

Boekbesprekingen

Redactie

Arend Oosterhoorn

Postadres

Van Doorne's Transmissie bv

Postbus 500

5000 AM TILBURG

telefoon 013 - 640319 (dinsdag, woensdag en vrijdag)

076 - 417440 (maandag en donderdag)

Besproken boeken in Kwantitatieve Methoden nr. 43

R.M. Oliver en J.Q. Smith (editors)	187
<i>Influence diagrams, belief nets and decision analysis</i>	
Roland Caulett	188
<i>Statistics in Research and Development (2nd edition)</i>	
F. Schreuder	189
<i>SPSS/PC+ Tables</i>	
Ochi, Michel K.	190
<i>Applied Probability & Stochastic Processes in Engineering & Physical Sciences</i>	
Harold R. Lindman	191
<i>Analysis of Variance in Experimental Design</i>	
Todorovic, Petar	192
<i>An Introduction to Stochastic Processes and Their Applications</i>	
Y.L. Tong	194
<i>The multivariate normal distribution</i>	
Shayle Rarle, George Casella en Charles E. McCulloch	196
<i>Variance components</i>	
Raoul Lepage & Lynne Billard	198
<i>Exploring the Limits of Bootstrap</i>	
J.-D. Deuschel en D.W. Stroock	199
<i>Large deviations</i>	

Jos Pompen	205
<i>HBO Statistiek</i>	
Leonard Fortuin, Mynt Zijlstra, Hieke Besselaar en Karin de Groot	206
<i>De Kroonjuwelen, kwantitatieve modellen en methoden in verleden, heden en toekomst.</i>	
A.D. Lunn en D.R. McNeill	207
<i>Computer-Interactive Data Analysis</i>	
B.F.J. Manly	209
<i>Randomisation and Monte Carlo methods in Biology</i>	

R.M. Oliver en J.Q. Smith (editors)

Influence diagrams, belief nets and decision analysis

John Wiley & Sons, New York, 1990, xxix + 465 pag., ISBN 0-471-92381-8, £ 34.95.

Dit boek betreft de proceedings van een workshop aan de Universiteit van Californië, Berkeley, op 9-11 mei 1988. De bijdragen zijn van 20 geïnviteerde sprekers. Daarnaast is er commentaar van 7 discussianten. Twee derde van de auteurs van de proceedings komt uit Californië. Onder de meer bekende auteurs bevinden zich R.E. Barlow, A.P. Dempster, K. Doksum, D.M. Eddy, W. Edwards, R.A. Howard, J. Pearl, L.D. Philips, H. Rubin, J.Q. Smith, T.P. Speed, D.J. Spiegelhalter en L. Zadeh.

De bijdragen gaan over de relatie tussen grafische modellen voor afhankelijkheden tussen variabelen, en de daarmee corresponderende waarschijnlijkheidsmodellen en -berekeningen. Een bepaald type graaf, het in de beslissingsanalyse ontwikkelde "influence diagram", neemt bij de grafische representaties een belangrijke plaats in.

Het boek is geformuleerd in 5 delen:

Part I: Model Formulation and Analysis

Part II: Theoretical Foundations

Part III Problems and Applications: Industrial

Part IV: Problems and Applications: Medical

Part V: Efficiency and Computational Issues

De kwaliteit van de bijdragen is over het algemeen zeer goed. Daar staat tegenover dat het niet duidelijk is hoeveel van het gepresenteerde materiaal werkelijk origineel is, en hoeveel via gericht zoeken in de literatuur minstens even goed te vinden zou zijn.

Het boek is met name stimulerend doordat het uit verschillende perspectieven laat zien hoe grafische modellen, besliskunde, waarschijnlijkheidsrekening en (Bayesiaanse) statistiek samengebracht kunnen worden. Er worden zowel benaderingen besproken die het mogelijk maken om meer theoretische causale modellen te ontwikkelen, als ook benaderingen die op belangrijke praktische problemen van toepassing zijn. Onder de toepassingen die bestudeerd worden bevinden zich klassieke reliability problemen, modellen voor nuclear power accidents, evaluatie van medische technologieën, problemen in de genetica, en, misschien enigszins achterhaald, het opsporen en blijven volgen van militaire objecten in de ruimte in het kader van strategic defence systems. Interessante theoretische bijdragen betreffen optimaliseren van experimental design met het oog op daarna te nemen beslissingen, met name het gebruik van influence diagrams daarbij; influence diagrams en multiattribute utility analysis; en influence and belief adjustment. Al met al een

interessant en kwalitatief hoogstaand de proceedings. Het is ongetwijfeld te kostbaar voor de meeste individuele statistici. Het is echter wel zeer aan te raden voor afdelingsbibliotheken, met name als er ook besliskundig gericht onderzoek loopt, of als er interesse is in moderne methodologische ontwikkelingen.

D. Habbema, Erasmus Universiteit Rotterdam

Roland Caulcutt

Statistics in Research and Development (2nd edition)

Chapman & Hall, London, 1991, xi + 471 pag., ISBN 0-412-35890-5, £ 37.50

Dit is een boek dat sterk kan worden aangeraden aan iedere ontwikkelaar, ontwerper of onderzoeker, die nog niet zo veel statistische kennis heeft. Ook voor iemand die de eerste beginselen wel machtig is, maar een inleiding behoeft op het terrein van proefopzetten is dit boek aan te bevelen.

De auteur bouwt op uiterst heldere, en ik mag bijna zeggen: 'spannende', wijze het statistische gebouw op voorzover mensen in de industrie en in laboratoria dat nodig hebben. Hij heeft daarbij een rode draad in de vorm van een zekere chemische industrie; overigen komen al zijn voorbeelden uit de chemische industrie. Voor lezers met een andere achtergrond hoeft dat niet direct een bezwaar te zijn, maar zij zullen hun fantasie wel extra moeten aanspreken om in hun eigen omgeving toepassingen te zien.

In 16 hoofdstukken behandelt de schrijver de volgende onderwerpen:

- steekproeven en populaties: hoe ze te beschrijven;
- schatten en toetsen;
- regelkaarten;
- regressie;
- proefopzetten;
- variantie-analyse;
- Taguchi methoden.

De hoofdstukken worden afgesloten met een opgaven-sectie; de oplossingen staan achter in het boek. De invalshoek van de auteur is die van een toepasser: wanneer kan hij welke techniek toepassen, welke valkuilen staan hem te wachten, enz.

Het boek is bovendien zeer goed verzorgd. Ik heb nauwelijks drukfouten en helemaal geen denkfouten kunnen constateren. Wel dacht ik tijdens het lezen een paar keren dat Caulcutt niet helemaal correct was, maar dat bleek hij dan later te herstellen. Zo vond ik het bijvoorbeeld vreemd dat allerlei hoge managers van een fabriek zich met de uitvoering van een experiment

bemoeiden. Maar al ras bleek de reden: de kwaliteit van het experiment was bedroevend, zodat de auteur amper gelegenheid had om de grondbeginselen van het experimenteren te illustreren. Bovendien kon hij nu laten zien hoe een slecht experiment toch nog ten goede kan worden verbeterd.

Mijn ernstigste bezwaar tegen het boek is de zopas geïllustreerde voorliefde van de auteur om in eerste instantie met een niet-optimale oplossing te komen, om die dan verderop te verbeteren. Dat biedt hem echter wel didactische mogelijkheden, en omdat die ten volle benut worden staat bovengenoemd bezwaar een sterke aanbeveling geenszins in de weg.

Albert Trip

Centre for Quantitative Methods CQM

F. Schreuder

SPSS/PC+ Tables

Academic Service, Schoonhoven, 1992, 124 pag., ISBN 90-5261-086-X, f 35.00.

Het voorwoord van dit boekje doet het ergste vrezen: 'De moderne manager van vandaag maakt veelvuldig gebruik van de computer voor de analyse van zijn onderzoeksgegevens en de weergave van zijn onderzoeksresultaten. Door de efficiëntie van de computer zijn geavanceerde statistische methoden mogelijk en bereikbaar geworden, oogverblindende tabellen en grafieken zijn dikwijls het gevolg'. De doorsnee-lezer van KM zal zich vermoedelijk weinig aangesproken voelen door deze aanhef; daarnaast is de indruk dat het boek mede het produceren van *grafieken* behandelt (zoals ook door het omslag wordt gesuggereerd) apert onjuist. Het boek beperkt zich namelijk tot het maken van tabellen.

Het boekje beschrijft het maken van tabellen met behulp van de SPSS/PC+-procedure TABLES. Eerst wordt TABLES vergeleken met enkele andere SPSS/PC+-procedures waarmee tabellen en rechte tellingen kunnen worden gemaakt. Daarna wordt stap voor stap aan de hand van concrete voorbeelden duidelijk gemaakt hoe de layout van tabellen door de gebruiker kan worden gemanipuleerd. Tevens worden er per hoofdstuk enkele opdrachten gegeven; de uitwerking hiervan is achterin het boek te vinden.

Een diskette met daarop het databestand dat in het boek als voorbeeld wordt gebruikt is gratis bij de uitgever verkrijgbaar, na insturing van de in het boekje gevoegde antwoordkaart. Proefondervindelijk bleek dat de beloofde diskette inderdaad per ommegaande werd toegestuurd. Deze diskette is niet beveiligd, waardoor dit bestand gewoon naar harde schijf kan worden gekopieerd;

dat werkt wel zo aangenaam. Het bestand is een SPSS/PC+-systeembestand, met keurig alle labels en missende waarden gedefinieerd. Prima!

Het boekje bevat tevens een sectie over de wijze waarop de geproduceerde tabellen uitgevoerd en in een gewone tekst opgenomen kunnen worden. Daarbij wordt aangenomen dat de gebruiker werkt met WordPerfect 5.0 of hoger; procedures voor andere tekstverwerkingsprogramma's worden niet beschreven. Vermoedelijk is dit geen groot probleem, omdat het importeren van ASCII-bestanden in andere tekstverwerkers niet moeilijk is. Dat de tabel in een niet-proportioneel lettertype moet worden afgedrukt zal door sommigen als een probleem worden ervaren (lelijk!), maar is inherent aan het gebruik van TABLES en heeft met het boekje als zodanig niets te maken.

Kan dit boekje worden aangeraden aan mensen die overwegen SPSS/PC+ TABLES te gaan gebruiken (studenten, onderzoekers, moderne managers van vandaag, en dergelijke)? De tekst is over het algemeen duidelijk en prettig leesbaar. Wel moet me van het hart dat ik grote moeite had de voorbeelden exact te reproduceren. In de meeste gevallen waren er typefoutjes in het spel, in andere gevallen klopte het voorbeeld zelf niet helemaal. In het voorwoord wordt zulks ook gemeld ('de tabellen ... hoeven echter niet precies overeen te komen met de tabellen in de SPSS-uitvoer'), maar zo'n disclaimer hoort in een dergelijk boek natuurlijk niet thuis. Verder ontbreekt een kort overzicht van de syntaxis en mogelijke commando's bij TABLES. Ik vind dit zeker een gemis aangezien het boekje daardoor praktisch gesproken niet als 'quick reference' gebruikt kan worden, en zulks m.i. de gebruikswaarde toch wel doet afnemen; bij andere uitgaven in deze reeks (het 'Leerboek gegevensverwerking met SPSS/PC+ en Data entry' van Van Essen, 1992) is wel gezorgd voor een dergelijk overzicht. Wie wil afzien van deze en andere - kleinere - schoonheidsfoutjes doet met dit boekje vermoedelijk een goede keus.

Toon Taris & Inge Bok,
Vrije Universiteit Amsterdam,
Vakgroep Arbeids- en Organisationspsychologie.

Ochi, Michel K.

Applied Probability & Stochastic Processes in Engineering & Physical Sciences

John Wiley & Sons, New York, 1990, xvi + 499 pag., ISBN 0-471-85742-4, £ 54.65.

Het boek is bedoeld als leerboek voor ouderejaars studenten in natuurwetenschappen en ingenieursstudies. Het bevat de kansrekening die relevant is voor stochastische processen en verschillende onderwerpen binnen stochastische processen. De nadruk ligt hierbij op spectraalanalyse. Ook geavanceerde onderwerpen komen aan bod.

Het deel kansrekening begint met elementaire onderwerpen als een inleiding in kansen, random variabelen en hun kansverdelingen, momenten, moment genererende functies en karakteristieke functies. Daarna worden discrete en continue stochasten met hun verdelingen gegeven. Ook komen transformaties van stochasten en extreme waarden statistieken aan de orde.

Het tweede deel behandelt stochastische processen. Het begint met een inleiding, waarna spectraalanalyse aan de orde komt. Na een behandeling van analyses van tijdreeksen komen Wiener-Lévy en Markov processen aan de orde. Zowel lineaire systemen als niet-lineaire systemen worden behandeld, evenals niet-normale processen en discrete processen.

Vier appendices behandelen enkele nodige wiskundige technieken.

Enige kennis van geavanceerde calculus en functionaalanalyse is nodig. Ook wordt enige kennis van elementaire statistiek verondersteld. Maattheorie wordt niet gebruikt.

Door het hele boek worden veel voorbeelden gegeven. Ook bevat het boek na elk hoofdstuk enkele opgaven. Uitwerkingen of uitkomsten worden niet gegeven. Het boek is zowel geschikt als naslagwerk als als studieboek.

Elly Jeurissen

Hoogovens Technical Services Management & Operations bv

Harold R. Lindman

Analysis of Variance in Experimental Design

Springer Verlag, New York, 1991, ix + 531 pag., ISBN 0-387-97571-3, DM 98.00.

Dit boek is een inleidende tekst in de variantie-analyse dat ook voor onderzoekers zonder een wiskundige achtergrond maar met voldoende kennis in statistiek, kan worden gebruikt. Het boek begint op een relatief laag niveau dat echter van hoofdstuk tot hoofdstuk opklimt in moeilijkheidsgraad. De eerste acht hoofdstukken moeten achter elkaar worden gelezen omdat deze hoofdstukken achtereenvolgens opgebouwd zijn.

In het eerste hoofdstuk worden basisconcepten van de statistiek herhaald zoals scheefheid en kurtosis en basis kansverdelingen (ook niet-centrale). Het tweede hoofdstuk geeft een inleiding in de eenvoudige variantie-analyse met vaste effecten. In de hoofdstukken 3 en 4 zijn de meervoudige vergelijkingen behandeld. Hoofdstuk 5 geeft een beschrijving van de tweevoudige variantie-analyse met vaste effecten.

Toevallige effecten worden in hoofdstuk 6 ingevoerd voor het eenvoudige en tweevoudige geval ('mixed' en 'random models'). Modellen voor de variantie-analyse met 3 en meer factoren worden in hoofdstuk 7 opgevoerd. In de hoofdstukken 5 t/m 7 worden proefschemata's met twee indelingen

bekeken en in hoofdstuk 8 wordt de hiërarchische indeling behandeld. Daarna volgen in hoofdstuk 9 onvolledige proefschemata's (BIB, Youden Squares, Latin Squares), in de hoofdstukken 10 en 11 de trend analyse in eenvoudige en meervoudige proefschemata's, in hoofdstuk 13 de meerdimensionale variantie-analyse, in hoofdstuk 14 de covariantie-analyse en in hoofdstuk 15 een inleiding in het lineaire model. Hoofdstuk 12 (matrix algebra) zou beter in de appendix geplaatst kunnen worden waar in Appendix A stellingen en bewijzen, in appendix B tabellen, en in appendix C en D variantie-analyse met behulp van SAS en SPSS te vinden zijn.

Afgezien van de discussie van de verschillende methoden van meervoudige vergelijkingen en de afwijzing van de Newman Keuls toets door de auteur kan deze inleidende tekst als een goede mogelijkheid worden gezien om een snelle toegang tot de methoden van de variantie-analyse te krijgen.

D. Rasch

Landbouw Universiteit Wageningen

Todorovic, Petar

An Introduction to Stochastic Processes and Their Applications

Springer Verlag, Berlijn, 1992, xiv + 289 pag., ISBN 3-540-97783-X. DM 98.00.

Dit is een heel goed studieboek. Het is bedoeld voor gevorderde studenten (wiskunde, natuur- en technische wetenschappen, econometrie) die een cursus "grondslagen waarschijnlijkheidsrekening" gebaseerd op enige inleidende maat- en integratietheorie achter de rug hebben. Het boek vormt de neerslag van jarenlange doceerervaring van de auteur aan de universiteit Santa Barbara (Californië). Aan de orde komen de volgende onderwerpen: stochastische processen in de continue tijd (basisbegrippen); puntprocessen (Poissonproces en verwante processen); Brownse Beweging en Gaussische processen; tweede orde processen en stationaire processen (spectraalanalyse), met een inleidend hoofdstuk over L2-theorie; Markov-processen in de continue tijd en met algemene toestandruimte), inclusief een hoofdstuk gewijd aan de benadering van dit onderwerp vanuit de theorie der contractie-halfgroepen (Hille-Yoshida); tenslotte een hoofdstuk over martingaaltheorie in discrete tijd.

Het boek is zeer helder geschreven met veel aandacht voor toepassingen. Elk hoofdstuk bevat een ruime collectie zorgvuldig geselecteerd illustratief oefenmateriaal. Een naar hoofdstuk ingedeelde adequate bibliografie bevindt zich achterin het boek. De tekst maakt een verzorgde indruk; drukfouten zijn er niet veel. Van een misschien wat storende onzorgvuldigheid wil ik hier melding maken: de slotconclusie in Example 10.6.1 (p276) is natuurlijk niet juist.

Thans een paar wat meer specifieke opmerkingen. Het boek is in principe "self-contained" met het hoofdstuk L2-ruimten als belangrijke uitvalsbasis. Zo wordt ten behoeve van de martingaaltheorie de voorwaardelijke verwachting ingevoerd via de "projectstelling" in L2 (geen beroep derhalve op de stelling van Radon-Nikodym). Wel worden door de auteur de algemene resultaten uit de maat- en integratietheorie en de grondslagen waarschijnlijkheidsrekening bij de lezer paraat ondersteld en wordt bij verwijzing naar bijvoorbeeld de stelling van Dini (over uniforme convergentie), de tweede stelling van Helly, de stelling van Cramer (over de factorisatie van de normale verdeling) geen nadere omschrijving nodig geacht. Mogelijk toch zou in dit opzicht de opname van een korte appendix het boek nog iets completer maken.

Een zijdelingse opmerking: de auteur kwalificeert terecht de convergentiestelling voor submartingale als een van de meest fundamentele resultaten binnen de martingaaltheorie (binnen de waarschijnlijkheidsrekening mag rustig gezegd worden). Zijn toelichtende opmerking echter: "... and can be considered an analog to the well-known fact in real analysis that every bounded monotone sequence has finite limit" ervaart uw recensent als een misleidende simplificatie.

Het boek kan warm worden aanbevolen aan diegenen die zich een beknopt maar gedegen beeld willen vormen omtrent de basisnotities van de belangrijkste typen continuïetijd stochastische processen.

De door Springer verzorgde uitvoering van het boek is uitstekend.

PJ Holewijn

Y.L. Tong

The multivariate normal distribution

Springer-Verlag, Berlijn, 1990, xiii + 271 pag., ISBN 3-540-97062-2, DM 98.00

"The multivariate normal distribution is undoubtedly one of the most well-known and useful distributions in statistics, playing a dominant role in many areas of applications" (hoofdstuk 1). Dit boek gaat in op verdelingsaspecten van die multivariate normale verdeling. Aspecten betreffende statistische gevolgtrekkingen en multivariate data-analyse worden niet behandeld. De bedoeling van de schrijver is niet een uitputtende verzameling van resultaten te presenteren, maar een bondige en systematische beschrijving te geven van een aantal onderwerpen. Ruime aandacht wordt geschonken aan de theorie van multivariate waarschijnlijkheidsongelijkheden. De doelstelling wordt gehaald.

Echter, dat betekent dat de auteur de titel van zijn boek te algemeen heeft gekozen: "Distributional aspects of the multivariate normal distribution" had ik een betere titel gevonden.

De onderwerpen die in het boek aan de orde komen, zijn gegroepeerd in de volgende hoofdstukken:

1. *Introduction*

2. *The Bivariate Normal Distribution*

Dit is een kort hoofdstuk met bekende resultaten betreffende marginale en conditionele verdelingen, transformaties, steekproefverdelingen en waarschijnlijkheidsongelijkheden.

3. *Fundamental Properties and Sampling Distributions of the Multivariate Normal Distribution*

Dit hoofdstuk is een uitbreiding van het vorige naar het n -dimensionale geval, met o.a. een aantal equivalente definities van de multivariatie normale verdeling. Er wordt kort ingegaan op regressie, partiële en canonieke correlatie en principale componenten. Het hoofdstuk sluit af met de Wishart verdeling en de T^2 verdeling van Hotelling.

4. *Other Related Properties*

Een aantal klassen van kansdichtheden wordt bestudeerd: dichtheden met elliptische contouren, log-concave dichtheden, dichtheden die "multivariate-totally-positive-of-order-2" (MTP₂) zijn, Schur-concave dichtheden en dichtheden die "arrangement permutation-symmetric" zijn. Door de kansdichtheid van de multivariatie normale verdeling te zien als een element van die klassen kunnen fraaie resultaten worden bereikt. De volgende hoofdstukken gaan daar op in.

5. *Positively Dependent and Exchangeable Normal Variables*

Eerst wordt positieve afhankelijkheid bestudeerd, ondermeer m.b.v. de MTP₂ eigenschap. Vervolgens wordt ingegaan op de klasse van permutatie-symmetrische normaal verdeelde stochasten, waarvan een belangrijke subklasse de klasse van uitwisselbare normaal verdeelde stochasten is. Het aardige van deze subklasse is dat de verdelingsfunctie kan worden uitgedrukt in een enkelvoudige integraal (i.p.v. een meervoudige).

6. *Order Statistics of Normal Variables*

De belangrijkste gedeelten van dit hoofdstuk zijn de kansdichtheden van de order statistics van (uitwisselbare) normaal verdeelde stochasten en de verdeling van lineaire combinaties van die order statistics. Gebruik wordt gemaakt van o.a. de MTP₂ eigenschap.

7. *Related Inequalities*

Diverse majorizatie ongelijkheden worden bestudeerd, met name voor uitwisselbare normaal verdeelde stochasten en voor de kansmassa op een rechthoekig, elliptisch of anders-convex gebied als de dichtheid Schur-concaaf is. Speciale aandacht wordt besteed aan n -dimensionale rechthoekige gebieden met vaste omtrek en meetkundige gebieden met vast volume.

8. *Statistical Computing Related to the Multivariate Normal Distribution*

Dit hoofdstuk gaat in op het genereren van multivariaat normaal verdeelde stochasten en op het uitrekenen van bepaalde integralen van de multivariatie normale verdeling als de covariantiematrix een speciale structuur heeft. Diverse tabellen zijn toegevoegd.

9. *The Multivariate t Distribution*

Het laatste hoofdstuk bespreekt de multivariate t verdeling: verdelingstheorie, momenten en waarschijnlijkheidsongelijkheden worden besproken. De relatie met de multivariatie normale verdeling wordt aangegeven.

Algemene resultaten worden beschreven als "propositions" en resultaten die alleen op de multivariatie normale verdeling betrekking hebben als "theorems". ("Proposition 3.5.1" had m.i. een "theorem" moeten zijn.) De doelgroep wordt gevormd door "graduate students and advanced undergraduate students in statistics, mathematics, and related applied areas". Onderzoekers die (veel) te maken hebben met de multivariatie normale verdeling, zullen dit boek als een waardevolle referentie beschouwen. Het boek is helder geschreven, geeft op het juiste moment een voorbeeld en vormt een bron van andere literatuur. Ieder hoofdstuk wordt gevolgd door een aantal opgaven. Deze postieve punten wegen ruimschoots op tegen het kritiekpuntje dat het boek enkele typfouten bevat.

Siebrand J. Wierda

Vakgroep Econometrie, Rijksuniversiteit Groningen

Shayle R. Searle, George Casella en Charles E. McCulloch

Variance components

John Wiley, New York, 1992, xxiii + 501 pag., ISBN 0-471-62162-5, \$ 56.00.

Modellen met variantiecomponenten kennen een lange geschiedenis en gaan minstens terug tot een boek van de astronoom Airy uit 1861 waarin een 1-weg random effecten model voor telescoop-waarnemingen wordt geformuleerd. Modellen met variantiecomponenten of mixed modellen, zoals ik ze gemakshalve zal noemen, worden in veel toepassingsgebieden van de statistiek gebruikt. Toch is dit pas het tweede boek waarin diepgaand en uitsluitend op mixed modellen wordt ingegaan. Het eerste boek waarin dit het geval is, is 'Estimation of variance components and applications' van Rao en Kleffe (1988, North-Holland Amsterdam). Rao en Kleffe beperken zich vrijwel uitsluitend tot één schattingsmethode, namelijk MINQUE en geven een zeer theoretische beschouwing. Het lezerspubliek is ongetwijfeld klein. Searle, Casella en McCulloch nemen nadrukkelijk afstand van de MINQUE methode (met als bezwaar de noodzaak tot het gebruik van a priori waarden) en

concentreren zich veel meer op maximum likelihood (ML) en restricted maximum likelihood (REML) onder normaliteit. Dit boek is toegankelijk voor iedereen die enige kennis van matrix algebra heeft en zo nu en dan aan armlange matrix expressies is blootgesteld. Veel van de benodigde matrix rekening (zoals Kroneckerproduct, Moore-Penrose inverse en differentiatie van matrices en determinanten) is in een appendix weergegeven.

Van het trio auteurs is Searle ongetwijfeld de bekendste, vooral door z'n boek 'Linear models' (1971, John Wiley), maar bij doorwinterde gebruikers van mixed modellen ook door z'n rapport 'Notes on variance components estimation. A detailed account of maximum likelihood and kindred methodology' (1979, rapport Cornell University). Veel van dit rapport is in dit boek terug te vinden.

De eerste drie hoofdstukken zijn inleidend: wat over de begrippen (wanneer zijn effecten random en wanneer fixed), over de geschiedenis en over wat er nog komen gaat, toegelicht aan de hand van het 1-weg random effecten model. De volgende twee hoofdstukken gaan over de ANOVA methode voor gebalanceerde en ongebalanceerde data. ANOVA voor ongebalanceerde data betekent Hendersons methoden I, II en III. Dit is voornamelijk van historische interesse. Vanaf hoofdstuk 6 wordt het vooral interessant: ML en REML (Hfd. 6), predictie van randomeffecten (Hfd. 7), numerieke methoden (Hfd. 8), Bayesiaanse methoden (Hfd.9), 0-1 waarnemingen (Hfd. 10), MINQUE, MIVQUE en dergelijke (Hfd. 11) en tenslotte het dispersion-mean model (Hfd. 12).

Gezien het belang van slimme numerieke methoden, zoals die vooral voor de analyse van melkproductie gegevens zijn ontwikkeld in het veeteeltkundig onderzoek, voor het berekenen van ML en REML schattingen is hoofdstuk 8 'Computing ML and REML estimates' wat aan de magere kant. In § 3.4 wordt aan de hand van IQ scores van college freshman Ronny Fysher de predictie van random effecten geïntroduceerd. In hoofdstuk 7 wordt hier verder op in gegaan. Dit hoofdstuk en hoofdstuk 9 over Bayesiaanse methoden maakt veel duidelijk over de verschillen tussen fixed en random effecten. Hoofdstuk 9 maakt overigens een nogal rommelige indruk: er wordt voortdurend van de ene op de andere a priori verdeling over gesprongen. Mogelijk dat dit een afspiegeling is van de moeilijkheden die men ondervindt met de empirical Bayes methode voor niet triviale a priori verdelingen. Van uitdrukking (16) op pagina 320 in dit hoofdstuk kan ik noch op z'n Bayesiaans, noch op z'n frequentistisch iets anders dan onzin maken. Een zwak moment van de auteur (of van mij)? Dit hoofdstuk had in ieder geval beter gekunt. Hoofdstuk 10 daarentegen geeft een beknopte maar uitstekende indruk van de stand van zaken met betrekking tot mixed modellen voor 0-1 gegevens rond begin 1991.

Hoofdstuk 11 geeft een goede indruk van de relaties tussen verschillende methoden zoals MINQUE, MIVQUE en REML. Hoofdstuk 12 gooit daar nog een schepje bovenop met de presentatie van het dispersion-mean model, waarin verschillende schattingsmethoden als kleinste kwadraten technieken zijn af te leiden.

Er wordt nauwelijks aandacht besteed aan co-variantie componenten. Ook uitbreidingen in de richting van tijdreeksachtige structuren of ruimtelijke interpolatie komen niet aan bod. Hetzelfde

geldt voor specifieke modellen als het animal model in de veefokkerij. De schrijvers beperken zich tot het traditionele mixed model met onafhankelijke (normaal) verdeelde random effecten.

Het is jammer dat uitgewerkte analyses van praktijkdata, a la 'Generalized linear models' van McCullagh en Nelder, ontbreken. Dit zou de lezer zeker meer inspireren tussen de vaak deprimerend lange aaneenschakelingen van matrices en veel van de praktijkproblemen die zich bij mixed models voordoen kunnen illustreren.

Het boek kan niettemin basis zijn voor een college, maar dan aangevuld met enig praktisch werk, bijvoorbeeld met behulp van de pakketten Genstat 5 of SAS.

Samengevat: een goed boek dat zeker in een behoefte voorziet en waar veel naar zal worden verwezen.

Bas Engel

Groep Landbouwwiskunde (GLW-DLO), Wageningen.

Raoul Lepage & Lynne Billard

Exploring the Limits of Bootstrap

John Wiley & Sons, New York, 1992, xi + 462 pag., ISBN 0-471-53631-8, £ 47.50.

This book contains papers which were read at the special topics meeting of the Institute of Mathematical Statistics held in East Lansing in May 1990. The papers all address the theme "Exploring the Limits of Bootstrap". Since the book contains such a variety of papers the best way to give an impression of its contents is just to list the papers.

Part 1. Introduction

- Introduction to the Bootstrap, B. Efron and R. Lepage

Part 2. General principles of the Bootstrap

- On the Bootstrap of M-estimators and other statistical functionals, M.A. Arcones and E. Giné.
- Bootstrapping Markov chains, K.B. Athreya and C.D. Fuh.
- Theoretical comparison of different bootstrap t confidence bounds, P.J. Bickel.
- Bootstrap for a finite state Markov chain based on i.i.d. resampling, S. Datta and W.P.Mc.Cormick.
- Six questions raised by Bootstrap, B. Efron.

- Efficient Bootstrap simulation, P.Hall
- Bootstrapping U -quantiles, R. helmers, P. Janssen and N. Veraverbeke.
- An invariance principle applicable to the Bootstrap, J.G. Kinatered.
- Edgeworth correction by "moving block" Bootstrap for stationary and nonstationary data, S.N. Lahiri.
- Bootstrapping signs, R. Le Page.
- Moving blocks jackknife and Bootstrap capture weak dependence, R.Y. Liu and K. Singh.
- Bootstrap bandwidth selection, J.S. Maron
- A circular block -resampling procedure for stationary data, D.N. Politis and J.P. Romano.
- Some applications of the Bootstrap in complex problems, R. Tibshirani.
- Approximating the distribution of a general standardized functional statistic with that of jackknife pseudo values, D.S. Tsu.

Part 3. Applications of the Bootstrap.

- Bootstrapping for order statistics sans random numbers (operational bootstrapping), W.A. Bailey.
- A generalized Bootstrap, E.J. Bedrick and J.R. Hill.
- A hazard process of survival analysis, J.J. Hsieh.
- Bootstrap assesment of prediction in exploratory regression analysis, V. Kipnis.
- Bootstrapping likelihood ratios in quantitative genetics, N. Schork.
- A nonparametric density estimation based resampling algorithm, M.S. Taylor and J.R. Thomson.
- Nonparametric rank estimation using bootstrap resampling and canonical correlation analysis, X.M. Tu, D.S. Burdick and B.C. Mitchell.

Although with such a rapidly developing subject as the Bootstrap one cannot expect that the limits in 1990 remain limits in 1993, the book contains many interesting papers by most of the well known researchers in the field. It also contains a couple of nice introductions.

A.J. van Es

Universiteit van Amsterdam

J.-D. Deuschel and D.W. Stroock

Large deviations

Academic Press, Boston, 1989, xiv + 307 pag., ISBN 0-12-213150-9, \$ 34.95.

This research monograph is the most complete and up to date book on the theory of large deviations. It covers all the key theoretical concepts as well as a number of important applications. It has many exercises, some historical notes and an extensive list of references.

Large deviation theory originated from works by Khinchin, Smirnov and Cramér (1920's and 30's) and came to full bloom in a series of papers by Donsker and Varadhan (1970's). Since then it has become a central part of statistics and probability, and has expanded into several other areas.

Before reviewing the content of the book I shall give a brief introduction to large deviations, aimed at the non-expert.

1 Brief introduction

The key concept in the theory of large deviation is the so-called *large deviation principle* (LDP). Given are a Polish space X , a σ -algebra $\mathcal{F}$ of Borel subsets of X , and a sequence of probability measures (P_n) on $\mathcal{F}$. The (P_n) is said to satisfy the LDP with rate function $I: X \rightarrow [0, \infty]$ if

$$\text{I.} \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n(C) \leq -\inf_{x \in C} I(x) \text{ for } C \in \mathcal{F} \text{ closed}$$

$$\text{II.} \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n(O) \geq -\inf_{x \in O} I(x) \text{ for } O \in \mathcal{F} \text{ open}$$

III. $I \neq \infty$, I is lower semicontinuous and has compact level sets.

Properties I and II are a kind of weak convergence but on an exponential scale. They imply that for all $S \in \mathcal{F}$ such that $\inf_{x \in \text{cl}(S)} I(x) = \inf_{x \in \text{int}(S)} I(x)$ (closure resp. interior) the limit law $\lim_{n \rightarrow \infty} \frac{1}{n} \log P_n(S) = -\inf_{x \in S} I(x)$

holds. Property III is important for applications (see (*) below). The function I corresponding to (P_n) is unique and satisfies $\min_{x \in X} I(x) = 0$.

The LDP gives a precise description of $P_n(S)$ as $n \rightarrow \infty$ for a large class of events S . For all S containing a minimizer of I it follows that $\lim_{n \rightarrow \infty} \frac{1}{n} \log P_n(S) = 0$. If the minimizer is unique, then this

typically corresponds to a law of large numbers or an ergodic theorem.

The main application of the LDP is to the asymptotics of exponential integrals. Namely, given $F: X \rightarrow \mathbb{R}$ sufficiently nice (e.g. bounded and continuous), I-III imply that:

$$(*) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \int_X e^{nF(x)} P_n(dx) = \sup_{x \in X} [F(x) - I(x)].$$

Integrals of this type appear in many concrete situations. To mention just a few areas: interacting particle systems, Gibbs measures, transport processes, biological models. One should think of (*) as a far reaching generalization of Laplace's method of steepest descent.

Basic example: Let $Y_1, Y_2, \dots$ be i.i.d. $\mathbb{R}$ -valued random variables with $E(e^{tY_1}) < \infty$ for t in a neighbourhood of $t = 0$. The LDP holds in the following three situations:

- (i) $X = \mathbb{R}, P_n(\cdot) = P(S_n \in \cdot)$ with $S_n = \frac{1}{n} \sum_{i=1}^n Y_i$ the empirical mean
- (ii) $X = \mathcal{P}(\mathbb{R}), P_n(\cdot) = P(L_n \in \cdot)$ with $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$ the empirical distribution.
- (iii) $X = \mathcal{P}_s(\mathbb{R}^{\mathbb{Z}}), P_n(\cdot) = P(Q_n \in \cdot)$ with $Q_n = \frac{1}{n} \sum_{i=1}^n \delta_{T^i Y}$ the empirical process

(T = shift, $Y = (Y_1, Y_2, \dots, Y_n)$ periodically repeated).

($\mathcal{P}(\mathbb{R})$ denotes the set of probability measures on $\mathbb{R}$, $\mathcal{P}_s(\mathbb{R}^{\mathbb{Z}})$ the set of stationary processes, both endowed with the weak topology). The corresponding rate functions can be explicitly expressed in terms of the distribution of Y_1 (see below). (In the literature (i) - (iii) are often called large deviations on level 1, 2 resp. 3, but for no apparent reason the book systematically avoids this terminology.)

Variations on the above example replace the i.i.d. sequence (Y_i) by a Markov chain, or by a stationary process that is sufficiently mixing. Much of the theory centers around trying to prove the LDP for the three empirical functional in (i)-(iii) for as large a class of processes (Y_i) as possible and trying to identify the corresponding rate function. Once this is achieved one can proceed to apply (*) for a concrete choice of F , or to try and extend (*) by replacing F by a sequence (F_n) with nice convergence properties.

These are also variations on the LDP with discrete time n replaced by continuous time t in I and II, or when weaker restrictions on the rate functions in III.

2 Outline of the book

CHAPTER I: The book starts in section 1.1 with a "soft" introduction to the LDP, explaining in what sense it describes deviations from ergodic behaviour (see part 1 above). In section 12 the classical Cramer theorem is formulated and proved, i.e., Example (i) above. Here the large

deviations are around the law of large numbers: $S_n \rightarrow EY_1$ a.s.. The rate function for this case is shown to be given by the *Legendre transform* of $\Lambda(t) = \log E(e^{tY_1})$, namely

$$I(\alpha) = \sup_{t \in \mathbb{R}} [t\alpha - \Lambda(t)].$$

Section 1.3 discusses Schilder's theorem for Brownian motion. Here

$$X = \{\theta \in C([0, \infty); \mathbb{R}^d) : \theta(0) = 0, \lim_{t \rightarrow \infty} \theta(t)/t = 0\}$$

is path space with norm $\|\theta\| = \sup_{t \geq 0} |\theta(t)|/(t+1)$, W is the Wiener measure on X , and W_ϵ is the distribution of the scaled path $\in^{1/2}\theta$ under W . Clearly $W_\epsilon \rightarrow \delta_{\{\theta=0\}}$ as $\epsilon \rightarrow 0$.

Schilder's theorem states that (W_ϵ) satisfies the LDP (with $1/n$ replaced by ϵ) with rate function

$$\begin{aligned} I(\psi) &= \frac{1}{2} \int_{\mathbb{R}^d} |\psi(\alpha)|^2 d\alpha \text{ if } \psi \in L^2([0, \infty); \mathbb{R}^d) \text{ exists} \\ &= \infty \text{ otherwise} \end{aligned}$$

Chapter I closes in section 1.4 with a discussion of two applications of Schilder's theorem: (1) Strassen's functional law of the iterated logarithm, (2) Ventcel- Freidlin large deviations for a randomly perturbed dynamical system.

Chapter II: This chapter is where a start is made with building up a systematic theory. Section 2.1 describes some generalities: elementary properties of rate functions, weak vs. strong LDP, Varadhan's theorem (*), and covariance under continuous maps. (In the literature the latter is often called the contraction principle, but again the book avoids this terminology.) Section 2.2 contains some elementary convex analysis. This leads to a formula identifying I for a general (P_n) as the

Legendre transform of Λ given by $\Lambda(t) = \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_{P_n}(nt)$, with Λ_{P_n} the logarithmic moment generating function of P_n (compare with the result mentioned above), assuming the limit exists. This formula, however, only holds if I is convex, and it also required some additional restrictions on the space X .

CHAPTER III: This chapter describes the general Cramer theory for i.i.d. random variables, i.e., Examples (i) and (ii) above for general spaces X . Section 3.1 contains some preliminary remarks. Section 3.2 formulates and proves Sanov's theorem for Example (ii). Here $X = P(\Sigma)$ with Σ a Polish space and X is endowed with the weak topology. If μ is the law of δ_{Y_1} , then the deviations are around $L_n \Rightarrow \mu$. The rate function for this case is shown to be given by the relative entropy

$$\begin{aligned} I(\nu) &= \int_{\Sigma} \log\left(\frac{d\nu}{d\mu}\right) d\nu \text{ if } \nu \ll \mu \\ &= \infty \text{ otherwise} \end{aligned}$$

Section 3.3 extends Cramér's theorem for Example (i) to Banach spaces, and the result is used in section 3.4 to generalize Schilder's theorem to Gaussian measures.

CHAPTER IV: This chapter deals with the Markov case. In section 4.1 the sequence (Y_i) is no longer i.i.d. but is a (stationary) Markov chain with a (recurrent) Markov kernel $\Lambda: X \rightarrow [0,1]$. A Sanov type theorem is proved for the empirical distribution (recall Example (ii) under the stringent assumption that Λ is uniformly ergodic. Roughly this means that the probabilities are uniform in the initial point, which implies exponentially fast convergence to equilibrium. (E.g. if X is finite then Λ is in the class.) As a consequence of this assumption the LDP can even be proved in the uniform sense where I and II are replaced by

$$I'. \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log(\sup_{x \in X} P_n(C | Y_0 = x)) \leq -\inf_{v \in C} I(v) \text{ for } C \in \mathcal{F} \text{ closed}$$

$$II''. \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log(\inf_{x \in X} P_n(O | Y_0 = x)) \geq -\inf_{v \in O} I(v) \text{ for } O \in \mathcal{F} \text{ open}$$

The rate function is identified as

$$I(v) = \sup_{u \in C_b(\Sigma; \mathbb{R})} \left[-\int_{\Sigma} \log\left(\frac{\Pi u}{u}\right) dv \right]$$

In section 4.2 the same program is carried out for a continuous time Markov process (Y_t) with the empirical distribution $L_t = \frac{1}{t} \int_{[0,t]} \delta_{Y_s} ds$. Under a similar condition of uniform ergodicity a Sanov type theorem is derived, this time with the rate function

$$I(v) = \sup_{u \in C_b(\Sigma; [1, \infty))} \left[-\int_{\Sigma} \frac{Lu}{u} dv \right],$$

where L is the generator of the process. The absence of the logarithm in this expression as compared to the previous one has an advantage when the Markov process is reversible. Namely, in that case the rate function reduces to

$$I(v) = \mathcal{E}(f^{\frac{1}{2}} f^{\frac{1}{2}}) \text{ with } f = \frac{dv}{dm} \text{ if } v \ll m, \\ = \infty \text{ otherwise}$$

where $\mathcal{E}$ is the Dirichlet form associated with L and m is the reversible measure. Section 4.3 discusses the application of section 4.2 to the celebrated Wiener Sausage, i.e., the process (σ_t^ϵ) with $\sigma_t^\epsilon =$ Lebesgue volume of the ϵ -neighbourhood of the Wiener process in $\mathbb{R}^d$ in the time interval $[0,t]$. The main result is that for any $\gamma, \epsilon > 0$ and $d \geq 1$

$$\lim_{t \rightarrow \infty} \frac{1}{t^{d/(d+2)}} \log E(e^{-\gamma \sigma_t^{(e)}}) = -\kappa(d, \gamma)$$

where $\kappa(d, \gamma)$ is a positive constant that comes out of variational formula involving the principal Dirichlet eigenvalue of $-\frac{1}{2}\Delta$ on the unit ball. The above equation is a result of the type (*) but with a scaled time parameter. In order for the results in section 4.2 to be applicable to the Wiener Sausage the LDP in Schilder's theorem has to be extended from the weak topology to the uniform (= variation norm) topology after an appropriate "mollification" procedure. This part of the proof is fairly technical. Section 4.4. extends the LDP of sections 4.1 and 4.2 to the empirical process (compare Example (iii)), again under the assumption of uniform ergodicity. This analysis requires a digression in ergodic theory. The resulting rate function is expressed in terms of the kernel Λ and the Kolmogorov Sinai entropy. The proofs are again somewhat technical, particularly in the continuous time case.

CHAPTER V: This chapter tries to weaken the assumption of uniform ergodicity used in Chapter IV. Section 5.1 contains some preliminaries and section 5.2 a further digression in ergodic theory. In section 5.3 it is shown that the LDP in I-III holds for reversible continuous time Markov processes under a number of restrictions that are equivalent to some weak dependence. In section 5.4 a very general LDP is proved for the empirical process under a condition that is called hypermixing. Again the Kolmogorov-Sinai entropy appears in the rate function. In section 5.5 it is shown that stationary processes with finite memory and a certain contraction property are hypermixing.

CHAPTER VI: The book closes with some analytical considerations. Section 6.1 gives conditions under which a Markov process is hypermixing. Again the Dirichlet form appears, as well as the important tool of logarithmic Sobolev inequalities. The results are applied in section 5.2 to symmetric diffusions on manifolds. Section 6.3 applies the results of section 4.2 to "hypoelliptic" diffusions on a compact manifold. The LDP is found to hold for the empirical distribution in the form with I' and II', and a fairly explicit expression for the rate function is derived.

3 Conclusion

To be able to fully appreciate large deviation theory one needs, besides a solid knowledge of probability theory, a certain skill in topology, functional analysis and ergodic theory. This is why some mathematicians experience the field as somewhat difficult and, unfortunately, most physicists even as totally inaccessible. On the other hand, the underlying principles (as listed in part 1) are really easy to grasp and they form the compass on which one can safely sail. The authors do a very

good job in bringing these principles to the surface, but the reader must not be put off by the many technicalities that come with it. The book is a research monograph, not a textbook. The researcher will find it an excellent source to learn from, given enough motivation. The book is too difficult for students, and even for the casually browsing mathematician. My advice to the newcomer in the field is: first Varadhan, and then go for Chapters I-IV of Deuschel and Stroock. Many more applications can be found in Ellis and Bucklew.

References:

- (1) J.A. Bucklew (1990). *Large deviation techniques in Decision, Simulation, and Estimations*. Wiley, New York.
- (2) R.S. Ellis (1985). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York.
- (3) D. Stroock (1984). *An Introduction to the theory of large Deviations*. Springer, new York.
- (4) S.R.S. Varadhan (1984). *Large deviations and Applications*. SIAM, Philadelphia.

M. Hollander

Technische Universiteit Delft

Jos Pompen

HBO Statistiek

Academic Service, Schoonhoven, 1992, xii + 277 pag., ISBN 90-5261-034-7, f 45.00.

Dit boek heeft kan als leerboek niet worden aanbevolen. Het is zeer beknopt, wat misschien voor een enkeling nog een voordeel is. Aangezien HBO-studenten de doelgroep vormen, zullen de meeste lezers toch te weinig toelichting aantreffen. Erger is nog dat de auteur op een aantal punten de statistiek geweld aandoet. Zo zijn enkele definities onvolledig (b.v. van aselekt trekken) of onduidelijk (b.v. van kansfunctie) en worden termen onjuist gebruikt (betrouwbaarheids grens in plaats van kritieke waarde). Ook de stijl is hier en daar hinderlijk ongewoon, of men moet gecharmeerd zijn van grappen als het gebruik van het woord 'toetsen' in plaats van toetsingsgrootte.

Elk hoofdstuk wordt besloten met een lijst 'kernbegrippen', een stel beweringen (juist/onjuist), en een aantal vraagstukken. De beweringen zijn soms slecht geformuleerd, soms inderdaad juist of onjuist en soms zeer discutabel. Sommige beweringen wekten de indruk als kwinkslag bedoeld te zijn, hetgeen niet bevorderlijk is voor het zelfvertrouwen van de onzekere student. Bovendien was een deel van de gegeven antwoorden fout, terwijl ik over enkele waarbij discussie mogelijk was,

met de auteur van mening verschilde. De uitwerkingen van de vraagstukken bevatten ook wel enkele fouten, maar lieten toch minder te wensen over dan de beweringen met hun antwoord.

Een overzicht van de inhoud: basisbegrippen, maten voor ligging en spreiding, kansrekening, binomiale, hypergeometrische, Poisson- en normale verdeling, centrale-limietstelling, beknopt schatten en toetsen, indexcijfers, iets over tijdreeksen, kleinste-kwadratenmethode bij één en twee variabelen. De lengte van de tekst van de negentien hoofdstukken varieert van vier tot acht bladzijden. Vele lezers zullen zich getroost voelen door de opdracht voorin: 'Omdat je niet alles kunt begrijpen'.

J.M. Buhrman

Hogeschool voor Economische Studies, Amsterdam

Leonard Fortuin, Mynt Zijlstra, Hieke Besselaar en Karin de Groot

De Kroonjuwelen, kwantitatieve modellen en methoden in verleden, heden en toekomst.

CQM, Eindhoven, 1992, vi + 255 pag., f 49.00

Dit boek is het liber amicorum voor Dr J.P.M (Jos) de Kroon, aangeboden ter gelegenheid van zijn afscheid van het CQM op 11 december 1992. In dit boek zijn de bijdragen bijeengebracht van 39 auteurs, die allemaal een kort stukje hebben geschreven ter gelegenheid van het genoemde afscheid. Een aantal auteurs hebben een stuk geschreven over het onderzoeksgebied waar zij bezig zijn, sommige hebben theorie en toepassing gemixed, waarbij de toepassing voor een deel over het functioneren van Jos de Kroon gaat. Veel mensen voegen een warm woord van dank en vriendschap toe en dat maakt het boek wat het moet zijn, een verzameling van persoonlijke blijken. Als alle namen in het boek onbekend voorkomen, zal de inhoud van het boek ook niet zo aanspreken. Maar als de naam Jos de Kroon bekend voorkomt en de schrijvers voor een deel bekend zijn, dan voegt dit boek iets extra's toe aan de niet aflatende stroom van boeken. Het is het persoonlijke wat spreekt in dit boek en dat is ook de reden voor het koopadvies.

Bijdragen zijn geleverd door

A.G.H. Rinnooy Kan	Emile H.L. Aarts	Paul Banens
P.M. Upperman	Peter J.M. van Laarhoven	Ronald J.M.M. Does
Willem Schaafsma	Paul van Beek	M. Hasselaar
W. Albers	R. Doornbos	J. Oudemans
Gerard Leppink	Peter Sander	B. Schriever
Fred Steutel	Luit Hoving	Stan van Hoesel
Paul van der Laan	T.J. Terpstra	Cor Hurkens
F.A. Lootsma	Carel. L. Scheffer	Jan Karel Lenstra
Albert Trip	Eduard de Jager	Martin Savelsbergh
Ivo W. Molenaar	Roel van Strik	Ton de Kok
Bouke van der Veen	B.B. van der Genugten	Jan Engel
Jaap Wessels	W.H.M. Zijm	Mynt Zijlstra
Leonard Fortuin	H.A. Fleuren	Jan B.M. Doremalen

Het boek is verkrijgbaar bij het CQM B.V., Postbus 414, 5600 AK Eindhoven tel 040-757747.

Arend Oosterhoorn

A.D. Lunn en D.R. McNeill

Computer-Interactive Data Analysis

John Wiley & Sons, New York, 1991, xiv + 374 pag., ISBN 0-471-93114-4, £ 39.95, inclusief twee diskettes.

Op het eerste gezicht doet de titel van dit boek vermoeden dat het gaat om een leerboek over interactieve data-analyse met behulp van de computer. Dit is echter maar ten dele waar. Het boek beschrijft hoe je data-analyse kunt uitvoeren met behulp van het statistische pakket SPIDA. Je zou het daarom met evenveel recht een handleiding kunnen noemen voor het betreffende computerprogramma.

Bij het boek zitten twee diskettes. De eerste bevat het statistische pakket SPIDA. Het is een gelimiteerde versie van het pakket dat eveneens commercieel wordt gedistribueerd. De beperking zit hem erin dat een te analyseren data-matrix slechts 1200 waarden mag bevatten. De tweede diskette bevat een groot aantal datasets die ook in het boek worden gebruikt voor het illustreren van allerlei statistische analysetechnieken.

SPIDA draait op microcomputers onder MS-DOS. Het is een commando-gestuurd pakket. De taal voor het specificeren van de commando's doet soms wat primitief aan. Met name de nadere specificatie van de parameters is niet altijd even inzichtelijk. SPIDA bevat een krachtige taal voor het manipuleren van vectoren en matrices. Dat maakt het mogelijk om niet aanwezige technieken eventueel toch te programmeren.

Het boek beschrijft een flink aantal technieken. De nadruk ligt iets meer op verklarende dan op exploratieve analysetechnieken. Toch worden deze laatste technieken zeker niet vergeten, met name daar waar het gaat om het controleren van modelveronderstellingen.

Het boek gaat niet erg diep in op de statistische theorie. Hier en daar worden wel enkele formules gegeven, maar ze worden amper uitgelegd, laat staan afgeleid. Het aardige van het boek is dat het de statistische technieken illustreerd aan de hand van een groot aantal datasets. Veel dan die voorbeeld-datasets zijn afkomstig uit het boek Data van Andrews en Herzberg (1985).

Het eerste hoofdstuk is gewijd aan de werking van het programma en de wijze waarop de command's moeten worden ingevoerd.

Het tweede hoofdstuk gaat over 1- of 2-steekproef problemen. Aan de orde komen chi-kwadraat-toetsen, t-toetsen, en enkele verdelingsvrije toetsen. Ook transformaties worden behandeld als een middel om de gegevens aan de modelveronderstellingen te laten voldoen. Een voorbeeld van de wat merkwaardige structuur van de SPIDA-commando's is dat de toets van Wilcoxon kan worden toegepast door aan het commando ANOVA de parameter $t=1$ mee te geven.

Het derde hoofdstuk is getiteld 'Multiple comparisons'. In dit hoofdstuk wordt met name ingegaan op de toepassing van 1-weg variantie-analyse en de toets van Kruskal en Wallis. Box-and-Whisker-plots zijn het belangrijke instrument voor het controleren van de onderliggende verdeling.

Hoofdstuk vier beschrijft simpele regressie-analyse met één onafhankelijke variabele. Terecht wordt eerst aanbevolen om spreidingsdiagrammen te maken ten einde te kunnen controleren of het verband wel lineair is. Tevens wordt het nut van transformaties onderstreept om dit eventueel te bereiken. Hoofdstuk vijf gaat dan verder met multiële regressie-analyse. De matrix van puntenwolken en de grafische analyse van de residuen worden hierin ten tonele gevoerd als belangrijke diagnostische instrumenten. Tenslotte wordt enige aandacht besteed aan de analyse van gegevens waarin autocorrelatie optreedt.

Hoofdstuk zes gaat over 2-weg variantie-analyse. Dat gebeurt op vrij uitgebreide wijze. Het boek laat zien dat ANOVA-schattingen via regressie-analyse kunnen worden berekend, als eerst dummy-variabelen worden ingevoerd. Ook wordt in dit hoofdstuk de toets van Friedman behandeld.

Hoofdstuk zeven is gewijd aan de analyse van kwalitatieve variabelen. Eerst wordt de kruistabel behandeld met de bijbehorende chi-kwadraat-toets. Daarna wordt het loglineaire model behandeld.

Hoofdstuk acht bespreekt twee technieken voor het analyseren van epidemiologische gegevens: logistische regressie-analyse en survival analyse.

Hoofdstuk 9 en 10 beschrijven de analyse van multivariate gegevens. Aan de orde komen principale componentenanalyse, factoranalyse, manova, discriminantanalyse en schalingstechnieken. Het is prettig te zien dat het boek uitgebreid ingaat op de conceptuele verschillen tussen principale componentenanalyse en factoranalyse.

Het boek laat een breed scala van statistische technieken in actie zien. De behandeling van de statistische theorie zelf is niet de sterkste kant van dit boek. De gebruiker van het boek (en het bijbehorende pakket) zullen die kennis elders moeten hebben opgedaan. Wel biedt het boek een goede leidraad voor het in de praktijk toepassen van allerlei technieken. Terecht legt het boek veel nadruk op het controleren van modelveronderstellingen. De voorbeeld-datasets in het boek kunnen een bron van inspiratie zijn voor statistici die betrokken zijn bij het onderwijs in de statistiek. Het is ook heel goed denkbaar dat boek en programma te samen kunnen worden ingezet in een onderwijs situatie.

Referentie:

Andrews, D.F. & A.M. Herzberg (1985): *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer Verlag: New York.

Jelke Bethlehem
CBS / Universiteit van Amsterdam

B.F.J. Manly

Randomisation and Monte Carlo methods in Biology

Chapman and Hall, London, 1991, xiii + 281 pag, ISBN 0-412-36710-6, £ 27.50

De doelgroep wordt omschreven met: "It is assumed that readers are familiar with the methods that are usually covered in a first university course in statistics for students who are specializing in other subjects. A somewhat fuller knowledge of analysis of variance, regression and multivariate methods will be helpful in some places". Inderdaad beweegt het boek zich veelal op een introducerend nivo, d.w.z. de technieken worden aan de hand van enige voorbeelden toegelicht en voor een grondiger bespreking wordt verwezen naar andere bronnen, bijv Edgington (1987). Toch is het blijkbaar wel de bedoeling van de auteur, dat de lezer zelf aan de slag gaat, want ieder hoofdstuk wordt afgesloten met FORTRAN-subroutines.

Het boek bevat zeer veel voorbeelden. Dit heeft als voordeel, dat de toepassing goed geïllustreerd wordt en werkt enthousiasmerend voor het gebruik van randomisatie-methoden. Nadeel is, dat de theorie en de praktijk niet meer gescheiden zijn en daardoor het lezen soms veel van het inlevingsvermogen vergt.

Het boek begint met de algemene beschrijving van randomisatie-tests en betrouwbaarheidsintervallen. Zonder veel details wordt de randomisatie-test geïntroduceerd en afgezet tegen de klassieke. Na lezing van dit hoofdstuk bleef bij mij de indruk achter, dat deze test een alomwerkend middel is dat je altijd kan toepassen. Iets meer aandacht aan randomisatie bij het uitvoeren van een experiment en op onafhankelijkheid was m.i. op zijn plaats geweest. Hoofdstuk 2 beschrijft Monte Carlo methoden, wat je ook een vorm van de parametrische bootstrap kan noemen, de jackknife en de bootstrap. Het hoofdstuk beslaat 8 bladzijden en kan dus niet meer dan een eerste introductie in deze methoden zijn. Na nog een paar algemene opmerkingen, bijv over het benodigd aantal randomisaties, volgt de uitwerking voor randomisatie-methoden in een aantal gebieden.

De toepassingen die aan bod komen, zijn: een- en twee-steekproeven toetsen, ANOVA, lineaire regressie, afstandsmatrices, ruimtelijke gegevens, tijdreeksen, multivariate gegevens en nog enige ad hoc situaties. De meeste onderwerpen zijn ook terug te vinden in Edgington maar worden wat lichtvoetiger behandeld. Soms neemt Manly een ander standpunt in, bijv over interactie-toetsen bij ANOVA, zonder veel diepgang, overigens. Hierover heeft Cajo Ter Braak gesproken tijdens een LBS-bijeenkomst op 1 maart 1991.

Om de associatie tussen twee afstandsmatrices te bekijken, wordt Mantels test (1967) behandeld, die neerkomt op de correlatie tussen de overeenkomende elementen uit beide matrices. De behandeling van ruimtelijke gegevens is een aanvulling op Edgington en stamt uit de ruimtelijke statistiek literatuur. Dit hoofdstuk komt erg fragmentarisch over.

Het boek ziet er goed verzorgd uit en lijkt weinig drukfouten te bevatten.

Conclusie

Randomisatie-testen is als onderwerp belangrijk genoeg om een boek over te laten verschijnen. Voor statistici levert dit boek een scala aan voorbeelden. Door de beperkte diepgang kan het niet als verbetering van Edgington gezien worden, wel als aanvulling op een aantal gebieden.

Voor niet-statistici kan dit boek een redelijk leesbare introductie vormen. Probleem blijft natuurlijk, dat de methodologie zelf vrij simpel uit te leggen is, maar de toepassingen in diverse gebieden van de statistiek de moeilijkheidsgraad verhogen. Ik denk dan ook, dat voor de omschreven doelgroep het boek behoorlijk zwaar zal zijn.

Doordat, noodgedwongen, een dwarsdoorsnede ontstaat van de statistische discipline, komt het boek nogal eens fragmentarisch over.

Peter van Ewijk
Solvay Duphar bv

> > > > > > > > > **Binnengekomen boeken** < < < < < < < < <

Om de nieuws waarde van de boekbesprekingsrubriek te verhogen wordt een lijst gepresenteerd van boeken die in de afgelopen periode bij de redactie zijn binnengekomen. Omdat er soms nogal wat tijd kan zitten tussen de uitgave van het boek en de publicatie van de recensie, hoopt de redactie hiermee de dienstverlening aan de lezers te vergroten.

Boeken waarvoor een $\boxtimes$ symbool staat zijn nog niet ondergebracht bij een recensent. Mocht er belangstelling bestaan voor het bespreken van een boek, dan kan contact opgenomen worden met de boekbesprekingsredacteur. Met nadruk wordt gesteld dat niet zeker is of het boek nog 'in de aanbieding' is.

Senn, S.

Cross-over trials in clinical research

John Wiley & Sons, New York, 1993, xiii + 266 pag., ISBN 0-471-93493-3, £ 24.95

Thompson, S.K.

Sampling

John Wiley & Sons, New York, 1992, xv + 339 pag., ISBN 0-471-54045-5, £ 41.95

Coffman, E.G., J.K. Lenstra & A.H.G. Rinnooy Kan

Computing

North Holland, Amsterdam, 1992, x + 679 pag., ISBN 0-444-88097-6, f 310.00

$\boxtimes$ **Biswas, S.**

Topics in statistical methodology

John Wiley & Sons, New York, 1991, xvi + 611 pag., ISBN 0-470-21153-9, £ 20.95

$\boxtimes$ **White, D.J.**

Markov Decision Processes

John Wiley & Sons, New York, 1993, xiv + 224 pag., ISBN 0-471-93627-8, £ 29.95

$\boxtimes$ **Rubinstein, R.Y. & A. Shapiro**

Discrete event systems

John Wiley & Sons, New York, 1993, xv + 334 pag., ISBN 0-471-93419-4, £ 39.95

$\boxtimes$ **Csörgö, M. & L. Horváth**

Weighted approximations in probability and statistics

John Wiley & Sons, New York, 1993, xiii + 442 pag., ISBN 0-471-93635-9, £ 55.00

>>>>>>>> **Proefschriften, CWI uitgaven** <<<<<<<<<

▣ Savelsbergh, M.W.P.

Computer aided routing

CWI, Amsterdam, 1992, iv + 132 pag., ISBN 90-6196-412-1, f 40.00

▣ Kalfs, N.

Hour by hour: effects of the data collection mode in time use research

NIMMO, Amsterdam, 1993, ix + 174 pag., ISBN 90-74598-01-3

In het onderzoek dat leidde tot dit proefschrift worden drie methoden van onderzoek, het schriftelijk interview, het 'face-to-face' interview en het of telefonische interview, naast elkaar gezet om te bepalen welke methode het meest geschikt is voor het onderwerp tijdsbesteding.

De kwaliteit van gegevens verzameld met een elektronisch dagboek staat in dit onderzoek centraal, waarbij de kwaliteit beoordeeld is op basis van validiteit, betrouwbaarheid, respons en kosten. Twee methoden worden bekeken: 'self-administered' interview (CASI) en telefonisch interview (CATI). Bovendien wordt de kwaliteit van deze computergestuurde methoden vergeleken met het traditionele 'paper-and-pencil' interview (PAPI).

Voor het uitvoeren van het onderzoek is voor elke dataverzamelingsmethode het meest efficiënte design gekozen. Indien noodzakelijk en mogelijk is gecorrigeerd voor implementatiekenmerken.

In hoofdstuk 4 worden de verschillen tussen CASI en CATI gepresenteerd en in hoofdstuk 5 tussen PAPI en CASI. In hoofdstuk 6 worden de tijdsbestedingsschattingen vergeleken en wordt onderzocht of de gevonden verschillen door de methoden kunnen worden verklaard. Omdat dit niet zo blijkt te zijn wordt in hoofdstuk 7 bepaald welke activiteiten nader onderzoek vereisen. De resultaten daarvan worden in hoofdstuk 8 gepresenteerd.

De algemene conclusie is dat geen van de methoden als de beste kan worden aangeduid. In PAPI zijn de gegevens het meest valide en het betrouwbaarst, in CATI worden de minste fouten gemaakt en is de respons het hoogst. In CASI is de minste tijd gemoeid met het uitvoeren van het onderzoek. Verder worden er nog wat gedetailleerder verschillen aangegeven.

Het proefschrift is verkrijgbaar bij NIMMO, Universiteit van Amsterdam, Oude Turfmarkt 151, 1012 GC Amsterdam.

▣ **Kort, J. de**

Edge-disjointness in combinatorial optimization

Thesis Publishers, Amsterdam, 1992, xii + 222 pag., ISBN 90-5170-178-0, f 37.50

Dit is het proefschrift van de Kort, uitgegeven door Thesis Publishers, Postnus 14791, 1001 LG Amsterdam. De door de auteur meegestuurde samenvatting luidt als volgt;

Naast de gebruikelijke doelstelling van kostenminimalisatie in een netwerk, raken onderzoekers meer en meer geïnteresseerd in netwerken die een zekere graad van betrouwbaarheid garanderen. Deze garantie wordt bijvoorbeeld gegeven door kant-disjunctheid restricties op te leggen aan bestaande netwerk problemen.

Deel I van het proefschrift geeft een overzicht van formuleringen en algoritmen voor netwerk problemen die rekening houden met kant-disjunctheid restricties. Een overzicht van complexiteitsresultaten wordt gegeven.

Deel II behandelt een klassiek branch-and-bound algoritme om twee kant disjuncte Hamiltonian cycles van minimale totale lengte in een graaf te bepalen. Aandacht wordt besteed aan bepaling van ondergrenzen, bovengrens-heuristieken, kant-eliminatie testen en splitsingsregels.

Deel III beschrijft een polyhedraal brach-and-cut algoritme voor het cycle vraagstuk van deel II, met aandacht voor facet-inducerende ongelijkheden en geïmplementeerde facet-identificatie procedures en splitsingsregels. Rekenregels illustreren de praktische toepasbaarheid van procedures beschreven in deel I en deel II.

▣ **E.D. de Leeuw**

Data quality in mail, telephone and face to face surveys

TT-publicaties, Amsterdam, 1992, xii + 168 pag., ISBN 90-801073-1-X, f 30.00

Met dit proefschrift is Edith de Leeuw op 22 oktober 1992 gepromoveerd. In het proefschrift worden drie belangrijke dataverzamelmethode voor sociaal-wetenschappelijk survey onderzoek, te weten postenquête, het telefonisch interview en het 'face-to-face' interview met elkaar vergeleken. Centraal staat de vraag of, en zo ja, in hoeverre de gegevens verkregen via deze drie dataverzamelmethode van elkaar verschillen.

Na een korte omschrijving wordt in hoofdstuk 2 een overzicht gegeven van de verschillende theoretische overwegingen omtrent het ontstaan van mogelijk verschillen. In hoofdstuk 3 wordt de bestaande empirische onderzoeksliteratuur samengevat, gebruik makend van de methode van meta-analyse.

A probabilistic analysis of the k-location problem

Dean P. Foster & Rakesh V. Vohra

De met cijfers bedekte negentiende eeuw

Toepassingen van statistiek en waarschijnlijkheidsrekening in Nederland en Vlaanderen tussen 1840 en 1920.

Bij de redactie is de 15e jaargang van het tijdschrift GEWINA, tijdschrift voor de geschiedenis der geneeskunde, natuurwetenschappen, wiskunde en techniek binnengekomen. Nummer 3 van deze jaargang is een themanummer met het in de aanhef vermelde onderwerp. De redactie van dit nummer lag in handen van I.H. Stamhuis en A. de Knecht - van Ekelen.

Na een inleiding met een kort overzicht van de artikelen volgen

De 'probabilistic revolution' in de wetenschappen

Ida. H. Stamhuis

Een procesverbaal van verhoor

Elly Dekker

Opvattingen over geneeskundige statistiek in het 'Nederlandsch tijdschrift voor geneeskunde' rond de eeuwwisseling

A. de Knecht - van Ekelen

'Als 't ware een nieuwe wetenschap'. De toepassing van de statistische methode door de Gentse botanici rond 1900

Geert Vanpaemel

Statistiek en waarschijnlijkheidsrekening in het werk van Tine Tammes (1871-1947)

Ida H. Stamhuis

De andere drie nummers bevatten bijdragen over geschiedenis van andere vakgebieden.

Het bestaan van deze vereniging was mij niet bekend, zo ook het tijdschrift niet. Maar van het lezen van een aantal van deze bijdragen heb ik een zeer goede indruk overgehouden. Het is waarlijk zeer leuk om te lezen hoe de statistiek en de kansrekening in de geschiedenis is toegepast en in sommige gevallen weer aan de kant geschoven, zoals in het verhaal over Tine Taammes, Gronings biologe en later hoogleraar in Groningen. Zij paste de statistiek eerst toe in een onderzoek naar vlas, maar zag in later werk af van het gebruik van de statistiek.

Het lezen van dit soort verhalen is goed voor de algemene ontwikkeling en de kijk op de strijd die het vakgebied in de geschiedenis heeft moeten voeren om volledig uitgenut te worden bij het

doen van onderzoek. Alhoewel, 'heeft moeten voeren' is veel te voorbarig gesteld, gezien de hedendaagse invloed van statistiek en statistici.

Nadere informatie en naar ik vermoed ook losse nummers zijn te verkrijgen bij

GEWINA

Instituut voor geschiedenis der natuurwetenschappen

Nieuwegracht 187

3512 LM UTRECHT

Oproep voor boekbespekers

Als redacteur van de boekbesprekingsrubriek van de VVS, welke gepubliceerd wordt in Kwantitatieve Methoden, ben ik steeds op zoek naar mensen die bereid zijn een pas verschenen boek te willen lezen met het doel daarover een oordeel te geven. Dat oordeel wordt op prijs gesteld door de leden van de VVS. Het verschaft inzicht in de kwaliteiten van het boek. Voor de schrijver(s) kan een oordeel van het publiek waarvoor het boek geschreven is een mogelijkheid bieden tot verdere verbetering. Ook de uitgever heeft er uiteraard baat bij. Als tegenprestatie voor het beoordelen mogen de recensenten de besproken boeken behouden.

Behalve de algemene boeken over statistiek, waar veel mensen zinnige dingen over kunnen zeggen, betreft het vaak ook boeken met een specialistisch onderwerp. Als ik, soms op de gok, iemand een dergelijk boek toestuur, dan hoor ik een enkele keer als reactie dat 'het boek precies aansluit bij het onderzoek' dat de recensent verricht. Dat is prettig, omdat veel mensen daarbij voordeel hebben. De recensent, omdat hij/zij het nieuw verschenen boek op het onderzoeksgebied ter beschikking heeft, en de auteur(s) omdat er een terzake kundig oordeel wordt gegeven.

Om nu de succesans van het toesturen van een boek te vergroten, wil ik graag in contact komen met veel collega statistici, kansrekenaars en OR-beoefenaren, al zijn de boeken op dit laatste gebied schaars. Stuur het onderstaande antwoordstrookje naar het adres van de boekbesprekingsrubriek als u interesse heeft om in de toekomst een boek te bespreken. Als er dan boeken binnenkomen heb ik een grotere keuze van mogelijke recensenten en u wellicht een gloednieuw boek op uw onderzoeksgebied.

Ja, ik ben graag bereid een boek op mijn interessegebied te bespreken voor de leden van de VVS

naam

adres

interessegebied (graag enigszins nauwkeurig omschrijven)
