstrap-Anpassungstest 163
5.3.4 Exkurs: Maximum-Likelihood-Schätzer der ursprünglichen, ungruppierten Stichprobe 170
5.4 Zwei Modellfamilien: Test über die Differenz der Diskrepanzen durch Approximation 174

6 Bootstrap-Tests für andere Diskrepanzen 178
6.1 Einleitung 178
6.2 Tests über $\Delta(\theta_0)$ 179
6.2.1 Test 1 179
6.2.2 Test 2 180
6.2.3 Bootstrap-Anpassungstests 182
6.3 Zwei Modellfamilien: Test über die Differenz der Diskrepanzen durch Approximation 188

7 Schlußwort 191

A Beweise **192**
A.1 Beweis von Satz 2.3 192
A.2 Beweis von Satz 3.1 194
A.3 Beweis des Lemmas 5.4 196

B Tabellen zu Abschnitt 2.5 **199**

C Mathematica-Programm zur Berechnung der Verteilungsfunktionen aus Kapitel 3 **212**
C.1 Das Programm 212
C.2 Eine Mustersitzung 220

Literaturverzeichnis **225**

Abbildungsverzeichnis

2.1 Plot des Fehlers $\beta_G - \beta$ in Abhängigkeit von der asymptotischen Macht β 41
2.2 Plot des Fehlers $\beta_N - \beta$ in Abhängigkeit von der asymptotischen Macht β 42
2.3 Abbildung A aus Broffitt & Randles (1977) 43
2.4 Plot des Fehlers $\beta_{AN} - \beta$ in Abhängigkeit von der asymptotischen Macht β 45
2.5 Exakte und angenäherte Macht – Beispiel 1 47
2.6 Exakte und angenäherte Macht – Beispiel 2 48
2.7 Plot von 100 erzeugten Vektoren $h(\theta_0)$ 51
2.8 Exakte Macht $\leftrightarrow$ App. Macht, LKC-Approximation, (I) 54
2.9 Exakte Macht $\leftrightarrow$ App. Macht, LKC-Approximation, (II) . . . 55
2.10 Exakte Macht $\leftrightarrow$ App. Macht, GLK-Approximation, (I) 56
2.11 Exakte Macht $\leftrightarrow$ App. Macht, GLK-Approximation, (II) . . . 57
2.12 Exakte Macht $\leftrightarrow$ App. Macht, GAM-Approximation, (I) . . . 58
2.13 Exakte Macht $\leftrightarrow$ App. Macht, GAM-Approximation, (II) . . . 59
2.14 Exakte Macht $\leftrightarrow$ App. Macht, NC-Approximation, (I) 60
2.15 Exakte Macht $\leftrightarrow$ App. Macht, NC-Approximation, (II) 61
2.16 Exakte Macht $\leftrightarrow$ App. Macht, NOR-Approximation, (I) . . . 62
2.17 Exakte Macht $\leftrightarrow$ App. Macht, NOR-Approximation, (II) . . . 63
2.18 Exakte Macht $\leftrightarrow$ App. Macht, AN-Approximation, (I) 64
2.19 Exakte Macht $\leftrightarrow$ App. Macht, AN-Approximation, (II) 65

3.1 Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 1, (I) 99
3.2 Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 1, (II) 100
3.3 Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 1, (III) 101
3.4 Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 2, (I) 104
3.5 Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 2, (II) 105
3.6 Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 3, (I) 109

3.7 Exakte Macht ↔ App. Macht, Bsp. 3, (II) 110
3.8 Exakte Macht ↔ App. Macht, Bsp. 3, (III) 111
3.9 Exakte Macht ↔ App. Macht, Bsp. 3, (IV) 112

5.1 Schätzer der wahren Signifikanzniveaus - Test 1 - n=100 . . . 155
5.2 Schätzer der wahren Signifikanzniveaus - Test 1 - n=200 . . . 156
5.3 Schätzer der wahren Signifikanzniveaus - Test 2 - n=100 . . . 157
5.4 Schätzer der wahren Signifikanzniveaus - Test 2 - n=100 . . . 158
5.5 Schätzer der wahren Signifikanzniveaus - Test 3 - n=100 . . . 159
5.6 Schätzer der wahren Signifikanzniveaus - Test 3 - n=200 . . . 160
5.7 Schätzer der wahren Signifikanzniveaus - Test 3 - n=2000 . . . 161
5.8 Schätzer der wahren Signifikanzniveaus - Test 3 - n=10000 . . 162
5.9 Schätzer der wahren Signifikanzniveaus - Bsp. 1 - n=50, 100 . 165
5.10 Schätzer der wahren Signifikanzniveaus - Bsp. 1 - n=200, 400 . 166
5.11 Schätzer der wahren Signifikanzniveaus - Bsp. 2 - n=50, 100 . 168
5.12 Schätzer der wahren Signifikanzniveaus - Bsp. 2 - n=200, 400 . 169
5.13 Schätzer der wahren Signifikanzniveaus, Bsp. von Chernoff & Lehmann (1954) 173

6.1 Schätzer der wahren Signifikanzniveaus - Bsp. 1 - n=50, 200 . 186
6.2 Schätzer der wahren Signifikanzniveaus - Bsp. 2 - n=50, 400 . 187

Tabellenverzeichnis

2.1 Auswertung: $\alpha = 0.10$, $n = 50$, $k = 4$. 66
2.2 Auswertung: $\alpha = 0.10$, $n = 10$, $k = 6$. 67
2.3 Auswertung: $\alpha = 0.10$, $n = 100$, $k = 10$. 67
2.4 Auswertung: $\alpha = 0.01$, $n = 500$, $k = 6$. 67

3.1 Auswertung: $\alpha = 0.10$, $n = 30$, 2×2-Tafel. 102
3.2 Auswertung: $\alpha = 0.10$, $n = 100$, 4×4-Tafel. 106
3.3 Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$. 113
3.4 Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$. 113
3.5 Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$. 113
3.6 Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$. 114

5.1 Schätzer der Macht bei Durchführung des χ^2-Anpassungstests zum Signifikanzniveau $\alpha = 0.10$. 151

B.1 Auswertung: $\alpha = 0.10$, $n = 10$, $k = 3$. 199
B.2 Auswertung: $\alpha = 0.10$, $n = 30$, $k = 3$. 199
B.3 Auswertung: $\alpha = 0.10$, $n = 50$, $k = 3$. 200
B.4 Auswertung: $\alpha = 0.10$, $n = 100$, $k = 3$. 200
B.5 Auswertung: $\alpha = 0.10$, $n = 500$, $k = 3$. 201
B.6 Auswertung: $\alpha = 0.10$, $n = 10$, $k = 4$. 201
B.7 Auswertung: $\alpha = 0.10$, $n = 30$, $k = 4$. 202
B.8 Auswertung: $\alpha = 0.10$, $n = 50$, $k = 4$. 202
B.9 Auswertung: $\alpha = 0.10$, $n = 10$, $k = 6$. 203
B.10 Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$. 203
B.11 Auswertung: $\alpha = 0.10$, $n = 100$, $k = 6$. 204
B.12 Auswertung: $\alpha = 0.10$, $n = 500$, $k = 6$. 204
B.13 Auswertung: $\alpha = 0.10$, $n = 100$, $k = 10$. 205
B.14 Auswertung: $\alpha = 0.01$, $n = 10$, $k = 3$. 205

B.15 Auswertung: $\alpha = 0.01$, $n = 30$, $k = 3$. 206
B.16 Auswertung: $\alpha = 0.01$, $n = 50$, $k = 3$. 206
B.17 Auswertung: $\alpha = 0.01$, $n = 100$, $k = 3$. 207
B.18 Auswertung: $\alpha = 0.01$, $n = 500$, $k = 3$. 207
B.19 Auswertung: $\alpha = 0.01$, $n = 10$, $k = 4$. 208
B.20 Auswertung: $\alpha = 0.01$, $n = 30$, $k = 4$. 208
B.21 Auswertung: $\alpha = 0.01$, $n = 50$, $k = 4$. 209
B.22 Auswertung: $\alpha = 0.01$, $n = 10$, $k = 6$. 209
B.23 Auswertung: $\alpha = 0.01$, $n = 50$, $k = 6$. 210
B.24 Auswertung: $\alpha = 0.01$, $n = 100$, $k = 6$. 210
B.25 Auswertung: $\alpha = 0.01$, $n = 500$, $k = 6$. 211
B.26 Auswertung: $\alpha = 0.01$, $n = 100$, $k = 10$. 211

Kapitel 1

Einleitung

Anpassungstests spielen in der Statistik eine wichtige Rolle. In vielen Fällen kann die Überprüfung der Hypothese, daß das den beobachteten Daten zugrundeliegende wahre Modell in einer vorgegebenen Modellfamile enthalten ist, auf das Testen von Hypothesen über die Parameter einer Multinomialverteilung zurückgeführt werden.

Im Mittelpunkt der vorliegenden Arbeit steht Pearsons χ^2, die in der Literatur wohl am häufigsten eingesetzte Statistik zum Test derartiger Hypothesen. Ausgehend von der χ^2-Diskrepanz, einem Maß für die Abweichung eines hypothetischen vom wahren Parametervektor, lassen sich die Hypothesen zum klassischen Anpassungstest als Hypothesen über die χ^2-Diskrepanz neu formulieren.

Es existieren neben Pearsons χ^2 andere Prüfgrößen für Anpassungstests. Zu nennen wären z. B. die Log-Likelihood-Ratio-Statistik G^2, das modifizierte χ^2 nach Neyman (1949) und, etwas seltener eingesetzt, die Freeman-Tukey-Statistik (Fienberg, 1979). Jeder dieser Statistiken kann eine Diskrepanz zugeordnet werden, und die verschiedenen Diskrepanzen können als unterschiedliche Maße des „Abstands“ des wahren Modells von der approximierenden Modellfamilie interpretiert werden. Unter der Nullhypothese besitzen die genannten Prüfgrößen die gleiche asymptotische Verteilung (Read & Cressie, 1988), bei Fehlspezifikation sind sie i. allg. nicht mehr asymptotisch äquivalent. Eine interessante Aufgabe ist es nun zu untersuchen, welcher Test in welcher Situation zu bevorzugen ist. Dazu werden Aussagen über die Macht benötigt.

In Kapitel 2 wird die Verteilung der χ^2-Prüfgröße bei einfachen Hypothesen untersucht, wobei das asymptotische Verhalten bei Fehlspezifikation

des Modells im Mittelpunkt steht. Es werden klassische Approximationen betrachtet und diesen einige neue bzw. alternativ abgeleitete Approximationen gegenübergestellt. Es folgt ein Vergleich, zunächst aufgrund der ersten beiden Momente, dann in bezug auf die Annäherung an die exakte Verteilungsfunktion mit wachsender Stichprobengröße. In einem weiteren Abschnitt wird die Eignung der Approximationen zur Berechnung der exakten Macht durch einen neuen theoretischen Ansatz überprüft. Es zeigt sich, daß einige Verteilungen *prinzipiell* nicht oder nur mit großen Einschränkungen zur Powerapproximation geeignet sind. Als am theoretisch befriedigendsten erweist sich eine aus den Ergebnissen von Drost et al. (1989) abgeleitete Verteilung. Abgeschlossen wird das Kapitel durch eine breit angelegte Simulationsstudie, die die aufgestellten Behauptungen stützt.

In Kapitel 3 werden die Untersuchungen des zweiten Kapitels, soweit möglich, auf zusammengesetzte Hypothesen übertragen. Zunächst besteht das Problem der Wahl des Schätzverfahrens für die auftretenden Parameter. Hierzu wird eine sehr allgemeine Klasse von Diskrepanzen konstruiert, über die dann sogenannte Minimum-Diskrepanz-Schätzer hergeleitet werden. Die üblicherweise verwendeten Maximum-Likelihood- bzw. Minimum-χ^2-Schätzer sind Spezialfälle dieser Klasse von Schätzern. Mittels der multivariaten Delta-Methode und des Satzes über implizite Funktionen läßt sich dann die asymptotische Verteilung von χ^2 unter der Nullhypothese als auch unter „local alternatives“ (Mitra, 1958) ableiten. Im Gegensatz zu den klassischen Ansätzen können *beliebige* Minimum-Diskrepanz-Schätzer zu Diskrepanzen aus der obengenannten Klasse eingesetzt werden. Die Idee des Ansatzes von Drost et al. (1989) wird für die χ^2-Prüfgröße auf die zusammengesetzte Hypothese übertragen. Durch diese Erweiterung ist es möglich, die erzielbare Macht bei Anwendung des χ^2-Anpassungstests für ein vorgegebenes wahres Modell sehr genau vorauszusagen. Anhand von Simulationen wird die Überlegenheit dieser neuen Approximation gegenüber alten Annäherungen demonstriert.

Die Ideen der Kapitel 2 und 3 werden in Kapitel 4 auf Diskrepanzen aus der oben erwähnten Klasse angewendet, solange dies möglich ist. Es zeigt sich, daß die verbesserte Approximation der wahren Macht u. a. für den modifizierten χ^2-Test nach Neyman (1949) und für den Anpassungstest unter Verwendung der Freeman-Tukey-Statistik herzuleiten ist, wobei im Fall der zusammengesetzten Hypothese beliebige Minimum-Diskrepanz-Schätzer eingesetzt werden dürfen.

Durch die Verwendung der χ^2-Diskrepanz wird die formale Basis geschaf-

fen, die Fehlspezifikation der Modellfamilie von vornherein zuzulassen. In Kapitel 5 wird die realitätsnähere Hypothese überprüft, daß der „Abstand“ zwischen approximierender Modellfamilie und wahrem Modell gemäß der χ^2-Diskrepanz eine vorgegebene Grenze nicht überschreitet. Für die Konstruktion von passenden Tests eignen sich die Anfang der achtziger Jahre entwickelten Bootstrap-Methoden. Nach einer Einführung in das Bootstrap-Verfahren werden zwei Bootstrap-Tests zur genannten Hypothese konstruiert. Ein weiterer Abschnitt zeigt die Schwierigkeiten auf, einen Bootstrap-*Anpassungs*test herzuleiten. In den anschließenden Simulationen wird überprüft, ob das angenommene Signifikanzniveau der Tests zufriedenstellend eingehalten wird. Weiterhin erfolgt die Herleitung eines Bootstrap-Tests zur Hypothese, daß eine Modellfamilie gemäß der χ^2-Diskrepanz einen größeren „Abstand“ zum wahren Modell besitzt als eine alternative Modellfamilie.

In Kapitel 6 wird aufgezeigt, wie Bootstrap-Tests für andere Diskrepanzen aufgebaut werden können. Als konkretes Anwendungsbeispiel wird ein Anpassungstest für die aus der euklidischen Distanz abgeleitete Gauß-Diskrepanz konstruiert.

Kapitel 2

Die Verteilung der χ^2-Prüfgröße, einfache Hypothese

2.1 Einleitung

Der von Karl Pearson (1900) eingeführte χ^2-Anpassungstest steht bis heute im Mittelpunkt von zahlreichen Veröffentlichungen. Es wäre müßig, die Anwendungen und statistischen Eigenschaften aufzuzählen, zu umfangreich sind die Modifikationen und Erweiterungen, die der χ^2-Anpassungstest im Laufe der Zeit erfahren hat. Cressie & Read (1989) geben für den interessierten Leser einen historischen Überblick und weitere Literaturhinweise.

Auffallend ist jedoch, daß sich das Gros der Autoren nur mit der Verteilung der χ^2-Prüfgröße *unter* der zu testenden Hypothese beschäftigt. Relativ wenige Arbeiten behandeln den Fall, daß die Modellfamilie nicht korrekt spezifiziert wurde.

Es ist das Verdienst von Cressie & Read (1984), erkannt zu haben, daß die χ^2-Prüfgröße nur ein Spezialfall einer stetig von einem reellwertigen Parameter abhängigen Familie von Teststatistiken ist. Diese Familie enthält auch so bekannte Statistiken wie die Log-Likelihood-Ratio-Statistik G^2, das modifizierte χ^2 nach Neyman (1949) sowie einige weitere allgemein verwendete Prüfgrößen. *Unter* der Hypothese besitzen die Cressie-Read-Prüfgrößen die gleiche asymptotische χ^2-Verteilung und könnten deshalb prinzipiell parallel eingesetzt werden. Wurde das Modell bzw. die Modellfamilie fehlspezifiziert,

unterscheiden sich die Verteilungen jedoch. Sprechen somit nicht theoretische Gründe für die Verwendung einer speziellen Teststatistik, so sollte der Test durchgeführt werden, der bei nicht zutreffender Hypothese eine möglichst große Macht erzielt.

Die klassischen Ansätze sind jedoch zu grob, um bei Fehlspezifikation der Modellfamilie zwischen den einzelnen Teststatistiken differenzieren zu können. Genauere Approximationen der Macht sind vonnöten, um für vorgegebene Konstellationen eine Auswahl der Teststatistik zu ermöglichen.

In dieser Arbeit sollen für die Verteilung der χ^2-Prüfgröße verbesserte Approximationen hergeleitet werden, wobei sich das Hauptinteresse auf das Verhalten bei Fehlspezifikation konzentriert. Es wird sich zeigen, daß sich die meisten Ansätze für die einfache Hypothese auf die zusammengesetzte Hypothese übertragen lassen. Es lohnt sich somit, die Verteilung der χ^2-Teststatistik für den Fall der *einfachen* Hypothese genauer zu untersuchen. In diesem Kapitel werden bekannte und neue bzw. alternativ abgeleitete Approximationen der Verteilung der χ^2-Prüfgröße im Falle der *einfachen* Hypothese anhand verschiedener Kriterien verglichen.

Der χ^2-Anpassungstest wird in dieser Arbeit nicht in der klassischen Weise gesehen. Durch die Verwendung einer Diskrepanz, einem Maß für die Abweichung zweier Modelle, kann der Anpassungstest als Test über die Diskrepanz neu formuliert werden. Ein Vorteil gegenüber dem klassischen Ansatz besteht darin, eine formale Basis geschaffen zu haben, die die Beschreibung der Verteilung bei Fehlspezifikation erleichtert.

2.2 Definitionen und verwendete allgemeine Resultate

2.2.1 Definitionen der Modelle und der Diskrepanzbegiffe im Falle der χ^2-Diskrepanz

Zunächst sei die verwendete Terminologie genauer beschrieben. In der ursprünglichen Form des χ^2-Anpassungstests liegen Beobachtungen $y_1, \ldots, y_n$ von unabhängig und identisch verteilten Zufallsvariablen bzw. -vektoren $Y_1, \ldots, Y_n$ vor. Der Wertebereich der Y_i wird in k disjunkte Klassen $C_1, \ldots, C_k$ aufgeteilt. Setzt man nun

$$X_i = j \Leftrightarrow Y_i \in C_j, \quad i = 1, \ldots, n,$$

und

$$\pi_j = P(Y_i \in C_j), \quad j = 1, \ldots, k,$$

so kann die folgende, vereinfachende Beschreibung gewählt werden:

Gegeben seien die Realisationen $x_1, \ldots, x_n$ von unabhängig und identisch verteilten, diskreten Zufallsvariablen $X_1, \ldots, X_n$. Unter dem „wahren“ Modell, dem sogenannten *Grundmodell*, besitze jedes X_j, $j = 1, \ldots, n$, die Wahrscheinlichkeitsfunktion

$$f(i) = \begin{cases} \pi_i & : \quad i \in \{1, \ldots, k\} \\ 0 & : \quad sonst \end{cases},$$

wobei $\pi_i > 0$, $i = 1, \ldots, k$, und $\sum_{i=1}^{k} \pi_i = 1$ gelten soll. Außerdem wird vereinbart:

Definition 2.1

$$\begin{array}{llll} a) & n_i & := & card_{j=1,\ldots,n}\{x_j : x_j = i\}, \quad i = 1, \ldots, k \\ b) & z_i & := & n_i/n, \qquad\qquad\qquad\qquad\quad i = 1, \ldots, k \\ c) & \pi & := & (\pi_1, \ldots, \pi_k)^T \\ d) & z & := & (z_1, \ldots, z_k)^T \end{array}$$

Die absoluten Häufigkeiten n_i, mit denen ein Wert i aufgetreten ist, sind dann multinomialverteilt:

$$P(n_1, \ldots, n_k) = \frac{n!}{n_1! \ldots n_k!} \pi_1^{n_1} \ldots \pi_k^{n_k}$$

Sei weiterhin F die Verteilungsfunktion einer Zufallsvariablen X_i und F_n die *empirische* Verteilungsfunktion zu den Daten $x_1, \ldots, x_n$:

$$F_n(x) = n^{-1} card_{j=1,\ldots,n}\{x_j : x_j \leq x\}$$

Jedes approximierende Modell $H_\theta, \theta = (\theta_1, \ldots, \theta_s)^T \in \Theta \subset I\!R^s$, sei durch die Angabe einer Wahrscheinlichkeitsfunktion

$$h_\theta(i) = \begin{cases} h_i(\theta) & : \quad i \in \{1, \ldots, k\} \\ 0 & : \quad sonst \end{cases}$$

definiert. Sei Θ zunächst eine beliebige Teilmenge des $\mathbb{R}^s$, $s < k$. Es sollen wieder für jedes $\theta \in \Theta$ die Bedingungen $h_i(\theta) > 0$, $i = 1, \ldots, k$, und $\sum_{i=1}^{k} h_i(\theta) = 1$ erfüllt sein, sowie zur Vereinfachung der Schreibweise die Definition

$$h(\theta) := (h_1(\theta), \ldots, h_k(\theta))^T$$

getroffen werden. Die Menge der Modelle $\{h(\theta) : \theta \in \Theta\}$ nennt man auch die *approximierende Modellfamilie.*

Das Grundmodell F, das approximierende Modell H_θ, $\theta \in \Theta$, und F_n sind durch die Angabe von π, $h(\theta)$ bzw. z bestimmt. Die χ^2-*Diskrepanz* zwischen dem Grundmodell und einem approximierenden Modell ist dann folgendermaßen definiert (vgl. Linhart & Zucchini, 1986):

$$\Delta(\pi, h(\theta)) := \sum_{i=1}^{k} \frac{(\pi_i - h_i(\theta))^2}{h_i(\theta)}$$

Die χ^2-Diskrepanz nimmt nur nichtnegative Werte an. Je mehr die einzelnen Komponenten der Vektoren π und $h(\theta)$ voneinander abweichen, um so größer wird der Wert der χ^2-Diskrepanz. Die *empirische Diskrepanz* ist gegeben durch:

$$\Delta(z, h(\theta)) := \sum_{i=1}^{k} \frac{(z_i - h_i(\theta))^2}{h_i(\theta)}$$

Die *Diskrepanz durch Approximation* ist definiert durch:

$$\Delta(\pi, \theta_0) := \min_{\theta \in \Theta} \Delta(\pi, h(\theta))$$

Für $\Delta(\pi, h(\theta_0))$ wird häufiger die Kurzform $\Delta(\theta_0)$ verwendet, falls es keinen Zweifel gibt, welches Grundmodell betrachtet wird. Existiert fast sicher zu z ein $\hat{\theta}(z)$ mit

$$\Delta(z, h(\hat{\theta}(z))) = \min_{\theta \in \Theta} \Delta(z, h(\theta)),$$

so wird $\hat{\theta}(z)$ als *Minimum-χ^2-Schätzer* bezeichnet. Die Diskrepanz zwischen dem Grundmodell π und $h(\hat{\theta}(z))$ wird dann *Gesamtdiskrepanz* genannt.

Beim χ^2-Anpassungstest wird die Hypothese überprüft, daß das Grundmodell ein Element der approximierenden Modellfamilie ist. Besteht die approximierende Modellfamilie nur aus einem Element H_{θ_0}, so ist dies ein Test

der einfachen Hypothese

$$H_0 : \pi = h(\theta_0),$$

und die Prüfgröße ist dann durch

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - nh_i(\theta_0))^2}{nh_i(\theta_0)} \tag{2.1}$$

gegeben. Für einen Test der zusammengesetzten Hypothese

$$H_0 : \pi \in \{h(\theta), \theta \in \Theta\}$$

wird

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - nh_i(\hat{\theta}))^2}{nh_i(\hat{\theta})}$$

verwendet, wobei der Schätzer $\hat{\theta}$ auf unterschiedliche Art und Weise gebildet werden kann.

Der χ^2-Anpassungstest läßt sich nun in Beziehung zur χ^2-Diskrepanz setzen:

- Getestet wird die Hypothese $H_0 : \Delta(\theta_0) = 0$
- Für die einfache Hypothese ist die Prüfgröße durch $PG = n\Delta(z, h(\theta_0))$, für die zusammengesetzte Hypothese durch $PG = n\Delta(z, h(\hat{\theta}))$ gegeben.

2.2.2 Weitere Definitionen

Für eine übersichtlichere und knappere Darstellung sollen die folgenden Vereinbarungen gelten:

Definition 2.2 *I_n, $n \in \mathbb{N}$, bezeichne die n-dimensionale Einheitsmatrix. Ergibt sich die Dimension aus dem Zusammenhang, wird die einfachere Schreibweise I verwendet.*

Definition 2.3 *Für $t \in \mathbb{R}^k$ und $\lambda \in \mathbb{R}$ sei D_t^λ die $k \times k$-Diagonalmatrix mit Elementen t_i^λ. Ist $\lambda = 1$, wird die Kurzform D_t benutzt.*

Definition 2.4 *Mit $\chi_m^2(\delta)$ sei eine nichtzentral χ^2-verteilte Zufallsvariable mit m Freiheitsgraden und Nichtzentralitätsparameter $\delta \geq 0$ bezeichnet. Für $\delta = 0$ wird die Kurzform χ_m^2 verwendet.*

Definition 2.5 *Mit $N(\mu, \Sigma)$ sei ein normalverteilter Zufallsvektor mit Erwartungsvektor μ und Kovarianzmatrix Σ bezeichnet. Die Dimension ergibt sich jeweils aus dem Zusammenhang. Im eindimensionalen Fall wird die Schreibweise $N(\mu, \sigma^2)$ benutzt.*

Die verwendeten abkürzenden Schreibweisen der verschiedenen Konvergenzen von Folgen von Zufallsvariablen bzw. von Verteilungsfunktionen, sind an Serflings (1980, Abschnitt 1.2) Nomenklatur angelehnt:

Definition 2.6 *Sei X_n, $n = 1, 2, 3, \ldots$, eine Folge von Zufallsvariablen und seien X bzw. Y weitere Zufallsvariablen. Dann bedeutet*

a) $X_n \xrightarrow{p} X$ **:** *X_n konvergiert stochastisch gegen X.*

b) $X_n \xrightarrow{wp1} X$ **:** *X_n konvergiert fast sicher gegen X.*

c) $X_n \xrightarrow{d} X$ **:** *X_n konvergiert in der Verteilung gegen X.*

d) $X \sim Y$ **:** *X besitzt die gleiche Verteilung wie Y.*

2.2.3 Bereitstellung von allgemeinen Resultaten

Es sei nun eine Reihe von allgemeinen Resultaten gegeben, auf die im Folgenden wiederholt zurückgegriffen wird.

Satz 2.1 *Sei X_n, $n = 1, 2, 3, \ldots$, eine Folge von k-dimensionalen Zufallsvektoren, X ein weiterer Zufallsvektor mit*

$$X_n \xrightarrow{d} X \quad ,$$

A eine $m \times k$-Matrix und B eine $k \times k$-Matrix, dann folgt:

$$\begin{array}{ll} 1. & X_n^T B X_n \xrightarrow{d} X^T B X \\ 2. & A X_n \xrightarrow{d} A X \end{array}$$

Beweis:
Dies ist eine Aussage des Korollars 1.7 von Serfling (1980, S. 25). □

Satz 2.2 *Sei X ein multivariat normalverteilter k-dimensionaler Zufallsvektor:*

$$X \sim N(\mu, \Sigma)$$

Dann ist $X^T X$ verteilt wie eine Linearkombination von unabhängig, möglicherweise nichtzentral χ^2-verteilten Zufallsvariablen plus einem additiven Glied:

$$X^T X \sim \sum_{\lambda_i > 0} \lambda_i \chi_1^2 (\frac{(S\mu)_i^2}{\lambda_i}) + \sum_{\lambda_i = 0} (S\mu)_i^2 \quad ,$$

wobei die transponierten Zeilen der $k \times k$-Matrix S aus einer orthonormalen Basis aus Eigenvektoren von Σ bestehen und die $\lambda_i, i = 1, \ldots, k$, die zugehörigen Eigenwerte bezeichnen.

Bemerkung: Die Aussage ergäbe sich nach einigen Berechnungen auch als Spezialfall des Theorems 2.1 von Dik & de Gunst (1985). Da die genannten Autoren von einer allgemeineren Problemstellung ausgehen, ist ihr Beweis für die vorliegende Aussage überdimensioniert. Es wird deshalb ein alternativ abgeleiteter, kürzerer Beweis gegeben.

Beweis:

Die Matrix Σ ist symmetrisch, d. h., es gibt eine orthonormale Basis aus Eigenvektoren. Diese Vektoren bilden die Zeilen der orthogonalen Matrix S. Außerdem ist Σ als Kovarianzmatrix positiv semidefinit, die zugehörigen Eigenwerte λ_i sind deshalb nicht-negativ. Ohne Beschränkung der Allgemeinheit kann zunächst angenommen werden, daß die Zeilen von S so angeordnet sind, daß $\lambda_1, \ldots, \lambda_r > 0$ und $\lambda_{r+1}, \ldots, \lambda_k = 0$ gilt. Die Matrix S wird analog partitioniert in die $r \times k$-Matrix S_r und in die $(k-r) \times k$-Matrix S_{k-r}. Sei Λ die $k \times k$-Diagonalmatrix der Eigenwerte $\lambda_1, \ldots, \lambda_k$ und Λ_r die $r \times r$-Diagonalmatrix der positiven Eigenwerte $\lambda_1, \ldots, \lambda_r$. Es gilt dann:

$$\begin{aligned} \Sigma &= S^T \Lambda S \\ &= S_r^T \Lambda_r S_r \end{aligned}$$

Seien weiterhin Y und Z multivariat normalverteilt:

$$\begin{aligned} Y &\sim N(0, \Lambda_r) \\ Z &\sim N(0, I_r) \end{aligned}$$

Es ergibt sich dann:

$$\begin{aligned}
X &\sim \mu + S_r^T Y \\
\Rightarrow X^T X &\sim (\mu + S_r^T Y)^T (\mu + S_r^T Y) \\
&= \mu^T \mu + 2\mu^T S_r^T Y + Y^T Y \\
&= \mu^T \mu + (Y + S_r\mu)^T (Y + S_r\mu) - (S_r\mu)^T (S_r\mu) \\
&= (S\mu)^T (S\mu) + (Y + S_r\mu)^T (Y + S_r\mu) - (S_r\mu)^T (S_r\mu) \\
&= (Y + S_r\mu)^T (Y + S_r\mu) + (S_{k-r}\mu)^T (S_{k-r}\mu) \\
&= \sum_{i=1}^{r} (Y_i + (S\mu)_i)^2 + \sum_{i=r+1}^{k} (S\mu)_i^2
\end{aligned}$$

Da $Y \sim \Lambda_r^{\frac{1}{2}} Z$, folgt nun

$$X^T X \sim \sum_{i=1}^{r} \lambda_i \left(Z_i + \frac{(S\mu)_i}{\sqrt{\lambda_i}} \right)^2 + \sum_{i=r+1}^{k} (S\mu)_i^2,$$

d.h., da die einzelnen Z_i unabhängig verteilt sind,

$$X^T X \sim \sum_{i=1}^{r} \lambda_i \chi_1^2 (\frac{(S\mu)_i^2}{\lambda_i}) + \sum_{i=r+1}^{k} (S\mu)_i^2 \quad .$$

Bei einer beliebigen Anordnung der Zeilen von S ergibt sich

$$X^T X \sim \sum_{\lambda_i > 0} \lambda_i \chi_1^2 (\frac{(S\mu)_i^2}{\lambda_i}) + \sum_{\lambda_i = 0} (S\mu)_i^2 \quad .$$

Es folgt die Behauptung. □

Besonders wichtig für dieses Kapitel ist die Delta-Methode. Sie sei an dieser Stelle in einer modifizierten Form gegeben und wird im Anhang A.1 bewiesen:

Satz 2.3 *Sei $x_0 \in I\!R^k$ und sei $x_n \in I\!R^k$, $n = 1, 2, 3, \ldots,$ eine Folge von Vektoren mit $lim_{n\to\infty} x_n = x_0$. Sei weiterhin X_n, $n = 1, 2, 3, \ldots,$ eine Folge von k-dimensionalen Zufallsvektoren mit einer asymptotischen Normalverteilung in dem Sinne, daß*

$$\sqrt{n}(X_n - x_n) \xrightarrow{d} N(\mu(x_0), \Sigma(x_0))$$

gilt. Nun sei $f(x)$ eine Abbildung vom $\mathbb{R}^k$ in den $\mathbb{R}^l$, die in einer offenen Umgebung $U(x_0) \subset \mathbb{R}^k$ stetig differenzierbar ist. Bezeichne $F(x)$ die Jacobi-Matrix für $x \in U(x_0)$, dann folgt für die asymptotische Verteilung:

$$\sqrt{n}(f(X_n) - f(x_n)) \xrightarrow{d} N(F(x_0)\mu(x_0), F(x_0)\Sigma(x_0)F(x_0)^T)$$

Der Beweis zu diesem Satz zeigt weiterhin, daß für den Fall $x_n = x_0$, $n = 1, 2, 3, \ldots$, nur die totale Differenzierbarkeit von f im Punkt x_0 erforderlich ist. Damit ist auch der folgende, häufiger benutzte Satz bewiesen:

Satz 2.4 *Sei $x_0 \in \mathbb{R}^k$ und sei weiterhin X_n, $n = 1, 2, 3, \ldots$, eine Folge von k-dimensionalen Zufallsvektoren mit einer asymptotischen Normalverteilung in dem Sinne, daß*

$$\sqrt{n}(X_n - x_0) \xrightarrow{d} N(\mu(x_0), \Sigma(x_0))$$

gilt. Nun sei $f(x)$ eine Abbildung vom $\mathbb{R}^k$ in den $\mathbb{R}^l$, die in einer offenen Umgebung $U(x_0) \subset \mathbb{R}^k$ definiert und in x_0 total differenzierbar ist. Bezeichne $F = F(x_0)$ die Jacobi-Matrix in x_0, dann folgt für die asymptotische Verteilung:

$$\sqrt{n}(f(X_n) - f(x_0)) \xrightarrow{d} N(F\mu(x_0), F\Sigma(x_0)F^T)$$

2.3 Approximationen der Verteilung der χ^2-Prüfgröße, einfache Hypothese

2.3.1 Einleitung

In diesem Abschnitt werden für die Verteilung der χ^2-Prüfgröße Approximationen gegeben. Es sei nur der klassische, der sogenannte „fixed-cells“-Fall betrachtet. Ansätze mit steigender Zellenzahl bei wachsender Stichprobengröße werden in dieser Arbeit aus Gründen der besseren Überschaubarkeit nicht behandelt.

Die zahlreichen in der Literatur erwähnten Modifikationen und Korrekturfaktoren für die Verteilung *unter* der Hypothese sollen hier nicht im Mittelpunkt stehen.

Der Einfachheit halber sei $z^{(n)}$ mit z bezeichnet, solange es nicht zu Mißverständnissen kommen kann.

2.3.2 Korrekt spezifiziertes Modell

Die approximierende Modellfamilie besteht in diesem Fall nur aus einem Element $h(\theta_0)$, und die χ^2-Prüfgröße ist durch

$$n\Delta(z, h(\theta_0)) = n \sum_{i=1}^{k} \frac{(z_i - h_i(\theta_0))^2}{h_i(\theta_0)}$$

gegeben (vgl. (2.1)). Da unter der Annahme korrekter Modellspezifikation $\pi = h(\theta_0)$ ist, folgt für $n \to \infty$ das klassische Resultat (siehe z. B. Read & Cressie, 1988):

$$n\Delta(z, h(\theta_0)) \xrightarrow{d} \chi^2_{k-1}$$

Es gibt nun Situationen, in denen es nicht ratsam ist, bei Durchführung eines Anpassungstests die kritischen Werte aus der asymptotischen Verteilung abzulesen. Hauptsächlich für kleine Stichproben bzw. bei kleinen erwarteten Häufigkeiten unter der Hypothese kann das angenommene vom tatsächlichen Signifikanzniveau stark abweichen. Diverse Autoren haben sich dieses Problems angenommen und Verbesserungen vorgeschlagen. Cressie & Read (1989) geben in ihrem Artikel einen Überblick.

2.3.3 Verteilung unter einer festen Alternative

Unter der Voraussetzung, daß das Modell falsch spezifiziert wird, d. h. $\pi \neq h(\theta_0)$, gilt für beliebiges $c \in I\!R$

$$\lim_{n\to\infty} P(n\Delta(z, h(\theta_0)) \geq c) = 1.$$

Der Anpassungstest ist somit konsistent (Read & Cressie, 1988, S. 54), gleichzeitig existiert jedoch keine asymptotische Verteilung für $n\Delta(z, h(\theta_0))$. Für $\sqrt{n}(\Delta(z, h(\theta_0)) - \Delta(\pi, h(\theta_0)))$ läßt sich hingegen mittels der Delta-Methode eine asymptotische Normalverteilung herleiten.

Für den Vektor z der relativen Häufigkeiten folgt als Konsequenz der multivariaten Version des zentralen Grenzwertsatzes (Bishop et al., 1975):

$$\sqrt{n}(z - \pi) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T) \tag{2.2}$$

Da

$$\Delta(z, h(\theta_0)) = \sum_{i=1}^{k} \frac{(z_i - h_i(\theta_0))^2}{h_i(\theta_0)} = \sum_{i=1}^{k} \frac{z_i^2}{h_i(\theta_0)} \quad - 1$$

ist, genügt es, die Funktion

$$\Delta(p, h(\theta_0)) = \sum_{i=1}^{k} \frac{p_i^2}{h_i(\theta_0)} \quad - 1$$

zu betrachten. Nun ist

$$\left.\frac{\partial \Delta(p, h(\theta_0))}{\partial p_j}\right|_{p=\pi} = 2\frac{\pi_j}{h_j(\theta_0)}, \quad j = 1, \ldots, k,$$

die $1 \times k$-Jacobi-Matrix F, D_h steht abkürzend für $D_{h(\theta_0)}$, also durch

$$F = 2\pi^T D_h^{-1}$$

gegeben. Die Anwendung des Satzes 2.4 auf (2.2) liefert:

$$\sqrt{n}(\Delta(z, h(\theta_0)) - \Delta(\pi, h(\theta_0))) \xrightarrow{d} N(0, 4\pi^T D_h^{-1}(D_\pi - \pi\pi^T)D_h^{-1}\pi) \tag{2.3}$$

Die asymptotische Varianz läßt sich auch als

$$\sigma^2 = 4\left[\sum_{i=1}^{k} \frac{\pi_i^3}{h_i^2(\theta_0)} - \left(\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)}\right)^2\right] \tag{2.4}$$

schreiben. Es muß nun sichergestellt werden, daß σ^2 tatsächlich positiv ist. $D_\pi - \pi\pi^T$ ist als Kovarianzmatrix positiv semidefinit. Es läßt sich zeigen, daß alle Eigenvektoren zum Eigenwert $\lambda = 0$ durch den Vektor

$$(1, \ldots, 1)^T$$

erzeugt werden. σ^2 kann demnach nur dann 0 sein, falls

$$\pi^T D_h^{-1} = c(1, \ldots, 1), \quad c \in I\!R,$$

ist. Dies ist jedoch nur für $\pi = h(\theta_0)$ der Fall.

Als Approximation für die Verteilung der χ^2-Prüfgröße wird wegen 2.3 die folgende Normalverteilung verwendet:

$$N(n\Delta(\pi, h(\theta_0)), n\sigma^2) \tag{2.5}$$

Dieses Resultat benutzen Broffitt & Randles (1977) zur Herleitung einer weiteren Approximation, die die Erwartung μ_1' und Varianz μ_2 der χ^2-Prüfgröße verwendet. Sie zeigen, daß

$$(n\Delta(z, h(\theta_0)) - \mu_1')/\sqrt{\mu_2} \xrightarrow{d} N(0, 1)$$

gilt. Die als Approximation verwendete Normalverteilung lautet in diesem Fall:

$$N(\mu_1', \mu_2) \tag{2.6}$$

2.3.4 „Local alternatives“

2.3.4.1 Nichtzentrale χ^2-Verteilung unter „local alternatives“

Die Idee der „local alternatives“ besteht darin, die Alternative, also das Grundmodell, nicht festzuhalten, sondern in bestimmter Weise gegen das hypothetische Modell konvergieren zu lassen:

$$\pi^{(n)} = h(\theta_0) + \frac{c}{\sqrt{n}}, \quad c \in I\!R^k, \sum_{i=1}^{k} c_i = 0$$

Zu gegebenem $\pi^{(n)}$ sei jeweils eine endliche Folge $X_1^{(n)}, \ldots, X_n^{(n)}$ von unter dem Modell $\pi^{(n)}$ unabhängig verteilten Zufallsvariablen gegeben. Der Vektor $z^{(n)}$ gebe wieder die zugehörigen relativen Häufigkeiten an. Nach Cochran (1952) folgt dann für $n \to \infty$:

$$n\Delta(z^{(n)}, h(\theta_0)) \xrightarrow{d} \chi^2_{k-1}(\sum_{i=1}^{k} \frac{c_i^2}{h_i(\theta_0)}).$$

Um nun für ein bestimmtes Modell π eine Approximation der Verteilung der χ^2-Prüfgröße zu erhalten, wird $\pi^{(n)}$, n ist durch die vorliegende Stichprobengröße bestimmt, dem Modell π gleichgesetzt, und der Nichtzentralitätsparameter $\delta = \sum_{i=1}^{k} c_i^2/h_i(\theta_0)$ berechnet sich dann zu

$$\delta = n \sum_{i=1}^{k} \frac{(\pi_i - h_i(\theta_0))^2}{h_i(\theta_0)} = n\Delta(\pi, h(\theta_0)).$$

Als Approximation für die Verteilung der χ^2-Prüfgröße wird deshalb in diesem Fall die folgende nichtzentrale χ^2-Verteilung verwendet:

$$\chi^2_{k-1}(n\Delta(\pi, h(\theta_0))) \tag{2.7}$$

Patnaik (1949) gibt ebenfalls diese Approximation an, er entwickelt sie aber nicht im Rahmen von „local alternatives“.

2.3.4.2 Approximationen der nichtzentralen χ^2-Verteilung

Da um 1950 nur wenig ergiebige Tabellen für die nichtzentrale χ^2- Verteilung existierten, bemühten sich in den folgenden Jahren einige Statistiker, Approximationen der nichtzentralen χ^2-Verteilung zu erhalten. Patnaik (1949)

selbst gibt eine einfache Approximation. Er erkannte durch Betrachtung der Momente der nichtzentralen χ^2-Verteilung, daß eine angepaßte Gammaverteilung mit den gleichen ersten beiden Momenten eine gute Approximation liefern könnte. Das Vorgehen wird durch die folgende Betrachtung gestützt:

Sei mit ρ_3 die Schiefe bzw. mit ρ_4 der Exzeß einer Verteilung bezeichnet. Im System nach Pearson (siehe z. B. Johnson & Kotz, 1970, Vol. 1) ist dann $\beta_1 := \rho_3^2$ bzw. $\beta_2 := \rho_4 + 3$. Für eine beliebige Gammaverteilung gilt (Johnson & Kotz, 1970, Vol. 1, S. 14):

$$2\beta_2 - 3\beta_1 = 6$$

Für eine $\chi^2_\nu(\lambda)$-verteilte Zufallsvariable erhält man die folgende Gleichung (Johnson & Kotz, 1970, Vol. 2, S. 135):

$$2\beta_2 - 3(1 - \frac{\lambda^2}{(\nu + 3\lambda)^2})\beta_1 = 6$$

Es ist erkennbar, daß sich für kleine Nichtzentralitätsparameter λ bzw. für eine große Anzahl von Freiheitsgraden die Gleichungen nur wenig unterscheiden, d. h. die nichtzentrale χ^2-Verteilung durch eine Gammaverteilung approximiert werden könnte.

Von E. S. Pearson (1959) stammt die Erweiterung, als Approximation eine dreiparametrige Gammaverteilung zu verwenden, so daß die ersten drei Momente übereinstimmen.

Abdel-Aty (1954) und Sankaran (1959, 1963) geben Transformationen für nichtzentralverteilte χ^2-Variablen an und erhalten dann Normalverteilungen als Approximationen.

Die genannten Approximationen stellen nur eine Auswahl dar. Ihnen ist gemein, daß sie im Hinblick auf eine bessere Anpassung an die exakte Verteilung der χ^2-Prüfgröße nur bedingt eine Hilfe darstellen, da die in (2.7) gegebene nichtzentrale χ^2-Verteilung selbst nur eine möglicherweise mit deutlichen Abweichungen behaftete Approximation der exakten Verteilung ist.

Da inzwischen fast jedes Statistikprogrammpaket die Berechnung der Verteilungsfunktion der nichtzentralen χ^2-Verteilung ermöglicht, ist die Bedeutung dieser Approximationen zusätzlich gemindert.

2.3.5 Der Ansatz von Patnaik & Frosini

Von Patnaik (1949) stammt die Idee, zur Approximation der Verteilung der χ^2-Prüfgröße wieder eine Gammaverteilung zu verwenden, jedoch die Para-

meter so zu wählen, daß die ersten beiden Momente mit den entsprechenden Momenten der *exakten* Verteilung der χ^2-Prüfgröße übereinstimmen. Die Anpassung einer Gammaverteilung wird dabei nicht durch ein asymptotisches Ergebnis gestützt, kombiniert wurde hier eher die Erkenntnis, daß für eine geringe Diskrepanz zwischen dem wahren Modell π und dem hypothetischen Modell $h(\theta_0)$ eine nichtzentrale χ^2-Verteilung eine gute Approximation für die exakte Verteilung liefert, mit der Vermutung, daß die nichtzentrale χ^2-Verteilung unter diesen Voraussetzungen (kleiner Nichtzentralitätsparameter!) durch eine Gammaverteilung angenähert werden kann. Da die Momente der exakten Verteilung angegeben werden können, ist es naheliegend, direkt eine Gammaverteilung anzupassen.

Erwartung und Varianz von $n\Delta(z, h(\theta_0))$ sind nach Patnaik (1949) durch

$$\begin{aligned}
\mu_1' &= (n-1)\sum_{i=1}^{k}\frac{\pi_i^2}{h_i(\theta_0)} + \sum_{i=1}^{k}\frac{\pi_i}{h_i(\theta_0)} - n \\
&= n\Delta(\pi, h(\theta_0)) + \sum_{i=1}^{k}\frac{\pi_i(1-\pi_i)}{h_i(\theta_0)} \quad \text{und} \qquad (2.8)\\
\mu_2 &= 1/n\{(n-1)(6-4n)[\sum_{i=1}^{k}\frac{\pi_i^2}{h_i(\theta_0)}]^2 + 4(n-1)(n-2)\sum_{i=1}^{k}\frac{\pi_i^3}{h_i^2(\theta_0)} \\
&\quad -4(n-1)\sum_{i=1}^{k}\frac{\pi_i^2}{h_i(\theta_0)}\sum_{i=1}^{k}\frac{\pi_i}{h_i(\theta_0)} + 6(n-1)\sum_{i=1}^{k}\frac{\pi_i^2}{h_i^2(\theta_0)} \qquad (2.9)\\
&\quad -[\sum_{i=1}^{k}\frac{\pi_i}{h_i(\theta_0)}]^2 + \sum_{i=1}^{k}\frac{\pi_i}{h_i^2(\theta_0)}\}
\end{aligned}$$

gegeben. Patnaik selbst hielt den Ausdruck für μ_2 zu kompliziert, und er hat den von ihm vorgeschlagenen Ansatz nicht weiterverfolgt.

Wird für die Dichtefunktion einer gammaverteilten Zufallsvariablen X ($X \sim G(\alpha, \beta)$) die Schreibweise

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} \qquad (2.10)$$

verwendet, dann ergibt sich:

$$\alpha = \mu_1'^2/\mu_2 \quad \beta = \mu_2/\mu_1'$$

Frosini (1976) greift den Ansatz von Patnaik (1949) wieder auf. Er beschränkt sich auf den Fall $h_i(\theta_0) = 1/k$, $i = 1, \ldots, k$. Im direkten Vergleich

mit der nichtzentralen χ^2-Verteilung als Approximation der exakten Verteilung zeigt sich die Gammaverteilung bei 40 verschiedenen Kombinationen von n, k und π in den meisten Fällen als mindestens gleichwertig, in manchen Fällen als deutlich überlegen. Erstaunlich ist, daß auch für große Diskrepanzen zwischen π und $h(\theta_0)$ die Gammaverteilung eine gute Approximation darstellt, während die nichtzentrale χ^2-Verteilung normalerweise unter dieser Voraussetzung nicht als Approximation geeignet ist. Das Versagen der $\chi^2_{k-1}(n\Delta(\pi, h(\theta_0)))$-Verteilung für solche Fälle führt Frosini darauf zurück, daß i. allg. der Quotient aus der wahren Varianz und der Varianz einer $\chi^2_{k-1}(n\Delta(\pi, h(\theta_0)))-$Variablen selbst für $n \to \infty$ nicht gegen 1 strebt.

Montinaro (1988) erweitert die Betrachtung auf verschiedene Hypothesen, und seine Resultate stützen die Ergebnisse von Frosini.

Einschränkend muß erwähnt werden, daß beide Autoren eine relativ kleine Zellenanzahl betrachten ($k \leq 7$) und der Stichprobenumfang den Wert 50 nicht überschreitet.

2.3.6 Eine verbesserte Approximation

2.3.6.1 Theoretische Ableitung der Approximation

Drost, Kallenberg, Moore und Oosterhoff (1989) leiten für den Fall der einfachen Hypothese eine verbesserte Approximation her. Sie betrachten dabei nicht nur die Verteilung der χ^2-Prüfgröße, sondern auch die Verteilungen der Prüfgrößen, die sich aus den allgemeineren Cressie-Read-Diskrepanzen ergeben. Eine ausführliche Beschreibung dieser Klasse von Diskrepanzen erfolgt in Abschnitt 3.2.2.

Da sich dieses Kapitel auf die χ^2-Diskrepanz beschränkt, soll eine vereinfachte, alternative Herleitung für diesen Spezialfall gegeben werden, für die einige der von Drost et al. (1989) für den allgemeinen Ansatz verwendeten Konstruktionen nicht benötigt werden.

Sei $h(\theta_0)$ das hypothetische Modell. Unter dem Grundmodell π kann zunächst mittels der Delta-Methode eine asymptotische Verteilung für

$$\sqrt{n}\left(\frac{z_1 - h_1(\theta_0)}{\sqrt{h_1(\theta_0)}}, \ldots, \frac{z_k - h_k(\theta_0)}{\sqrt{h_k(\theta_0)}}\right)^T$$

abgeleitet werden:

Sei $w^*(p) : [0,1]^k \to I\!R^k$ die Abbildung mit

$$w_i^*(p) = \frac{p_i - h_i(\theta_0)}{\sqrt{h_i(\theta_0)}}, \quad i = 1, \ldots, k.$$

Die Ableitung von $w_i^*(p)$ nach p_j in π ist gegeben durch:

$$\left.\frac{\partial w_i^*}{\partial p_j}\right|_{p=\pi} = \frac{\delta_{ij}}{\sqrt{h_i(\theta_0)}}, \quad i, j = 1, \ldots, k$$

Die Jacobi-Matrix W^* nimmt deshalb die einfache Gestalt

$$W^* = D_h^{-\frac{1}{2}}$$

an. Die Anwendung der Delta-Methode auf (2.2) ergibt somit:

$$\sqrt{n}(w^*(z) - w^*(\pi)) \xrightarrow{d} N(\mathbf{0}, D_h^{-\frac{1}{2}}(D_\pi - \pi\pi^T)D_h^{-\frac{1}{2}})$$

Die Idee ist nun, die Verteilung der χ^2-Prüfgröße $nw^*(z)^T w^*(z)$ durch die Verteilung von nY^TY zu approximieren, wobei

$$\sqrt{n}(Y - w^*(\pi)) \sim N(\mathbf{0}, D_h^{-\frac{1}{2}}(D_\pi - \pi\pi^T)D_h^{-\frac{1}{2}}), \tag{2.11}$$

also eine exakte Normalverteilung angenommen wird. Aus (2.11) folgt nun

$$\sqrt{n}Y \sim N(\sqrt{n}w^*(\pi), D_h^{-\frac{1}{2}}(D_\pi - \pi\pi^T)D_h^{-\frac{1}{2}})$$

Die Anwendung des Satzes 2.2 ergibt:

$$nY^TY \sim \sum_{\lambda_i > 0} \lambda_i \chi_1^2\left(\frac{n(Sw^*(\pi))_i^2}{\lambda_i}\right) + \sum_{\lambda_i = 0} n(Sw^*(\pi))_i^2 \tag{2.12}$$

S ist hier eine orthogonale Matrix, deren transponierte Zeilen eine orthonormale Basis von Eigenvektoren der Matrix $V := D_h^{-\frac{1}{2}}(D_\pi - \pi\pi^T)D_h^{-\frac{1}{2}}$ bilden.

Es wird nun der konstante, nichtstochastische Term aus (2.12) näher betrachtet. Die Matrix V ist symmetrisch und positiv semidefinit, und da $D_\pi - \pi\pi^T$ den Rang $k-1$ besitzt, ist der Rang von V ebenfalls $k-1$. Es gibt deshalb genau einen Eigenvektor s_i aus der genannten Basis zum Eigenwert 0, o. B. d. A. sei i=k. Für diesen Eigenvektor gilt:

$$D_h^{-\frac{1}{2}}(D_\pi - \pi\pi^T)D_h^{-\frac{1}{2}} s_k = \mathbf{0}$$

$$\Rightarrow (D_\pi - \pi\pi^T) D_h^{-\frac{1}{2}} s_k = \mathbf{0}$$

Multiplikation mit $h(\theta_0)^T D_\pi^{-1}$ von links ergibt:

$$\begin{aligned} h(\theta_0)^T (I - (1,\ldots,1)^T \pi^T) D_h^{-\frac{1}{2}} s_k &= 0 \\ \Rightarrow (h(\theta_0)^T - \pi^T) D_h^{-\frac{1}{2}} s_k &= 0 \end{aligned}$$

Da

$$\begin{aligned} w^*(\pi) &= \left(\frac{\pi_1 - h_1(\theta_0)}{\sqrt{h_1(\theta_0)}}, \ldots, \frac{\pi_k - h_k(\theta_0)}{\sqrt{h_k(\theta_0)}} \right)^T \\ &= D_h^{-\frac{1}{2}} (\pi - h(\theta_0)) \end{aligned}$$

ist, folgt

$$(Sw^*(\pi))_k = s_k^T D_h^{-\frac{1}{2}} (\pi - h(\theta)) = 0 \quad .$$

Es kann nun die endgültige Approximation für die Verteilung der χ^2-Prüfgröße angeschrieben werden:

$$nY^TY \sim \sum_{i=1}^{k-1} \lambda_i \chi_1^2 \left(\frac{n(Sw^*(\pi))_i^2}{\lambda_i} \right) \tag{2.13}$$

Die besondere Eignung dieser Approximation ergibt sich aus der Tatsache, daß sowohl korrekt als auch falsch spezifizierte Modelle zugelassen sind. Ist das Modell korrekt spezifiziert, d. h. $\pi = h(\theta_0)$, so folgt aus $w^*(\pi) = \mathbf{0}$ und $V = I - \sqrt{\pi}\sqrt{\pi}^T$ sofort das klassische Resultat aus Abschnitt 2.3.2, denn $I - \sqrt{\pi}\sqrt{\pi}^T$ hat als idempotente Matrix mit Spur $k-1$ genau $k-1$ Eigenwerte $\lambda_i = 1$.

Auch für diese Approximation waren die Vorüberlegungen von Patnaik (1949) grundlegend. Er schrieb $\Delta(z, h(\theta_0))$ in der Form

$$\begin{aligned} n\Delta(z, h(\theta_0)) &= \sum_{i=1}^{k} \frac{(n(z_i - h_i(\theta_0)))^2}{nh_i(\theta_0)} \\ &= \sum_{i=1}^{k} \frac{\pi_i}{h_i(\theta_0)} \left(\sqrt{n} \left(\frac{z_i - \pi_i}{\sqrt{\pi_i}} + \frac{\pi_i - h_i(\theta_0)}{\sqrt{\pi_i}} \right) \right)^2 \end{aligned}$$

und bemerkte, daß eine brauchbare Approximation entsteht, falls man für

$$\sqrt{n} \frac{z_i - \pi_i}{\sqrt{\pi_i}}, \quad i = 1, \ldots, k,$$

gemäß der asymptotischen Verteilung eine multivariate Normalverteilung $N(\mathbf{0}, I - \sqrt{\pi}\sqrt{\pi}^T)$ annehmen könnte. Dieser Ansatz führt ebenfalls zu der hergeleiteten Approximation (2.13). Die alternative Herleitung läßt sich jedoch einfacher auf den Fall der zusammengesetzten Hypothese übertragen.

Patnaik benutzte seinen Ansatz nur zur Ermittlung der Momente dieser Approximation und schlug vor, wiederum eine Gammaverteilung anzupassen. Erst Drost et al. (1989) haben die exakte Verteilung explizit bestimmt.

2.3.6.2 Numerische Berechnungen

Es bereitet keinerlei Schwierigkeiten, Zufallszahlen mit der in (2.13) angegebenen Verteilung zu erzeugen. Die Berechnung der Eigenwerte und -vektoren kann leicht programmiert werden, und da viele Statistikprogramme wie z. B. das Programm „Mathematica“ (Wolfram Research, Inc. , 1991) die Möglichkeit zur Erzeugung von normalverteilten bzw. sogar von nichtzentral χ^2-verteilten Zufallszahlen beinhalten, wird man vor kein großes Problem gestellt.

Etwas schwieriger ist es, die Verteilungsfunktion numerisch zu berechnen. Sei $G(t)$ die Verteilungsfunktion von nY^TY und $F(\nu, t)$ die Verteilungsfunktion einer zentral χ^2-verteilten Zufallsvariablen mit ν Freiheitsgraden, dann folgt aus den Resultaten von Ruben (1962) und Kotz et al. (1967), die die Verteilungsfunktion einer positiv semidefiniten quadratischen Form in multivariat normalverteilten Zufallsvariablen berechnen, für die gegebene Situation:

$$G(t) = \begin{cases} \sum_{i=0}^{\infty} c_i F(k-1+2i, t/\beta), & t > 0 \\ 0 & t \leq 0 \end{cases} \tag{2.14}$$

Wie β und die Koeffizienten c_i bestimmt werden, läßt sich bei Kotz et al. (1967) nachlesen. Der Algorithmus wurde bereits von Sheil & Muircheartaigh (1977) in ein Programm umgesetzt. Die Autoren geben eine Abschätzung des Fehlers an, falls die Reihe nach $i = I$ abgeschnitten wird.

Der genannte Algorithmus konvergiert sehr schnell, falls die Eigenwerte die gleiche Größenordnung besitzen. Ist dies nicht der Fall, kann es passieren, daß zur Berechnung eines Wertes der Verteilungsfunktion bis zur gewünschten Genauigkeit die Anzahl der benötigten Summanden rapide ansteigt. Für die konkrete Situation ergibt sich deshalb eine schnelle Konvergenz, falls sich π und $h(\theta_0)$ nicht allzuviel unterscheiden, für Extremfälle, wenn z. B. ein

h_i sehr klein ist, aber das zugehörige π_i nicht, konvergiert das Verfahren bedeutend langsamer.

2.3.6.3 Anpassung einer Gammaverteilung

Da sich die ersten beiden Momente der Approximation aus (2.13) berechnen lassen (Patnaik, 1949), bietet es sich wieder an, eine Gammaverteilung mit gleicher Erwartung bzw. Varianz anzupassen. Die Momente ergeben sich zu

$$\mu_1' = n\Delta(\pi, h(\theta_0)) + \sum_{i=1}^{k} \frac{\pi_i(1-\pi_i)}{h_i(\theta_0)} \tag{2.15}$$

$$\begin{aligned} \mu_2 &= 4(n-1)\sum_{i=1}^{k} \frac{\pi_i^3}{h_i^2(\theta_0)} - 2(2n-1)\left[\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)}\right]^2 \\ &\quad +2\sum_{i=1}^{k} \frac{\pi_i^2}{h_i^2(\theta_0)} \end{aligned} \tag{2.16}$$

Mit der Schreibweise aus (2.10) folgt für die Parameter α und β:

$$\alpha = \mu_1'^2/\mu_2 \quad \beta = \mu_2/\mu_1'$$

2.4 Vergleich der Approximationen, einfache Hypothese

2.4.1 Zusammenstellung der untersuchten Approximationen

Bevor der eigentliche Vergleich der approximierenden Verteilungen durchgeführt wird, sollen zwecks einer einfacheren Schreibweise für gegebene Werte von n, π, $h(\theta_0)$ und k die folgenden Definitionen getroffen werden:

X_{AN} : Zufallsvariable mit der Normalverteilung (2.5), basierend auf dem asymptotischen Ergebnis aus Abschnitt 2.3.3.

X_{NOR} : Zufallsvariable mit der Normalverteilung (2.6) von Broffitt & Randles (1977), Erwartung und Varianz stimmen mit den exakten Werten überein.

X_{NC} : Zufallsvariable mit der nichtzentralen χ^2-Verteilung (2.7) aus Abschnitt 2.3.4.

X_{GAM} : Zufallsvariable mit der Gammaverteilung aus Abschnitt 2.3.5, Erwartung und Varianz stimmen mit den exakten Werten überein.

X_{LKC} : Zufallsvariable mit der Verteilung der Linearkombination von χ^2-verteilten Zufallsvariablen (2.13) aus Abschnitt 2.3.6.

X_{GLK} : Zufallsvariable mit der Gammaverteilung aus Abschnitt 2.3.6, Erwartung und Varianz stimmen mit den entsprechenden Momenten von X_{LKC} überein.

χ^2 : χ^2-Prüfgröße

Es wird auch auf eine Folge von Zufallsvariablen Bezug genommen, bei denen sich mit der Stichprobengröße n auch das Grundmodell ändert. In diesem Fall wird z. B. die Schreibweise $X_{NC}^{(n)}$ verwendet.

2.4.2 Ein Vergleich aufgrund der Momente

2.4.2.1 Berechnung der Erwartungen und Varianzen

Obwohl aus der Übereinstimmung der ersten beiden Momente einer Approximation mit den exakten Momenten nicht zwingend eine gute Anpassung

der Verteilungsfunktion folgt, können Abweichungen bei den Momenten zumindest Hinweise auf die Güte der Anpassung geben. Die exakte Erwartung und die exakte Varianz (nach einer Umordnung der Summanden) sind nach (2.8) und (2.9) durch

$$E(\chi^2) = n\Delta(\pi, h(\theta_0)) + \sum_{i=1}^{k} \frac{\pi_i(1-\pi_i)}{h_i(\theta_0)} \quad \text{und} \tag{2.17}$$

$$\begin{aligned} Var(\chi^2) &= 4n \left[\sum_{i=1}^{k} \frac{\pi_i^3}{h_i^2(\theta_0)} - [\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)}]^2 \right] \\ &+10[\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)}]^2 - 12\sum_{i=1}^{k} \frac{\pi_i^3}{h_i^2(\theta_0)} - 4\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)} \sum_{i=1}^{k} \frac{\pi_i}{h_i(\theta_0)} + 6\sum_{i=1}^{k} \frac{\pi_i^2}{h_i^2(\theta_0)} \\ &+\frac{1}{n} \left[8\sum_{i=1}^{k} \frac{\pi_i^3}{h_i^2(\theta_0)} - 6[\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)}]^2 + 4\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)} \sum_{i=1}^{k} \frac{\pi_i}{h_i(\theta_0)} \right. \\ &\left. -6\sum_{i=1}^{k} \frac{\pi_i^2}{h_i^2(\theta_0)} - [\sum_{i=1}^{k} \frac{\pi_i}{h_i(\theta_0)}]^2 + \sum_{i=1}^{k} \frac{\pi_i}{h_i^2(\theta_0)} \right] \end{aligned} \tag{2.18}$$

gegeben.

Die Erwartungen der Approximationen berechnen sich zu

$$\begin{aligned} E(X_{AN}) &= n\Delta(\pi, h(\theta_0)) \\ E(X_{NC}) &= n\Delta(\pi, h(\theta_0)) + k - 1 \\ E(X_{LKC}) &= n\Delta(\pi, h(\theta_0)) + \sum_{i=1}^{k} \frac{\pi_i(1-\pi_i)}{h_i(\theta_0)} \qquad (2.19) \\ &= E(\chi^2) \\ E(X_{GLK}) &= E(X_{LKC}) = E(\chi^2). \qquad (2.20) \end{aligned}$$

Die Erwartungen von X_{GAM} und X_{NOR} stimmen natürlich mit der exakten Erwartung überein. Setzt man die Erwartungen in Beziehung zur exakten Erwartung, so folgt:

$$E(X_{AN}) = E(\chi^2) - \sum_{i=1}^{k} \frac{\pi_i(1-\pi_i)}{h_i(\theta_0)} \tag{2.21}$$

$$E(X_{NC}) = E(\chi^2) - \sum_{i=1}^{k} \frac{\pi_i(1-\pi_i)}{h_i(\theta_0)} + k - 1 \tag{2.22}$$

$$E(X_{LKC}) = E(X_{GLK}) = E(\chi^2) \tag{2.23}$$

Für die Varianzen erhält man:

$$Var(X_{AN}) = 4n\left[\sum_{i=1}^{k}\frac{\pi_i^3}{h_i^2(\theta_0)} - [\sum_{i=1}^{k}\frac{\pi_i^2}{h_i(\theta_0)}]^2\right] \tag{2.24}$$

$$Var(X_{NC}) = 4n\Delta(\pi, h(\theta_0)) + 2(k-1) \tag{2.25}$$

$$\begin{aligned} Var(X_{LKC}) &= Var(X_{GLK}) \\ &= 4(n-1)\sum_{i=1}^{k}\frac{\pi_i^3}{h_i^2(\theta_0)} - 2(2n-1)\left[\sum_{i=1}^{k}\frac{\pi_i^2}{h_i(\theta_0)}\right]^2 \\ &\quad +2\sum_{i=1}^{k}\frac{\pi_i^2}{h_i^2(\theta_0)} \\ &= 4n\left[\sum_{i=1}^{k}\frac{\pi_i^3}{h_i^2(\theta_0)} - [\sum_{i=1}^{k}\frac{\pi_i^2}{h_i(\theta_0)}]^2\right] \\ &\quad +2[\sum_{i=1}^{k}\frac{\pi_i^2}{h_i(\theta_0)}]^2 - 4\sum_{i=1}^{k}\frac{\pi_i^3}{h_i^2(\theta_0)} + 2\sum_{i=1}^{k}\frac{\pi_i^2}{h_i^2(\theta_0)} \end{aligned} \tag{2.26}$$

Die Varianzen von X_{GAM} und X_{NOR} sind natürlich identisch mit der exakten Varianz.

2.4.2.2 Die Momente von X_{AN} im Vergleich zu den exakten Momenten

Für X_{AN} ergibt sich nun, daß nach (2.21) die exakte Erwartung grundsätzlich unterschätzt wird. Die Differenz der Erwartungen ist vernachlässigbar, falls ein π_i dicht bei 1, die π_j, $j \neq i$, sehr klein sind und die Werte der $h_i(\theta_0)$, $i = 1, \ldots, k$, sich voneinander nicht allzusehr unterscheiden. Da für eine solche Konstellation $n\Delta(\pi, h(\theta))$ verhältnismäßig große Werte annimmt, sinkt auch die relative Abweichung der Erwartungen. Die Differenz der Erwartungen wird jedoch sehr groß, falls es ein $h_i(\theta_0)$ dicht bei 0 gibt und das zugehörige π_i einen moderaten Wert besitzt.

Ein Vergleich der Varianzen zeigt, daß $Var(X_{AN})$ nach (2.24) mit dem ersten Term aus (2.18) übereinstimmt und zumindest für $n \to \infty$ und $\pi \neq h(\theta_0)$ die relative Abweichung der Varianzen gegen 0 strebt. Zwar ist der Fall $\pi = h(\theta_0)$ für die asymptotische Ableitung aus Abschnitt 2.3.3 nicht zugelassen, in dieser Situation würde $Var(X_{AN})$ nach (2.24) verschwinden

und (2.18) sich zu

$$Var(\chi^2) = 2(k-1)(1-\frac{1}{n}) + \frac{1}{n}(\sum_{i=1}^{k}\frac{1}{\pi_i} - k^2) \tag{2.27}$$

vereinfachen, aber auch für eine geringe Diskrepanz zwischen π und $h(\theta_0)$ ist $Var(X_{AN})$ sehr klein, während $Var(\chi^2)$ Werte in der Nähe von $2(k-1)$ annimmt, größere Werte von n vorausgesetzt.

Aus den genannten Betrachtungen läßt sich nun die Schlußfolgerung ziehen, daß die exakten Momente besonders gut angenähert werden, wenn die π_i bis auf einen Wert sehr klein sind und die $h_i(\theta_0)$ eine moderate Größenordnung besitzen. Die Momente werden schlecht angepaßt, falls sich π und $h(\theta_0)$ nur wenig unterscheiden, und die Anpassung verschlechtert sich zusehends für den Fall einer oder mehrerer kleinerer Komponenten $h_i(\theta)$, relativ zu π_i gesehen.

2.4.2.3 Die Momente von X_{NC} im Vergleich zu den exakten Momenten

Die Gleichung (2.22) zeigt, daß sich die exakte Erwartung i. allg. von $E(X_{NC})$ unterscheidet, für $\pi = h(\theta_0)$ fallen die Erwartungen jedoch zusammen. Die exakte Erwartung wird durch $E(X_{NC})$ um höchstens $k-1$ überschätzt. Dieser Fall tritt angenähert ein, wenn die π_i bis auf einen Wert dicht bei 0 liegen und die $h_i(\theta_0)$ moderate Werte annehmen. Für bis auf einen Wert sehr kleine $h_i(\theta_0)$ und moderate π_i wird die exakte Erwartung jedoch sehr stark unterschätzt. Für $n \to \infty$ strebt die relative Abweichung der Erwartungen immer gegen null.

Für $\pi = h(\theta_0)$ vereinfacht sich $Var(X_{NC})$ zu

$$Var(X_{NC}) = 2(k-1), \tag{2.28}$$

und für größere Werte von n ist dann nach (2.27) der Unterschied der Varianzen vernachlässigbar, sofern die π_i moderate Werte annehmen. Falls sich π und $h(\theta_0)$ unterscheiden, verschwindet die relative Abweichung der Varianzen i. allg. selbst für $n \to \infty$ nicht. Für Frosini (1976), der allerdings nur die Hypothese $h_i(\theta_0) = 1/k$, $i = 1, \ldots, k$, betrachtet, ist dann der Quotient

$$\lim_{n\to\infty} \frac{Var(\chi^2)}{Var(X_{NC})} = \left[\sum_{i=1}^{k}\frac{\pi_i^3}{h_i^2(\theta_0)} - [\sum_{i=1}^{k}\frac{\pi_i^2}{h_i(\theta_0)}]^2\right] / \Delta(\pi, h(\theta_0)) \tag{2.29}$$

entscheidend. Er geht sogar so weit, die mögliche Qualität einer Approximation am Wert dieses Quotienten zu messen, da er bei seinen Untersuchungen feststellt, daß für verschiedene Paare von π und $h(\theta_0)$ mit relativ großer, aber annähernd gleicher Diskrepanz die Macht in den Fällen gut angenähert wird, in denen der Quotient dicht bei 1 liegt.

Für X_{NC} ergibt sich insgesamt eine gute Anpassung der Momente, falls sich π und $h(\theta_0)$ nicht allzusehr unterscheiden. Es ist jedoch nach den Ergebnissen von Frosini anzunehmen, daß der Wert der Diskrepanz $\Delta(\pi, h(\theta_0))$ für die Güte der Anpassung der Momente nicht die einzige einfließende Komponente ist.

2.4.2.4 Die Momente von X_{LKC} und X_{GLK} im Vergleich zu den exakten Momenten

Da X_{GLK} die gleichen Momente besitzt wie X_{LKC}, genügt es, sich auf X_{LKC} zu beschränken. Nach (2.19) stimmt die Erwartung von X_{LKC} mit der exakten Erwartung überein.

Die Varianz von X_{LKC} nimmt unter der Annahme $\pi = h(\theta_0)$ (X_{LKC} ist unter dieser Voraussetzung nach Abschnitt 2.3.6 χ^2_{k-1}-verteilt) den Wert $2(k-1)$ an und unterscheidet sich somit von der exakten Varianz aus (2.27) für größere n nur unwesentlich, es sei denn, es treten sehr kleine Werte π_i auf. Der Quotient aus der exakten Varianz und $Var(X_{LKC})$ strebt somit für $n \to \infty$ sowohl unter der Nullhypothese, als auch nach (2.18) und (2.26) unter jeder Alternative für $n \to \infty$ gegen eins. Der Vergleich der Varianzen für kleinere Stichproben läßt sich aufgrund der Komplexität von (2.18) und (2.26) nur für Spezialfälle durchführen.

2.4.3 Ein Vergleich aufgrund der asymptotischen Verteilung für eine feste Alternative

2.4.3.1 Einleitung

In Abschnitt 2.3.3 wurde gezeigt, daß für eine feste Alternative $\pi \neq h(\theta_0)$ die χ^2-Prüfgröße asymptotisch normalverteilt ist,

$$\begin{aligned} n\Delta(z, h(\theta_0)) &\sim AN(n\Delta(\pi, h(\theta_0)), n\sigma^2) \quad \text{bzw.} \\ n\Delta(z, h(\theta_0)) &\sim AN(\mu_1', \mu_2), \end{aligned}$$

wobei μ_1' und μ_2 die exakte Erwartung und Varianz bezeichnen und σ^2 wie in (2.4) definiert ist. Für die Zufallsvariablen X_{NOR} und X_{AN} bedeutet dies, daß sich ihre Verteilungsfunktionen mit wachsendem n der exakten Verteilungsfunktion der χ^2-Prüfgröße beliebig genau annähern lassen.

In den nächsten Unterabschnitten wird überprüft, ob sich diese Aussage auf die restlichen Approximationen übertragen läßt. Es genügt zu zeigen, daß für eine gegebene Zufallsvariable X_{App} eine der beiden asymptotischen Normalverteilungen zutrifft, nach Lemma A von Serfling (1980, S. 20) ist dann automatisch die andere erfüllt.

Die weiter unten angegebenen momenterzeugenden Funktionen lassen sich bei Johnson & Kotz (1970, Vol. 1 und Vol. 2) nachschlagen.

2.4.3.2 Asymptotische Verteilung von X_{NC} für eine feste Alternative

Für gegebene Werte von π und $h(\theta_0)$ gilt nach (2.7):

$$X_{NC} \sim \chi^2_{k-1}(n\Delta(\pi, h(\theta_0)))$$

Sei $\Delta := \Delta(\pi, h(\theta_0))$ und

$$Y_n = \frac{X_{NC} - n\Delta}{\sqrt{n}\sigma},$$

dann ist zu zeigen:

$$Y_n \xrightarrow{d} N(0,1)$$

Die momenterzeugende Funktion von Y_n ergibt sich zu

$$\begin{aligned} M_n(t) &= \exp(-\sqrt{n}\frac{\Delta}{\sigma}t)\left(\frac{1}{1-\frac{2t}{\sqrt{n}\sigma}}\right)^{\frac{k-1}{2}} \exp\left(\frac{\sqrt{n}\frac{\Delta}{\sigma}t}{1-\frac{2t}{\sqrt{n}\sigma}}\right) \\ &= \exp\left(\frac{2t^2\Delta}{\sigma^2}\frac{1}{1-\frac{2t}{\sqrt{n}\sigma}}\right)\left(\frac{1}{1-\frac{2t}{\sqrt{n}\sigma}}\right)^{\frac{k-1}{2}}. \end{aligned}$$

Für $n \to \infty$ folgt nun

$$\lim_{n\to\infty} M_n(t) = \exp(\frac{2t^2\Delta}{\sigma^2}).$$

Nur für $4\Delta = \sigma^2$ wird die exakte Verteilungsfunktion der χ^2-Prüfgröße für wachsendes n immer genauer angenähert. Ein Vergleich mit (2.29) zeigt, daß diese Bedingung dann und nur dann erfüllt ist, wenn der von Frosini betrachtete Quotient der Varianzen gegen 1 strebt.

2.4.3.3 Asymptotische Verteilung von X_{GAM} für eine feste Alternative

Die Parameter der Gammaverteilung werden in diesem Fall über die Momente angepaßt. Die momenterzeugende Funktion von X_{GAM} lautet

$$M(t) = (1 - \beta t)^{-\alpha},$$

mit $\alpha = (\mu_1')^2/\mu_2$ und $\beta = \mu_2/\mu_1'$. Im vorliegenden Fall ist es einfacher, die Folge der Zufallsvariablen

$$Y_n = \frac{X_{GAM} - \mu_1'}{\sqrt{\mu_2}}$$

zu betrachten. Die momenterzeugende Funktion von Y_n ergibt sich zu

$$M_n(t) = \exp(-\frac{\mu_1'}{\sqrt{\mu_2}}t)\left(\frac{1}{1 - \frac{\sqrt{\mu_2}}{\mu_1'}t}\right)^{\frac{(\mu_1')^2}{\mu_2}}.$$

Für $n \to \infty$ strebt $(\mu_1')/\sqrt{\mu_2}$ gegen unendlich. Es genügt also,

$$\begin{aligned}\lim_{c\to\infty} \exp(-ct)\left(\frac{1}{1-\frac{t}{c}}\right)^{c^2} &= \lim_{c\to\infty} \exp(-ct)\left(\frac{c}{c-t}\right)^{c^2} \\ &= \lim_{c\to\infty} \exp(c(c\log\frac{c}{c-t} - t))\end{aligned}$$

zu betrachten. Nun zeigt man über eine Reihenentwicklung

$$\log(1+x) = x - \frac{x^2}{2} + o(x^2), \quad x \to 0,$$

und dann folgt für den Grenzwert des Exponenten:

$$\begin{aligned}\lim_{c\to\infty} c(c\log\frac{c}{c-t} - t) &= \lim_{c\to\infty}\left[\frac{c^2t}{c-t} - \frac{t^2c^2}{2(c-t)^2} - ct + o(\frac{c^2t^2}{(c-t)^2})\right] \\ &= \lim_{c\to\infty}(\frac{1}{2}t^2 - \frac{t^3}{c})/(1-\frac{t}{c})^2 + o(1) \\ &= \frac{1}{2}t^2\end{aligned}$$

Für die momenterzeugende Funktion von Y_n gilt somit:

$$\lim_{n\to\infty} M_n(t) = \exp(\frac{t^2}{2})$$

Y_n ist asymptotisch standardnormalverteilt, die Verteilung von X_{GAM} nähert sich der exakten Verteilung bei wachsender Stichprobengröße an.

2.4.3.4 Asymptotische Verteilung von X_{GLK} für eine feste Alternative

Die Parameter der Gammaverteilung werden hier durch die Momente der Zufallsvariablen X_{LKC} bestimmt. Betrachtet man zunächst die Folge der Zufallsvariablen

$$Y_n = \frac{X_{GLK} - E(X_{GLK})}{\sqrt{Var(X_{GLK})}},$$

so folgt analog zum Vorgehen des letzten Unterabschnitts für die momenterzeugende Funktion von Y_n:

$$\lim_{n\to\infty} M_n(t) = \exp(\frac{t^2}{2}).$$

Nun stimmt nach (2.20) $E(X_{GLK})$ mit der exakten Erwartung μ_1' überein und weiterhin ist

$$\lim_{n\to\infty} \frac{Var(X_{GLK})}{\mu_2} = 1.$$

Unter Verwendung des Lemmas A von Serfling (1980, S. 20) folgt

$$X_{GLK} \sim AN(\mu_1', \mu_2),$$

die Verteilung von X_{GLK} eignet sich, zumindest für großes n, als Approximation der exakten Verteilung der χ^2-Prüfgröße.

2.4.3.5 Asymptotische Verteilung von X_{LKC} für eine feste Alternative

Mit den Bezeichnungen des Abschnitts 2.3.6 läßt sich X_{LKC} als

$$X_{LKC} = \sum_{i=1}^{k-1} \lambda_i \chi_1^2(\frac{nc_i^2}{\lambda_i})$$

darstellen. Die $c_i := (Sw^*(\pi))_i, i = 1, \ldots, k-1$, erfüllen die Nebenbedingung

$$\sum_{i=1}^{k-1} c_i^2 = \Delta(\pi, h(\theta_0)). \tag{2.30}$$

Sei D_λ die $k \times k$-Diagonalmatrix mit den Eigenwerten λ_i, $i = 1, \ldots, k$, und $\lambda_k = 0$, dann folgt weiterhin

$$\begin{aligned}
4\sum_{i=1}^{k-1} \lambda_i c_i^2 &= 4\sum_{i=1}^{k} \lambda_i c_i^2 \\
&= 4(w^*(\pi))^T S^T D_\lambda S w^*(\pi) \\
&= 4(\pi - h(\theta_0))^T D_h^{-\frac{1}{2}} (D_h^{-\frac{1}{2}} (D_\pi - \pi\pi^T) D_h^{-\frac{1}{2}}) D_h^{-\frac{1}{2}} (\pi - h(\theta_0)) \\
&= 4(\pi - h(\theta_0))^T D_h^{-1} (D_\pi - \pi\pi^T) D_h^{-1} (\pi - h(\theta_0)) \\
&= 4\pi^T D_h^{-1} (D_\pi - \pi\pi^T) D_h^{-1} \pi = \sigma^2
\end{aligned} \tag{2.31}$$

Man betrachte nun die Folge

$$Y_n = \frac{X_{LKC} - n\Delta}{\sqrt{n}}.$$

Die momenterzeugende Funktion für Y_n ist durch

$$\begin{aligned}
M_n(t) &= \exp(-\sqrt{n}\Delta t) \exp(-\frac{1}{2}\sum_{i=1}^{k-1} \frac{nc_i^2}{\lambda_i}) \exp(\frac{1}{2}\sum_{i=1}^{k-1} \frac{nc_i^2}{\lambda_i(1 - 2t\lambda_i/\sqrt{n})}) \\
&\quad * \prod_{i=1}^{k-1} \left(1 - \frac{2t\lambda_i}{\sqrt{n}}\right)^{-\frac{1}{2}}
\end{aligned}$$

gegeben (vgl. Johnson & Kotz, 1970, Vol. 2). Wird Gleichung (2.30) ausgenutzt, so ergibt sich nach einigen Umformungen

$$M_n(t) = \exp\left(\sum_{i=1}^{k-1} \frac{2\lambda_i c_i^2 t^2}{1 - 2\lambda_i t/\sqrt{n}}\right) \prod_{i=1}^{k-1} \left(1 - \frac{2t\lambda_i}{\sqrt{n}}\right)^{-\frac{1}{2}}.$$

Man sieht nun leicht, daß unter Verwendung von (2.31)

$$\lim_{n\to\infty} M_n(t) = \exp(\frac{t^2}{2}(4\sum_{i=1}^{k-1} \lambda_i c_i^2)) = \exp(\frac{\sigma^2 t^2}{2})$$

folgt. Somit gilt:

$$X_{LKC} \sim AN(n\Delta(\pi, h(\theta_0)), n\sigma^2)$$

Die Verteilung von X_{LKC} eignet sich somit zur Approximation der exakten Verteilung.

2.4.4 Ein Vergleich aufgrund der Eignung zur Approximation der Macht, Verteilung unter „local alternatives“

2.4.4.1 Einleitung

In den letzten Jahren sind einige Simulationen durchgeführt worden, um die Eignung von Verteilungen zur Powerapproximation zu überprüfen. Während die meisten Standardwerke immer noch die nichtzentrale χ^2-Verteilung als wichtigste Approximation nennen (z. B. Johnson & Kotz, 1970, Vol. 2), sind einige Autoren diesbezüglich anderer Meinung.

Frosini (1976) hält seine Approximation mittels einer Gammaverteilung für empfehlenswert, denn sie sei der nichtzentralen χ^2-Verteilung mindestens ebenbürtig, in manchen Fällen sogar deutlich überlegen. Die Ergebnisse von Montinaro (1988) stützen diese Aussage.

Broffitt & Randles (1977) vergleichen die Eignung der Verteilungen von X_{NC} bzw. von X_{NOR} zur Approximation der wahren Macht. Sie kommen zu dem Schluß, daß sich zumindest für große Diskrepanzen zwischen π und $h(\theta_0)$ die nichtzentrale χ^2-Verteilung nicht anwenden läßt.

Drost et al. (1989) untersuchen die Verteilungen von X_{LKC} und X_{NC} sowie einigen weiteren Zufallsvariablen. Sie empfehlen für genaue Berechnungen die LKC-Powerapproximation.

Leider wird die Güte der einzelnen Approximationen von den genannten Autoren auf unterschiedliche Art und Weise gemessen, und ein direkter Vergleich ist somit nicht möglich. Außerdem hat noch kein Autor die Verteilungen von X_{LKC} und X_{GAM}, der erfolgversprechendsten Approximationen, direkt gegenübergestellt.

Das entscheidende Manko der meisten genannten Artikel (nur Drost et al. bilden die Ausnahme) ist jedoch das Fehlen einer Untersuchung, ob sich die einzelnen Verteilungen *prinzipiell* zur Approximation der wahren Macht eignen. Es wird vorbehaltlos davon ausgegangen, daß für wachsende Stichproben die exakte Macht, auch wenn sie deutlich kleiner als 1 ist, genauer bestimmt werden kann. Für einige der genannten Verteilungen ist dies jedoch ein Trugschluß, wie nun gezeigt werden soll.

Für die Zufallsvariablen X_{AN}, X_{NOR}, X_{LKC}, X_{GLK} und X_{GAM} wurde bewiesen, daß sich für wachsende Stichprobengrößen die Verteilungsfunktionen an die exakte Verteilungsfunktion der χ^2-Prüfgröße annähern. Dabei wird jedoch von einem festen Grundmodell $\pi \neq h(\theta_0)$ ausgegangen. Für die Po-

werapproximation folgt hieraus nur die wenig hilfreiche Aussage

$$\begin{aligned}
\lim_{n\to\infty} P(n\Delta(z^{(n)}, h(\theta_0)) > c_\alpha) &= \lim_{n\to\infty} P(X_{AN} > c_\alpha) \\
&= \lim_{n\to\infty} P(X_{NOR} > c_\alpha) \\
&= \lim_{n\to\infty} P(X_{LKC} > c_\alpha) \\
&= \lim_{n\to\infty} P(X_{GLK} > c_\alpha) \\
&= \lim_{n\to\infty} P(X_{GAM} > c_\alpha) = 1,
\end{aligned}$$

wobei c_α den $1-\alpha$-Prozentpunkt der χ^2_{k-1}-Verteilung bezeichnet. Für Folgen von X_{NC} läßt sich eine äquivalente Aussage herleiten, ohne daß auf eine allmähliche Anpassung der Verteilungsfunktionen Bezug genommen werden muß.

Möchte man interessantere Resultate über die Eignung der Verteilungen zur Approximation der Macht bei wachsender Stichprobengröße erhalten, so darf das Grundmodell π nicht festgehalten werden, sondern es muß eine Folge von Grundmodellen betrachtet werden, die in geeigneter Weise gegen das hypothetische Modell $h(\theta_0)$ konvergiert. Ein Weg zur Konstruktion einer solchen Folge ist nun gerade durch den Ansatz der „local alternatives“ gegeben. Für

$$\pi^{(n)} = h(\theta_0) + \frac{c}{\sqrt{n}} \quad , c \in I\!R^k, \sum_{i=1}^{k} c_i = 0,$$

folgte in Abschnitt 2.3.4

$$n\Delta(z^{(n)}, h(\theta_0)) \xrightarrow{d} \chi^2_{k-1}(\sum_{i=1}^{k} \frac{c_i^2}{h_i(\theta_0)}).$$

Für die Macht bedeutet dies, daß

$$\lim_{n\to\infty} P(n\Delta(z^{(n)}, h(\theta_0)) > c_\alpha) = P(\chi^2_{k-1}(\sum_{i=1}^{k} \frac{c_i^2}{h_i(\theta_0)}) > c_\alpha)$$

ist. Nichts liegt nun näher, als für die einzelnen Verteilungen zu untersuchen, wie sich

$$P(X^{(n)}_{App} > c_\alpha), \quad App \in \{AN, NOR, NC, GAM, LKC, GLK\},$$

für $n \to \infty$ unter „local alternatives“ verhält. Dieser Ansatz stellt ein überaus nützliches Hilfsmittel zum Verständnis der Funktionsweise der verschiedenen Approximationen dar und gibt darüberhinaus wichtige Hinweise, in welchen Situationen welche Approximation zu bevorzugen ist.

2.4.4.2 Die asymptotische Verteilung von X_{AN} bei „local alternatives“

Für gegebene Werte von π, $h(\theta_0)$ und n ist die Verteilung von X_{AN} nach (2.5) durch die Normalverteilung

$$N(n\Delta(\pi, h(\theta_0)), n\sigma^2)$$

mit

$$\sigma^2 = 4\left[\sum_{i=1}^{k} \frac{\pi_i^3}{h_i^2(\theta_0)} - \left(\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)}\right)^2\right]$$

gegeben. Für eine Folge

$$\pi^{(n)} = h(\theta_0) + \frac{c}{\sqrt{n}}, \quad c \in I\!R^k \setminus \{\mathbf{0}\}, \sum_{i=1}^{k} c_i = 0,$$

ergibt sich somit:

$$\begin{aligned} E(X_{AN}^{(n)}) &= n\sum_{i=1}^{k} \frac{(\pi_i^{(n)} - h_i(\theta_0))^2}{h_i(\theta_0)} \\ Var(X_{AN}^{(n)}) &= 4n\left[\sum_{i=1}^{k} \frac{(\pi_i^{(n)})^3}{h_i^2(\theta_0)} - \left(\sum_{i=1}^{k} \frac{(\pi_i^{(n)})^2}{h_i(\theta_0)}\right)^2\right] \end{aligned}$$

Elementare Berechnungen ergeben nun:

$$\begin{aligned} \lim_{n\to\infty} E(X_{AN}^{(n)}) &= \sum_{i=1}^{k} \frac{c_i^2}{h_i(\theta_0)} \\ \lim_{n\to\infty} Var(X_{AN}^{(n)}) &= 4\sum_{i=1}^{k} \frac{c_i^2}{h_i(\theta_0)} \end{aligned}$$

Da die momenterzeugende Funktion einer Normalverteilung stetig in den Parametern ist, folgt somit

$$X_{AN}^{(n)} \xrightarrow{d} N(\delta, 4\delta) \tag{2.32}$$

mit $\delta = \sum_{i=1}^{k} c_i^2/h_i(\theta_0)$. Dies ist jedoch nicht die Grenzverteilung der χ^2-Prüfgröße, d. h. selbst für beliebig großes n wird die exakte Macht falsch approximiert.

2.4.4.3 Die asymptotische Verteilung von X_{NOR} bei „local alternatives“

Die Verteilung von X_{NOR} wurde über die Momente angepaßt. Für die Folge $\pi^{(n)}$ ergibt sich deshalb (vgl. (2.8) und (2.9))

$$E(X_{NOR}^{(n)}) \quad = \quad n\Delta(\pi^{(n)}, h(\theta_0)) + \sum_{i=1}^{k} \frac{\pi_i^{(n)}(1-\pi_i^{(n)})}{h_i(\theta_0)}$$

und

$$\begin{aligned}
&Var(X_{NOR}^{(n)}) = \\
&1/n\{(n-1)(6-4n)[\sum_{i=1}^{k} \frac{(\pi_i^{(n)})^2}{h_i(\theta_0)}]^2 + 4(n-1)(n-2)\sum_{i=1}^{k} \frac{(\pi_i^{(n)})^3}{h_i^2(\theta_0)} \\
&-4(n-1)\sum_{i=1}^{k} \frac{(\pi_i^{(n)})^2}{h_i(\theta_0)} \sum_{i=1}^{k} \frac{\pi_i^{(n)}}{h_i(\theta_0)} + 6(n-1)\sum_{i=1}^{k} \frac{(\pi_i^{(n)})^2}{h_i^2(\theta_0)} \\
&-[\sum_{i=1}^{k} \frac{\pi_i^{(n)}}{h_i(\theta_0)}]^2 + \sum_{i=1}^{k} \frac{\pi_i^{(n)}}{h_i^2(\theta_0)}\}.
\end{aligned}$$

Mit

$$\delta = \sum_{i=1}^{k} \frac{c_i^2}{h_i(\theta_0)}$$

ergibt sich:

$$\lim_{n\to\infty} E(X_{NOR}^{(n)}) \quad = \quad k-1+\delta \tag{2.33}$$

$$\lim_{n\to\infty} Var(X_{NOR}^{(n)}) \quad = \quad 2(k-1+2\delta) \tag{2.34}$$

Hieraus folgt sofort

$$X_{NOR}^{(n)} \xrightarrow{d} N(k-1+\delta, 2(k-1+2\delta))$$

Die asymptotische Verteilung unter „local alternatives“ ist somit nicht korrekt.

2.4.4.4 Die asymptotische Verteilung von X_{GAM} bei „local alternatives“

X_{GAM} besitzt ebenfalls Erwartung und Varianz der exakten Verteilung. Setzt man, analog zu (2.13),

$$\begin{aligned} \alpha_n &= \frac{E(X_{GAM}^{(n)})^2}{Var(X_{GAM}^{(n)})} \quad \text{und} \\ \beta_n &= \frac{Var(X_{GAM}^{(n)})}{E(X_{GAM}^{(n)})}, \end{aligned}$$

so folgt aus (2.33) und (2.34) sowie der Stetigkeit der momenterzeugenden Funktion der Gammaverteilung in ihren Parametern

$$X_{GAM}^{(n)} \xrightarrow{d} G(\alpha, \beta),$$

mit

$$\begin{aligned} \alpha &= \frac{(k-1+\delta)^2}{2(k-1+2\delta)}, \\ \beta &= \frac{2(k-1+2\delta)}{k-1+\delta}. \end{aligned} \tag{2.35}$$

Auch für diese Verteilung erhält man nicht die korrekte asymptotische Verteilung unter „local alternatives“.

2.4.4.5 Die asymptotische Verteilung von X_{LKC} bei „local alternatives“

Für gegebenes $\pi^{(n)}$ ist $X_{LKC}^{(n)}$ nach (2.13) verteilt wie die Linearkombination

$$\sum_{i=1}^{k-1} \lambda_i^{(n)} \chi_1^2 \left(\frac{n(S^{(n)} w^*(\pi^{(n)}))_i^2}{\lambda_i^{(n)}} \right),$$

wobei $S^{(n)}$ eine orthogonale Matrix ist, deren transponierte Zeilen aus einer orthonormalen Basis von Eigenvektoren der Matrix

$$V^{(n)} = D_h^{-\frac{1}{2}} (D_{\pi^{(n)}} - (\pi^{(n)})(\pi^{(n)})^T) D_h^{-\frac{1}{2}}$$

bestehen. Die $\lambda_i^{(n)}$ sind die zugehörigen Eigenwerte, und $w^*(\pi^{(n)})$ ist durch

$$w^*(\pi^{(n)}) = D_h^{-\frac{1}{2}}(\pi^{(n)} - h(\theta_0))$$

gegeben. Die charakteristische Funktion von $X_{LKC}^{(n)}$ lautet (siehe z. B. Johnson & Kotz, 1970, Vol. 2):

$$\phi_n(t) = \exp\left(-\frac{1}{2}\sum_{i=1}^{k-1} d_i^{(n)}\right)\exp\left(\frac{1}{2}\sum_{i=1}^{k-1}\frac{d_i^{(n)}}{1-2it\lambda_i^{(n)}}\right)\prod_{i=1}^{k-1}(1-2it\lambda_i^{(n)})^{-\frac{1}{2}},$$

wobei

$$d_i^{(n)} := \frac{n(S^{(n)}w^*(\pi^{(n)}))_i^2}{\lambda_i^{(n)}}$$

ist. Für $n \to \infty$ ergibt sich nun

$$\lim_{n\to\infty} V^{(n)} = V^* = I - \sqrt{h(\theta_0)}\sqrt{h(\theta_0)}^T,$$

und da V^* idempotent mit Spur $k-1$ ist und die Eigenwerte einer Matrix stetig von den Komponenten abhängen, folgt daraus

$$\lim_{n\to\infty} \lambda_i^{(n)} = 1, \quad i = 1, \ldots, k-1.$$

Nutzt man noch die Bedingung $(S^{(n)}w^*(\pi^{(n)}))_k = 0$ aus, folgt weiterhin

$$\begin{aligned}
\lim_{n\to\infty}\sum_{i=1}^{k-1} n(S^{(n)}w^*(\pi^{(n)}))_i^2 &= \lim_{n\to\infty}(\sqrt{n}w^*(\pi^{(n)}))^T(S^{(n)})^T S^{(n)}(\sqrt{n}w^*(\pi^{(n)})) \\
&= \lim_{n\to\infty} n w^*(\pi^{(n)})^T w^*(\pi^{(n)}) \\
&= \lim_{n\to\infty} n(\pi^{(n)} - h(\theta_0))^T D_h^{-1}(\pi^{(n)} - h(\theta_0)) \\
&= \sum_{i=1}^{k}\frac{c_i^2}{h_i(\theta_0)} = \delta.
\end{aligned}$$

Für die charakteristische Funktion ergibt sich hiermit

$$\begin{aligned}
\lim_{n\to\infty}\phi_n(t) &= \exp\left(-\frac{1}{2}\delta\right)\exp\left(\frac{1}{2(1-2it)}\delta\right)(1-2it)^{-\frac{k-1}{2}} \\
&= (1-2it)^{-\frac{k-1}{2}}\exp\left[\frac{it\delta}{1-2it}\right].
\end{aligned}$$

Dies ist aber die charakteristische Funktion einer $\chi^2_{k-1}(\delta)$- verteilten Zufallsvariable (siehe z. B. Johnson & Kotz, 1970, Vol. 2). Somit folgt:

$$X^{(n)}_{LKC} \xrightarrow{d} \chi^2_{k-1}(\delta)$$

Die aus dem Ansatz von Drost et al. (1989) gewonnene Verteilung läßt sich also auch für „local alternatives“ anwenden.

2.4.4.6 Die asymptotische Verteilung von X_{GLK} bei „local alternatives“

Die gammaverteilte Zufallsvariable X_{GLK} besitzt die gleiche Erwartung und Varianz wie X_{LKC}. Wird in die Formeln (2.15) und (2.16)

$$\pi = \pi^{(n)} = h(\theta_0) + \frac{c}{\sqrt{n}}$$

gesetzt, so folgt nach einer kurzen Rechnung

$$\begin{aligned} \lim_{n\to\infty} E(X^{(n)}_{GLK}) &= k-1+\delta \quad \text{bzw.} \\ \lim_{n\to\infty} Var(X^{(n)}_{GLK}) &= 2(k-1+2\delta). \end{aligned}$$

Die asymptotische Verteilung von $X^{(n)}_{GLK}$ ist deshalb mit der asymptotischen Verteilung von $X^{(n)}_{GAM}$ unter „local alternatives“ identisch.

2.4.4.7 Die asymptotische Verteilung von X_{NC} bei „local alternatives“

Aufgrund der Konstruktion der „local alternatives“ folgt für die asymptotische Verteilung von

$$X^{(n)}_{NC} = \chi^2_{k-1}(n\Delta(\pi^{(n)}, h(\theta_0)))$$

sofort

$$X^{(n)}_{NC} \xrightarrow{d} \chi^2_{k-1}(\delta).$$

2.4.4.8 Interpretation der asymptotischen Verteilungen unter „local alternatives“

Die bisherigen Ergebnisse über die asymptotischen Verteilungen zeigen, daß nur die Verteilung von X_{LKC} uneingeschränkt geeignet ist, die exakte Verteilung der χ^2-Prüfgröße sowohl unter einer festen Alternative als auch unter „local alternatives“ für wachsende Stichprobengrößen immer genauer zu approximieren. Verlangt man von einer Approximation, daß sie für größer werdende Stichproben Werte der Macht, die deutlich kleiner als 1 sind, immer genauer annähert, so ist außerdem nur noch die Zufallsvariable X_{NC} geeignet. Für diese beiden Zufallsvariablen ist es nur eine Frage der Größenordnung der Stichprobe, bei der eine Approximation der Macht zufriedenstellende Ergebnisse liefert.

Für die übrigen Approximationen gilt, daß sie für „local alternatives“ in dem Sinne versagen, daß die exakte Macht nicht beliebig genau approximiert werden kann. Trotzdem können fast alle mit gewissen Einschränkungen verwendet werden, wie die folgenden Überlegungen zeigen:

Die exakte Verteilung der χ^2-Prüfgröße strebe unter „local alternatives“ gegen eine $\chi^2_{k-1}(\delta)$-Verteilung. Die Grenzverteilungen von $X^{(n)}_{GAM}$, $X^{(n)}_{GLK}$ und $X^{(n)}_{NOR}$ sind durch eine Gammaverteilung bzw. eine Normalverteilung gegeben, deren Erwartungen und Varianzen mit den entsprechenden Größen der nichtzentralen χ^2-Verteilung übereinstimmen.

Unterscheiden sich diese Grenzverteilungen nur wenig von der nichtzentralen χ^2-Verteilung, so wird auch der Grenzwert der exakten Macht

$$\beta = P(\chi^2_{k-1}(\delta) > c_\alpha)$$

nur geringfügig von den entsprechenden Grenzwerten der Approximationen abweichen.

Nun ist bekannt (siehe z. B. Johnson & Kotz, 1970, Vol. 2), daß sich die nichtzentrale χ^2-Verteilung in vielen Situationen durch eine Gammaverteilung approximieren läßt. Der Fehler, der durch die Anpassung einer Gammaverteilung entsteht, sollte sich deshalb in den meisten Fällen nicht entscheidend bemerkbar machen. Die Normalverteilung mit Erwartung und Varianz der $\chi^2_{k-1}(\delta)$-Verteilung eignet sich nur dann als Approximation, falls die Anzahl der Freiheitsgrade groß ist.

Eine nützliche Hilfe zur Beurteilung der voraussichtlichen Eignung einer Approximation für gegebene Werte von α und k ergibt sich über die Analyse der Grenzverteilungen. Dazu führe man die folgenden Schritte durch:

Schritt 1: Berechnung des kritischen Wertes c_α mit $P(\chi^2_{k-1} > c_\alpha) = \alpha$.

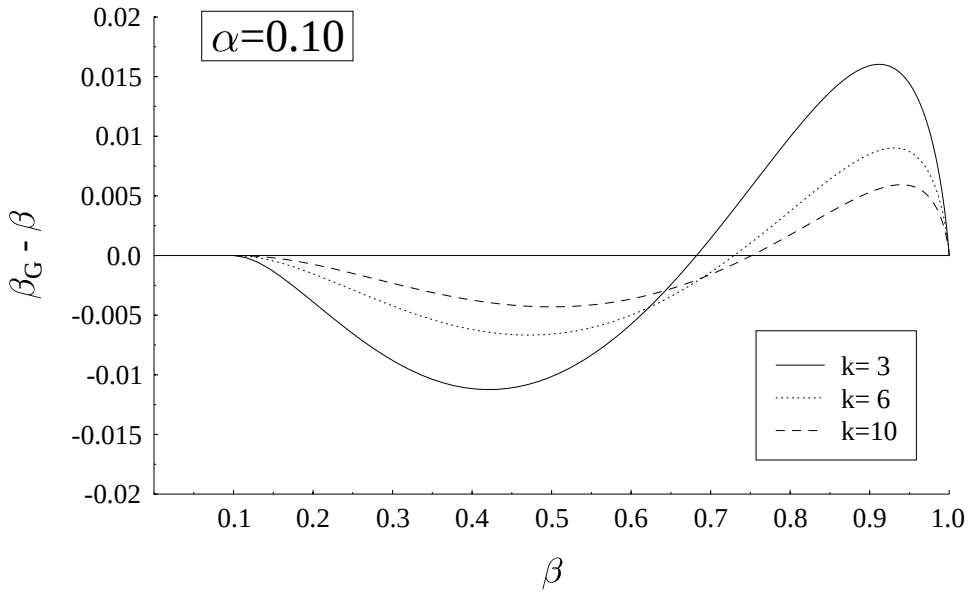
Schritt 2: Für β mit $\alpha < \beta < 1$ berechne man δ mit $P(\chi^2_{k-1}(\delta) > c_\alpha) = \beta$.

Schritt 3: Berechnung von $\beta_G = P(G_{k,\delta} > c_\alpha)$, beziehungsweise von $\beta_N = P(N_{k,\delta} > c_\alpha)$.
Bemerkung: $G_{k,\delta}$ bezeichne eine gammaverteilte Zufallsvariable, deren Parameter durch die Gleichungen aus (2.35) gegeben sind. Analog sei $N_{k,\delta}$ eine normalverteilte Zufallsvariable mit Erwartung und Varianz wie in (2.33) und (2.34).

Werden die Schritte 2 und 3 für verschiedene Werte von β wiederholt, so gewinnt man einen umfassenden Eindruck, für welche Bereiche der Macht gute Ergebnisse zu erwarten sind. Es darf natürlich nicht vergessen werden, daß es sich hier nur um asymptotische Aussagen handelt, und möglicherweise weichen die für kleineres n erhaltenen Schätzer eines Wertes der exakten Macht deutlich von den zugehörigen Grenzwerten ab. Die Abbildungen 2.1 und 2.2 zeigen nun die Differenzen $\beta_G - \beta$ bzw. $\beta_N - \beta$ in Abhängigkeit von β für die Werte von $k = 3$, $k = 6$ und $k = 10$, wobei die Signifikanzniveaus $\alpha = 0.10$ und $\alpha = 0.01$ betrachtet wurden. Man beachte dabei die unterschiedlichen Maßstäbe.

Diese Grafiken zeigen, daß sich mit wachsender Zellenzahl k die Verteilungen von X_{GAM} und X_{GLK} bzw. von X_{NOR} immer besser zur Powerapproximation eignen, zumindest aus dem asymptotischen Blickwinkel der „local alternatives“. X_{GAM} und X_{GLK} eignen sich schon für $k = 3$, X_{NOR} sollte nur dann zur Powerapproximation herangezogen werden, wenn die Zellenzahl deutlich größer als 10 ist. Natürlich darf nicht vergessen werden, daß die Genauigkeit der Schätzung der exakten Macht im Gegensatz zur X_{LKC}-Approximation mit wachsender Stichprobengröße nicht beliebig erhöht werden kann.

Interessant ist in diesem Zusammenhang die Grafik 2.3, die aus Broffitt & Randles (1977) entnommen ist. Sie zeigt für neun Alternativen die Fehler der X_{NOR}- bzw. der X_{NC}-Powerapproximation in Abhängigkeit von der exakten Macht. Für die relativ kleine Stichprobengröße von $n = 45$ ist die Übereinstimmung der aus diesen Punkten gebildeten EN-Kurve mit der entsprechenden Kurve aus Abbildung 2.2 ($k = 3$, $\alpha = 0.01$) erstaunlich. Broffitt & Randles vermuten fälschlicherweise, daß sich für eine größere Stichprobe die EN-Kurve der Abszisse nähert.



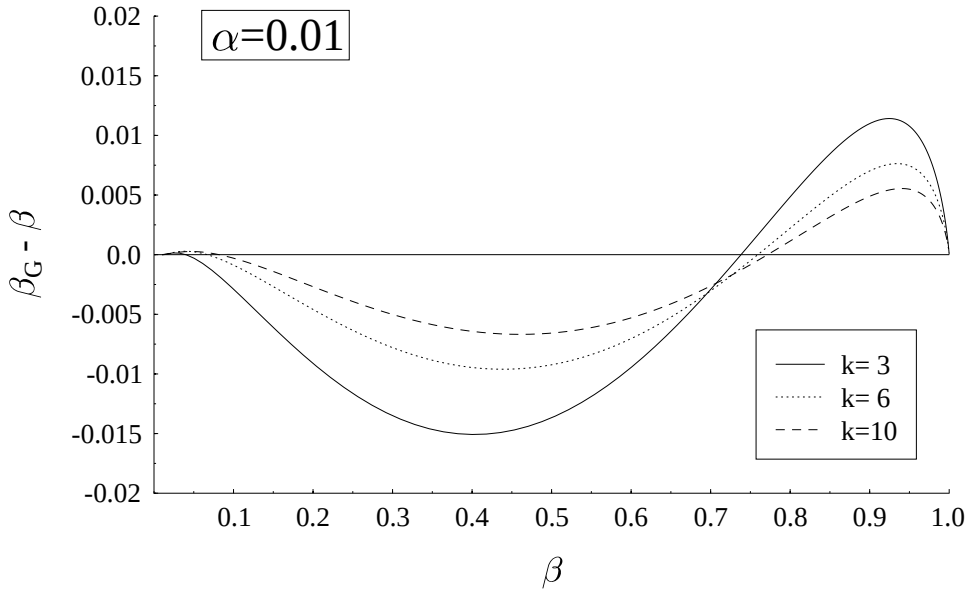


Abb. 2.1: Plot des Fehlers $\beta_G - \beta$ in Abhängigkeit von der asymptotischen Macht β.

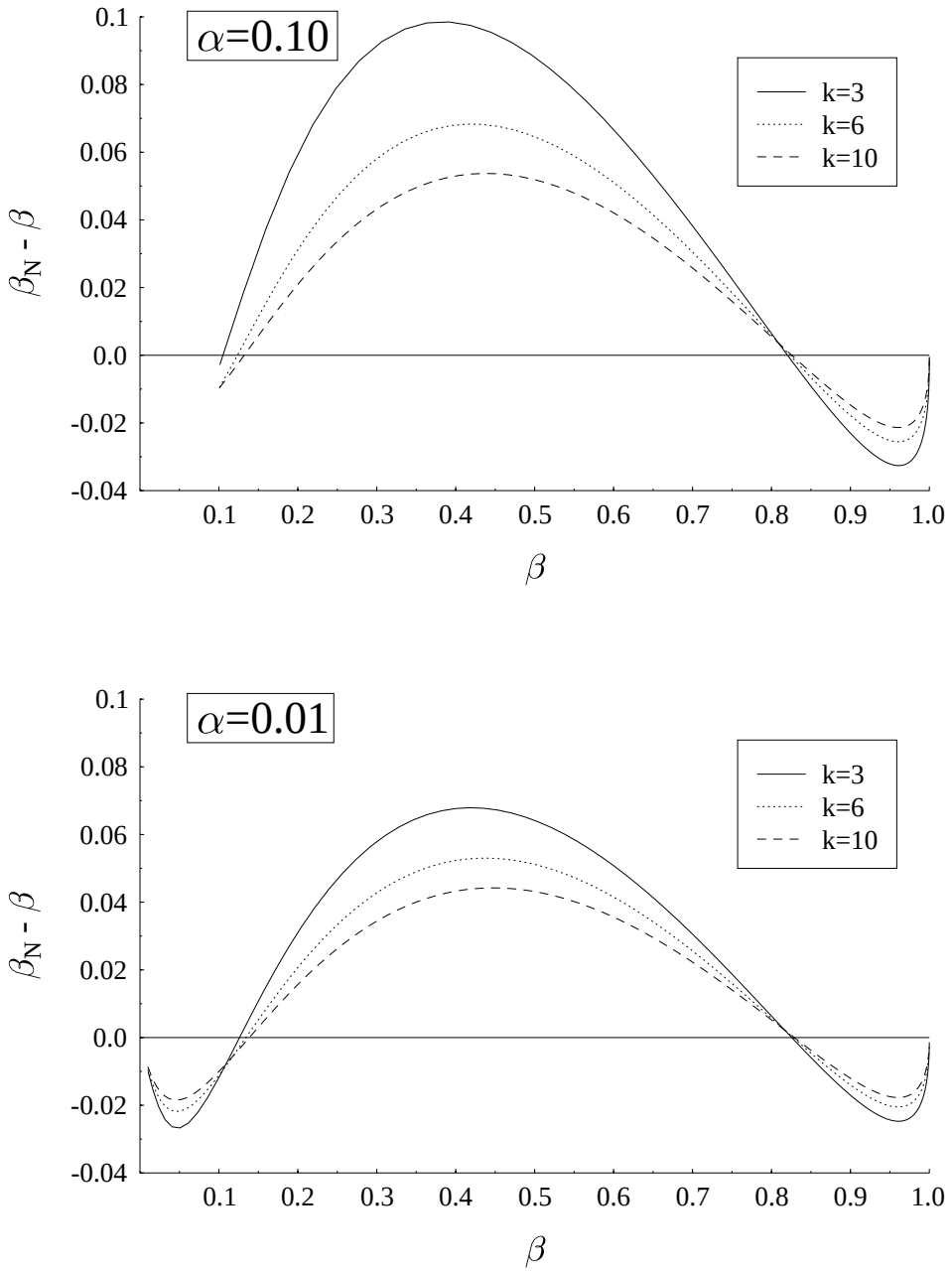


Abb. 2.2: Plot des Fehlers $\beta_N - \beta$ in Abhängigkeit von der asymptotischen Macht β.

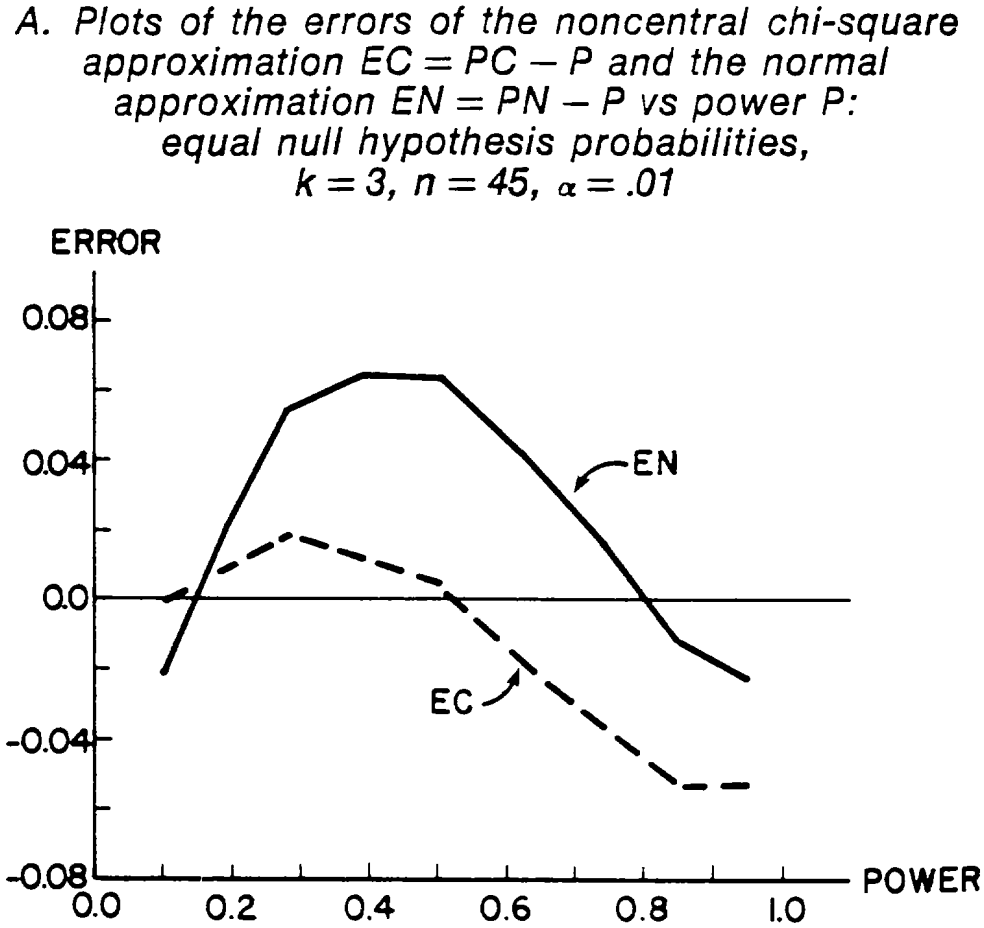
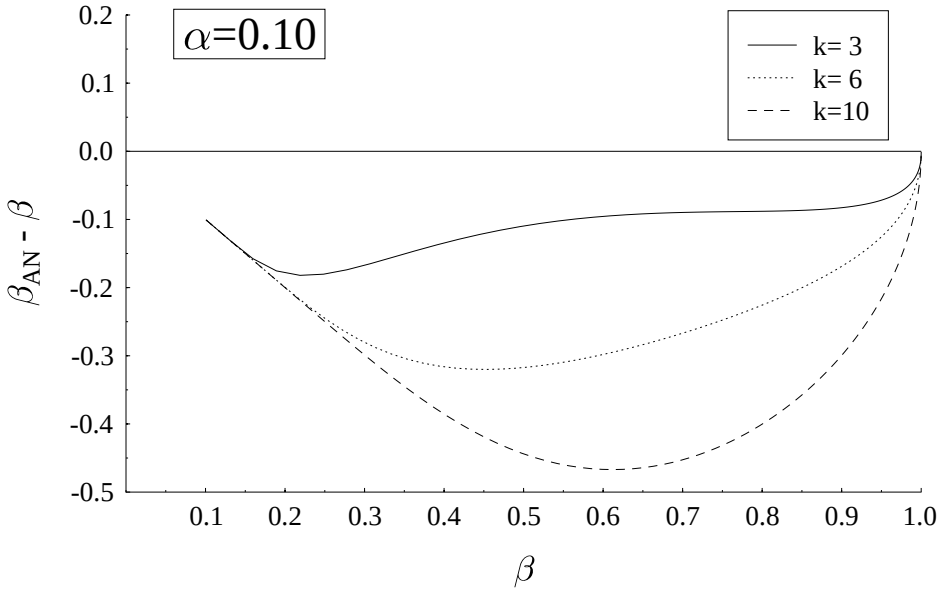


Abb. 2.3: Abbildung A aus Broffitt & Randles (1977)

Sind die Verteilungen von X_{GAM}, X_{NOR} und X_{GLK} prinzipiell noch für eine breitere Palette von Situationen anwendbar, so kann dies von der Verteilung von X_{AN} nicht behauptet werden. Allein die Tatsache, daß die asymptotische Verteilung unter „local alternatives“ noch nicht einmal von der Zellenzahl k abhängt, lassen Zweifel an der Verwendbarkeit aufkommen. Mit steigendem k werden die Unterschiede der $N(\delta, 4\delta)$-Verteilung aus (2.32) zur $\chi^2_{k-1}(\delta)$-Verteilung immer dramatischer. Nur für sehr großes δ ähneln sich die beiden Verteilungsfunktionen, hiermit ist aber ein dicht bei 1 liegender Wert der Macht verbunden.

Die Abbildung 2.4 zeigt deutlich, daß selbst für $k = 3$ die Verteilung von X_{AN} für die Schätzung der exakten Macht unbrauchbar ist, zumindest falls der Anspruch erhoben wird, daß für eine sehr große Stichprobe der Wert der exakten Macht, auch wenn er kleiner als 1 ist, mit kleinem Fehler geschätzt wird. Über das Verhalten bei kleinen Stichproben kann keine sichere Aussage getroffen werden.



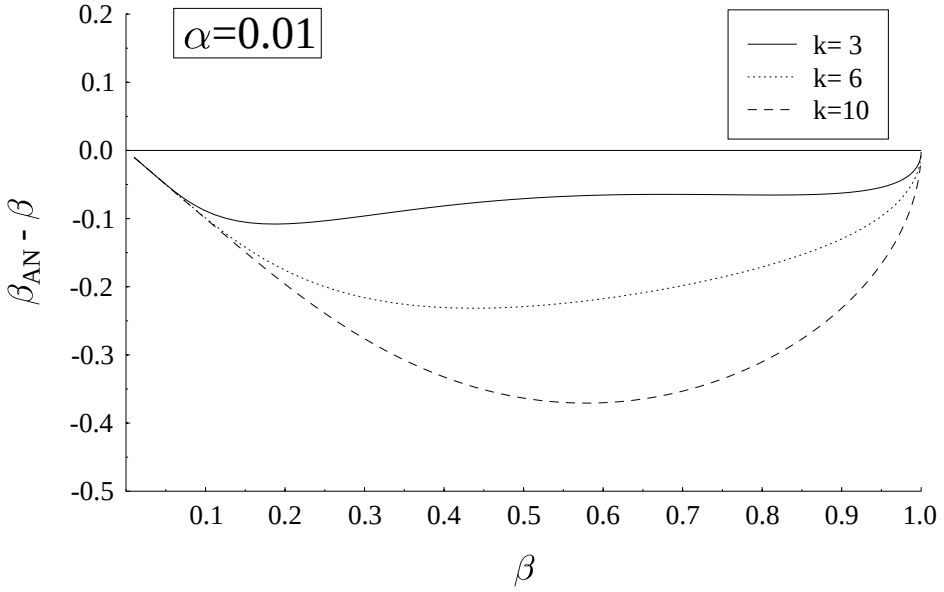


Abb. 2.4: Plot des Fehlers $\beta_{AN} - \beta$ in Abhängigkeit von der asymptotischen Macht β.

2.5 Simulationen zum Vergleich der Approximationen der exakten Macht, einfache Hypothese

2.5.1 Durchführung der Simulationen

2.5.1.1 Probleme bei der Auswertung der Ergebnisse

Zunächst sollen einige zusätzliche Probleme genannt werden, die dadurch entstehen, daß die Verteilung von χ^2 diskret ist. Da die Powerbetrachtungen hauptsächlich für kleine Stichproben durchgeführt werden sollen, ist hier eine kritische Beurteilung besonders wichtig. Üblicherweise wird zu gegebenem Niveau α mit $0 < \alpha < 1$ der kritische Wert aus der asymptotischen Verteilung der χ^2-Prüfgröße abgelesen. Man nimmt dann in Kauf, daß das tatsächliche Niveau nur angenähert α entspricht. Störend wirkt sich jedoch aus, daß die exakte Macht keine stetige Funktion des Niveaus α ist. Die Konsequenzen werden nun anhand eines Beispiels erläutert.

Sei dazu folgende Situation gegeben:

Beispiel 1:

$$\begin{aligned} \pi &= (0.24, 0.22, 0.20, 0.18, 0.16)^T \\ h(\theta_0) &= (0.2, 0.2, 0.2, 0.2, 0.2)^T \\ n &= 10 \\ \Delta(\pi, h(\theta_0)) &= 0.02 \end{aligned}$$

Sei weiterhin $c(\alpha)$ der kritische Wert zu gegebenem Niveau α:

$$P(\chi^2_{k-1} > c(\alpha)) = \alpha$$

Die exakte Macht ist dann durch

$$Pow_{\chi^2}(\alpha) = P(\chi^2 > c(\alpha))$$

und eine Approximation dieses Wertes z. B. durch

$$Pow_{LKC}(\alpha) = P(X_{LKC} > c(\alpha))$$

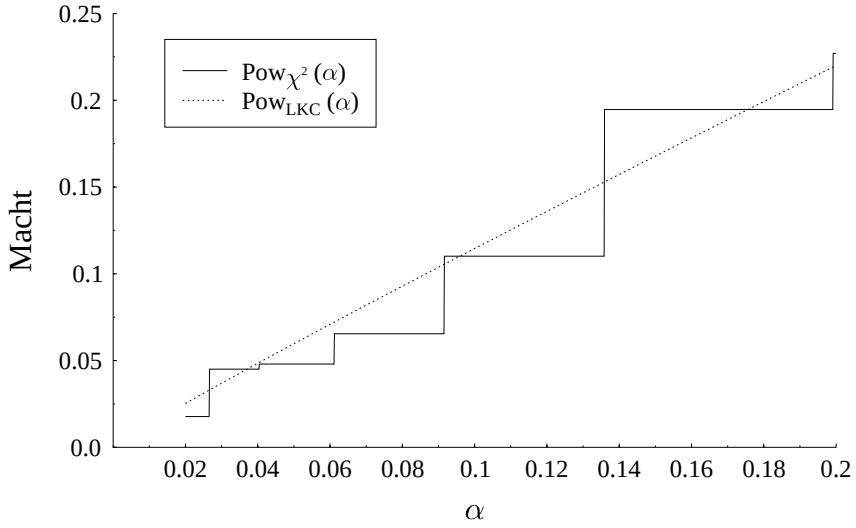


Abb. 2.5: Exakte und angenäherte Macht für Beispiel 1

gegeben. Abbildung 2.5 zeigt nun die Verläufe dieser Funktionen für $\alpha \in [0.02, 0.2]$.

Wählt man nun z. B. $\alpha = 0.10$, so ergibt sich:

$$\begin{aligned} Pow_{\chi^2}(0.10) &= 0.1102 \\ Pow_{LKC}(0.10) &= 0.1146 \end{aligned}$$

In dieser Situation würde man eine sehr gute Approximation der Macht festhalten. Wird jedoch das Signifikanzniveau $\alpha = 0.09$ verwendet, so ändert sich das Bild schlagartig:

$$\begin{aligned} Pow_{\chi^2}(0.09) &= 0.0655 \\ Pow_{LKC}(0.09) &= 0.1038 \end{aligned}$$

Die eben betrachtete Nullhypothese gleicher Wahrscheinlichkeiten der Zellen,

$$H_0: \quad h_i(\theta_0) = \frac{1}{k}, \quad i = 1, \ldots, k,$$

wird in der Literatur (z. B. Frosini, 1976) häufig betrachtet. In diesem Fall wird jedoch der diskrete Charakter der χ^2-Prüfgröße besonders deutlich. Eine nur scheinbar ähnliche Situation zu Beispiel 1 liegt vor, wenn die Rollen von π und $h(\theta_0)$ vertauscht werden:

Beispiel 2:

$$\begin{aligned} \pi &= (0.2, 0.2, 0.2, 0.2, 0.2)^T \\ h(\theta_0) &= (0.24, 0.22, 0.20, 0.18, 0.16)^T \\ n &= 10 \\ \Delta(\pi, h(\theta_0)) &= 0.0207 \end{aligned}$$

Die Diskrepanz unterscheidet sich nur wenig von dem Wert des ersten Beispiels, und es wäre plausibel, einen ähnlichen Verlauf der Funktionen wie in Abbildung 2.5 zu erwarten.

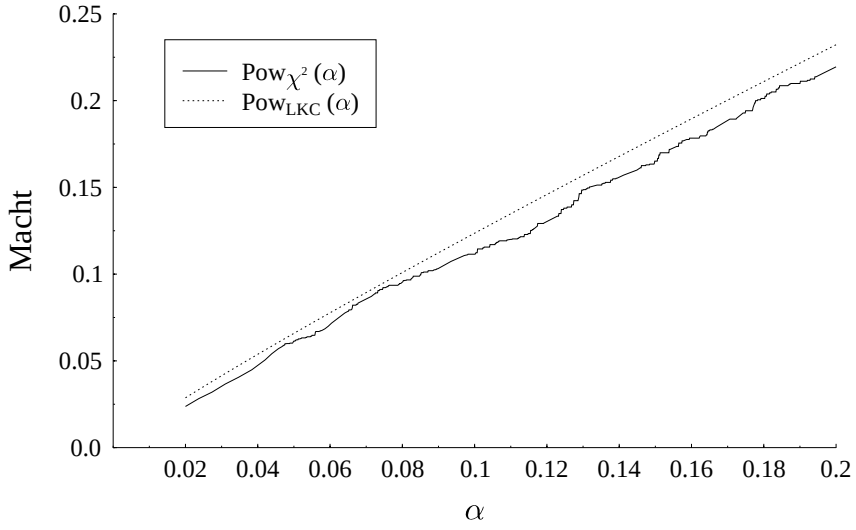


Abb. 2.6: Exakte und angenäherte Macht für Beispiel 2

Die Abbildung 2.6 zeigt jedoch einen auf den ersten Blick überraschend

glatten Verlauf der Funktion der exakten Macht und eine relativ gute Anpassung durch die LKC-Approximation. Die Erklärung ist im folgenden Sachverhalt zu finden:

Für $h_i(\theta_0) = 1/k$, $i = 1, \ldots, k$, führt jede Permutation der relativen Häufigkeiten zum gleichen Wert der Prüfgröße. Im Beispiel 1 werden nur 22, im zweiten Beispiel hingegen 925 verschiedene Werte angenommen. Es gibt für gegebenes n und k

$$\binom{n+k-1}{k-1} \tag{2.36}$$

Möglichkeiten (1001 für n=10, k=5) für den Vektor der relativen Häufigkeiten (siehe z.B. Krengel [1988, S. 9]), und ebenso viele verschiedene Werte der Prüfgröße können prinzipiell auftreten. Bei geeigneter Wahl von π und $h(\theta_0)$ kann dann wie in Beispiel 2 ein selbst für kleines n relativ glatter Verlauf der Funktion eintreten. Für großes n wird natürlich der Verlauf der Funktion auch unter $h_i(\theta_0) = 1/k$, $i = 1, \ldots, k$, glatter.

Für die Güte der Approximationen ist also nicht nur die ungefähre Größenordnung der Werte $h_i(\theta_0)$ entscheidend. Relativ kleine Änderungen der Hypothese können durchaus einen großen Einfluß auf die gemessene Abweichung ausüben.

Eine Möglichkeit, eine stetige Funktion der Macht in Abhängigkeit des Signifikanzniveaus zu erhalten, wäre durch Randomisierung gegeben. In der Praxis würde jedoch kaum jemand bei Anwendung des χ^2-Tests zu diesem Verfahren greifen. Deshalb werden die Simulationen in der Form durchgeführt, daß die kritischen Werte weiterhin aus der asymptotischen Verteilung der χ^2-Prüfgröße abgelesen werden.

2.5.1.2 Konstruktion der Hypothesen und Alternativen

Um möglichst viele Konstellationen zu erfassen, werden sowohl Hypothesen als auch Alternativen zufällig erzeugt. Die Unterstellung einer Verteilung ist zwar problematisch, aber es soll von dem üblichen Weg abgewichen werden, besonders „glatte“, willkürlich gewählte Werte wie z.B. $h(\theta_0) = (0.2, 0.3, 0.5)^T$ oder $h(\theta_0) = (1/4, 1/4, 1/4, 1/4)^T$ zu verwenden, die, wie im letzten Abschnitt gesehen, zu deutlichen Sprüngen in der Funktion der exakten Macht führen können. Eventuell gemessene Abweichungen sind dann eher mit dem Aussehen der exakten Funktion der Macht als mit der fehlenden Güte der Approximationen zu begründen. Natürlich können bei

zufälliger Erzeugung auch solche Fälle entstehen, sie werden aber nicht überbewertet. Außerdem werden die vorläufig erzeugten Alternativen „korrigiert“, damit die auftretenden Werte der exakten Macht möglichst gleichmäßig verteilt sind.

Die Simulation ist in Anlehnung an Slakter (1968) und Broffitt & Randles (1977) so gestaltet, daß formal davon ausgegangen wird, daß die nichtzentrale χ^2-Verteilung eine gute Approximation liefert.

Es wird die Tatsache ausgenutzt, daß die Erzeugung von Zufallszahlen mit einer Rechteckverteilung auf dem Computer besonders einfach ist, da bei den meisten Programmiersprachen diese Funktion implementiert ist. Für gegebene Werte n, k und α sind dann die Schritte zur Ermittlung einer Hypothese und einer Alternative wie folgt gegeben:

Schritt 1: Berechnung des kritischen Wertes $c(\alpha)$ aus der asymptotischen Verteilung der χ^2-Prüfgröße:

$$P(\chi^2_{k-1} > c(\alpha)) = \alpha$$

Schritt 2: Erzeugung einer Zufallszahl w_α im Intervall $]\alpha, 1[$.

Schritt 3: Berechnung des Wertes δ mit

$$P(\chi^2_{k-1}(\delta) > c(\alpha)) = w_\alpha$$

Hier wird angenommen, daß die nichtzentrale χ^2-Verteilung als Approximation geeignet ist, um den möglichen Bereich der Werte der exakten Macht gleichmäßig abzudecken.

Schritt 4: Erzeugung von k Zufallszahlen $u_1, \ldots, u_k$ aus dem Intervall $]0, 1[$. Setze dann

$$h_i(\theta_0) = u_i/(u_1 + \ldots + u_k), \quad i = 1, \ldots, k.$$

Schritt 5: Erzeugung von k Zufallszahlen $v_1, \ldots, v_k$ aus dem Intervall $]0, 1[$. Der vorläufige Vektor π^v ist definiert durch

$$\pi^v_i = v_i/(v_1 + \ldots + v_k), \quad i = 1, \ldots, k.$$

Schritt 6: Berechnung von $n\Delta^v = n\Delta(\pi^v, h(\theta_0))$.

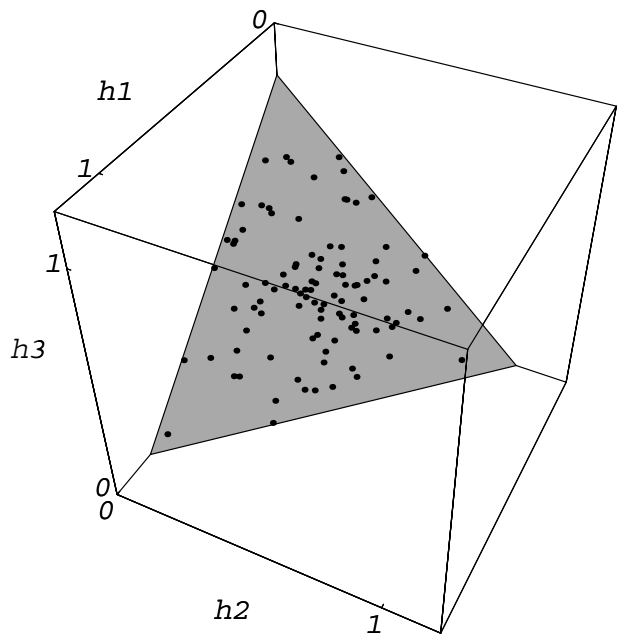


Abb. 2.7: Plot von 100 erzeugten Vektoren $h(\theta_0)$

Schritt 7: Bestimmung eines λ, so daß sich für

$$\pi = h(\theta_0) + \lambda(\pi^v - h(\theta_0))$$

eine Diskrepanz von $n\Delta(\pi, h(\theta_0)) = \delta$ ergibt:

$$\lambda = \sqrt{\frac{\delta}{n\Delta^v}}$$

Schritt 8: Kontrolle, ob alle Komponenten des Vektors π positiv sind. Falls dies nicht der Fall ist, wird bei Schritt 2 neu begonnen.

Abbildung 2.7 zeigt in einem $h_1 - h_2 - h_3$-Diagramm die Lage von 100 nach Schritt 4 erzeugten $h(\theta_0)$ für $k = 3$. Zumindest für $k = 3$ wird mit der Vorschrift aus Schritt 4 der Bereich der möglichen Hypothesen gut abgedeckt. Außerdem ist erkennbar, daß auch Vektoren $h(\theta_0)$ erzeugt werden, die eine oder mehrere sehr kleine Komponenten besitzen.

Die folgenden Kombinationen von n und k wurden in der Untersuchung betrachtet:

$$k = 3 \qquad n = 10, 30, 50, 100^*, 500^* \tag{2.37}$$

$$
\begin{aligned}
k = 4 \qquad & n = 10, 30, 50 \\
k = 6 \qquad & n = 10, 50^*, 100^*, 500^* \\
k = 10 \qquad & n = 100^*
\end{aligned}
$$

Für jedes Paar (n, k) wurden jeweils 100 Hypothesen bzw. Alternativen erzeugt, sowohl für $\alpha = 0.10$ als auch für $\alpha = 0.01$. Insgesamt ergeben sich somit 2600 zu untersuchende Fälle.

2.5.1.3 Beurteilung der Güte der Approximation

Für jedes der 2600 Tupel $(n, k, \alpha, \pi, h(\theta_0))$ werden die folgenden Berechnungen durchgeführt:

Schritt 1: Berechnung des kritischen Wertes $c(\alpha)$ aus der asymptotischen Verteilung der χ^2-Prüfgröße:

$$P(\chi^2_{k-1} > c(\alpha)) = \alpha$$

Schritt 2: Berechnung der exakten Macht und ihrer Approximationen:

$$
\begin{aligned}
Pow_{\chi^2}(\alpha) &= P(\chi^2 > c(\alpha)) \\
Pow_{App}(\alpha) &= P(X_{App} > c(\alpha)), \\
& \quad App \in \{AN, NOR, NC, GAM, LKC, GLK\}
\end{aligned}
$$

Bemerkung: Für die Kombinationen, bei denen die Stichprobengröße in (2.37) mit einem Sternchen gekennzeichnet ist, wurde die exakte Macht mit Monte-Carlo-Methoden geschätzt. Für $n = 500$ wurden dazu unter der jeweiligen Alternative 20000 Prüfgrößen erzeugt und der relative Anteil der Werte, die $c(\alpha)$ übertrafen, berechnet. Für die übrigen Fälle wurden 10000 Werte erzeugt.

Schritt 3a: Berechnung der Abweichung

$$\mid Pow_{\chi^2}(\alpha) - Pow_{App}(\alpha) \mid$$

für jede Approximation.

Schritt 3b: Berechnung der relativen Abweichung

$$\frac{\mid Pow_{\chi^2}(\alpha) - Pow_{App}(\alpha) \mid}{(Pow_{\chi^2}(\alpha)(1 - Pow_{\chi^2}(\alpha)))^{\frac{1}{2}}}$$

für jede Approximation.

In der Literatur wird sowohl mit dem Verfahren aus Schritt 3a (z. B. Drost et al., 1989) als auch mit dem Verfahren aus Schritt 3b gearbeitet (z. B. Broffitt & Randles, 1977). Die Messung der absoluten Abweichung besitzt den Vorteil der leichteren Interpretierbarkeit, die Beurteilung der Güte der Approximation auf der Grundlage der speziellen relativen Abweichung aus Schritt 3b entspringt der Überlegung, daß absolute Abweichungen stärker berücksichtigt werden sollten, falls die wahre Macht entweder sehr klein oder aber sehr groß ist. Abweichungen, die sich zu einer exakten Macht von ungefähr 0.50 ergeben, sollen weniger stark ins Gewicht fallen als gleichgroße Differenzen zu einer wahren Macht von beispielsweise 0.10 oder 0.90. Daß auch kleine Abweichungen bei sehr großer Macht höher gewichtet werden sollen, folgt aus der Überlegung, daß als duales Problem auch der Fehler 2. Art, der zu $1 - Pow_{\chi^2}$ berechnet wird, hätte bestimmt werden können. Deshalb wird der Nenner symmetrisch zu 0.5 gewählt.

2.5.2 Zusammenfassung der Ergebnisse

Es ist natürlich an dieser Stelle unmöglich, für jeden der 2600 Fälle die berechnete Macht und die Abweichungen der Approximationen explizit anzugeben.

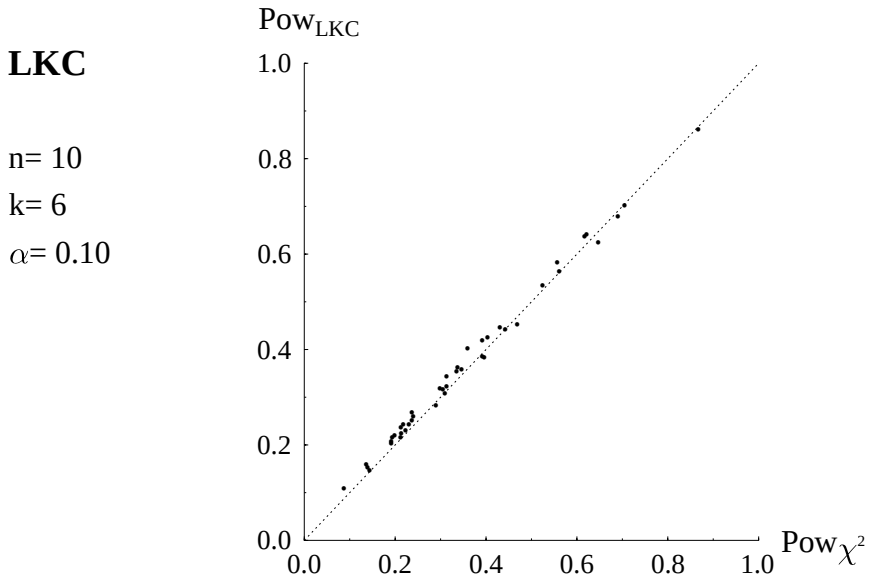
Wie in Abbildung 2.7 ersichtlich war, werden auch Hypothesen mit extrem kleinen Werten $h_i(\theta_0)$ erzeugt. Nun wird man eine Approximation nur dann verwenden, wenn die Rahmenbedingungen für die Anwendung des χ^2-Tests gegeben sind. Es existieren Faustregeln, die, je nach Autor, ein Minimum der Erwartungen $nh_i(\theta)$ von mindestens 1, 2 oder sogar 5 verlangen (Moore, 1986, gibt dazu einen kurzen Überblick). Wie einige Vorstudien ergaben, sind diese Bedingungen in vielen Situationen zu restriktiv, und es wurden deshalb sämtliche Konstellationen mit

$$nh_i(\theta_0) \geq 0.5, \quad i = 1, \ldots, k,$$

ausgewertet. Es werden nun für jede Approximation die Simulationsergebnisse zu den Kombinationen

$$(n, k, \alpha) \in \{(10, 6, 0.10), (50, 4, 0.10), (100, 10, 0.10), (500, 6, 0.01)\}$$

graphisch veranschaulicht.



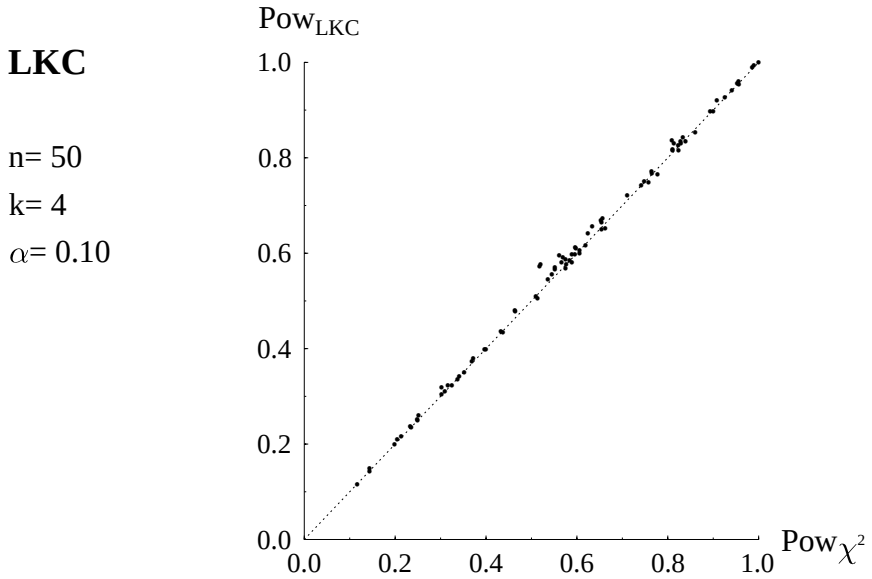


Abb. 2.8: Exakte Macht ↔ App. Macht, LKC-Approximation, (I)

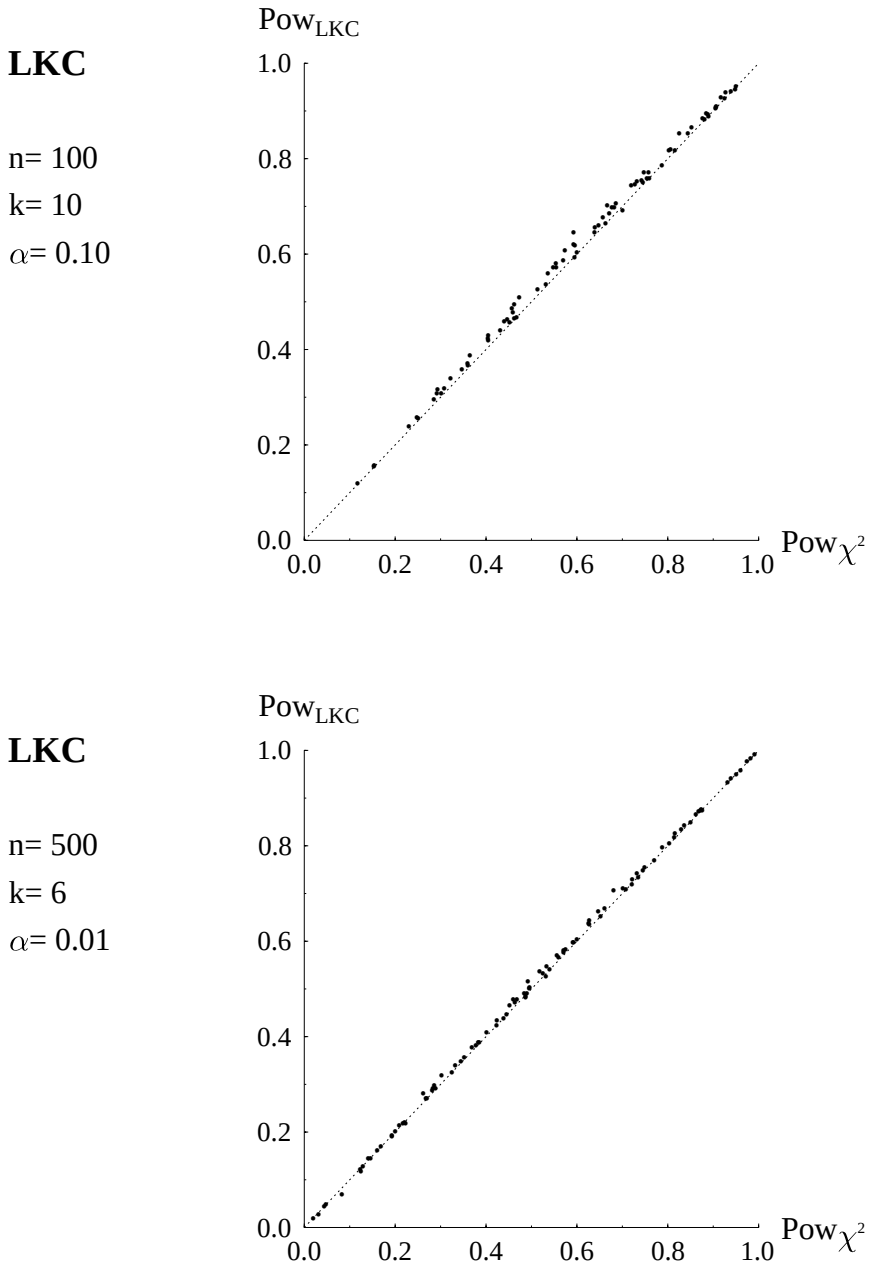
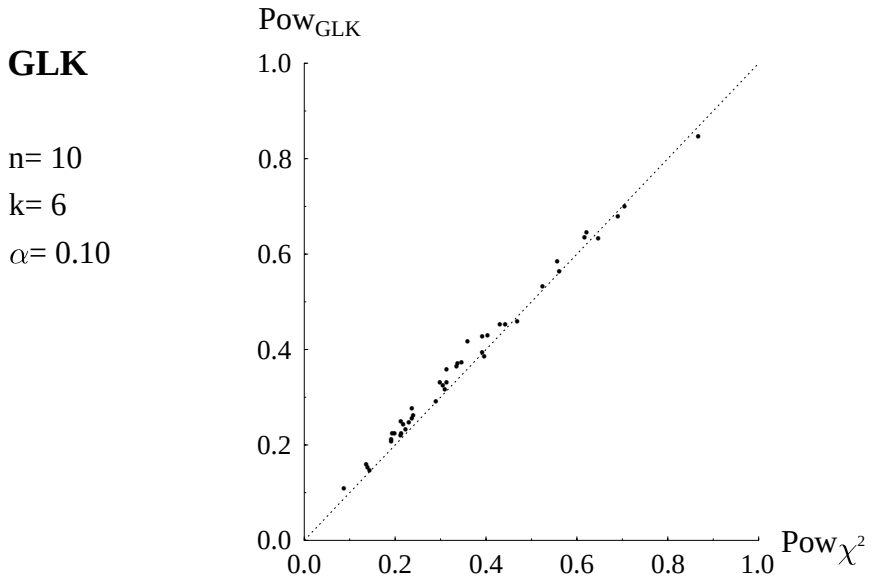


Abb. 2.9: Exakte Macht ↔ App. Macht, LKC-Approximation, (II)



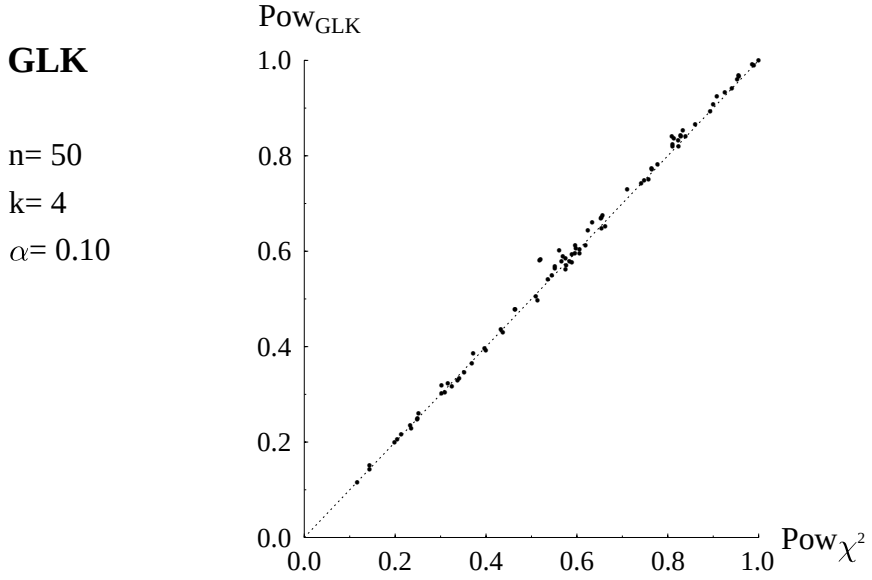


Abb. 2.10: Exakte Macht ↔ App. Macht, GLK-Approximation, (I)

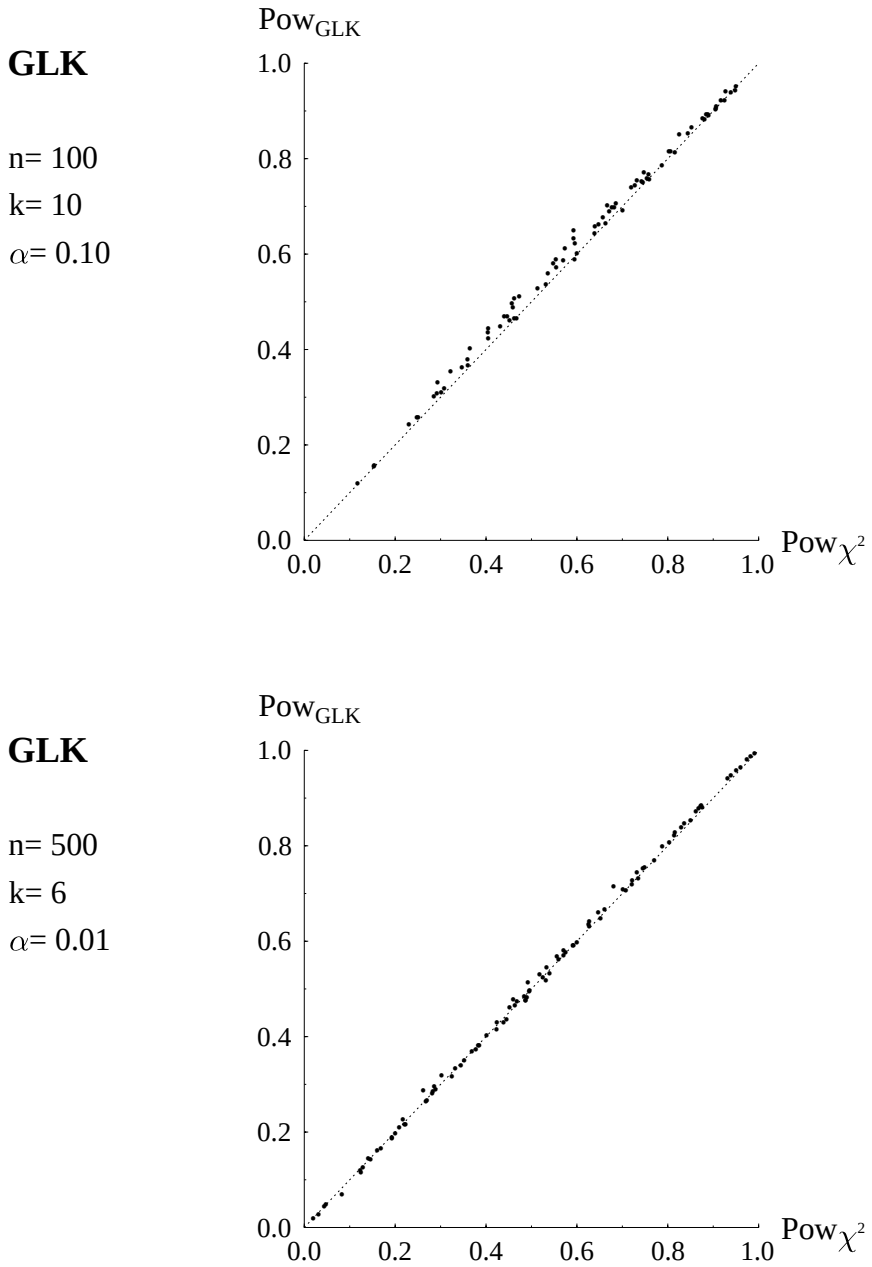
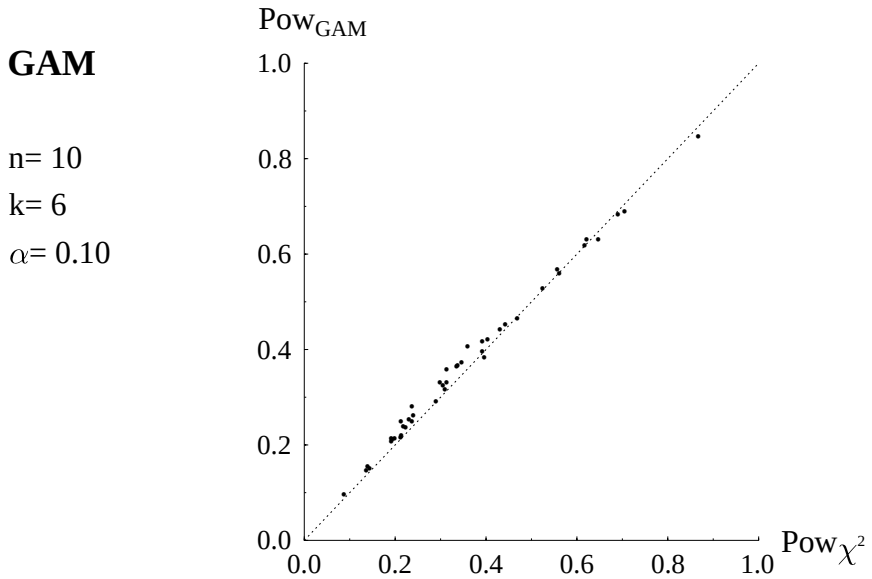


Abb. 2.11: Exakte Macht ↔ App. Macht, GLK-Approximation, (II)



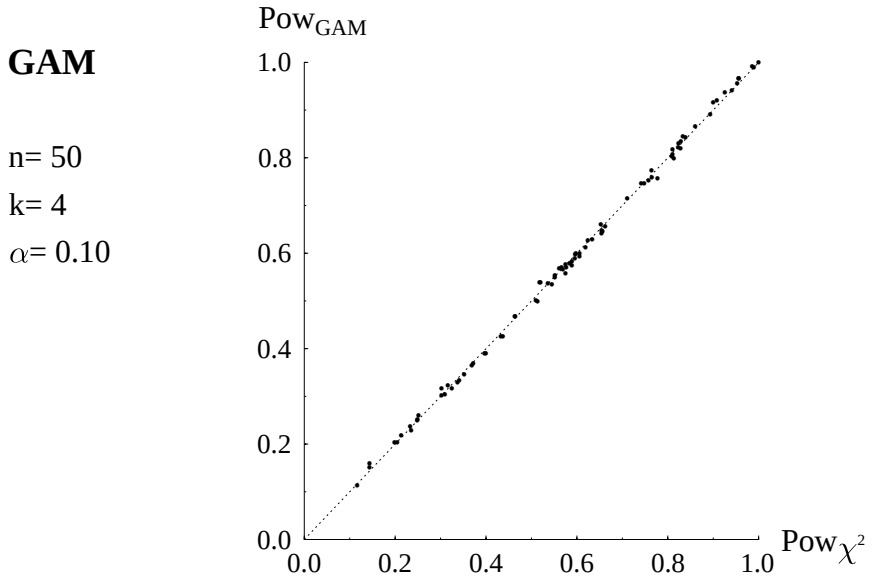
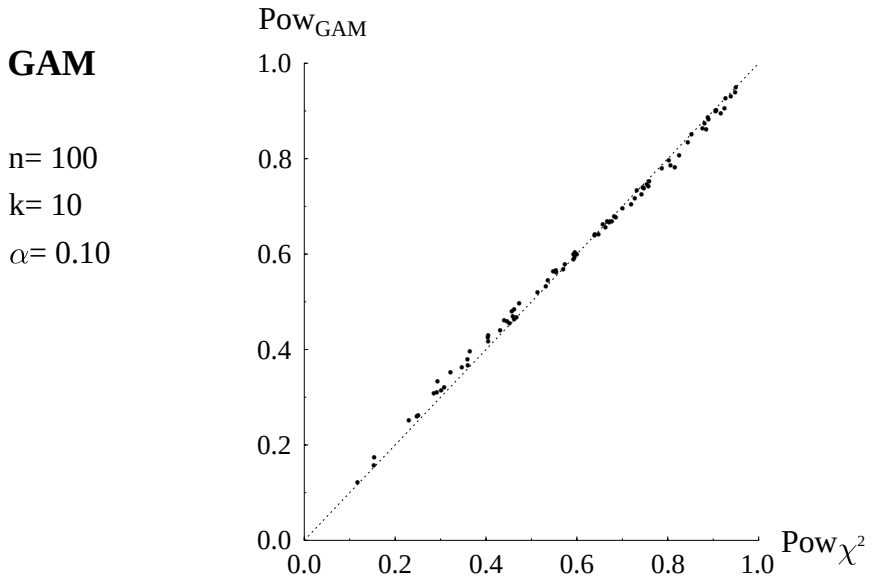


Abb. 2.12: Exakte Macht ↔ App. Macht, GAM-Approximation, (I)



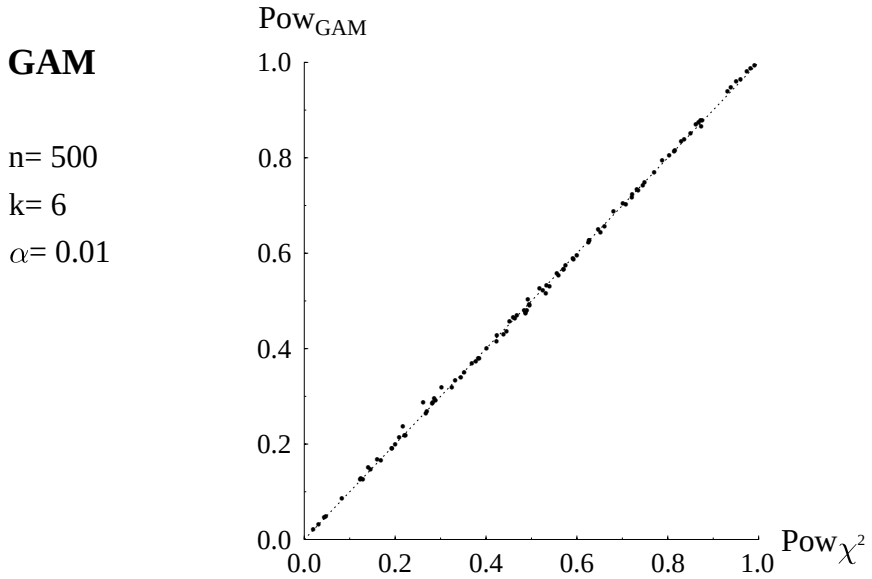
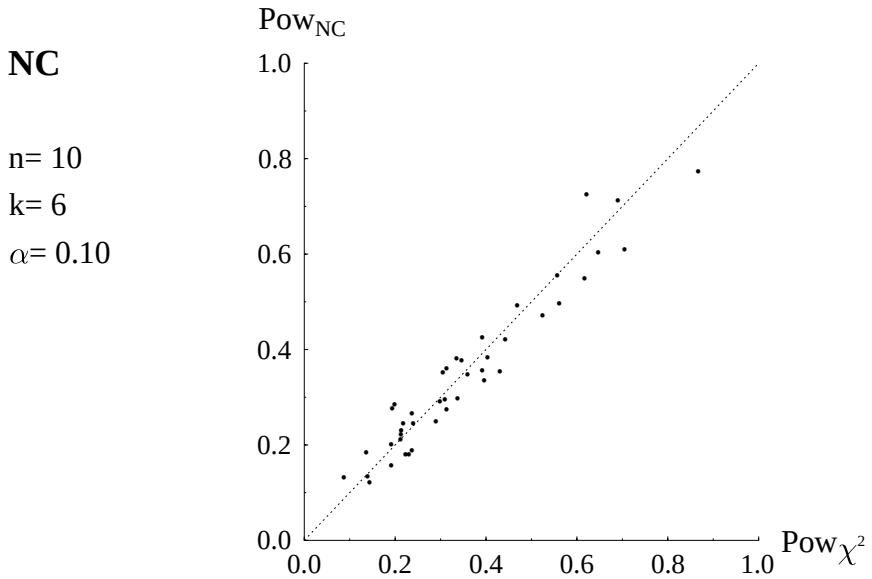


Abb. 2.13: Exakte Macht ↔ App. Macht, GAM-Approximation, (II)



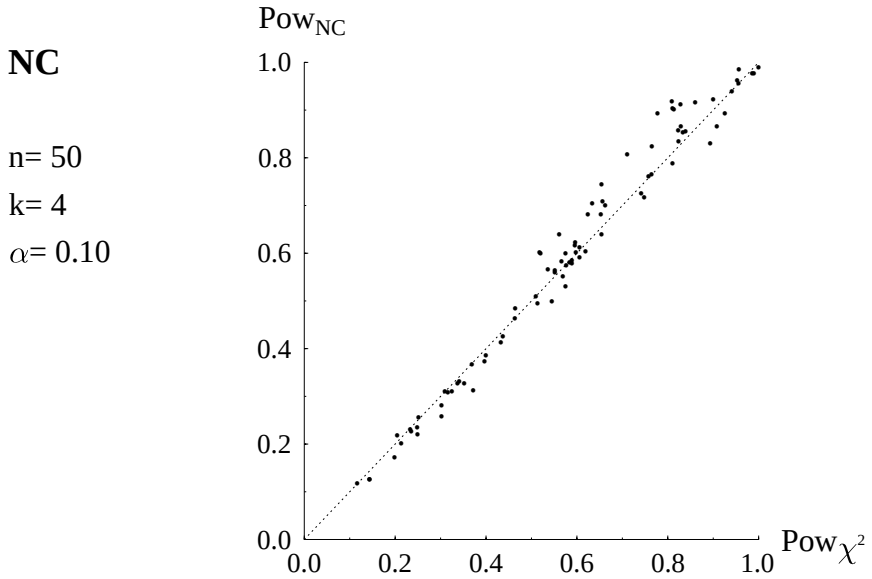


Abb. 2.14: Exakte Macht ↔ App. Macht, NC-Approximation, (I)

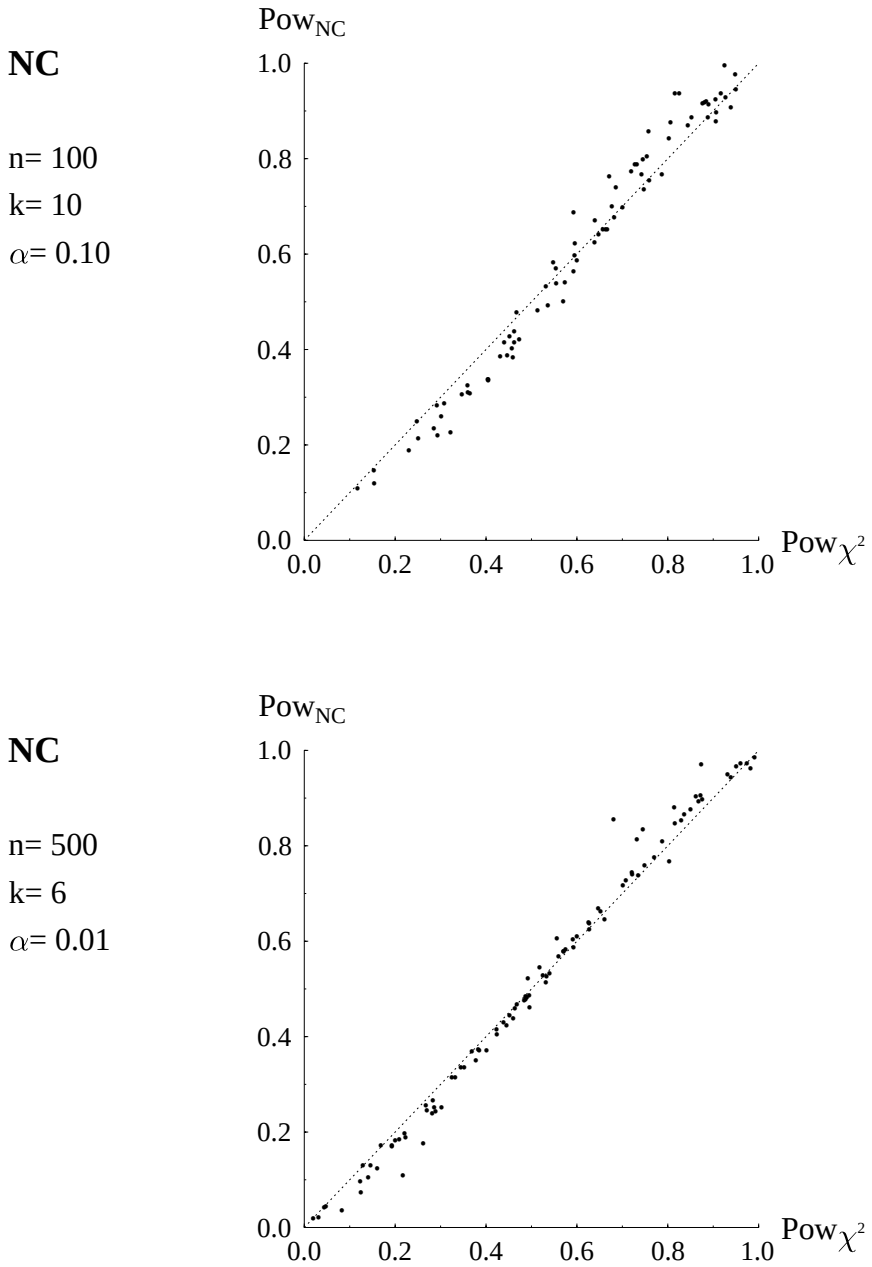
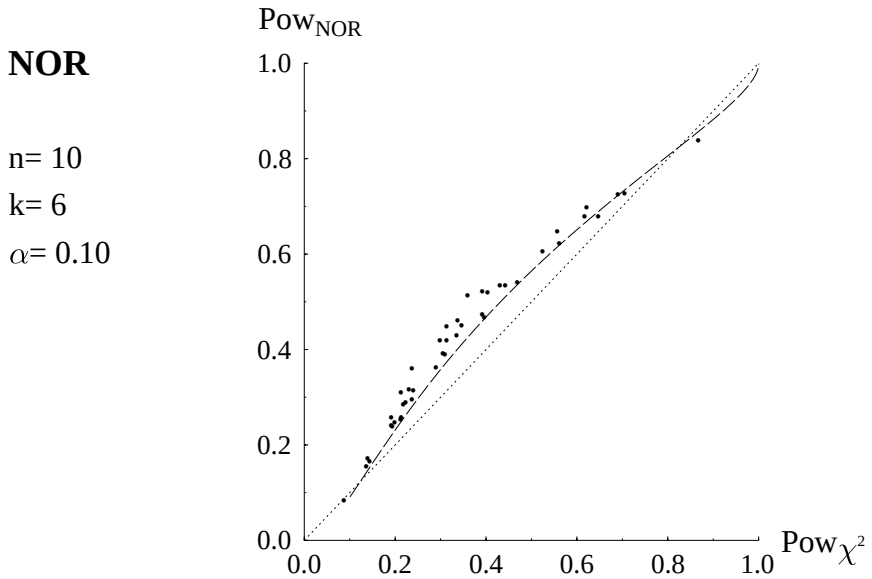


Abb. 2.15: Exakte Macht ↔ App. Macht, NC-Approximation, (II)



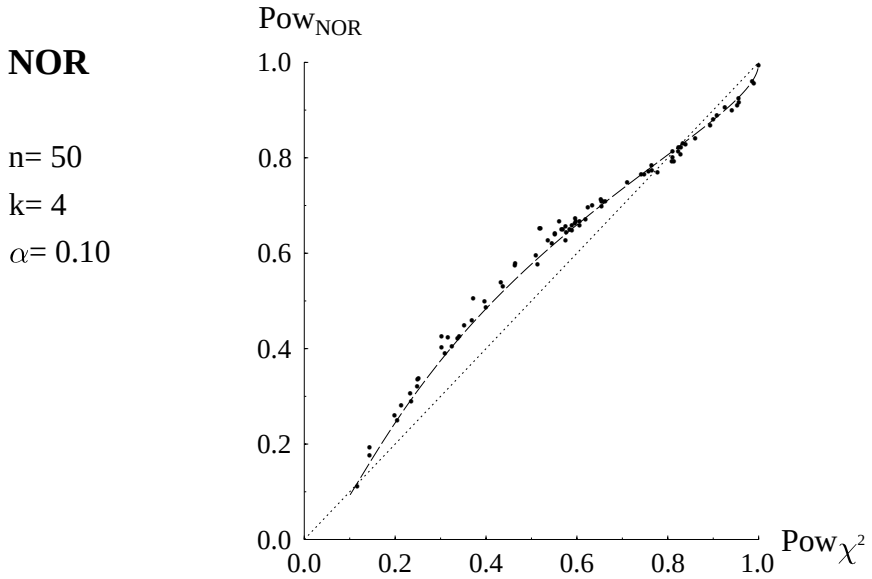
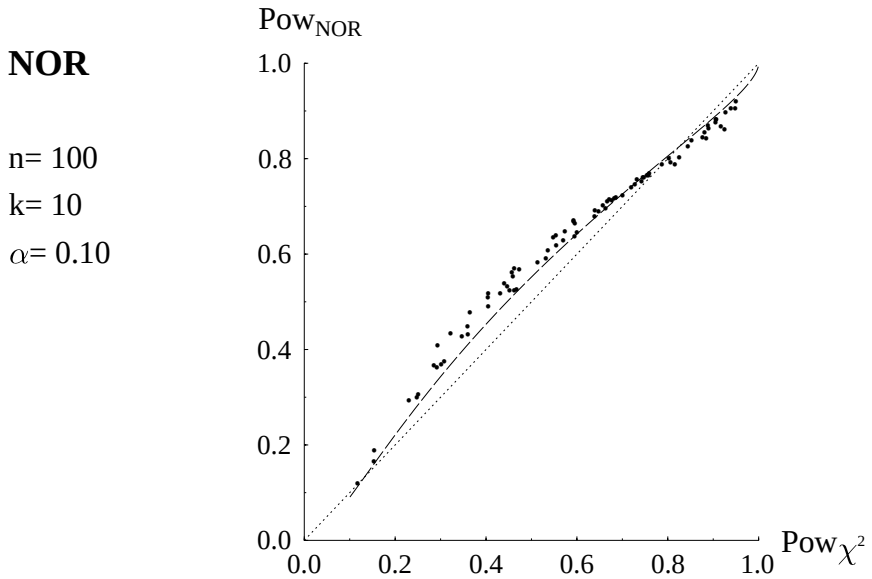


Abb. 2.16: Exakte Macht ↔ App. Macht, NOR-Approximation, (I)



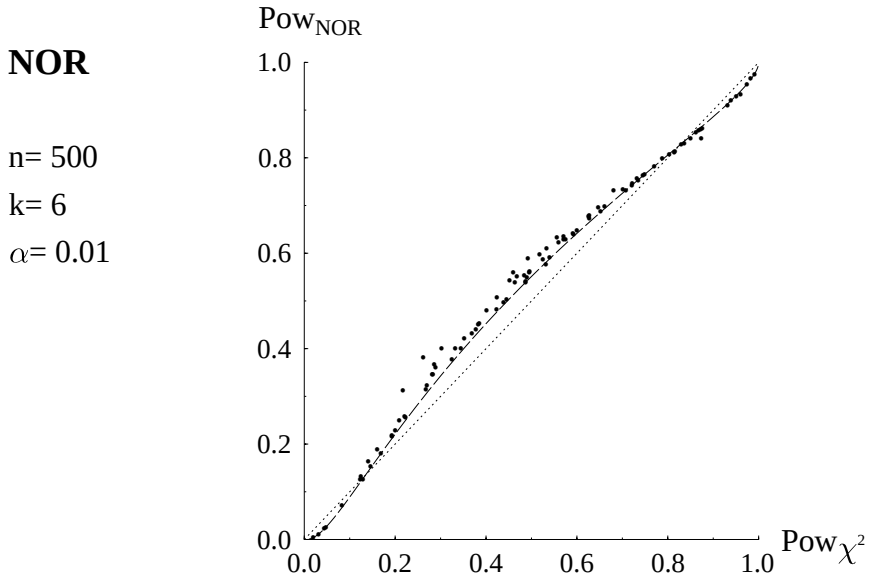


Abb. 2.17: Exakte Macht ↔ App. Macht, NOR-Approximation, (II)

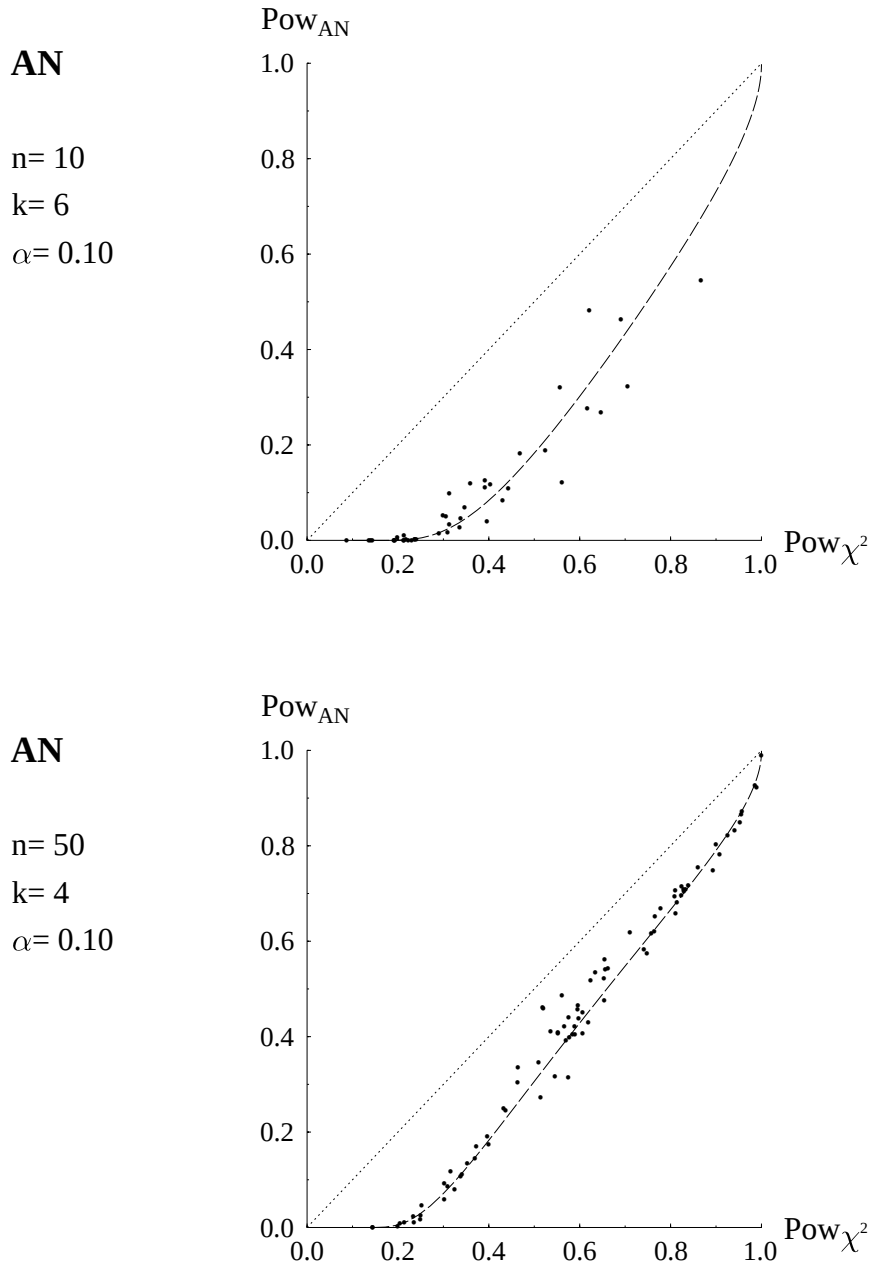


Abb. 2.18: Exakte Macht ↔ App. Macht, AN-Approximation, (I)

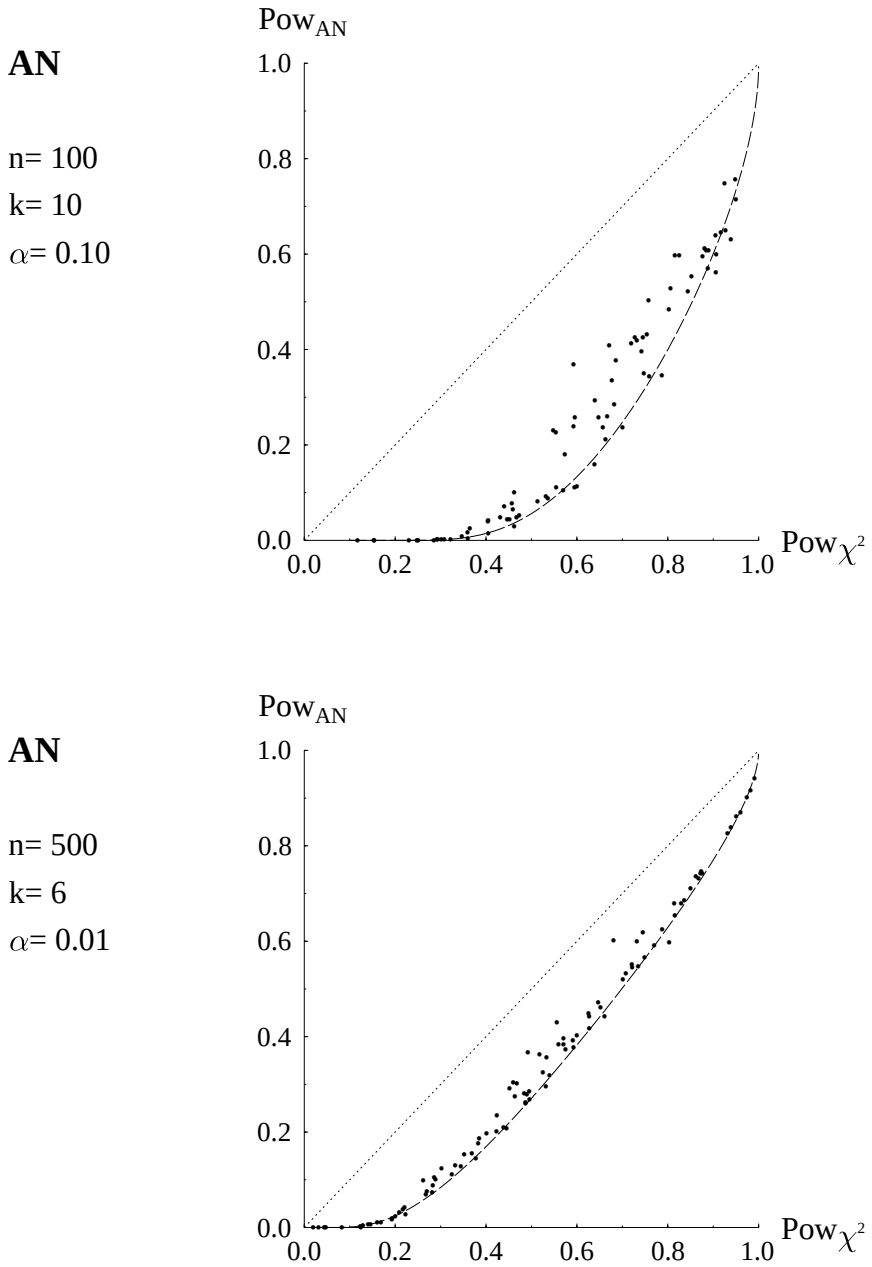


Abb. 2.19: Exakte Macht ↔ App. Macht, AN-Approximation, (II)

In den Abbildungen 2.8 bis 2.19 sind die Paare $(Pow_{\chi^2}(\alpha), Pow_{App}(\alpha))$ zusammen mit einer Hilfslinie eingezeichnet. In den Diagrammen der X_{AN}- und der X_{NOR}-Verteilung ist zusätzlich jeweils die Kurve der Paare (β, β_{AN}) bzw. (β, β_N) eingetragen, die man über die Betrachtung der asymptotischen Macht unter „local alternatives“ erhält (vgl. Abschnitt 2.4.4.8). Für die Gammaverteilungen sind die analogen Kurven nicht angegeben, da sie in der Grafik kaum von der Hilfslinie zu unterscheiden wären.

Es werden nun die Ergebnisse für die vier betrachteten Konstellationen in Tabellenform zusammengefaßt. Im Anhang B werden die Tabellen zu sämtlichen Konstellationen gegeben. Die Plots legen nahe, den Bereich der möglichen Werte der Macht für die Auswertung zu unterteilen:

$$\begin{array}{lcccccc}
\text{niedrige Macht (n. M.)} & : & & & Pow_{\chi^2}(\alpha) & < & 0.30 \\
\text{mittlere Macht (m. M.)} & : & 0.30 & \leq & Pow_{\chi^2}(\alpha) & < & 0.70 \\
\text{hohe Macht (h. M.)} & : & 0.70 & \leq & Pow_{\chi^2}(\alpha) & &
\end{array}$$

Für jede Kombination von n, k und α wird die durchschnittliche absolute ($\emptyset$ a. A.) bzw. relative ($\emptyset$ r. A.) Abweichung nach Schritt 3 sowie die Anzahl der aufgetretenen Fälle für die drei Bereiche der Macht festgehalten. Dabei werden nur die Hypothesen mit

$$\min_{i=1,\ldots,k} nh_i(\theta_0) \geq 0.5$$

betrachtet. Die jeweils niedrigsten Werte sind **fett** gedruckt.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	0.0030	**0.0027**	0.0049	0.0138	0.0574	0.1906
11	$\emptyset$ r. A.	0.0187	**0.0175**	0.0341	0.0880	0.3437	1.1878
m. M.	$\emptyset$ a. A.	0.0108	0.0129	**0.0073**	0.0262	0.0821	0.1697
46	$\emptyset$ r. A.	0.0448	0.0540	**0.0309**	0.1107	0.3471	0.7213
h. M.	$\emptyset$ a. A.	**0.0060**	0.0088	0.0070	0.0385	0.0193	0.1128
29	$\emptyset$ r. A.	0.0974	0.1028	**0.0952**	1.6333	1.0967	2.7572

Tab. 2.1: Auswertung: $\alpha = 0.10$, $n = 50$, $k = 4$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0161**	0.0195	0.0170	0.0322	0.0599	0.2000
19	$\emptyset$ r. A.	0.1065	0.1268	**0.1057**	0.2151	0.3520	1.2376
m. M.	$\emptyset$ a. A.	0.0163	0.0209	**0.0162**	0.0409	0.0905	0.2915
22	$\emptyset$ r. A.	**0.0704**	0.0912	0.0713	0.1760	0.3909	1.2525
h. M.	$\emptyset$ a. A.	**0.0051**	0.0134	0.0175	0.0937	0.0254	0.3519
2	$\emptyset$ r. A.	**0.0364**	0.1051	0.1212	0.6295	0.1788	2.3140

Tab. 2.2: Auswertung: $\alpha = 0.10$, $n = 10$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0085**	0.0117	0.0166	0.0292	0.0537	0.2245
9	$\emptyset$ r. A.	**0.0447**	0.0608	0.0930	0.1646	0.2871	1.2964
m. M.	$\emptyset$ a. A.	0.0171	0.0216	**0.0102**	0.0379	0.0706	0.3804
43	$\emptyset$ r. A.	0.0721	0.0910	**0.0431**	0.1606	0.2967	1.6063
h. M.	$\emptyset$ a. A.	0.0088	**0.0082**	0.0106	0.0383	0.0232	0.3020
31	$\emptyset$ r. A.	0.0646	**0.0616**	0.0925	0.3075	0.2535	2.6281

Tab. 2.3: Auswertung: $\alpha = 0.10$, $n = 100$, $k = 10$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0039**	0.0048	0.0046	0.0291	0.0365	0.1449
26	$\emptyset$ r. A.	**0.0320**	0.0382	0.0324	0.2101	0.2861	1.0525
m. M.	$\emptyset$ a. A.	0.0081	0.0073	**0.0052**	0.0180	0.0632	0.1945
45	$\emptyset$ r. A.	0.0338	0.0306	**0.0218**	0.0772	0.2630	0.8109
h. M.	$\emptyset$ a. A.	**0.0035**	0.0070	0.0041	0.0282	0.0158	0.1375
28	$\emptyset$ r. A.	**0.0318**	0.0864	0.0648	0.2728	0.2608	1.4654

Tab. 2.4: Auswertung: $\alpha = 0.01$, $n = 500$, $k = 6$.

2.5.3 Interpretation der Simulationsergebnisse

Zunächst sollte festgehalten werden, daß die Kommentare und Empfehlungen auf den folgenden Prämissen aufbauen:

- Es wird angenommen, daß in der Anwendung eine Vielzahl von verschiedenen Situationen auftritt. Dies betrifft hauptsächlich die Hypothesen. Berücksichtigt wird diese Annahme durch die Art und Weise der Konstruktion der Hypothesen und Alternativen.

- Es gibt keine Vorabinformation über die Größenordnung der exakten Macht.

Gesucht wurde also eine Approximation, die unabhängig vom genauen Aussehen der Hypothese und der Alternative ohne größere Einschränkungen eine brauchbare Schätzung der exakten Macht ermöglicht. Es kann aber andererseits nicht ausgeschlossen werden, daß Approximationen, die global nicht geeignet sind, für Spezialfälle durchaus anwendbar sind.

Betrachtet man die Plots und die Tabellen des letzten Abschnitts (siehe auch im Anhang), so wird mehr als deutlich, daß die Verteilung von X_{NC} für Powerberechnungen gemäß der genannten Prämissen *nicht* verwendet werden sollte. Während die Verteilungen von X_{GAM}, X_{LKC} und X_{GLK} schon für kleine Stichprobenumfänge hervorragend geeignet sind, um die exakte Macht zu approximieren, ist der bei Verwendung der nichtzentralen χ^2-Verteilung entstehende Fehler besonders für hohe Werte der Macht nicht vernachlässigbar.

Aber auch für niedrige Werte der wahren Macht ist der durchschnittliche absolute Fehler bei Benutzung von X_{NC} im Vergleich zu den besten Approximationen deutlich größer, selbst oder gerade für großes n. Eine geringe Diskrepanz zwischen π und $h(\theta_0)$ bedeutet noch nicht eine gute Approximation, genausowenig kann allerdings aus einer großen Diskrepanz auf eine schlechte Approximation geschlossen werden. Hier wäre eine genauere Analyse notwendig, um festzustellen, welche Faktoren die Güte der Approximation beeinflussen. Der in Gleichung (2.29) gegebene Quotient kann als weiteres Hilfsmittel dienen zu entscheiden, ob die nichtzentrale χ^2-Verteilung verwendet werden kann. In einigen hier nicht erwähnten Simulationen hatte ein bei 1 liegender Quotient in der überwiegenden Anzahl der Fälle eine gute Approximation zur Folge.

Mit Abstand am schlechtesten schneidet die auf der Verteilung von X_{AN} beruhende Approximation ab. Nahezu für alle Kombinationen von n, k und α

erhält man miserable Werte. Nur in dem Fall, daß die exakte Macht sehr dicht bei 1 liegt, ist die Approximation zu gebrauchen. Schon für kleine Stichprobengrößen verläuft die über die Grenzverteilungen unter „local alternatives“ gewonnene Kurve (siehe Abbildungen 2.18 und 2.19) durch die Punktwolke. Die Zuhilfenahme der Grenzverteilungen unter “local alternatives“ ist deshalb ein wichtiges Hilfsmittel zur Überprüfung der Eignung einer Verteilung zur Powerapproximation. Wie in Abschnitt 2.4.4.8 vorausgesagt, wachsen die Fehler mit steigender Zellenzahl k an. Das kleinere Signifikanzniveau $\alpha = 0.01$ führt zu einer geringfügigen Verbesserung. Nur bei extrem kleinen Signifikanzniveau mag die X_{AN}-Approximation zu gebrauchen sein.

Die Powerapproximation auf der Basis der Verteilung von X_{NOR} ist nur bedingt verwendbar. Die Abbildungen 2.16 und 2.17 zeigen, daß auch in diesem Fall die Punktwolke entlang der gestrichelten Kurve verläuft. Da sich diese Kurve mit steigender Zellenzahl nur langsam der Winkelhalbierenden nähert, folgt als Konsequenz nur der Verzicht auf diese Approximation bei den betrachteten Werten von k. Immerhin ist in fast allen Fällen der durchschnittliche absolute bzw. relative Fehler dieser Approximation für den Bereich hoher Macht geringer als der der X_{NC}-Approximation.

Nur die Verteilungen von X_{LKC}, X_{GLK} und X_{GAM} eignen sich dazu, die exakte Macht für kleine bis moderate Stichprobengrößen ohne weitere Vorkenntnisse zu approximieren. Sie liefern für den gesamten Bereich möglicher Werte der Macht verwertbare Ergebnisse. Dabei sind die mit Hilfe der X_{GLK}-Verteilung ermittelten Approximationen meistens geringfügig schlechter. Soll eine Gammaverteilung verwendet werden, so sollte man die X_{GAM}-Verteilung vorziehen. Zwar konvergieren die Schätzer der exakten Macht für die X_{GLK}- und die X_{GAM}-Verteilungen gemäß Abschnitt 2.4.4.8 unter „local alternatives“ mit wachsender Stichprobengröße gegen die falschen Werte, aber die Unterschiede zu den exakten Werten sind so gering, daß sie kaum ins Gewicht fallen. Für sehr großes n sollte man jedoch nur die X_{LKC}-Verteilung verwenden, im anderen Falle spart man unnötig an der Genauigkeit.

Der Unterschied zwischen den beiden besten Approximationen läßt sich so beschreiben: Die mittels der X_{GAM}-Verteilung gewonnenen Schätzer nähern sich den exakten Werten sehr schnell, beinhalten aber einen kleinen systematischen Fehler, der sich erst bei großen Stichproben bemerkbar macht. Die X_{LKC}-Verteilung liefert mit wachsender Stichprobengröße immer genauere Schätzer, nur für kleineres n kann es größere Abweichungen geben, wenn die

Annahme der Normalität der

$$w_i^*(z) = \frac{z_i - h_i(\theta_0)}{\sqrt{h_i(\theta_0)}}, \quad i = 1, \ldots, k,$$

nicht gegeben ist. Dies ist der Fall, wenn mindestens ein π_i sehr klein ist. Bemerkbar macht sich eine Abweichung aber nur dann, wenn die zugehörigen Komponenten von $h(\theta_0)$ ebenfalls klein sind. Testet man z. B. die Hypothese, daß alle Zellen gleichwahrscheinlich sind, so ist in der Nähe der Hypothese die Annahme der Normalität gerechtfertigt. Tritt nun ein kleiner Wert eines π_i auf, ist zwar die Annahme verletzt, aber die Modelle sind dann so unterschiedlich, daß ein Wert der Macht in der Nähe von 1 zu erwarten ist, und die X_{LKC}-Verteilung liefert dann einen Schätzer, der ebenfalls in dieser Region liegt, obwohl sie von der exakten Verteilung der χ^2-Prüfgröße abweicht.

Kapitel 3

Die Verteilung der χ^2-Prüfgröße bei zusammengesetzten Hypothesen

3.1 Einleitung

Im Falle der zusammengesetzten Hypothese existieren nur wenige Arbeiten, die sich mit der Verteilung der χ^2-Prüfgröße bei Fehlspezifikation der Modellfamilie befassen. Üblicherweise wird eine nichtzentrale χ^2-Verteilung, die auf Mitra (1958) zurückgeht, angegeben (z. B. Bishop et al. , 1975). Überträgt man die Ergebnisse des zweiten Kapitels, so ist davon auszugehen, daß die erzielbare Annäherung an die exakte Macht verbesserungswürdig ist. In diesem Kapitel werden daher einige neue Approximationen der Verteilung hergeleitet und die jeweilige Eignung zur näherungsweisen Berechnung der exakten Macht anhand von Simulationen überprüft.

Liegt eine zusammengesetzte Hypothese vor, so hat man sich für Schätzer der Parameter, genauer gesagt für ein Schätzverfahren, zu entscheiden. Die in dieser Arbeit betrachteten Schätzer werden über Minimierung von empirischen Diskrepanzen hergeleitet. Es handelt sich dann um sogenannte Minimum-Diskrepanz-Schätzer. In dieser Klasse liegen z. B. die häufig verwendeten Maximum-Likelihood- und die Minimum-χ^2-Schätzer. Da die theoretische Vorgehensweise für alle betrachteten Schätzer identisch ist, können mit einem einzigen Ansatz Approximationen der Verteilung der χ^2-Teststatistik hergeleitet werden.

3.2 Diskrepanzen zur Herleitung von Minimum-Diskrepanz-Schätzern

3.2.1 Eine allgemeine Klasse von Diskrepanzen

Eine noch offenstehende Aufgabe ist die konkrete Konstruktion von Schätzern $\hat{\theta}(z)$. Dazu wird eine sehr allgemein gehaltene Klasse $\mathcal{D}$ von Diskrepanzen betrachtet, die die χ^2-Diskrepanz als Spezialfall enthält:

Definition 3.1 *Sei $\mathcal{G}$ definiert als die Menge aller Funktionen g auf $\mathbb{D} = (0,1] \times (0,1]$ mit den Eigenschaften*

1. *g ist zweimal stetig differenzierbar auf $\mathbb{D}$*
2. $\dfrac{\partial g}{\partial y}(x,y) = const \quad \forall (x,y) \in \mathbb{D} \cap \{(x,y) : x = y\}$
3. $\dfrac{\partial^2 g}{\partial y^2}(x,y) > 0 \quad \forall (x,y) \in \mathbb{D}$
4. $\dfrac{\partial^2 g}{\partial x \partial y}(x,y) < 0 \quad \forall (x,y) \in \mathbb{D}$

Lemma 3.1 *Sei $p := (p_1, \ldots, p_k)^T$ mit $p_i > 0, i = 1, \ldots, k$, und $\sum_{i=1}^k p_i = 1$. Für $g \in \mathcal{G}$ sei*

$$\Delta_g(\pi, p) := \sum_{i=1}^k g(\pi_i, p_i).$$

Dann ist

$$\Delta_g(\pi, p) \geq \Delta_g(\pi, \pi).$$

Beweis:
Unter den Voraussetzungen für $\mathcal{G}$ existiert für $g \in \mathcal{G}$ die Taylorentwicklung:

$$\begin{aligned} g(\pi_i, p_i) &= g(\pi_i, \pi_i) + \frac{\partial g}{\partial y}(\pi_i, \pi_i)(p_i - \pi_i) + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}(\pi_i, p_i^*)(p_i - \pi_i)^2, \\ &\quad mit \mid p_i^* - \pi_i \mid \leq \mid p_i - \pi_i \mid, \quad i = 1, \ldots, k. \end{aligned}$$

$$\Rightarrow \quad \sum_{i=1}^k g(\pi_i, p_i) = \sum_{i=1}^k g(\pi_i, \pi_i) \quad + \quad \underbrace{\sum_{i=1}^k \frac{\partial g}{\partial y}(\pi_i, \pi_i)(p_i - \pi_i)}_{=0}$$

$$+ \quad \frac{1}{2} \underbrace{\sum_{i=1}^{k} \frac{\partial^2 g}{\partial y^2}(\pi_i, p_i^*)(p_i - \pi_i)^2}_{\geq 0}$$

$$\Rightarrow \quad \sum_{i=1}^{k} g(\pi_i, p_i) \geq \sum_{i=1}^{k} g(\pi_i, \pi_i),$$

wobei Gleichheit genau dann eintritt, wenn $\pi = p$ ist. □

Definition 3.2 *Sei Δ_g wie in Lemma 3.1 definiert. Die Menge $\mathcal{D}$ der Diskrepanzen sei nun definiert als*

$$\mathcal{D} := \{\Delta_g : g \in \mathcal{G}\}.$$

Jede Diskrepanz $\Delta_g \in \mathcal{D}$ kann somit als ein Maß für die Unterschiedlichkeit der Vektoren π und p interpretiert werden. Eine besondere Eigenschaft der Funktionen $g \in \mathcal{G}$ sei an dieser Stelle festgehalten:

Lemma 3.2 *Für $g(x, y)$ wie in Definition 3.1 gilt*

$$\frac{\partial^2 g}{\partial x \partial y}(x_0, x_0) \equiv -\frac{\partial^2 g}{\partial y^2}(x_0, x_0) \quad \forall x_0 \in (0, 1]$$

Beweis:
Nach Def. 3.1 ist für beliebiges $x_0 \in (0, 1)$

$$\frac{\partial g}{\partial y}(x_0, x_0) = const$$

$\Rightarrow$ Die Richtungsableitung in (x_0, x_0) in Richtung $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ ist null.

$$\Rightarrow \left(\frac{\partial^2 g}{\partial x \partial y}(x_0, x_0), \frac{\partial^2 g}{\partial y^2}(x_0, x_0) \right) \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)^T = 0$$

Es folgt die Behauptung. □

3.2.2 Spezielle Klassen von Diskrepanzen

Wird die Bedingung 4. der Definition 3.1 zu

$$
\begin{aligned}
4.^*a) \qquad & \frac{\partial^2 g}{\partial x \partial y}(x,y) < 0 \qquad \forall (x,y) \in I\!\!D \\
b) \quad & x_0 \frac{\partial^2 g}{\partial x \partial y}(x_0,x_0) \equiv -d \quad \forall x_0 \in (0;1], \quad d > 0,
\end{aligned} \tag{3.1}
$$

verschärft, so ergibt sich eine wichtige Klasse von Diskrepanzen $\mathcal{D}^* \subset \mathcal{D}$. Die Bedingung $4.^*b)$ erscheint nur auf den ersten Blick restriktiv. Sie wird z. B. von den von Read & Cressie (1988) behandelten „divergence measures“ erfüllt. Diese haben, abgesehen vom konstanten Faktor 2, die folgende Form:

$$
\begin{aligned}
\Delta_g^\lambda(\pi, h(\theta)) &= \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k \pi_i [(\frac{\pi_i}{h_i(\theta)})^\lambda - 1], \\
& \quad \lambda \in I\!\!R \setminus \{0; -1\} \\
\text{bzw. } \Delta_g^0(\pi, h(\theta)) &= 2 \sum_{i=1}^k \pi_i \log(\frac{\pi_i}{h_i(\theta)}) \\
\text{und } \Delta_g^{-1}(\pi, h(\theta)) &= 2 \sum_{i=1}^k h_i(\theta) \log(\frac{h_i(\theta)}{\pi_i}).
\end{aligned}
$$

Für $\lambda = 1$ ergibt sich als Spezialfall die χ^2-Diskrepanz. Die Log-Likelihood-Ratio-Statistik G^2 und die modifizierte Neyman-χ^2-Statistik lassen sich in Beziehung zu $\Delta_g^0(\pi, h(\theta))$ bzw. $\Delta_g^{-2}(\pi, h(\theta))$ setzen, wie es analog mit der χ^2-Prüfgröße in Abschnitt 2.2.1 geschehen ist. Die Menge $\mathcal{D}^*$ enthält weitere Diskrepanzen, denen eine bekannte Prüfgröße zuzuordnen ist.

Mit der bisher verwendeten Schreibweise folgt:

$$
\begin{aligned}
g^\lambda(x,y) &= \frac{2}{\lambda(\lambda+1)} x[(\frac{x}{y})^\lambda - 1] \quad , \lambda \in I\!\!R \setminus \{0; -1\} \\
\text{bzw. } g^0(x,y) &= 2x \log\,(\frac{x}{y}) \\
\text{und } g^{-1}(x,y) &= 2y \log\,(\frac{y}{x}).
\end{aligned}
$$

Der Nachweis, daß die Menge $\mathcal{D}^{\mathcal{CR}}$ der Cressie-Read-Diskrepanzen eine Teilmenge von $\mathcal{D}^*$ bildet, bereitet keinerlei Schwierigkeiten.

Falls die reelle Funktion $s(t)$ zweimal stetig differenzierbar auf $I\!R^+$ ist und für alle $t \in I\!R^+$ die Ungleichung

$$2s'(t) + ts''(t) > 0 \tag{3.2}$$

erfüllt ist, ist selbst die allgemeinere Klasse von Diskrepanzen der Form

$$\Delta_g(\pi, h(\theta)) = \sum_{i=1}^{k} \pi_i s(\frac{\pi_i}{h_i(\theta)}) \tag{3.3}$$

in $\mathcal{D}^*$ enthalten. Die genannte Ungleichung wird z. B. von allen konvexen und monoton wachsenden Funktionen erfüllt, zweifache Differenzierbarkeit vorausgesetzt. Die Menge dieser Diskrepanzen soll mit $\mathcal{D}^{\mathcal{Q}}$ bezeichnet werden. Es ist leicht zu zeigen, daß $\mathcal{D}^{\mathcal{CR}}$ eine Teilmenge von $\mathcal{D}^{\mathcal{Q}}$ ist und wiederum $\mathcal{D}^{\mathcal{Q}}$ selbst eine Teilmenge von $\mathcal{D}^*$.

Bezeichnet Δ wieder die χ^2-Diskrepanz, so ergibt sich die folgende Schachtelung der bisher erwähnten Diskrepanzen:

$$\Delta \in \mathcal{D}^{\mathcal{CR}} \subset \mathcal{D}^{\mathcal{Q}} \subset \mathcal{D}^* \subset \mathcal{D}$$

3.3 Annahmen und asymptotische Resultate

3.3.1 Hinreichende Bedingungen für die Ableitung der asymptotischen Resultate

Die Konstruktion von Schätzern erfolgt nun durch Minimierung von *empirischen* Diskrepanzen, d. h. Diskrepanzen zwischen dem vorliegenden Vektor z der relativen Häufigkeiten und $h(\theta)$.

Sei $\Delta_g \in \mathcal{D}$. Es wird nun eine Liste von Bedingungen für die approximierende Modellfamilie gegeben, die für die Existenz von Schätzern und die Ableitungen der asymptotischen Resultate benötigt werden (vgl. Linhart & Zucchini, 1986, S. 238):

B1 $\Theta \subset I\!R^s$ ist kompakt.

B2 $\theta_0 = \arg\min\{\Delta_g(\pi, h(\theta)), \theta \in \Theta)\}$ ist eindeutig und innerer Punkt von Θ. Es existiert eine s-dimensionale Umgebung von θ_0, die vollständig in Θ enthalten ist.

B3 $h(\theta)$ ist stetig auf Θ.

B4 $h(\theta)$ ist zweimal stetig differenzierbar in einer offenen Umgebung $H^*(\theta_0) \subset \Theta$.

B5 Ω ist positiv definit mit $\Omega_{ij} := \left.\dfrac{\partial^2 \Delta_g(\pi, h(\theta))}{\partial\theta_i \partial\theta_j}\right|_{\theta=\theta_0}$, $i, j = 1, \ldots, s$.

In der Literatur wird häufig auf Birch (1964) verwiesen. Er untersucht Bedingungen, unter denen die χ^2-Prüfgröße unter der Hypothese asymptotisch χ^2_{k-s-1}-verteilt ist. Die von ihm gestellten Anforderungen sind geringer als **B1-B5**, sind jedoch auf den Fall korrekter Modellspezifikation beschränkt. Außerdem sind sie zu schwach, um in Kapitel 5 für die Konstruktion von Bootstrap-Tests verwendet zu werden.

Die Voraussetzungen von Linhart & Zucchini (1986) sind im Vergleich zu **B1-B5** allgemein formuliert und beschränken sich nicht auf spezielle Diskrepanzen. Ihre Bedingungen dienen hauptsächlich dazu, die Existenz und Eigenschaften von bestimmten Momenten herzuleiten.

In dieser Arbeit steht jedoch die *Verteilungsfunktion* von Prüfgrößen im Mittelpunkt. Die Herleitung der erforderlichen Resultate basiert auf dem Satz über implizite Funktionen und auf der Anwendung der multivariaten Delta-Methode. Die gleiche Grundidee verwendet Cox (1984) für den Spezialfall des Maximum-Likelihood-Schätzers bei korrekter Modellspezifikation.

Die weitere Vorgehensweise sei nun kurz skizziert. **B4** sichert in Verbindung mit der Definition von Δ_g die Existenz der Matrix Ω. Aus **B5** folgt, daß $\Delta_g(\pi, h(\theta))$ in θ_0 ein *lokales* Minimum annimmt, das nach **B2** gleichzeitig das *globale* Minimum ist. Der Satz über implizite Funktionen zeigt, daß für p in einer Umgebung von π eine differenzierbare Funktion $\hat{\theta}(p)$ existiert, sodaß $\Delta(p, h(\theta))$ ein eindeutiges *lokales* Minimum in $\hat{\theta}(p)$ annimmt. Um sicherstellen zu können, daß dieses Minimum auch das eindeutige globale ist, wird Bedingung **B1** angenommen. **B1** kann durch jede andere Bedingung ersetzt werden, aus der die Eindeutigkeit des globalen Minimums von $\Delta_g(p, h(\theta))$ für Werte von p in einer Umgebung von π folgt. Obwohl **B3** für die Herleitung der asymptotischen Resultate nicht erforderlich ist, wurde diese Bedingung in die Liste aufgenommen, damit für beliebiges $p \in (0, 1]^k$ das Minimum von $\Delta_g(p, h(\theta))$ angenommen wird.

Die Bedingungen **B1-B5** *werden ab jetzt, falls nicht anders vermerkt, für sämtliche folgenden Ableitungen und Beweise vorausgesetzt.*

Es folgt nun die ausführliche Ableitung. Dazu wird zunächst das rein analytische Problem betrachtet, die Diskrepanz $\Delta_g(p, h(\theta))$ für fest vorgegebenes $p \in [0,1]^k$ zu minimieren. Jedes θ, für das das Minimum angenommen wird, soll Minimum-Diskrepanz-Parameter (MDP) genannt werden. Zunächst soll die Frage der Existenz von MDP beantwortet werden:

Lemma 3.3 *Sei $\Delta_g \in \mathcal{D}$ und* **B1** *und* **B3** *erfüllt. Dann existiert für beliebiges $p \in (0,1]^k$ mindestens ein Minimum-Diskrepanz-Parameter θ.*

Beweis: $\Delta_g(p, h(\theta)) = \sum_{i=1}^k g(p_i, h_i(\theta))$ ist für gegebenes $p \in (0,1]^k$ stetig auf der kompakten Menge Θ und nimmt daher sein Minimum für mindestens ein $\theta \in \Theta$ an. □

Das Lemma erfaßt nicht den Fall, daß ein $p_i = 0$ ist. Dieser Ausnahmefall ist jedoch für die asymptotischen Ableitungen ohne Bedeutung. Läßt sich die Funktion g auf $\mathbb{D}^* = [0,1] \times (0,1]$ stetig fortsetzen, so ist Lemma 3.3 auch für $p \geq \mathbf{0}$ richtig. Existiert diese stetige Fortsetzung nicht, so sei im Fall, daß ein $p_i = 0$ ist, per definitionem jedes $\theta \in \Theta$ Minimum-Diskrepanz-Parameter.

Die Anwendung des Satzes über implizite Funktionen liefert nun die Eindeutigkeit des MDP in einer Umgebung von π:

Satz 3.1 *Unter* **B1-B5** *gibt es eine offene δ^*-Umgebung $U_{\delta^*}(\pi) \subset (0,1)^k$, so daß für jedes $p \in U_{\delta^*}(\pi)$ genau ein MDP $\hat{\theta}(p)$ existiert.*

Beweis:
Siehe Anhang A.2. □

Es sei zunächst festgehalten, daß für p nicht notwendigerweise die Bedingung $\sum_{i=1}^k p_i = 1$ gelten muß. Satz 3.1 liefert nun in Verbindung mit Lemma 3.3 die Existenz einer Abbildung $\hat{\theta} : [0,1]^k \to \mathbb{R}$, die jedem p den MDP $\hat{\theta}(p)$ zuordnet. Diese Abbildung ist auf $U_{\delta^*}(\pi)$ eindeutig, kann aber außerhalb dieser Umgebung bei nicht eindeutigen MDP auf mehrere Arten definiert werden. Möglicherweise kann unter den gegebenen Voraussetzungen erreicht werden, daß $\hat{\theta}(p)$ stetig ist, aber dieser Gedanke soll nicht weiter

verfolgt werden, da im vorliegenden Fall nur das Aussehen von $\hat{\theta}(p)$ in einer Umgebung von π für die asymptotischen Verteilungen eine Rolle spielt.

Für die Anwendung des multivariaten Delta-Theorems muß nun gesichert werden, daß $\hat{\theta}(p)$ in π total differenzierbar ist. Unter **B1-B5** folgt die stärkere, also hinreichende Aussage:

Satz 3.2 *Die im Satz 3.1 definierte Abbildung $\hat{\theta}(p)$ ist unter* **B1-B5** *in einer Umgebung von π stetig differenzierbar.*

Beweis:
Die Behauptung folgt aus dem Satz 170.2 in Heuser (1981). Die Matrix der Ableitungen im Punkte π läßt sich explizit angeben. Die $s \times k$-Matrix B sei definiert durch

$$B_{ij} := \left.\frac{\partial^2 \Delta_g}{\partial \theta_i \partial p_j}\right|_{(p,\theta)=(\pi,\theta_0)} = \frac{\partial^2 g}{\partial x \partial y}(\pi_j, h_j(\theta_0))\frac{\partial h_j}{\partial \theta_i}, \qquad \begin{array}{l} i = 1, \ldots, s \\ j = 1, \ldots, k, \end{array}$$

dann kann die gesuchte $s \times k$-Matrix T mit

$$T_{ij} = \left.\frac{\partial \hat{\theta}_i(p)}{\partial p_j}\right|_{p=\pi} \qquad i = 1, \ldots, s \quad j = 1, \ldots, k,$$

durch

$$T = -\Omega^{-1} B$$

ausgedrückt werden. □

Die für die Anwendung der Delta-Methode benötigte Jacobi-Matrix T kann somit berechnet werden, ohne daß eine explizite Darstellung von $\hat{\theta}(p)$ in Abhängigkeit von p notwendig ist.

3.3.2 Asymptotische Resultate

Sei nun eine Folge $X_1, X_2, X_3, \ldots$ von unabhängig und identisch unter dem Grundmodell verteilten Zufallsvariablen gegeben. Mit F_n werde wieder die empirische Verteilungsfunktion der Realisationen $x_1, \ldots, x_n$ bezeichnet. Der Vektor $z^{(n)}$ der relativen Häufigkeiten sei für gegebenes n wie in Definition 2.1 gebildet. Nach den Ausführungen des letzten Abschnitts existiert zu gegebenem $\Delta_g \in \mathcal{D}$ ein Minimum-Diskrepanz-Schätzer $\hat{\theta}(z^{(n)})$.

Für den Vektor $z^{(n)}$ der relativen Häufigkeiten folgt als Konsequenz der multivariaten Version des zentralen Grenzwertsatzes (Rao, 1973, S. 128):

$$\sqrt{n}(z^{(n)} - \pi) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T) \tag{3.4}$$

Fast jede in diesem und den folgenden Abschnitten genannte asymptotische Verteilung hat diese Aussage als Grundlage und wird durch wiederholte Anwendung der multivariaten Delta-Methode hergeleitet.

Zum Zwecke einer knapperen Darstellung soll die Vereinbarung gelten, daß Werte von Abbildungen bzw. Funktionen, bei denen keine Variablen angegeben werden, immer an der Stelle π und / oder θ_0 gebildet werden:

$$\text{Bsp. :} \qquad \frac{\partial h_i}{\partial \theta_m} \text{ steht für } \left.\frac{\partial h_i(\theta)}{\partial \theta_m}\right|_{\theta=\theta_0}$$
$$\frac{\partial \hat{\theta}_m}{\partial p_j} \text{ steht für } \left.\frac{\partial \hat{\theta}_m(p)}{\partial p_j}\right|_{p=\pi}$$

Definition 3.3 *Sei $C = C(\theta_0)$ die $s \times k$-Matrix mit $(C(\theta))_{ij} := \dfrac{\partial h_j(\theta)}{\partial \theta_i}$, $i = 1, \ldots, s$, $j = 1, \ldots, k$.*

Definition 3.4 *Sei G die $k \times k$-Diagonalmatrix mit den Elementen*

$$G_{ii} = \frac{\partial^2 g}{\partial x \partial y}(\pi_i, h_i(\theta_0)), \quad i = 1, \ldots k$$

Satz 3.3 *Seien B und T wie im Beweis zu Satz 3.2 definiert und $\hat{\theta}_n := \hat{\theta}(z^{(n)})$. Unter* **B1-B5** *folgt*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(\mathbf{0}, A),$$

mit

$$A = \Omega^{-1} C G (D_\pi - \pi\pi^T) G C^T \Omega^{-1} \quad .$$

Beweis:
1. Seien C und G wie in den Definitionen 3.3 und 3.4, dann folgt nach Satz 3.2:

$$B = CG,$$
$$T = -\Omega^{-1} C G$$

Da $\hat{\theta}(p)$ nach Satz 3.2 in einer Umgebung von π *stetig* differenzierbar ist, liefert nun die Delta-Methode aus Satz 2.4, angewandt auf (3.4) die gewünschte Aussage. $\square$

Die χ^2-Prüfgröße $n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))$ läßt sich als quadratische Form $Y_n^T Y_n$ mit

$$Y_n := \sqrt{n}\left(\frac{z_1^{(n)} - h_1(\hat{\theta}(z^{(n)}))}{\sqrt{h_1(\hat{\theta}(z^{(n)}))}}, \ldots, \frac{z_k^{(n)} - h_k(\hat{\theta}(z^{(n)}))}{\sqrt{h_k(\hat{\theta}(z^{(n)}))}}\right)^T$$

schreiben. Für die Ermittlung der asymptotischen Verteilung dieser quadratischen Form wird die asymptotische Verteilung von Y_n benötigt:

Satz 3.4 *Sei T wie in Satz 3.2 definiert und* **B1-B5** *erfüllt. Die Matrix W sei definiert als*

$$W = D_h^{-\frac{1}{2}}[I - \frac{1}{2}(I + D_\pi D_h^{-1})C^T T]$$

Dann läßt sich folgende asymptotische Aussage ableiten:

$$\sqrt{n}\left[\begin{pmatrix} \frac{z_1^{(n)} - h_1(\hat{\theta}(z^{(n)}))}{\sqrt{h_1(\hat{\theta}(z^{(n)}))}} \\ \vdots \\ \frac{z_k^{(n)} - h_k(\hat{\theta}(z^{(n)}))}{\sqrt{h_k(\hat{\theta}(z^{(n)}))}} \end{pmatrix} - \begin{pmatrix} \frac{\pi_1 - h_1(\theta_0)}{\sqrt{h_1(\theta_0)}} \\ \vdots \\ \frac{\pi_k - h_k(\theta_0)}{\sqrt{h_k(\theta_0)}} \end{pmatrix}\right] \xrightarrow{d} N(0, W(D_\pi - \pi\pi^T)W^T)$$

Beweis:
Sei $w(p) : [0,1]^k \to I\!R^k$ die Abbildung mit

$$w_i(p) = \frac{p_i - h_i(\hat{\theta}(p))}{\sqrt{h_i(\hat{\theta}(p))}}, \quad i = 1, \ldots, k. \tag{3.5}$$

Nach Satz 3.2 und **B4** ist $w(p)$ in einer Umgebung von π stetig differenzierbar, und durch Anwendung der Kettenregel folgt:

$$\left.\frac{\partial w_i}{\partial p_j}\right|_{p=\pi} = \frac{1}{\sqrt{h_i(\theta_0)}}\left[\delta_{ij} - (\frac{1}{2}\sum_{m=1}^{s}\frac{\partial h_i}{\partial \theta_m}\frac{\partial \hat{\theta}_m}{\partial p_j})(1 + \frac{\pi_i}{h_i(\theta_0)})\right], \quad i, j = 1, \ldots, k$$

Die Matrix W der Ableitungen läßt sich nun durch die bisher verwendeten Matrizen ausdrücken:

$$W = D_h^{-\frac{1}{2}}[I - \frac{1}{2}(I + D_\pi D_h^{-1})C^T T]$$

Die Behauptung folgt dann aus der Anwendung der Delta-Methode auf (3.4).

3.4 Approximationen der Verteilung der χ^2-Prüfgröße, zusammengesetzte Hypothese

3.4.1 Korrekt spezifizierte Modellfamilie

Sei Δ_g eine Diskrepanz aus $\mathcal{D}$, und die Modellfamilie sei korrekt spezifiziert, d. h., es gibt ein $\theta_0 \in \Theta$ mit $h(\theta_0) = \pi$. Unter **B1-B5** folgt aus Satz 3.4:

$$\sqrt{n} w(z^{(n)}) \xrightarrow{d} N(\mathbf{0}, W(D_\pi - \pi\pi^T)W^T), \tag{3.6}$$

wobei $w(p)$ wie in (3.5) definiert ist und W sich zu

$$W = D_\pi^{-\frac{1}{2}}[I + C^T \Omega^{-1} C G]$$

vereinfacht. Um die asymptotische Verteilung von $n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) = nw(z^{(n)})^T w(z^{(n)})$ zu bestimmen, genügt es nach Satz 2.1, die Verteilung von $Y^T Y$ zu ermitteln, wobei

$$Y \sim N(\mathbf{0}, W(D_\pi - \pi\pi^T)W^T).$$

Sei $Z := W(D_\pi - \pi\pi^T)W^T$, dann ist $Y^T Y$ nach Satz 2.2 verteilt wie

$$\sum_{\lambda_i > 0} \lambda_i \chi_1^2, \tag{3.7}$$

wobei die λ_i die Eigenwerte von Z sind. Abgesehen von der Tatsache, daß mindestens ein λ_i den Wert null annimmt, dies folgt aus der Singularität von Z, lassen sich a priori keine weiteren Aussagen über die Eigenwerte von Z machen. Entstammt die Diskrepanz Δ_g jedoch der Menge $\mathcal{D}^*$, dann ergibt sich die stärkere Aussage:

Satz 3.5 *Ist* $\Delta_g \in \mathcal{D}^*$ *und sind* **B1-B5** *erfüllt, so gilt*

$$n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) \xrightarrow{d} \chi^2_{k-s-1}$$

Beweis:
Zu zeigen ist: Es gibt $k-s-1$ von 0 verschiedene Eigenwerte von Z mit dem Wert eins.
1. Da unter der Annahme $\pi = h(\theta_0)$

$$\Omega_{ij} = \sum_{l=1}^{k} \frac{\partial^2 g}{\partial y^2}(\pi_l, h_l(\theta_0)) \frac{\partial h_l}{\partial \theta_i} \frac{\partial h_l}{\partial \theta_j}, \quad i,j = 1, \ldots, s,$$

ist, folgt nach Lemma 3.2

$$\Omega = -CGC^T$$

Da $-G$ positiv definit ist, muß C den maximalen Zeilenrang s besitzen.
2. Ist $\Delta_g \in \mathcal{D}^*$, so ist nach (3.1) für beliebiges $x_0 \in (0,1]$ die zusätzliche Bedingung

$$x_0 \frac{\partial^2 g}{\partial x \partial y}(x_0, x_0) = -d \quad , d > 0,$$

erfüllt:

$$\begin{aligned} \Rightarrow \quad & G = -dD_\pi^{-1} \\ \Rightarrow \quad & \Omega = dCD_\pi^{-\frac{1}{2}} D_\pi^{-\frac{1}{2}} C^T \end{aligned}$$

3. Aus der Tatsache, daß $\sum_{i=1}^{k} h_i(\theta) = 1$ gilt, folgt weiterhin $CG\pi = -dCD_\pi^{-1}\pi = \mathbf{0}$
4. Setzt man $M := D_\pi^{-\frac{1}{2}} C^T$ und $\sqrt{\pi} := (\sqrt{\pi_1}, \ldots, \sqrt{\pi_k})^T$, so folgt unter Verwendung von 1., 2. und 3. nach einer kurzen Rechnung :

$$\begin{aligned} W &= [I - M(M^TM)^{-1}M^T]D_\pi^{-\frac{1}{2}} \\ Z &= W(D_\pi - \pi\pi^T)W^T = \ldots = I - \sqrt{\pi}\sqrt{\pi}^T - M(M^TM)^{-1}M^T \end{aligned}$$

Dies ist die gleiche Matrix wie die von Bishop et al. (1975, S. 517) für den Fall des Maximum-Likelihood-Schätzers hergeleitete Matrix. In ihrem Beweis zeigen sie, daß Z idempotent und symmetrisch ist, daher nur Eigenwerte besitzt, die 0 oder 1 sind. Weiterhin ergibt sich die Spur von Z zu $tr(Z) = k-s-1$, und damit folgt unter Verwendung von (3.7) die Behauptung. □

An dieser Stelle wird die Bedeutung der Menge $\mathcal{D}^*$ deutlich. Unter der Annahme korrekter Modellspezifikation sind die so gewonnenen Minimum-Diskrepanz-Schätzer „best asymptotic normal“ (BAN), d. h., daß (siehe z. B. Read & Cressie, 1988, S. 164)

$$\hat{\theta}(z^{(n)}) = \hat{\theta}(\pi) + (M^T M)^{-1} M^T D_\pi^{-\frac{1}{2}} (z^{(n)} - \pi) + o_p(n^{-\frac{1}{2}}) \tag{3.8}$$

gilt, wobei M wie oben definiert ist. Daß diese Bedingung für Schätzer zu $\Delta_g \in \mathcal{D}^*$ unter **B1-B5** erfüllt ist, folgt aus der Gleichung (A.1) aus dem Beweis zu Satz 2.3 im Anhang und den Vereinfachungen im Beweis zu Satz 3.5.

Read & Cressie (1988) zeigen, daß unter den bereits in Abschnitt 3.3.1 erwähnten Annahmen von Birch folgt, daß die Minimum-Diskrepanz-Schätzer zu den Cressie-Read-Diskrepanzen $\mathcal{D}^{\mathcal{CR}}$, welche ja in $\mathcal{D}^*$ enthalten sind, BAN sind und zeigen weiterhin, daß die χ^2-Prüfgröße asymptotisch χ^2_{k-s-1}-verteilt ist, wenn BAN-Schätzer eingesetzt werden.

Wie bereits erwähnt, sind die Annahmen von Birch (1964) auf den Fall korrekter Modellspezifikation beschränkt. Es sei deshalb an dieser Stelle angemerkt, daß es nur bei Beschränkung der Diskrepanzen auf $\mathcal{D}^*$ und zutreffender Hypothese einfacher gewesen wäre zu zeigen, daß BAN-Schätzer vorliegen, um dann die Ergebnisse von Read & Cressie (1988) zu verwenden. Der Ansatz von Cox (1984), der den Satz über implizite Funktionen verwendet, sich ebenfalls auf korrekte Modelle beschränkt, nur den Maximum-Likelihood-Schätzer untersucht und auch stärkere Voraussetzungen benötigt, erscheint von der Idee her leichter übertragbar auf den Fall, daß die Modellfamilie nicht korrekt spezifiziert wird. Der Vorteil der Delta-Methode, die schnelle Ableitung von asymptotischen Resultaten unter zugegebenermaßen stärkeren Voraussetzungen, wird für $\pi \notin \{h(\theta), \theta \in \Theta\}$ besonders deutlich.

3.4.2 Verteilung unter einer festen Alternative

3.4.2.1 Allgemeiner Fall

Ist die Hypothese zusammengesetzt, aber $\pi \notin \{h(\theta) : \theta \in \Theta\}$, dann kann unter **B1-B5** für $\sqrt{n}(\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0)))$ eine asymptotische Verteilung abgeleitet werden:

Satz 3.6 *Sei* $\Delta_g \in \mathcal{D}$. *Unter* **B1-B5** *und unter der Annahme, daß* $\pi \notin \{h(\theta) : \theta \in \Theta\}$ *ist, folgt*

$$\begin{aligned} & \sqrt{n}(\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))) \xrightarrow{d} N(0, \sigma^2), \\ \textit{mit} \quad & \sigma^2 = F(D_\pi - \pi\pi^T)F^T \\ \textit{und} \quad & F := 2\pi^T D_h^{-1}[I + \frac{1}{2} D_h^{-1} D_\pi C^T \Omega^{-1} CG] \end{aligned}$$

Beweis:
Da

$$\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) = \sum_{i=1}^{k} \frac{(z_i^{(n)} - h_i(\hat{\theta}(z^{(n)})))^2}{h_i(\hat{\theta}(z^{(n)}))} = \sum_{i=1}^{k} \frac{(z_i^{(n)})^2}{h_i(\hat{\theta}(z^{(n)}))} \quad - 1$$

ist, genügt es, analog zum Abschnitt 2.3.3, die Funktion

$$\Delta(p, h(\hat{\theta}(p))) = \sum_{i=1}^{k} \frac{p_i^2}{h_i(\hat{\theta}(p))} \quad - 1$$

zu betrachten. Diese Funktion ist als Folgerung von Satz 3.2 in einer Umgebung von π stetig differenzierbar. Unter Anwendung der Kettenregel folgt

$$\left.\frac{\partial \Delta}{\partial p_j}\right|_{p=\pi} = 2\frac{\pi_j}{h_j(\theta_0)} - \sum_{i=1}^{k} \frac{\pi_i^2}{h_i^2(\theta_0)} \sum_{m=1}^{s} \frac{\partial h_i}{\partial \theta_m} \frac{\partial \hat{\theta}_m}{\partial p_j}, \quad j = 1, \ldots, k. \tag{3.9}$$

In Matrixschreibweise ergibt sich für die Jacobi-Matrix F:

$$F = 2\pi^T D_h^{-1}[I - \frac{1}{2} D_h^{-1} D_\pi C^T T]$$

Berücksichtigt man noch

$$T = -\Omega^{-1} CG,$$

so folgt

$$F = 2\pi^T D_h^{-1}[I + \frac{1}{2} D_h^{-1} D_\pi C^T \Omega^{-1} CG]$$

Mittels der Delta-Methode und (3.4) ergibt sich nun die Behauptung. □

Für gegebenes n wird deshalb die Verteilung der χ^2-Prüfgröße $n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))$ durch die Verteilung von

$$N(n\Delta(\pi, h(\theta_0)), n\sigma^2) \tag{3.10}$$

approximiert. Leider ist es schwierig, Bedingungen herzuleiten, die sicherstellen, daß σ^2 tatsächlich positiv ist.

3.4.2.2 Die Sonderstellung des Minimum-χ^2-Schätzers

Ist $\Delta_g = \Delta$, sind die erhaltenen Schätzer also die Minimum-χ^2-Schätzer, so vereinfacht sich die asymptotische Verteilung erheblich. Der Subtrahend aus (3.9) läßt sich auch als

$$\sum_{m=1}^{s} \frac{\partial \hat{\theta}_m}{\partial p_j} \sum_{i=1}^{k} \frac{\pi_i^2}{h_i^2(\theta_0)} \frac{\partial h_i}{\partial \theta_m}$$

schreiben. Nun ist

$$-\sum_{i=1}^{k} \frac{\pi_i^2}{h_i^2(\theta_0)} \frac{\partial h_i}{\partial \theta_m}$$

gerade die Ableitung von $\Delta(\pi, h(\theta))$ nach θ_m in θ_0, d. h. dieser Ausdruck verschwindet. Die Jacobi-Matrix F vereinfacht sich zu

$$F = 2\pi^T D_h^{-1},$$

und der Vergleich mit (2.3) ergibt das folgende Lemma:

Lemma 3.4 *Unter den Annahmen $\pi \neq h(\theta_0)$, $\Delta_g = \Delta$ und* **B1-B5** *besitzen*

$$\sqrt{n}(\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0)))$$

und

$$\sqrt{n}(\Delta(z^{(n)}, h(\theta_0)) - \Delta(\pi, h(\theta_0)))$$

die gleiche asymptotische Verteilung.

Werden Minimum-χ^2-Schätzer verwendet, kann zusätzlich zur einfacheren asymptotischen Verteilung der Vorteil genutzt werden, daß σ^2 nach Abschnitt 2.3.3 nicht verschwinden kann.

3.4.3 „Local alternatives“

Analog zu Abschnitt 2.3.4.1 sei eine Folge von alternativen Modellen gegeben, die in bestimmter Weise gegen ein $\pi = h(\theta_0) \in \{h(\theta) : \theta \in \Theta\}$ konvergiert:

$$\pi^{(n)} = \pi + \frac{c}{\sqrt{n}} \quad , c \in I\!R^k \setminus \{\mathbf{0}\}, \sum_{i=1}^{k} c_i = 0 \tag{3.11}$$

Der Vektor $z^{(n)}$ der relativen Häufigkeiten sei analog zu Abschnitt 2.3.4.1 definiert. Nach Serfling (1980, S. 132) existiert dann die folgende asymptotische Verteilung:

$$\sqrt{n}(z^{(n)} - \pi) \xrightarrow{d} N(c, D_\pi - \pi\pi^T) \tag{3.12}$$

Mit den bisherigen Bezeichnungen kann nun der folgende Satz bewiesen werden:

Satz 3.7 *Sei eine Folge von alternativen Modellen wie in (3.11) gegeben. Unter* **B1-B5** *gilt mit den bisherigen Bezeichnungen:*

$$n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) \xrightarrow{d} \sum_{\lambda_i > 0} \lambda_i \chi_1^2(\frac{(SWc)_i^2}{\lambda_i}) + \sum_{\lambda_i = 0} (SWc)_i^2,$$

wobei $W = D_\pi^{-\frac{1}{2}}[I + C^T\Omega^{-1}CG]$ *ist, die Zeilen von* S *aus einer orthonormalen Basis aus Eigenvektoren von* $W(D_\pi - \pi\pi^T)W^T$ *bestehen und die* λ_i *die zugehörigen Eigenwerte bezeichnen.*

Beweis:
1. Die im Beweis zu Satz 3.4 definierte Abbildung $w(p)$ ist in einer Umgebung von π stetig differenzierbar , d. h. die Voraussetzungen des Satzes 2.4 zur Anwendung auf (3.12) sind erfüllt.
2. Die Jacobi-Matrix W im Punkte π ist unter der Annahme $\pi = h(\theta_0)$ durch (vgl. Abschnitt 3.4.1)

$$W = D_\pi^{-\frac{1}{2}}[I + C^T\Omega^{-1}CG]$$

gegeben.
3. Satz 2.4 liefert nun:

$$\sqrt{n}(w(z^{(n)}) - w(\pi)) \xrightarrow{d} N(Wc, W(D_\pi - \pi\pi^T)W^T),$$

Da $w(\pi) = 0$ ist und die χ^2-Prüfgröße $n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))$ sich als $nw(z^{(n)})^T w(z^{(n)})$ schreiben läßt, genügt es nach Satz 2.1, die Verteilung von Y^TY zu bestimmen, mit

$$Y \sim N(Wc, W(D_\pi - \pi\pi^T)W^T).$$

4. Satz 2.2 sagt aus, daß die Verteilung von Y^TY durch

$$Y^TY \sim \sum_{\lambda_i > 0} \lambda_i \chi_1^2(\frac{(SWc)_i^2}{\lambda_i}) + \sum_{\lambda_i = 0} (SWc)_i^2$$

gegeben ist, wobei S und die λ_i wie oben definiert sind. $\square$

Wird die Diskrepanz Δ_g wieder aus der Menge $\mathcal{D}^*$ gewählt, so vereinfacht sich die asymptotische Verteilung erheblich:

Satz 3.8 *Ist $\Delta_g \in \mathcal{D}^*$ und sind* **B1-B5** *erfüllt, so gilt*

$$n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) \xrightarrow{d} \chi^2_{k-s-1}(\delta)$$

mit $\delta = c^T D_\pi^{-\frac{1}{2}}(I - M(M^TM)^{-1}M^T)D_\pi^{-\frac{1}{2}}c$. M ist wieder als $M := D_\pi^{-\frac{1}{2}}C^T$ definiert.

Beweis:

1. Für $\Delta_g \in \mathcal{D}^*$ vereinfacht sich die Matrix W zu (vgl. Beweis zu Satz 3.5)

$$W = D_\pi^{-\frac{1}{2}}[I + C^T\Omega^{-1}CG] = [I - M(M^TM)^{-1}M^T]D_\pi^{-\frac{1}{2}}.$$

2. Für $Z := W(D_\pi - \pi\pi^T)W^T$ folgt

$$Z = I - \sqrt{\pi}\sqrt{\pi}^T - M(M^TM)^{-1}M^T.$$

Z hat nach Satz 3.5 $k - s - 1$ von 0 verschiedene Eigenwerte mit dem Wert eins.

3. Nach Satz 2.2 gibt es eine orthonormale Basis aus Eigenvektoren, deren Transponierte die Zeilen der Matrix S bilden. Sei s_i ein Eigenvektor aus dieser Basis zu einem Eigenwert $\lambda_i = 0$

$$\begin{aligned}
\Rightarrow \quad & (I - \sqrt{\pi}\sqrt{\pi}^T - M(M^TM)^{-1}M^T)s_i = \mathbf{0} \\
\Rightarrow (SWc)_i &= s_i^T Wc = s_i^T[I - M(M^TM)^{-1}M^T]D_\pi^{-\frac{1}{2}}c \\
&= s_i^T[I - \sqrt{\pi}\sqrt{\pi}^T - M(M^TM)^{-1}M^T]D_\pi^{-\frac{1}{2}}c \\
&\quad + s_i^T\sqrt{\pi}\sqrt{\pi}^T D_\pi^{-\frac{1}{2}}c \\
&= s_i^T\sqrt{\pi}(1,\ldots,1)c = 0 \quad \text{, da } \sum_{i=1}^k c_i = 0 \text{ ist.}
\end{aligned}$$

4. Da S orthogonal ist, gilt:

$$\begin{aligned}
\sum_{i=1}^k (SWc)_i^2 &= \sum_{i=1}^k (Wc)_i^2 = c^T W^T W c \\
&= c^T D_\pi^{-\frac{1}{2}}[I - M(M^TM)^{-1}M^T][I - M(M^TM)^{-1}M^T]D_\pi^{-\frac{1}{2}}c \\
&= c^T D_\pi^{-\frac{1}{2}}[I - M(M^TM)^{-1}M^T]D_\pi^{-\frac{1}{2}}c
\end{aligned}$$

$\Rightarrow$ Aus 1.–4. folgt mit dem Ergebnis des letzten Satzes die Behauptung. $\square$

Für Minimum-Diskrepanz-Schätzer zu beliebigem $\Delta_g \in \mathcal{D}^*$ ist deshalb die asymptotische Verteilung der χ^2-Prüfgröße unter „local alternatives“ identisch.

Es soll nun veranschaulicht werden, wie unter den eben genannten Voraussetzungen die Verteilung der χ^2-Prüfgröße für eine unter dem Modell $\pi^* \notin \{h(\theta) : \theta \in \Theta\}$ entstandene Stichprobe der Größe n approximiert wird. Dazu wird π^* als das Element $\pi^{(n)}$ der in (3.11) genannten Folge von Modellen angesehen. Diese Folge strebt gegen ein Modell $\pi = h(\theta_0) \in \{h(\theta) : \theta \in \Theta\}$. Dieses Modell ist *nicht* eindeutig festgelegt, denn es liegt ja eine zusammengesetzte Hypothese vor. Sei $h(\theta_0)$ zunächst ein beliebiges Element aus $\{h(\theta) : \theta \in \Theta\}$. Setzt man $\pi^{(n)} = \pi^*$, so nimmt c den Wert

$$c = \sqrt{n}(\pi^* - h(\theta_0))$$

an. Der Nichtzentralitätsparameter ergibt sich dann zu

$$\begin{aligned}
\delta &= n\Delta(\pi^*, h(\theta_0)) \\
&\quad -n(\pi^* - h(\theta_0))^T D_\pi^{-\frac{1}{2}} M(M^T M)^{-1} M^T D_\pi^{-\frac{1}{2}} (\pi^* - h(\theta_0)) \\
&= n\Delta(\pi^*, h(\theta_0)) - n\pi^{*T} D_\pi^{-\frac{1}{2}} M(M^T M)^{-1} M^T D_\pi^{-\frac{1}{2}} \pi^*
\end{aligned}$$

Für $\Delta_g \in \mathcal{D}^*$ wird nun die χ^2-Verteilung mit $k - s - 1$ Freiheitsgraden und dem Nichtzentralitätsparameter δ als Approximation für die Verteilung der χ^2-Prüfgröße verwendet.

Mitra (1958) leitet die asymptotische Verteilung bei „local alternatives“ in einem etwas anderen Rahmen für den Fall des Maximum-Likelihood-Schätzers her. Während er zum gleichen Nichtzentralitätsparameter wie in Satz 3.8 gelangt, wird er meines Erachtens von Bishop et al. (1975) und später von Read & Cressie (1988) ungenau zitiert, diese Autoren verwenden den Nichtzentralitätsparameter $\delta = c^T D_\pi^{-1} c$ bzw $n\Delta(\pi^*, h(\theta_0))$, der sich nur dann ergibt, falls $h(\theta_0)$ so gewählt wird, daß die in Abschnitt 3.2.2 erwähnte, zu Maximum-Likelihood-Schätzern führende Diskrepanz

$$\Delta_g^0(\pi^*, h(\theta)) = 2\sum_{i=1}^{k} \pi_i^* \log \frac{\pi_i^*}{h_i(\theta)},$$

minimiert wird:

$$\begin{aligned} 0 \stackrel{!}{=} \frac{\partial \Delta_g^0}{\partial \theta_j} &= -2\sum_{i=1}^{k} \frac{\pi_i^*}{h_i(\theta)} \frac{\partial h_i(\theta)}{\partial \theta_j}, j = 1, \ldots, p \\ \Rightarrow & \quad CD_h^{-1}\pi^* = \mathbf{0} \\ \Rightarrow & \quad M^T D_\pi^{-\frac{1}{2}} \pi^* = CD_\pi^{-1}\pi^* = CD_h^{-1}\pi^* = \mathbf{0} \\ \Rightarrow & \quad \delta = n\Delta(\pi^*, h(\theta_0)) \end{aligned}$$

Während bei Mitra $h(\theta_0)$ ein beliebiges Element der zusammengesetzten Hypothese sein kann, ist es also bei Bishop et al. bzw. bei Read & Cressie implizit festgelegt; die letztgenannten Autoren äußern sich jedoch nicht zur Wahl von $h(\theta_0)$, und deshalb sind die letzten Zeilen nur ein Versuch, ihr Ergebnis zu retten.

Naheliegender ist es aber, zur Bestimmung von $h(\theta_0)$ entweder die für die Ermittlung der Schätzer verwendete Diskrepanz $\Delta_g(\pi^*, h(\theta))$ oder die χ^2-Diskrepanz $\Delta(\pi^*, h(\theta))$ zu minimieren.

3.4.4 Eine verbesserte Approximation für die zusammengesetzte Hypothese

3.4.4.1 Theoretische Ableitung

Der abgewandelte Ansatz von Drost et al. (1989), der in Abschnitt 2.3.6 für den Fall der einfachen Hypothese anhand der χ^2-Diskrepanz erläutert wurde, läßt sich mit den bisher hergeleiteten Resultaten leicht auf den Fall der zusammengesetzten Hypothese übertragen. Seien π das Grundmodell, $\{h(\theta) : \theta \in \Theta\}$ die Menge der möglichen approximierenden Modelle, $\Delta_g \in \mathcal{D}$ und **B1-B5** erfüllt. In Satz 3.4 wurde mit den dort gegebenen Bezeichnungen die Gültigkeit der asymptotischen Aussage

$$\sqrt{n}(w(z^{(n)}) - w(\pi)) \xrightarrow{d} N(\mathbf{0}, W(D_\pi - \pi\pi^T)W^T)$$

mit $W = D_h^{-\frac{1}{2}}[I + \frac{1}{2}(I + D_\pi D_h^{-1})C^T\Omega^{-1}CG)]$ bewiesen. Die Verteilung der χ^2-Prüfgröße $nw(z^{(n)})^T w(z^{(n)})$ kann nun durch die Verteilung von nY^TY approximiert werden, wobei gilt:

$$\begin{aligned} &\sqrt{n}(Y - w(\pi)) \sim N(\mathbf{0}, W(D_\pi - \pi\pi^T)W^T). \qquad (3.13) \\ &\Rightarrow \sqrt{n}Y \sim N(\sqrt{n}w(\pi), W(D_\pi - \pi\pi^T)W^T) \end{aligned}$$

Die Anwendung des Satzes 2.2 liefert:

$$nY^TY \sim \sum_{\lambda_i>0} \lambda_i \chi_1^2(\frac{n(Sw(\pi))_i^2}{\lambda_i}) + \sum_{\lambda_i=0} n(Sw(\pi))_i^2 \tag{3.14}$$

S ist hier wieder eine orthogonale Matrix, deren transponierte Zeilen aus einer orthonormalen Basis von Eigenvektoren der Matrix $V := W(D_\pi - \pi\pi^T)W^T$ bestehen.

3.4.4.2 Numerische Berechnungen

Die Erzeugung von Zufallszahlen mit der in (3.14) angegebenen Verteilung stellt kein großes Problem dar, nur die Berechnung des besten Parameters θ_0, der allerdings für alle genannten Approximationen zur Verfügung stehen muß, kann Schwierigkeiten bereiten.

Leider ist die Berechnung der Verteilungsfunktion mit dem in Abschnitt 2.3.6 genannten Algorithmus für den Fall der zusammengesetzten Hypothese mit erheblichem Aufwand verbunden. Die in (2.14) gegebene Darstellung der Verteilungsfunktion als unendliche Reihe führt zu einer sehr großen Anzahl von benötigten Summanden, um eine angemessene Genauigkeit (z. B. 0.0001) zu erzielen.

Es sei der Einfachheit halber $\Delta_g \in \mathcal{D}^*$. Unter der Hypothese nehmen dann die $k - s - 1$ von 0 verschiedenen Eigenwerte λ_i nur den Wert 1 an, und es wird nur ein einziger Summand zur exakten Berechnung benötigt, da unter dieser Voraussetzung die Verteilung der Prüfgröße durch eine zentrale χ^2-Verteilung angenähert wird.

Ist jedoch $\pi \notin \{h(\theta) : \theta \in \Theta\}$, so tritt im allgemeinen die Situation ein, daß ein Eigenwert den Wert null annimmt, $k - s - 1$ Eigenwerte in etwa die gleiche Größenordnung besitzen (in der Nähe der Hypothese betragen die Werte ungefähr 1) und die restlichen Eigenwerte, obwohl größer als 0, im Vergleich deutlich kleiner sind. Bei dieser Konstellation konvergiert der Algorithmus extrem langsam.

Nun ist die Darstellung der Verteilungsfunktion einer quadratischen Form in normalverteilten Zufallsvariablen als χ^2-Reihe nicht der einzig mögliche Ansatz zur numerischen Berechnung. Wie einige Simulationen ergaben, leiden allerdings auch die anderen von Johnson & Kotz (1970, Vol. 2) genannten Ansätze unter Verwendung von Reihenentwicklungen zur Berechnung der

Verteilungsfunktion einer quadratischen Form

$$Q = \sum_{i=1}^{I} \alpha_i (Z_i - c_i)^2,$$

in unabhängig standardnormalverteilten Zufallsvariablen $Z_1, \ldots, Z_I$ unter der gleichen Anfälligkeit gegenüber extrem unterschiedlichen Größenordnungen der Koeffizienten α_i, die ja im vorliegendem Falle den Eigenwerten λ_i entsprechen.

Einen anderen Weg gehen Davies (1973, 1980) bzw. Farebrother (1985, 1990), die die Inversionsmethode nach Gil-Pelaez (1951), aufbauend auf der Arbeit von Imhof (1961), unter anderem auf die charakteristische Funktion einer Linearkombination

$$\sum_{i=1}^{I} \alpha_i \chi^2_{n_i}(c_i^2) \tag{3.15}$$

anwenden. Die von ihnen genannten Algorithmen konvergieren i. allg. relativ schnell, jedoch nicht in dem Fall, daß (3.15) von Termen dominiert wird, bei denen die Freiheitsgrade n_i kleiner als 4 sind (Davies [1973, 1984]). Somit ist auch dieser Weg zur Berechnung der Verteilungsfunktion im vorliegenden Fall nicht geeignet.

Damit die Berechnung der Verteilungsfunktion der Approximation nicht mehr Zeit in Anspruch nimmt als die Berechnung der exakten Verteilungsfunktion, was zumindest für kleinere Werte von n und k bei den erwähnten Ansätzen der Fall ist, erscheint es notwendig, für die Verteilungsfunktion des stochastischen Teils der rechten Seite von (3.14) selbst eine Approximation zu finden.

Mit dem Problem, die Verteilungsfunktion einer quadratischen Form in normalverteilen Zufallsvektoren zu approximieren, haben sich schon diverse Autoren beschäftigt (z. B. Jensen & Solomon [1972], Solomon & Stephens [1977] und Walker [1979]). Der Schwerpunkt dieser Arbeiten liegt jedoch entweder auf dem zentralen Fall, d. h. alle c_i sind 0, oder die Koeffizienten sind zwar verschieden, jedoch sind die Größenordnungen nicht so extrem unterschiedlich wie in der gegebenen Situation.

Der einzige dem Verfasser bekannte erfolgversprechende Ansatz stammt von Konishi et al. (1988), der auf der Arbeit von Jensen & Solomon (1972) aufbaut. Er benutzt Edgeworth-Expansionen, angewendet nach einer Transformation der quadratischen Form. Mit dieser Methode kann gemäß der Aussage der Autoren eine Genauigkeit von 0.0001 erreicht werden, wenn die

Schiefe von Q kleiner als eins ist. Einige Berechnungen ergaben, daß auch für größere Werte der Schiefe die Übereinstimmung mit der Verteilungsfunktion, die aus der χ^2-Reihenentwicklung näherungsweise gewonnen wurde, für eine Anwendung ausreicht. Die Formeln von Konishi et al. (1988) sind zwar umfangreich, erlauben aber eine direkte und somit schnelle Berechnung einer Approximation der Verteilungsfunktion von

$$\sum_{\lambda_i>0} \lambda_i \chi_1^2(\frac{n(Sw(\pi))_i^2}{\lambda_i}).$$

Die so erhaltene Funktion muß anschließend nur noch um den konstanten Term aus (3.14) verschoben werden. Der einzige Nachteil liegt in der Tatsache begründet, daß es sich bei dieser Funktion um keine „echte“ Verteilungsfunktion handelt, da negative Werte auftreten können. Für Powerapproximationen ist dieser Schwachpunkt i. allg. ohne Relevanz, da dieser Fall nur dann auftritt, wenn die wahre Macht sehr dicht bei 1 liegt.

3.4.4.3 Anpassung einer Gammaverteilung

Analog zum Fall der einfachen Hypothese (siehe Abschnitt 2.3.6.3) soll auch eine Gammaverteilung angepaßt werden. Damit der Träger der angepaßten Verteilungsfunktion mit dem Träger der Verteilungsfunktion der Linearkombination

$$\sum_{\lambda_i>0} \lambda_i \chi_1^2(\frac{n(Sw(\pi))_i^2}{\lambda_i}) + \sum_{\lambda_i=0} n(Sw(\pi))_i^2 \tag{3.16}$$

übereinstimmt, wird ein dritter Parameter γ mit

$$\gamma = \sum_{\lambda_i=0} n(Sw(\pi))_i^2$$

verwendet. Sei X eine Zufallsvariable mit einer Gamma(α, β)-Verteilung. Der Term

$$\sum_{\lambda_i>0} \lambda_i \chi_1^2(\frac{n(Sw(\pi))_i^2}{\lambda_i})$$

besitzt die Erwartung

$$\begin{aligned} \mu_1' &= \sum_{\lambda_i>0} \lambda_i (1 + \frac{n(Sw(\pi))_i^2}{\lambda_i}) \\ &= \sum_{\lambda_i>0} (\lambda_i + n(Sw(\pi))_i^2) \\ &= tr(V) + n\Delta(\pi, h(\theta_0)) - \gamma \end{aligned}$$

und Varianz

$$\begin{aligned}
\mu_2 &= \sum_{\lambda_i>0} \lambda_i^2 (2 + 4(\frac{n(Sw(\pi))_i^2}{\lambda_i})) \\
&= \sum_{\lambda_i>0} (2\lambda_i^2 + 4n\lambda_i (Sw(\pi))_i^2) \\
&= 2tr(V^2) + 4nw(\pi)^T V w(\pi).
\end{aligned}$$

Setzt man

$$\begin{aligned}
\alpha &= \frac{\mu_1'^2}{\mu_2} \text{ und} \\
\beta &= \frac{\mu_2}{\mu_1'},
\end{aligned}$$

so stimmen Erwartung und Varianz der Zufallsvariablen $X_{GLK} = X + \gamma$, sowie der Träger der Verteilungsfunktion mit den entsprechenden Größen von (3.16) überein.

3.4.5 Anpassung von Verteilungen über die Momente

Im Fall der einfachen Hypothese lieferte die Gammaverteilung mit exakter Erwartung und Varianz erstaunlich gute Ergebnisse, und die über die Momente angepaßte Normalverteilung war der asymptotischen Normalverteilung aus (2.5) deutlich überlegen. Für festes π und $h(\theta_0)$ ließen sich Erwartung und Varianz (siehe (2.17) und (2.18)) als

$$\begin{aligned}
E(\chi^2) &= e_1 n + e_2 \text{ bzw. als} \\
Var(\chi^2) &= v_1 n + v_2 + v_3/n
\end{aligned}$$

mit den entprechend gewählten Konstanten e_1, e_2, v_1, v_2 und v_3 schreiben. Wie einige hier nicht näher beschriebene Simulationen ergaben, ist der Term v_3/n entscheidend für die sehr gute Anpassung gerade bei kleinen Stichproben. Paßt man eine Gammaverteilung mit $Var = v_1 n + v_2$ an, so ist die Approximation zwar in den meisten Fällen noch besser als die durch die nichtzentrale χ^2-Verteilung, doch die Güte der Anpassung läßt besonders für kleines n nach, und für eine gewisse Anzahl von Konstellationen nimmt die X_{LKC}-Verteilung die Position der besten Approximation ein. Verzichtet man auch noch auf den Ausdruck v_2, so wird die Approximation schlecht.

Nun kann im Falle der zusammengesetzten Hypothese normalerweise weder Erwartung noch Varianz der χ^2-Prüfgröße exakt bestimmt werden. Unter der Annahme, daß sich die Ergebnisse der einfachen Hypothese übertragen lassen, sollte die Varianz zumindest bis auf einen $o(1)$-, möglicherweise sogar bis auf einen $o(n^{-1})$-Term bestimmt werden. Aber selbst eine Darstellung der exakten Varianz als

$$Var(\chi^2) = v_1^* n + v_2^* + o(1)$$

führt zu recht komplizierten Ausdrücken und im Gegensatz zu den anderen bisher genannten Approximationen zu zusätzlichen Annahmen. Diese Schwierigkeiten werden nun kurz angedeutet.

Eine Taylorentwicklung von $n\Delta(p, h(\hat{\theta}(p)))$ um π ergibt:

$$\begin{aligned} n\Delta(p, h(\hat{\theta}(p))) &= n\Delta(\pi, h(\theta_0)) + n(p-\pi)^T \Delta'(p, h(\hat{\theta}(p)))\mid_{p=\pi} \\ &+ \frac{n}{2}(p-\pi)^T \Delta''(p, h(\hat{\theta}(p)))\mid_{p=\pi} (p-\pi) + R_n, \end{aligned}$$

wobei mit Δ' der Vektor der ersten Ableitungen nach $p_1, \ldots, p_k$ und mit Δ'' die Matrix der zweiten Ableitungen gemeint ist.

Es müßten nun zusätzliche Annahmen getroffen werden, um beispielsweise sicherstellen zu können, daß, nach Einsetzen der relativen Häufigkeiten $z^{(n)}$ für p, $E(R_n) = o(1)$ gilt. Die näherungsweise Berechnung der Varianz gestaltet sich zusätzlich als schwierig. Die einfache Anwendbarkeit der Anpassung einer Verteilung über die Momente geht deshalb bei der zusammengesetzten Hypothese verloren.

Als Konsequenz dieser auftretenden Schwierigkeiten soll im Rahmen dieser Arbeit auf die Anpassung einer Gamma- bzw. einer Normalverteilung über die Momente für den Fall der zusammengesetzten Hypothese verzichtet werden.

3.5 Simulationen zum Vergleich der Approximationen der Macht, zusammengesetzte Hypothese

3.5.1 Einleitung

Im Vergleich zur einfachen Hypothese existiert für die zusammengesetzte Hypothese außer der nichtzentralen χ^2-Verteilung keine weitere in der Literatur

verwendete Verteilung zur näherungsweisen Berechnung der exakten Macht. Dem Verfasser sind auch keine Untersuchungen bekannt, die die Eignung der nichtzentralen χ^2-Verteilung für diesen Zweck überprüfen. Die durchgeführten Berechnungen und Simulationen dienen deshalb einerseits dazu, die Güte der Anpassung durch die nichtzentrale χ^2-Verteilung zu beurteilen, gleichzeitig soll natürlich untersucht werden, ob durch die anderen hergeleiteten Approximationen die Macht möglicherweise genauer bestimmt werden kann.

Die Idee, das Verhalten der Verteilungen unter „local alternatives“ theoretisch zu untersuchen, läßt sich zwar prinzipiell durchführen, doch die auftretenden Matrix-Produkte werden so komplex und unüberschaubar, daß eine Untersuchung nur für konkrete Situationen, nicht mehr aber ohne weiteres allgemein durchgeführt werden kann.

Da die Wahl der zusammengesetzten Hypothese entscheidenden Einfluß auf die erzielten Ergebnisse hat, werden verschiedene Konstellationen betrachtet. Dazu werden Modellfamilien ausgesucht, die sowohl linear als auch nichtlinear in den Parametern sind und sich in der Anzahl der Parameter und in den Werten von k unterscheiden können. Außerdem werden verschiedene Minimum-Diskrepanz-Schätzer verwendet, um auch die Art des Schätzverfahrens in die Untersuchung einfließen zu lassen.

Solange es möglich ist, werden die Werte der wahren Macht und die ihrer Approximationen exakt berechnet, sind die Berechnungen zu aufwendig, werden Monte-Carlo-Methoden verwendet.

Für eine einfachere Beschreibung der Simulationen erscheinen die folgenden Definitionen als zweckmäßig:

X_{AN} : Zufallsvariable mit der Normalverteilung aus (3.10), basierend auf dem asymptotischen Ergebnis aus Abschnitt 3.4.2.1.

X_{NC1} : Zufallsvariable mit der nichtzentralen χ^2-Verteilung aus Abschnitt 3.4.3, $h(\theta_0)$ wird über Δ_g bestimmt.

X_{NC2} : Zufallsvariable mit der nichtzentralen χ^2-Verteilung aus Abschnitt 3.4.3, $h(\theta_0)$ wird über Δ bestimmt.

X_{LKC} : Zufallsvariable mit der Verteilung der Linearkombination von χ^2-verteilten Zufallsvariablen aus (3.14) aus Abschnitt 3.4.4.1.

X_{GLK} : Zufallsvariable mit der Gammaverteilung aus Abschnitt 3.4.4.3, Erwartung, Varianz und Träger stimmen mit den entsprechenden Termen von X_{LKC} überein.

χ^2 : χ^2-Prüfgröße

$F_{KON}(t)$: Geschätzte Verteilungsfunktion von X_{LKC}, ermittelt unter Verwendung des Ansatzes von Konishi et al. (1988).

Für die Berechnung der Approximationen der exakten Macht wurde das Programm „Mathematica“ (Wolfram Research, Inc., 1991) eingesetzt. Im Anhang ist der Quellcode eines vom Verfasser geschriebenen Programms gegeben, und es wird eine Mustersitzung durchgeführt, um die einfache Anwendung zu demonstrieren.

3.5.2 Beispiel 1: Unabhängigkeit bei zweidimensionalen Kontingenztafeln, 2×2-Tafel

Betrachtet wird eine 2×2-Kontingenztafel; die Hypothese sei, daß Unabhängigkeit vorliegt (siehe z. B. Bishop et al. [1975]). Verwendet man die Parameter $\theta_1 \in [0;1]$ und $\theta_2 \in [0;1]$ und faßt die Zeilen der Tafel zu einem Vektor zusammen, so kann $h(\theta)$ als

$$\begin{aligned} h(\theta) &= (h_1(\theta), h_2(\theta), h_3(\theta), h_4(\theta))^T \\ &= (\theta_1\theta_2, \theta_1(1-\theta_2), (1-\theta_1)\theta_2, (1-\theta_1)(1-\theta_2))^T \end{aligned}$$

geschrieben werden. Das Grundmodell ist durch

$$\pi = (\pi_1, \pi_2, \pi_3, \pi_4)^T$$

gegeben. Es werden Maximum-Likelihood-Schätzer verwendet, d. h. die benutzte Diskrepanz Δ_g ist

$$\Delta_g = 2\sum_{i=1}^{4} \pi_i \log \frac{\pi_i}{h_i(\theta)},$$

und θ_0, C, Ω und G werden allein unter Verwendung dieser Diskrepanz abgeleitet.

Für $n = 30$ und $\alpha = 0.10$ wurden 100 Konstellationen, ähnlich wie im Fall der einfachen Hypothese, erzeugt. Die einzelnen Schritte sind wie folgt gegeben:

Schritt 1: Berechnung des kritischen Wertes $c(\alpha)$ aus der asymptotischen Verteilung der χ^2-Prüfgröße:

$$P(\chi_1^2 > c(\alpha)) = \alpha$$

Schritt 2: Erzeugung einer Zufallszahl w_α im Intervall $]\alpha, 1[$.

Schritt 3: Berechnung des Wertes δ mit

$$P(\chi_1^2(\delta) > c(\alpha)) = w_\alpha$$

Hier wird angenommen, daß die nichtzentrale χ^2-Verteilung als Approximation geeignet ist, um den möglichen Bereich der Werte der exakten Macht gleichmäßig abzudecken.

Schritt 4: Erzeugung von 4 Zufallszahlen $v_1, \ldots, v_4$ aus dem Intervall $]0, 1[$. Der vorläufige Vektor π^v ist definiert durch

$$\pi_i^v = v_i/(v_1 + \ldots + v_4), \quad i = 1, \ldots, 4.$$

Schritt 5: Berechnung von $n\Delta^v = n\Delta(\pi^v, h(\theta_0))$. Bemerkung: θ_0 minimiert $\Delta_g(\pi, h(\theta))$.

Schritt 6: Bestimmung eines λ, so daß sich für

$$\pi = h(\theta_0) + \lambda(\pi^v - h(\theta_0))$$

eine Diskrepanz von $n\Delta(\pi, h(\theta_0)) = \delta$ ergibt:

$$\lambda = \sqrt{\frac{\delta}{n\Delta^v}}$$

Bemerkung: Man kann zeigen, daß sich für alle π^* auf der Verbindungsgeraden von π^v nach $h(\theta_0)$ das gleiche θ_0 ergibt.

Schritt 7: Kontrolle, ob alle Komponenten des Vektors π positiv sind. Falls dies nicht der Fall ist, wird bei Schritt 2 neu begonnen.

Der Unterschied zur einfachen Hypothese liegt darin, daß nun nur der Vektor π zufällig erzeugt wird, $h(\theta_0)$ wird dann gemäß der verwendeten Diskrepanz Δ_g bestimmt.

Vielleicht mag das Beispiel etwas klein aussehen, aber es darf nicht vergessen werden, daß schon für $n = 30$ und $k = 4$ gemäß der Formel (2.36) 5456 verschiedene Vektoren der relativen Häufigkeiten auftreten; für eine 3×4-Tafel wären es schon mehr als drei Milliarden Möglichkeiten, und eine Berechnung der exakten Verteilungsfunktion ist dann selbst auf einem schnellen Rechner unmöglich.

Die Güte der Approximationen wurde jeweils durch die absoluten und die relativen Abweichungen gemessen:

Schritt 1: Berechnung des kritischen Wertes $c(\alpha)$ aus der asymptotischen Verteilung der χ^2-Prüfgröße:

$$P(\chi_1^2 > c(\alpha)) = \alpha$$

Schritt 2: Berechnung der exakten Macht und ihrer Approximationen:

$$\begin{aligned} Pow_{\chi^2}(\alpha) &= P(\chi^2 > c(\alpha)) \\ Pow_{App}(\alpha) &= P(X_{App} > c(\alpha)), \\ & \quad App \in \{GLK, AN, NC1, NC2\} \\ Pow_{KON}(\alpha) &= 1 - F_{KON}(c(\alpha)) \end{aligned}$$

Schritt 3a: Berechnung der Abweichung

$$\mid Pow_{\chi^2}(\alpha) - Pow_{App}(\alpha) \mid$$

für jede Approximation.

Schritt 3b: Berechnung der relativen Abweichung

$$\frac{\mid Pow_{\chi^2}(\alpha) - Pow_{App}(\alpha) \mid}{(Pow_{\chi^2}(\alpha)(1 - Pow_{\chi^2}(\alpha)))^{\frac{1}{2}}}$$

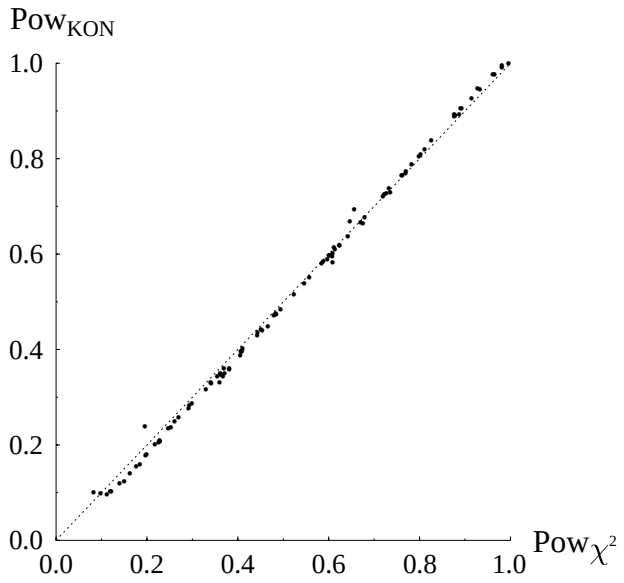
für jede Approximation.

Alle erzeugten Hypothesen besaßen erwartete Häufigkeiten, die größer als 0.5 waren. Für einige Fälle ließ sich die exakte Verteilungsfunktion von X_{LKC} zwar bis auf die gewünschte Genauigkeit berechnen, aber da es auch Situationen gab, in denen selbst 5000 Summanden der χ^2-Reihe nicht ausreichten, wurde anstelle der X_{LKC}-Verteilungsfunktion nur die $F_{KON}(t)$-Approximation nach Konishi et al. betrachtet.

Die Plots der exakten Macht gegen ihre Approximationen sind in den Abbildungen 3.1, 3.2 und 3.3 für die erwähnten Approximationen dargestellt. Die Tabelle 3.1 gibt die durchschnittlichen absoluten bzw. relativen Abweichungen der betrachteten Approximationen für die verschiedenen Bereiche der exakten Macht (niedrig, mittel, hoch). Aus den Plots und der Tabelle wird ersichtlich, daß die Funktion $F_{KON}(t)$ am besten geeignet ist, Schätzer der exakten Macht zu berechnen. Für den Bereich niedriger Macht werden zwar von den restlichen Approximationen, abgesehen von der Normalverteilung, im Durchschnitt geringfügig niedrigere Abweichungen erzielt, ansonsten

KON

2×2-Tafel
n= 30
α= 0.10



GLK

2×2-Tafel
n= 30
α= 0.10

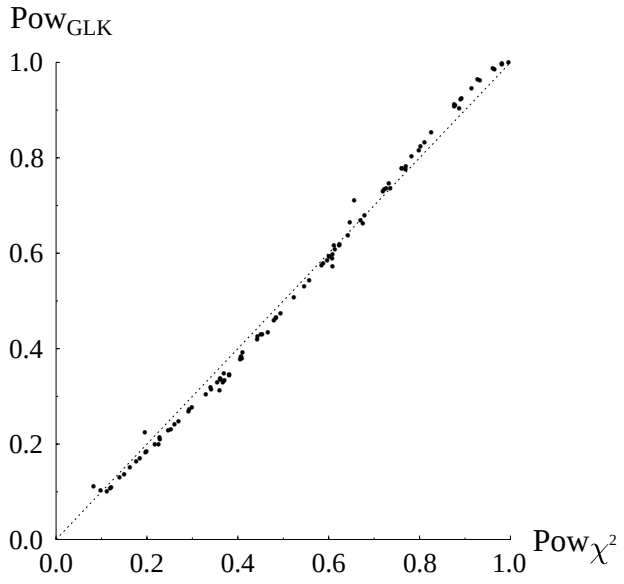
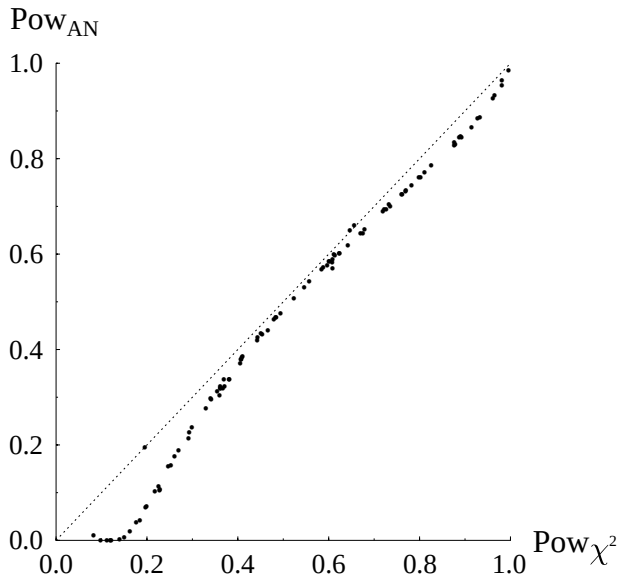


Abb. 3.1: Exakte Macht ↔ App. Macht, Bsp. 1, (I)

AN

2×2-Tafel
n= 30
α= 0.10



NC1

2×2-Tafel
n= 30
α= 0.10

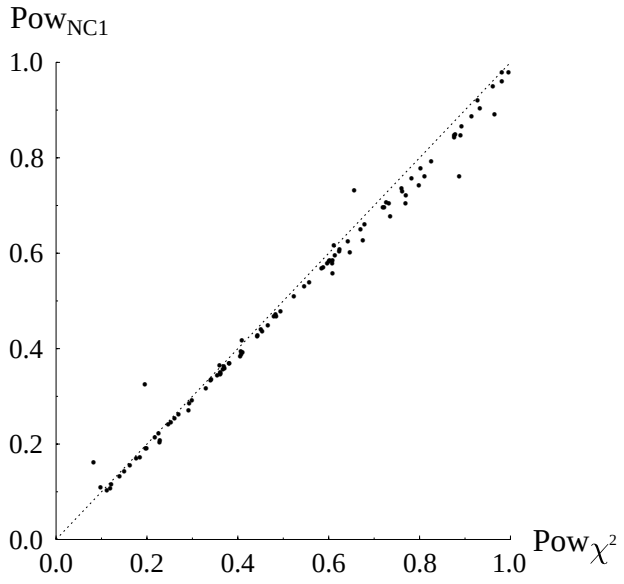


Abb. 3.2: Exakte Macht ↔ App. Macht, Bsp. 1, (II)

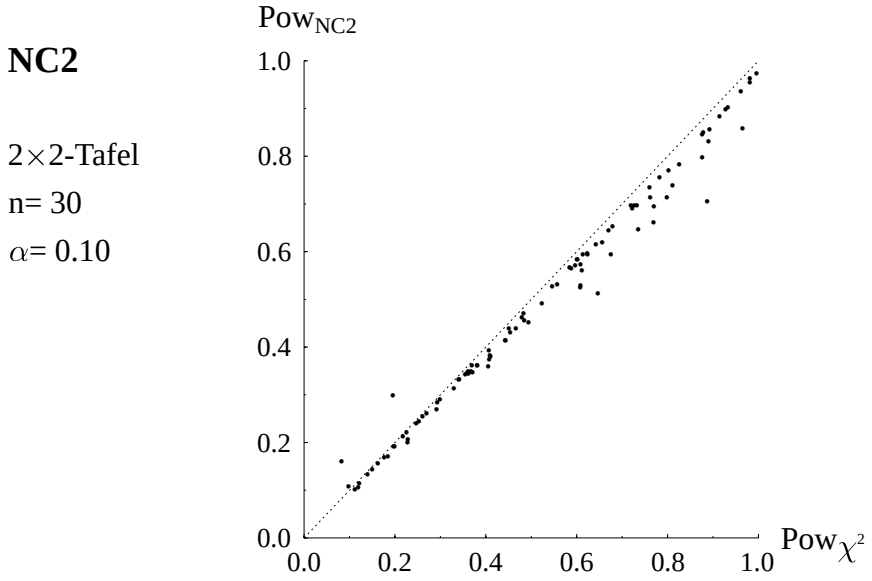


Abb. 3.3: Exakte Macht ↔ App. Macht, Bsp. 1, (III)

		KON	GLK	AN	NC1	NC2
n. M.	$\emptyset$ a. A.	0.0182	0.0173	0.1051	0.0175	**0.0167**
24	$\emptyset$ r. A.	0.1263	**0.1159**	0.7447	0.1365	0.1318
m. M.	$\emptyset$ a. A.	**0.0119**	0.0208	0.0272	0.0183	0.0293
48	$\emptyset$ r. A.	**0.0505**	0.0880	0.1157	0.0776	0.1238
h. M.	$\emptyset$ a. A.	**0.0084**	0.0201	0.0371	0.0356	0.0511
28	$\emptyset$ r. A.	**0.1617**	0.2975	0.5212	0.5233	0.7513

Tab. 3.1: Auswertung: $\alpha = 0.10$, $n = 30$, 2×2-Tafel.

erhält man aber unter Verwendung der Funktion $F_{KON}(t)$ für die gegebene Stichprobengröße von $n = 30$ die mit Abstand besten Annäherungen an die wahren Werte der Macht.

Verwendet man im Falle der nichtzentralen χ^2-Verteilung für die Berechnung von $h(\theta_0)$ (siehe Abschnitt 3.4.3) nicht die Diskrepanz $\Delta_g(\rightarrow NC1)$, sondern die χ^2-Diskrepanz ($\rightarrow NC2$), so ergibt sich global gesehen eine schlechtere Approximation. Für die vorliegende Form der Hypothese ist es deshalb günstiger, die Verteilung von X_{NC1} zu verwenden. Für den Bereich hoher Macht sind beide Approximationen nicht zu empfehlen.

Die X_{AN}-Approximation versagt im Bereich niedriger Macht. Die wahre Macht wird stark unterschätzt.

3.5.3 Beispiel 2: Unabhängigkeit bei zweidimensionalen Kontingenztafeln, 4×4-Tafel

Es wird nun eine 4×4-Kontingenztafel betrachtet, die Hypothese sei wieder die der Unabhängigkeit. Verwendet man $\theta = (\theta_1, \ldots, \theta_6)^T$ mit $\theta_i \geq 0$, $i = 1, \ldots, 6$ und $\theta_1 + \theta_2 + \theta_3 \leq 1$ bzw. $\theta_4 + \theta_5 + \theta_6 \leq 1$, so lassen sich die Komponenten des Vektors $h(\theta)$ aus der folgenden Tabelle ablesen, wobei jeweils das Produkt der beiden Größen an den Rändern genommen wird:

$h_1(\theta)$	$h_2(\theta)$	$h_3(\theta)$	$h_4(\theta)$	θ_1
$h_5(\theta)$	$h_6(\theta)$	$h_7(\theta)$	$h_8(\theta)$	θ_2
$h_9(\theta)$	$h_{10}(\theta)$	$h_{11}(\theta)$	$h_{12}(\theta)$	θ_3
$h_{13}(\theta)$	$h_{14}(\theta)$	$h_{15}(\theta)$	$h_{16}(\theta)$	$1 - \theta_1 - \theta_2 - \theta_3$
θ_4	θ_5	θ_6	$1 - \theta_4 - \theta_5 - \theta_6$	

Die betrachtete Methode zur Schätzung der Parameter ist wieder die Maximum-Likelihood-Methode. Die Komponenten von θ_0, welches

$$\Delta_g(\pi, h(\theta)) = 2\sum_{i=1}^{16} \pi_i \log \frac{\pi_i}{h_i(\theta)}$$

minimiert, werden bekanntermaßen durch Aufsummieren von Zeilen- bzw. Spaltenwahrscheinlichkeiten berechnet:

π_1	π_2	π_3	π_4	$\sum \Rightarrow \theta_{01}$
π_5	π_6	π_7	π_8	$\sum \Rightarrow \theta_{02}$
π_9	π_{10}	π_{11}	π_{12}	$\sum \Rightarrow \theta_{03}$
π_{13}	π_{14}	π_{15}	π_{16}	
$\sum \Rightarrow \theta_{04}$	$\sum \Rightarrow \theta_{05}$	$\sum \Rightarrow \theta_{06}$		

Analog bildet man den Maximum-Likelihood-Schätzer $\hat{\theta}(z)$ aus den relativen Häufigkeiten $z_1, \ldots, z_{16}$.

Für $n = 100$ und $\alpha = 0.10$ wurden 100 Grundmodelle erzeugt. Das Vorgehen ist analog zur 2×2-Tafel, natürlich mit den entsprechend geänderten Werten von $k = 16$ und neun Freiheitsgraden für die χ^2-Verteilung.

Die exakte Macht wurde mit Monte-Carlo-Methoden geschätzt. Für jedes Grundmodell wurden 50000 Prüfgrößen erzeugt und der relative Anteil der Werte, die in den Ablehnungsbereich fielen, bestimmt. Dieser Aufwand ließ sich nur betreiben, weil $\hat{\theta}(z)$ explizit durch die relativen Häufigkeiten ausgedrückt werden kann.

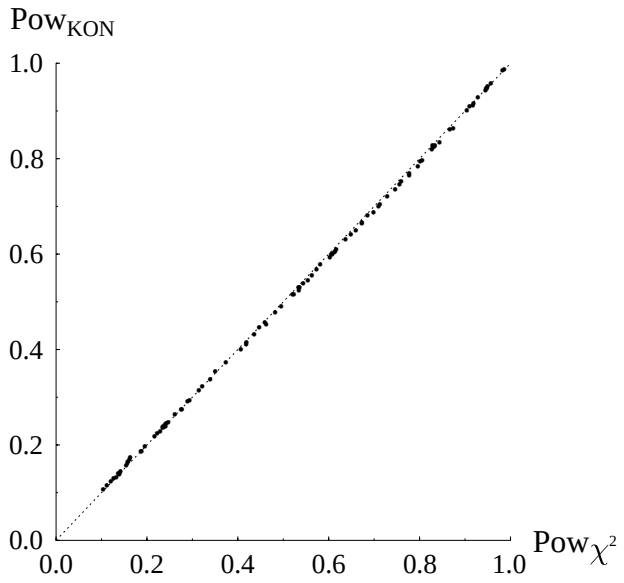
Die X_{NC2}-Approximation wird an dieser Stelle nicht betrachtet, da kein zuverlässig arbeitender Algorithmus zur Minimierung von $\Delta(\pi, h(\theta))$ zur Verfügung stand.

Für die übrigen Approximationen ergaben sich die Plots aus den Abbildungen 3.4 und 3.5. Tabelle 3.2 gibt eine Zusammenfassung der durchschnittlichen absoluten und relativen Abweichungen.

Wiederum liefert die Verwendung von $F_{KON}(t)$ die besten Ergebnisse, dicht gefolgt von der X_{GLK}-Verteilung, wobei, rein optisch gesehen, die Unterschiede in den Plots kaum zu erkennen sind. Die Approximation unter Benutzung der nichtzentralen χ^2-Verteilung ist etwas, aber erkennbar schlechter. Gänzlich ungeeignet ist in diesem Falle die aus der Normalverteilung gewonnenene Annäherung. Der Verlauf der Punktwolke deutet darauf hin,

KON

4×4-Tafel
n= 100
α= 0.10



GLK

4×4-Tafel
n= 100
α= 0.10

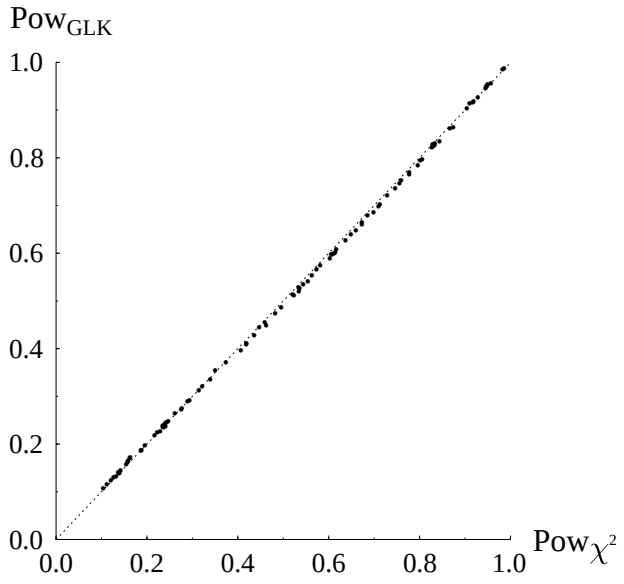
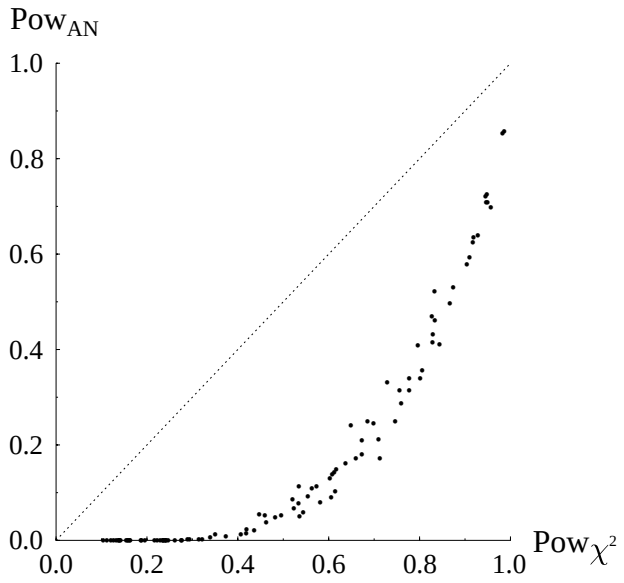


Abb. 3.4: Exakte Macht ↔ App. Macht, Bsp. 2, (I)

AN

4×4-Tafel
n= 100
α= 0.10



NC1

4×4-Tafel
n= 100
α= 0.10

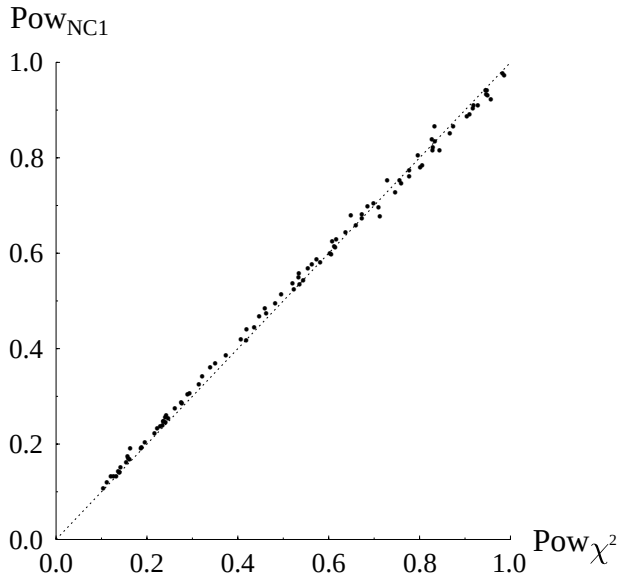


Abb. 3.5: Exakte Macht ↔ App. Macht, Bsp. 2, (II)

		KON	GLK	AN	NC1
n. M.	$\emptyset$ a .A.	**0.0024**	0.0026	0.1956	0.0086
32	$\emptyset$ r. A.	**0.0175**	0.0182	1.2487	0.0560
m. M.	$\emptyset$ a. A.	**0.0062**	0.0087	0.4361	0.0114
37	$\emptyset$ r. A.	**0.0260**	0.0366	1.8378	0.0481
h. M.	$\emptyset$ a. A.	**0.0057**	0.0058	0.3551	0.0155
31	$\emptyset$ r. A.	**0.0473**	0.0500	3.6455	0.1970

Tab. 3.2: Auswertung: $\alpha = 0.10$, $n = 100$, $k = 16$.

daß erst bei einer sehr großen Diskrepanz des wahren Modells zur approximierenden Modellfamilie die Macht richtig angenähert wird. Dieser Fall ist jedoch uninteressant, da dann a priori eine Macht von nahezu 1 anzunehmen ist.

3.5.4 Beispiel 3: Eine einparametrige Modellfamilie

Es wird nun die Hypothese betrachtet, daß die ersten $k-1$ Zellen die gleiche Wahrscheinlichkeit besitzen, d. h.

$$\begin{aligned} h(\theta) &= (h_1(\theta), \ldots, h_{k-1}(\theta), h_k(\theta))^T \\ &= (\theta, \ldots, \theta, 1-(k-1)\theta)^T. \end{aligned}$$

Dieses Beispiel hat den Vorteil, daß die Minimum-Diskrepanz-Schätzer für Diskrepanzen aus $\mathcal{D}^{\mathcal{CR}}$ (siehe Abschnitt 3.2.2) explizit angegeben werden können. Für $\lambda \neq -1$ berechnet sich $\hat{\theta}(z)$ zu

$$\hat{\theta}(z) = \frac{(\sum_{i=1}^{k-1} z_i^{\lambda+1})^{\frac{1}{\lambda+1}}}{(k-1)^{\frac{1}{\lambda+1}} z_k + (k-1) * (\sum_{i=1}^{k-1} z_i^{\lambda+1})^{\frac{1}{\lambda+1}}}.$$

Da für $\lambda < -1$ Definitionsprobleme auftreten, falls mindestens ein z_i den Wert null annimmt, seien nur Minimum-Diskrepanz-Schätzer zu Diskrepanzen mit $\lambda > -1$ zugelassen. In diesem Fall kann $\Delta(z, \hat{\theta}(z))$ für $z = (0, \ldots, 0, 1)^T$ stetig mit $\Delta = 0$ ergänzt werden, so daß für θ der Definitionsbereich $\mathbb{D} = [0; 1/(k-1)]$ gewählt werden kann.

Es werden nun die vier Fälle $\lambda = 0$ ($\Rightarrow$ Maximum-Likelihood-Schätzer), $\lambda = 1$ ($\Rightarrow$ Minimum-χ^2-Schätzer), $\lambda = -0.5$ ($\Rightarrow$ Minimum-Diskrepanz-Schätzer der Freeman-Tukey-Diskrepanz) und $\lambda = 10$ betrachtet.

Für $\lambda = 0$ vereinfacht sich der Schätzer zu

$$\hat{\theta}(z) = \frac{1}{k-1} \sum_{i=1}^{k-1} z_i.$$

Für $n = 50$, $k = 6$ und $\alpha = 0.10$ wurden 200 Alternativen erzeugt, die Vorgehensweise ist analog zu den bisherigen Beispielen. Es wurde wieder die zu $\lambda = 0$ gehörige Diskrepanz

$$\Delta_g^0(\pi, h(\theta)) = 2 \sum_{i=1}^{k} \pi_i \log \frac{\pi_i}{h_i(\theta)}$$

verwendet, um die Alternativen zu bestimmen. Diese erzeugten Alternativen wurden der Einfachheit halber auch für die Fälle $\lambda = 1$, $\lambda = -0.5$ und $\lambda = 10$ verwendet.

Die exakte Macht wurde mit Monte-Carlo-Methoden geschätzt. Für jedes der 200 Grundmodelle wurden 20000 Prüfgrößen $n\Delta(z, h(\hat{\theta}(z)))$ mit eingesetztem Maximum-Likelihood-Schätzer $\hat{\theta}(z)$ erzeugt und der relative Anteil der Werte, die in den Ablehnungsbereich fielen, bestimmt.

Dieses Vorgehen wurde anschließend für die anderen drei Diskrepanzen wiederholt, wobei der eingesetzte Schätzer durch die zugehörige Diskrepanz bestimmt wird.

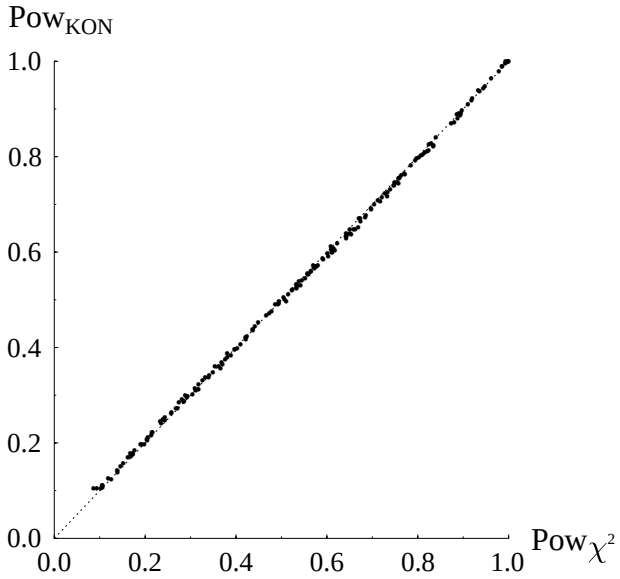
Für die Approximation mit Hilfe der nichtzentralen χ^2-Verteilung wurde $h(\theta_0)$ jeweils durch Minimierung der verwendeten Diskrepanz Δ_g bestimmt (siehe Abschnitt 3.4.3).

Die Abbildungen 3.6 bis 3.9 zeigen die Plots der approximierenden gegen die (geschätzte) exakte Macht für den Fall des Minimum-χ^2-Schätzers bzw. des Minimum-Diskrepanz-Schätzers zu $\lambda = 10$. Die Tabellen geben die durchschnittlichen absoluten bzw. relativen Abweichungen für die 196 Fälle an, in denen das Minimum der erwarteten Häufigkeiten 0.5 übertraf.

In den Tabellen 3.4 und 3.5 konnte für den Bereich hoher Macht die durchschnittliche relative Abweichung nicht unmittelbar berechnet werden, da die exakte Macht in einem Fall durch 1 geschätzt wurde. Der angegebene Wert des durchschnittlichen relativen Fehlers wurde dann aus den übrigen Fällen ermittelt.

KON

λ= 1
n= 50
k= 6
α= 0.10



GLK

λ= 1
n= 50
k= 6
α= 0.10

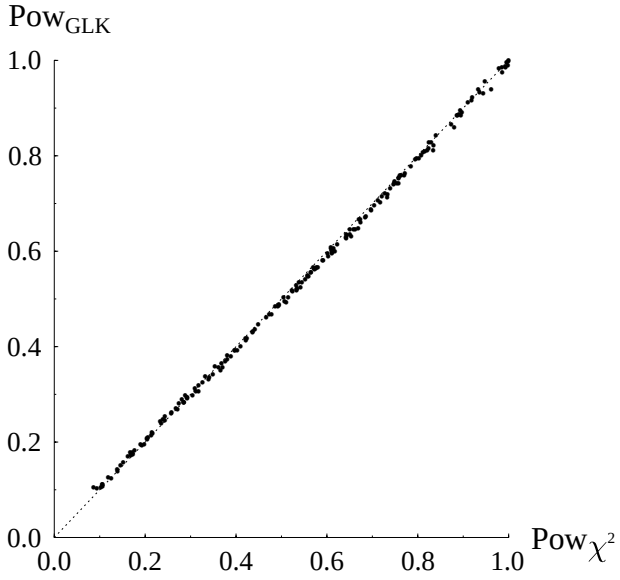
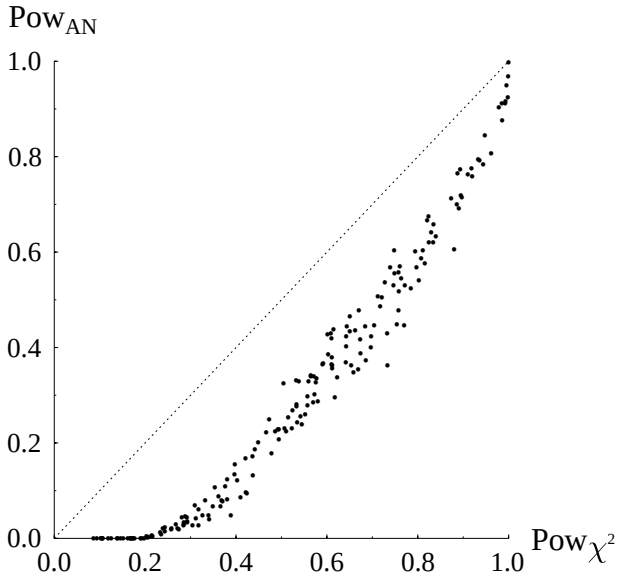


Abb. 3.6: Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 3, (I)

AN

λ= 1
n= 50
k= 6
α= 0.10



NC1

λ= 1
n= 50
k= 6
α= 0.10

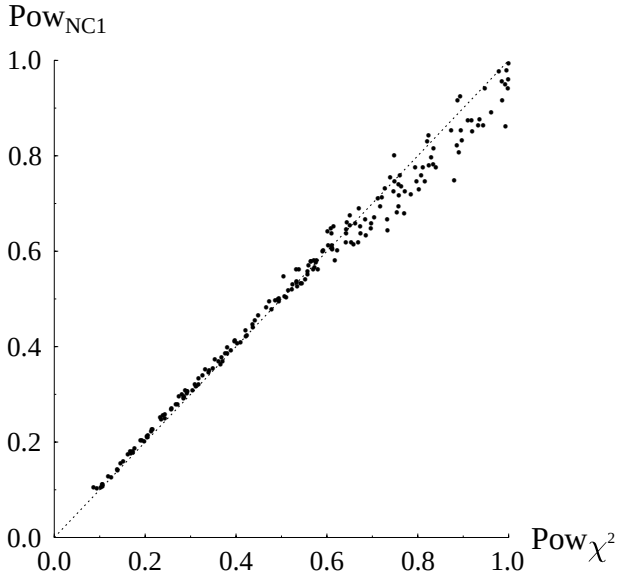
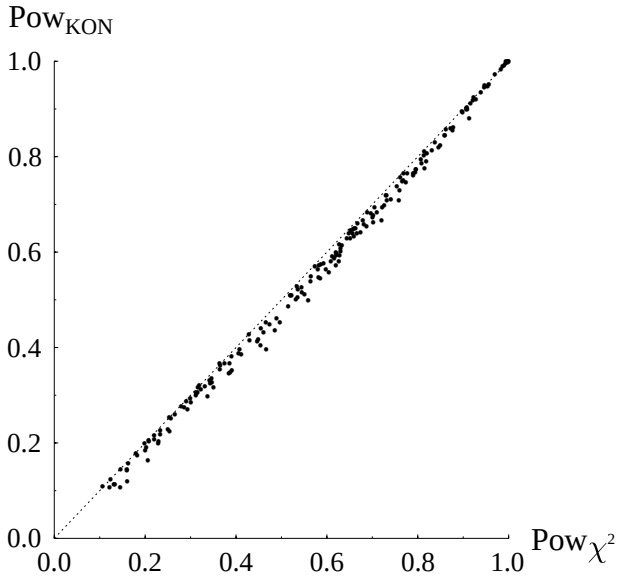


Abb. 3.7: Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 3, (II)

KON

λ= 10
n= 50
k= 6
α= 0.10



GLK

λ= 10
n= 50
k= 6
α= 0.10

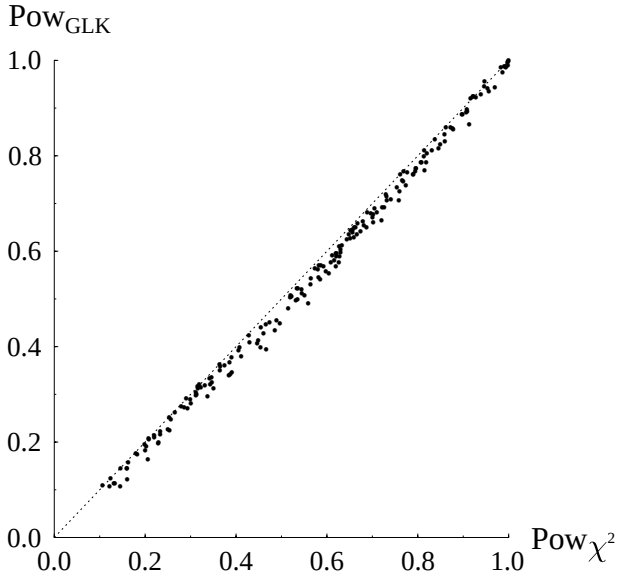
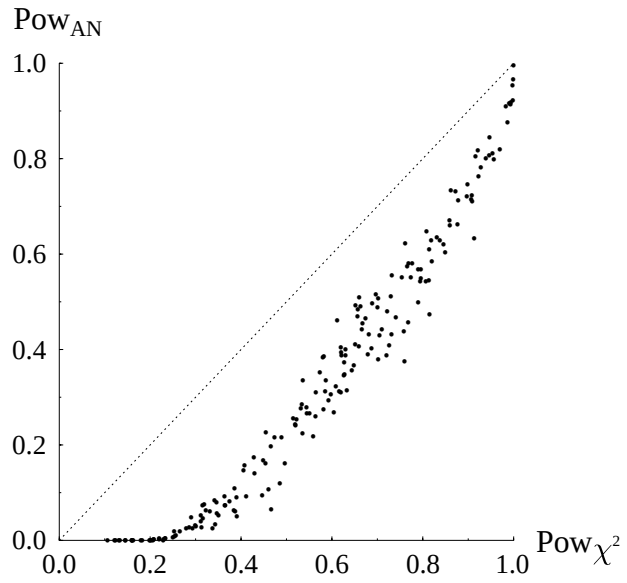


Abb. 3.8: Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 3, (III)

AN

λ= 10
n= 50
k= 6
α= 0.10



NC1

λ= 10
n= 50
k= 6
α= 0.10

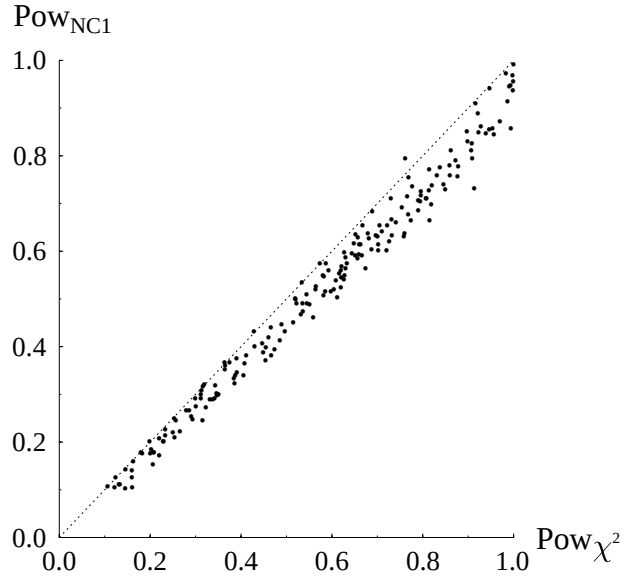


Abb. 3.9: Exakte Macht $\leftrightarrow$ App. Macht, Bsp. 3, (IV)

		KON	GLK	AN	NC1
n. M.	$\emptyset$ a. A.	0.0043	**0.0034**	0.1930	0.0074
46	$\emptyset$ r. A.	0.0288	**0.0240**	1.2079	0.04630
m. M.	$\emptyset$ a. A.	**0.0052**	0.0092	0.2611	0.0135
90	$\emptyset$ r. A.	**0.0223**	0.0390	1.1038	0.0577
h. M.	$\emptyset$ a. A.	**0.0055**	0.0076	0.1856	0.0431
60	$\emptyset$ r. A.	**0.1084**	0.3122	5.1707	3.9915

Tab. 3.3: Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$, $\lambda = 0$.

		KON	GLK	AN	NC1
n. M.	$\emptyset$ a. A.	0.0055	**0.0044**	0.1907	0.0096
46	$\emptyset$ r. A.	0.0390	**0.0331**	1.2045	0.0622
m. M.	$\emptyset$ a. A.	**0.0055**	0.0091	0.2610	0.0138
92	$\emptyset$ r. A.	**0.0234**	0.0386	1.1058	0.0593
h. M.	$\emptyset$ a. A.	**0.0048**	0.0072	0.1823	0.0446
58/57	$\emptyset$ r. A.	**0.0800**	0.2430	3.6382	2.2888

Tab. 3.4: Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$, $\lambda = 1$.

		KON	GLK	AN	NC1
n. M.	$\emptyset$ a. A.	0.0048	0.0057	0.1919	**0.0042**
40	$\emptyset$ r. A.	0.0329	0.0376	1.2166	**0.0297**
m. M.	$\emptyset$ a. A.	**0.0112**	0.0156	0.2673	0.0124
95	$\emptyset$ r. A.	**0.0476**	0.0662	1.1374	0.0531
h. M.	$\emptyset$ a. A.	**0.0082**	0.0102	0.1877	0.0451
61/60	$\emptyset$ r. A.	**0.1056**	0.2603	4.0839	2.3606

Tab. 3.5: Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$, $\lambda = -0.5$.

		KON	GLK	AN	NC1
n. M.	$\emptyset$ a. A.	**0.0136**	0.0139	0.1999	0.0223
35	$\emptyset$ r. A.	**0.0886**	0.0902	1.2321	0.1383
m. M.	$\emptyset$ a. A.	**0.0222**	0.0251	0.2680	0.0473
93	$\emptyset$ r. A.	**0.0936**	0.1057	1.1401	0.2011
h. M.	$\emptyset$ a. A.	**0.0154**	0.0185	0.1944	0.0779
68	$\emptyset$ r. A.	**0.1602**	0.3528	5.0207	5.5971

Tab. 3.6: Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$, $\lambda = 10$.

Wieder wird deutlich, daß die Funktion $F_{KON}(t)$ am besten geeignet ist, die wahre Macht näherungsweise zu berechnen. In fast allen Fällen wurden im Durchschnitt die geringsten Fehler erzielt. Die Approximation mittels der nichtzentralen χ^2-Verteilung fällt im Vergleich vor allem für den Bereich mittlerer und hoher Macht deutlich ab. Die X_{AN}-Approximation ist, dies sollte jetzt nicht mehr überraschen, nicht zu gebrauchen.

Für den extremen Wert $\lambda = 10$ werden sämtliche Approximationen schlechter, trotzdem erzielte die $F_{KON}(t)$-Approximation in allen Bereichen die besten Ergebnisse.

Kapitel 4

Erweiterung der Ergebnisse auf andere Diskrepanzen

4.1 Einleitung

Die theoretischen Ergebnisse und die Simulationen haben gezeigt, daß für die Approximation der wahren Macht deutlich bessere Verteilungen als die nichtzentrale χ^2-Verteilung existieren. Dabei zeichnet sich $F_{KON}(t)$ (bzw. die Verteilungsfunktion von X_{LKC}) durch eine besonders gute Annäherung der exakten Macht aus. Es bleibt zu untersuchen, welchen Einfluß die Approximation der Verteilungsfunktion von X_{LKC} durch $F_{KON}(t)$ hat; zumindest bei einzelnen, willkürlich gewählten Beispielen waren die Unterschiede vernachlässigbar.

Jeder Diskrepanz $\Delta \in \mathcal{D}$ läßt sich eine „Prüfgröße“ zuordnen. In diesem Abschnitt wird untersucht, inwiefern sich die für die χ^2-Diskrepanz erhaltenen Resultate auf andere Diskrepanzen übertragen lassen. Dabei werden der Vollständigkeit halber auch einige asymptotische Resultate genannt, die für die Approximation der wahren Macht *nicht* geeignet sind, aber in Kapitel 6 für die Bootstrap-Tests wieder aufgegriffen werden.

Die betrachteten Diskrepanzen sind Elemente der Menge $\mathcal{D}$, die jeweils verwendete Diskrepanz wird mit Δ, die Diskrepanz zur Ermittlung der Schätzer im Fall der zusammengesetzten Hypothese mit Δ_g bezeichnet.

Die Diskrepanzen aus der Cressie-Read-Familie (vgl. Abschnitt 3.2.2) seien mit Δ^λ, $\lambda \in \mathbb{R}$, bezeichnet.

4.2 Einfache Hypothesen

4.2.1 Verteilung unter der Hypothese

4.2.1.1 Cressie-Read-Diskrepanzen

Es ist im allgemeinen schwierig, für beliebiges Δ die Verteilung der Prüfgröße $n\Delta(z^{(n)}, h(\theta_0))$ unter der Hypothese $\pi = h(\theta_0)$ anzugeben. Beschränkt man sich auf Diskrepanzen $\Delta = \Delta^\lambda \in \mathcal{D}^{\mathcal{CR}}$, so folgt (Read & Cressie, 1988, S. 47):

$$n\Delta^\lambda(z^{(n)}, h(\theta_0)) \xrightarrow{d} \chi^2_{k-1}$$

4.2.1.2 Gauß-Diskrepanz

Verwendet man die Gauß-Diskrepanz

$$\Delta(\pi, h(\theta_0)) = \sum_{i=1}^{k} (\pi_i - h_i(\theta_0))^2,$$

welche zur Menge $\mathcal{D} \setminus \mathcal{D}^*$ gehört, so läßt sich eine asymptotische Verteilung unter der Hypothese ableiten. Da

$$\sqrt{n}(z^{(n)} - \pi) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T)$$

gilt, folgt unter Verwendung der Sätze 2.1 und 2.2 unter der Hypothese

$$n\Delta(z^{(n)}, h(\theta_0)) \xrightarrow{d} \sum_{i=1}^{k-1} \lambda_i \chi^2_1,$$

wobei die λ_i, $i = 1, \ldots, k-1$, die positiven Eigenwerte der Matrix $D_\pi - \pi\pi^T$ bezeichnen sollen. Dieses zunächst theoretisch anmutende Ergebnis wird in seiner Erweiterung für zusammengesetzte Hypothesen in Kapitel 6 für die Konstruktion von Bootstrap-Tests an Bedeutung gewinnen. An dieser Stelle sei es der Vollständigkeit halber erwähnt.

4.2.2 Verteilung unter einer festen Alternative

4.2.2.1 Allgemeiner Fall

Sei $\pi \neq h(\theta_0)$, $\Delta(p, h(\theta_0)) = \sum_{i=1}^{k} g(p_i, h_i(\theta_0))$, und bezeichne G_x die $1 \times k$-Matrix der Ableitungen

$$\left.\frac{\partial \Delta(p, h(\theta_0))}{\partial p_j}\right|_{p=\pi} = \frac{\partial g}{\partial x}(\pi_j, h_j(\theta_0)), \quad j = 1, \ldots, k.$$

Mittels der Delta-Methode folgt nun sofort:

$$\sqrt{n}(\Delta(z, h(\theta_0)) - \Delta(\pi, h(\theta_0))) \xrightarrow{d} N(0, G_x(D_\pi - \pi\pi^T)G_x^T). \tag{4.1}$$

4.2.2.2 Cressie-Read-Diskrepanzen

Entstammt die Diskrepanz Δ der Menge $\mathcal{D}^{\mathcal{CR}}$ der Cressie-Read-Diskrepanzen (vgl. Abschnitt 3.2.2), d. h., ist $g(x, y)$ von der Form

$$g(x, y) = \frac{2}{\lambda(\lambda+1)} x[(\frac{x}{y})^\lambda - 1] \quad , \lambda \in I\!R \setminus \{0; -1\},$$

so folgt

$$\frac{\partial g(x, y)}{\partial x} = \frac{2}{\lambda}(\frac{x}{y})^\lambda - \frac{2}{\lambda(\lambda+1)} \quad , \lambda \in I\!R \setminus \{0; -1\},$$

und somit

$$G_x = \frac{2}{\lambda}\pi^T D_\pi^{\lambda-1} D_h^{-\lambda} - \frac{2}{\lambda(\lambda+1)}(1, \ldots, 1).$$

Es ergibt sich dann nach kurzer Rechnung:

$$\sqrt{n}(\Delta^\lambda(z, h(\theta_0)) - \Delta^\lambda(\pi, h(\theta_0))) \xrightarrow{d} N(0, \sigma_\lambda^2)$$

mit

$$\sigma_\lambda^2 = \frac{4}{\lambda^2}\pi^T D_\pi^{\lambda-1} D_h^{-\lambda}(D_\pi - \pi\pi^T) D_h^{-\lambda} D_\pi^{\lambda-1}\pi \tag{4.2}$$

Wie man direkt zeigen kann, ist dieses Resultat auch für den Spezialfall $\lambda = -1$ richtig. Für $\lambda = 1$ erhält man natürlich das Ergebnis aus (2.3).

Für zwei k-dimensionale Vektoren a und b sei $\log(\frac{a}{b})$ definiert als

$$\log(\frac{a}{b}) := (\log(\frac{a_1}{b_1}), \ldots, \log(\frac{a_k}{b_k}))^T.$$

Es ergibt sich für den Spezialfall $\lambda = 0$:

$$\sqrt{n}(\Delta(z, h(\theta_0)) - \Delta(\pi, h(\theta_0))) \xrightarrow{d} N(0, 4\log(\frac{\pi}{h(\theta_0)})^T(D_\pi - \pi\pi^T)\log(\frac{\pi}{h(\theta_0)})).$$

4.2.2.3 Gauß-Diskrepanz

Für die Gauß-Diskrepanz ist $G_x = 2(\pi - h(\theta_0))^T$. Es ergibt sich somit

$$\sqrt{n}(\Delta(z, h(\theta_0)) - \Delta(\pi, h(\theta_0))) \xrightarrow{d} N(0, 4(\pi - h(\theta_0))^T(D_\pi - \pi\pi^T)(\pi - h(\theta_0))).$$

4.2.3 Verteilung unter „local alternatives“

Bei „local alternatives“ ist es nicht sinnvoll, für jede Diskrepanz die asymptotische Verteilung der zugehörigen Prüfgröße herleiten zu wollen. Beschränkt man sich auf die Cressie-Read-Diskrepanzen, so ergibt sich für eine Folge

$$\pi^{(n)} = h(\theta_0) + \frac{c}{\sqrt{n}} \quad , \sum_{i=1}^{k} c_i = 0,$$

daß die asymptotische Verteilung von $n\Delta^\lambda(z^{(n)}, h(\theta_0))$, unabhängig vom Parameter λ, gegen eine nichtzentrale χ^2-Verteilung strebt (Read & Cressie, 1988):

$$n\Delta^\lambda(z^{(n)}, h(\theta_0)) \xrightarrow{d} \chi^2_{k-1}(\sum_{i=1}^{k} \frac{c_i^2}{h_i(\theta_0)})$$

Für eine beliebige Diskrepanz aus der Cressie-Read-Familie würde man zu gegebenen Werten von n, π und $h(\theta_0)$ die Macht mit dem gleichen Wert schätzen, den man auch schon über die χ^2-Diskrepanz erhalten hat, ein Umstand, der nicht gerade für die Eignung zur Powerapproximation spricht.

Drost et al. (1989) schlagen vor, ersatzweise $n\Delta^\lambda(\pi, h(\theta_0))$ als Nichtzentralitätsparameter zu verwenden.

4.2.4 Der Ansatz von Drost et al.

Drost et al. (1989) betrachten die Cressie-Read-Prüfgrößen

$$n\Delta^\lambda(z^{(n)}, h(\theta_0)) = \frac{2n}{\lambda(\lambda+1)} \sum_{i=1}^{k} z_i^{(n)} [\left(\frac{z_i^{(n)}}{h_i(\theta_0)}\right)^\lambda - 1] \quad , \lambda \in I\!R \setminus \{-1; 0\}.$$

Mittels einer Taylorentwicklung und der Definition

$$Y_n := \sqrt{n} D_\pi^{-\frac{1}{2}} (z^{(n)} - \pi)$$

können Sie dann die Prüfgröße als

$$n\Delta^\lambda(z^{(n)}, h(\theta_0)) = A^\lambda(Y_n) + O_p(n^{-1/2})$$

darstellen. Der Ausdruck $A^\lambda(Y_n)$ ergibt sich zu

$$A^\lambda(Y_n) \quad = \quad \sum_{i=1}^{k} (\frac{\pi_i}{h_i(\theta_0)})^\lambda \left\{ Y_{in} + \frac{1}{\lambda}\sqrt{n\pi_i} \left(1 - (\frac{\pi_i}{h_i(\theta_0)})^{-\lambda}\right) \right\}^2$$

$$-\frac{n}{\lambda^2}\sum_{i=1}^{k}\pi_i\left(1-(\frac{\pi_i}{h_i(\theta_0)})^{-\lambda}\right)^2(\frac{\pi_i}{h_i(\theta_0)})^{\lambda}$$
$$+n\Delta^{\lambda}(\pi,h(\theta_0)).$$

$A^0(Y_n)$ und $A^{-1}(Y_n)$ können durch Grenzwertbildung ermittelt werden.

Y_n ist asymptotisch normalverteilt:

$$Y_n \xrightarrow{d} N(\mathbf{0}, I-\sqrt{\pi}\sqrt{\pi}^T)$$

Sei U_n ein Zufallsvektor mit dieser Normalverteilung. Die Approximation ist nun durch die Verteilung von $A^{\lambda}(U_n)$ gegeben. Man erhält eine Linearkombination von möglicherweise nichtzentral χ^2-verteilten Zufallsvariablen plus einer Konstanten. Für die ausführliche Ableitung sei auf den Originalartikel verwiesen.

Diese Approximation ist für einen recht großen Wertebereich des Parameters λ anwendbar. Sie ist sowohl für „local alternatives“ als auch für feste Alternativen geeignet. Drost et al. (1989) geben in ihrem Artikel Fehlergrenzen an. Für die einfache Hypothese führt dieser Ansatz zur theoretisch und praktisch befriedigendsten Approximation.

Ein entscheidender Nachteil liegt aber in der Tatsache begründet, daß eine unmittelbare Verallgemeinerung auf den Fall der zusammmengesetzten Hypothese nicht möglich ist.

4.2.5 Abgewandelter Ansatz von Drost et al.

4.2.5.1 Einleitung

Die in Abschnitt 2.3.6 gegebene alternative Ableitung führt im Fall der χ^2-Diskrepanz ($\lambda = 1$) zur gleichen Approximation, die man nach dem Ansatz von Drost et al. (1989) erhält. Die alternative Ableitung besitzt den Vorteil, ohne Schwierigkeiten auf die zusammengesetzte Hypothese übertragbar zu sein. Die Idee läßt sich auch auf andere Prüfgrößen übertragen. Eine notwendige Voraussetzung ist, daß sich die Funktion $g(x, y)$ als Quadrat einer weiteren Funktion darstellen läßt. Es sei noch einmal betont, daß die folgenden Herleitungen, sofern Cressie-Read-Diskrepanzen betrachtet werden, für $\lambda \neq 1$ im allgemeinen nicht zu den Resultaten von Drost et al. führen, und die angegebenen Approximationen sollten den Approximationen von Drost et al. leicht unterlegen, jedoch der über „local alternatives“ gewonnenen nichtzentralen χ^2-Verteilung überlegen sein.

4.2.5.2 Neyman-χ^2-Diskrepanz

Ein Spezialfall der Cressie-Read-Diskrepanzen ist für $\lambda = -2$ durch die Neyman-χ^2-Diskrepanz gegeben:

$$\Delta^{-2}(\pi, h(\theta_0)) = \sum_{i=1}^{k} \frac{(\pi_i - h_i(\theta_0))^2}{\pi_i}$$

Sei $w(p) : (0,1]^k \to I\!R^k$ die Abbildung mit

$$w_i(p) := \frac{p_i - h_i(\theta_0)}{\sqrt{p_i}}, \quad i = 1, \ldots, k.$$

Die Ableitung von $w_i(p)$ nach p_j an der Stelle π ergibt sich zu

$$\left.\frac{\partial w_i}{\partial p_j}\right|_{p=\pi} = \frac{\delta_{ij}}{2\sqrt{\pi_i}} \left(1 + \frac{h_i(\theta_0)}{\pi_i}\right), \quad i, j = 1, \ldots, k.$$

Die Jacobi-Matrix W berechnet sich zu

$$W = \frac{1}{2} D_\pi^{-\frac{1}{2}} [I + D_\pi^{-1} D_h].$$

Mittels der Delta-Methode erhält man nun:

$$\sqrt{n}(w(z^{(n)}) - w(\pi)) \xrightarrow{d} N(\mathbf{0}, W(D_\pi - \pi\pi^T)W^T)$$

Die Verteilung der Prüfgröße $n\Delta(z^{(n)}, h(\theta_0)) = nw(z^{(n)})^T w(z^{(n)})$ wird durch die Verteilung von nY^TY approximiert, wobei gilt:

$$\sqrt{n}(Y - w(\pi)) \sim N(\mathbf{0}, W(D_\pi - \pi\pi^T)W^T)$$

Die Anwendung des Satzes 2.2 ergibt:

$$nY^TY \sim \sum_{\lambda_i > 0} \lambda_i \chi_1^2 (\frac{n(Sw(\pi))_i^2}{\lambda_i}) + \sum_{\lambda_i = 0} n(Sw(\pi))_i^2$$

S ist eine orthogonale Matrix, deren transponierte Zeilen eine orthonormale Basis aus Eigenvektoren der Matrix $V := W(D_\pi - \pi\pi^T)W^T$ bilden. Der konstante Term verschwindet im allgemeinen nicht.

4.2.5.3 Freeman-Tukey-Diskrepanz

Die Freeman-Tukey-Diskrepanz ist wiederum ein Spezialfall der Cressie-Read-Diskrepanzen. Für $\lambda = -1/2$ ergibt sich:

$$\Delta^{-\frac{1}{2}}(\pi, h(\theta_0)) = 4\sum_{i=1}^{k}(\sqrt{\pi_i} - \sqrt{h_i(\theta_0)})^2$$

Setzt man

$$w_i(p) := 2(\sqrt{p_i} - \sqrt{h_i(\theta_0)}), \quad i = 1, \ldots, k,$$

so folgt für die Jacobi-Matrix W an der Stelle π:

$$W = D_\pi^{-\frac{1}{2}}$$

Die Delta-Methode ergibt:

$$\sqrt{n}(w(z^{(n)}) - w(\pi)) \xrightarrow{d} N(\mathbf{0}, I - \sqrt{\pi}\sqrt{\pi}^T)$$

Deshalb wird die Verteilung von $n\Delta(z^{(n)}, h(\theta_0)) = nw(z^{(n)})^T w(z^{(n)})$ durch die Verteilung von nY^TY mit

$$\sqrt{n}(Y - w(\pi)) \sim N(\mathbf{0}, I - \sqrt{\pi}\sqrt{\pi}^T)$$

approximiert. Die Matrix $I - \sqrt{\pi}\sqrt{\pi}^T$ ist idempotent mit $k-1$ Eigenwerten eins. Es ergibt sich nun:

$$\sqrt{n}Y \sim N(\sqrt{n}w(\pi), I - \sqrt{\pi}\sqrt{\pi}^T)$$

Aus Satz 2.2 folgt

$$nY^TY \sim \sum_{\lambda_i=1} \chi_1^2(n(Sw(\pi))_i^2) + \sum_{\lambda_i=0} n(Sw(\pi))_i^2 \quad . \tag{4.3}$$

Ohne Beschränkung der Allgemeinheit sei s_k, die transponierte k-te Zeile von S, ein normierter Eigenvektor zum Eigenwert null. Der Eigenvektor ist bis auf das Vorzeichen durch

$$s_k = \sqrt{\pi}$$

gegeben. Für den konstanten Term aus (4.3) folgt:

$$\begin{aligned}\sum_{\lambda_i=0} n(Sw(\pi))_i^2 &= n(s_k^T w(\pi))^2 \\ &= 4n\left(\sqrt{\pi}^T(\sqrt{\pi}-\sqrt{h(\theta_0)})\right)^2 \\ &= 4n\left(1-\sum_{i=1}^k \sqrt{\pi_i h_i(\theta_0)}\right)^2\end{aligned}$$

Da

$$\begin{aligned}\sum_{i=1}^k (Sw(\pi))_i^2 &= \sum_{i=1}^k (w(\pi))_i^2 \\ &= \Delta^{-\frac{1}{2}}(\pi, h(\theta_0))\end{aligned}$$

ist, und elementare Berechnungen ergeben, daß

$$\Delta^{-\frac{1}{2}}(\pi, h(\theta_0)) - 4\left(1-\sum_{i=1}^k \sqrt{\pi_i h_i(\theta_0)}\right)^2 = \sigma^2/4 \tag{4.4}$$

ist, wobei

$$\sigma^2 = 16\pi^T D_\pi^{-\frac{3}{2}} D_h^{\frac{1}{2}}(D_\pi - \pi\pi^T) D_h^{\frac{1}{2}} D_\pi^{-\frac{3}{2}}\pi$$

sich aus (4.2) für $\lambda = -1/2$ ergibt, läßt sich die Approximation aus (4.3) einfacher als

$$nY^TY \sim \chi^2_{k-1}(n\sigma^2/4) + \left[n\Delta^{-\frac{1}{2}}(\pi, h(\theta_0)) - n\sigma^2/4\right]$$

schreiben.

4.2.5.4 Gauß-Diskrepanz

Für die Gauß-Diskrepanz

$$\Delta(\pi, h(\theta_0)) = \sum_{i=1}^k (\pi_i - h_i(\theta_0))^2$$

ist die Ableitung der Approximation besonders einfach. Aus

$$\sqrt{n}(z^{(n)} - \pi) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T)$$

folgt sofort:

$$\sqrt{n}((z^{(n)} - h(\theta_0)) - (\pi - h(\theta_0))) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T)$$

Sei Y ein Zufallsvektor mit

$$\sqrt{n}(Y - (\pi - h(\theta_0))) \sim N(\mathbf{0}, D_\pi - \pi\pi^T),$$

dann ergibt die Anwendung des Satzes 2.2:

$$nY^TY \sim \sum_{\lambda_i > 0} \lambda_i \chi_1^2 (\frac{n(S(\pi - h(\theta_0)))_i^2}{\lambda_i}) + \sum_{\lambda_i = 0} n(S(\pi - h(\theta_0)))_i^2 \tag{4.5}$$

S ist wieder eine orthogonale Matrix, deren transponierte Zeilen eine orthonormale Basis aus Eigenvektoren der Matrix $D_\pi - \pi\pi^T$ bilden, die zugehörigen Eigenwerte werden mit λ_i, $i = 1, \ldots, k$, bezeichnet.

Ohne Beschränkung der Allgemeinheit sei die k-te Zeile, s_k, der transponierte Eigenvektor zum einzigen Eigenwert $\lambda_k = 0$. Es folgt nun

$$\begin{aligned}
& (D_\pi - \pi\pi^T)s_k^T = \mathbf{0} \\
\Rightarrow\ & h(\theta_0)^T D_\pi^{-1}(D_\pi - \pi\pi^T)s_k^T = 0 \\
\Rightarrow\ & h(\theta_0)^T(I - (1, \ldots, 1)^T\pi^T)s_k^T = 0 \\
\Rightarrow\ & (h(\theta_0) - \pi)^T s_k^T = 0 \\
\Rightarrow\ & (S(h(\theta_0) - \pi))_k = 0,
\end{aligned}$$

d. h. der konstante Term aus (4.5) verschwindet.

4.2.6 Anpassung von Verteilungen über die Momente

Analog zum Fall der χ^2-Diskrepanz ist es möglich, Verteilungen über die Momente anzupassen. Für die Cressie-Read-Prüfgrößen $n\Delta^\lambda(z^{(n)}, h(\theta_0))$ können Erwartung und Varianz angegeben werden, falls $\lambda > -1$ ist, in den anderen Fällen existieren die Momente nicht (Read & Cressie, 1988). Eine angepaßte Gammaverteilung sollte zufriedenstellende Ergebnisse liefern. Für beliebige Diskrepanzen muß von Fall zu Fall entschieden werden, welche Verteilungen für eine Approximation in Frage kommen.

4.3 Zusammengesetzte Hypothesen

4.3.1 Verteilung unter der Hypothese

4.3.1.1 Cressie-Read-Diskrepanzen

Die Verteilung der aus einer beliebigen Diskrepanz $\Delta \in \mathcal{D}$ abgeleiteten Prüfgröße kann äußerst kompliziert sein. Für die Cressie-Read-Diskrepanzen folgen hingegen einfache asymptotische Resultate. Sei $\Delta_g \in \mathcal{D}^*$ und $\Delta = \Delta^\lambda \in \mathcal{D}^{CR}, \lambda \in I\!R$. Unter der Annahme, daß **B1-B5** erfüllt sind, folgt aus (3.8), daß die durch Minimierung von $\Delta_g(z^{(n)}, h(\theta))$ gewonnenen Schätzer $\hat{\theta}(z^{(n)})$ „best asymptotic normal“ sind. Nach Read & Cressie (1988) ergibt sich dann:

$$n\Delta^\lambda(z^{(n)}, h(\hat{\theta}(z^{(n)}))) \xrightarrow{d} \chi^2_{k-s-1}$$

4.3.1.2 Gauß-Diskrepanz

Sei Δ die Gauß-Diskrepanz und Δ_g eine beliebige Diskrepanz aus $\mathcal{D}$. Sei $w : [0,1]^k \to I\!R^k$ die Abbildung mit

$$w_i(p) := p_i - h_i(\hat{\theta}(p)), \quad i = 1, \ldots, k. \tag{4.6}$$

Die Jacobi-Matrix W der Ableitungen

$$\left.\frac{\partial w_i}{\partial p_j}\right|_{p=\pi} = \delta_{ij} - \sum_{m=1}^{p} \frac{\partial h_i}{\partial \theta_m} \frac{\partial \hat{\theta}_m}{\partial p_j}, \quad i, j = 1, \ldots, k, \tag{4.7}$$

ergibt sich zu

$$W = I + C^T \Omega^{-1} C G, \tag{4.8}$$

wobei die Bezeichnungen des Abschnitts 3.3.2 verwendet wurden. Mittels der Delta-Methode folgt nun:

$$\sqrt{n}(w(z^{(n)}) - w(\pi)) \xrightarrow{d} N(\mathbf{0}, W(D_\pi - \pi\pi^T)W^T) \tag{4.9}$$

Unter der Hypothese $\pi = h(\theta_0)$ verschwindet $w(\pi)$, und die Anwendung des Satzes 2.2 ergibt:

$$n w(z^{(n)})^T w(z^{(n)}) \sim \sum_{i=1}^{k} \lambda_i \chi_1^2$$

Die λ_i, $i = 1, \ldots, k$, bezeichnen die Eigenwerte der Matrix $Z = W(D_\pi - \pi\pi^T)W^T$.

4.3.2 Verteilung unter einer festen Alternative

4.3.2.1 Allgemeiner Fall

Der Schätzer $\hat{\theta}(z^{(n)})$ bzw. der Parameter θ_0 seien durch Minimierung der empirischen Diskrepanz $\Delta_g(z^{(n)}, h(\theta))$ bzw. von $\Delta_g(\pi, h(\theta))$ ermittelt worden. Sei die Modellfamilie falsch spezifiziert, d. h. $\pi \neq h(\theta_0)$. Für die Diskrepanz $\Delta = \sum_{i=1}^{k} g^*(\pi_i, h_i(\theta))$ seien die $1 \times k$-Matrizen G_x und G_y folgendermaßen definiert:

$$\begin{aligned}
(G_x)_j &= \left.\frac{\partial g^*(x,y)}{\partial x}\right|_{(x,y)=(\pi_j, h_j(\theta_0))}, \quad j = 1, \ldots, k \\
(G_y)_j &= \left.\frac{\partial g^*(x,y)}{\partial y}\right|_{(x,y)=(\pi_j, h_j(\theta_0))}, \quad j = 1, \ldots, k
\end{aligned}$$

Durch Anwendung der Delta-Methode kann eine asymptotische Verteilung für $\sqrt{n}(\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0)))$ abgeleitet werden. Für die Funktion $\Delta(p, h(\hat{\theta}(p)))$ folgt zunächst:

$$\begin{aligned}
\left.\frac{\partial \Delta(p, h(\hat{\theta}(p)))}{\partial p_j}\right|_{p=\pi} &= \left.\frac{\partial}{\partial p_j} \sum_{i=1}^{k} g^*(p_i, h_i(\hat{\theta}(p)))\right|_{p=\pi} \\
&= \sum_{i=1}^{k} \delta_{ij} \left.\frac{\partial g^*(x,y)}{\partial x}\right|_{(x,y)=(\pi_i, h_i(\theta_0))} \\
&\quad + \sum_{i=1}^{k} \left.\frac{\partial g^*(x,y)}{\partial y}\right|_{(x,y)=(\pi_i, h_i(\theta_0))} \sum_{m=1}^{s} \frac{\partial h_i}{\partial \theta_m} \frac{\partial \hat{\theta}_m}{\partial p_j}, \\
&\quad j = 1 \ldots, k.
\end{aligned}$$

In Matrixschreibweise ergibt sich für die $1 \times k$-Jacobi-Matrix W

$$W = G_x - G_y C^T \Omega^{-1} C G \quad .$$

Mittels der Delta-Methode folgt weiterhin:

$$\sqrt{n}(\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))) \xrightarrow{d} N(0, W(D_\pi - \pi\pi^T)W^T)$$

4.3.2.2 Die Sonderstellung der Minimum–Diskrepanz–Schätzer

Ist $\Delta = \Delta_g$, wird also die betrachtete Diskrepanz gleichzeitig zur Ermittlung der Schätzer verwendet, so erhält man ein sehr einfaches asymptotisches Ergebnis. Werden die Ableitungen von $\Delta(\pi, h(\theta))$ nach $\theta_j, j = 1, \ldots, s$, gebildet,

so ergibt sich:

$$\begin{aligned}\left.\frac{\partial \Delta(\pi, h(\theta))}{\partial \theta_j}\right|_{\theta=\theta_0} &= \left.\frac{\partial}{\partial \theta_j} \sum_{i=1}^{k} g^*(\pi_i, h_i(\theta))\right|_{\theta=\theta_0} \\ &= \sum_{i=1}^{k} \left.\frac{\partial g^*(x,y)}{\partial y}\right|_{(x,y)=(\pi_i, h_i(\theta_0))} \frac{\partial h_i}{\partial \theta_j}\end{aligned}$$

Für den Gradienten gilt somit für $\theta = \theta_0$:

$$\text{grad} = CG_y^T = \mathbf{0}$$

Die oben betrachtete Jacobi-Matrix vereinfacht sich zu $W = G_x$, und es folgt in Verbindung mit (4.1), daß

$$\sqrt{n}(\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0)))$$

und

$$\sqrt{n}(\Delta(z^{(n)}, h(\theta_0)) - \Delta(\pi, h(\theta_0)))$$

asymptotisch äquivalent sind und gegen die gleiche Normalverteilung streben.

4.3.3 Verteilung unter „local alternatives“

Im Fall der „local alternatives“ seien wieder nur die Cressie-Read-Diskrepanzen betrachtet. Nach Theorem A8.1 von Read & Cressie (1988, S. 171) folgt, in Verbindung mit Satz 3.8, daß es erlaubt ist, einen Minimum-Diskrepanz-Schätzer zu einer beliebigen Diskrepanz $\Delta_g \in \mathcal{D}^*$ in eine beliebige Diskrepanz aus der Cressie-Read-Familie einzusetzen, ohne daß sich die asymptotische Verteilung unter „local alternatives“ ändert. Es ergibt sich die in Satz 3.8 genannte nichtzentrale χ^2-Verteilung.

Bemerkung: Auch wenn der von Read & Cressie angegebene Nichtzentralitätsparameter i. A. nicht korrekt ist, kann das Theorem A8.1 trotzdem verwendet werden. Read & Cressie beziehen sich zwar auf die Regularitätsbedingungen von Birch (1964), aber die für ihren Beweis notwendigen Voraussetzungen folgen auch aus den Bedingungen **B1-B5**.

4.3.4 Ein verbesserter Ansatz zur Approximation der Verteilung von speziellen Prüfgrößen

4.3.4.1 Allgemeiner Ansatz

Es wird nun das Vorgehen aus Abschnitt 4.2.5 auf den Fall der zusammengesetzten Hypothese übertragen. Sei $\Delta(\pi, h(\theta)) \in \mathcal{D}$ eine Diskrepanz der Form

$$\Delta(\pi, h(\theta)) = \sum_{i=1}^{k} g^*(\pi_i, h_i(\theta)),$$

wobei sich $g^*(x, y)$ als Quadrat einer Funktion $w^*(x, y)$ darstellen läßt. Weiterhin seien die Schätzer $\hat{\theta}(z^{(n)})$ durch Minimierung einer empirischen Diskrepanz $\Delta_g(z^{(n)}, h(\theta))$ gebildet.

Sei nun die Abbildung $w : (0; 1)^k \to I\!R^k$ durch

$$w_i(p) := w^*(p_i, h_i(\hat{\theta}(p))), \quad i = 1, \ldots, k,$$

definiert. Bezeichne W die Jacobi-Matrix dieser Abbildung in π. Nach Anwendung der Delta-Methode ergibt sich nun:

$$\sqrt{n}(w(z^{(n)}) - w(\pi))) \xrightarrow{d} N(0, W(D_\pi - \pi\pi^T)W^T)$$

Hier muß nun nur vorausgesetzt werden, daß die Matrix $W(D_\pi - \pi\pi^T)W^T$ nicht verschwindet.

Da weiterhin

$$\begin{aligned} nw(z^{(n)})^T w(z^{(n)}) &= n\sum_{i=1}^{k}(w_i(z^{(n)}))^2 \\ &= n\sum_{i=1}^{k}(w^*(z_i^{(n)}, h_i(\hat{\theta}(z^{(n)}))))^2 \\ &= n\sum_{i=1}^{k} g^*(z_i^{(n)}, h_i(\hat{\theta}(z^{(n)}))) \\ &= n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) \end{aligned}$$

ist, liegt es nahe, als Approximation die Verteilung von nY^TY mit

$$\sqrt{n}(Y - w(\pi))) \sim N(0, W(D_\pi - \pi\pi^T)W^T)$$

zu verwenden. Nach Satz 2.2 ergibt sich:

$$nY^TY \sim \sum_{\lambda_i>0} \lambda_i \chi_1^2(\frac{n(Sw(\pi))_i^2}{\lambda_i}) + \sum_{\lambda_i=0} n(Sw(\pi))_i^2$$

Die Spalten von S^T bilden eine Basis aus Eigenvektoren zur Matrix $V = W(D_\pi - \pi\pi^T)W^T$, die λ_i bezeichnen die zugehörigen Eigenwerte. Im folgenden wird die Jacobi-Matrix W für spezielle Diskrepanzen berechnet.

4.3.4.2 Neyman-χ^2-Diskrepanz

Für die Neyman-χ^2-Diskrepanz

$$\Delta^{-2}(\pi, h(\theta)) = \sum_{i=1}^{k} \frac{(\pi_i - h_i(\theta))^2}{\pi_i}$$

gilt:

$$w_i(p) := \frac{p_i - h_i(\hat{\theta}(p))}{\sqrt{p_i}}, \quad i = 1, \ldots, k$$

Die partiellen Ableitungen nach p_j ergeben sich für $p = \pi$ zu

$$\frac{\partial w_i}{\partial p_j} = \frac{\delta_{ij}}{2\sqrt{\pi_i}} - \frac{1}{\pi_i}\left(\sum_{m=1}^{s} \frac{\partial h_i}{\partial \theta_m}\frac{\partial \hat{\theta}_m}{\partial p_j}\sqrt{\pi_i} - \delta_{ij}\frac{h_i(\theta_0)}{2\sqrt{\pi_i}}\right) \quad .$$

Die Jacobi-Matrix W berechnet sich zu

$$W = \frac{1}{2} D_\pi^{-\frac{1}{2}}[I + D_\pi^{-1} D_h + 2C^T\Omega^{-1}CG] \quad .$$

4.3.4.3 Freeman-Tukey-Diskrepanz

Für die Freeman-Tukey-Diskrepanz

$$\Delta^{-\frac{1}{2}}(\pi, h(\theta)) = \sum_{i=1}^{k} 4(\sqrt{\pi_i} - \sqrt{h_i(\theta)})^2$$

ergibt sich:

$$w_i(p) := 2(\sqrt{p_i} - \sqrt{h_i(\hat{\theta}(p))}), \quad i = 1, \ldots, k$$

Für die partiellen Ableitungen nach p_j, $j = 1, \ldots, k$, in $p = \pi$ folgt:

$$\frac{\partial w_i}{\partial p_j} = \frac{\delta_{ij}}{\sqrt{\pi_i}} - \frac{1}{\sqrt{h_i(\theta_0)}} \sum_{m=1}^{s} \frac{\partial h_i}{\partial \theta_m} \frac{\partial \hat{\theta}_m}{\partial p_j}$$

In Matrixschreibweise erhält man nun:

$$W = D_\pi^{-\frac{1}{2}} + D_h^{-\frac{1}{2}} C^T \Omega^{-1} C G$$

4.3.4.4 Gauß-Diskrepanz

Für die Gauß-Diskrepanz

$$\Delta(\pi, h(\theta)) = \sum_{i=1}^{k} (\pi_i - h_i(\theta))^2$$

ist

$$w_i(p) := (p_i - h_i(\hat{\theta}(p))), \quad i = 1, \ldots, k.$$

Somit folgt:

$$\frac{\partial w_i}{\partial p_j} = \delta_{ij} - \sum_{m=1}^{s} \frac{\partial h_i}{\partial \theta_m} \frac{\partial \hat{\theta}_m}{\partial p_j}$$

W ergibt sich zu

$$W = I + C^T \Omega^{-1} C G.$$

Kapitel 5

Bootstrap-Tests

5.1 Das Bootstrap-Verfahren

Das von Efron (1979) eingeführte Bootstrap-Verfahren stellt ein universell einsetzbares Instrument zur Schätzung der Verteilung von Statistiken dar. Durch Verwendung von Bootstrap-Methoden werden die technischen Schwierigkeiten umgangen, die bei einer exakten Ableitung der Verteilung einer Statistik auftreten können. Die Vorgehensweise soll nun kurz umrissen werden. Gegeben seien die Realisationen $x_1, \ldots, x_n$ von unabhängig und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$ mit gemeinsamer, unbekannter Verteilungsfunktion F. Sei F_n die empirische Verteilungsfunktion

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(x_i \leq x),$$

wobei mit $I(A)$ die Indikatorfunktion bezeichnet wird. Gegenstand der Untersuchung sei eine Statistik $R_n(X_1, \ldots, X_n; F)$. Die Bootstrap-Methode besteht nun darin, die Verteilung von $R_n(X_1, \ldots, X_n; F)$ durch die Verteilung von $R_n(X_1^*, \ldots, X_n^*; F_n)$ zu approximieren, wobei $X_1^*, \ldots, X_n^*$ unabhängig und identisch verteilte Zufallsvariablen mit gemeinsamer Verteilungsfunktion F_n sind. Die Verteilungsfunktion der Statistik $R_n(X_1, \ldots, X_n; F)$,

$$J_n(x, F) := P_F(R_n(X_1, \ldots, X_n; F) \leq x),$$

wird durch

$$J_n(x, F_n) := P_{F_n}(R_n(X_1^*, \ldots, X_n^*; F_n) \leq x)$$

geschätzt.

Die in den folgenden Abschnitten vorgestellten Tests beruhen auf dem gemeinsamen Prinzip, bestimmte Prozentpunkte der Verteilung einer Statistik unter F durch Bootstrap-Methoden anzunähern.

Für beliebiges $\alpha \in (0,1)$ definiere man (vgl. Beran, 1984)

$$\begin{aligned} c_{n,L}(\alpha, F) &:= inf\{x : J_n(x, F) \geq 1 - \alpha\}, \\ c_{n,U}(\alpha, F) &:= sup\{x : J_n(x, F) \leq 1 - \alpha\}. \end{aligned}$$

Die Bootstrap-Schätzer sind dann durch

$$\begin{aligned} c_{n,L}(\alpha, F_n) &:= inf\{x : J_n(x, F_n) \geq 1 - \alpha\}, \\ c_{n,U}(\alpha, F_n) &:= sup\{x : J_n(x, F_n) \leq 1 - \alpha\}. \end{aligned} \tag{5.1}$$

gegeben. $c_{n,L}(\alpha, F_n)$ und $c_{n,U}(\alpha, F_n)$ lassen sich zwar prinzipiell berechnen, doch der Aufwand ist in den meisten Fällen unvertretbar hoch. Durch Verwendung von Monte-Carlo-Methoden können Schätzer für $c_{n,L}(\alpha, F_n)$ und $c_{n,U}(\alpha, F_n)$ angegeben werden. Dazu werden wiederholt Stichproben $x^*_{b1}, \ldots, x^*_{bn}$ $(b = 1, \ldots, B)$ der Größe n unter F_n erzeugt. Sei $J^*_{n,B}(x)$ die empirische Verteilungsfunktion zu den erzeugten Werten $R_n(x^*_{11}, \ldots, x^*_{1n}; F_n), \ldots, R_n(x^*_{B1}, \ldots, x^*_{Bn}; F_n)$:

$$J^*_{n,B}(x) := B^{-1} \sum_{b=1}^{B} I(R_n(x^*_{b1}, \ldots, x^*_{bn}; F_n) \leq x)$$

Mit

$$\begin{aligned} c_{n,L,B}(\alpha, F_n) &:= inf\{x : J^*_{n,B}(x) \geq 1 - \alpha\}, \\ c_{n,U,B}(\alpha, F_n) &:= sup\{x : J^*_{n,B}(x) \leq 1 - \alpha\} \end{aligned}$$

sind dann Schätzer für $c_{n,L}(\alpha, F_n)$ und $c_{n,U}(\alpha, F_n)$ gegeben.

Sinnvoll einsetzbar sind diese Bootstrap-Schätzer nur dann, wenn sichergestellt werden kann, daß sie mit wachsender Stichprobengröße genügend schnell gegen die entsprechenden Prozentpunkte der Statistik unter F konvergieren. Sei für jedes F_n eine reelle Zahl $c_n(\alpha, F_n)$ definiert mit der Eigenschaft:

$$c_{n,L}(\alpha, F_n) \leq c_n(\alpha, F_n) \leq c_{n,U}(\alpha, F_n) \tag{5.2}$$

Bei den auftretenden Statistiken muß dann gezeigt werden, daß folgende Konvergenzaussage zutrifft:

$$\lim_{n \to \infty} J_n(c_n(\alpha, F_n), F) = 1 - \alpha$$

Dazu wird jeweils Theorem 1 von Beran (1984) verwendet. Für die Anwendung dieses Satzes müssen die folgenden Voraussetzungen erfüllt sein:

V1 Es muß eine Menge $\mathcal{G}_F$, bestehend aus Folgen (G_n) von Verteilungsfunktionen definiert werden, dergestalt, daß

V2 die Folge (F_n) der empirischen Verteilungsfunktionen unter dem Grundmodell fast sicher der Menge $\mathcal{G}_F$ angehört:

$$P_F[(F_n) \in \mathcal{G}_F] = 1$$

V3 Für jede Folge $(G_n) \in \mathcal{G}_F$ konvergiert $J_n(x, G_n)$ schwach gegen eine Grenzverteilungsfunktion $J(x, F)$, die nur von $\mathcal{G}_F$ abhängt.

5.2 Bootstrap-Tests für die χ^2-Diskrepanz

5.2.1 Vorbereitungen

Zunächst seien die Voraussetzungen für die auftretenden Größen zusammengefaßt. Sei $X_1, X_2, X_3, \ldots,$ eine Folge von unabhängig und identisch verteilten diskreten Zufallsvariablen unter dem Grundmodell F, d. h. mit Wahrscheinlichkeitsfunktion

$$f(i) = \begin{cases} \pi_i & i \in \{1, \ldots, k\} \\ 0 & sonst \end{cases}$$

und Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x < 1 \\ \sum_{i=1}^{[x]} \pi_i & 1 \leq x < k \\ 1 & x \geq k \end{cases} .$$

(Es sei $[x]$:=„größte ganze Zahl $\leq x$“.) Sei $z^{(n)} \in I\!R^k$ wieder der Vektor der relativen Häufigkeiten und F_n die empirische Verteilungsfunktion zu gegebenen Beobachtungen $x_1, \ldots, x_n$. Zwischen $F_n(x)$ und $z^{(n)}$ besteht der Zusammenhang

$$F_n(x) = \begin{cases} 0 & x < 1 \\ \sum_{i=1}^{[x]} z_i^{(n)} & 1 \leq x < k \\ 1 & x \geq k \end{cases} \tag{5.3}$$

Zu gegebenem n kann deshalb einerseits aus $F_n(x)$ der Vektor $z^{(n)}$ rekonstruiert, andererseits aus $z^{(n)}$ die empirische Verteilungsfunktion ermittelt werden.

Nun wird für den konkret vorliegenden Fall die in Abschnitt 5.1 auf S. 132 erwähnte Menge $\mathcal{G}_F$ konstruiert.

Definition 5.1 *Sei $p^{(1)}, p^{(2)}, p^{(3)}, \ldots$ eine Folge von Vektoren des $\mathbb{R}^k$ mit $p_i^{(n)} \geq 0, i = 1, \ldots, k$, und $\sum_{i=1}^k p_i^{(n)} = 1$. Sei dadurch eine Folge (G_n) von Verteilungsfunktionen mit*

$$G_n(x) = \begin{cases} 0 & x < 1 \\ \sum_{i=1}^{[x]} p_i^{(n)} & 1 \leq x < k \\ 1 & x \geq k \end{cases} \tag{5.4}$$

definiert. Die Menge $\mathcal{G}_F^*$ bestehe nun aus allen möglichen Folgen (G_n) dieser Art. □

Lemma 5.1 *Sei $(G_n) \in \mathcal{G}_F^*$ und $D(F, G_n)$ die Kolmogoroff-Smirnov-Distanz zwischen $F(x)$ und $G_n(x), n = 1, 2, 3, \ldots$:*

$$D(F, G_n) := \sup_{x \in \mathbb{R}} \mid F(x) - G_n(x) \mid$$

Dann ist die Bedingung $\lim_{n\to\infty} p^{(n)} = \pi$ notwendig und hinreichend für

$$\lim_{n\to\infty} D(F, G_n) = 0$$

Beweis:

$$\begin{aligned} D(F, G_n) &= \sup_{x \in \mathbb{R}} \mid F(x) - G_n(x) \mid \\ &= \sup_{x \in [1,k)} \mid \sum_{i=1}^{[x]} (\pi_i - p_i^{(n)}) \mid \\ &= \max_{x \in \{1,\ldots,k-1\}} \mid \sum_{i=1}^{x} (\pi_i - p_i^{(n)}) \mid \end{aligned}$$

$$
\begin{aligned}
\Rightarrow \quad & \lim_{n\to\infty} D(F, G_n) = 0 \\
\Leftrightarrow \quad & \lim_{n\to\infty} \mid \sum_{i=1}^{x} (\pi_i - p_i^{(n)}) \mid = 0 \quad \forall x \in \{1, \ldots, k-1\} \\
\Leftrightarrow \quad & \lim_{n\to\infty} \mid \pi_i - p_i^{(n)} \mid = 0 \quad \forall i \in \{1, \ldots, k-1\} \\
\Leftrightarrow \quad & \lim_{n\to\infty} p^{(n)} = \pi
\end{aligned}
$$

□

Die Menge $\mathcal{G}_F$ ist nun folgendermaßen definiert:

Definition 5.2 *Sei* $\mathcal{G}_F := \{(G_n) \in \mathcal{G}_F^* : \lim_{n\to\infty} D(F, G_n) = 0\}$

Da eine beliebige Folge der F_n nach (5.3) und (5.4) in $\mathcal{G}_F^*$ liegt, genügt es zu zeigen, daß $P_F[(F_n) : \lim_{n\to\infty} D(F, F_n) = 0] = 1$ gilt. Dies ist aber die Aussage des Satzes von Glivenko-Cantelli (Serfling, 1980, S. 61). Somit sind **V1** und **V2** erfüllt.

Für jeden in den nächsten Abschnitten vorgestellten Test ist die dritte Bedingung **V3** herzuleiten. Da die Statistiken $R_n(X_1, \ldots, X_n; F)$ von Fall zu Fall unterschiedlich aussehen, ist an dieser Stelle ein allgemeiner Ansatz nicht möglich. Sämtliche Tests bauen jedoch auf dem folgenden Resultat auf:

Satz 5.1 *Sei* $(G_n) \in \mathcal{G}_F$ *und sei für jedes* n *eine endliche Folge von unabhängig und identisch verteilten Zufallsvariablen* $X_1^{(n)}, \ldots, X_n^{(n)}$ *mit Verteilungsfunktion* G_n *gegeben. Sei* $z_*^{(n)}$ *der Vektor der relativen Häufigkeiten zu gegebenen Realisationen* $x_1^{(n)}, \ldots, x_n^{(n)}$. *Dann gilt*

$$\sqrt{n}(z_*^{(n)} - p^{(n)}) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T)$$

Beweis:
1. Die Folge $(G_n) \in \mathcal{G}_F$ ist durch die zugehörige Folge der $p^{(n)}$ eindeutig beschrieben. Eine Zufallsvariable mit Verteilungsfunktion G_n besitzt dann die Wahrscheinlichkeitsfunktion

$$g_n(i) = \begin{cases} p_i^{(n)} & i \in \{1, \ldots, k\} \\ 0 & sonst \end{cases}$$

$\Rightarrow$ Der Vektor der absoluten Häufigkeiten $nz^{(n)}$ ist multinomialverteilt mit Parametern n und $p^{(n)}$.
$\Rightarrow$ Die momenterzeugende Funktion zu

$$Y^{(n)} := \sqrt{n}(z_*^{(n)} - p^{(n)})$$

ist (vgl. Bishop et al., 1975) durch

$$M_{Y^{(n)}}(\mathbf{t}) = \left(\sum_{i=1}^{k} p_i^{(n)} \exp\left(n^{-\frac{1}{2}}(t_i - \mathbf{t}^T p^{(n)})\right)\right)^n$$

gegeben ($\mathbf{t} := (t_1, \ldots, t_k)^T$).

2. Aus $\{G_n\} \in \mathcal{G}_F$ folgt $\lim_{n\to\infty} p^{(n)} = \pi$, d. h. $p^{(n)} = \pi + o(1)$. Da

$$\exp(x) = 1 + x + \frac{1}{2}x^2 + o(x^2), \quad x \to 0,$$

folgt:

$$M_{Y^{(n)}}(\mathbf{t}) = \left(\sum_{i=1}^{k} p_i^{(n)} \left(1 + n^{-\frac{1}{2}}(t_i - \mathbf{t}^T p^{(n)}) + \frac{1}{2}n^{-1}(t_i - \mathbf{t}^T p^{(n)})^2 + o(n^{-1})\right)\right)^n$$

Nun ist

$$\begin{aligned}\sum_{i=1}^{k} p_i^{(n)}(t_i - \mathbf{t}^T p^{(n)}) &= \sum_{i=1}^{k} p_i^{(n)} t_i - \mathbf{t}^T p^{(n)} \sum_{i=1}^{k} p_i^{(n)} \\ &= \mathbf{t}^T p^{(n)} - \mathbf{t}^T p^{(n)} = 0,\end{aligned}$$

also folgt weiter:

$$\begin{aligned}M_{Y^{(n)}}(\mathbf{t}) &= \left(1 + \frac{1}{2}n^{-1}\sum_{i=1}^{k} p_i^{(n)}(t_i - \mathbf{t}^T p^{(n)})^2 + o(n^{-1})\right)^n \\ &= \left(1 + \frac{1}{2}n^{-1}\sum_{i=1}^{k}(\pi_i + o(1))(t_i - \mathbf{t}^T \pi + o(1))^2 + o(n^{-1})\right)^n \\ &= \left(1 + \frac{1}{2}n^{-1}\sum_{i=1}^{k}\pi_i(t_i - \mathbf{t}^T \pi)^2 + o(n^{-1})\right)^n\end{aligned}$$

Nach Bishop et al. (1975) ergibt sich nun

$$\lim_{n\to\infty} M_{Y^{(n)}}(\mathbf{t}) = \exp(\frac{1}{2}\mathbf{t}^T(D_\pi - \pi\pi^T)\mathbf{t})$$

$$\Rightarrow \sqrt{n}(z_*^{(n)} - p^{(n)}) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T)$$

Damit ist die Behauptung bewiesen. □

5.2.2 Ein Test über $\Delta(\theta_0)$

Gegeben sei ein Grundmodell π und eine approximierende Modellfamilie $\{h(\theta) : \theta \in \Theta\}$. Häufig besteht ein berechtigter Grund zur Annahme, daß das Grundmodell nicht in der approximierenden Modellfamilie enthalten ist. Unter dieser Voraussetzung nimmt die χ^2-Diskrepanz $\Delta(\pi, h(\theta))$ nur positive Werte an. Von Interesse ist dann die Frage, ob wenigstens die Diskrepanz durch Approximation, also

$$\Delta(\pi, h(\theta_0)) = \min_{\theta\in\Theta} \Delta(\pi, h(\theta))$$

eine vorgegebene Schranke $d > 0$ nicht überschreitet.

Es bereitet keinerlei Schwierigkeiten, die genannte Problemstellung in der folgenden Weise zu erweitern:

Sei für $\Delta_g \in \mathcal{D}$

$$\theta_0 := \arg\min_{\theta\in\Theta} \Delta_g(\pi, h(\theta)).$$

Dann sei zu testen, ob die Aussage

$$\Delta(\pi, h(\theta_0)) \leq d$$

zutrifft. Für $\Delta_g = \Delta$ ist dieser Ansatz mit dem vorhergehenden identisch.

Es wird nun ein Test zu dieser Hypothese unter Verwendung von Bootstrap-Methoden hergeleitet:

Sei $\pi \notin \{h(\theta) : \theta \in \Theta\}$ und **B1-B5** für eine Diskrepanz $\Delta_g \in \mathcal{D}$ erfüllt. Mit der bisherigen Notation dieses Kapitels sei die Statistik

$$R_n(X_1, \ldots, X_n; F) := \sqrt{n}[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))]$$

definiert. Unter Verwendung der Definitionen des Abschnitts 5.1 kann nun der folgende Satz bewiesen werden:

Satz 5.2 *Unter der Hypothese*

$$H_0^*: \quad \Delta(\pi, h(\theta_0)) = d, \quad d > 0,$$

gilt

$$\lim_{n\to\infty} P_F(\sqrt{n}[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - d] \geq c_n(\alpha, F_n)) = \alpha.$$

Beweis:
Der Beweis verwendet Theorem 1 von Beran (1984). Sei $\mathcal{G}_F$ wie in Definition 5.2 gebildet, und sei $(G_n) \in \mathcal{G}_F$ eine beliebige Folge von Verteilungsfunktionen.
Zu zeigen ist: **V3** ist erfüllt, d.h. $J_n(x, G_n)$ konvergiert schwach gegen eine Grenzverteilungsfunktion, die nur von $\mathcal{G}_F$ abhängt. Nun ist

$$J_n(x, G_n) = P_{G_n}(R_n(X_1^{(n)}, \ldots, X_n^{(n)}; G_n) \leq x)$$

und

$$R_n(X_1^{(n)}, \ldots, X_n^{(n)}; G_n) = \sqrt{n}[\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(p^{(n)}, h(\hat{\theta}(p^{(n)})))].$$

$X_1^{(n)}, \ldots, X_n^{(n)}$ sind unabhängig und identisch verteilte, diskrete Zufallsvariablen mit Verteilungsfunktion G_n, $p^{(n)}$ der zu G_n zugehörige Wahrscheinlichkeitsvektor und $z_*^{(n)}$ der Vektor der relativen Häufigkeiten. In Satz 5.1 wurde gezeigt, daß

$$\sqrt{n}(z_*^{(n)} - p^{(n)}) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T)$$

gilt. Die Anwendung der multivariaten Deltamethode (Satz 2.3) in Verbindung mit Satz 3.6 ergibt nun

$$\sqrt{n}[\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(p^{(n)}, h(\hat{\theta}(p^{(n)})))] \xrightarrow{d} N(0, \sigma^2), \tag{5.5}$$

wobei $\sigma^2 > 0$ wie in Satz 3.6 definiert ist und nicht von der speziellen Folge (G_n) abhängt.
$\Rightarrow$ $J_n(x, G_n)$ konvergiert schwach gegen eine Normalverteilungsfunktion $J(x, F) = \Phi(x/\sigma)$. Die Anwendung des Theorems 1 von Beran (1984) ergibt nun:

$$\lim_{n\to\infty} P_F(\sqrt{n}[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))] \leq c_n(\alpha, F_n)) = 1 - \alpha$$

Da die Grenzverteilungsfunktion stetig ist, folgt unter der Hypothese H_0^* : $\Delta(\pi, h(\theta_0)) = d$

$$\lim_{n\to\infty} P_F(\sqrt{n}[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - d] \geq c_n(\alpha, F_n)) = \alpha,$$

womit der Satz bewiesen ist. $\square$

Der so erzeugte kritische Wert $c_n(\alpha, F_n)$ kann natürlich auch zum Test der Hypothese

$$H_0 : \Delta(\pi, \theta_0) \leq d, \quad d > 0,$$

verwendet werden, da ein einseitiger Ablehnungsbereich festgelegt wurde.

Somit können die Schritte zur Durchführung des Testes zur Hypothese

$$H_0 : \quad \Delta(\pi, h(\theta_0)) \leq d, \quad d > 0,$$

nun genannt werden:

Schritt 1 Getestet wird die Hypothese $H_0 : \Delta(\theta_0) \leq d, \quad d > 0$.

Schritt 2 Gegeben seien die Realisationen $x_1, \ldots, x_n$ von unter dem Grundmodell F unabhängig und identisch verteilten Zufallsvariablen. Aus diesen Beobachtungen bilde man F_n und $z^{(n)}$.

Schritt 3 Man bestimme die Verteilung von

$$\sqrt{n}[\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))]$$

unter F_n.

Schritt 4 Man berechne $c_{n,L}(\alpha, F_n)$ und $c_{n,U}(\alpha, F_n)$ und bestimme einen Wert $c_n(\alpha, F_n)$ mit

$$c_{n,L}(\alpha, F_n) \leq c_n(\alpha, F_n) \leq c_{n,U}(\alpha, F_n)$$

Schritt 5 Die Hypothese wird verworfen, falls

$$\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) > n^{-\frac{1}{2}} c_n(\alpha, F_n) + d$$

ist.

Der in dieser Art und Weise konstruierte Test hat als Folgerung des letzten Satzes zumindest das asymptotische Niveau α. Ob sich das Testverfahren in der Praxis bewähren kann, wird in Abschnitt 5.3.2.1 anhand von Beispielen untersucht.

Schritt 3 läßt sich prinzipiell wegen der endlichen Anzahl der angenommenen Werte exakt durchführen. Sind die Berechnungen zu aufwendig, werden, wie in Abschnitt 5.1 erwähnt, Monte-Carlo-Methoden verwendet.

5.2.3 Ein alternativer Test über $\Delta(\theta_0)$

Hall und Wilson (1991) geben allgemeine Regeln zur Konstruktion von Bootstrap-Tests. Überträgt man ihre Anweisungen auf die vorliegende Situation, so sollte als Statistik nicht

$$R_n^{(1)}(X_1,\ldots,X_n;F) = \sqrt{n}[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))] \quad ,$$

sondern

$$R_n^{(2)}(X_1,\ldots,X_n;F) = \frac{\sqrt{n}[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))]}{\sigma(z^{(n)})}$$

verwendet werden, wobei $\sigma(z^{(n)})$ einen Schätzer der asymptotischen Standardabweichung aus (5.5) bezeichnet.

Zunächst wird gezeigt, daß sich wieder ein Test zur Hypothese H_0 : $\Delta(\theta_0) \leq d$ mit asymptotischen Niveau α herleiten läßt, anschließend werden einige Probleme betrachtet, die bei kleinen Stichproben auftreten können.

Obwohl die gewählte Diskrepanz Δ_g eine beliebige aus $\mathcal{D}$ sein kann, soll nur der Fall $\Delta_g = \Delta$ betrachtet werden, da dann ein Schätzer $\sigma(z^{(n)})$ ohne Schwierigkeiten angegeben werden kann. Aus Lemma 3.4 und (2.4) ergibt sich für die asymptotische Varianz der Statistik $R_n^{(1)}$:

$$\sigma^2 = 4(\sum_{i=1}^{k} \frac{\pi_i^3}{h_i^2(\theta_0)} - (\sum_{i=1}^{k} \frac{\pi_i^2}{h_i(\theta_0)})^2)$$

Nun läßt sich die Funktion $\sigma^2(p)$ mit

$$\sigma^2(p) = 4(\sum_{i=1}^{k} \frac{p_i^3}{h_i^2(\hat{\theta}(p))} - (\sum_{i=1}^{k} \frac{p_i^2}{h_i(\hat{\theta}(p))})^2)$$

definieren. $\sigma^2(p)$ ist stetig in π mit $\sigma^2(\pi) = \sigma^2$.

Sei nun (G_n) eine beliebige Folge aus $\mathcal{G}_F$. Dann kann das folgende Lemma bewiesen werden:

Lemma 5.2 *Für beliebiges* $(G_n) \in \mathcal{G}_F$ *gilt* $\sigma(z_*^{(n)}) \xrightarrow{p} \sigma(\pi)$

Beweis:
1. Sei $\epsilon > 0$ gegeben, aus $(G_n) \in \mathcal{G}_F$ folgt:

$$\exists N_0 \in I\!N \text{ mit } \| p^{(n)} - \pi \| < \frac{\epsilon}{2} \quad \forall n \geq N_0$$

$$
\begin{aligned}
\Rightarrow P(\| z_*^{(n)} - \pi \| \geq \epsilon) &= P(\| z_*^{(n)} - p^{(n)} + p^{(n)} - \pi \| \geq \epsilon) \\
&\leq P(\| z_*^{(n)} - p^{(n)} \| + \| p^{(n)} - \pi \| \geq \epsilon) \\
&\leq P(\| z_*^{(n)} - p^{(n)} \| \geq \frac{\epsilon}{2}) \quad \forall n \geq N_0
\end{aligned}
$$

2. Nach Satz 5.1 gilt

$$\sqrt{n}(z_*^{(n)} - p^{(n)}) \xrightarrow{d} N(\mathbf{0}, D_\pi - \pi\pi^T)$$

Nach Serfling (1980, S. 26) folgt:

$$z_*^{(n)} - p^{(n)} \xrightarrow{p} \mathbf{0}$$

3. Die Kombination von 1. und 2. ergibt nun:

$$z_*^{(n)} - \pi \xrightarrow{p} \mathbf{0} \Rightarrow z_*^{(n)} \xrightarrow{p} \pi$$

Da $\sigma(p) = (\sigma^2(p))^{\frac{1}{2}}$ stetig in π ist, folgt die Behauptung. □

Ist $\pi \notin \{h(\theta) : \theta \in \Theta\}$, dann ist $\sigma(\pi) > 0$. Aus Slutskys Theorem (Serfling, 1980, S. 19) folgt nun sofort das gewünschte Resultat:

Lemma 5.3 *Sei* **B1-B5** *erfüllt und* $\Delta(\theta_0) = d > 0$. *Für* $(G_n) \in \mathcal{G}_F$ *besitzt* $R_n^{(2)}$ *eine asymptotische Verteilung:*

$$\sqrt{n}\frac{\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(p^{(n)}, h(\hat{\theta}(p^{(n)})))}{\sigma(z_*^{(n)})} \xrightarrow{d} N(0,1)$$

Dies bedeutet, daß **V3** aus Abschnitt 5.1 für die Statistik $R_n^{(2)}$ erfüllt ist. Der Satz von Beran (1984) läßt sich anwenden und ein Test mit asymptotischen Niveau α konstruieren.

Obwohl für $n \to \infty$ die Wahrscheinlichkeit, daß $\sigma(z^{(n)})$ verschwindet, gegen null strebt, kann für kleine Stichproben dieser Fall durchaus einen größeren Einfluß besitzen. Es müssen zwei Situationen unterschieden werden:

1. $\sigma(z^{(n)}) = 0$, d. h. $\sigma(z^{(n)})$ verschwindet für die ursprüngliche Stichprobe.
2. $\sigma(z_*^{(n)}) = 0$, d. h. $\sigma(z_*^{(n)})$ verschwindet für eine Bootstrap-Stichprobe.

Der erste Fall läßt sich einfach behandeln. Zunächst stellt man fest, daß nur für $p \in \{h(\theta) : \theta \in \Theta\}$ die Ausdrücke $\sigma(p)$ und $\Delta(p, h(\hat{\theta}(p)))$ verschwinden. Tritt der Fall $\sigma(z^{(n)}) = 0$ bei der gegebenen Stichprobe ein, so wird die Hypothese $H_0 : \Delta(\theta_0) \leq d$ gestützt, also nicht verworfen. Ist $\sigma(z^{(n)}) > 0$, dann muß für gegebenes $z^{(n)}$ die Verteilung von

$$\sqrt{n}\frac{\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))}{\sigma(z_*^{(n)})} \tag{5.6}$$

bestimmt werden. Da $\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) > 0$ ist, setzt man für $\sigma(z_*^{(n)}) = 0$ einfach den Wert der Statistik auf eine negative Zahl mit sehr großem Betrag. Diese Art der Festlegung beeinflußt nicht die kritischen Werte bei kleinen Signifikanzniveaus.

Die Schritte zur Durchführung des Tests bei Verwendung der modifizierten Statistik können nun wie folgt angegeben werden:

Schritt 1 Getestet wird die Hypothese $H_0 : \Delta(\theta_0) \leq d, \quad d > 0$

Schritt 2 Aus den Beobachtungen $x_1, \ldots, x_n$ bilde man F_n, $z^{(n)}$ und $\sigma(z^{(n)})$.

Schritt 3 a) Ist $\sigma(z^{(n)}) = 0$, dann wird die Hypothese nicht verworfen.

Schritt 3 b) Ist $\sigma(z^{(n)}) > 0$, dann berechne man die Verteilung von (5.6) unter F_n, wobei der Fall $\sigma(z_*^{(n)}) = 0$ wie oben genannt behandelt wird.

Schritt 4 Man ermittle zu gegebenem Signifikanzniveau α die Werte $c_{n,L}(\alpha, F_n)$ und $c_{n,U}(\alpha, F_n)$, sowie einen Wert $c_n(\alpha, F_n)$ mit

$$c_{n,L}(\alpha, F_n) \leq c_n(\alpha, F_n) \leq c_{n,U}(\alpha, F_n)$$

Schritt 5 Die Hypothese wird verworfen, falls

$$\sqrt{n}\frac{\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - d}{\sigma(z^{(n)})} > c_n(\alpha, F_n)$$

ist.

Sind die Berechnungen zu Schritt 4 zu aufwendig, können Monte-Carlo-Methoden zur Bestimmung der kritischen Werte verwendet werden. In Abschnitt 5.3.2.1 wird die Eignung dieses Tests untersucht und ein Vergleich mit dem Test des vorhergehenden Abschnitts durchgeführt.

5.2.4 Ein Bootstrap-Anpassungstest

5.2.4.1 Ein „falscher“ Bootstrap-Test

Bei den bishergenannten Tests über $\Delta(\theta_0)$ mußte vorausgesetzt werden, daß das Grundmodell nicht in der approximierenden Modellfamilie enthalten ist. Es wird zunächst erläutert, weshalb sich die Vorgehensweise dieser Tests ohne Modifikation nicht auf den Fall korrekter Modellspezifikation übertragen läßt.

Auf den ersten Blick wäre es naheliegend, als Statistik

$$R_n(X_1, \ldots, X_n; F) = n[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))] \tag{5.7}$$

zu verwenden. Beschränkt man sich auf $\Delta_g \in \mathcal{D}^*$, so wäre diese Statistik unter der Nullhypothese asymptotisch χ^2_{k-s-1}-verteilt (siehe Satz 3.5). Bei Verwendung von Bootstrap-Methoden könnte der kritische Wert durch den $(1-\alpha)$-Prozentpunkt der Verteilung der Statistik

$$R_n(X_1^*, \ldots, X_n^*; F_n) = n[\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))] \tag{5.8}$$

geschätzt werden. Das Theorem 1 von Beran (1984) läßt sich jedoch nicht anwenden, da keine Menge $\mathcal{G}_F$ gefunden werden kann, so daß die Bedingungen $\mathbf{V1} - \mathbf{V3}$ erfüllt sind. Dieser Sachverhalt soll nun kurz angedeutet werden:

Sei $\mathcal{G}_F$ wieder wie in Definition (5.2) gebildet. Für jede Folge $(G_n) \in \mathcal{G}_F$ gilt für die zugehörige Folge $(p^{(n)})$ von Wahrscheinlichkeitsvektoren:

$$\lim_{n\to\infty} p^{(n)} = \pi$$

Sei

$$w(p^{(n)}) := \left(\frac{p_1^{(n)} - h_1(\hat{\theta}(p^{(n)}))}{\sqrt{h_1(\hat{\theta}(p^{(n)}))}}, \ldots, \frac{p_k^{(n)} - h_k(\hat{\theta}(p^{(n)}))}{\sqrt{h_k(\hat{\theta}(p^{(n)}))}} \right)^T,$$

und sei $z_*^{(n)}$ der Vektor der relativen Häufigkeiten zu einer unter $p^{(n)}$ entstandenen Stichprobe der Größe n. Es folgt nun unter Verwendung der multivariaten Deltamethode aus Satz 2.3

$$\sqrt{n}(w(z_*^{(n)}) - w(p^{(n)})) \xrightarrow{d} N(\mathbf{0}, Z), \tag{5.9}$$

wobei die idempotente Matrix Z wie im Beweis zu Satz 3.5 definiert ist. Aus (5.9) ergibt sich nun

$$n(w(z_*^{(n)}) - w(p^{(n)}))^T(w(z_*^{(n)}) - w(p^{(n)})) \xrightarrow{d} \chi^2_{k-s-1}, \tag{5.10}$$

eine asymptotische Aussage, die von dem gewünschten Ergebnis

$$n[\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(p^{(n)}, h(\hat{\theta}(p^{(n)})))] \xrightarrow{d} \chi^2_{k-s-1} \tag{5.11}$$

$$\Leftrightarrow \quad n(w(z_*^{(n)})^T w(z_*^{(n)}) - w(p^{(n)})^T w(p^{(n)})) \xrightarrow{d} \chi^2_{k-s-1} \tag{5.12}$$

abweicht. Eine Umformung von (5.12), in Verbindung mit (5.10), zeigt, daß Folgen $(p^{(n)})$ gesucht sind, die sowohl

$$n(w(z_*^{(n)}) - w(p^{(n)}))^T (w(z_*^{(n)}) - w(p^{(n)})) \xrightarrow{d} \chi^2_{k-s-1}$$

als auch

$$n(w(z_*^{(n)}) + w(p^{(n)}))^T (w(z_*^{(n)}) - w(p^{(n)})) \xrightarrow{d} \chi^2_{k-s-1}$$

erfüllen. Die Bedingung

$$\lim_{n\to\infty} n w(p^{(n)})^T w(p^{(n)}) = 0 \tag{5.13}$$

ist hinreichend, leider ist aber kein Beweis gelungen, daß sie auch notwendig ist. Zumindest läßt sich aus den Ergebnissen des Abschnitts 3.4.3 herleiten, daß

$$\lim_{n\to\infty} p^{(n)} = \pi \tag{5.14}$$

nicht hinreichend ist, da für Folgen der Form

$$p^{(n)} = \pi + \frac{c}{\sqrt{n}}, \quad c \in I\!R^k, \ \sum_{i=1}^{k} c_i = 0,$$

für

$$n\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)})))$$

im allgemeinen eine nichtzentrale χ^2-Verteilung als Grenzverteilung existiert und somit (5.11) nicht erfüllt sein kann. Somit kann die Menge $\mathcal{G}_F$ nicht wie bisher definiert werden, und es ist fraglich, ob die Folgen (F_n) auch einer eingeschränkten Menge mit Wahrscheinlichkeit eins angehören. Würde sich erweisen, daß (5.13) nicht nur hinreichend, sondern auch notwendig ist, könnte gefolgert werden, daß die Menge der Folgen $(z^{(n)})$, für die die bedingte Verteilung von

$$n[\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))],$$

gegeben $z^{(n)}$, gegen die χ^2-Verteilung mit $k-s-1$ Freiheitsgraden strebt, *nicht* die Wahrscheinlichkeit 1 besitzt, da dies ansonsten im Widerspruch zu

$$n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) \xrightarrow{d} \chi^2_{k-s-1}$$

stehen würde.

Es sei nun ein weiterer Versuch unternommen, das Versagen dieses Bootstrap-Tests zu begründen. Sei $R_n := R_n(X_1, \ldots, X_n; F)$ wie in (5.7) definiert und $c_n(\alpha, F_n)$ ein Bootstrap-Schätzer des $(1-\alpha)$-Quantils. Es lassen sich nun, mit den weiter unten genannten Einschränkungen, neue Erkenntnisse über

$$\lim_{n\to\infty} P_F(R_n > c_n(\alpha, F_n))$$

herleiten. Dazu wird das folgende Lemma benötigt:

Lemma 5.4 *Sei die Funktion* $h : I\!N \times [0, \infty) \times I\!R \to [0; 1]$ *definiert durch*

$$h(n, c, d) = P(\chi^2_n(c) \leq 2c - d).$$

Dann ist $h(n, c, 0)$ *für festes, aber beliebiges* $n \in I\!N$ *streng monoton steigend in* c *mit* $h(n, 0, 0) = 0$ *und* $\lim_{c\to\infty} h(n, c, 0) = 1$. *Weiterhin gibt es zu* α *mit* $0 < \alpha < 1$ *genau ein* $c^* := c^*(n, \alpha)$ *mit* $h(n, c^*, 0) = 1 - \alpha$.

Beweis: Im Anhang A.3. □

Da einerseits die korrekte mathematische Ableitung des folgenden Ergebnisses aufgrund der auftretenden singulären Normalverteilungen und der fehlenden asymptotischen Verteilungen fast unüberwindbare Schwierigkeiten bereitet, andererseits das Resultat sehr gut die Simulationsergebnisse beschreibt, wird nur eine Beweisidee gegeben und die Aussage nur als Vermutung formuliert.

Vermutung *Sei*

$$R_n := R_n(X_1, \ldots, X_n; F) = n[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))]$$

und für $\alpha \in (0;1)$ sei $c_n(\alpha, F_n)$ ein Bootstrap-Schätzer des $(1-\alpha)$-Quantils. Unter der Nullhypothese $H_0 : \Delta(\theta_0) = 0$, $\Delta_g \in \mathcal{D}^*$, gilt dann

$$\lim_{n\to\infty} P_F(R_n \leq c_n(\alpha, F_n)) = P(\chi^2_{k-s-1} \leq c_\alpha),$$

wobei c_α durch $P(\chi^2_{k-s-1}(c_\alpha) \leq 2c_\alpha) = 1-\alpha$ definiert ist.

Beweisidee:

1. Die Beweisidee basiert auf der Vermutung, daß für genügend großes n sowohl die Verteilung von

$$n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))$$

durch eine χ^2_{k-s-1}-Verteilung als auch die bedingte Verteilung von

$$\sqrt{n}(w(z_*^{(n)}) - w(z^{(n)})),$$

gegeben $z^{(n)}$, $z^{(n)}$ aber ansonsten beliebig, gemäß (5.9) durch eine $N(\mathbf{0}, Z)$-Verteilung ersetzt werden kann. An dieser Stelle wird eine Art gleichmäßiger Konvergenz vorausgesetzt, deren Nachweis leider nicht gelungen ist. Zunächst stellt man fest, daß

$$P_F(R_n \leq c_n(\alpha, F_n)) = P_F(2n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) \leq d_n(\alpha, F_n))$$

gilt, wobei

$$d_n(\alpha, F_n) := c_n(\alpha, F_n) + n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))$$

den Schätzer des $1-\alpha$-Quantils der bedingten Verteilung von $n\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)})))$, gegeben $z^{(n)}$, bezeichnen möge. Führt man die oben genannten Ersetzungen durch, so folgt:

$$\begin{aligned}
\sqrt{n}(w(z_*^{(n)}) - w(z^{(n)})) \mid z^{(n)} &\sim N(\mathbf{0}, Z) \\
\Rightarrow \sqrt{n}w(z_*^{(n)}) \mid z^{(n)} &\sim N(\sqrt{n}w(z^{(n)}), Z) \\
\Rightarrow n\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) \mid z^{(n)} &\sim \chi^2_{k-p-1}(n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))))
\end{aligned}$$

2. Aus Lemma 5.4 folgt die Existenz eines c_α mit

$$P(\chi^2_{k-s-1}(c_\alpha) \leq 2c_\alpha) = 1-\alpha.$$

Für $c > c_\alpha$ ist $P(\chi^2_{k-s-1}(c) \leq 2c) > 1-\alpha$, d. h. für ein $c^* < c$ ist $P(\chi^2_{k-s-1}(c) \leq 2c^*) = 1-\alpha$. Für die gegebene Situation bedeutet dies,

daß aus $n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) > c_\alpha$ folgt, daß $2n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))$ größer ist als der $1-\alpha$-Prozentpunkt einer $\chi^2_{k-s-1}(n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))))$-verteilten Zufallsvariablen. Die analoge Aussage gilt natürlich auch, falls „<“ durch „>“ und „größer“ durch „kleiner“ ersetzt wird.
3. Aus 1. und 2. folgt:

$$\lim_{n\to\infty} P_F(R_n \leq c_n(\alpha, F_n)) = P(\chi^2_{k-s-1} \leq c_\alpha)$$

□

Auch wenn der lückenlose Nachweis dieser Behauptung nicht gelungen ist, so scheinen die Simulationsergebnisse in Abschnitt 5.3.3 die Vermutung zu bestätigen.

5.2.4.2 Modifikation zur Herleitung eines korrekten Tests

Um nun trotzdem einen Bootstrap-Anpassungstest durchführen zu können, muß eine Modifikation vorgenommen werden. Gesucht ist eine Statistik, die unter der Nullhypothese mit der χ^2-Prüfgröße zusammenfällt und die außerdem die Voraussetzungen **V1-V3** des Satzes von Beran erfüllt.

Die Diskrepanz, die zur Ermittlung der Schätzer der Parameter verwendet wird, darf nun wieder eine beliebige aus der Menge $\mathcal{D}$ sein. Sei zunächst π ein beliebiges Grundmodell, das nicht notwendigerweise in der approximierenden Modellfamilie enthalten sein muß.

Für Folgen $(G_n) \in \mathcal{G}_F$ kann unter Benutzung von Satz 3.4 und Satz 2.3 gezeigt werden, daß

$$n(w(z_*^{(n)}) - w(p^{(n)}))^T(w(z_*^{(n)}) - w(p^{(n)})) \xrightarrow{d} \sum_{\lambda_i>0} \lambda_i \chi_1^2 \tag{5.15}$$

gilt, wobei die λ_i, $i = 1, \ldots, k$, die Eigenwerte der Matrix

$$Z := W(D_\pi - \pi\pi^T)W^T$$

bezeichnen, die nicht von der speziellen Folge (G_n) abhängen. Die Matrix W ist (vgl. Satz 3.4) durch

$$W := D_h^{-\frac{1}{2}}[I + \frac{1}{2}(I + D_\pi D_h^{-1})C^T\Omega^{-1}CG]$$

definiert. Setzt man nun

$$R_n(X_1,\ldots,X_n;F) := n\sum_{i=1}^{k}\left(\frac{z_i^{(n)} - h_i(\hat\theta(z^{(n)}))}{\sqrt{h_i(\hat\theta(z^{(n)}))}} - \frac{\pi_i - h_i(\theta_0)}{\sqrt{h_i(\theta_0)}}\right)^2,$$

so sind die Voraussetzungen **V1-V3** von Beran erfüllt. Die Verteilung von $R_n(X_1,\ldots,X_n;F)$ kann durch die bedingte Verteilung von

$$R_n(X_1^*,\ldots,X_n^*;F_n) := n\sum_{i=1}^{k}\left(\frac{(z_*^{(n)})_i - h_i(\hat\theta(z_*^{(n)}))}{\sqrt{h_i(\hat\theta(z_*^{(n)}))}} - \frac{z_i^{(n)} - h_i(\hat\theta(z^{(n)}))}{\sqrt{h_i(\hat\theta(z^{(n)}))}}\right)^2,$$

gegeben $z^{(n)}$, vernünftig geschätzt werden.

Ist nun $c_n(\alpha,F_n)$ ein Schätzer des $(1-\alpha)$-Quantils der Verteilung von $R_n(X_1^*,\ldots,X_n^*;F_n)$, nach den Regeln des Abschnitts 5.1 gebildet, so folgt nun das Lemma:

Lemma 5.5 *Ist $\pi = h(\theta_0)$, so gilt:*

$$\lim_{n\to\infty} P_F(n\Delta(z^{(n)}, h(\hat\theta(z^{(n)}))) > c_n(\alpha,F_n)) = \alpha$$

Ist die Hypothese falsch, so konvergiert für fast alle Folgen $(z^{(n)})$ die bedingte Verteilung von $R_n(X_1^*,\ldots,X_n^*;F_n)$ gemäß (5.15) gegen eine Linearkombination

$$\sum_{\lambda_i>0}\lambda_i\chi_1^2.$$

Wichtig ist nur, daß somit auch die Folge der $c_n(\alpha,F_n)$ fast sicher gegen einen konstanten Wert konvergiert. Somit folgt unter einer festen Alternative

$$\lim_{n\to\infty} P_F(n\Delta(z^{(n)}, h(\hat\theta(z^{(n)}))) > c_n(\alpha,F_n)) = 1,$$

d. h. der in dieser Weise konstruierte Anpassungstest ist konsistent.

Für $\Delta_g \in \mathcal{D}^*$ ist die Prüfgröße unter der Hypothese asymptotisch χ^2-verteilt, in diesem Fall könnte man natürlich die entsprechenden Prozentpunkte der asymptotischen Verteilung verwenden. Interessant ist es, in diesem Zusammenhang zu überprüfen, wie sich der Bootstrap-Anpassungstest bei kleinen erwarteten Häufigkeiten verhält, also in einer Situation, in der man den klassischen χ^2-Anpassungstest nicht verwenden würde. Der

Bootstrap-Test besitzt außerdem den Vorteil, daß auch Schätzer zu Diskrepanzen $\Delta_g \in \mathcal{D} \setminus \mathcal{D}^*$ eingesetzt werden dürfen. Hier wäre es im klassischen Fall gar nicht möglich, kritische Werte anzugeben, da die asymptotische Verteilung unter der Hypothese vom unbekannten, wahren Modell abhängt.

Die Idee dieses Bootstrap-Tests kann auf andere Diskrepanzen erweitert werden, vorzugsweise auf solche, bei denen die Verteilung unter der Hypothese vom wahren Modell abhängt. Nähere Erläuterungen hierzu werden in Kapitel 6.2.3 gegeben.

5.3 Simulationen

5.3.1 Probleme bei der Durchführung der Simulationen

Es werden nun einige Simulationen durchgeführt, um die Leistungsfähigkeit der genannten Bootstrap-Tests zu überprüfen. Bevor die eigentlichen Ergebnisse genannt werden, ist es angebracht, einige Faktoren zu nennen, die die Resultate zumindest für kleinere Stichproben stark beeinflussen können. Hauptverantwortlich für die entstehenden Schwierigkeiten ist der diskrete Charakter der vorliegenden Verteilungen. Es sei zunächst angenommen, daß die jeweilige Hypothese erfüllt ist.

Für die Bestimmung der kritischen Werte benötigt man die Verteilung der Statistik

$$R_n(X_1^*, \ldots, X_n^*; F_n) \quad .$$

Ist diese Verteilung berechenbar, so ist ein gültiger Bootstrap-Schätzer des $1-\alpha$-Quantils gemäß (5.2) durch jeden Wert $c_n(\alpha, F_n)$ mit

$$c_{n,L}(\alpha, F_n) \leq c_n(\alpha, F_n) \leq c_{n,U}(\alpha, F_n)$$

gegeben. Da die Verteilungsfunktion eine rechtsstetige Treppenfunktion ist, ergibt sich:

$$\begin{aligned} c_{n,L}(\alpha, F_n) &= \inf\{x : J_n(x, F_n) \geq 1-\alpha\} = \min\{x : J_n(x, F_n) \geq 1-\alpha\} \\ c_{n,U}(\alpha, F_n) &= \sup\{x : J_n(x, F_n) \leq 1-\alpha\} = \min\{x : J_n(x, F_n) > 1-\alpha\} \end{aligned}$$

$c_{n,L}(\alpha, F_n)$ und $c_{n,U}(\alpha, F_n)$ sind nur dann verschieden, falls ein x mit $J_n(x, F_n) = 1-\alpha$ existiert. Im allgemeinen wird deshalb $c_n(\alpha, F_n)$ eindeutig festgelegt sein, und das Problem der Wahl von $c_n(\alpha, F_n)$ besteht dann nicht.

Nun ist es nur selten praktikabel, die exakte Verteilungsfunktion zu bestimmen. Vielmehr wird man versuchen, den $1-\alpha$-Prozentpunkt mit Monte-Carlo-Methoden zu ermitteln. Es werden B Stichproben $x_{b1}^*, \ldots, x_{bn}^*$, $b = 1, \ldots, B$, unter F_n erzeugt und die Werte $R_n(x_{11}^*, \ldots, x_{1n}^*; F_n), \ldots, R_n(x_{B1}^*, \ldots, x_{Bn}^*; F_n)$ betrachtet (vgl. Abschnitt 5.1). Seien

$$R^1 \leq R^2 \leq \ldots \leq R^B$$

die geordneten Werte, sei weiterhin $[x]$ wieder definiert als „größte ganze Zahl $\leq x$“, dann können die Monte-Carlo-Quantils-Schätzer mit

$$\begin{aligned} c_{n,L,B}(\alpha, F_n) &= R^{-[-(1-\alpha)B]} \quad \text{bzw.} \\ c_{n,U,B}(\alpha, F_n) &= R^{[(1-\alpha)B]+1} \end{aligned}$$

angegeben werden. Ist beispielsweise $\alpha = 0.10$ und $B = 200$, so erhält man

$$\begin{aligned} c_{n,L,B}(\alpha, F_n) &= R^{180} \quad \text{und} \\ c_{n,U,B}(\alpha, F_n) &= R^{181} \quad . \end{aligned}$$

Hat man sich nun z. B. für $c_{n,L,B}(\alpha, F_n)$ als Schätzer des Quantils entschieden, so würde derselbe Schätzer (in dem genannten Fall also R^{180}) auch für jedes andere Signifikanzniveau $\alpha \in [0.10, 0.105)$ verwendet werden. Diese Ungenauigkeit wirkt sich besonders störend bei kleinen Signifikanzniveaus aus. Soll ein Bootstrap-Test für $\alpha = 0.001$ durchgeführt werden, so reichen 200 Wiederholungen sicherlich nicht aus, da die Durchführung des Tests unabhängig von der Größe der Originalstichprobe $x_1, \ldots, x_n$ für jedes α aus dem Intervall $(0, 0.005)$ identisch wäre.

Während bei der praktischen Durchführung der Tests nur *eine* Stichprobe vorliegt, und das Problem nur darin besteht, einen vernünftigen Wert für B anzusetzen, führt die Überprüfung, wie genau das Signifikanzniveau eingehalten wird, zu weiteren Komplikationen. Soll die Wahrscheinlichkeit

$$P_F(R_n(X_1, \ldots, X_n; F_n) > c_n(\alpha, F_n))$$

unter der Hypothese geschätzt werden, müssen unter dem Grundmodell zunächst wiederholt Stichproben der Größe n erzeugt werden. Für jede Stichprobe werden dann B Bootstrap-Stichproben ermittelt, aus ihnen ein kritischer Wert gewonnen, und anschließend wird überprüft, ob die Hypothese verworfen wird.

Der Anteil der Fälle, in denen die Hypothese abgelehnt wurde, stellt dann den Schätzer des Signifikanzniveaus dar. Die Überprüfung der Eignung der Bootstrap-Tests unter Verwendung von Monte-Carlo-Methoden wird demnach durch den Einfluß des Wertes von B bzw. der Anzahl der Stichproben unter dem Grundmodell entschieden erschwert.

5.3.2 Simulationen zur Hypothese $H_0 : \Delta(\theta_0) \leq d$

5.3.2.1 Durchführung der Simulationen und Ergebnisse

Wie bereits im letzten Unterabschnitt angedeutet, sind Simulationen zur Überprüfung der Einhaltung des Signifikanzniveaus sehr zeitaufwendig. Es darf deshalb nicht verwundern, daß Resultate nur für eine, willkürlich ausgewählte Modellfamilie gegeben werden. Die betrachtete Modellfamilie wurde schon in Abschnitt 3.5.4 verwendet, der Vektor $h(\theta)$ ist durch

$$h(\theta) = (\theta, \theta, \theta, \theta, \theta, 1 - 5\theta)^T, \quad \theta \in [0, 0.2],$$

gegeben.

Es wurden Minimum-χ^2-Schätzer verwendet, die Diskrepanz Δ_g ist also identisch mit der χ^2-Diskrepanz Δ. Für den willkürlich gewählten vorläufigen Vektor

$$\pi = (0.10, 0.13, 0.15, 0.19, 0.23, 0.2)^T$$

ergibt sich θ_0 zu

$$\theta_0 = 0.161235 \quad .$$

Anschließend wurden die endgültig betrachteten Alternativen so gewählt, daß sie auf der Verbindungsgeraden

$$\pi_t = h(\theta_0) + t(\pi - h(\theta_0)), \quad t \in I\!R,$$

zwischen π und $h(\theta_0)$ liegen. Der Parameter t nimmt dabei solche Werte an, daß für die Diskrepanz durch Approximation $\Delta(\pi_t, h(\hat{\theta}(\pi_t)))$ die folgenden Werte entstehen:

$$i * 0.002, \quad i = 0, \ldots, 5,$$

und

$$0.012 + i * 0.008, \quad i = 0, \ldots, 7.$$

Um ein Gefühl für die Lage der Alternativen zu bekommen, sind in der Tabelle 5.1 für $n = 100$ bzw. $n = 200$ zusätzlich Schätzer der Macht bei

Durchführung des χ^2-Anpassungstests zum Signifikanzniveau $\alpha = 0.10$ gegeben, ermittelt mit der in Abschnitt 3.5.1 erwähnten Funktion $F_{KON}(t)$.

$\Delta(\theta_0)$ / n	0.000	0.002	0.004	0.006	0.008	0.010	0.012
100	0.100	0.115	0.131	0.148	0.164	0.181	0.197
200	0.100	0.132	0.165	0.198	0.233	0.268	0.303

$\Delta(\theta_0)$ / n	0.020	0.028	0.036	0.044	0.052	0.060	0.068
100	0.266	0.335	0.403	0.468	0.531	0.589	0.643
200	0.439	0.563	0.670	0.757	0.825	0.877	0.916

Tab. 5.1: Schätzer der Macht bei Durchführung des χ^2-Anpassungstests zum Signifikanzniveau $\alpha = 0.10$.

Ist $t = 0$, so dürften die Tests aus den Abschnitten 5.2.2 und 5.2.3 eigentlich nicht angewendet werden. Dieser Fall ist aber an dieser Stelle mitaufgeführt, um das Versagen der Bootstrap-Tests demonstrieren zu können.

Für jede der genannten 14 Alternativen π_t wurden dann zu der jeweils gegebenen Stichprobengröße 10000 Stichproben erzeugt und jedes Mal der Wert von $z^{(n)}$ berechnet. Für jede dieser Stichproben wurden 200 Bootstrap-Stichproben generiert. Anschließend wurden dann simultan für die theoretischen Signifikanzniveaus $\alpha = 0.05, 0.10, 0.15, \ldots, 0.45, 0.50$ mit Hilfe dieser Bootstrapstichproben die entsprechenden $1-\alpha$-Prozentpunkte der Verteilung der Statistik

$$R_n(X_1^*, \ldots, X_n^*; F_n) = \sqrt{n}[\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))]$$

gemäß des Vorgehens des Testverfahrens aus Abschnitt 5.2.2, hier kurz Test 1 genannt, für gegebenes $z^{(n)}$ mittels der Werte

$$c_{n,L,200}(\alpha, F_n), \quad \alpha = 0.05, 0.10, \ldots, 0.50,$$

geschätzt (siehe Abschnitt 5.3.1, S. 149). Die von der jeweiligen Alternative abhängige Hypothese

$$H_0 : \Delta(\theta_0) \leq d,$$

mit $d = \Delta(\pi_t, \hat{\theta}(\pi_t))$, wurde zum Signifikanzniveau α verworfen, falls

$$\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) > n^{-\frac{1}{2}} c_{n,L,200}(\alpha, F_n) + d$$

war. Danach wurde festgehalten, wie oft für festes Signifikanzniveau und feste Alternative die Hypothese verworfen wurde. Diese Anzahl, dividiert durch 10000, ergibt dann einen Schätzer des tatsächlichen Signifikanzniveaus. In den Graphiken 5.1 und 5.2 sind diese Schätzer für alle Alternativen zur Stichprobengröße $n = 100$ bzw. $n = 200$ eingetragen. Dabei sind Schätzer zum gleichen Signifikanzniveau, aber verschiedenen Alternativen, verbunden, um eine übersichtlichere Darstellung zu erreichen.

Mit denselben erzeugten Stichproben wurden analoge Schätzer für das Testverfahren aus Abschnitt 5.2.3 (Test 2) gebildet. Natürlich wurde nun die Statistik

$$R_n(X_1^*, \ldots, X_n^*; F_n) = \sqrt{n} \frac{\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))}{\sigma(z_*^{(n)})}$$

betrachtet und die Quantilsschätzer $c_{n,L,200}(\alpha, F_n)$ neu berechnet. Die erhaltenen Schätzer der tatsächlichen Signifikanzniveaus sind in den Abbildungen 5.3 und 5.4 wiedergegeben.

Diesen beiden Bootstrap-Tests wurde weiterhin ein asymptotischer Test (Test 3) gegenübergestellt. Mit dem Ergebnis aus Lemma 3.4 läßt sich schnell zeigen, daß

$$\sqrt{n} \frac{\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))}{\sigma(z^{(n)})} \xrightarrow{d} N(0, 1)$$

gilt. Ein Test zur Hypothese $H_0 : \Delta(\theta_0) \leq d$ und Signifikanzniveau α kann nun durchgeführt werden, indem der $1 - \alpha$-Prozentpunkt der Standard-Normalverteilung als kritischer Wert bestimmt und die Hypothese verworfen wird, falls

$$\sqrt{n} \frac{\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - d}{\sigma(z^{(n)})}$$

größer als dieser kritische Wert ist. Der Schätzer für das tatsächliche Signifikanzniveau α wurde wieder aus dem Anteil der 10000 Stichproben berechnet,

bei denen die Hypothese verworfen wurde. Die Schätzer der tatsächlichen Signifikanzniveaus sind in den Abbildungen 5.5 und 5.6 gegeben. Die Abbildungen 5.7 und 5.8 zeigen zusätzlich die erhaltenen Schätzer für die sehr großen Stichprobengrößen von $n = 2000$ und $n = 10000$.

5.3.2.2 Interpretation der Simulationsergebnisse

Betrachtet man die Abbildungen, so ist sofort ersichtlich, daß sowohl die Bootstrap-Tests als auch der asymptotische Test für größere Diskrepanzen in der Hinsicht besser funktionieren, daß das gewünschte Signifikanzniveau genauer eingehalten wird. Entscheidende Hinweise über die Schnelligkeit der Konvergenz der Schätzer des tatsächlichen Signifikanzniveaus gegen das gewünschte theoretische Signifikanzniveau kann die Analyse der Testverfahren für den Fall $\Delta(\theta_0) = 0$ geben. Es sei noch einmal betont, daß dann die Durchführung der Tests unzulässig ist, da die Schätzer im allgemeinen selbst für $n \to \infty$ nicht gegen die korrekten Werte konvergieren.

Beispielsweise wurden in Abschnitt 5.2.4.1 mit den dort genannten Einschränkungen in der Vermutung für Test 1 die Grenzwerte der tatsächlichen Signifikanzniveaus hergeleitet. Unterscheiden sich diese Grenzwerte nur wenig vom theoretischen Signifikanzniveau, kann davon ausgegangen werden, daß dieser Bootstrap-Test auch in einer Umgebung $\Delta(\theta_0) \leq \epsilon$, $\epsilon > 0$, gute Annäherungen an das gewünschte Signifikanzniveau ergibt. Andererseits sind schlechte Ergebnisse zu erwarten, falls dieser Grenzwert stark vom theoretischen Signifikanzniveau abweicht. Zum Beispiel ist der genannte Grenzwert zum Signifikanzniveau $\alpha = 0.05$ für das gewählte Beispiel etwa 0.00026, und aus Abbildung 5.1 ist ersichtlich, daß das tatsächliche Signifikanzniveau selbst beim Test der Hypothese $H_0 : \Delta(\theta_0) \leq 0.068$ erst in der Größenordnung von 0.01 liegt. Auch die Verdoppelung der Stichprobengröße auf $n = 200$ liefert nur einen Schätzer, der 0.02 nicht übertrifft. Für das Signifikanzniveau $\alpha = 0.45$ hingegen liegt der Grenzwert unter der Hypothese $H_0 : \Delta(\theta_0) = 0$ nahe bei 0.449, und das theoretische Signifikanzniveau wird dann über dem gesamten betrachteten Bereich der Werte von $\Delta(\theta_0)$ gut eingehalten.

Für die Tests 2 und 3 konnten zwar die Grenzwerte der Signifikanzniveaus zur Hypothese $H_0 : \Delta(\theta_0) = 0$ nicht berechnet werden, aber die eben genannten Ausführungen lassen sich übertragen. Liegt der Grenzwert dicht am theoretischen Signifikanzniveau, so ist eine gute Annäherung auch für Tests $H_0 : \Delta(\theta_0) \leq \epsilon$ für kleine $\epsilon > 0$ unter der Hypothese zu erwarten. Dies ist zum Beispiel bei Test 2 für $\alpha = 0.10$ der Fall.

Eine Überlegenheit des Tests 2 über den Test 1 aufgrund der Ergebnisse dieses einen Beispieles abzuleiten, erscheint zweifelhaft. Zwar werden bei Test 2 die üblicherweise verwendeten kleinen Werte von Signifikanzniveaus genauer eingehalten, aber für größere Werte von α liefert Test 1 die befriedigenderen Ergebnisse. Sollte es möglich sein, auch für Test 2 die Grenzwerte zu $H_0 : \Delta(\theta_0) = 0$ zu ermitteln, so sollte man das Testverfahren vorziehen, bei dem der Grenzwert näher am theoretischen Signifikanzniveau liegt.

Trotzdem scheinen beide Bootstrap-Tests, zumindest bei diesem Beispiel, dem asymptotischen Test 3 überlegen zu sein. Die Konvergenz bei Test 3 ist bei wachsender Stichprobengröße sehr langsam.

Ein Manko, das für alle drei Testverfahren in unterschiedlicher Ausprägung vorhanden ist, kann selbst durch sehr große Stichproben nicht ausgeräumt werden. Für jede noch so umfangreiche Stichprobe existieren Alternativen in der Nähe der approximierenden Modellfamilie, bei denen die Tests zu stark von den theoretischen Signifikanzniveaus abweichenden Werten führen können. Der Bruch im asymptotischen Verhalten aller betrachteten Statistiken, abhängig davon, ob die Alternative zur approximierenden Modellfamilie gehört oder nicht, führt zwangsläufig zur Existenz einer Umgebung, in der die Anwendung der Tests nicht ratsam ist. Leider läßt sich nicht a priori sagen, welche Stichprobengröße vorliegen sollte, um einen bestimmten Test $\Delta(\theta_0) \leq d$ durchzuführen.

Wünschenswert wäre es, falls dies überhaupt möglich ist, Statistiken $R(X_1, \ldots, X_n; F)$ zu finden, deren asymptotische Verteilungen *stetig* von der Alternative π abhängen, auch im Problembereich am Rand der approximierenden Modellfamilie. Diese Statistiken müßten trotzdem die Möglichkeit bieten, Tests über $\Delta(\theta_0)$ durchzuführen.

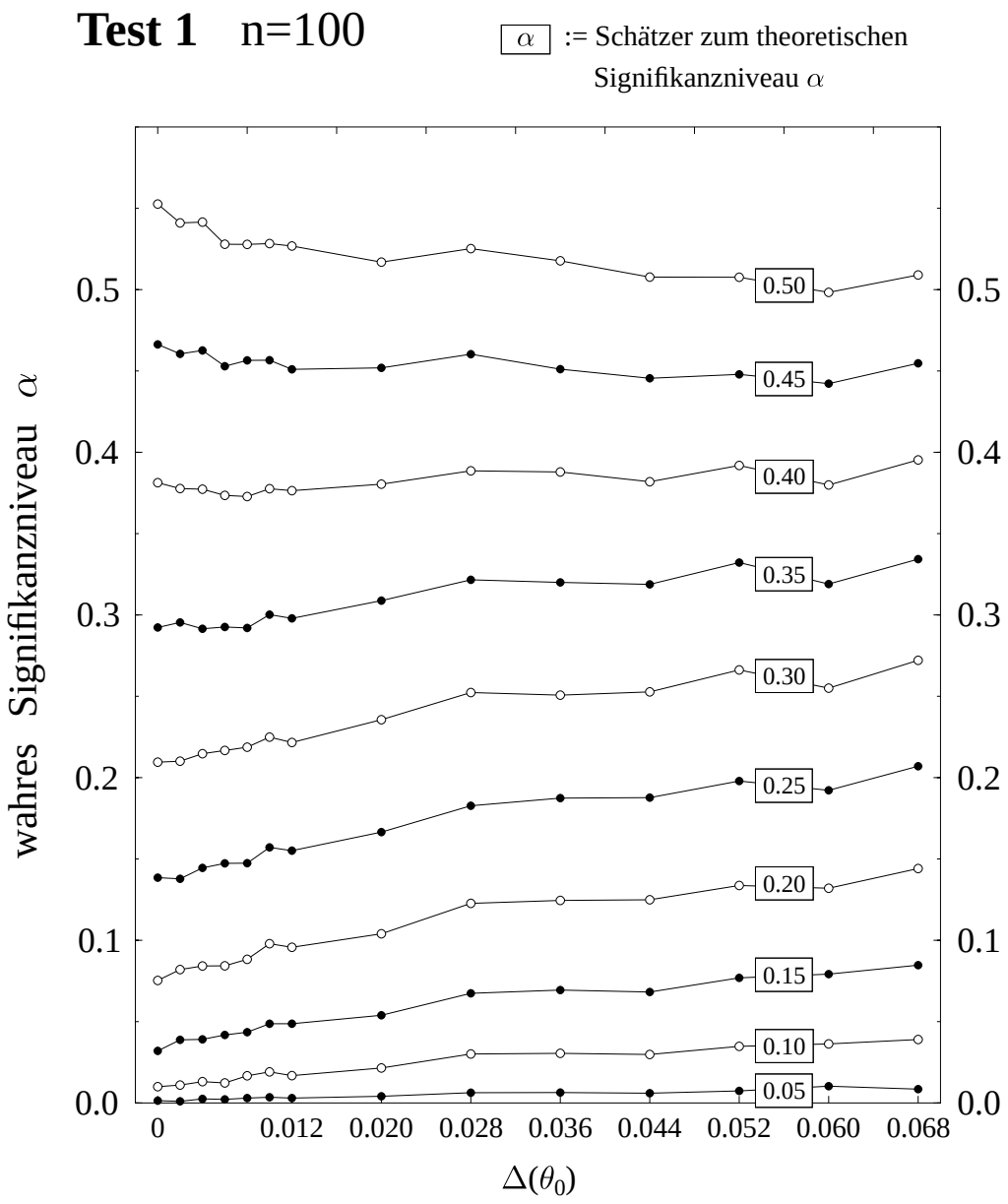


Abb. 5.1: Schätzer der wahren Signifikanzniveaus - Test 1 - n=100

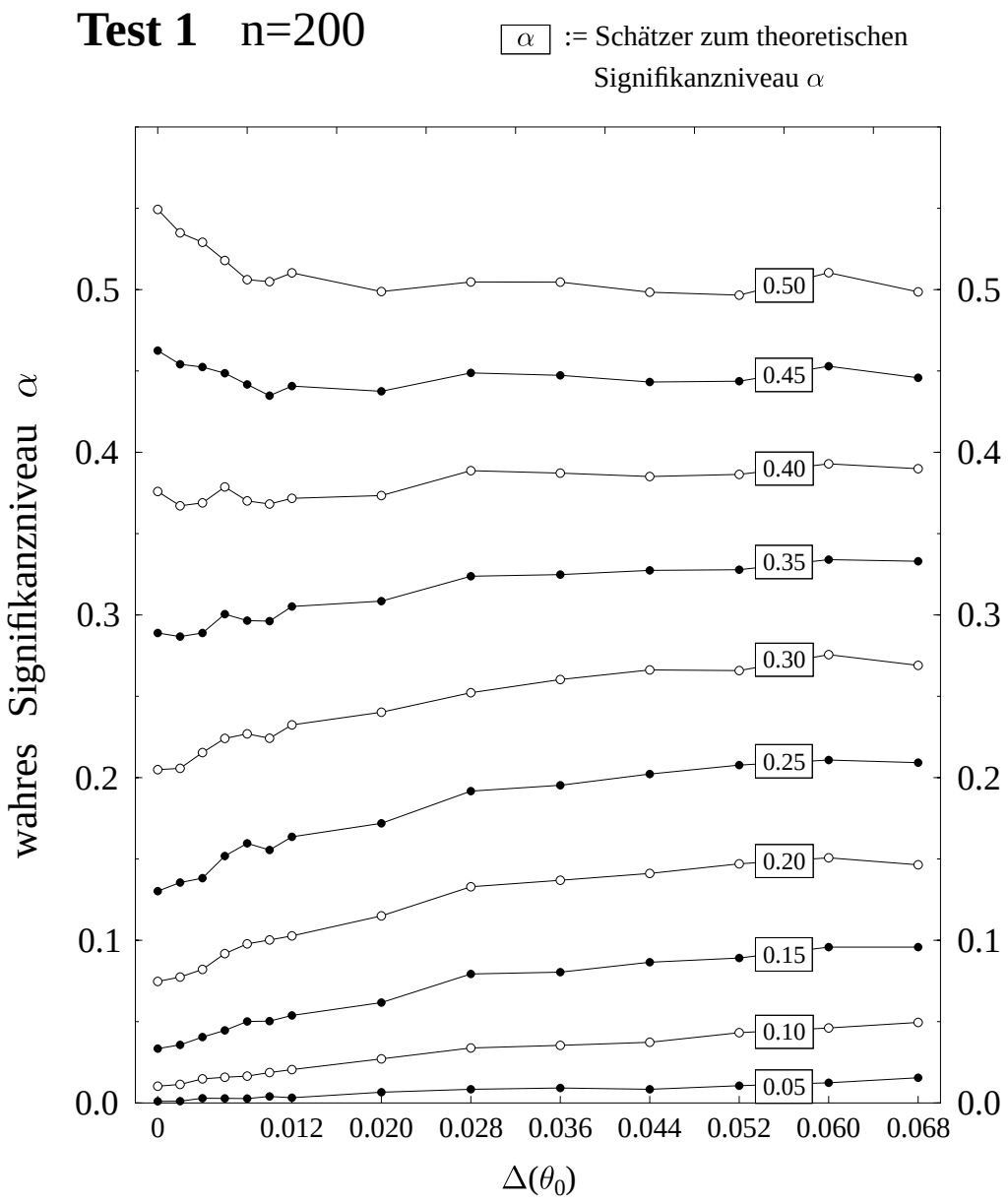


Abb. 5.2: Schätzer der wahren Signifikanzniveaus - Test 1 - n=200

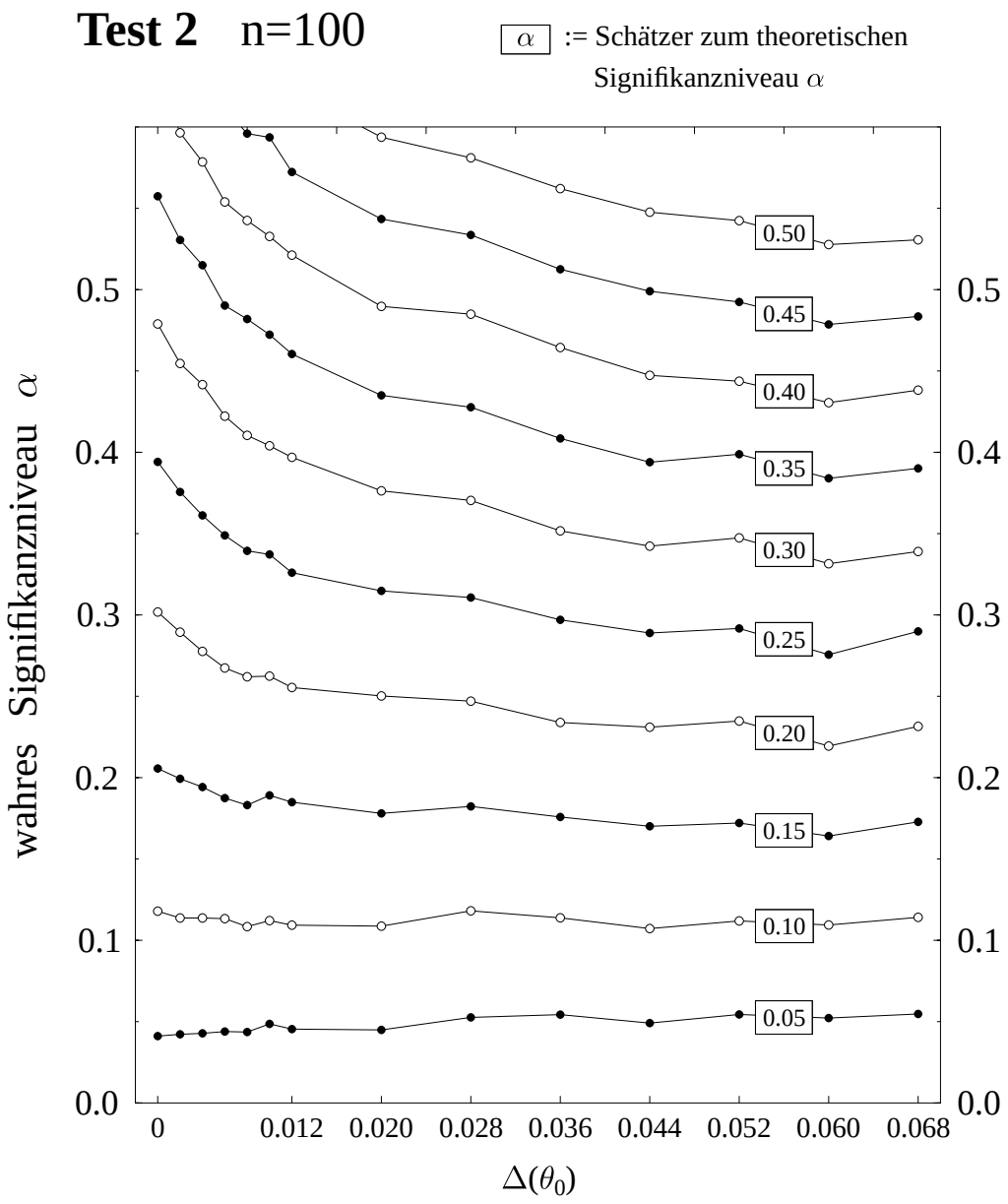


Abb. 5.3: Schätzer der wahren Signifikanzniveaus - Test 2 - n=100

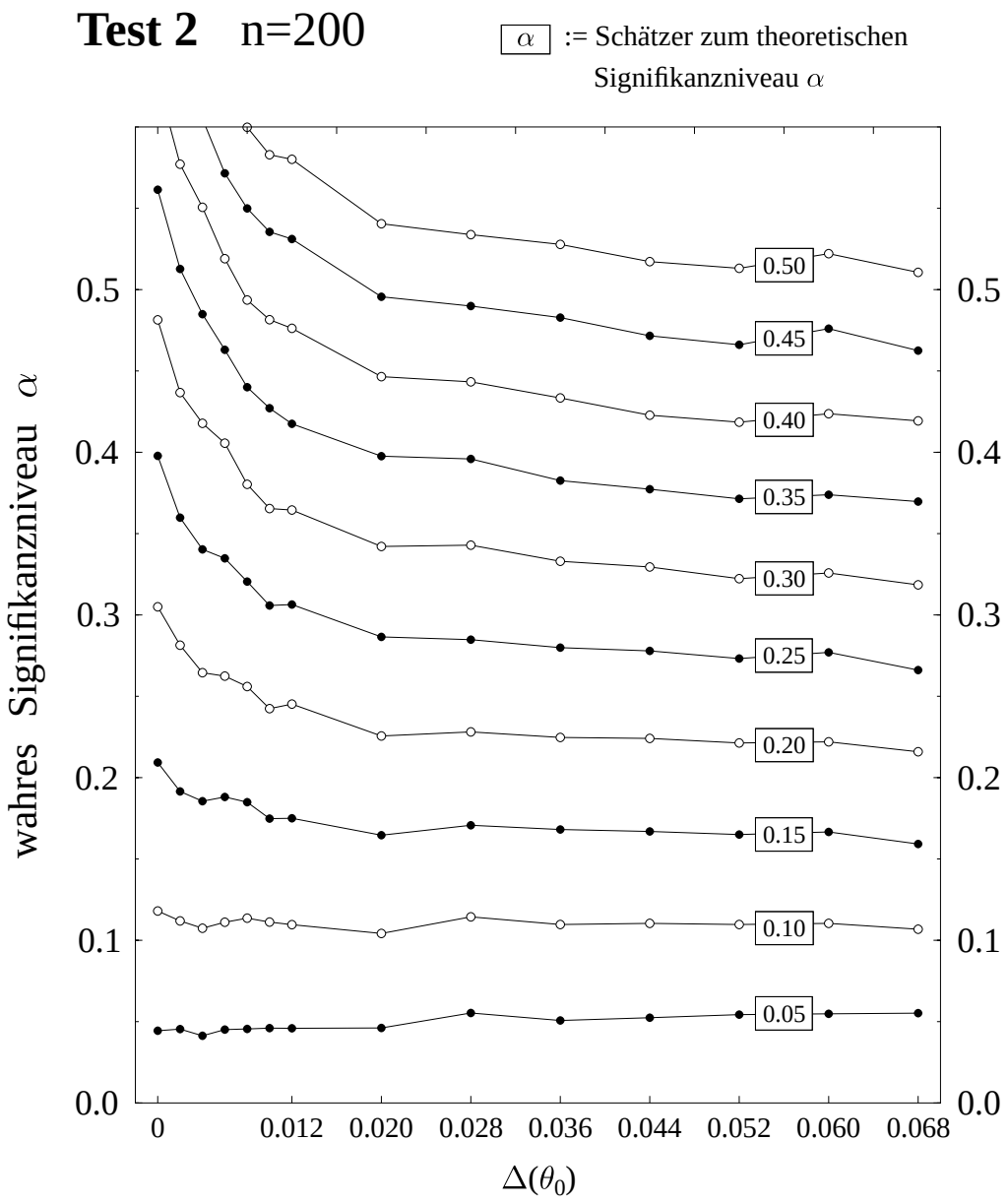


Abb. 5.4: Schätzer der wahren Signifikanzniveaus - Test 2 - n=100

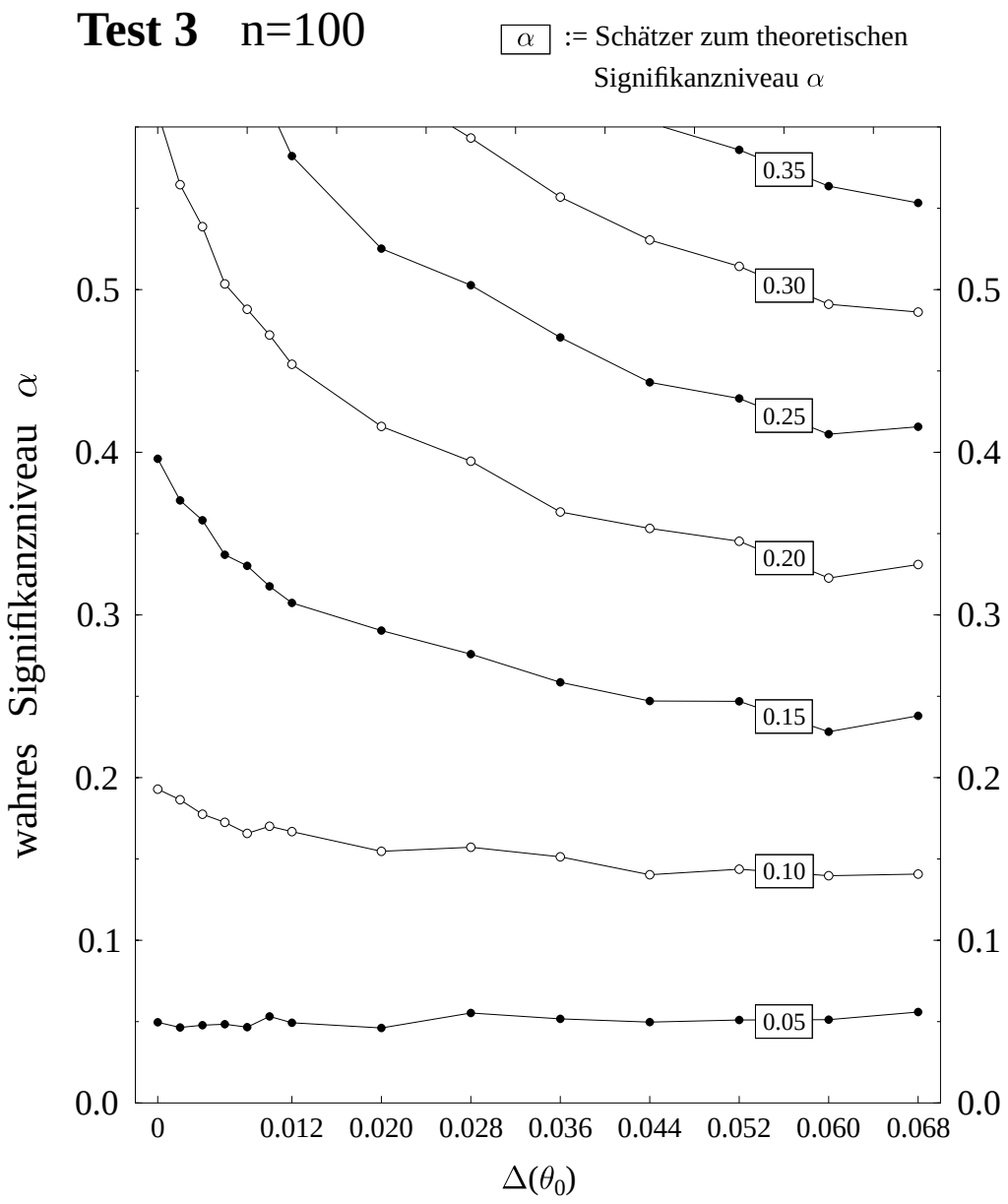


Abb. 5.5: Schätzer der wahren Signifikanzniveaus - Test 3 - n=100

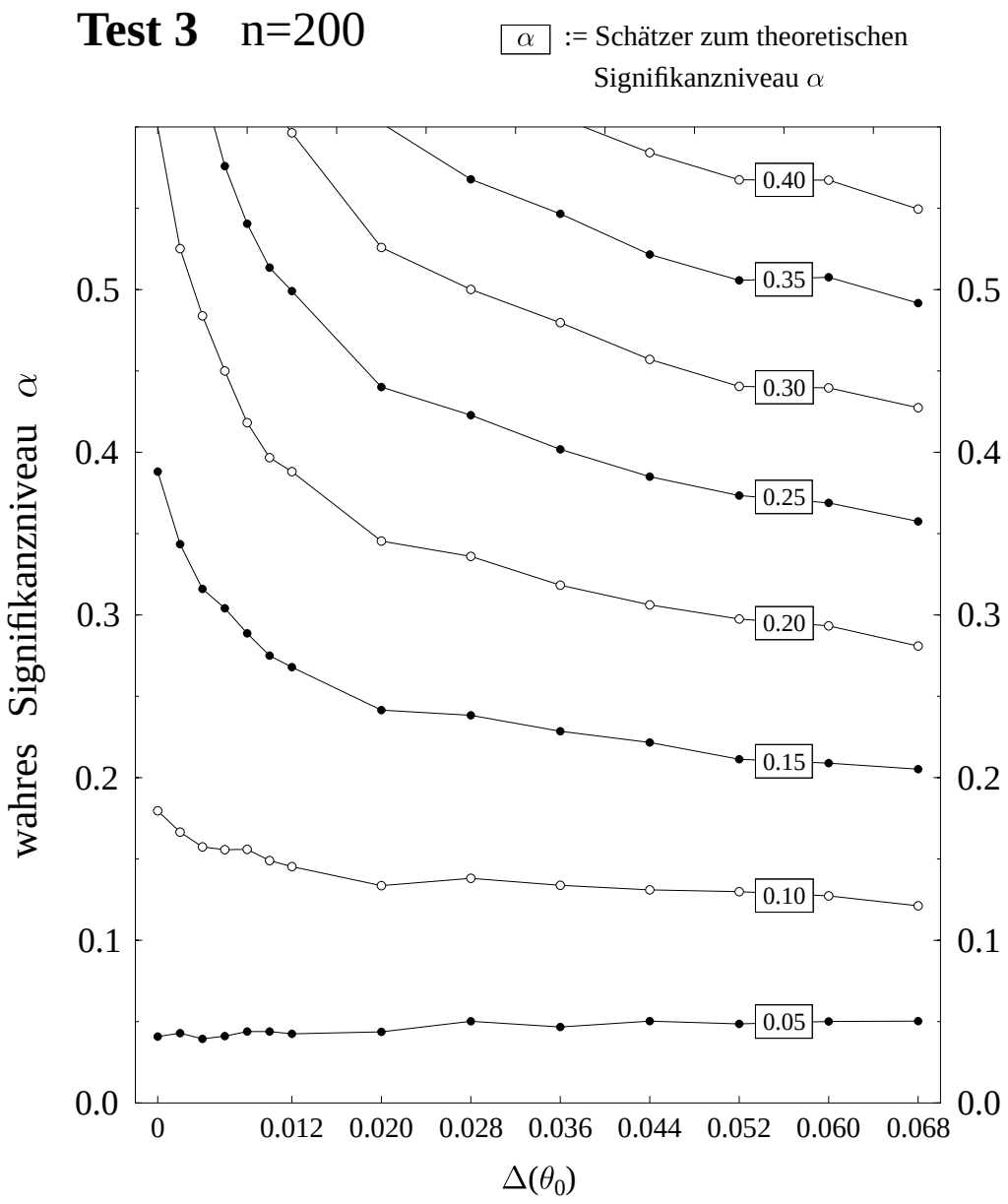


Abb. 5.6: Schätzer der wahren Signifikanzniveaus - Test 3 - n=200

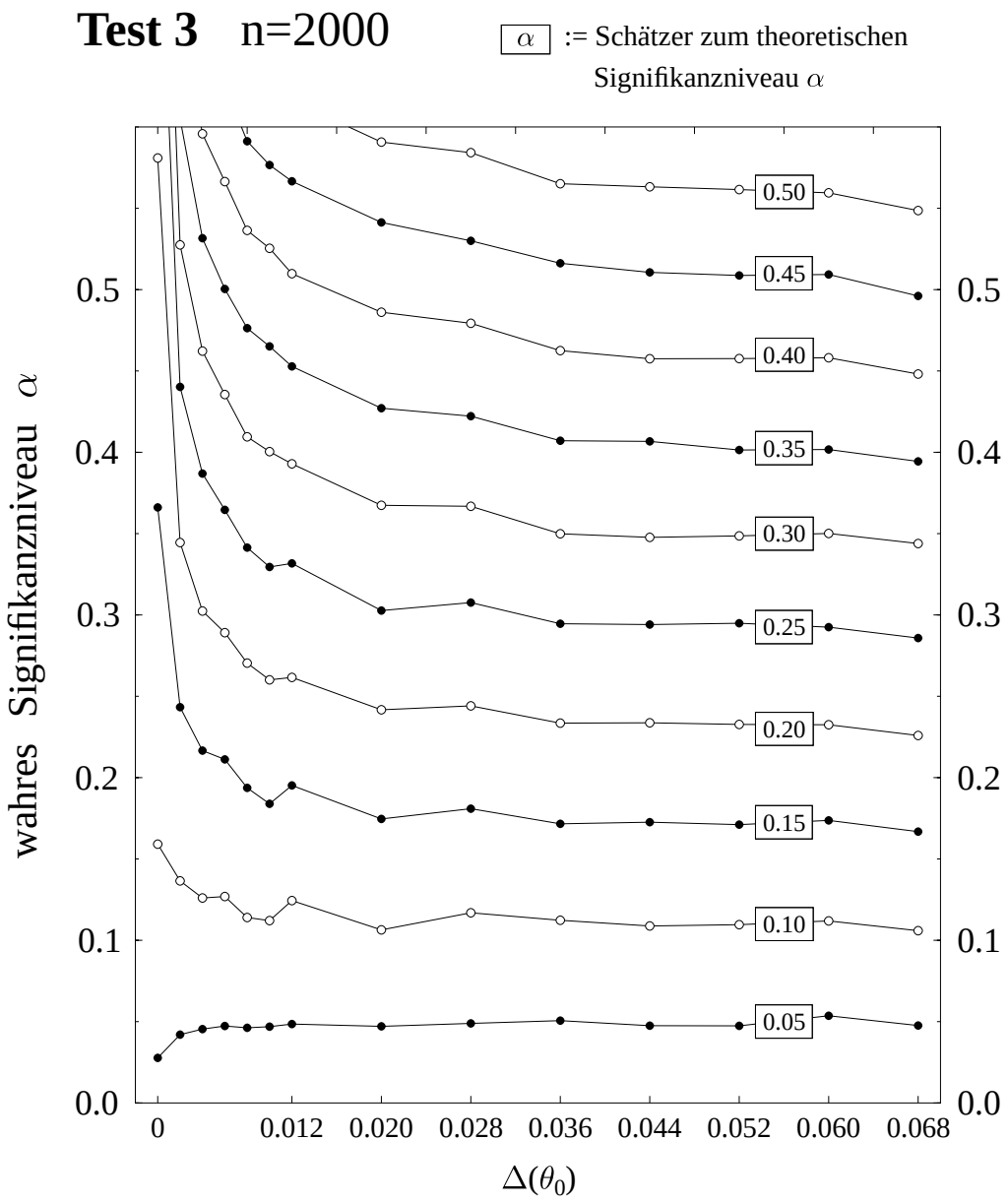


Abb. 5.7: Schätzer der wahren Signifikanzniveaus - Test 3 - n=2000

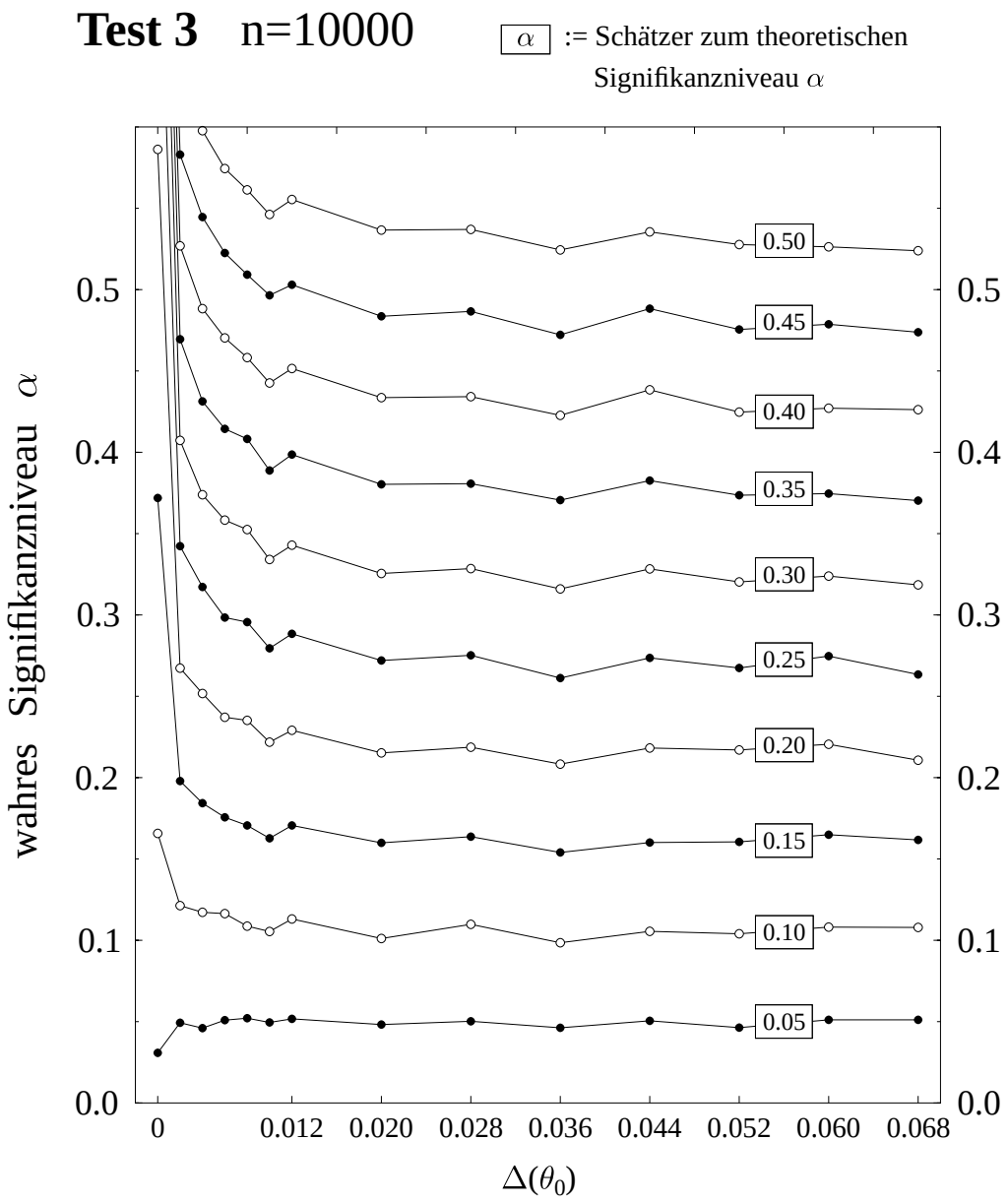


Abb. 5.8: Schätzer der wahren Signifikanzniveaus - Test 3 - n=10000

5.3.3 Simulationen zum Bootstrap-Anpassungstest

5.3.3.1 Durchführung der Simulationen

In Abschnitt 5.2.4.2 wurde ein Bootstrap-Anpassungstest (B-A-Test) unter Verwendung der Statistik

$$R_n(X_1, \ldots, X_n; F) := n \sum_{i=1}^{k} \left(\frac{z_i - h_i(\hat{\theta}(z))}{\sqrt{h_i(\hat{\theta}(z))}} - \frac{\pi_i - h_i(\theta_0)}{\sqrt{h_i(\theta_0)}} \right)^2,$$

hergeleitet. Es soll nun anhand von Simulationen überprüft werden, in welchem Rahmen sich die Abweichungen vom theoretischen Signifikanzniveau bewegen. Für $\theta \in [0, 0.2]$ ist ein Element der approximierenden Modellfamilie wieder durch

$$h(\theta) = (\theta, \theta, \theta, \theta, \theta, 1 - 5\theta)^T$$

gegeben. Ebenso wie im letzten Abschnitt werden Minimum-χ^2-Diskrepanz-Schätzer verwendet. Das Vorgehen bei den Simulationen ist analog gestaltet. Für ein festes $\pi = h(\theta_0)$ werden 10000 Stichproben der Größe n erzeugt. Die Anzahl der Bootstrap-Stichproben beträgt 200. Die kritischen Werte für den Test werden wieder durch $c_{n,L,200}(\alpha, F_n)$ geschätzt. Anschließend wird der Anteil der 10000 Fälle bestimmt, bei denen die Hypothese verworfen wurde.

Unter der Hypothese ist $R_n(X_1, \ldots, X_n, F)$ asymptotisch χ^2_4-verteilt, bei den Simulationen wurde deshalb gleichzeitig festgehalten, wie oft die Hypothese nach dem klassischen Verfahren verworfen wurde.

Außerdem wurde der eigentlich nicht zulässige Test 1 aus Abschnitt 5.2.2 durchgeführt, um die Vermutung aus Abschnitt 5.2.4.1 zu belegen.

5.3.3.2 Simulationsergebnisse zu Beispiel 1

Zunächst sei $\theta_0 = 0.161235$ und

$$\pi = h(\theta_0) = (0.161235, 0.161235, 0.161235, 0.161235, 0.161235, 0.193825)^T.$$

Dieses Modell ist identisch mit dem zu $\Delta(\theta_0) = 0$ gehörigen Modell der vorangegangenen Simulation. Für $n = 50$, $n = 100$, $n = 200$ und $n = 400$ sind die erhaltenen Schätzer für die Signifikanzniveaus $\alpha = 0.05, 0.10, \ldots, 0.45, 0.50$ in den Abbildungen 5.9 und 5.10 dargestellt. Zusätzlich ist für das „falsche“ Testverfahren 1 die Kurve der Grenzwerte der tatsächlichen Signifikanzniveaus eingetragen, die aus dem Ergebnis der Vermutung für $n \to \infty$ ermittelt wurde. Unter dem Begriff „theoretisches Signifikanzniveaußei wieder das

Signifikanzniveau verstanden, das eigentlich bei Durchführung der Tests a priori angenommen wird.

Die Abbildungen zeigen durchgehend gute Resultate für das klassische Verfahren. Dies ist nicht überraschend, denn die minimalen erwarteten Häufigkeiten sind schon bei $n = 50$ genügend groß, um eine Anwendung des klassischen χ^2-Anpassungstests zu rechtfertigen. Aber auch der Bootstrap-Anpassungstest (B-A-Test) liefert mit wachsender Stichprobengröße brauchbare Annäherungen an das gewünschte, theoretische Signifikanzniveau. Nur für $n = 50$ sind die durchschnittlichen Abweichungen von etwa 0.03 vielleicht etwas zu groß. Bei $n = 400$ sind die beiden Tests ebenbürtig.

Das Testverfahren 1 hingegen, das in Abschnitt 5.2.2 zur Hypothese $H_0 : \Delta(\theta_0) = d$, $d > 0$, hergeleitet wurde, versagt für den Fall $\Delta(\theta_0) = 0$. Die Schätzer der wahren Signifikanzniveaus konvergieren, soweit aus den Abbildungen ersichtlich wird, gegen die eingezeichnete Grenzkurve, die aus der Vermutung ermittelt wurde.

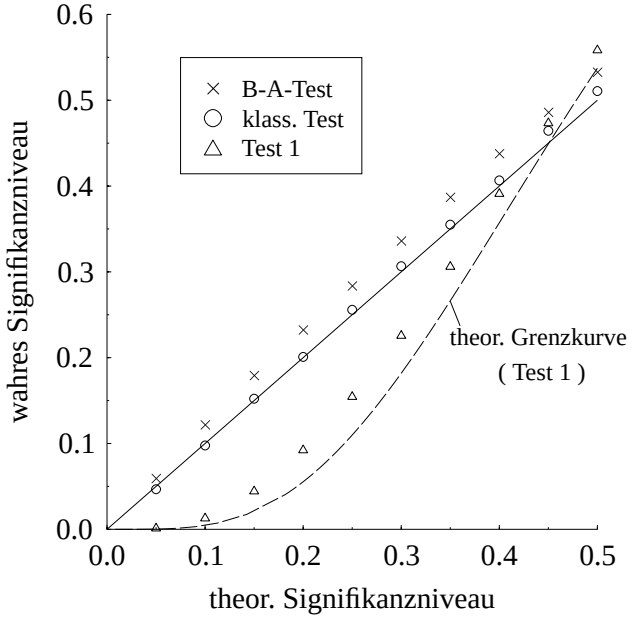
Unterstellt man die Richtigkeit dieser Vermutung, ergibt sich, zusammen mit den Ergebnissen der Simulationen aus Abschnitt 5.3.2.1, daß ein enger Zusammenhang zwischen dem Verlauf der Grenzkurve und den Resultaten des Testverfahrens 1 für Tests zur Hypothese $H_0 : \Delta(\theta_0) \leq d$, $d > 0$, besteht. Für kleine, positive Werte von d wird Test 1 weit vom gewünschten Signifikanzniveau entfernt liegende tatsächliche Signifikanzniveaus liefern, falls die Grenzkurve einen großen Abstand von der Winkelhalbierenden besitzt. Die in den Abbildungen 5.1 und 5.2 zu beobachtende gute Annäherung an das gewünschte Signifikanzniveau $\alpha = 0.45$ läßt sich mit dem Verlauf der Grenzkurve ebenfalls plausibel erklären.

Beispiel 1

n = 50

θ_0 = 0.161235

B = 200



Beispiel 1

n = 100

θ_0 = 0.161235

B = 200

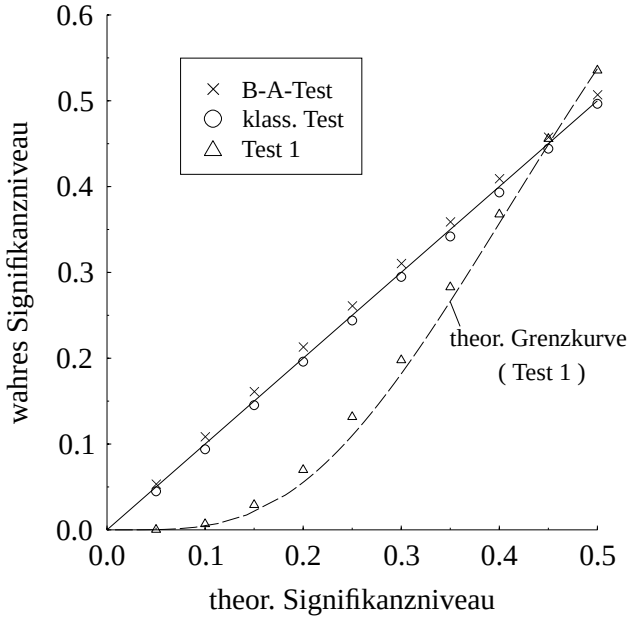


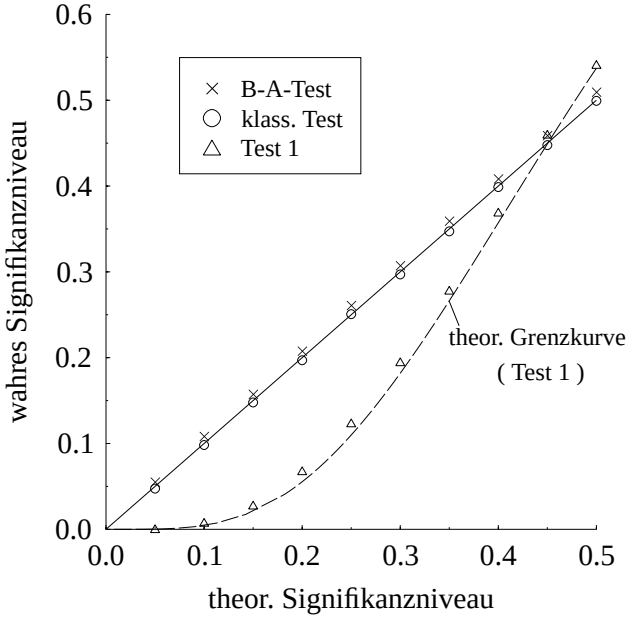
Abb. 5.9: Schätzer der wahren Signifikanzniveaus - Bsp. 1 - n=50, 100

Beispiel 1

n = 200

θ_0 = 0.161235

B = 200



Beispiel 1

n = 400

θ_0 = 0.161235

B = 200

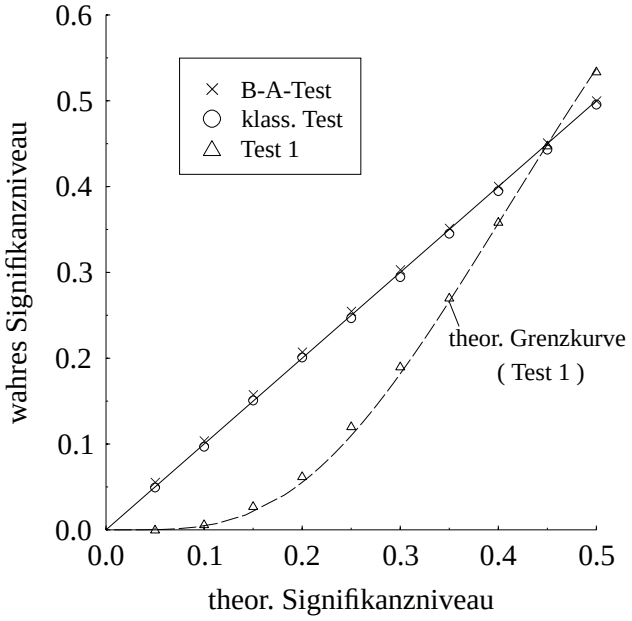


Abb. 5.10: Schätzer der wahren Signifikanzniveaus - Bsp. 1 - n=200, 400

5.3.3.3 Simulationsergebnisse zu Beispiel 2

Es wird nun eine weitere Simulation durchgeführt. Für die Anwendung des klassischen χ^2-Anpassungstests wird häufig gefordert, daß das Minimum der erwarteten Häufigkeiten, je nach Autor, 1, 2 oder sogar 5 übertreffen sollte (Moore, 1986). Seien deshalb ein Modell und Stichprobengrößen betrachtet, bei denen die Verwendung der kritischen Werte aus der asymptotischen χ^2_4-Verteilung zweifelhaft ist:

$$\pi = h(\theta_0) = (0.02, 0.02, 0.02, 0.02, 0.02, 0.9)^T$$

Für $n = 50$, $n = 100$, $n = 200$ und $n = 400$ wurden, analog zu Beispiel 1, Schätzer der wahren Signifikanzniveaus berechnet. In den Abbildungen 5.11 und 5.12 sind diese Schätzer für das klassische und das Bootstrap-Verfahren eingezeichnet. Das Testverfahren 1 wurde nun nicht mehr betrachtet.

Zumindest bei diesem Beispiel hat die Verwendung von Bootstrap-Methoden zur Erzeugung der kritischen Werte keine Verbesserung, sondern im Gegenteil eine Verschlechterung im Vergleich zum klassischen Test erbracht. Zwar konvergieren die Schätzer der tatsächlichen Signifikanzniveaus allmählich gegen die gewünschten Werte, doch während der übliche χ^2-Test spätestens für $n = 400$ ohne Einschränkungen angewendet werden darf, müßte die Stichprobe im Falle des Bootstrap-Tests weiter vergrößert werden, um von einer zufriedenstellenden Annäherung an das theoretische Signifikanzniveau zu sprechen.

Die Vorteile des Bootstrap-Anpassungstests liegen deshalb nicht in dem Bereich, bei denen die Prüfgröße unter der Hypothese eine bekannte asymptotische Verteilung besitzt.

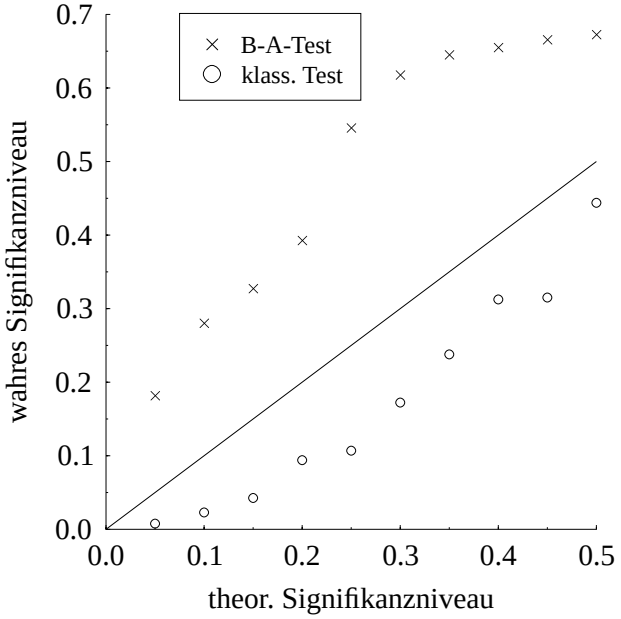
Im Falle der χ^2-Prüfgröße stellt sich die asymptotische χ^2_{k-s-1}-Verteilung nur dann ein, wenn BAN-Schätzer eingesetzt werden. Werden jedoch z. B. Minimum-Diskrepanz-Schätzer zu Diskrepanzen $\Delta_g \in \mathcal{D} \setminus \mathcal{D}^*$ verwendet, so ergibt sich im allgemeinen eine von π abhängige asymptotische Verteilung unter der Hypothese. Nur bei Anwendung des Bootstrap-Anpassungstest werden dann die kritischen Werte automatisch erzeugt und für $n \to \infty$ die korrekten Signifikanzniveaus angenommen.

Beispiel 2

n = 50

θ_0 = 0.02

B = 200



Beispiel 2

n = 100

θ_0 = 0.02

B = 200

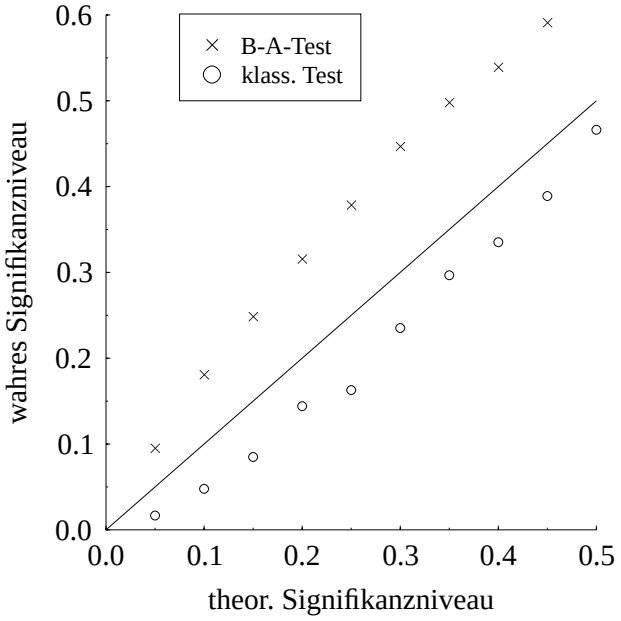


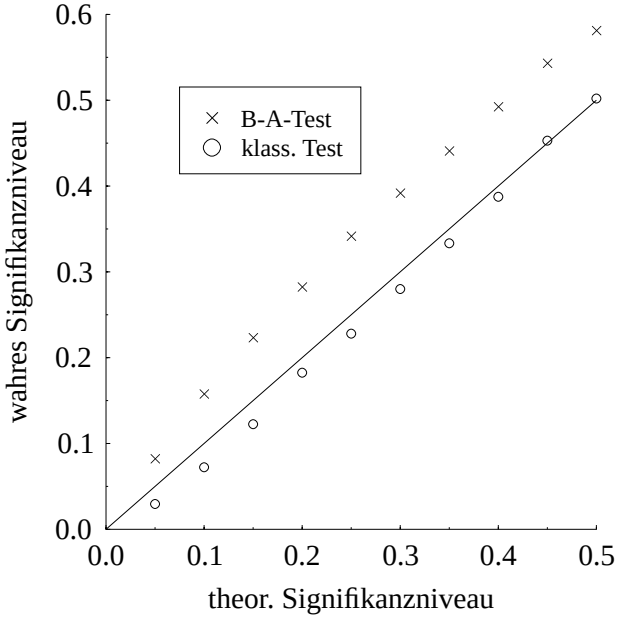
Abb. 5.11: Schätzer der wahren Signifikanzniveaus - Bsp. 2 - n=50, 100

Beispiel 2

n = 200

$\theta_0 = 0.02$

B = 200



Beispiel 2

n = 400

$\theta_0 = 0.02$

B = 200

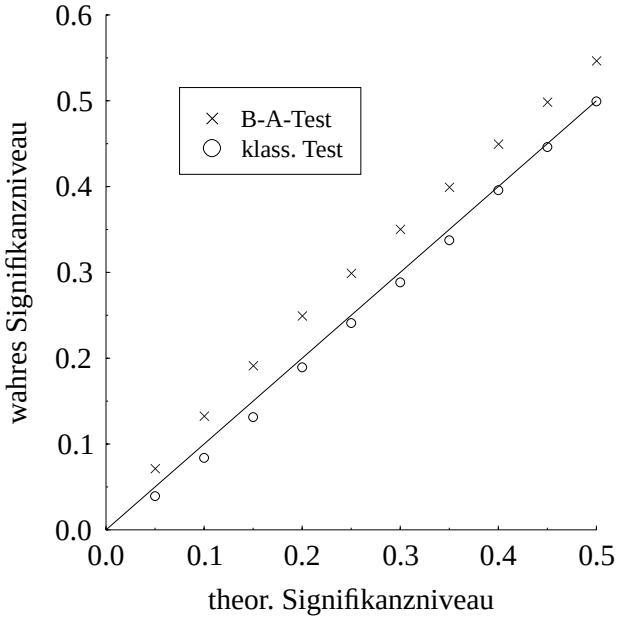


Abb. 5.12: Schätzer der wahren Signifikanzniveaus - Bsp. 2 - n=200, 400

5.3.4 Exkurs: Maximum-Likelihood-Schätzer der ursprünglichen, ungruppierten Stichprobe

5.3.4.1 Problemstellung

Eine weitere interessante Anwendungsmöglichkeit sei in diesem Abschnitt beschrieben. Häufig liegen Beobachtungen von identisch und unabhängig verteilten Zufallsvariablen mit einer Dichte $f_G(x)$ vor. Soll nun getestet werden, ob $f_G(x)$ in der Modellfamilie $\{f_\theta(x) : \theta \in \Theta\}$ enthalten ist, muß zunächst eine Gruppierung vorgenommen werden, um den χ^2-Anpassungstest durchführen zu können. Ist z. B. die reelle Achse durch Angabe von $k-1$ Werten

$$a_1 < a_2 < \ldots < a_{k-2} < a_{k-1}$$

in k Intervalle aufgeteilt, und setzt man weiterhin $a_0 := -\infty$ bzw. $a_k := +\infty$, so lassen sich π und $h(\theta)$ mit

$$\begin{aligned} \pi_i &:= \int_{a_{i-1}}^{a_i} f_G(x)dx, \quad i = 1, \ldots, k, \text{ bzw.} \\ h_i(\theta) &:= \int_{a_{i-1}}^{a_i} f_\theta(x)dx, \quad i = 1, \ldots, k, \end{aligned}$$

vernünftig definieren. Die relativen Häufigkeiten z_i, $i = 1, \ldots, k$, bestimmt man anschließend aus dem Anteil der Werte der gegebenen ursprünglichen Stichprobe, die in das Intervall $(a_{i-1}; a_i]$ fallen. Mit diesen Definitionen ließen sich prinzipiell alle in dieser Arbeit genannten Verfahren anwenden, solange die jeweils verlangten Voraussetzungen an π und $h(\theta)$ erfüllt sind.

Aufgrund der involvierten Integrale ist es jedoch häufig schwierig, Minimum-Diskrepanz-Schätzer der Parameter zu berechnen. Eine gebräuchliche Praxis besteht darin, aus der *ungruppierten* Stichprobe $x_1, \ldots, x_n$ gebildete Maximum-Likelihood-Schätzer $\hat{\theta}(x_1, \ldots, x_n)$ zu verwenden. Werden diese jedoch in die χ^2-Prüfgröße eingesetzt, so ist die asymptotische Verteilung nicht mehr durch eine χ^2_{k-s-1}-Verteilung gegeben. Chernoff & Lehmann (1954) haben gezeigt, daß, bei entsprechenden Anforderungen an die auftretenden Dichten, die asymptotische Verteilung durch eine Linearkombination von unabhängig verteilten χ^2-Variablen gegeben ist:

$$n\Delta(z, h(\hat{\theta}(x_1, \ldots, x_n))) \xrightarrow{d} \chi^2_{k-s-1} + \sum_{i=1}^{s} \lambda_i \chi_1^2$$

Die Koeffizienten λ_i liegen im Intervall $[0, 1]$ und hängen vom unbekannten, wahren Modell ab. In der Praxis haben sich zwei Varianten der Anwendung

des Tests ergeben. Die Vertreter der einen Richtung verwenden weiterhin kritische Werte der χ^2_{k-s-1}-Verteilung, nehmen dabei aber in Kauf, daß das tatsächliche Signifikanzniveau viel höher ist als das angenommene. Von Vorsicht geprägt ist die Denkweise der zweiten Abteilung. Sie benutzt die χ^2_{k-1}-Verteilung zum Ablesen der kritischen Werte und verschenkt damit zuweilen die Möglichkeit falsche Hypothesen zu entdecken.

Besonders kritisch sind diese beiden Vorgehensweisen, wenn die Anzahl der Parameter nur wenig kleiner als k ist. An dieser Stelle bietet sich die Anwendung von Bootstrap-Verfahren an. Da sich der theoretische Hintergrund von dem der bisherigen Abschnitte unterscheidet und einen gänzlich neuen Ansatz erfordern würde, sei nun einfach eine Simulation vorgestellt. Die theoretische Absicherung, d. h. die Zusammenstellung der Anforderungen an die Dichtefunktionen und der Beweis der korrekten Konvergenz der Schätzer der Signifikanzniveaus, ist ein noch zu lösendes Problem.

Die Vorgehensweise bei der Durchführung im Vergleich zum Bootstrap-Anpassungstest aus Abschnitt 5.2.4.2 unterscheidet sich nur in zwei Punkten. Die Schätzer werden zum einen aus der ursprünglichen Stichprobe ermittelt, nicht über den Vektor z der relativen Häufigkeiten, zum anderen werden die Bootstrap-Stichproben ebenfalls aus den ungruppierten Daten erzeugt.

5.3.4.2 Simulation

Chernoff & Lehmann (1954) betrachten zur Illustration ihres theoretischen Resultates die folgende Konstellation: An gegebene Daten $x_1, \ldots, x_n$ soll eine Normalverteilung $N(\mu, \sigma^2)$ angepaßt und anschließend ein Anpassungstest durchgeführt werden. Die reelle Achse ist in vier Zellen aufgeteilt:

$$(a_0, a_1, a_2, a_3, a_4) := (-\infty, -1, 0, 1, +\infty)$$

Es werden die Maximum-Likelihood-Schätzer

$$\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{bzw.} \\ \hat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \end{aligned}$$

verwendet. Das Grundmodell ist durch eine $N(0, 6.25)$-Verteilung gegeben. Sei nun z. B. das gewünschte Signifikanzniveau $\alpha = 0.05$. Chernoff & Lehmann (1954, S.586) schätzen das tatsächliche asymptotische Signifikanzni-

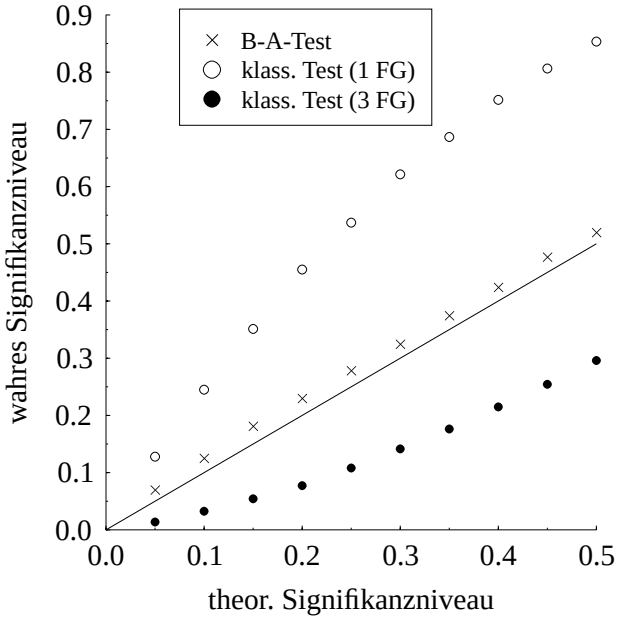
veau nach unten mit 0.12 ab, falls der kritische Wert aus der χ_1^2-Verteilung abgelesen wird. Es tritt somit eine Abweichung um mehr als 100 Prozent auf.

Anders verhält es sich bei Anwendung des Bootstrap-Anpassungstests. Unter dem Grundmodell wurden nun 10000 Stichproben der Größe 50 bzw. 200 erzeugt. Für jede Stichprobe wurden wiederum 200 Bootstrap-Stichproben generiert. Neben den durch das Bootstrap-Verfahren ermittelten kritischen Werten zu $\alpha = 0.05, 0.10, \ldots, 0.50$ wurden auch die kritischen Werte betrachtet, die sich aus der χ_1^2- und der χ_3^2-Verteilung ergeben. Die Schätzer der tatsächlichen Signifikanzniveaus sind in Abbildung 5.13 dargestellt.

Die Graphiken zeigen deutlich, daß für den Bootstrap-Test die tatsächlichen Signifikanzniveaus dicht bei den gewünschten liegen, selbst für die relativ kleine Stichprobengröße von $n = 50$. Der Bootstrap-Anpassungstest kann somit angewendet werden, obwohl die asymptotische Verteilung der χ^2-Prüfgröße überhaupt nicht bekannt ist. Die Benutzung der kritischen Werte der χ_1^2- bzw. χ_3^2-Verteilung führt hingegen zu deutlichen Abweichungen.

n = 50

B = 200



n = 200

B = 200

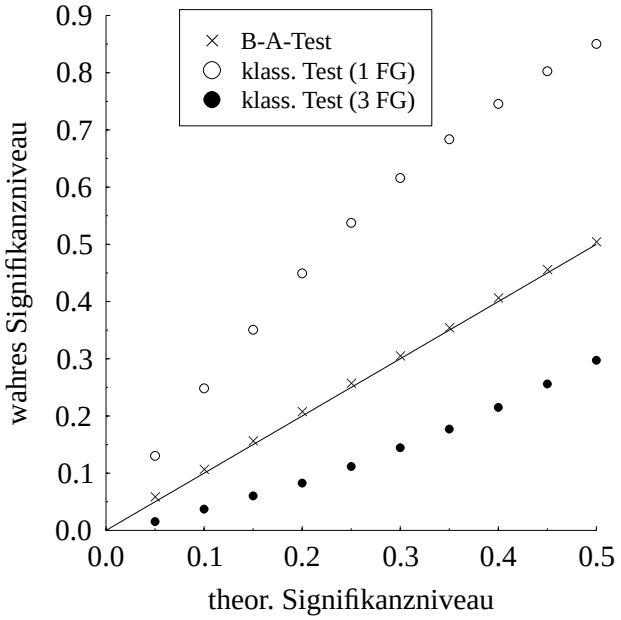


Abb. 5.13: Schätzer der wahren Signifikanzniveaus, Bsp. von Chernoff & Lehmann (1954)

5.4 Zwei Modellfamilien: Test über die Differenz der Diskrepanzen durch Approximation

Die bisherigen Bootstrap-Tests dienten dazu, Aussagen über die Diskrepanz durch Approximation *einer* Modellfamilie zu treffen. Liegen *zwei* Modellfamilien $\{h^{(1)}(\theta^{(1)}) : \theta^{(1)} \in \Theta^{(1)}\}$ und $\{h^{(2)}(\theta^{(2)}) : \theta^{(2)} \in \Theta^{(2)}\}$ vor, besteht Interesse, die Modellfamilie mit der geringeren Diskrepanz durch Approximation auszuwählen und die getroffene Wahl statistisch abzusichern. Ein Bootstrap-Test kann nun für dieses Problem konstruiert werden.

Der Einfachheit halber sei $\Delta_g = \Delta$. Sei $\Delta^{(i)} := \Delta(\pi, h^{(i)}(\theta_0^{(i)}))$, $i = 1, 2$. Eine plausible Hypothese lautet nun:

$$H_0^* : \Delta^{(1)} \leq \Delta^{(2)}$$

Diese Hypothese läßt sich als Spezialfall der allgemeineren Hypothese

$$H_0 : \Delta^{(1)} - \Delta^{(2)} \leq d, \quad d \in I\!R,$$

auffassen. Um nun einen Bootstrap-Test zu H_0 herzuleiten, definiere man zunächst die vorläufige Statistik

$$T_n(X_1, \ldots, X_n; F) := \sqrt{n}\left((\Delta_n^{(1)}(\hat{\theta}^{(1)}) - \Delta^{(1)}) - (\Delta_n^{(2)}(\hat{\theta}^{(2)}) - \Delta^{(2)})\right),$$

wobei $\Delta_n^{(i)}(\hat{\theta}^{(i)})$, $i = 1, 2$, abkürzend für $\Delta(z^{(n)}, h^{(i)}(\hat{\theta}^{(i)}(z^{(n)})))$ stehen soll. Unter der Annahme, daß keine der beiden Modellfamilien das wahre Modell π enthalten soll, folgt mittels der Delta-Methode:

$$\sqrt{n}\begin{pmatrix} \Delta_n^{(1)}(\hat{\theta}^{(1)}) - \Delta^{(1)} \\ \Delta_n^{(2)}(\hat{\theta}^{(2)}) - \Delta^{(2)} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right),$$

wobei

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = 4 * \begin{pmatrix} \pi^T D_{h^{(1)}}^{-1} \\ \pi^T D_{h^{(2)}}^{-1} \end{pmatrix} \left(D_\pi - \pi\pi^T\right) \begin{pmatrix} D_{h^{(1)}}^{-1}\pi \\ D_{h^{(2)}}^{-1}\pi \end{pmatrix}$$

ist. Somit ergibt sich

$$T_n(X_1, \ldots, X_n; F) \xrightarrow{d} N(0, \sigma^2)$$

mit

$$\begin{aligned}
\sigma^2 &= \sigma_1^2 - 2\sigma_{12} + \sigma_2^2 \\
&= 4 * \left[\sum_{i=1}^{k} \frac{\pi_i^3}{(h_i^{(1)}(\theta_0^{(1)}))^2} - \left(\sum_{i=1}^{k} \frac{\pi_i^2}{h_i^{(1)}(\theta_0^{(1)})} \right)^2 \right. \\
&\quad -2 \left(\sum_{i=1}^{k} \frac{\pi_i^3}{h_i^{(1)}(\theta_0^{(1)}) h_i^{(2)}(\theta_0^{(2)})} - \sum_{i=1}^{k} \frac{\pi_i^2}{h_i^{(1)}(\theta_0^{(1)})} \sum_{i=1}^{k} \frac{\pi_i^2}{h_i^{(2)}(\theta_0^{(2)})} \right) \\
&\quad \left. + \sum_{i=1}^{k} \frac{\pi_i^3}{(h_i^{(2)}(\theta_0^{(2)}))^2} - \left(\sum_{i=1}^{k} \frac{\pi_i^2}{h_i^{(2)}(\theta_0^{(2)})} \right)^2 \right]
\end{aligned}$$

Sei nun die Funktion $\sigma^2(p)$ mit

$$\begin{aligned}
\sigma^2(p) := 4 * &\left[\sum_{i=1}^{k} \frac{p_i^3}{(h_i^{(1)}(\hat{\theta}^{(1)}(p)))^2} - \left(\sum_{i=1}^{k} \frac{p_i^2}{h_i^{(1)}(\hat{\theta}^{(1)}(p))} \right)^2 \right. \\
-2 &\left(\sum_{i=1}^{k} \frac{p_i^3}{h_i^{(1)}(\hat{\theta}^{(1)}(p)) h_i^{(2)}(\hat{\theta}^{(2)}(p))} - \sum_{i=1}^{k} \frac{p_i^2}{h_i^{(1)}(\hat{\theta}^{(1)}(p))} \sum_{i=1}^{k} \frac{p_i^2}{h_i^{(2)}(\hat{\theta}^{(2)}(p))} \right) \\
+ &\left. \sum_{i=1}^{k} \frac{p_i^3}{(h_i^{(2)}(\hat{\theta}^{(2)}(p)))^2} - \left(\sum_{i=1}^{k} \frac{p_i^2}{h_i^{(2)}(\hat{\theta}^{(2)}(p))} \right)^2 \right]
\end{aligned} \tag{5.16}$$

definiert. Analog zur Vorgehensweise des Lemmas 5.2 läßt sich der folgende Satz beweisen:

Satz 5.3 *Seien* **B1-B5** *für die beiden Modellfamilien erfüllt. Sei* (G_n) *eine Folge aus* $\mathcal{G}_F$ *und* $(p^{(n)})$ *die dazugehörige Folge von Wahrscheinlichkeitsvektoren. Sei weiterhin* $(z_*^{(n)})$ *eine Folge von Vektoren der relativen Häufigkeiten, wobei jedes* $z_*^{(n)}$ *unter* G_n *erzeugt wird. Sei* $\sigma^2(p)$ *wie in (5.16) definiert. Dann gilt*

$$\sqrt{n} \frac{(\Delta_n^{*(1)} - \Delta_n^{*(2)}) - (\Delta_n^{(1)} - \Delta_n^{(2)})}{(\sigma^2(z_*^{(n)}))^{\frac{1}{2}}} \xrightarrow{d} N(0,1),$$

wobei die folgenden Definitionen getroffen werden:

$$\begin{aligned}
\Delta_n^{*(i)} &:= \Delta(z_*^{(n)}, h^{(i)}(\hat{\theta}^{(i)}(z_*^{(n)}))), \quad i = 1, 2. \\
\Delta_n^{(i)} &:= \Delta(p^{(n)}, h^{(i)}(\hat{\theta}^{(i)}(p^{(n)}))), \quad i = 1, 2.
\end{aligned}$$

Beweis: Zunächst zeigt man, analog zu der Herleitung der asymptotischen Verteilung von $T_n(X_1, \ldots, X_n; F)$, mittels der Delta-Methode aus Satz 2.3:

$$\sqrt{n}\left((\Delta_n^{*(1)} - \Delta_n^{*(2)}) - (\Delta_n^{(1)} - \Delta_n^{(2)})\right) \xrightarrow{d} N(0, \sigma^2).$$

Anschließend folgt, analog zu Lemma 5.2,

$$(\sigma^2(z_*^{(n)}))^{\frac{1}{2}} \xrightarrow{p} (\sigma^2)^{\frac{1}{2}}.$$

Aus Slutskys Theorem (Serfling, 1980, S. 19) folgt dann die Behauptung. □

Die Voraussetzungen des Theorems 1 von Beran sind deshalb erfüllt. Definiert man nun

$$R_n(X_1, \ldots, X_n; F) := T_n(X_1, \ldots, X_n; F)/(\sigma^2(z^{(n)}))^{\frac{1}{2}} \quad ,$$

dann konvergiert die bedingte Verteilung von

$$R_n(X_1^*, \ldots, X_n^*; F_n) \;\; := \;\; T_n(X_1^*, \ldots, X_n^*; F_n)/(\sigma^2(z_*^{(n)}))^{\frac{1}{2}},$$

gegeben z_n, für fast alle Folgen (z_n) gegen die asympotische Verteilung von $R_n(X_1, \ldots, X_n; F)$. Ist $c_n(\alpha, F_n)$ wieder ein Schätzer des $1-\alpha$-Quantils der Verteilung von $R_n(X_1^*, \ldots, X_n^*; F_n)$, so folgt nach Beran:

$$\lim_{n\to\infty} P_F(R_n(X_1, \ldots, X_n; F) > c_n(\alpha, F_n)) = \alpha$$

Unter der Hypothese

$$H_0 : \Delta^{(1)} = \Delta^{(2)} + d, \quad d \in I\!R$$

ergibt sich dann die Konvergenzaussage

$$\lim_{n\to\infty} P_F(\sqrt{n}\frac{\Delta_n^{(1)}(\hat{\theta}^{(1)}) - \Delta_n^{(2)}(\hat{\theta}^{(2)}) - d}{(\sigma^2(z^{(n)}))^{\frac{1}{2}}} > c_n(\alpha, F_n)) = \alpha.$$

Verwirft man die Hypothese, falls

$$\sqrt{n}\frac{\Delta_n^{(1)}(\hat{\theta}^{(1)}) - \Delta_n^{(2)}(\hat{\theta}^{(1)}) - d}{(\sigma^2(z^{(n)}))^{\frac{1}{2}}} > c_n(\alpha, F_n)$$

ist, so ist dies gleichzeitig ein Test zur Hypothese

$$H_0 : \Delta^{(1)} \leq \Delta^{(2)} + d, \quad d \in I\!R.$$

Hauptsächlich wird wohl der Test zum Spezialfall $d = 0$ Anwendung finden.

Die Zuverlässigkeit des in dieser Art und Weise konstruierten Tests in bezug auf die Einhaltung des angenommenen Signifikanzniveaus läßt sich schwer beurteilen. Vermutlich liegt das tatsächliche Signifikanzniveau um so näher am gewünschten, je größer die Diskrepanzen durch Approximation der betrachteten Modellfamilien sind und je größer σ^2 ist.

Kapitel 6

Bootstrap-Tests für andere Diskrepanzen

6.1 Einleitung

Die Idee der in Kapitel 5 hergeleiteten Tests ist natürlich nicht an die χ^2-Diskrepanz gebunden. Die Tests zur Hypothese $\Delta(\theta_0) \leq d$ der Abschnitte 5.2.2 und 5.2.3 lassen sich fast ohne Einschränkungen auf beliebige Diskrepanzen der Menge $\mathcal{D}$ übertragen. Aber auch im Falle des Bootstrap-Anpassungstests existieren Diskrepanzen, bei denen der Test analog zum χ^2-Fall konstruiert wird.

Eine andere Frage ist, welche Bedingungen an die betrachteten Diskrepanzen gestellt werden müssen, damit die Tests überhaupt sinnvoll eingesetzt werden können. Wird z. B. eine Diskrepanz verwendet, bei der $\Delta(p, p)$ vom Modell p abhängt, läßt sich zwar formal ein Test zur Hypothese $H_0 : \Delta(\theta_0) \leq d$ durchführen, doch führt die Angabe eines Wertes von d unweigerlich zu Schwierigkeiten. Sei beispielsweise

$$\Delta(\pi, h(\theta)) = \sum_{i=1}^{k} \left((\pi_i - h_i(\theta))^2 + \pi_i^2 \right),$$

sei außerdem $\pi \in \{h(\theta) : \theta \in \Theta\}$, aber ansonsten beliebig, dann ergibt sich in diesem Fall:

$$\Delta(\pi, h(\theta_0)) = \sum_{i=1}^{k} \pi_i^2$$

Obwohl π in der approximierenden Modellfamilie enthalten ist, kann die Diskrepanz durch Approximation prinzipiell, je nach genauem Aussehen der betrachteten Modellfamilie, jeden Wert $1/k \leq \Delta(\theta_0) \leq 1$ annehmen. Es kann dann vorkommen, daß die Diskrepanz durch Approximation für ein $\pi \in \{h(\theta) : \theta \in \Theta\}$ größer ist als der entsprechende Wert eines Modells, das nicht in dieser Menge enthalten ist. Als Konsequenz läßt sich festhalten, daß nur Diskrepanzen mit

$$\Delta(p,p) \equiv 0, \quad p \in [0,1]^k, \ \sum_{i=1}^{k} p_i = 1, \tag{6.1}$$

betrachtet werden sollten. Diskrepanzen mit

$$\Delta(p,p) \equiv c, \quad c \in I\!R,$$

lassen sich durch Subtraktion von c in die Gestalt (6.1) überführen.

Die eben genannten Ausführungen verlieren ihre Gültigkeit, falls zwei Modellfamilien betrachtet werden und Aussagen über die Differenz der Diskrepanzen durch Approximation getroffen werden sollen, also nur die relative Anordnung von Wichtigkeit ist. Hypothesen der Form (vgl. Abschnitt 5.4)

$$H_0 : \Delta^{(1)} \leq \Delta^{(2)}$$

lassen sich überprüfen, da die absoluten Werte der Diskrepanzen in diesem Fall keine Rolle spielen.

6.2 Tests über $\Delta(\theta_0)$

6.2.1 Test 1

Sei $\Delta \in \mathcal{D}$ eine beliebige Diskrepanz, die die Bedingung (6.1) erfüllt. Für die Ermittlung der Schätzer $\hat{\theta}(z)$ wird eine Diskrepanz $\Delta_g \in \mathcal{D}$ verwendet. Analog zu Abschnitt 5.2.2 wird die betrachtete Statistik mit

$$R_n(X_1, \ldots, X_n; F) := \sqrt{n}[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))] \tag{6.2}$$

definiert. Sei $\mathcal{G}_{\mathcal{F}}$ die Menge aus Definition 5.2 und $(G_n) \in \mathcal{G}_{\mathcal{F}}$, sowie $(p^{(n)})$ die Folge der zugehörigen Wahrscheinlichkeitsvektoren. Seien weiterhin $X_1^{(n)}, \ldots, X_n^{(n)}$ unabhängig und identisch verteilte Zufallsvariablen mit

Verteilungsfunktion G_n und $z_*^{(n)}$ der zugehörige Vektor der relativen Häufigkeiten. Ist $\pi \neq h(\theta_0)$, so ergibt sich unter Verwendung der Deltamethode aus Satz 5.1 und den Ausführungen des Abschnittes 4.3.2.1 mit den dort verwendeten Bezeichnungen:

$$\sqrt{n}[\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(p^{(n)}, h(\hat{\theta}(p^{(n)})))] \xrightarrow{d} N(0, W(D_\pi - \pi\pi^T)W^T) \quad (6.3)$$

Die asymptotische Varianz ist unabhängig von der genauen Gestalt der Folge $\{G_n\}$. Kann vorausgesetzt werden, daß

$$\sigma^2 := W(D_\pi - \pi\pi^T)W^T \quad (6.4)$$

positiv ist, so ist die Bedingung **V3** des Theorems 1 von Beran (1984) erfüllt. Es folgt der Satz:

Satz 6.1 *Unter der Hypothese*

$$H_0^* : \Delta(\pi, h(\theta_0)) = d, \quad d > 0,$$

ergibt sich

$$\lim_{n\to\infty} P_F(\sqrt{n}[\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - d] \geq c_n(\alpha, F_n)) = \alpha.$$

Der Test zur Hypothese

$$H_0 : \Delta(\pi, \theta_0) \leq d, \quad d > 0,$$

ist dann analog zu Abschnitt 5.2.2 durchzuführen.

6.2.2 Test 2

Mit dem Ergebnis des letzten Unterabschnitts läßt sich nun ein weiterer Test zur Hypothese $\Delta(\theta_0) \leq d$ herleiten, der analog zum Test aus Abschnitt 5.2.3 aufgebaut ist. Es genügt dazu, einen konsistenten Schätzer der asymptotischen Standardabweichung σ aus (6.4) zu finden. Besonders einfach läßt sich ein solcher Schätzer angeben, falls $\Delta_g = \Delta$ ist, d. h die betrachtete Diskrepanz gleichzeitig zur Ermittlung der Schätzer $\hat{\theta}(z^{(n)})$ verwendet wird. Mit den Bezeichnungen des Abschnitts 4.3.2.1 vereinfacht sich σ^2 dann zu

$$\sigma^2 = G_x(D_\pi - \pi\pi^T)G_x^T.$$

Sei nun

$$G_x(p) := \left(\frac{\partial g^*(x,y)}{\partial x}\bigg|_{(x,y)=(p_1,h_1(\hat{\theta}(p)))}, \ldots, \frac{\partial g^*(x,y)}{\partial x}\bigg|_{(x,y)=(p_k,h_k(\hat{\theta}(p)))} \right)$$

und

$$\sigma^2(p) := G_x(p)(D_p - pp^T)(G_x(p))^T,$$

dann ist $\sigma^2(p)$ in einer Umgebung von π stetig mit $\sigma^2(\pi) = \sigma^2$. Sei $z_*^{(n)}$ wie in Abschnitt 6.2.1 definiert. Genauso wie in Lemma 5.2 folgt nun:

$$\sigma(z_*^{(n)}) \xrightarrow{p} \sigma(\pi)$$

Aus Slutskys Theorem (Serfling, 1980) und (6.3) ergibt sich:

$$\sqrt{n}\frac{\Delta(z_*^{(n)}, h(\hat{\theta}(z_*^{(n)}))) - \Delta(p^{(n)}, h(\hat{\theta}(p^{(n)})))}{(\sigma^2(z_*^{(n)}))^{\frac{1}{2}}} \xrightarrow{d} N(0,1)$$

Definiert man deshalb

$$R_n(X_1, \ldots, X_n; F) := \sqrt{n}\frac{\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))}{(\sigma^2(z^{(n)}))^{\frac{1}{2}}},$$

so sind die Voraussetzungen **V1**–**V3** des Theorems 1 von Beran (1984) erfüllt, und es folgt unmittelbar:

Satz 6.2 *Unter der Hypothese*

$$H_0^* : \Delta(\pi, h(\theta_0)) = d, \quad d > 0$$

ergibt sich

$$\lim_{n\to\infty} P_F(\sqrt{n}\frac{\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - d}{(\sigma^2(z^{(n)}))^{\frac{1}{2}}} \geq c_n(\alpha, F_n)) = \alpha$$

Der Test 2 zur Hypothese

$$H_0 : \Delta(\pi, \theta_0) \leq d, \quad d > 0$$

ist dann analog zu Abschnitt 5.2.3 durchzuführen.

6.2.3 Bootstrap-Anpassungstests

6.2.3.1 Allgemeine Herleitung

In Abschnitt 5.2.4.1 wurde dargelegt, daß die intuitive Erweiterung des Tests zur Hypothese $H_0 : \Delta(\theta_0) \leq d$, $d > 0$, auf den Fall $d = 0$ leider nicht funktioniert. Dieses Problem tritt bei allen Diskrepanzen auf, bei denen die asymptotische Verteilung der Statistik

$$R_n(X_1, \ldots, X_n; F) = \sqrt{n}(\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) - \Delta(\pi, h(\theta_0))$$

unter $H_0 : \pi = h(\theta_0)$ nicht mehr existiert und durch Verwendung des Faktors n anstelle von $\sqrt{n}$ eine qualitativ abweichende asymptotische Verteilung entsteht.

In diesem Abschnitt wird gezeigt, wie sich die Idee des modifizierten Bootstrap-Anpassungstests aus Abschnitt 5.2.4.2 auf einige spezielle Diskrepanzen übertragen läßt. Es ergibt sich, daß sich für Diskrepanzen, bei denen der verbesserte Ansatz zur Berechnung der exakten Macht im Falle eines Anpassungstests aus Abschnitt 4.3.4 möglich ist, ein zu Abschnitt 5.2.4.2 analoger Test herleiten läßt.

Sei deshalb $\Delta(\pi, h(\theta))$ eine Diskrepanz der Form, wie sie in Abschnitt 4.3.4 angenommen wird:

$$\Delta(\pi, h(\theta)) = \sum_{i=1}^{k} g^*(\pi_i, h_i(\theta))$$

$g^*(x, y)$ läßt sich als Quadrat einer Funktion $w^*(x, y)$ ausdrücken. Die Abbildung $w : [0; 1]^k \to I\!R^k$ sei durch

$$w_i(p) := w^*(p_i, h_i(\hat{\theta}(p))), \quad i = 1, \ldots, k,$$

definiert. Sei weiterhin $(G_n) \in \mathcal{G}_{\mathcal{F}}$ und $(p^{(n)})$ die Folge der zugehörigen Wahrscheinlichkeitsvektoren. Bezeichne $z_*^{(n)}$ wieder den Vektor der relativen Häufigkeiten unter G_n. Mittels der Delta-Methode aus Satz 2.3 folgt nun:

$$\sqrt{n}(w(z_*^{(n)}) - w(p^{(n)})) \xrightarrow{d} N(\mathbf{0}, W(D_\pi - \pi\pi^T)W^T)$$

Die Anwendung der Sätze 2.1 und 2.2 ergibt:

$$n(w(z_*^{(n)}) - w(p^{(n)}))^T(w(z_*^{(n)}) - w(p^{(n)})) \xrightarrow{d} \sum_{\lambda_i > 0} \lambda_i \chi_1^2(\frac{n(Sw(\pi))_i^2}{\lambda_i}) \quad + \quad \sum_{\lambda_i = 0} n(Sw(\pi))_i^2 \quad (6.5)$$

S ist eine orthogonale Matrix, deren transponierte Zeilen eine Basis aus Eigenvektoren der Matrix $W(D_\pi - \pi\pi^T)W^T$ bilden und die nicht von der speziellen Folge der G_n abhängt. Wird die Statistik

$$R_n(X_1, \ldots, X_n; F) := n(w(z^{(n)}) - w(\pi))^T(w(z^{(n)}) - w(\pi))$$

verwendet, so kann ihre Verteilung durch die Verteilung von

$$R_n(X_1^*, \ldots, X_n^*; F_n) := n(w(z_*^{(n)}) - w(z^{(n)}))^T(w(z_*^{(n)}) - w(z^{(n)})),$$

gegeben $z^{(n)}$, vernünftig geschätzt werden, da die Verteilung von $R_n(X_1^*, \ldots, X_n^*; F_n)$, gegeben F_n, nach dem Theorem 1 von Beran (1984) für fast alle Folgen (F_n) gegen die Verteilung der rechten Seite aus (6.5) strebt.

Unter der Nullhypothese $\pi = h(\theta_0)$ vereinfacht sich $R_n(X_1, \ldots, X_n; F)$ zur Prüfgröße $n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)})))$. Sei $c_n(\alpha, F_n)$ wieder ein über die Verteilung von $R_n(X_1^*, \ldots, X_n^*; F_n)$ gebildeter Schätzer des $1-\alpha$-Quantils, dann folgt:

Lemma 6.1 *Ist $\pi = h(\theta_0)$, so gilt:*

$$\lim_{n\to\infty} P_F(n\Delta(z, h(\hat{\theta}(z))) > c_n(\alpha, F_n)) = \alpha$$

Die Hypothese $H_0 : \pi = h(\theta_0)$ wird somit verworfen, falls

$$n\Delta(z, h(\hat{\theta}(z))) > c_n(\alpha, F_n)$$

ist.

6.2.3.2 Gauß-Diskrepanz

Aus den Ergebnissen des Abschnitts 4.3.4 folgt, daß sich in der eben genannten Weise Bootstrap-Anpassungstests z. B. für die Neyman-χ^2- und die Freeman-Tukey-Diskrepanz konstruieren lassen. Die zugehörigen Prüfgrößen sind jedoch, zumindest, wenn BAN-Schätzer eingesetzt werden, nach Abschnitt 4.3.1.1 asymptotisch χ^2_{k-s-1}-verteilt, die kritischen Werte ließen sich somit aus der asymptotischen Verteilung ablesen. Nur wenn Schätzer zu einer Diskrepanz $\Delta \in \mathcal{D} \setminus \mathcal{D}^*$ eingesetzt werden, ist die asymptotische Verteilung nicht mehr bekannt, so daß nur der Bootstrap-Anpassungstest angewendet werden kann.

Für die Gauß-Diskrepanz hingegen konnte bisher kein Anpassungstest hergeleitet werden, da die asymptotische Verteilung der Prüfgröße

$$n\Delta(z^{(n)}, h(\hat{\theta}(z^{(n)}))) = n\sum_{i=1}^{k}(z_i^{(n)} - h_i(\hat{\theta}(z^{(n)})))^2$$

auch unter der Nullhypothese vom unbekannten Modell π abhängt (siehe 4.3.1.2). Setzt man nun

$$R_n(X_1, \ldots, X_n; F) := n\sum_{i=1}^{k}\Big((z_i^{(n)} - h_i(\hat{\theta}(z^{(n)}))) - (\pi_i - h_i(\theta_0))\Big)^2, \qquad (6.6)$$

so kann als kritischer Wert ein Schätzer $c_n(\alpha, F_n)$ des $1-\alpha$-Prozentpunktes der Verteilung von

$$R_n(X_1^*, \ldots, X_n^*; F_n) := n\sum_{i=1}^{k}\Big((z_i^{*(n)} - h_i(\hat{\theta}(z^{*(n)}))) - (z_i^{(n)} - h_i(\hat{\theta}(z^{(n)})))\Big)^2,$$

gegeben $z^{(n)}$, verwendet werden. Die Nullhypothese wird verworfen, falls

$$n\sum_{i=1}^{k}(z_i^{(n)} - h_i(\hat{\theta}(z^{(n)})))^2 > c_n(\alpha, F_n)$$

ist.

Die Verwendung der Gauß-Diskrepanz gegenüber der χ^2-Diskrepanz besitzt außerdem den Vorteil, daß keinerlei Probleme entstehen, falls Komponenten des Vektors $h(\hat{\theta}(z))$ verschwinden. Der Wert der Prüfgröße ist immer definiert.

6.2.3.3 Simulationen zum Gauß-Bootstrap-Anpassungstest

Es werden nun Simulationsergebnisse zum Gauß-Bootstrap-Anpassungstest gegeben. Die Durchführung der Simulation ist analog zu Abschnitt 5.3.3 gestaltet, und es wird wieder die Modellfamilie $\{h(\theta) : \theta \in [0; 0.2]\}$ mit

$$h(\theta) = (\theta, \theta, \theta, \theta, \theta, 1 - 5\theta)^T$$

betrachtet. Abweichend zu Abschnitt 5.3.3 werden Minimum-Gauß-Schätzer eingesetzt, die in diesem Fall identisch mit den Maximum-Likelihood-Schätzern

$$\hat{\theta}(z) = \frac{1}{5}\sum_{i=1}^{5} z_i$$

sind. Im ersten Beispiel ist

$$\pi = (0.161235, 0.161235, 0.161235, 0.161235, 0.161235, 0.193825)^T,$$

im zweiten wird wieder

$$\pi = (0.02, 0.02, 0.02, 0.02, 0.02, 0.9)^T$$

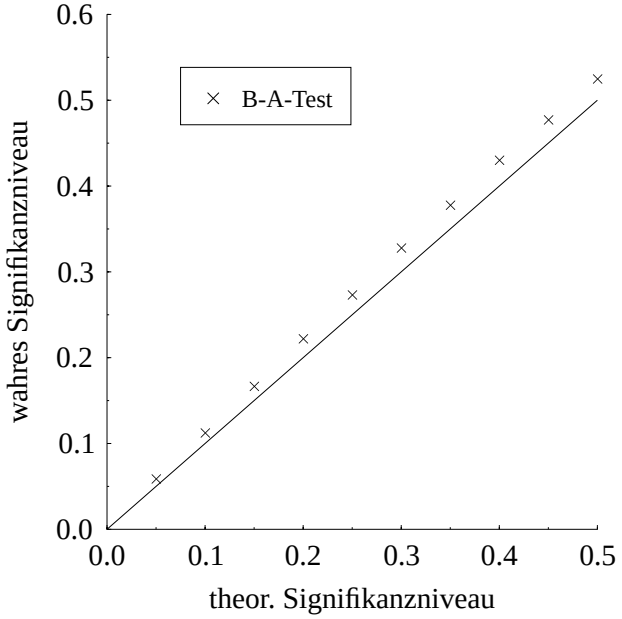
verwendet. In den Abbildungen 6.1 und 6.2 sind die Schätzer der wahren Signifikanzniveaus zu den Stichprobengrößen $n = 50$ und $n = 200$ (Beispiel 1) bzw. $n = 50$ und $n = 400$ (Beispiel 2) wiedergegeben. Die Plots zeigen, daß die Schätzer der tatsächlichen Signifikanzniveaus bei Beispiel 1 recht schnell gegen die gewünschten Niveaus konvergieren. Wird das Grundmodell mit kleinen Wahrscheinlichkeiten aus Beispiel 2 genommen, so ergibt sich eine im Vergleich zum χ^2-Bootstrap-Anpassungstest (vgl. Abbildungen 5.11 und 5.12) weitaus bessere Annäherung an die angenommenen Signifikanzniveaus.

Beispiel 1

n = 50

θ_0 = 0.161235

B = 200



Beispiel 1

n = 200

θ_0 = 0.161235

B = 200

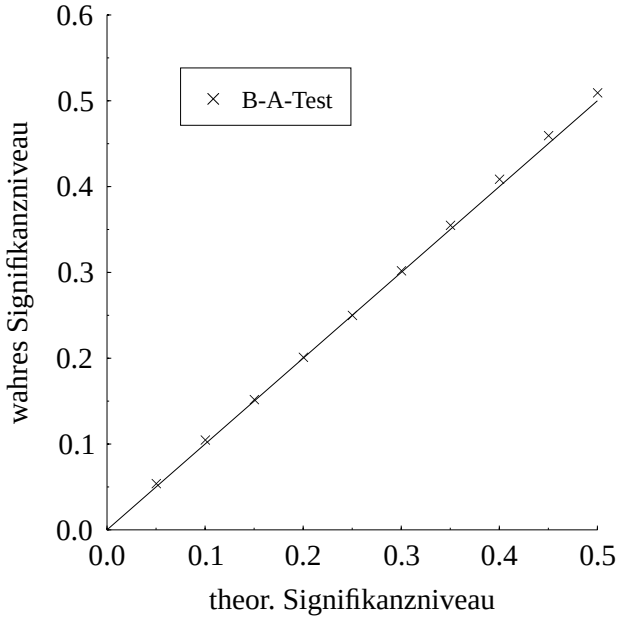


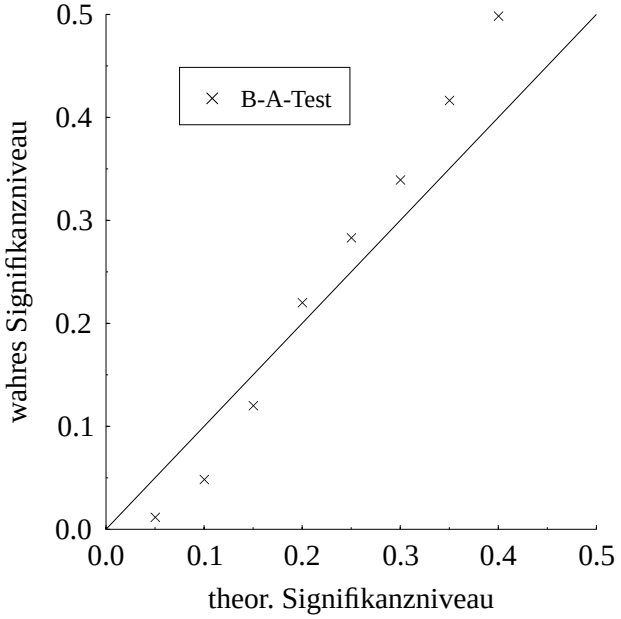
Abb. 6.1: Schätzer der wahren Signifikanzniveaus - Bsp. 1 - n=50, 200

Beispiel 2

n = 50

θ_0 = 0.02

B = 200



Beispiel 2

n = 400

θ_0 = 0.02

B = 200

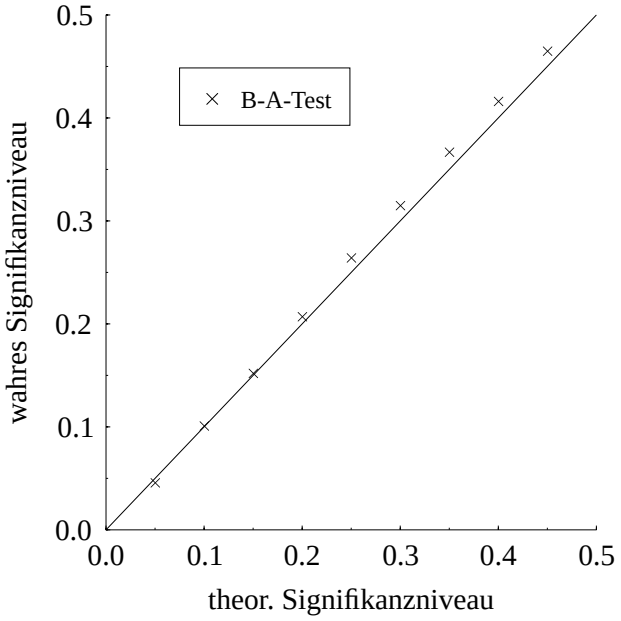


Abb. 6.2: Schätzer der wahren Signifikanzniveaus - Bsp. 2 - n=50, 400

6.3 Zwei Modellfamilien: Test über die Differenz der Diskrepanzen durch Approximation

Im Falle zweier Modellfamilien $\{h^{(i)}(\theta^{(i)}) : \theta^{(i)} \in \Theta^{(i)}\}$, $i = 1, 2$, wurde in Abschnitt 5.4 für die χ^2-Diskrepanz ein Bootstrap-Test zur Hypothese

$$H_0 : \Delta^{(1)} - \Delta^{(2)} \leq d, \quad d \in I\!R,$$

hergeleitet. Ein analoger Test läßt sich für beliebige Diskrepanzen $\Delta \in \mathcal{D}$ mit

$$\Delta(\pi, h(\theta)) = \sum_{i=1}^{k} g(\pi_i, h_i(\theta))$$

ableiten. Der Einfachheit halber sei $\Delta_g = \Delta$. Mit den abkürzenden Schreibweisen

$$\begin{aligned} \Delta^{(i)} &:= \Delta(\pi, h^{(i)}(\theta_0^{(i)})), \quad i = 1, 2, \text{ und} \\ \Delta_n^{(i)}(\hat{\theta}^{(i)}) &= \Delta(z^{(n)}, h^{(i)}(\hat{\theta}^{(i)}(z^{(n)}))), \quad i = 1, 2, \end{aligned}$$

definiere man die vorläufige Statistik

$$T_n(X_1, \ldots, X_n; F) := \sqrt{n} \left((\Delta_n^{(1)}(\hat{\theta}^{(1)}) - \Delta^{(1)}) - (\Delta_n^{(2)}(\hat{\theta}^{(2)}) - \Delta^{(2)}) \right),$$

Unter der Voraussetzung, daß das wahre Modell π in keiner der beiden Modellfamilien enthalten ist, folgt mittels der Delta-Methode und dem Ergebnis aus Abschnitt 4.3.2 zunächst:

$$\sqrt{n} \begin{pmatrix} \Delta_n^{(1)}(\hat{\theta}^{(1)}) - \Delta^{(1)} \\ \Delta_n^{(2)}(\hat{\theta}^{(2)}) - \Delta^{(2)} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right),$$

wobei

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} G_x^{(1)} \\ G_x^{(2)} \end{pmatrix} \left(D_\pi - \pi\pi^T \right) \begin{pmatrix} (G_x^{(1)})^T \\ (G_x^{(2)})^T \end{pmatrix}$$

ist. Die $1 \times k$-Matrizen $G_x^{(i)}$, $i = 1, 2$, sind analog zu Abschnitt 4.3.2 durch

$$(G_x^{(i)})_j = \left. \frac{\partial g(x, y)}{\partial x} \right|_{(x,y)=(\pi_j, h_j^{(i)}(\theta_0^{(i)}))}, \quad j = 1, \ldots, k, \tag{6.7}$$

definiert. Es ergibt sich somit

$$T_n(X_1, \ldots, X_n; F) \xrightarrow{d} N(0, \sigma^2)$$

mit $\sigma^2 = \sigma_1^2 - 2\sigma_{12} + \sigma_2^2$. Es muß nun angenommen werden, daß σ^2 positiv ist.

Seien nun die von p abhängigen $1 \times k$-Matrizen $G_x^{(i)}(p)$, $i = 1, 2$, mit

$$(G_x^{(i)}(p))_j = \left.\frac{\partial g(x,y)}{\partial x}\right|_{(x,y)=(p_j, h_j^{(i)}(\hat{\theta}^{(i)}(p)))}, \quad j = 1, \ldots, k,$$

definiert. Seien weiterhin die Funktionen $\sigma_1^2(p)$, $\sigma_2^2(p)$ und $\sigma_{12}(p)$ durch

$$\begin{aligned}
\sigma_1^2(p) &:= (G_x^{(1)}(p))(D_p - pp^T)(G_x^{(1)}(p))^T \\
\sigma_2^2(p) &:= (G_x^{(2)}(p))(D_p - pp^T)(G_x^{(2)}(p))^T \\
\sigma_{12}(p) &:= (G_x^{(1)}(p))(D_p - pp^T)(G_x^{(2)}(p))^T
\end{aligned}$$

definiert, dann ist die Funktion

$$\sigma^2(p) := \sigma_1^2(p) - 2\sigma_{12}(p) + \sigma_2^2(p) \tag{6.8}$$

stetig in π mit $\sigma^2(\pi) = \sigma^2$. Man definiere nun die endgültige Statistik

$$R_n(X_1, \ldots, X_n; F) := T_n(X_1, \ldots, X_n; F)/(\sigma^2(z^{(n)}))^{\frac{1}{2}}.$$

Es ergibt sich dann der Satz:

Satz 6.3 *Seien* **B1-B5** *für die beiden Modellfamilien erfüllt. Sei* (G_n) *eine Folge aus* $\mathcal{G}_F$ *und* $(p^{(n)})$ *die dazugehörige Folge von Wahrscheinlichkeitsvektoren. Sei weiterhin* $(z_*^{(n)})$ *eine Folge von Vektoren der relativen Häufigkeiten, wobei jedes* $z_*^{(n)}$ *unter* G_n *erzeugt wird. Sei* $\sigma^2(p)$ *wie in (6.8) definiert. Dann gilt:*

$$\sqrt{n}\frac{(\Delta_n^{*(1)} - \Delta_n^{*(2)}) - (\Delta_n^{(1)} - \Delta_n^{(2)})}{(\sigma^2(z_*^{(n)}))^{\frac{1}{2}}} \xrightarrow{d} N(0,1),$$

wobei die folgenden Definitionen getroffen werden:

$$\begin{aligned}
\Delta_n^{*(i)} &:= \Delta(z_*^{(n)}, h^{(i)}(\hat{\theta}^{(i)}(z_*^{(n)}))), \quad i = 1, 2, \\
\Delta_n^{(i)} &:= \Delta(p^{(n)}, h^{(i)}(\hat{\theta}^{(i)}(p^{(n)}))), \quad i = 1, 2.
\end{aligned}$$

Beweis: Zunächst zeigt man, analog zu der Herleitung der asymptotischen Verteilung von $T_n(X_1, \ldots, X_n; F)$, mittels der Delta-Methode aus Satz 2.3:

$$\sqrt{n}\left((\Delta_n^{*(1)} - \Delta_n^{*(2)}) - (\Delta_n^{(1)} - \Delta_n^{(2)})\right) \xrightarrow{d} N(0, \sigma^2).$$

Wie in Lemma 5.2 folgt:

$$(\sigma^2(z_*^{(n)}))^{\frac{1}{2}} \xrightarrow{p} (\sigma^2)^{\frac{1}{2}}$$

Die Anwendung von Slutskys Theorem (Serfling, 1980, S. 19) beendet den Beweis. □

Somit sind die Voraussetzungen des Theorems 1 von Beran (1984) erfüllt. Sei $c_n(\alpha, F_n)$ ein Schätzer des $1 - \alpha$-Quantils der Verteilung von $R_n(X_1^*, \ldots, X_n^*; F_n)$, dann ergibt sich die Konvergenzaussage:

$$\lim_{n\to\infty} P_F(R_n(X_1, \ldots, X_n; F) > c_n(\alpha, F_n)) = \alpha$$

Unter der Hypothese

$$H_0 : \Delta^{(1)} = \Delta^{(2)} + d, \quad d \in I\!R$$

folgt

$$\lim_{n\to\infty} P_F(\sqrt{n}\frac{\Delta_n^{(1)}(\hat{\theta}^{(1)}) - \Delta_n^{(2)}(\hat{\theta}^{(2)}) - d}{(\sigma^2(z^{(n)}))^{\frac{1}{2}}} > c_n(\alpha, F_n)) = \alpha$$

Da ein einseitiger Verwerfungsbereich festgelegt wurde, ist der genannte Test auch für die Hypothese

$$H_0 : \Delta^{(1)} \leq \Delta^{(2)} + d, \quad d \in I\!R$$

verwendbar. Die Wahl eines geeigneten Wertes von d kann wieder zu Schwierigkeiten führen, für den Fall $d = 0$ läßt sich der Test ohne Probleme durchführen.

Kapitel 7

Schlußwort

Im ersten Teil der vorliegenden Arbeit sollten verbesserte Approximationen zur Berechnung der exakten Macht bei Anwendung des χ^2-Anpassungstests hergeleitet werden. Dieser Anspruch kann nun mit Hilfe des aus den Ergebnissen von Drost et al. (1989) entwickelten Ansatzes befriedigt werden. Im Falle der einfachen Hypothese kann man alternativ die in der Literatur vernachlässigte Approximation von Frosini (1976) einsetzen. Die Betrachtung der asymptotischen Verteilungen unter „local alternatives“ ermöglicht eine Beurteilung der Eignung zur Approximation der Macht. Diese asymptotischen Verteilungen konnten für die einfache Hypothese ermittelt werden.

Der Gebrauch der multivariaten Delta-Methode in Verbindung mit dem Satz über implizite Funktionen gestattet eine allgemeine Herleitung von asymptotischen Resultaten. Die explizite Darstellung der Schätzer $\hat{\theta}(z)$ in Abhängigkeit von z ist für eine Anwendung nicht erforderlich.

Die am Beispiel der χ^2-Diskrepanz illustrierten Ideen konnten auf andere Diskrepanzen bzw. Prüfgrößen für Anpassungstests übertragen werden.

Für das Testen von Hypothesen über die Diskrepanz durch Approximation konnten Bootstrap-Tests konstruiert werden, die zumindest im Prinzip funktionieren. Etwaige Schwächen sind bei alternativen asymptotischen Tests stärker ausgeprägt, so daß der Einsatz von Bootstrap-Methoden als Fortschritt anzusehen ist.

Es gelang, Bootstrap-Anpassungstests zu entwickeln, die auch dann eingesetzt werden können, wenn klassische Methoden versagen.

Anhang A

Beweise

A.1 Beweis von Satz 2.3

Der Beweis von Satz 2.3 benötigt einige formale Vorbereitungen. Für $n = 1, 2, 3, \ldots$, sei $(\Omega_n, \mathcal{A}_n, \mathcal{P}_n)$ ein Wahrscheinlichkeitsraum, bestehend aus einem Stichprobenraum $\Omega_n \subset I\!R^k$, einer σ-Algebra $\mathcal{A}_n$ von Teilmengen von Ω_n und einem auf $\mathcal{A}_n$ definierten Wahrscheinlichkeitsmaß $\mathcal{P}_n$. Sei ω_n eine abstrakte Zufallsvariable mit Werten in Ω_n, die nach $\mathcal{P}_n$ verteilt ist. Außerdem sei H_n eine meßbare Abbildung von Ω_n in den $I\!R^l$. Die Theoreme 14.4-4 und 14.4-5 von Bishop et al. (1975) besagen nun:

Lemma A.1 *Sei b_n eine Folge reeller Zahlen. Unter den genannten Voraussetzungen ist $\tau_n = H_n(\omega_n) = o_p(b_n)[bzw. O_p(b_n)]$ genau dann, wenn für jedes $\eta > 0$ Mengen $S_n \in \mathcal{A}_n$ existieren mit*

1. $\mathcal{P}_n(\omega \in S_n) \geq 1 - \eta \quad \forall n$
2. *Ist w_n eine nichtstochastische Folge, so daß $w_n \in S_n \quad \forall n$, dann ist $t_n := H_n(w_n) = o(b_n)[bzw. O(b_n)]$*

Nun kann der Satz 2.3 bewiesen werden:
Satz 2.3 *Sei $x_0 \in I\!R^k$ und sei $x_n \in I\!R^k$, $n = 1, 2, 3, \ldots$ eine Folge von Vektoren mit $lim_{n\to\infty} x_n = x_0$. Sei weiterhin $X_n, n = 1, 2, 3, \ldots$, eine Folge von k-dimensionalen Zufallsvektoren mit einer asymptotischen Normalverteilung in dem Sinne, daß*

$$\sqrt{n}(X_n - x_n) \xrightarrow{d} N(\mu(x_0), \Sigma(x_0))$$

ist. Nun sei $f(x)$ eine Abbildung vom $\mathbb{R}^k$ in den $\mathbb{R}^l$, die in einer offenen Umgebung $U(x_0) \subset \mathbb{R}^k$ stetig differenzierbar ist. Bezeichne $F(x)$ die Jacobi-Matrix für $x \in U(x_0)$, dann folgt für die asymptotische Verteilung:

$$\sqrt{n}(f(X_n) - f(x_n)) \xrightarrow{d} N(F(x_0)\mu(x_0), F(x_0)\Sigma(x_0)F(x_0)^T)$$

Beweis:
1. Sei für $n = 1, 2, 3, \ldots$, $(\Omega_n, \mathcal{A}_n, \mathcal{P}_n)$ ein Wahrscheinlichkeitsraum und sei X_n mit ω_n identifiziert. Aus der Voraussetzung

$$\sqrt{n}(X_n - x_n) \xrightarrow{d} N(\mu(x_0), \Sigma(x_0))$$

folgt nach Theorem 14.4.-2 von Bishop et al. (1975):

$$\sqrt{n}(X_n - x_n) = O_p(1)$$

Dies ist nach Lemma A.1 äquivalent zur Existenz von Mengen $S_n \in \mathcal{A}_n$, so daß für jedes $\eta > 0$ gilt:

a) $\mathcal{P}_n(X_n \in S_n) \geq 1 - \eta \quad \forall n$

b) Ist w_n eine nichtstochastische Folge, so daß $w_n \in S_n \quad \forall n$, dann ist $\sqrt{n}(w_n - x_n) = O(1)$

2. Sei w_n eine solche Folge. Aus $\lim_{n\to\infty} x_n = x_0$ folgt $\lim_{n\to\infty} w_n = x_0$. Es folgt die Existenz eines N, so daß x_n, $w_n \in U(x_0) \quad \forall n \geq N$. Für $n \geq N$ gilt dann für die Komponenten f_i der Abbildung f die Taylorentwicklung

$$f_i(w_n) = f_i(x_n) + \sum_{j=1}^{k} \frac{\partial f_i}{\partial x_j}(v_n^i)((w_n)_j - (x_n)_j),$$

mit $v_n^i = w_n + \lambda_{ni}[x_n - w_n]$, $\lambda_{ni} \in [0, 1]$, $i = 1, \ldots, l$. Da $\lim_{n\to\infty}(x_n - w_n) = 0$ ist, ergibt sich $\lim_{n\to\infty} v_n^i = x_0$, $i = 1, \ldots, l$.
3. Da $f(x)$ nach Voraussetzung in $U(x_0)$ stetig differenzierbar ist, folgt somit

$$f_i(w_n) = f_i(x_n) + \sum_{j=1}^{k}(\frac{\partial f_i}{\partial x_j}(x_0) + o(1))((w_n)_j - (x_n)_j), \quad w_n, x_n \to x_0$$

4. Kombiniert man 1.-3., dann folgt

$$\begin{aligned} f(w_n) &= f(x_n) + F(x_0)(w_n - x_n) + o(1)O(n^{-\frac{1}{2}}), \quad w_n, x_n \to x_0 \\ &= f(x_n) + F(x_0)(w_n - x_n) + o(n^{-\frac{1}{2}}), \quad w_n, x_n \to x_0 \end{aligned}$$

5. Definiere

$$H_n(w_n) := f(w_n) - f(x_n) - F(x_0)(w_n - x_n),$$

dann ist $H_n(w_n) = o(n^{-\frac{1}{2}})$. Die Anwendung des obenstehenden Lemmas, diesmal in umgekehrter Richtung, ergibt

$$\begin{aligned} H_n(X_n) &= f(X_n) - f(x_n) - F(x_0)(X_n - x_n) + o_p(n^{-\frac{1}{2}}) \\ \Rightarrow \sqrt{n}(f(X_n) - f(x_n)) &= F(x_0)\sqrt{n}(X_n - x_n) + o_p(1) \end{aligned} \tag{A.1}$$

6. Es folgt aus Satz 2.1

$$F(x_0)\sqrt{n}(X_n - x_n) \xrightarrow{d} N(F(x_0)\mu, F(x_0)\Sigma(x_0)F(x_0)^T)$$

und aus (A.1) die Behauptung. □

Ist $x_n = x_0$, $\forall n$, dann muß Schritt 2 des Beweises in

$$f(w_n) = f(x_0) + F(x_0)(w_n - x_0) + o(\| w_n - x_0 \|), \quad w_n \to x_0,$$

geändert werden, Schritt 3 wird dann nicht mehr benötigt, und es genügt die totale Differenzierbarkeit von $f(x)$ in x_0. Anschließend wird der Beweis ab der letzten Zeile aus 4. analog fortgesetzt.

A.2 Beweis von Satz 3.1

Satz 3.1 *Unter* **B1-B5** *gibt es eine offene δ^*-Umgebung $U_{\delta^*}(\pi) \subset (0,1)^k$, so daß für jedes $p \in U_{\delta^*}(\pi)$ genau ein MDP $\hat{\theta}(p)$ existiert.*

Beweis:
1. Sei $G^* = G^*(\pi) \subset (0,1)^k$ eine offene Umgebung von π bezüglich des $I\!R^k$, und $H^* = H^*(\theta_0)$ sei wie in **B4** definiert. Sei $F^* : G^* \times H^* \to I\!R^s$ definiert durch:

$$F_i^*(p,\theta) := \frac{\partial}{\partial \theta_i}\Delta_g(p, h(\theta)), \quad i = 1, \ldots, s.$$

F^* ist nach Definition 3.1 und **B4** stetig differenzierbar mit

$$F^*(p,\theta)|_{(p,\theta)=(\pi,h(\theta_0))} = \mathbf{0}.$$

Sei $\Omega(p,\theta)$ die $s \times s$-Matrix der zweiten Ableitungen mit

$$(\Omega(p,\theta))_{ij} := \frac{\partial^2 \Delta_g(p,h(\theta))}{\partial\theta_i \partial\theta_j}, \quad i,j = 1,\ldots,s,$$

dann ist $\Omega = \Omega(\pi,\theta_0)$ nach **B5** invertierbar.

2. $\Omega(p,\theta)$ ist, als Abbildung gesehen, stetig auf $G^* \times H^*$. Da $\Omega(p,\theta)$ symmetrisch ist, eine symmetrische Matrix genau dann positiv definit ist, wenn ihre Eigenwerte positiv sind (Press, 1972) und die Eigenwerte einer Matrix stetig von den Komponenten abhängen (Werner, 1992), gibt es eine offene Umgebung $G = G(\pi) \subset G^*(\pi)$ und eine offene Umgebung $H = H(\theta_0) \subset H^*(\theta_0)$, so daß $\Omega(p,\theta)$ für alle $(p,\theta) \in G \times H$ positiv definit ist.

3. Sei $F := F^*|G \times H$, so existieren nach Heuser (1981, S. 292) die offenen Umgebungen $U_\delta(\pi) \subset G$, $V_\epsilon(\theta_0) \subset H$ und genau eine stetige Funktion

$$\begin{array}{lll} & & \hat\theta : U_\delta(\pi) \to V_\epsilon(\theta_0) \\ \text{mit} & a) & \hat\theta(\pi) = \theta_0 \\ & b) & F(p,\hat\theta(p)) = \mathbf{0} \quad \forall p \in U_\delta(\pi). \end{array}$$

4. Sei $p \in U_\delta(\pi)$, $\hat\theta = \hat\theta(p)$ und $\theta \in V_\epsilon(\theta_0)$, dann gilt die Taylorentwicklung:

$$\Delta_g(p,h(\theta)) = \Delta_g(p,h(\hat\theta)) + (\theta-\hat\theta)^T \Delta_g'(p,h(\hat\theta)) + \frac{1}{2}(\theta-\hat\theta)^T \Omega(p,\theta^*)(\theta-\hat\theta),$$

mit $\theta^* = \theta + t(\hat\theta - \theta)$ für ein $t \in [0,1]$. Für festes $p \in U_\delta(\pi)$ nimmt $\Delta_g(p,h(\theta))|V_\epsilon(\theta_0)$ daher das Minimum eindeutig in $\theta = \hat\theta$ an, denn $\Omega(p,\theta^*)$ ist positiv definit und $\Delta_g'(p,h(\hat\theta)) = F(p,\hat\theta(p))$ verschwindet.

5. Sei $V^C := \Theta \setminus V_\epsilon(\theta_0)$, V^C ist dann kompakt, und sei die Funktion $m : U_\delta(\pi) \to I\!R$ definiert durch

$$m(p) := \min_{\theta \in V^C} \Delta_g(p,h(\theta)).$$

$m(p)$ existiert für alle $p \in U_\delta(\pi)$, da $\Delta_g(p,h(\theta))$ stetig in θ ist. Außerdem ist $m(p)$ selbst stetig auf der lokalkompakten Menge $U_\delta(\pi)$.

6. Da nach **B2** θ_0 eindeutig ist, gibt es ein $\alpha > 0$, so daß

$$m(\pi) > \Delta_g(\pi,h(\theta_0)) + \alpha$$

ist.

7. Aus 5. folgt die Existenz einer δ_1-Umgebung $U_{\delta_1}(\pi) \subset U_\delta(\pi)$ mit

$$m(p) > m(\pi) - \frac{\alpha}{2}, \quad \forall p \in U_{\delta_1}(\pi).$$

8. Aus der Stetigkeit von $\hat{\theta}(p)$ kann auf die Existenz einer offenen Umgebung $U_{\delta^*}(\pi) \subset U_{\delta_1}(\pi)$ mit

$$\Delta_g(p, h(\hat{\theta}(p))) < \Delta_g(\pi, h(\theta_0)) + \frac{\alpha}{2}$$

für alle $p \in U_{\delta^*}(\pi)$ geschlossen werden.

9. Sei nun $p \in U_{\delta^*}(\pi)$ und $\theta \in \Theta \setminus \{\hat{\theta}(p)\}$ beliebig.
Für $\theta \in V_\epsilon(\theta_0)$ ist nach 4.

$$\Delta_g(p, h(\theta)) > \Delta_g(p, h(\hat{\theta}(p))).$$

Für $\theta \in V^C$ ist nach 5., 6., 7. und 8.

$$\Delta_g(p, h(\theta)) \geq m(p) > m(\pi) - \frac{\alpha}{2} > \Delta_g(\pi, h(\theta_0)) + \frac{\alpha}{2} > \Delta_g(p, h(\hat{\theta}(p))).$$

$\Rightarrow$ Für $p \in U_{\delta^*}(\pi)$ wird das absolute Minimum von $\Delta_g(p, h(\theta))$ eindeutig in $\theta = \hat{\theta}(p)$ angenommen. $\square$

A.3 Beweis des Lemmas 5.4

Lemma 5.4 *Sei die Funktion* $h : I\!N \times [0, \infty) \times I\!R \to [0; 1]$ *definiert durch*

$$h(n, c, d) = P(\chi_n^2(c) \leq 2c - d).$$

Dann ist $h(n, c, 0)$ *für festes, aber beliebiges* $n \in I\!N$ *streng monoton steigend in* c *mit* $h(n, 0, 0) = 0$ *und* $\lim_{c\to\infty} h(n, c, 0) = 1$. *Weiterhin gibt es zu* α *mit* $0 < \alpha < 1$ *genau ein* $c^* := c^*(n, \alpha)$ *mit* $h(n, c^*, 0) = 1 - \alpha$.
Beweis:
1. Seien $n = 1$, $d \geq 0$ und Z eine standardnormalverteilte Zufallsvariable mit Verteilungsfunktion $\Phi(z)$. Dann gilt für $c \geq d/2$:

$$\begin{aligned}
\Rightarrow h(1, c, d) &= P(\chi_1^2(c) \leq 2c - d) \\
&= P((Z - \sqrt{c})^2 \leq 2c - d) \\
&= P(-\sqrt{2c - d} + \sqrt{c} \leq Z \leq \sqrt{2c - d} + \sqrt{c}) \\
&= \Phi(\sqrt{2c - d} + \sqrt{c}) + \Phi(\sqrt{2c - d} - \sqrt{c}) - 1
\end{aligned}$$

Nun sei $g(c) := \sqrt{2c-d} - \sqrt{c}$ für $c > d/2 > 0$.

$$\Rightarrow g'(c) = \frac{1}{\sqrt{2c-d}} - \frac{1}{2\sqrt{c}} > \frac{1}{\sqrt{2c-d}} - \frac{1}{\sqrt{2c}} > 0$$

$\Rightarrow h(1,c,d)$ ist für festes d streng monoton steigend für $c \in (d/2, \infty)$. Außerdem ist $h(1, \frac{d}{2}, d) = 0$ und, da

$$\lim_{c\to\infty} g(c) = \lim_{c\to\infty} \sqrt{c}(\sqrt{2 - \frac{d}{c}} - 1) = +\infty$$

ist, folgt ebenso:

$$\lim_{c\to\infty} h(1,c,d) = 1$$

Diese Aussage behält natürlich auch für $d = 0$ Gültigkeit. Da $\Phi(x)$ stetig in x ist, ist $h(1,c,d)$ stetig in c. Somit gibt es zu $\alpha \in (0;1)$ genau ein $c^* = c^*(1,\alpha)$ mit $h(1,c^*,0) = 1-\alpha$.

2. Für den Fall $n > 1$ nutzt man die Tatsache aus, daß die Summe von unabhängig χ^2-verteilten Zufallsvariablen wieder χ^2-verteilt ist, wobei sich die Freiheitsgrade und die Nichtzentralitätsparameter einfach summieren. Im vorliegenden Fall wird eine $\chi^2_n(c)$-verteilte Zufallsvariable dargestellt als Summe einer χ^2_{n-1}- und einer $\chi^2_1(c)$-verteilten Zufallsvariablen. Bezeichne $f_{n,c}(x)$ die Dichtefunktion einer $\chi^2_n(c)$-verteilten Zufallsvariablen. Dann läßt sich $h(n,c,0)$ folgendermaßen darstellen:

$$\begin{aligned} h(n,c,0) &= P(\chi^2_n(c) \le 2c) \\ &= \int_0^{2c}\int_0^{2c-y} f_{1,c}(x) f_{n-1,0}(y)\,dx\,dy \\ &= \int_0^{2c} f_{n-1,0}(y) \int_0^{2c-y} f_{1,c}(x)\,dx\,dy \\ &= \int_0^{2c} f_{n-1,0}(y) h(1,c,y)\,dy \end{aligned}$$

Für $0 < c_1 < c_2$ folgt dann unter der Verwendung der Ergebnisse aus 1. die Abschätzung:

$$\begin{aligned} h(n,c_1,0) &= \int_0^{2c_1} f_{n-1,0}(y) h(1,c_1,y)\,dy \\ &< \int_0^{2c_1} f_{n-1,0}(y) h(1,c_2,y)\,dy \\ &< \int_0^{2c_2} f_{n-1,0}(y) h(1,c_2,y)\,dy = h(n,c_2,0) \end{aligned}$$

$h(n, c, 0)$ ist deshalb streng monoton steigend in c, außerdem folgt aus der Definition von h sofort $h(n, 0, 0) = 0$.
Sei nun $\epsilon > 0$ vorgegeben. Da $f_{n-1,0}(y)$ eine Dichtefunktion ist, gibt es ein c_1 mit

$$\int_0^{c_1} f_{n-1,0}(y)dy = \sqrt{1-\epsilon}$$

Da $h(1, c, c_1)$ für festes c_1 monoton steigend mit Grenzwert 1 ist, gibt es ein $c_2 > c_1$ mit

$$h(1, c_2, c_1) > \sqrt{1-\epsilon}.$$

Diese Abschätzung besitzt auch für $h(1, c_2, y)$, $y \in [0, c_1]$, Gültigkeit, da $h(1, c, y)$ monoton fallend in y ist. Insgesamt ergibt sich deshalb die folgende Abschätzung:

$$\begin{aligned} h(n, c_2, 0) &= \int_0^{2c_2} f_{n-1,0}(y)h(1, c_2, y)dy \\ &> \int_0^{c_1} f_{n-1,0}(y)h(1, c_2, y)dy \\ &> \int_0^{c_1} f_{n-1,0}(y)\sqrt{1-\epsilon}dy \\ &\geq \sqrt{1-\epsilon}\sqrt{1-\epsilon} = 1-\epsilon \end{aligned}$$

Da $\epsilon > 0$ beliebig war und $h(n, c, 0)$ streng monoton steigend in c ist, bedeutet dies, daß

$$\lim_{c\to\infty} h(n, c, 0) = 1$$

sein muß. Außerdem ist $h(n, c, 0)$ stetig in c, d. h., es gibt zu $\alpha \in (0; 1)$ genau ein $c^* = c^*(n, \alpha)$ mit $h(n, c^*, 0) = 1 - \alpha$. $\square$

Anhang B

Tabellen zu Abschnitt 2.5

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	0.0247	0.0256	**0.0233**	0.0385	0.0905	0.1538
24	$\emptyset$ r. A.	0.1614	0.1663	**0.1534**	0.2486	0.5600	0.9744
m. M.	$\emptyset$ a. A.	0.0326	0.0347	**0.0305**	0.0440	0.0939	0.1232
50	$\emptyset$ r. A.	0.1377	0.1462	**0.1289**	0.1848	0.3933	0.5241
h. M.	$\emptyset$ a. A.	**0.0312**	0.0384	0.0390	0.0828	0.0334	0.1039
19	$\emptyset$ r. A.	**0.2701**	0.3144	0.3839	1.5709	0.4205	1.1084

Tab. B.1: Auswertung: $\alpha = 0.10$, $n = 10$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0111**	0.0124	0.0129	0.0203	0.0734	0.1765
13	$\emptyset$ r. A.	**0.0718**	0.0790	0.0821	0.1247	0.4449	1.1011
m. M.	$\emptyset$ a. A.	0.0217	0.0240	**0.0187**	0.0437	0.0918	0.1020
51	$\emptyset$ r. A.	0.0902	0.0996	**0.0775**	0.1840	0.3852	0.4299
h. M.	$\emptyset$ a. A.	**0.0093**	0.0130	0.0114	0.0421	0.0205	0.0848
29	$\emptyset$ r. A.	**0.1365**	0.1788	0.1797	2.6191	0.2927	0.9764

Tab. B.2: Auswertung: $\alpha = 0.10$, $n = 30$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	∅ a. A.	**0.0063**	0.0072	0.0067	0.0160	0.0732	0.1643
16	∅ r. A.	**0.0372**	0.0421	0.0389	0.0974	0.4254	1.0452
m. M.	∅ a. A.	0.0152	0.0170	**0.0116**	0.0287	0.0990	0.1114
52	∅ r. A.	0.0650	0.0728	**0.0496**	0.1238	0.4217	0.4797
h. M.	∅ a. A.	**0.0057**	0.0120	0.0115	0.0285	0.0258	0.0851
28	∅ r. A.	**0.0980**	0.1959	0.1839	2.7090	1.2713	2.5075

Tab. B.3: Auswertung: $\alpha = 0.10$, $n = 50$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	∅ a. A.	**0.0061**	0.0075	0.0070	0.0135	0.0705	0.1633
28	∅ r. A.	**0.0363**	0.0433	0.0427	0.0777	0.4121	1.0535
m. M.	∅ a. A.	**0.0087**	0.0124	0.0110	0.0153	0.0930	0.1168
48	∅ r. A.	**0.0365**	0.0524	0.0464	0.0660	0.3965	0.5026
h. M.	∅ a. A.	**0.0046**	0.0096	0.0098	0.0249	0.0200	0.0771
22	∅ r. A.	**0.0954**	0.2002	0.2090	0.8283	0.9829	1.9884

Tab. B.4: Auswertung: $\alpha = 0.10$, $n = 100$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0050**	0.0067	0.0067	0.0060	0.0558	0.1555
19	$\emptyset$ r. A.	**0.0356**	0.0448	0.0451	0.0422	0.3533	1.0860
m. M.	$\emptyset$ a. A.	**0.0051**	0.0073	0.0077	0.0096	0.0814	0.1063
47	$\emptyset$ r. A.	**0.0216**	0.0308	0.0312	0.0411	0.3433	0.4525
h. M.	$\emptyset$ a. A.	**0.0032**	0.0134	0.0106	0.0209	0.0203	0.0813
34	$\emptyset$ r. A.	**0.0321**	0.1599	0.1414	0.1908	0.3201	0.9338

Tab. B.5: Auswertung: $\alpha = 0.10$, $n = 500$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0142**	0.0162	0.0166	0.0347	0.0688	0.1654
17	$\emptyset$ r. A.	**0.0917**	0.1056	0.1089	0.2218	0.4250	1.1080
m. M.	$\emptyset$ a. A.	0.0280	0.0288	**0.0224**	0.0496	0.0966	0.1663
32	$\emptyset$ r. A.	0.1177	0.1211	**0.0938**	0.2087	0.4033	0.6992
h. M.	$\emptyset$ a. A.	**0.0129**	0.0266	0.0233	0.0694	0.0298	0.1390
15	$\emptyset$ r. A.	**0.1085**	0.3409	0.2473	0.8230	0.4995	1.6052

Tab. B.6: Auswertung: $\alpha = 0.10$, $n = 10$, $k = 4$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0061**	0.0063	0.0092	0.0192	0.0571	0.1758
26	$\emptyset$ r. A.	**0.0411**	0.0428	0.0680	0.1158	0.3527	1.1603
m. M.	$\emptyset$ a. A.	0.0182	0.0195	**0.0106**	0.0298	0.0913	0.1728
46	$\emptyset$ r. A.	0.0768	0.0824	**0.0450**	0.1262	0.3871	0.7401
h. M.	$\emptyset$ a. A.	**0.0080**	0.01326	0.0094	0.0405	0.0277	0.1260
22	$\emptyset$ r. A.	**0.0786**	0.1954	0.1903	0.4540	0.8655	2.0735

Tab. B.7: Auswertung: $\alpha = 0.10$, $n = 30$, $k = 4$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	0.0030	**0.0027**	0.0049	0.0138	0.0574	0.1906
11	$\emptyset$ r. A.	0.0187	**0.0175**	0.0341	0.0880	0.3437	1.1878
m. M.	$\emptyset$ a. A.	0.0108	0.0129	**0.0073**	0.0262	0.0821	0.1697
46	$\emptyset$ r. A.	0.0448	0.0540	**0.0309**	0.1107	0.3471	0.7213
h. M.	$\emptyset$ a. A.	**0.0060**	0.0088	0.0070	0.0385	0.0193	0.1128
29	$\emptyset$ r. A.	0.0974	0.1028	**0.0952**	1.6333	1.0967	2.7572

Tab. B.8: Auswertung: $\alpha = 0.10$, $n = 50$, $k = 4$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0161**	0.0195	0.0170	0.0322	0.0599	0.2000
19	$\emptyset$ r. A.	0.1065	0.1268	**0.1057**	0.2151	0.3520	1.2376
m. M.	$\emptyset$ a. A.	0.0163	0.0209	**0.0162**	0.0409	0.0905	0.2915
22	$\emptyset$ r. A.	**0.0704**	0.0912	0.0713	0.1760	0.3909	1.2525
h. M.	$\emptyset$ a. A.	**0.0051**	0.0134	0.0175	0.0937	0.0254	0.3519
2	$\emptyset$ r. A.	**0.0364**	0.1051	0.1212	0.6295	0.1788	2.3140

Tab. B.9: Auswertung: $\alpha = 0.10$, $n = 10$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0063**	0.0091	0.0118	0.0299	0.0622	0.2180
22	$\emptyset$ r. A.	**0.0351**	0.0495	0.0654	0.1648	0.3333	1.2688
m. M.	$\emptyset$ a. A.	0.0164	0.0189	**0.0091**	0.0297	0.0824	0.2744
46	$\emptyset$ r. A.	0.0686	0.0787	**0.0382**	0.1246	0.3455	1.1544
h. M.	$\emptyset$ a. A.	**0.0061**	0.0075	0.0132	0.0330	0.0220	0.1999
19	$\emptyset$ r. A.	**0.0493**	0.0827	0.1378	0.2791	0.4049	1.9433

Tab. B.10: Auswertung: $\alpha = 0.10$, $n = 50$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0044**	0.0062	0.0110	0.0285	0.0565	0.2094
22	$\emptyset$ r. A.	**0.0270**	0.0369	0.0680	0.1703	0.3325	1.2614
m. M.	$\emptyset$ a. A.	0.0145	0.0162	**0.0066**	0.0306	0.0764	0.2701
46	$\emptyset$ r. A.	0.0614	0.0684	**0.0278**	0.1327	0.3234	1.1526
h. M.	$\emptyset$ a. A.	0.0082	0.0091	**0.0066**	0.0355	0.0161	0.1910
23	$\emptyset$ r. A.	0.0677	0.0758	**0.0633**	0.3922	0.1959	1.6573

Tab. B.11: Auswertung: $\alpha = 0.10$, $n = 100$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0022**	0.0023	0.0033	0.0122	0.0361	0.1964
23	$\emptyset$ r. A.	**0.0158**	0.0161	0.0222	0.0763	0.2062	1.2417
m. M.	$\emptyset$ a. A.	0.0077	0.0079	**0.0055**	0.0129	0.0747	0.2870
47	$\emptyset$ r. A.	0.0320	0.0329	**0.0229**	0.0550	0.3128	1.2019
h. M.	$\emptyset$ a. A.	0.0049	0.0090	**0.0045**	0.0321	0.0169	0.1843
29	$\emptyset$ r. A.	**0.0369**	0.0882	0.0585	0.2455	0.2410	1.6776

Tab. B.12: Auswertung: $\alpha = 0.10$, $n = 500$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	∅ a. A.	**0.0085**	0.0117	0.0166	0.0292	0.0537	0.2245
9	∅ r. A.	**0.0447**	0.0608	0.0930	0.1646	0.2871	1.2964
m. M.	∅ a. A.	0.0171	0.0216	**0.0102**	0.0379	0.0706	0.3804
43	∅ r. A.	0.0721	0.0910	**0.0431**	0.1606	0.2967	1.6063
h. M.	∅ a. A.	0.0088	**0.0082**	0.0106	0.0383	0.0232	0.3020
31	∅ r. A.	0.0646	**0.0616**	0.0925	0.3075	0.2535	2.6281

Tab. B.13: Auswertung: $\alpha = 0.10$, $n = 100$, $k = 10$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	∅ a. A.	0.0242	0.0235	**0.0231**	0.0417	0.0483	0.0712
40	∅ r. A.	0.2366	0.2367	**0.2216**	0.4194	0.4430	0.7297
m. M.	∅ a. A.	**0.0451**	0.0474	0.0468	0.0576	0.0686	0.0829
29	∅ r. A.	**0.1885**	0.1979	0.1957	0.2409	0.2859	0.3450
h. M.	∅ a. A.	0.0387	0.0387	0.0346	0.0913	**0.0304**	0.0996
14	∅ r. A.	0.2617	0.2539	**0.2448**	0.6702	0.3243	0.8813

Tab. B.14: Auswertung: $\alpha = 0.01$, $n = 10$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	0.0108	0.0117	**0.0106**	0.0412	0.0466	0.0919
22	$\emptyset$ r. A.	0.0926	0.0977	**0.0804**	0.3034	0.3235	0.7367
m. M.	$\emptyset$ a. A.	0.0191	0.0208	**0.0171**	0.0292	0.0796	0.0681
52	$\emptyset$ r. A.	0.0800	0.0869	**0.0712**	0.1226	0.3322	0.2853
h. M.	$\emptyset$ a. A.	**0.0099**	0.0123	0.0135	0.0694	0.0159	0.0716
24	$\emptyset$ r. A.	**0.1393**	0.2281	0.2281	1.4637	0.2705	0.8099

Tab. B.15: Auswertung: $\alpha = 0.01$, $n = 30$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0058**	0.0073	0.0060	0.0315	0.0343	0.0853
27	$\emptyset$ r. A.	0.0615	0.0730	**0.0519**	0.2861	0.3697	0.8504
m. M.	$\emptyset$ a. A.	0.0150	0.0164	**0.0142**	0.0356	0.0717	0.0644
42	$\emptyset$ r. A.	0.0630	0.0694	**0.0599**	0.1547	0.3008	0.2721
h. M.	$\emptyset$ a. A.	**0.0063**	0.0084	0.0082	0.0386	0.0141	0.0601
27	$\emptyset$ r. A.	0.1379	0.1266	**0.1316**	$> 10^7$	958.87	53.343

Tab. B.16: Auswertung: $\alpha = 0.01$, $n = 50$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0056**	0.0069	0.0067	0.0267	0.0415	0.0856
31	$\emptyset$ r. A.	**0.0431**	0.0506	0.0556	0.2229	0.3642	0.8216
m. M.	$\emptyset$ a. A.	**0.0102**	0.0134	0.0133	0.0190	0.0728	0.0679
34	$\emptyset$ r. A.	**0.0428**	0.0566	0.0562	0.0807	0.3047	0.2887
h. M.	$\emptyset$ a. A.	**0.0047**	0.0087	0.0083	0.0355	0.0160	0.0631
32	$\emptyset$ r. A.	**0.0453**	0.1115	0.1138	0.3787	0.2755	0.7604

Tab. B.17: Auswertung: $\alpha = 0.01$, $n = 100$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0037**	0.0070	0.0059	0.0123	0.0359	0.0854
25	$\emptyset$ r. A.	**0.0317**	0.0532	0.0424	0.1040	0.3474	0.7795
m. M.	$\emptyset$ a. A.	**0.0054**	0.0091	0.0102	0.0094	0.0631	0.0656
40	$\emptyset$ r. A.	**0.0228**	0.0379	0.0425	0.0399	0.2644	0.2769
h. M.	$\emptyset$ a. A.	**0.0035**	0.0084	0.0067	0.0283	0.0158	0.0609
34	$\emptyset$ r. A.	**0.0290**	0.0984	0.0881	0.2417	0.2244	0.6653

Tab. B.18: Auswertung: $\alpha = 0.01$, $n = 500$, $k = 3$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	∅ a. A.	0.0125	0.0140	**0.0124**	0.0439	0.0372	0.1061
23	∅ r. A.	0.1343	0.1401	**0.0979**	0.4654	0.3491	0.9660
m. M.	∅ a. A.	**0.0279**	0.0301	0.0326	0.0473	0.0708	0.1276
30	∅ r. A.	**0.1162**	0.1251	0.1360	0.2008	0.2972	0.5378
h. M.	∅ a. A.	0.0161	0.0160	**0.0144**	0.0669	0.0180	0.1069
7	∅ r. A.	**0.0938**	0.0973	0.1024	0.4232	0.1037	0.6696

Tab. B.19: Auswertung: $\alpha = 0.01$, $n = 10$, $k = 4$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	∅ a. A.	**0.0048**	0.0060	0.0066	0.0303	0.0407	0.1145
31	∅ r. A.	**0.0495**	0.0586	0.0577	0.2613	0.3732	0.9351
m. M.	∅ a. A.	0.0146	0.0149	**0.0115**	0.0457	0.0868	0.0975
40	∅ r. A.	0.0620	0.0636	**0.0494**	0.1973	0.3696	0.4199
h. M.	∅ a. A.	0.0066	0.0090	**0.0061**	0.0573	0.0215	0.0900
22	∅ r. A.	0.0598	0.0773	**0.0529**	0.4488	0.2624	0.9467

Tab. B.20: Auswertung: $\alpha = 0.01$, $n = 30$, $k = 4$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	∅ a. A.	0.0081	0.0079	**0.0074**	0.0351	0.0563	0.1090
31	∅ r. A.	0.0571	0.0563	**0.0495**	0.2532	0.4021	0.8487
m. M.	∅ a. A.	0.0130	0.0147	**0.0126**	0.0378	0.0716	0.1123
43	∅ r. A.	0.0551	0.0621	**0.0533**	0.1630	0.3030	0.4780
h. M.	∅ a. A.	0.0052	0.0074	**0.0041**	0.0456	0.0194	0.0912
19	∅ r. A.	0.0442	0.0727	**0.0402**	0.3832	0.3655	1.0868

Tab. B.21: Auswertung: $\alpha = 0.01$, $n = 50$, $k = 4$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	∅ a. A.	**0.0105**	0.0130	0.0146	0.0300	0.0327	0.1161
17	∅ r. A.	**0.1019**	0.1217	0.1435	0.3031	0.2985	1.0801
m. M.	∅ a. A.	**0.0129**	0.0168	0.0142	0.0372	0.0824	0.2115
11	∅ r. A.	**0.0535**	0.0698	0.0593	0.1538	0.3407	0.8784
h. M.	∅ a. A.	0.0140	0.0123	**0.0013**	0.0849	0.0266	0.3065
1	∅ r. A.	0.0670	0.0586	**0.0060**	0.4058	0.1272	1.4650

Tab. B.22: Auswertung: $\alpha = 0.01$, $n = 10$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0081**	0.0091	0.0108	0.0426	0.0524	0.1461
28	$\emptyset$ r. A.	**0.0563**	0.0611	0.0746	0.2867	0.3611	1.0430
m. M.	$\emptyset$ a. A.	0.0156	0.0159	**0.0078**	0.0504	0.0728	0.1869
43	$\emptyset$ r. A.	0.0658	0.0674	0.0331	0.2131	0.3073	0.7863
h. M.	$\emptyset$ a. A.	0.0082	0.0092	**0.0061**	0.0665	0.0192	0.1394
21	$\emptyset$ r. A.	**0.0575**	0.0760	0.0605	0.5307	0.2588	1.3502

Tab. B.23: Auswertung: $\alpha = 0.01$, $n = 50$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0048**	0.0055	0.0066	0.0365	0.0419	0.1400
26	$\emptyset$ r. A.	0.0549	0.0586	**0.0507**	0.2902	0.3288	1.0483
m. M.	$\emptyset$ a. A.	0.0122	0.0119	**0.0051**	0.0284	0.0719	0.1842
41	$\emptyset$ r. A.	0.0507	0.0494	**0.0214**	0.1199	0.2992	0.7736
h. M.	$\emptyset$ a. A.	**0.0064**	0.0067	0.0069	0.0462	0.0197	0.1416
23	$\emptyset$ r. A.	0.0839	0.0699	**0.0661**	1.2794	0.3561	1.7223

Tab. B.24: Auswertung: $\alpha = 0.01$, $n = 100$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0039**	0.0048	0.0046	0.0291	0.0365	0.1449
26	$\emptyset$ r. A.	**0.0320**	0.0382	0.0324	0.2101	0.2861	1.0525
m. M.	$\emptyset$ a. A.	0.0081	0.0073	**0.0052**	0.0180	0.0632	0.1945
45	$\emptyset$ r. A.	0.0338	0.0306	**0.0218**	0.0772	0.2630	0.8109
h. M.	$\emptyset$ a. A.	**0.0035**	0.0070	0.0041	0.0282	0.0158	0.1375
28	$\emptyset$ r. A.	**0.0318**	0.0864	0.0648	0.2728	0.2608	1.4654

Tab. B.25: Auswertung: $\alpha = 0.01$, $n = 500$, $k = 6$.

		LKC	GLK	GAM	NC	NOR	AN
n. M.	$\emptyset$ a. A.	**0.0079**	0.0114	0.0136	0.0535	0.0420	0.1589
24	$\emptyset$ r. A.	**0.0856**	0.1070	0.0956	0.4369	0.3113	1.1676
m. M.	$\emptyset$ a. A.	0.0141	0.0182	**0.0107**	0.0455	0.0695	0.2931
48	$\emptyset$ r. A.	0.0592	0.0767	**0.0459**	0.1954	0.2942	1.2453
h. M.	$\emptyset$ a. A.	**0.0056**	0.0062	0.0089	0.0426	0.0219	0.2405
15	$\emptyset$ r. A.	**0.0482**	0.1238	0.1886	0.6245	0.8058	4.3377

Tab. B.26: Auswertung: $\alpha = 0.01$, $n = 100$, $k = 10$.

Anhang C

Mathematica-Programm zur Berechnung der Verteilungsfunktionen aus Kapitel 3

C.1 Das Programm

Nachfolgend wird der bei den Beispielen aus Kapitel 3 verwendete Programmcode zur Ermittlung der Verteilungsfunktionen der verschiedenen Approximationen gegeben. Das Programm *Mathematica* (Wolfram Research, Inc., 1991) erlaubt symbolisches Ableiten von Funktionen und ist deshalb besonders geeignet zur Ermittlung der benötigten Matrizen und Vektoren. Die Eingaben des Anwenders können auf ein Minimum reduziert werden. Im anschließenden Abschnitt wird eine Mustersitzung durchgeführt, um die Funktionsweise zu illustrieren.

Programmcode:

```
(* Allgemeine Definitionen *)

krit[alpha_]:=Quantile[ChiSquareDistribution[Length[p]-1-
```

```
              Length[par]],1-alpha];

w[x_,y_]:=(x-y)/Sqrt[y];

delta[n_,x_,y_]:= n Sum[(w[x,y][[i]])^2, {i, Length[x]}];

deltag[g_,x_,y_]:= Sum[g[x[[i]],y[[i]]], {i, Length[x]}];

abl[g_,x_,y_]:=Table[D[deltag[g,x,y], par[[i]]],
               {i, Length[par]}];

hesse[g_,x_,y_] := Table[D[abl[g,x, y][[j]], par[[i]]],
                   {i,Length[par]},{j,Length[par]}];

norm[x_]:=x /Sqrt[Sum[x[[i]]^2,{i,Length[x]}]];

cmat[h_]:=Table[D[h[[j]],par[[i]]],{i,Length[par]},
          {j,Length[ht]}];

dmat[p_,exp_]:=DiagonalMatrix[p^exp];

id[p_]:=IdentityMatrix[Length[p]];

g2xy[x_,y_]:=D[g[x,y],x,y];

pp[p_]:=Table[p[[i]]p[[j]],{i,Length[p]},{j,Length[p]}];

dasy[p_]:=dmat[p,1]-pp[p];

konmom[x_]:=Apply[Plus,lambdapos^x*(1+x nzpos)];


(*CDF fuer die asymptotische Normalverteilung*)

cdfan[wert_]:=CDF[NormalDistribution[erwan,Sqrt[varan]],
                 wert];
```

```
(*CDF fuer die nichtzentrale Chiquadratverteilung *)

cdfnc[wert_]:=CDF[NoncentralChiSquareDistribution[Length[p]-
Length[par]-1,nzpchi],wert];


(* CDF fuer die Gammaverteilung mit den Momenten von LKC*)

cdfglk[wert_]:=
CDF[GammaDistribution[lkcgama,lkcgamb],wert-nznull];


(* CDF der Konishi-Approximation *)

cdfkon[x_]:=Module[{dum0,dum1,y,d1,d2},
y=Sqrt[m1]*( ((x-nznull)/m1)^hto-1-
1/m1*hto*(hto-1)*w2)/Sqrt[2 hto^2 *w2];
dum0=CDF[NormalDistribution[0,1],y]-
PDF[NormalDistribution[0,1],y]*
(a2[y]*m1^-1+a3[y]*m1^-1.5+a4[y]*m1^-2+a5[y]*m1^-2.5+
a6[y]*m1^-3);If[hto<0,dum1=1-dum0,dum1=dum0];
dum1];


(* Initialisierung fuer gegebenes n und p:
Matrizen und Vektoren berechnen, fuer alle Approximationen*)

init[p_,n_]:=(

gmat=Module[{x,y},dmat[Table[g2xy[x,y]/. {x->p[[i]],y->h[[i]]},
  {i,Length[p]}],1] /. tnull[p]];

hminhalb=dmat[h,-0.5] /. tnull[p];

hmineins=dmat[h,-1] /. tnull[p];

omega=hesse[g,p,h] /. tnull[p];
```

```
omegainv=Inverse[omega];

cmat2=cmat[ht] /. tnull[p];

wpi=w[p,h] /. tnull[p];

deltanull=delta[1,p,h] /.tnull[p];


(* Momente der AN-Verteilung*)

erwan=n* deltanull;

jacan= 2 p. hmineins.(id[p]+0.5 hmineins.dmat[p,1].
           Transpose[cmat2].omegainv.cmat2.gmat);

varan=n jacan.dasy[p].Transpose[jacan];


(* Berechnung des Nichtzentralitaetsparameters fuer die
CHI-Verteilung*)

chidum1=dmat[h,-0.5].Transpose[cmat2] /.tnull[p];

chidum2=Inverse[Transpose[chidum1].chidum1];

chidum3=Sqrt[n] p. dmat[h,-0.5].chidum1 /.tnull[p];

nzpchi=n deltanull - chidum3.chidum2.Transpose[chidum3];


(* Konstanten, Erwartung und Varianz / LKC*)

lkcdumw=hminhalb.(id[p]+0.5 (id[p] +dmat[p,1].hmineins).
                  Transpose[cmat2].omegainv.cmat2.gmat);

lkcdumv=lkcdumw.dasy[p].Transpose[lkcdumw];
```

```
lkcdumvquadrat=lkcdumv.lkcdumv;

lkctrace=Sum[lkcdumv[[i,i]],{i,Length[p]}];

lkcquadrattrace=Sum[lkcdumvquadrat[[i,i]],{i,Length[p]}];

eig=Eigenvectors[lkcdumv];

ortho=Table[norm[eig[[i]]],{i,Length[eig]}];

lam=Eigenvalues[lkcdumv];

nzp=n * (ortho.wpi)^2;

varcon1=2 Apply[Plus,lam^2];

varcon2=4 Apply[Plus,lam *nzp];

varcon=varcon1+varcon2;

lamnz=Table[{lam[[i]],nzp[[i]]},{i,Length[p]}];

poslam=Select[lamnz,#[[1]]>10^-8&];

nulllam=Complement[lamnz,poslam];

lambdapos=Transpose[poslam][[1]];

nzpos=Transpose[poslam][[2]]/lambdapos;

nznull=Apply[Plus,Transpose[nulllam][[2]]];

lkcerw=lkctrace+n *deltanull-nznull;

lkcvar=2 lkcquadrattrace+ 4 n wpi.lkcdumv.wpi;

lkcgama=lkcerw^2/lkcvar;
```

```
lkcgamb=lkcvar/lkcerw;


(*Berechnung der Konishi-Approximation*)

m1=konmom[1];
m2=konmom[2];
m3=konmom[3];
m4=konmom[4];
m5=konmom[5];
m6=konmom[6];
m7=konmom[7];
m8=konmom[8];

w2=m2/m1;
w3=m3/m1;
w4=m4/m1;
w5=m5/m1;
w6=m6/m1;
w7=m7/m1;
w8=m8/m1;

hto=1-2 m1*m3/(3*m2^2);

hp1[y_]:=y;
hp2[y_]:=y^2-1;
hp3[y_]:=y^3-3y;
hp4[y_]:=y^4-6y^2+3;
hp5[y_]:=y^5-10y^3+15 y;
hp6[y_]:=y^6-15 y^4+45 y^2 -15;
hp7[y_]:=y^7-21y^5+105y^3-105y;
hp8[y_]:=y^8-28 y^6+210 y^4-420 y^2+105;
hp9[y_]:=y^9-36 y^7+378 y^5-1260 y^3+945* y;
hp10[y_]:=y^10-45*y^8+630 y^6-3150 y^4+4725 y^2-945;
hp11[y_]:=y^11-55 y^9+990 y^7-6930 y^5+17325 y^3-10395 y;

a2[y_]:=w2^(-3)(hp3[y](0.5 w4 *w2 -20./27 w3^2+2./9*w3*w2^2)
+hp1[y]*w3*(-2./3.*w3+2/3.*w2^2));
```

```
a3[y_]:=Sqrt[2.]*w2^-4.5(hp4[y](0.4 w5*w2^2-4./3 w4*w3*w2+
76./81w3^3+1./9 * w3^2*w2^2-1./9 *w3*w2^4) +
 hp2[y]*w3*(-2.w4*w2+184./81 *w3^2+4./9 w3*w2^2-2./3*w2^4)
+w3*(2./9w3^2+1./9*w3 * w2^2-1/3.*w2^4));

a4[y_]:=w2^-6 *(hp7[y](1./8*w4^2*w2^2-10./27*w4*w3^2*w2+
1./9*w4*w3*w2^3+200./729*w3^4-40./243*w3^3*w2^2+
2./81*w3^2*w2^4)+hp5[y]*(2./3*w6*w2^3-8./3*w5*w3*w2^2+
7./3*w4*w3^2*w2+5./3*w4*w3*w2^3+56./405*w3^4-
1016./405*w3^3*w2^2+32./135*w3^2*w2^4+2/15.*w3*w2^6)+
hp3[y]*w3*(-16./3*w5*w2^2+104/9*w4*w3*w2+ 16./3*w4*w2^3-
1106./243*w3^3-76./9*w3^2*w2^2+2/27*w3*w2^4+4/3*w2^6)+
hp1[y]*w3*(28./9*w4*w3*w2+8/3*w4*w2^3-560./243*w3^3-
392/81.*w3^2*w2^2-20./27*w3*w2^4+2*w2^6));

a5[y_]:=Sqrt[2.]*w2^-7.5 *(hp8[y]*(0.2*w5*w4*w2^3-
8./27*w5*w3^2*w2^2+4./45*w5*w3*w2^4-2./3*w4^2*w3*w2^2+
118./81*w4*w3^3*w2-13./54*w4*w3^2*w2^3-1./18 *w4*w3*w2^5 -
 1520./2187 * w3^5 +92./729*w3^4*w2^2+26./243*w3^3*w2^4-
2/81*w3^2*w2^6)+hp6[y]*(4./7*w7*w2^4- 8./3*w6*w3*w2^3+
148./45*w5*w3^2*w2^2+8./5*w5*w3*w2^4-7./3*w4^2*w3*w2^2+
508./81*w4*w3^3*w2-14/3*w4*w3^2*w2^3-11./9*w4*w3*w2^5-
19016./3645*w3^5+70./27*w3^4*w2^2+556./243*w3^3*w2^4-
34./81*w3^2*w2^6-4./45*w3*w2^8)+hp4[y]*w3*(-20./3*w6*w2^3+
56/3*w5*w3*w2^2+20./3*w5*w2^4-67./9*w4*w3^2*w2-
391./18*w4*w3*w2^3-41/6*w4*w2^5-2168./405*w3^4+
3124./243*w3^3*w2^2+1012./81*w3^2*w2^4-4./3*w3*w2^6-
4./3*w2^8)+hp2[y]*w3*(80./9*w5*w3*w2^2+16./3*w5*w2^4-
328./27*w4*w3^2*w2-196./9*w4*w3*w2^3-28./3*w4*w2^5+
1316./729*w3^4+3394./243*w3^3*w2^2+1360./81*w3^2*w2^4+
14./27*w3*w2^6-4*w2^8)+w3*(-8./27*w4*w3^2*w2-
4./3*w4*w3*w2^3-4./3*w4*w2^5+88./729*w3^4+248/243*w3^3*w2^2
+ 194./81*w3^2*w2^4+22./27*w3*w2^6-4./3*w2^8));

a6[y_]:=w2^-9.(hp11[y]*(1./48*w4^3*w2^3-5./54*w4^2*w3^2*w2^2+
1./36*w4^2*w3*w2^4+100./729*w4*w3^4*w2-20./243*w4*w3^3*w2^3+
1./81*w4*w3^2*w2^5-4000./59049*w3^6+400./6561*w3^5*w2^2-
```

```
40./2187*w3^4*w2^4+4./2187*w3^3*w2^6)+hp9[y]*(1./3*w6*w4*w2^4-
40./81*w6*w3^2*w2^3+4./27*w6*w3*w2^5+4./25*w5^2*w2^4-
12./5*w5*w4*w3*w2^3+368./135*w5*w3^3*w2^2-
68./135*w5*w3^2*w2^4-4./45*w5*w3*w2^6+109./36*w4^2*w3^2*w2^2+
3./4*w4^2*w3*w2^4-5356./1215*w4*w3^4*w2-788./405*w4*w3^3*w2^3+
32./45*w4*w3^2*w2^5+1./15*w4*w3*w2^7+6304./6561*w3^6+
19736./10935*w3^5*w2^2-2927./3645*w3^4*w2^4-
106./1215*w3^3*w2^6+17./405*w3^2*w2^8)+
hp7[y]*(w8*w2^5-16./3*w7*w3*w2^4+76./9*w6*w3^2*w2^3+
28./9*w6*w3*w2^5-48./5*w5*w4*w3*w2^3+4336./405*w5*w3^3*w2^2-
1552./135*w5*w3^2*w2^4-104./45*w5*w3*w2^6 +
76./3*w4^2*w3^2*w2^2+16./3*w4^2*w3*w2^4-12233./243*w4*w3^4*w2-
682./81*w4*w3^3*w2^3+325./27*w4*w3^2*w2^5+2*w4*w3*w2^7+
4275352./229635*w3^6+1120492./76545*w3^5*w2^2-
285704/25515.*w3^4*w2^4-31132./8505*w3^3*w2^6+
2872./2835*w3^2*w2^8+8./63*w3*w2^10)+hp5[y]*w3*(-16*w7*w2^4+
496./9*w6*w3*w2^3+16*w6*w2^5-5416./135*w5*w3^2*w2^2-
1012./15*w5*w3*w2^4-244./15*w5*w2^6+266./9*w4^2*w3*w2^2+
28./3*w4^2*w2^4-7000./81*w4*w3^3*w2+220./9*w4*w3^2*w2^3+
214./3*w4*w3*w2^5+17*w4*w2^7+1732396./32805*w3^5+
23480./729*w3^4*w2^2-13954./243*w3^3*w2^4-7900./243*w3^2*w2^6+
778./135*w3*w2^8+8./3*w2^10)+hp3[y]*w3*(1040./27*w6*w3*w2^3+
160./9*w6*w2^5-1984./27*w5*w3^2*w2^2-896./9*w5*w3*w2^4-
32*w5*w2^6+272./81*w4*w3^3*w2+2668./27*w4*w3^2*w2^3+
132*w4*w3*w2^5+44*w4*w2^7+55616./2187*w3^5+
1040./243*w3^4*w2^2-2552./27*w3^3*w2^4-2216./27*w3^2*w2^6+
112./27*w3*w2^8+40./3*w2^10)+hp1[y]*w3*(-64./9*w5*w3^2*w2^2-
160./9*w5*w3*w2^4-32./3*w5*w2^6+160./27*w4*w3^3*w2+
784./27*w4*w3^2*w2^3+416./9*w4*w3*w2^5+24*w4*w2^7+
196./729*w3^5-3932./729*w3^4*w2^2-7081./243*w3^3*w2^4-
3358./81*w3^2*w2^6-197./27*w3*w2^8+40./3*w2^10));

Print["Initialisierung durchgefuehrt"]);
```

C.2 Eine Mustersitzung

Anhand einer Mustersitzung soll die leichte Anwendbarkeit des Programmes demonstriert werden. Angenommen, es sollen zum Beispiel 3 aus Abschnitt 3.5.4 verschiedene Berechnungen durchgeführt werden. Für das Grundmodell

$$\pi = (0.1, 0.2, 0.15, 0.1, 0.05, 0.4)^T$$

soll die Macht des χ^2-Anpassungstests bei Verwendung des Maximum-Likelihood-Schätzers für die Stichprobengrößen $n = 50$ und 100 approximiert werden. Betrachtet werden dabei die Signifikanzniveaus $\alpha = 0.10$ und $\alpha = 0.01$.

Anschließend sollen analoge Schätzer der Macht berechnet werden, wenn Mimimum-χ^2-Schätzer verwendet werden. Sei der Programmcode in der Datei `cdf.inp` enthalten. Kommentare zu den Befehlen werden durch „(*“ und „*)“ eingeschlossen, sie erscheinen bei der Programmausführung *nicht*. Nachfolgend ist die entstehende, interaktive Sitzung protokolliert.

```
Mathematica 2.0 for MS-DOS 386/7 (June 21, 1991)
Copyright 1988-91 Wolfram Research, Inc.

In[1]:= <<Statistics`ContinuousDistributions`
        (* Einlesen des Mathematica-Statistikpaketes *)
In[2]:= <<cdf.inp
        (* Einlesen des Programmes *)

        (* Die ersten beiden Eingaben brauchen nur einmal
           ausgefuehrt zu werden. *)

In[3]:= g[x_,y_]:=2x Log[x/y]
        (* Definition der Funktion g und somit der Art
        der Parameterschaetzung *)
In[4]:= par={t[1]};
        (* Definition des Parametervektors *)
In[5]:= h={t[1],t[1],t[1],t[1],t[1],1-5t[1]};
        (* approximierendes Modell *)
In[6]:= tnull[p_]:={t[1]->1/5Sum[p[[i]],{i,1,5}]};
        (* Formel fuer Theta, das Diskrepanz minimiert. *)
```

```
        (* Die Eingaben In[3]-In[6] reichen aus, um die
           approximierende Modellfamilie und die verwendeten
           Schaetzer festzulegen! *)

In[7]:= p={0.1,0.2,0.15,0.1,0.05,0.4};
        (* Eingabe des Grundmodells *)
In[8]:= init[p,50]
Out[8]= Initialisierung durchgefuehrt
        (* Initialisierung fuer n=50: Berechnung aller
        benoetigten Groessen. Anschliessend stehen
        saemtliche Verteilungsfunktionen zur Verfuegung:
           cdfkon[x],cdfglk[x], cdfnc[x] und cdfan[x] *)
In[9]:= kw1=krit[0.10]
        (* Kritischer Wert zum Signifikanzniveau 0.10. Die
        Anzahl der Freiheitsgrade wird automatisch
        berechnet. *)
Out[9]= 7.77944
In[10]:=kw2=krit[0.01]
        (* Kritischer Wert zum Signifikanzniveau 0.01 *)
Out[10]=  13.2767

     (* Zunaechst wird das Signifikanzniveau 0.10 betrachtet *)

In[11]:= 1-cdfkon[kw1]
Out[11]= 0.53878
        (* Schaetzer der Macht durch KON-Approximation *)

In[12]:= 1-cdfglk[kw1]
Out[12]= 0.536477
        (* Schaetzer der Macht durch GLK-Approximation *)

In[13]:= 1-cdfnc[kw1]
Out[13]= 0.555645
        (* Schaetzer der Macht durch NC1-Approximation *)

In[14]:= 1-cdfan[kw1]
Out[14]= 0.301531
        (* Schaetzer der Macht durch AN-Approximation *)
```

```
     (* Nun folgen die Schaetzer zum Signifikanzniveau 0.01 *)

In[15]:= 1-cdfkon[kw2]
Out[15]= 0.19827
In[16]:= 1-cdfglk[kw2]
Out[16]= 0.19801
In[17]:= 1-cdfnc[kw2]
Out[17]= 0.212606
In[18]:= 1-cdfan[kw2]
Out[18]= 0.0418286

In[19]:= init[p,100]
Out[19]= Initialisierung durchgefuehrt
        (* Initialisierung fuer n=100 *)

        (* Berechnungen zum Signifikanzniveau 0.10 *)

In[20]:= 1-cdfkon[kw1]
Out[20]= 0.845835
In[21]:= 1-cdfglk[kw1]
Out[21]= 0.847549
In[22]:= 1-cdfnc[kw1]
Out[22]= 0.842861
In[23]:= 1-cdfan[kw1]
Out[23]= 0.682696

        (* Berechnungen zum Signifikanzniveau 0.01 *)

In[24]:= 1-cdfkon[kw2]
Out[24]= 0.520424
In[25]:= 1-cdfglk[kw2]
Out[25]= 0.516808
In[26]:= 1-cdfnc[kw2]
Out[26]= 0.534427
In[27]:= 1-cdfan[kw2]
Out[27]= 0.351884
```

```
     (* Nun folgen die analogen Berechnungen bei Verwendung
           von Minimum-Chi-Quadrat-Schaetzern *)

In[28]:= g[x_,y_]:=(x-y)^2/y;
In[29]:= tnull[p_]:={t[1]->1/(5+p[[6]]*Sqrt[5/Sum[p[[i]]^2,
    {i,1,5}]])};
In[30]:= init[p,50]
Out[30]= Initialisierung durchgefuehrt

In[31]:= 1-cdfkon[kw1]
Out[31]= 0.524921
In[32]:= 1-cdfglk[kw1]
Out[32]= 0.522972
In[33]:= 1-cdfnc[kw1]
Out[33]= 0.542769
In[34]:= 1-cdfan[kw1]
Out[34]= 0.288732
In[35]:= 1-cdfkon[kw2]
Out[35]= 0.184498
In[36]:= 1-cdfglk[kw2]
Out[36]= 0.184461
In[37]:= 1-cdfnc[kw2]
Out[37]= 0.203035
In[38]:= 1-cdfan[kw2]
Out[38]= 0.0354787

In[39]:= init[p,100]
Out[39]= Initialisierung durchgefuehrt

In[40]:= 1-cdfkon[kw1]
Out[40]= 0.841249
In[41]:= 1-cdfglk[kw1]
Out[41]= 0.842646
In[42]:= 1-cdfnc[kw1]
Out[42]= 0.830902
In[43]:= 1-cdfan[kw1]
Out[43]= 0.677827
In[44]:= 1-cdfkon[kw2]
```

```
Out[44]= 0.504682
In[45]:= 1-cdfglk[kw2]
Out[45]= 0.501394
In[46]:= 1-cdfnc[kw2]
Out[46]= 0.514799
In[47]:= 1-cdfan[kw2]
Out[47]= 0.336767
```

Die einzige Arbeit, die dem Benutzer nicht abgenommen werden kann, ist die Herleitung eines Algorithmus zur Bestimmung des besten Parameters θ_0. Im vorliegenden Beispiel kann θ_0 direkt ermittelt werden (siehe `In[6]` und `In[29]`). Für kompliziertere Anwendungen reicht es jedoch immer aus, einen Algorithmus zu konstruieren, der zu gegebenem Wert von π gemäß der verwendeten Diskrepanz Δ_g das θ_0 findet, und sei es nur numerisch. Eine explizite Darstellung von θ_0 in Abhängigkeit von π ist *nicht* notwendig.

Sollen die Berechnungen für eine große Anzahl von Datensätzen durchgeführt werden, wie in der vorliegenden Arbeit geschehen, so bietet das Programmpaket Mathematica die Möglichkeit zur Programmierung eines Batch-Runs, d. h. einer Abarbeitung ohne weitere Einwirkung des Anwenders.

Literaturverzeichnis

Abdel-Aty, S. H. (1954). Approximate formulae for the percentage points and the probability integral of the non-central χ^2-distribution. *Biometrika*, **41**, 538–540.

Aithal, B. U. & Nagnur, B. N. (1988). A note on second order efficiency of minimum distance estimators for the multinomial distribution. *J. Indian Statist. Assoc.*, **26**, 1–9.

Beran, R. (1984). Bootstrap methods in statistics. *Jber. d. Dt. Math.-Verein*, **86**, 14–30.

Bhapkar, V. P. (1961). Some tests for categorical data. *Ann. Math. Statist.*, **32**, 72–83.

Bhapkar, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *J. Amer. Statist. Assoc.*, **61**, 228–235.

Bickel, P. J. & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196–1217.

Birch, M. W. (1964). A new proof of the Pearson-Fisher theorem. *Ann. Math. Statist.*, **35**, 817–824.

Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, MA.

Blacher, R. (1988). A new form for the chi-squared test of independence. *Statistics*, **19**, 519–536.

Bosq, D. (1980). Sur une classe de tests qui contient le test du χ^2. *Pub. Inst. Stat. Univ. Paris*, **XXV**, 1–16.

Broffitt, J. D. & Randles, R. H. (1977). A power approximation for the chi-square goodness-of-fit test: simple hypothesis case. *J. Amer. Statist. Assoc.*, **72**, 604–607.

Burke, M. D. & Gombay. E. (1988). On goodness-of-fit and the bootstrap. *Statist. Probab. Lett.*, **6**, 287–293.

Chernoff, H. & Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness-of-fit. *Ann. Math. Statist.*, **25**, 579–586.

Cochran, W. G. (1952). The χ^2 test of goodness of fit. *Ann. Math. Stat.*, **23**, 315–345.

Cox, C. (1984). An elementary introduction to maximum likelihood estimation for multinomial models: Birch's theorem and the delta method. *Amer. Stat.*, **38**, 283–287.

Cressie, N. & Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. R. Statist. Soc. B*, **46**, 440–464.

Cressie, N. & Read, T. R. C. (1989). Pearson's χ^2 and the log likelihood ratio statistic G^2: a comparative review. *Int. Statist. Rev.*, **57**, 19–43.

Davies, R. B. (1973). Numerical inversion of a characteristic function. *Biometrika*, **60**, 415–417.

Davies, R. B. (1980). Algorithm AS 155: The distribution of a linear combination of χ^2 random variables. *Appl. Statist.*, **29**, 323–333.

Dik, J. J. & de Gunst, M. C. M. (1985). The distribution of general quadratic forms in normal variables. *Statistica Neerlandica*, **39**, 14–26.

Drost, F. C. (1988). *Asymptotics for generalized chi-square goodness-of-fit tests*, Tract **48**. CWI, Stichting.

Drost, F.C., Kallenberg, W.C.M., Moore, D.S. & Oosterhoff, J. (1989). Power approximations to multinomial tests of fit. *J. Amer. Statist. Assoc.*, **84**, 130–141.

Easterling, R. G. (1976). Goodness-of-fit and parameter estimation. *Technometrics*, **18**, 1–9.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–26.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.* SIAM, Philadelphia.

Falk, M. (1991). Functional limit theorems for inverse bootstrap processes of sample quantiles. *Statist. Prob. Lett.*, **11**, 529–536.

Farebrother, R. W. (1985). Eigenvalue-free methods for computing the distribution of a quadratic form in normal variables. *Statist. Hft.*, **26**, 287–302.

Farebrother, R. W. (1990). Algorithm AS 256: The distribution of a quadratic form in normal variables. *Appl. Statist.*, **39**, 294–309.

Fienberg, S. E. (1979). The use of chi-squared statistics for categorical data problems. *J. R. Statist. Soc. B*, **41**, 54–64.

Frosini, B. V. (1976). On the power function of the χ^2 test. *Metron*, **34**, 3–36.

Gabler, S. & Wolff, C. (1987). A quick and easy approximation to the distribution of a sum of weighted chi-square variables. *Statist. Hft.*, **28**, 317–325.

Gideon, R. A. & Gurland, J. (1976). Series expansions for quadratic forms in normal variables. *J. Amer. Statist. Assoc.*, **71**, 227–232.

Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, **38**, 481–482.

Hall, P. & Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, **47**, 757–762.

Harris, R. R. & Kanji, G. K. (1983). On the use of minimum chi-square estimation. *Statistician*, **32**, 379–394.

Heuser, H. (1981). *Lehrbuch der Analysis 2.* Teubner, Stuttgart.

Hinkley, D. V. (1988). Bootstrap methods. *J. R. Statist. Soc. B*, **50**, 321–337.

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.*, **36**, 369–408.

Holland, P. W. (1967). A variation on the minimum chi-square test. *J. Mathematical Psychology*, **4**, 377–413.

Holst, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika*, **59**, 137–145.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48**, 419–426.

Jensen, D. R. & Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form. *J. Amer. Statist. Assoc.*, **67**, 898–902.

Johnson, N. L. & Kotz, S. (1970). *Continuous Univariate Distributions, Vol. 1 and 2*. Houghton Mifflin, Boston.

Kallenberg, W. C. M. , Oosterhoff, J. & Schriever, B. F. (1985). The number of classes in chi-squared goodness-of-fit tests. *J. Amer. Statist. Assoc.*, **80**, 959–968.

Kallenberg, W. C. M. (1990). Inequalities for noncentral chi-square distributions. *Statist. Probab. Lett.*, **9**, 273–278.

Konishi, S. , Niki, N. & Gupta, A. K. (1988). Asymptotic expansions for the distribution of quadratic forms in normal variables. *Ann. Inst. Statist. Math.*, **40**, 279–296.

Koorts, A. S. (1985). Second order efficiency for a class of estimators within the multinomial family. *Comm. Statist.– Theor. Meth*, **14**, 2331–2343.

Kotz, S. , Johnson, N. L. & Boyd, D. W. (1967). Series representations of distributions of quadratic forms in normal variables II. Noncentral case. *Ann. Math. Statist.*, **38**, 838–848.

Krengel, U. (1988). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, Braunschweig.

Levit, B. Ya. (1975). On the efficiency of a class of non-parametric estimates. *Theor. Probability Appl.*, **XX**, 723–740.

Linhart, H. (1988). A test whether two AIC's differ significantly. *S. Afr. Statist. J.*, **22**, 153–161.

Linhart, H. & Volkers, P. (1984). Asymptotic criteria for model selection. *OR Spektrum*, **6**, 161–165.

Linhart, H. & Zucchini, W. (1986). *Model selection.* Wiley, New York.

Mann, H. B. & Wald, A. (1942). On the choice of the number of class intervals in the application of the chi-squared test. *Ann. Math. Statist.*, **13**, 306–317.

van der Merve, L. & Crowther, N. A. S. (1982). An asymptotic representation of the distribution of a specific quadratic form. *S. Afr. Statist. J.*, **16**, 133–145.

Mitra, S. K. (1958). On the limiting power function of the frequency chi-square test. *Ann. Math. Statist.*, **29**, 1221–1233.

Montinaro, M. (1988). On a better approximation of the power function of the χ^2 test. *Metron*, **46**, 361–389.

Moore, D. S. (1984). Measures of lack of fit from tests of chi-squared type. *J. Statist. Plann. Inf.*, **10**, 151–166.

Moore, D. S. (1986). Tests of chi-squared type. In D'Agostino, R. B. D., *Goodness-of-fit techniques*, Dekker, New York.

Moore, D. S. & Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.*, **3**, 599–616.

Morris, C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.*, **3**, 165–188.

Neyman, J. (1949). Contributions to the theory of the χ^2 test. *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, 239–273.

Patnaik, P. B. (1949). The non-central χ^2- and F-distributions and their applications. *Biometrika*, **36**, 202–232.

Pearson, E. S. (1959). Note on an approximation to the distribution of non-central χ^2. *Biometrika*, **46**, 364.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine Series (5)*, **50**, 157–172.

Pederson, S. P. & Johnson, M. E. (1990). Estimating model discrepancy. *Technometrics*, **32**, 305–314.

Ponnapalli, R. (1976). Deficiencies of minimum discrepancy estimators. *Can. J. Statist.*, **4**, 33–50.

Press, S. J. (1972). *Applied Multivariate Analysis.* Holt, Rinehart and Winston, New York.

Rao, C. R. (1957). Theory of the method of estimation by minimum chi-square. *Bulletin of the International Statistical Institute*, **35**, 25–32.

Rao, C. R. (1962). Efficient estimates and optimum inference procedures in large samples. *J. R. Statist. Soc. B*, **24**, 46–72.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications (2nd edition).* Wiley, New York.

Read, T. R. C. & Cressie, N. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Springer, New York.

Robertson, C. A. (1972). On minimum discrepancy estimators. *Sankhyā A*, **34**, 133–144.

Ruben, H. (1962). Probability content of regions under spherical normal distributions, IV. *Ann. Math. Statist.*, **33**, 542–570.

Sankaran, M. (1959). On the non-central chi square distribution. *Biometrika*, **46**, 235–237.

Sankaran, M. (1963). Approximations to the non-central chi-square distribution. *Biometrika*, **50**, 199–204.

Selianov, B. I. (1984). Limit distributions of the χ^2 statistic. *Theor. Probability Appl.*, **29**, 133–134.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Sheil, J. & O'Muircheartaigh, I. (1977). Algorithm AS 106: The distribution of non-negative quadratic forms in normal variables. *Appl. Statist.*, **26**, 92–98.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.*, **9**, 1187–1195.

Slakter, M. J. (1968). Accuracy of an approximation to the power function of the chi-square goodness of fit test with small but equal expected frequencies. *J. Amer. Statist. Assoc.*, **63**, 912–918.

Solomon, H. & Stephens, M. A. (1977). Distribution of a sum of weighted chi-square variables. *J. Amer. Statist. Assoc.*, **72**, 881–885.

Steck, G. P. (1957). Limit theorems for conditional distributions. *Univ. Calif. Publ. Statist.*, **2**, 237–284.

Stephens, M. A. (1974). EDF Statistics for goodness of fit and some comparisons. *J. Amer. Statist. Assoc.*, **69**, 730–737.

Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *Ann. Statist.*, **4**, 357–369.

Swanepoel, J. W. H. (1990). Invited review paper. A review of bootstrap methods. *S. Afr. Statist. J.*, **24**, 1–34.

Walker, J. J. (1979). An asymptotic expansion for unbalanced quadratic forms in normal variables. *J. Amer. Statist. Assoc.*, **74**, 389–392.

Werner, J. (1992). *Numerische Mathematik II*. Vieweg, Braunschweig.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

Withers, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Ann. Statist*, **11**, 577–587.

Wolfram Research, Inc. (1991). *Mathematica, V. 2. 0.* Wolfram Research, Inc., Champaign, Illinois.

Zorin, V. A. & Iyubimov, A. K. (1989). The distribution of quadratic forms of Gaussian vectors. *Theor. Probability Appl.*, **33**, 557–560.