

## **Book Reviews**

Editor: F. Konecny

**Monte Carlo Methods in Bayesian Computation.** (M.-H. Chen, J.G. Ibrahim)

*Sylvia Frühwirth-Schnatter*

**Stat Labs, Mathematical Statistics Through Applications.** (D. Nolan, T. Speed)

*Werner G. Müller*

**Deskriptive Statistik. Eine Einführung mit SPSS für Windows mit Übungsaufgaben und Lösungen.** (H. Toutenburg)

*Christine Duller*

**Probability via Expectation.** (P. Whittle)

*Immanuel M. Bomze*

**Recursive Partitioning in the Health Sciences.** (H. Zhang und B. Singer)

*Michael G. Schimek*

**Einführung in die Mehrebenenanalyse.** (U. Engel)

*Wilfried Grossmann*

**Intelligent Data Analysis.** (M. Berthold, D.J./ Hand)

*Ernst Stadlober*

**Probabilistic Networks and Expert Systems.** (R.G. Cowell, P.A. Dawid, S.L. Lauritzen)

*Wolfgang Janko*

**Scan Statistics and Applications.** (J. Glatz)

*Walter Böhm*

**Modern Applied Statistical Learning Theory.** (W.N. Venables, B.D. Ripley)

*Norbert Kusolitsch*

**Dynamic Nonlinear Econometric Models.** (B.M. Pötscher, I.R. Prucha)

*Dietmar Trummer*

**Optimal Tests bei Intervallwahrscheinlichkeit.** (T. Augustin)

*Reinhard Viertl*

M.-H. CHEN, Q.-M. SHAO und J.G. IBRAHIM, **Monte Carlo Methods in Bayesian Computation**. Springer series in statistics. New York: Springer-Verlag, 2000, xiii+386 S., DM 144.-, ISBN 0-387-98935-8 .

In den letzten Jahren fanden auch in Österreich Bayesianische Methoden vermehrt Anwendung bei der Analyse von komplexen Problemstellungen unter anderem in der Marktforschung, in der Medizin oder bei der Analyse von Finanzmarktdaten. Mit der Komplexität der Modelle steigt allerdings die Anforderung an die Fertigkeit der Statistikerin oder des Substanzwissenschaftlers im Umgang mit modernen Monte Carlo Techniken. Ein Buch wie *Monte Carlo Methods in Bayesian Computation* von Chen/Shao/Ibrahim wird dann rasch zu einem unentbehrlichen Hilfsmittel.

Bei der Bayesianischen Analyse von komplexen Modellen muss man zunächst entscheiden, welche Simulationstechnik anzuwenden ist, um effizient aus der oft extrem hoch-dimensionalen a-posteriori Dichte zu simulieren. Die von den Autoren vorgestellten Markov Chain Monte Carlo Simulationstechniken gehen über die Standardtechnik des Gibbs sampling weit hinaus. Der hit-and-run-Algorithmus etwa adaptiert Ideen, die von deterministischen Optimierungsproblemen bekannt sind. Außerdem werden wichtige Aspekte wie der Einfluss der Parametrisierung auf die Effizienz der Simulationsmethode angesprochen.

Stehen Simulationen zur Verfügung, so muss man dann aus den Simulationen relevante Größen schätzen. Das Beispiel der Schätzung der Normierungskonstanten der a-posteriori-Dichte, die eine wichtige Rolle bei der Bayesianischen Modellwahl spielt, zeigt, dass diese Aufgabe zu einer echten Herausforderung werden kann. Die Autoren widmen der Schätzung der Normierungskonstante ein ganzes Kapitel und diskutieren neben Standardtechniken wie importance sampling auch hochaktuelle Methoden wie bridge sampling.

Ein besonders interessantes Kapitel ist der Bayesianischen Variablenwahl gewidmet. Man kann mit den vorgestellten Methoden Situationen bewältigen, bei denen sich die Dimension des Modelles ständig verändert. Das Problem tritt etwa bei einem klassischen Regressionsmodell auf, wenn man *simultan* Variablenwahl und Parameterschätzung durchführen möchte. Das Potential dieses Ansatzes ist auch in der internationalen community noch nicht voll ausgeschöpft worden. Die Lektüre dieses Abschnitts könnte daher den entscheidenden Impuls geben, sich an spannende Fragestellungen wie etwa die Schätzung von random-effects Modellen bei *simultaner* Entscheidung, ob die Effekte fixed oder random sind, heranzuwagen.

*Sylvia Frühwirth-Schnatter  
Institut für Statistik  
Wirtschaftsuniversität Wien*

D. NOLAN und T. SPEED, **Stat Labs, Mathematical Statistics Through Applications**. Springer Texts in Statistics. New York: Springer Verlag, 335 S., öS 504.-, ISBN 0-387-98974-9.

Das vorliegende Textbuch wendet sich an Studierende mit nur geringen Vorkenntnissen der mathematischen Statistik. Es ist von seiner Konzeption als eine Einführung in die Statistik gedacht, aufbauend auf aus konkreten Anwendungen entwickelten Fragestellungen.

Zwölf Sektionen behandeln unterschiedliche Themenkreise – die deskriptiven Methoden, einfache und geschichtete Zufallsstichproben, Schätz- und Testverfahren, Kontingenztafeln, Poisson Prozesse, einfache und multiple lineare Regression, die in den Politikwissenschaften zuletzt populäre „ökologische“ Regression, Varianzanalyse und Wirkungsflächenanalyse. Jedes Kapitel widmet sich einer konkreten Anwendung und der zugehörigen, bzw. zugeordneten, Methodologie. Dem Text liegt die – durchaus begrüßenswerte – Philosophie zu Grunde, dass ein zufriedenstellendes Verständnis der Statistik nur im engen Konnex mit tatsächlichen Fragestellungen entwickelt werden kann. In dieser Hinsicht ist das Buch in seiner Radikalität einzigartig und eine erfrischende „Erneuerung“ aufbauenden Statistikerunterrichts.

Die inhaltliche Umsetzung kann leider nicht immer diesem Anspruch der Modernität mithalten. Selten überzeugen die Relevanz und der natürliche Zusammenhang von Anwendung und Methode, oft wirkt der Theorieteil allzu bemüht. Details wie der nahezu anachronistisch anmutende Verzicht von Matrixnotation irritieren zudem. Die als Einstieg – und zur Illustration von deskriptiven Verfahren – gedachte Analyse der Kindersterblichkeit ist zunächst bestenfalls schlecht motiviert und entwirrt sich dem Leser erst bei neuerlicher Aufnahme in einem späteren Kapitel zur Regression. Die Schwerpunktsetzung auf schließende Statistik ist für ein Einführungsbuch außerdem zu bezweifeln und ist im Unterricht wohl weder wünschenswert noch realisierbar. Von einem konzeptorientierten Lehrprogramm, wie es in fast idealer Form in Griffiths et al. (1998) umgesetzt wird, sind die Autoren jedenfalls weit entfernt.

Als voll gelungen würde ich nur die beiden Kapitel, die sich mit etablierten Anwendungsbeispielen befassen, nämlich „Can She Taste the Difference“ (Kontingenztafeln) und „Helicopter Design“ (Wirkungsflächen) bezeichnen. Dem Rest fehlt entweder die innere Harmonie oder die Zugkraft für den Hörsaal (als Beispiel etwa das an Langweile kaum überbietbare „Who Plays Video Games?“). Wer ersteres sucht ist sicher mit dem Klassiker Cox und Snell (1981) besser bedient, mehr Spaß hat man jedenfalls mit Schaeffer et al. (1996).

Die Autoren haben sich ihre Lorbeeren sicher schon mit der Propagierung der Anwendungszentriertheit verdient. Ihr Lehrbuch wird verdienterweise und zum Glück für die Leser Nachahmer finden. In der Zwischenzeit muss das Buch für einen gelungenen Hörsaaleinsatz aber stark adaptiert werden. Allein zum Kennenlernen des Konzepts reicht vermutlich die Lektüre von Nolan und Speed (1999) aus.

### **Literatur**

- D.R. Cox and E.J. Snell. *Applied Statistics, Principles and Examples*. Chapman and Hall, London, 1981.
- D. Griffiths, W.D. Stirling, and K.L. Weldon. *Understanding Data, Principles and Practice of Statistics*. Jacaranda Wiley, Milton, 1998.
- D. Nolan and T.P. Speed. Teaching Statistics Theory Through Applications. *The American Statistician*, 53:370-375, 1999.
- R.L. Schaeffer, M. Gnanadesikan, A. Watkins, and J.A. Witmer. *Activity-Based Statistics*. Springer, New York, 1996.

Werner G. Müller  
Institut für Statistik  
Wirtschaftsuniversität Wien

H. TOUTENBURG, **Deskriptive Statistik. Eine Einführung mit SPSS für Windows mit Übungsaufgaben und Lösungen.** Berlin: Springer Verlag, 2000, x+285 S., öS 292.-, ISBN 3-540-67169-2.

Das Buch „Deskriptive Statistik: eine Einführung mit SPSS für Windows und Übungsaufgaben“ liegt nun als dritte, neu bearbeitete und erweiterte Auflage vor, wobei der Unterschied zur zweiten Auflage eher gering ist. Der Aufbau ist klar strukturiert und beinhaltet die Kapitel Grundlagen, Häufigkeitsverteilungen, Maßzahlen für eindimensionale Merkmale, Maßzahlen für den Zusammenhang zweier Merkmale, Zweidimensionale quantitative Merkmale: Lineare Regression, Zeitreihen und zuletzt Verhältniszahlen und Indizes. Im Kapitel Grundlagen werden die Grundbegriffe eingeführt, wobei zusätzlich zu den Standardbegriffen auch Begriffe eingeführt werden, die über den Standard hinausgehen, z.B. wird eine Unterscheidung quantitativer Merkmale in extensive und intensive Merkmale vorgenommen. Im zweiten Kapitel werden weitere Begriffe definiert, wobei bei diesen Definitionen Wert auf mathematisch präzise Formulierungen gelegt wird, beispielsweise wird der Begriff der absoluten Häufigkeit über die Indikatorfunktion eingeführt - statistisch oder mathematisch orientierte Leserinnen und Leser sehen dies sicher als Bereicherung. In diesem Kapitel wird auch erstmals der Bezug zu dem Programmpaket SPSS hergestellt, indem ein SPSS-Output zum Bereich Häufigkeiten erklärt wird. Interessant wäre in diesem Zusammenhang aber auch der Weg, um zu solch einem Output zu gelangen. Neben Möglichkeiten zur grafischen Darstellung von Häufigkeitsverteilungen wird in diesem Kapitel auch der Begriff der Kerndichteschätzer angerissen. Als Maßzahlen für eindimensionale Merkmale werden die gängigen Maßzahlen der Lage, der Streuung, der Schiefe, der Wölbung und der Konzentration dargestellt. Auch hier sind die Bezüge zu dem Programmpaket SPSS eher dürftig angesiedelt und auf Feinheiten, die im Umgang mit SPSS zu beachten sind, wird nicht eingegangen. Auch in diesem Kapitel werden Möglichkeiten zur grafischen Darstellung erklärt, wie z.B. Q-Q-Plots, Box-Plots oder Lorenzkurven. Im vierten Kapitel werden sehr umfassend die Maßzahlen für den Zusammenhang zweier Merkmale dargestellt, wobei auch sonst eher stiefmütterlich behandelte Maßzahlen, wie z.B. die Lamda-Maße ihre Beachtung finden. Das nächste Kapitel ist der linearen Regression gewidmet, auch dies sehr ausführlich und mit kleinen Abstechern zur multiplen linearen Regression, zur nichtlinearen und polynomialen Regression und zur linearen Regression mit kategorialen Regressoren. Das Kapitel Zeitreihen ist kurz gehalten, dafür ist dann das Kapitel der Verhältniszahlen und Indizes wiederum sehr ausführlich. Neben den üblichen Begriffen werden auch die Grundbegriffe der Lebensdaueranalyse eingeführt. Jedes dieser Kapitel wird mit Übungsaufgaben und Kontrollfragen abgeschlossen, die am Ende des Buches gelöst werden.

Die systematische Darstellung der Grundbegriffe der deskriptiven Statistik zählt auf jeden Fall zu den Vorteilen dieses Buches. Positiv fällt auf, dass auch Teilbereichen und Begriffen Beachtung geschenkt wird, die in anderen Büchern zu den Grundlagen der Statistik vielleicht manchmal etwas kurz kommen. Das Stoffgebiet wird leicht verständlich erklärt, ist mit vielen, einfach nachvollziehbaren Beispielen untermauert und wird durch Aufgaben mit Lösungen abgerundet. Das Buch eignet sich daher sehr gut zum Selbststudium oder als Ergänzung zu einem Einführungskurs Statistik. Als Kritik muss angeführt

werden, daß ich mir unter einer „Einführung mit SPSS für Windows“ - so der Untertitel - viel mehr Bezug zum Programmpaket erwartet habe. So gut dieses Buch tatsächlich ist, als Hilfestellung für SPSS ist es nicht geeignet. Der trotzdem positive Gesamteindruck wird ergänzt durch ein angenehmes Druckbild und gute optische Aufbereitung, so dass die Lektüre dieses Werkes durchaus empfohlen werden kann.

*Christine Duller  
Institut für Angewandte Statistik  
Johannes Kepler Universität Linz*

Peter WHITTLE, **Probability via Expectation**. 4th Edition, Springer Texts in Statistics. New York: Springer-Verlag, 2000, xvi+352 S., EURO 69.95, ISBN 0-387-8955-2.

Die nun vorliegende vierte Auflage eines erfolgreichen Werks, das 1970 entstand und 1982 auch ins Russische übersetzt wurde, bietet gegenüber der dritten, vollständig überarbeiteten Auflage aus dem Jahre 1992 nur kleinere Modifikationen, wurde jedoch um neue Kapitel über (stochastische) dynamische Optimierung, Allokationsoptimierung, Optionenbewertung und large deviations erweitert.

Nach wie vor geht der Autor den Weg der Axiomatik für Erwartungswerte anstelle der für Wahrscheinlichkeiten, was auch 1970 vielleicht für Praktiker unorthodox, jedoch in einer funktionalanalytisch orientierten Maßtheorie schon wohl ein halbes Jahrhundert früher durchaus üblich war (Stichwort Daniell-Integral als lineares, positives Funktional – leider wird dieser Name von Whittle nicht erwähnt). Dieses Konzept wird relativ konsequent durch das Buch verfolgt, wenn auch in Kauf genommen werden muss, dass so zentrale Begriffe wie fast sichere Konvergenz erst sehr spät behandelt werden können.

Das sehr dicht geschriebene Buch enthält für seinen Umfang eine beträchtliche Fülle an Information, die sich Nicht-Fachleuten manchmal nicht leicht erschließt, zumal der Aufbau leider nicht immer logisch ist (Martingale werden etwa erst nach dem Kapitel Optionen behandelt), und manch sinnstörender Druckfehler die diversen Neuauflagen überlebt hat. Der Index ist ebenfalls recht spartanisch gehalten. Dieses Werk wird wohl hauptsächlich für Interessenten an dynamischen (zeitabhängigen) stochastischen Phänomenen, auch unter Optimierungsaspekten – einem Gebiet, auf dem der Autor seit geraumer Zeit zurecht als Koryphäe weltweit berühmt ist – als ergänzende Lektüre nützlich sein.

*Immanuel M. Bomze  
Institut für Statistik und Decision Support Systems  
Universität Wien*

Heping ZHANG und Burton SINGER, **Recursive Partitioning in the Health Sciences**. New York: Springer-Verlag, 1999, Series Statistics for Biology and Health, xii+226 S., ATS 869.–, ISBN 0-387-98671-5.

Die Kernthemen dieses Buches sind Klassifikations- und Regressionsbäume (CART) sowie adaptive Spline-Ansätze in den Gesundheitswissenschaften. Besonderes Augenmerk

wird auf die Knotenwahl gelegt, wobei der rekursive Partitionierungsansatz breiten Raum einnimmt. Aber nicht nur deshalb ist der Buchtitel gerechtfertigt. Auch verwandte Ansätze werden unter diesem Aspekt dargestellt und diskutiert.

Ausgangspunkte für das Werk sind einerseits die Kritik an den klassischen parametrischen Regressionsverfahren und andererseits die grosse Bedeutung von Klassifikationsmethoden in der medizinischen Forschung. Ersteres wird an einer Betrachtung der logistischen Regression exemplifiziert und letzteres an der bekannten Yale Pregnancy Outcome Study (ein laufendes Beispiel im Buch) motiviert. Abgesehen davon ist mir seit dem Klassiker von Breiman et al. (1984) kein Buch bekannt, das sich vorrangig mit CART auseinandersetzt. Wie noch auszuführen sein wird, werden seither erzielte, wesentliche methodische Ergebnisse abgehandelt, die sonst nur über Zeitschriftenartikel zugänglich sind. Viele dieser Beiträge stammen von Prof. Zhang selbst. Ausserdem nimmt das Buch auf seine freie Software STREE (unter Unix, Linux und Windows – die genannte Webpage <http://peace.med.yale.edu/pub> für Downloads habe ich auf Richtigkeit geprüft) Bezug. Deshalb werden nicht wenige potentielle Leser bedauern, dass das vorliegende Werk alle Motivation aus dem medizinischen Bereich holt. Für medizinische Statistikerinnen und Statistiker ist es freilich ein ideales Buch, das eine gute Einführung in die Methodik leistet, um sich dann in Spezialthemen zu vertiefen, ohne an Lesbarkeit zu verlieren. Die gute Sprache, die klare Gliederung und das ansprechende Format sind hervorsteckende Merkmale.

Das Buch umfasst folgende Themen: (1) Klassifikationsbäume für binäre Responses, wobei auch ein rezenter Pruning-Ansatz und Methoden bei fehlenden Werten dargestellt werden (äußerst wichtig in der Medizin). (2) Risikofaktorenanalyse mittels baumbasierter Stratifikation. (3) Die Analyse zensierter Daten. Hierbei werden Überlebensbäume den klassischen Verfahren, wie Coxsche Proportionale Hazardregression, gegenübergestellt. Dabei wird auch auf Likelihood-basierte Aufteilungskriterien eingegangen. (4) Regressionsbäume und adaptive Splines für kontinuierliche Responses. Friedmans MARS-Methode („multivariate adaptive regression splines“) steht hier im Zentrum der Betrachtung. Zhang, selbst Stanford-Absolvent und an der Weiterentwicklung von MARS beteiligt, liefert hier meines Erachtens keine ausgewogene Sicht. Ich möchte in diesem Zusammenhang auf einen Beitrag von Smith, Kohn und Yau (2000) verweisen. (5) Die Analyse longitudinaler Daten. Diese Thematik mag auf den ersten Blick verwundern. Die Autoren belehren uns jedoch eines Besseren: auch hier gibt es neuerdings Partitionierungsansätze in Form der MASAL-Methode („multivariate adaptive splines for the analysis of longitudinal data“). Ebenso werden Modelle für gemischte Effekte und ein spezielles semi-parametrisches Modell gestreift. Schlussendlich wird (6) die Analyse multipler diskreter Responses behandelt. Dabei wird statistisches Neuland betreten. Die Diskussion nimmt ihren Ausgang bei den parametrischen Modellen (loglineare Modelle, marginale Modelle und Frailty Modelle). Die Entwicklung geeigneter Klassifikationsbäume steht hier erst am Anfang.

Insgesamt muss gesagt werden, dass die Autoren wirklich den Stand der aktuellen Forschung wiedergeben und mit zahlreichen Fallstudien aus den Gesundheitswissenschaften in gekonnter Weise illustrieren. Ich selbst habe aus der Lektüre viel Gewinn gezogen und empfehle sie uneingeschränkt allen Biostatistikern und Epidemiologen. Andere Leserschaften können ebenso Nutzen ziehen, weil die Beispiele auch für Fachfremde nach-

vollziehbar dargestellt sind. Etwas mehr Hinweise auf kommerzielle Software, insbesondere S-Plus (hat die umfangreichste CART-Implementation), wären von Vorteil. Auch zu STREE gibt es relativ wenig technische Informationen. Diese Schlussbemerkungen relativieren den Wert des Buches jedoch nur unwesentlich.

### Literatur

L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees*. Wadsworth, California, 1984.

M. Smith, R. Kohn, and P. Yau. Nonparametric Bayesian Bivariate Surface Estimation. In Schimek, M.G. (Hg.). *Smoothing and Regression. Approaches, Computation and Application*. Wiley, New York, 545–580, 2000.

*Michael G. Schimek*  
*Institut für Medizinische Informatik, Statistik und Dokumentation*  
*Karl–Franzens–Universität Graz*

Uwe ENGEL, **Einführung in die Mehrebenenanalyse**. Westdeutscher Verlag GmbH, Opladen, Wiesbaden, 1998, ISBN 3-531-22182-5.

Mehrstufige Auswahlverfahren spielen in der Surveystatistik traditionell eine große Rolle. Will man in der Modellierung und Analyse derartiger Daten das Stichprobendesign in einer adäquaten Art und Weise repräsentieren, so sind die Methoden der Mehrebenenanalyse, die auf der Theorie der Varianzkomponentenmodelle und der Random Coefficient Models aufbauen, vielfach ein geeignetes Werkzeug. In der englischsprachigen Literatur gibt es bereits eine Reihe von Lehrbüchern zu diesem Thema und es wurden auch spezielle Softwaresysteme für die Mehrebenenanalyse (Multilevel modelling) entwickelt, zum Beispiel das System ML-Win. Mit dem vorliegenden Buch liegt nun auch eine deutschsprachige Einführung in dieses aktuelle Gebiet vor, die sich in der Darstellung primär an Sozial- und Wirtschaftswissenschaftler und Statistiker in der Praxis wendet. Dabei wird die Mehrebenenanalyse in den allgemeinen Rahmen der Surveyforschung eingebettet. Der mathematische Formalismus wird in der Darstellung auf ein minimales Maß reduziert und der Umgang mit statistischer Terminologie ist manchmal auch ungenau (beispielsweise *Likelihood* für *Loglikelihood*). Das erste Kapitel des Buches gibt einen gelungenen Überblick über die Konzepte und Methoden bei der Durchführung von Surveys, insbesondere Surveydesign, Längsschnittforschung und Analyse von Aggregatdaten. Im zweiten Kapitel findet man eine kurze Einführung in struktursoziologische Analysen. Darauf aufbauend werden die Konzepte der Mehrebenenanalyse und deren Anwendung in der sozialwissenschaftlichen Datenmodellierung anhand von Fallbeispielen dargestellt, die auch alle im Detail durchgerechnet werden. Ein gewisser Nachteil ist die Tatsache, dass die Beispiele noch mit der etwas älteren Version des Programmsystems ML anstatt mit der neueren Version ML-Win analysiert werden. Dies ist vermutlich auch die Ursache, dass datenanalytische Überlegungen und grafische Darstellungen etwas zu kurz kommen. Zusammenfassend lässt sich feststellen, dass das Buch einen guten Überblick über die Anwendung der Methoden in der sozialwissenschaftlichen Praxis gibt. Es scheint aber weniger geeignet, dem Leser jene Fähigkeiten zu vermitteln, die zur selbständigen Ver-

wendung der Mehrebenenanalyse bei der Modellierung und Analyse von Daten notwendig sind. Der daran und an mathematisch-statistischen Fragen interessierte Leser sollte wohl besser auf die englischsprachige Literatur zurückgreifen, zum Beispiel auf das Standardwerk von H. Goldstein: *Multilevel Statistical Models*, Kendall's Library of Statistics 1999.

Wilfried Grossmann  
Institut für Statistik und Decision Support Systems  
Universität Wien

Michael BERTHOLD und David J. HAND: **Intelligent Data Analysis**. Berlin: Springer-Verlag, 1999, ix+400 S., ATS 716.-, ISBN 3-540-65808-4.

Moderne Datenanalyse ist eine sich rasch entwickelnde Technik, die in hohem Maße von den Fortschritten verschiedener wissenschaftlicher Disziplinen wie z.B. Statistik, Künstliche Intelligenz, *Machine Learning*, *Data Mining* und nicht zuletzt den Computerwissenschaften profitiert. „Intelligent Data Analysis“, das offenbar auf der Grundlage dieses ganzheitlichen Standpunktes entstanden ist, gibt in einer Serie von 10 Artikeln einen Überblick über die wichtigsten Methoden und aktuellsten Entwicklungen in der modernen Datenanalyse. Die ersten drei Artikel beinhalten eine Einführung in fundamentale statistische Begriffe und mittlerweile klassische statistische Verfahren. Viele Themen wie z.B. *Resampling* Methoden, Generalisierte Lineare Modelle sowie klassische Multivariate Methoden werden diskutiert, ohne natürlich erschöpfend behandelt werden zu können. Der immer größer werdenden Bedeutung Bayes'scher Statistik wird im vierten Artikel Rechnung getragen; insbesondere der Abschnitt über Bayes'sche Netze wird Interesse finden. Der Beitrag über Zeitreihenanalyse beschreibt neben linearen vor allem nichtlineare Ansätze. Die folgenden Artikel sind einem Methodenspektrum gewidmet, das sich insbesondere im Bereich Künstliche Intelligenz, *Machine Learning* und *Data Mining* etabliert hat, angefangen von *Rule Induction*, Neuronalen Netzen, *Fuzzy Logic* bis zu stochastischen Optimierungsverfahren, wie *Simulated Annealing* und Genetische Algorithmen. Abgerundet wird dieses Buch von der Beschreibung interessanter Anwendungen und von Software Paketen für intelligente Datenanalyse. „Intelligent Data Analysis“ ist allen zu empfehlen, die sich rasch einen Überblick über die aktuellen Entwicklungen in diesem Bereich verschaffen wollen. Die einzelnen Artikel sind klar und auch für den weniger versierten Leser verständlich geschrieben und können unabhängig voneinander gelesen werden. Wer sich intensiver mit den einzelnen Methoden auseinandersetzen möchte, kann dieses Buch, das mit einer umfangreichen Bibliographie ausgestattet ist, als Ausgangspunkt verwenden.

Ernst Stadlober  
Institut für Statistik  
Technische Universität Graz

Robert G. COWELL, Philip A. DAWID, Steffen L. LAURITZEN, David J. SPIEGEL-THALER: **Probabilistic Networks and Expert Systems**. Berlin: Springer-Verlag, 1999, xii+321 S., ATS 1015.–, ISBN 0-387-98767-3.

Das Buch behandelt Bayes'sche Netzwerke, ein Thema, das nach einer ausführlichen, Auseinandersetzung über die Modellierung von Unsicherheit in Expertensystemen in den frühen 80er Jahren schließlich durch Judea Pearl's<sup>1</sup> klassisches Buch bekannt wurde. Das Buch gliedert sich nach einer Einleitung in drei Hauptabschnitte: Die Kapitel 2 bis 5 („Logic, Uncertainty, and Probability“, „Building and Using Probabilistic Networks“, „Graph Theory“ and „Markov Properties on Graphs“) stellen die grundlegenden Ideen zur Bildung von probabilistischen Netzwerken vor. In den Kapiteln 6 bis 8 („Discrete Networks“, „Gaussian and Mixed Discrete-Gaussian Networks“ und „Discrete Multistage Decision Networks“), die den Kern des Buches darstellen, werden die wesentlichen Schlußalgorithmen und ihre Eigenschaften genau beschrieben und untersucht. In den Kapiteln 9 bis 11 („Learning About Probabilities“, „Checking Models Against Data“ und „Structural Learning“) werden verschiedene Themen inklusive einschlägiger bisher nicht publizierter Resultate dargestellt, die sich überwiegend mit Lernen und Modellkritik im Lichte von Beobachtungsdaten befassen. Das Buch schließt mit drei Appendizes: „Bayes'sche Analyse von diskreten Daten“, „Gibbs'sche Stichprobenziehungen“ und „Information und Software auf dem World Wide Web“.

Das Buch stellt nicht den Anspruch, ein durchgehender Leitfaden zu dem Gebiet „Graphische Modellierung“ zu sein. Es behandelt z.B. das Thema der Muster-Erkennungs-Algorithmen und andere verwandte Themen nicht. Es versucht vielmehr, im wesentlichen den Kern des Themengebietes, die Theorie und die Anwendungen von exakten Methoden in probabilistischen Netzwerken darzustellen.

*Wolfgang Janko  
Institut für Informatik  
Technische Universität Wien*

Joseph GLATZ: **Scan Statistics and Applications**. Basel: Birkhäuser Verlag, 1999, xv+324 S., ATS 1227.–, ISBN 3-7643-4041-X.

Ich hatte Gelegenheit, die Entstehung dieses bemerkenswerten Buches gewissermaßen aus der ersten Reihe mizu erleben. Im Sommer 1997 veranstaltete N. Balakrishnan anlässlich der Emeritierung unseres gemeinsamen Lehrers Sri Gopal Mohanty ein Meeting an der McMaster University in Hamilton, Kanada. Ein nicht unbeträchtlicher Teil der Vorträge war *Scan* Statistiken gewidmet, und so war es nur natürlich, daß in einer der Coffee Breaks die Idee geboren wurde, ein Resümee über den gegenwärtigen Stand der Forschung auf diesem Gebiet zu ziehen und diesen Band herauszugeben.

---

<sup>1</sup>J. Pearl, *Probabilistic Inference in Intelligent Systems*, Morgan Kaufmann, San Mateo, California, 1988

Mit dem Erscheinen dieses Buches wurde eine wichtige Lücke in der Fachliteratur geschlossen, denn obwohl *scan* Statistiken immer wieder in Anwendungen auftreten, insbesondere bei Tests auf Uniformität gegen Cluster-Alternativen, und auch eine Fülle von (teils noch ungelösten) technischen Problemen aufwerfen, gab es bisher kein Buch, das den Stand der Forschung in geeigneter Weise zusammengefaßt und damit einem breiteren Publikum zugänglich gemacht hätte. Das Bemühen, ein wenig bekanntes Forschungsgebiet populärer zu machen, ist auch die Leitidee in der gelungenen Gestaltung des Buches. Den Anfang bildet eine von Glaz und Balakrishnan verfaßte Einführung in *Scan* Statistiken, so daß jemand, der noch nie mit diesem Thema zu tun hatte, leichten Zugang findet. Insbesondere wird in diesem Kapitel die enge Beziehung zwischen *Order Statistics* und *Scan* Statistiken erläutert und eine umfangreiche Bibliographie (96 Quellen) über das Thema gegeben. Der Rest des Buches gliedert sich in drei Teile: Diskrete *Scan* Statistiken, stetige *Scan* Statistiken und Anwendungen.

Bei diskreten *Scan* Statistiken stellt sich das folgende Problem: es seien  $X_1, \dots, X_N$  identisch verteilte, unabhängige Zufallsvariable, nicht-negativ und ganzzahlig. Für ein fixes  $m, 2 \leq m \leq N$  bildet man die gleitenden Summen

$$Y_n = \sum_{n-m+1 \leq i \leq n} X_i, \quad n = m, m+1, \dots, N$$

und definiert die diskrete *Scan* Statistik

$$S_m = \max_n Y_n.$$

Das Problem ist nun, Verteilung und Momente von  $S_m$  zu bestimmen.

Ähnlich ist die Situation im stetigen Fall. Wenn etwa  $\{X_t, t \geq 0\}$  ein Poisson Prozeß ist, dann definiert man den *Scan* Prozeß durch  $\{Y_t(w) = X_{t+w} - X_t, t \geq 0\}$  und die *Scan* Statistik ist bei Fensterbreite  $w > 0$ :

$$S_w = \max_{0 \leq t \leq T-w} Y_t(w).$$

So einfach die Formulierung des Problems, so schwierig seine Lösung! Für einfache Spezialfälle stehen kombinatorische Methoden zur Verfügung, mit deren Hilfe man Verteilung und Momente von  $S_m$  bzw.  $S_w$  bestimmen kann. Allerdings sind diese Ergebnisse wegen ihrer Komplexität von verhältnismäßig geringem praktischen Wert. Deshalb sind Approximationen von großer Bedeutung. Sie bilden ein immer wiederkehrendes Thema in fast allen Beiträgen, sind aber insbesondere in den Aufsätzen von Chen und Glaz, von Alm, Huffer und Lin sowie Mansson im Zentrum des Interesses. Der letzte Teil des Buches ist diversen Anwendungen gewidmet. Zwei interessante Beiträge widmen sich dem Problem des *Pattern Matching* in DNA-Molekülen, der letzte Artikel (von Martin Kulldorf) diskutiert Anwendungen räumlicher *Scan* Statistiken in der Epidemiologie, Archäologie, Astronomie und Regionalplanung.

Eine kritische Bemerkung zum Abschluß: In den Beiträgen, in welchen Approximationsprobleme diskutiert werden (und dies ist ein Schwerpunkt des Buches), wird nirgends auf *asymptotically almost deterministic processes* im Sinne von P. Révész Bezug genommen. Das ist insofern bemerkenswert, als dieser methodische Ansatz auf überaus elegante Weise die erstaunliche Genauigkeit der Approximationen erklären würde. Aber das kann man schwerlich den Herausgebern anlasten.

Alles in allem ein ausgezeichnetes Buch, das nicht nur einen up to date Überblick über die aktuelle Forschung bietet, sondern auch geeignet ist, das Interesse so manchen Lesers und Leserin zu wecken, selbst in diese Richtung zu forschen.

*Walter Böhm  
Institut für Statistik  
Wirtschaftsuniversität Wien*

William N. VENABLES, Brian D. RIPLEY **Modern Applied Statistics with S-Plus**. Berlin: Springer-Verlag, 1999, xii+501 S., ATS 942.–, ISBN 0-387-98825-4.

„Modern Applied Statistics with S-PLUS“ ist der Klassiker unter den anwendungsorientierten Büchern über die Datenanalyseumgebung S-PLUS. Während sich die Bücher der Autoren der S-PLUS zugrunde liegenden Sprache S (z.B. Becker, Chambers & Wilks (1988): The New S Language) vor allem an Programmierer richten, haben Venables & Ripley den Daten analysierenden Benutzer als Zielpublikum. Nach einer grundlegenden Einführung in die Sprache S (Kapitel 1 und 2), Grafik (Kapitel 3) sowie Programmieren in S (Kapitel 4) behandeln die Kapitel 5 bis 17 viele Teilgebiete der klassischen und moderneren Statistik in einführender Form.

Die meisten dieser anwendungsorientierten Kapitel sind in sich geschlossen und zeigen am Beispiel konkreter Datensätze, wie man S-PLUS als statistisches Analysewerkzeug effizient verwenden kann. Statistische Theorie wird in Ansätzen geboten und macht meist ein paralleles Nachschlagen in der theoretischen Literatur, statistische Grundkenntnisse in den jeweiligen Gebieten werden aber vorausgesetzt. Alle Funktionen und Datensätze, die im Buch verwendet werden (und nicht Teil von S-PLUS selber sind) können von der Homepage der Autoren über das Internet bezogen werden. Dies erleichtert den Einsatz des Buches in der Lehre ungemein, da alle Fallbeispiele ohne großen Aufwand selber am Rechner nachvollzogen werden können.

Die zur Rezension vorliegende 2. Auflage behandelt nur die Kommandozeilen-orientierte Verwendung von S-PLUS, d.h., die Analyse und Visualisierung von Daten mittels Befehlseingabe am Prompt. Die mittlerweile erschienene 3. Auflage (Springer, 1999) behandelt auch die graphische Oberfläche der neueren Windows-Versionen wie S-PLUS 2000, behält aber trotzdem ihr Hauptaugenmerk auf der direkten Verwendung der Sprache S. Dies macht das Buch auch aus einer sehr subjektiven Sicht des Rezensenten sehr interessant: es kann über weite Strecken (de facto alles außer Trellis Grafiken) auch als Handbuch für die zweite verfügbare Implementierung der Sprache S verwendet werden, das Open Source Projekt R. Auf der Homepage des Buches findet sich eine knapp 50-seitige Ergänzung („R Complements“), die die Verwendung des Buches mit R erklären.

Obwohl mittlerweile mehrere andere einführende Bücher über S erschienen sind, bleibt das Buch von Venables und Ripley wohl weiterhin das Standardwerk, das viele Benutzer von S (oder R) ständig in Griffweite neben dem Computer stehen haben. Es kann sowohl dem Anfänger wie auch erfahreneren Benutzer empfohlen werden.

*Norbert Kusolitsch  
Institut für Statistik und Wahrscheinlichkeitstheorie  
Technische Universität Wien*

Benedikt M. PÖTSCHER, Ingmar R. PRUCHA: **Dynamic Nonlinear Econometric Models**. Berlin: Springer-Verlag, 1999, xix+501 S., ATS 942.–, ISBN 0-387-98825-4.

In der modernen Zeitreihenanalyse gewinnen nichtlineare Modelle immer mehr an Bedeutung. Bekanntlich ist die Schätzung der Parameter auch in einem linearen dynamischen System ein nichtlineares Problem. Da die Theorie für lineare dynamische Systeme (für stationäre Prozesse) bereits einen gewissen Grad an Vollständigkeit erreicht hat, stellt sich die Frage, ob und wie die Ergebnisse auf nichtlineare Systeme übertragbar sind. Man kann bei der Identifikation nichtlinearer dynamischer Systeme meist zwischen der Analyse spezieller sowie der Behandlung sehr allgemeiner Modellklassen unterscheiden: In diesem Buch wurde der zweite Weg gewählt und hier wiederum der parametrische Fall.

Die Autoren leiten in diesem Buch sehr umfassende Aussagen über Probleme der Parameterschätzung eines nichtlinearen dynamischen System ab. Es werden allgemeine M-Schätzer betrachtet (wie z.B. *Least Squares* - oder Maximum Likelihood Schätzer); der Schwerpunkt liegt bei Konsistenz und asymptotischer Normalität; auch wird auf die Schätzung der Varianz - Kovarianz Matrix eingegangen. Bewußt halten die Autoren die Modelle für den datenerzeugenden Prozeß sehr allgemein und führen in der Folge die getroffenen Annahmen nicht nur immer genau an, sondern weisen auch auf die Konsequenzen dieser Annahmen hin. Dieser letzte Aspekt wird zum Teil ausgiebig analysiert, was die Darstellung sehr gut abrundet.

Es gibt wohl wenige Bücher, die so klar strukturiert sind wie das hier besprochene. Die Beweise zu den jeweiligen Kapiteln wurden in einen Anhang gestellt, was das Lesen erheblich erleichtert, weil die Formulierung der Sätze technisch meist nicht einfach ist. Inhaltlich ist das Buch vor allem in die beiden Bereiche *Konsistenz* und *Asymptotische Normalität* gegliedert, die beide ähnlich aufgebaut sind.

Darüber hinaus ist hauptsächlich die exzellente inhaltliche Aufbereitung zu loben: Die Argumentation und die Entwicklung des Themas sind immer sehr klar und logisch. Wert wird vor allem darauf gelegt, unter welchen Voraussetzungen welche Aussagen gefolgert werden können. Meistens arbeitet man sich stufenweise von der allgemeinsten Aussage (z.B. beginnt das Buch mit einem sehr allgemeinen Konsistenzsatz) zu spezielleren Sätzen vor. Auch Beweise sind klar formuliert und kurz gehalten.

Dieses Buch ist wohl ein Standardwerk auf dem Gebiet der Statistik nichtlinearer dynamischer Modelle. Es bietet eine schön zu lesende und umfassende Grundlage für die Untersuchung der Parameterschätzung nichtlinearer Systeme und richtet sich damit an in diesem Gebiet arbeitende Forscher als Referenz ebenso wie an (Doktorats-)Studenten, die einen Einstieg in das Gebiet mit hohem mathematischen Niveau suchen. Der Text und erst recht die Beweise sind technisch anspruchsvoll, dank der sorgfältigen Aufbereitung durch die Autoren und der konsequent genauen Notation aber sehr verständlich. Die Qualität des Druckes und der Formelsetzung ist tadellos.

Dietmar Trummer  
Institut für Ökonometrie, OR und Systemtheorie  
Technische Universität Wien

T. AUGUSTIN: **Optimal Tests bei Intervallwahrscheinlichkeit**. Göttingen: Vandenhoeck & Ruprecht, 1998, xvi+290 S., ATS 619.20, ISBN 3-525-11411-37.

Seit längerer Zeit beschäftigt man sich mit allgemeineren Modellen zur Beschreibung stochastischer Unsicherheit, als es Wahrscheinlichkeitsräume im Kolmogoroff'schen Sinne sind. Neben unscharfen Wahrscheinlichkeiten, *Fuzzy Measures* und Kapazitäten sind Intervallwahrscheinlichkeiten von Interesse, bei denen Intervalle statt Zahlen zur Beschreibung der Unsicherheit herangezogen werden. Es erhebt sich die Frage, wieweit Resultate der klassischen statistischen Testtheorie auf Modelle mit Intervallwahrscheinlichkeiten übertragbar sind. Aufbauend auf Fragestellungen über die Existenz ungünstiger Paare für Tests, wie sie im Ergebnis von Huber und Strasser vorliegen, wird in der vorliegenden Untersuchung durch Modifikation der Begriffsbildung eine Existenzaussage für Intervallwahrscheinlichkeiten gewonnen. Außerdem werden Methoden zur Konstruktion optimaler Tests angegeben und die Beziehung zum Fundamentallemma hergestellt. Viele Aussagen setzen endliche Stichprobenräume voraus, zeigen aber trotzdem interessante Eigenschaften der verallgemeinerten Wahrscheinlichkeitsstrukturen bezüglich Testproblemen. Ein gutes Literaturverzeichnis, ein Autorenindex und ein umfangreicher Sachindex sowie der ansprechende Druck machen das Werk zu einem schönen Beitrag zur Stochastik. Ob die verallgemeinerten Wahrscheinlichkeiten und damit zusammenhängende statistische Analysen breite Anwendung finden, wird die Zukunft zeigen.

*Reinhard Viertl*  
*Institut für Statistik und Wahrscheinlichkeitstheorie*  
*Technische Universität Wien*