

Reihenherausgeber:

Prof. Dr. Holger Dette · Prof. Dr. Wolfgang Karl Härdle

Statistik und ihre Anwendungen

Weitere Bände dieser Reihe finden Sie unter <http://www.springer.com/series/5100>

Claudia Czado · Thorsten Schmidt

Mathematische Statistik

 Springer

Prof. Claudia Czado, Ph.D.
Technische Universität München
Lehrstuhl für Mathematische
Statistik
Boltzmannstraße 3
85748 Garching
Deutschland
cczado@ma.tum.de

Prof. Dr. Thorsten Schmidt
Technische Universität Chemnitz
Fakultät für Mathematik
Reichenhainer Straße 41
09126 Chemnitz
Deutschland
thorsten.schmidt@mathematik.tu-chemnitz.de

ISBN 978-3-642-17260-1 e-ISBN 978-3-642-17261-8
DOI 10.1007/978-3-642-17261-8
Springer Heidelberg Dordrecht London New York

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Springer-Verlag Berlin Heidelberg 2011

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Einbandentwurf: WMXDesign GmbH, Heidelberg

Gedruckt auf säurefreiem Papier

Springer ist Teil der Fachverlagsgruppe Springer Science+Business Media (www.springer.com)

Vorwort

Mit den wachsenden Möglichkeiten Daten zu erheben steht deren adäquate Auswertung und Bewertung im Mittelpunkt der Statistik. Dabei treten viele unterschiedliche Datenstrukturen auf, die eine komplexe Modellierung erforderlich machen. In weiteren Schritten sind statistische Verfahren zum Anpassen der Modelle oder zum Untersuchen von interessanten Fragestellungen notwendig. Dieses Buch stellt die dafür notwendigen mathematischen Grundlagen und Konzepte der Statistik zur Verfügung. Dabei wird Wert auf die Herleitung von statistischen Fragestellungen und deren probabilistische Behandlung gelegt. Um die Verständlichkeit zu erhöhen, werden viele Beispiele ausgearbeitet und elementare Beweise ohne maßtheoretische Hilfsmittel gezeigt. Genaue Literaturhinweise ermöglichen die weitergehende Vertiefung. Durch die kurze und präzise Darstellung wird darüber hinaus ein schneller Einstieg in das Fachgebiet ermöglicht. Dabei folgen wir dem Ansatz von Bickel und Doksum (1977, 2001) und Casella und Berger (2002). Bei der Auswahl der Themen orientieren wir uns an der Praxisrelevanz der Verfahren. Anhand einer umfangreichen Aufgabensammlung am Ende jedes Kapitels kann das Verständnis überprüft und vertieft werden.

Dieses Buch richtet sich an Studierende der Mathematik und Statistik im zweiten oder dritten Jahr des Bachelor-Studiums oder ersten Jahr des Master-Studiums. Für andere Fachrichtungen ist ein starker mathematischer Schwerpunkt notwendig. Das Buch setzt Grundlagen der Wahrscheinlichkeitstheorie voraus wie sie zum Beispiel in Dehling und Haupt (2004) oder Georgii (2004) zu finden sind. Das mathematische Niveau des Buches liegt zwischen Fahrmeir et. al (2004) und den englischen Standardwerken von Lehmann und Casella (1998), Lehmann und Romano (2006) und Shao (2008).

Das Buch ist aus einer vierstündigen Vorlesung „Mathematische Statistik“, die wir an der Technischen Universität München für Studierende in Mathematik mit Schwerpunkt Finanz- und Wirtschaftsmathematik gehalten haben, entstanden.

Der Inhalt des Buches gliedert sich wie folgt: Im ersten Kapitel werden die später benötigten Konzepte der Wahrscheinlichkeitstheorie kurz vorgestellt.

Der zentrale Begriff eines statistischen Modells und insbesondere die Klasse der exponentiellen Familien werden im zweiten Kapitel eingeführt. Neben dem klassischen statistischen Modellansatz wird auch der Bayesianische Modellansatz diskutiert, welcher mit der Entwicklung von Markov Chain Monte Carlo Verfahren in jüngster Zeit sehr an Bedeutung gewonnen hat.

Im dritten Kapitel wenden wir uns den Schätzverfahren zu, wobei wir die Momentenmethode, Kleinste-Quadrate-Verfahren und Maximum-Likelihood-Schätzer (MLS) in ein- und mehrdimensionalen Modellen beschreiben. Es schließen sich das numerische Fisher-Scoring-Verfahren und Bayesianische a-posteriori-Modusschätzer an.

Im vierten Kapitel werden Vergleichskriterien von Schätzverfahren entwickelt. Dabei folgen wir im ersten Teil der klassischen Theorie nach Lehmann-Scheffé und studieren den zentralen Begriff eines gleichförmig besten Schätzers (Uniformly Minimal Variance Unbiased Estimator - kurz UMVUE). Die Bestimmung solcher Schätzer wird anhand zahlreicher Beispiele gezeigt. Im zweiten Teil widmen wir uns der asymptotischen Theorie der Schätzfolgen und analysieren Konsistenz, asymptotische Normalität und asymptotische Effizienz. Im Weiteren wird die Fisher Information eingeführt und ihr Zusammenhang mit der Informationsungleichung aufgezeigt.

Zur Bestimmung der Präzision eines Schätzverfahrens wird im fünften Kapitel der Begriff eines Intervallschätzers eingeführt. Dieser entspricht im klassischen Ansatz dem Konfidenzintervall, und im Bayesianischen Ansatz dem „Credible Interval“. Anschließend entwickeln wir das Konzept des statistischen Hypothesentestes und schließen mit der Dualität zwischen Hypothesentests und Konfidenzintervallen.

In Kapitel 6 wird die Optimalitätstheorie nach Neyman und Pearson behandelt. Es zeigt sich, dass die Anwendbarkeit dieser Konstruktion von optimalen Tests auf eine kleine Klasse von Testproblemen beschränkt ist, weswegen im zweiten Teil der verallgemeinerte Likelihood-Quotienten-Test eingeführt und an mehreren Beispielen illustriert wird. Konfidenzintervalle können nun mit Hilfe der oben angesprochenen Dualität bestimmt werden.

Das abschließende Kapitel stellt lineare Modelle vor und wir zeigen, dass die klassisch auftretenden Kleinste-Quadrate Schätzer als UMVUE-Schätzer identifiziert werden können. Die Optimalität dieser Schätzer wird mit Hilfe des Theorems von Gauß und Markov bewiesen. Hiernach leiten wir verallgemeinerte Likelihood-Quotienten-Tests her und illustrieren in der Anwendung wichtige Modellklassen wie multiple lineare Regression und Varianzanalyse (ANOVA) an Datenbeispielen.

Zu guter Letzt möchten wir uns bei den Studierenden für die zahlreichen Rückmeldungen bezüglich der ersten Skriptversionen bedanken. Insbesondere danken wir Stephan Haug, Aleksey Min, Jan Mai, Eike Christian Brechmann und Jakob Stöber für ihre Korrekturhilfen und Damir Filipović für seinen wichtigen Hinweis. Ein ganz besonderer Dank gilt Susanne Vet-

ter für ihre fabelhafte und unermüdliche Hilfe mit welcher sie das Skriptum um viele Quantensprünge verbessert hat. Die Zusammenarbeit mit Clemens Heine vom Springer Verlag war sehr professionell und stets hilfreich.

München & Leipzig,
22. Januar 2011

Claudia Czado und Thorsten Schmidt

Inhaltsverzeichnis

1. Grundlagen der Wahrscheinlichkeitstheorie und Statistik	1
1.1 Grundbegriffe der Wahrscheinlichkeitstheorie	1
1.2 Klassische Verteilungen der Statistik	9
1.2.1 Die Multivariate Normalverteilung	18
1.3 Bedingte Verteilungen	20
1.4 Grenzwertsätze	24
1.4.1 Referenzen	28
1.5 Aufgaben	29
2. Statistische Modelle	37
2.1 Formulierung von statistischen Modellen	39
2.2 Suffizienz	43
2.3 Exponentielle Familien	49
2.4 Bayesianische Modelle	57
2.4.1 Referenzen	63
2.5 Aufgaben	63
3. Schätzmethoden	71
3.1 Substitutionsprinzip	72
3.1.1 Häufigkeitssubstitution	73
3.1.2 Momentenmethode	75
3.2 Methode der kleinsten Quadrate	77
3.2.1 Allgemeine und lineare Regressionsmodelle	78
3.2.2 Methode der kleinsten Quadrate	80
3.2.3 Gewichtete Kleinste-Quadrate-Schätzer	83
3.3 Maximum-Likelihood-Schätzung	83
3.3.1 Maximum-Likelihood in eindimensionalen Modellen	86
3.3.2 Maximum-Likelihood in mehrdimensionalen Modellen	92
3.3.3 Numerische Bestimmung des Maximum-Likelihood-Schätzers	93

3.4	Vergleich der Maximum-Likelihood-Methode mit anderen Schätzverfahren	95
3.5	Anpassungstests	96
3.6	Aufgaben	96
4.	Vergleich von Schätzern: Optimalitätstheorie	103
4.1	Schätzkriterien	103
4.2	UMVUE-Schätzer	108
4.3	Die Informationsungleichung	115
4.3.1	Anwendung der Informationsungleichung	118
4.4	Asymptotische Theorie	119
4.4.1	Konsistenz	120
4.4.2	Asymptotische Normalität und verwandte Eigenschaften	122
4.4.3	Asymptotische Effizienz und Optimalität	126
4.4.4	Asymptotische Verteilung von Maximum-Likelihood-Schätzern	128
4.5	Aufgaben	130
5.	Konfidenzintervalle und Hypothesentests	139
5.1	Konfidenzintervalle	139
5.1.1	Der eindimensionale Fall	140
5.1.2	Der mehrdimensionale Fall	145
5.1.3	Bayesianischer Intervallschätzer	146
5.2	Das Testen von Hypothesen	147
5.2.1	Fehlerwahrscheinlichkeiten und Güte	149
5.2.2	Der p-Wert: Die Teststatistik als Evidenz	154
5.2.3	Güte und Stichprobengröße: Indifferenzzonen	155
5.3	Dualität zwischen Konfidenzintervallen und Tests	157
5.3.1	Aus Konfidenzintervallen konstruierte Tests	158
5.3.2	Aus Tests konstruierte Konfidenzintervalle	158
5.4	Aufgaben	159
6.	Optimale Tests und Konfidenzintervalle, Likelihood-Quotienten-Tests und verwandte Methoden	163
6.1	Das Neyman-Pearson-Lemma	163
6.2	Uniformly Most Powerful Tests	171
6.2.1	Exponentielle Familien	172
6.3	Likelihood-Quotienten-Tests	177
6.3.1	Konfidenzintervalle	179
6.4	Aufgaben	185
7.	Lineare Modelle - Regression und Varianzanalyse (ANOVA)	191
7.1	Einführung	191
7.1.1	Das allgemeine lineare Modell	193
7.1.2	Die Matrixformulierung des linearen Modells	195

7.2	Schätzung in linearen Modellen	197
7.2.1	Die kanonische Form	198
7.2.2	UMVUE-Schätzer	200
7.2.3	Projektionen im linearen Modell	201
7.2.4	Der Satz von Gauß-Markov	209
7.2.5	Schätzung der Fehlervarianz	210
7.2.6	Verteilungstheorie und Konfidenzintervalle	211
7.3	Hypothesentests	213
7.3.1	Likelihood-Quotienten-Test	214
7.3.2	Beispiele: Anwendungen	220
7.4	Varianzanalyse	223
7.4.1	ANOVA im Einfaktorenmodell	224
7.4.2	ANOVA im Mehrfaktormodell	227
7.4.3	Referenzen	231
7.5	Aufgaben	232
A	Resultate über benutzte Verteilungsfamilien	235
A1	Liste der verwendeten Verteilungen	235
B	Tabellen	237
	Exponentielle Familien	237
C	Verzeichnisse	239
	Tabellenverzeichnis	239
	Abbildungsverzeichnis	240
	Liste der Beispiele	241
	Liste der Aufgaben	244
	Literaturverzeichnis	249
	Sachverzeichnis	251