

Boekbesprekingen

Redactie: Douwe-Frits Broens
 Vakgroep Marktkunde en Marktonderzoek
 Landbouwniversiteit Wageningen
 Hollandseweg 1
 6706 KN Wageningen

08370 - 84353 telefoon
 08370 - 84361 fax
 douwe-frits.broens@alg.menm.wau.nl e-mail

Besproken boeken in Kwantitatieve Methoden nr. 50

Jan Beran

Statistics for Long-Memory Processes

Chapman & Hall, New York, 1994, x + 315 pag., ISBN 0-412-04901-5

Bernardo, J.M. and Smith, A.F.M.

Bayesian Theory

Wiley, Chichester, 1994, xiv + 586 pag., ISBN 0-471-92416-4, £ 60.00.

T.F. Cox en M.A.A. Cox,

Multidimensional Scaling. Monographs on Statistics and Applied Probability, nr.59.

London: Chapman & Hall, 1994, pp. 213.

Peter J. Diggle, Kung-Yee Liang en Scott L. Zeger.

Analysis of Longitudinal data

Oxford University Press, Oxford, 1994, ISBN 0 19 852284 3

U. Engel en J. Reinecke.

Panelanalyse: Grundlagen, Techniken, Beispiele.

Walter de Gruyter, Berlin, 1994, ix + 327 pag., ISBN 3-11-013570-1, prijs onbekend.

P.R. Freeman en A.F.M. Smith (Editors)

Aspects of Uncertainty, A tribute to D.V. Lindley

Wiley, Chichester, 1994, xviii + 392 pag., ISBN 0-471-94347-9, £ 45.00.

N.L. Johnson, S. Kotz en A.W. Kemp

Univariate Discrete Distributions

John Wiley and Sons, New York, 1992, 2e ed., xx + 565 pag., ISBN 0-471-54897-7

Jan Beran

Statistics for Long-Memory Processes

Chapman & Hall, New York, 1994, x + 315 pag., ISBN 0-412-04901-5

Dit is het eerste boek over tijdreeksmodellen met een "lang-geheugen" (long-memory). Het geeft een nauwkeurige samenvatting van de statistische methoden die de afgelopen jaren zijn ontwikkeld om gegevens te analyseren met een "lang-geheugen". Hoewel het boek af en toe behoorlijk technisch is, is de auteur de praktijk niet uit het oog verloren. Met behulp van voorbeelden uit o.a. de hydrologie, informatica en economie wordt de lezer door elegante theorie van dit vakgebied geleid. Voor de lezer die beschikt over enige kennis van de tijdreeksanalyse, zeg op het niveau van Brockwell & Davis (1987), is dit een prima boek om in dit vakgebied thuis te raken.

Het boek bestaat uit de volgende hoofdstukken:

1. Introduction
2. Stationary processes with long memory
3. Limit theorems
4. Estimation of long memory: heuristische approaches
5. Estimation of long memory: time domain MLE
6. Estimation of long memory: frequency domain MLE
7. Robust estimation of long memory
8. Estimation of location and scale, forecasting
9. Regression
10. Goodness of fit tests and related topics
11. Miscellaneous topics
12. Programs and data sets

De hoofdstukken 1 en 2 dienen als inleiding. De voorbeelden die in hoofdstuk 1 besproken worden, komen later weer terug om de theorie te verduidelijken. De andere hoofdstukken staan op zich zelf. In hoofdstuk 11 komen processen met oneindige variantie, fractionele Gegenbauer ARMA's en simulatie aan de orde. Het boek bevat wel de stellingen, maar niet de bewijzen. Tevens bevat het boek Splus programma's om "lange-geheugen" modellen te schatten.

Het boek is een aanrader voor iedereen die kennis wil maken met, of onderzoek doet op het terrein van de "lange-geheugen" tijdreeksmodellen. De auteur is erin geslaagd om een gedegen beschrijving van dit vakgebied te geven, zonder te verzanden in allerlei theoretische en technische details. De toepasser van tijdreeksanalyse kan op basis van de vele voorbeelden al snel aan de slag, de onderzoeker kan via de zeer uitgebreide bibliografie zijn/haar weg vinden in de literatuur. Het is een mooi werk dat in de boekenkast van elke tijdreeksanalyst een plaats verdient.

Freek Huele

Vakgroep Wiskunde, Faculteit der Wiskunde en Informatica
Universiteit van Amsterdam

Bernardo, J.M. and Smith, A.F.M.

Bayesian Theory

Wiley, Chichester, 1994, xiv + 586 pag., ISBN 0-471-92416-4, £ 60.00.

Dit boek is het eerste uit een serie van drie, 'Bayesian Theory', 'Bayesian Computation' (A.E. Gelfand & Smith) en 'Bayesian Methods' (Bernardo & Smith), bedoeld om een zo volledig mogelijk overzicht te geven van het 'waarom, hoe en wat' van Bayesiaanse statistiek. Dit is een zeer omvangrijk en ambitieus project, dat waarschijnlijk ook zeer succesvol zal zijn. Dit eerste boek is ongetwijfeld het meest complete en beste overzicht van Bayesiaanse theorie dat momenteel verkrijgbaar is, en het is ook een verkoopssucces (men is al toe aan de tweede druk!). Het is jammer dat de twee volgende delen nog op zich laten wachten, zowel omdat de hoge kwaliteit en leesbaarheid van dit eerste boek doen uitzien naar het vervolg, als ook omdat de scheidslijnen tussen de drie onderwerpen niet scherp zijn.

Het Bayesiaans concept voor statistiek en beslissingstheorie is geheel opgebouwd vanaf enkele basis principes. Wanneer men een aantal axioma's accepteert (o.a. met betrekking tot preferenties en de centrale rol van nutsfuncties) leveren deze een algemene methode waarmee alle statistische problemen aangepakt kunnen worden. De nadruk van dit boek ligt op de axioma's en de daarop gebaseerde ontwikkeling van de algemene theorie. Het boek bestaat uit 6 hoofdstukken: Introduction, Foundations, Generalisations, Modelling, Inference, Remodelling, en 2 appendices (appendix B geeft een bespreking van niet-Bayesiaanse theorieën). Naast de volledigheid is het sterkste punt van het boek de verteltrant. De uitgever heeft de auteurs duidelijk de ruimte gegeven om een uitermate lezervriendelijk boek te schrijven, waardoor het voor de genteresseerde lezer geen zware opgave is het boek van begin tot eind te lezen. Nergens worden verwijzingen gebruikt voor belangrijke stappen in het verhaal. Ieder hoofdstuk wordt afgesloten met een uitgebreide sectie 'discussie en verdere referenties', de referentielijst telt 66 bladzijden. Mede hierdoor is het een prima up-to-date overzicht van wat er verschenen is op het gebied van Bayesiaanse theorie, en het boek zal een referentiewerk bij uitstek worden. Het lijkt dan ook onmisbaar voor iedere statisticus, zelfs de meest fervente tegenstanders van de Bayesiaanse theorie kunnen eigenlijk niet om het boek heen. Vanwege de duidelijke beschrijving en bespreking van de axioma's biedt het boek deze tegenstanders de uitgelezen mogelijkheid na te gaan en aan te geven waarom men deze theorie afwijst, waardoor het ook voor hen een goed en bruikbaar boek is. Het is een aanwinst voor iedere bibliotheek en tevens toegankelijk voor studenten, alhoewel het duidelijk niet bedoeld is als cursus materiaal. Het lijkt me ook dat het boek zeer geschikt is voor niet-statistici die eens wat van de Bayesiaanse theorie willen weten (alhoewel een goede basis kennis van de wiskunde nodig is). Omdat het Bayesiaans concept beslissingsgericht is denk ik hierbij met name aan beoefenaars van stochastische operations reseach en aanverwante gebieden.

Ook voor mensen met speciale interesse voor de grondslagen van statistiek heeft het boek veel te bieden. De laatste drie hoofdstukken verschaffen veel inzicht in mogelijke argumenten voor modelkeuze en aspecten van inferentie. Niet eerder zag ik een boek waarin zoveel fundamentele onderwerpen zo duidelijk besproken worden.

Samenvattend: dit boek is een van de belangrijkste en duidelijkste boeken over statistische theorie die tot nu toe verschenen zijn. Maar omdat statistiek een wetenschap is die bij uitstek bedoeld is voor toepassing zijn de vervolgdelen onmisbaar voor de totale waardering.

F.P.A. Coolen

Department of Mathematical Sciences,
University of Durham, England

T.F. Cox en M.A.A. Cox,

Multidimensional Scaling. Monographs on Statistics and Applied Probability, nr.59.

London: Chapman & Hall, 1994, pp. 213.

Multidimensional Scaling geeft een overzicht van een groot aantal technieken voor de analyse van multivariate gegevens. De besproken technieken hebben tot doel deze gegevens te vereenvoudigen in de vorm van een grafische weergave van de onderzochte objecten. In deze weergave dienen de afstanden tussen de objecten een zo adequaat mogelijke afspiegeling te zijn van de geobserveerde relaties tussen de objecten. De technieken verschillen wat betreft het type data waarop ze toegepast kunnen worden en de precieze aard van de doelstelling. De volgende technieken en programma's worden in het boek besproken:

- *Metrisch en nonmetrisch multidimensioneel schalen*: Representatie van ongelijkenisdata (metrisch of nonmetrisch) tussen elementen uit eenzelfde set middels afstanden in een ruimte van zo laag mogelijke dimensionaliteit.
- *Procrustes analyse*: Constructie van een puntenconfiguratie die zo goed mogelijk past bij een set van individuele puntenconfiguraties.
- *Ontvouwing*: Representatie van (conditionele) ongelijkenisdata tussen elementen uit verschillende sets middels afstanden in een ruimte van zo laag mogelijke dimensionaliteit.
- *Correspondentie analyse*: Ruimtelijke representatie van onder andere de interactie rij- en kolomvariabelen van een kruistabel.
- *"Individual difference scaling"*: Multidimensioneel schalen waarbij individuele verschillen gerepresenteerd worden (onder andere de programma's INDSCAL en IDIOSCAL).
- *ALSCAL en SMACOF*: Algemene programma's waarmee onder andere multidimensioneel schalen, ontvouwing en "individual difference scaling" uitgevoerd kan worden.
- *CANDECOP, PARAFAC, en CANDELINC*: Geavanceerde schaaltechnieken voor meer complexe datasets ("three-way three-mode" data).

Voor ieder van deze technieken presenteren de auteurs op een bondige wijze de onderliggende wiskundige ideeën. Vanwege de breedte van het boek, worden van sommige technieken alleen de meest basale aspecten gepresenteerd. In alle gevallen worden er voldoende verwijzingen gegeven naar relevante literatuur, zodat verdere verdieping mogelijk is. Het overzicht lijkt tamelijk compleet te zijn, waarbij ook meer recente technieken besproken worden.

De basisconcepten en wiskundige gereedschappen welke nodig zijn om de hoofdttekst te

begrijpen worden gegeven in een inleidend hoofdstuk. Gezien de bondigheid waarmee sommige zaken hier gepresenteerd worden, lijkt dit hoofdstuk meer bedoeld te zijn om zaken op te frissen dan om deze uit te leggen aan de lezer. Dit geldt vooral voor de matrix algebra welke een belangrijk gereedschap blijkt in de uitleg van de verschillende technieken. Voor een lezer die niet zo goed op de hoogte is van zaken als bijvoorbeeld singuliere waarden decompositie, spectrale decompositie en de conceptuele betekenis van deze wiskundige operaties is het aan te bevelen eerst een inleidend werk te lezen zoals dat van Green en Carroll (1976).

Het theoretische karakter van *Multidimensional Scaling* wordt gecompenseerd door de vele voorbeelden waarin de technieken toegepast worden op datasets. Voorbeelden uit verschillende onderzoeksdomeinen zoals de ethologie, de voedingsleer, en de kleurwaarneming worden gegeven. Alhoewel deze voorbeelden illustreren hoe de technieken toegepast kunnen worden en bovendien zorgen voor een aangename afwisseling, blijken deze delen, bezien vanuit de lezer die in praktische toepassingen geïnteresseerd is, het zwakst te zijn. De toepassingsmogelijkheden van de technieken worden niet voldoende besproken, en de interpretatie van de analyse-resultaten wordt vaak expliciet aan de lezer overgelaten. Dat de interpretatie geen triviale aangelegenheid hoeft te zijn, blijkt uit discussies die experts zelf soms op dit gebied voeren (zie bijvoorbeeld de discussie van Greenacre (1989) en Carroll et al. (1989) omtrent de interpretatie van de resultaten van correspondentie-analyse). Blijkbaar was het de bedoeling van de auteurs zich te richten op de wiskundige kant van de technieken. Vanuit die optiek ware het verstandig geweest indien zij de doelgroep of in ieder geval de veronderstelde voorkennis van de lezers in het voorwoord geëxpliciteerd hadden.

Bij het boek hoort een diskette met verscheidene computerprogramma's. Dit stelt de lezer in staat om de verschillende technieken zoals die in het boek behandeld worden zelf toe te passen. De programmatuur draagt aldus bij aan het verlevendigen van de stof. Voor meer serieuze toepassingen is de programmatuur niet gebruikersvriendelijk genoeg en zijn de beperkingen te groot. Een handleiding voor het gebruiken van de bijgesloten programmatuur alsmede nuttige referenties naar programmatuur die in de loop der jaren door anderen is ontwikkeld, worden gegeven in de appendix.

Samenvattend kan gesteld worden dat *Multidimensional Scaling* een gecomprimeerde bespreking geeft van een groot aantal technieken voor de analyse van multivariate data. Behoudens enkele kleine typografische fouten is de leesbaarheid goed te noemen. De waarde van dit boek is vooral gelegen in de compactheid van de uiteenzetting alsmede in het bij elkaar brengen en vergelijken van een groot aantal technieken op het gebied van meerdimensioneel schalen.

Literatuur:

- Carroll, J.D., Green, P.E., & Schaffer, C.M. (1989). Reply to Greenacre's commentary on the Carroll-Green-Schaffer scaling of two-way correspondence analysis solutions. *Journal of Marketing Research*, 26, 366-368.
- Green, P.E., & Carroll, J.D. (1976). *Mathematical Tools for Applied Multivariate Analysis*. New York: Academic Press.
- Greenacre, M.J. (1989). The Carroll-Green-Schaffer scaling in correspondence analysis: A theoretical and empirical appraisal. *Journal of Marketing Research*, 26, 358-365.

Math Candel
 Vakgroep Marktkunde en Marktonderzoek
 Landbouwwuniversiteit Wageningen

Peter J. Diggle, Kung-Yee Liang en Scott L. Zeger.
Analysis of Longitudinal data
 Oxford University Press, Oxford, 1994, ISBN 0 19 852284 3

Dit boek beschrijft statistische modellen voor de analyse van herhaalde waarnemingen. De schrijvers zijn alle drie zeer bekend door hun werk in longitudinale gegevens. Diggle (1990) heeft eerder al een boek over het gebruik van tijdreeksen in de biostatistiek geschreven. Liang en Zeger hebben baanbrekend werk verricht met de door hen geïntroduceerde generalized estimation equations (GEE).

Het boek leest prettig doordat de behandelde theorieën steeds geïllustreerd worden met voorbeelden. Naar mijn mening is een uitstekend tekstboek voor onderwijs aan (statistiek) studenten en voor hen die meer van de analyse van herhaalde gegevens wil weten. Erg nuttig vond ik de hoofdstukken over grafische weergave van herhaalde gegevens, en de aandacht die er gegeven wordt aan de problemen die optreden bij het modelleren met logistische regressie en andere gegeneraliseerde lineaire modellen. De nadruk ligt steeds op modelbouw en daarom worden (M)ANOVA-technieken slechts beknopt besproken.

Een nadeel is dat er geen opgaven zijn. Wel staan er bij elk hoofdstuk verwijzingen naar extra literatuur staan die men zou kunnen lezen. Wat mij tegen viel, gezien de achtergrond van de schrijvers, was dat er slechts drie bladzijden gewijd werden aan GEE.

De indeling is niet altijd overzichtelijk en dat maakt het soms moeilijk om snel dingen op te zoeken. Sommige onderwerpen staan verspreid over meerdere hoofdstukken. Zo gaat bijvoorbeeld hoofdstuk 4 over het modelleren van de gemiddelde respons met lineaire regressie modellen. Hoofdstuk 5 gaat over het modelleren van de covariantiestructuur maar bevat ook een paragraaf over het niet parametrisch modelleren van de responsfunctie. Deze paragraaf had ik eerder in hoofdstuk 4 verwacht.

Een ander nadeel vind ik dat het moeilijk is om veel van de besproken theorieën toe te passen op eigen gegevens. In het boek wordt geen statistische software besproken. Dat is aan de ene kant wel begrijpelijk omdat dit soort informatie snel veroudert. Maar ik kwam verscheidene malen methoden tegen die heel geschikt zijn voor gegevens die ik op dit moment aan het analyseren ben en ik vond het vervelend dat informatie over de beschikbare software ontbrak.

Een korte inhoudsbeschrijving:

Hoofdstuk 1 is een inleiding. De verschillende datasets die in het boek gebruikt worden staan beschreven. Er wordt ingegaan op de voordelen van longitudinaal onderzoek. De auteurs beschrijven het verschil tussen marginale, random effect modellen en transitie modellen. Ook maken zij onderscheid tussen problemen waarbij men voornamelijk geïnteresseerd is in de regressie van Y op X , en problemen waarbij de correlatie tussen de verschillende

observaties het belangrijkste interessepunt is.

Hoofdstuk 2 gaat over design overwegingen. De resultaten van cross-sectionele studies worden vergeleken met resultaten van longitudinale studies. Ook is er een paragraaf over steekproefgrootte berekeningen.

Hoofdstuk 3. Verschillende manieren om herhaalde gegevens grafisch weer te geven worden op een rijtje gezet. Er wordt gekeken hoe men in grafieken cross-sectionele en longitudinale effecten kan onderscheiden en methoden voor het grafisch weergeven van de correlatiestructuur (variogrammen) worden besproken.

In Hoofdstuk 4 gaat men dieper in op lineaire regressie modellen voor longitudinale gegevens. In dit hoofdstuk ligt de nadruk op het schatten van de regressieparameters. Het model dient zo opgeschreven te worden dat cross-sectionele effecten en longitudinale effecten gescheiden worden. Schattingsmethoden als gewogen kleinste kwadraten, maximum likelihood en REML, en het gebruik van de robuuste variantie schatter worden bekeken.

Hoofdstuk 5 gaat verder met lineaire regressie modellen. Nu wordt er gekeken naar verschillende parametrische modellen voor de covariantiestructuur. Het aanpassen en evalueren van modellen wordt besproken. Ook wordt gekeken naar niet parametrische modellering van de gemiddelde respons.

Hoofdstuk 6 beschrijft de klassieke ANOVA technieken met hun limitaties.

In hoofdstuk 7 tot en met 10 komen uitbreidingen naar gegeneralizeerde lineaire modellen aan bod.

Aandacht daarbij ligt grotendeels op de analyse van binaire uitkomsten (logistische regressie) en tellingen (Poisson regressie). Er bestaan drie verschillende type modellen om dit soort gegevens te analyseren: marginale modellen, random effect modellen en transitie modellen. In logistische regressie levert elk van deze modellen andere regressie coëfficiënten met een andere interpretatie op. Dit in tegenstelling tot lineaire regressie. In hoofdstuk 7 worden deze verschillen besproken. Hoofdstuk 8, 9 en 10 beschrijven de details van de drie aanpakken.

Hoofdstuk 11 tenslotte gaat in op het probleem van ontbrekende waarnemingen, waarbij de notaties en aanpak van Little en Rubin (1987) gevolgd worden. Belangrijkste boodschap in dit hoofdstuk is dat wanneer het ontbreken van gegevens afhangt van de gemeten uitkomst, dit grote problemen kan opleveren en dat er op het gebied van missende gegevens in longitudinale studies meer onderzoek nodig is voordat men adequate met ontbrekende gegevens kan omgaan.

Conclusie: een prettig boek om te lezen waarin veel recente technieken voor herhaalde waarnemingen beschreven staan. De hoofdstukindeling vond ik soms wat onlogisch en nergens wordt de software die nodig is om de methoden toe te passen beschreven.

Referenties:

- Little, R.J.A. en Rubin, D.B. (1987). *Statistical analysis with missing data*. John Wiley, New York.
- Diggle, P.J. (1990). *Times series: a biostatistical approach*. Oxford University Press, Oxford.

Saskia le Cessie,
Afdeling Medische Statistiek,
Rijksuniversiteit Leiden

U. Engel en J. Reinecke.

Panelanalyse: Grundlagen, Techniken, Beispiele.

Walter de Gruyter, Berlin, 1994, ix + 327 pag., ISBN 3-11-013570-1, prijs onbekend.

Bij het schrijven van een inleiding in statistische onderzoeksmethoden voor een niet-specialistisch publiek (zoals studenten) kunnen twee strategieën gevolgd worden. Enerzijds kan men proberen de lezer een overzicht te geven van wat er allemaal op een bepaald terrein te koop is. Deze strategie leidt vaak tot een brede en oppervlakkige inleiding, die veel auteurs -- immers specialist op hun gebied -- onbevredigend zullen vinden. Anderzijds kan men ook een keuze maken uit de mogelijke onderwerpen, wat de auteur in staat stelt om complexere (en daardoor interessantere) zaken te behandelen. Valkuil bij deze strategie is natuurlijk dat de keuze van de onderwerpen moet aansluiten bij wat de lezers interessant en relevant zullen vinden; het stokpaardje van de auteur wordt door anderen wellicht minder graag bereden.

In beide gevallen moet men echter twee eisen stellen aan een dergelijke inleiding. Ten eerste, bij de onderwerpkeuze zal men zich moeten laten leiden door het belang van de onderwerpen. Interessante, maar verder irrelevante zaken horen niet in een inleiding thuis. Lezers moeten het gevoel hebben iets te kunnen met het gebodene; het moet aansluiten bij het type vragen dat men aan de gegevens zou willen stellen (en dat beperkt het aantal mogelijke onderwerpen; statistisch interessante, maar praktisch onbruikbare onderwerpen moeten worden vermeden). Ten tweede, men moet de presentatie van het gebodene aanpassen aan de aard van het publiek. De dingen moeten zo toegankelijk mogelijk worden gepresenteerd: alleen dan mag men verwachten dat de lezers zich uitgenodigd voelen om de besproken technieken ook eens op de eigen gegevens toe te passen.

Engel en Reinecke presenteren met hun "Panelanalyse: Grundlagen, Techniken, Beispiele" een inleiding in discrete-tijd panel-analyse, die bestemd is voor hogerejaars studenten en promovendi in de sociale wetenschappen. Ze gaan daarbij uit van de tweede hierboven geschetste benadering. In een korte inleiding worden de voors en tegens van panel-onderzoek besproken, waarna wordt doorgestoomd naar meer geavanceerde onderwerpen. Het boek is gestructureerd op basis van twee problemen die kunnen spelen in longitudinaal onderzoek; (1) variabelen zijn vaak onbetrouwbaar gemeten; (2) men heeft vaak te maken met missende waarden. Beide problemen komen natuurlijk ook in ander onderzoek voor.

Hoofdstuk 2-4 bespreken het gebruik van latente variabele-modellen in de context van longitudinaal onderzoek. Hoofdstuk 2 bespreekt de analyse van continue variabelen (LISREL), terwijl hoofdstuk 3 en 4 de analyse van nominale variabelen behandelen (loglineaire en latente-klasse analyse). Gewoonlijk wordt eerst de techniek in het algemeen besproken, waarna enkele specifiek longitudinale toepassingen worden gegeven. De meerwaarde van deze hoofdstukken ten opzichte van inleidingen in de betreffende technieken is, dat in de loop van elk hoofdstuk ook

enkele voor longitudinale analyse relevante zaken worden besproken. In hoofdstuk 3 wordt zo niet alleen een uitgebreide inleiding in LISREL geboden, maar komen ook cross-lagged panel analyse en het gebruik van verschillscores aan de orde; in hoofdstuk 3 en 4 wordt ook ingegaan op Markov-modellen, enzovoort.

In hoofdstuk 5 wordt tenslotte ingegaan op het probleem van missende waarden in panel-onderzoek. Dat blijkt niet zozeer het probleem van selectiviteit te zijn, maar vooral dat er zo weinig observaties overblijven. De veel toegepaste strategie van het wege van data wordt in drie zinnen van tafel geveegd, en de rest van het hoofdstuk is vooral gewijd aan het imputeren van missende waarden, waarbij geleund wordt op Little & Rubin, en Vermunt.

Positieve punten van het boek zijn, dat (a) er überhaupt een poging wordt gewaagd om een inleiding in dit zo complexe en belangrijke onderzoeksterrein te geven, voor gebruik tijdens een cursus; (b) er vrij veel voorbeelden (soms vergezeld van de betreffende opdrachtjobs) worden gegeven; (c) de gebruikte technieken uitgebreid en grondig worden toegelicht en besproken; en (d) ook de analyse van missende waarden wordt besproken.

Helaas heeft het boek ook zwakke kanten, die ervoor zorgen dat het niet voldoet voor de beoogde doelgroep. Het boek is (op de inleiding na) ontoegankelijk door een enorme hoeveelheid (soms paginalange) formules, die meer praktisch ingestelde lezers al snel zullen ontmoedigen. Ook de 300 voetnoten (die soms zaken uitleggen die voor een goed begrip essentieel zijn) bevorderen de toegankelijkheid niet. Hinderlijk is dat gerefereerd wordt aan variabelen als "V34" en "S14", wat afbreuk doet aan de begrijpelijkheid van de voorbeelden. Ik heb niet de indruk dat men zich veel aan het beoogde publiek gelegen heeft laten liggen.

Verder ontbreekt een aantal onderwerpen/technieken waarin lezers geïnteresseerd zullen zijn (en waar men soms iets van moet weten om zinvol longitudinale gegevens te kunnen analyseren): repeated-measures ANOVA, gelijkheid van factorstructuren, discrete-tijd survival-analyse, multi-level analyse, enzovoort, maar dat heeft natuurlijk te maken met de beperkte ruimte. De keuze van de onderwerpen lijkt helaas niet zozeer op basis van praktische relevantie te hebben plaatsgevonden: er wordt bijvoorbeeld veel aandacht gevraagd voor de analyse van latente-klasse Markov-modellen, maar dat deel had ik graag ingeruild voor een bespreking van -- toch! -- het wege van steekproeven en het probleem van non-random uitval. Een trefwoorden/auteursregister ontbreekt, wat de waarde als naslagwerk beperkt.

Samenvattend: voor gebruik door studenten is het ongeschikt (ook na vertaling); geïnteresseerde vakgenoten zullen wellicht iets hebben aan de hoofdstukken waarin de analyse van nominale variabelen wordt besproken.

Toon Taris, Peter Kerkhof, Karin van Rijn, Marga de Weerd
Vakgroep Sociale Psychologie
Vrije Universiteit Amsterdam,

P.R. Freeman en A.F.M. Smith (Editors)

Aspects of Uncertainty, A tribute to D.V. Lindley

Wiley, Chichester, 1994, xviii + 392 pag., ISBN 0-471-94347-9, £ 45.00.

Op 25 juli 1993 vierde Dennis (Victor) Lindley zijn 70-ste verjaardag, en in plaats van een bescheiden *liber amicorum* is ter zijner ere een verzameling artikelen gebundeld van voorname vrienden en/of vakgenoten, en uitgegeven als prachtig boek in de *Wiley series in probability and mathematical statistics*. Het boek bestaat uit 22 hoofdstukken: 21 artikelen van wiskundige aard, en een alleraardigst eerste hoofdstuk, 'Dennis Lindley: the first 70 years' door P. Armitage, waarin we Dennis Lindley wat beter leren kennen.

Ondanks Dennis's vroege ambitie om architect te worden kwam hij toch in de wiskunde terecht, met een groot gedeelte van zijn opleiding gedurende de oorlogsjaren. Aan het eind van de oorlog werkte Dennis, zoals de meeste wiskundigen in Engeland, in dienst van de overheid, en daar werd hij aan het werk gezet op het gebied van de statistiek. Hij was een van de eersten die het voorrecht hadden te kunnen werken met een computer (ACE), die geconstrueerd was door Alan Turing. Gedurende zijn laatste jaar in Cambridge bevond hij zich tussen bekende statistici als Bartlett, Wishart, Jeffreys en Fisher. Tussen laatstgenoemde en Dennis Lindley zou later, gedurende de vijftiger jaren, wrijving ontstaan naar aanleiding van Dennis's kritiek op Fisher's gebruik van fiduciele verdelingen. Toen al kon Dennis Lindley als Bayesiaan beschouwd worden, en het zijn vooral het werk van Jeffreys en een bezoek aan Savage (midden jaren '50) geweest die Dennis overtuigden van de juistheid van de Bayesiaanse theorie. Daarvoor had hij zich ook nog een poosje bezig gehouden met wachtrijtheorie, leidend tot een befaamd artikel en 'Lindley's vergelijking'. Sedertdien heeft hij zijn aandacht continu gericht op Bayesiaanse theorie, en dit heeft naast talloze boeiende artikelen, 'Lindley's informatie maat' en 'Lindley's paradox', geleid tot enkele welbekende boeken: 'Introduction to Probability and Statistics (From a Bayesian Viewpoint)' (2 vol., 1965), 'Making Decisions' (1971, 2e ed. 1985) en 'Bayesian Statistics: A Review' (1971). Vooral het tweede genoemde werk is een indrukwekkend en duidelijk boek dat ook nu nog niets van zijn kracht heeft verloren.

Op 54-jarige ging Lindley plots met pensioen, maar hij bleef wetenschappelijk actief en reisde zowat de hele wereld rond. Tot op de dag van vandaag verschijnen er interessante artikelen van zijn hand, en zijn visie komt ook sterk naar voren in vele contributies aan discussies van artikelen. Verder is zijn invloed zeer sterk geweest op het werk van zijn vroegere studenten, die nu veelal tot de meest bekende en gevierde Bayesiaanse statistici behoren (b.v. Adrian Smith en José Bernardo), en die zelf ook weer hun kennis en visie verspreid hebben (bv. aan O'Hagan en Goldstein), wat er toe geleid heeft dat zo ongeveer alle bekende Bayesiaanse statistici die hun opleiding in Engeland hadden, beschouwd kunnen worden als wetenschappelijke afstammelingen van Lindley. Voor degenen die nog nooit het genoegen hadden Dennis persoonlijk te ontmoeten zijn er 5 mooie foto's (waarvan 4 in kleur) afgedrukt, waarop we naast Dennis ook zijn vrouw, familie en politieke voorkeur leren kennen.

En alsof dit alles nog niet genoeg is, zijn er dan ook nog de 21 artikelen. Zonder uitzondering zijn ze boeiend, hoewel sommige artikelen weinig nieuws bevatten voor degenen die het werk van de auteurs goed kennen. Doordat er wellicht minder druk is geweest de lengte van de artikelen tot een minimum te beperken, lezen ze over het algemeen prettig en geven een goed idee van recente ontwikkelingen op het gebied van de Bayesiaanse statistiek (enkele uitzonderingen hebben dit niet als onderwerp). De auteurs en titels zijn (met korte commentaren om mijn mening weer te geven):

- R.E. Barlow en M.B. Mendel, 'The operational Bayesian approach'. Interessant artikel met nadruk op waarneembare variabelen in plaats van parameters, overeenkomend met andere recente artikelen van de auteurs.
- G.A. Barnard, 'Pivotal inference illustrated on the Darwin maïze data'. Dit moet een poging zijn om Lindley te overtuigen van het nut van een niet-Bayesiaanse methode.
- D.J. Bartholomew, 'Bayes' theorem in latent variable modelling'. Niet echt Bayesiaans, wel gebruik makend van Bayes' theorem.
- J.A. Bather en P.J. Browne, 'Bayesian sampling schemes for auditors'. Een eenvoudig model voor auditing, generalisatie eerder werk zonder echte toepassing.
- J.M. Bernardo, 'Optimizing prediction with hierarchical models: Bayesian clustering'. Een aardige toepassing waaruit blijkt dat ook in Mexico verkiezingsuitslagen beter achteraf dan vooraf te voorspellen zijn.
- P.J. Brown, N.D. Le en J.V. Zidek, 'Inference for a covariance matrix'. Weer eens een artikel waarin een nieuwe prior voorgesteld wordt en de bijbehorende wiskunde netjes beschreven staat. Van echte inference is geen sprake.
- R.V. Brown, 'The role of statistical theory in decision aiding: measuring decision effectiveness in the light of the outcomes'. Het evalueren van beslissingen is zonder meer belangrijk, maar het blijft onduidelijk wat we ervan leren voor het nemen van nieuwe beslissingen.
- S. Campodónico en N.D. Singpurwalla, 'The signature as a covariate in reliability and biometry'. Introduceert een Bayesiaanse methode om het power spectrum als een covariate mee te nemen in survival analysis. Indrukwekkende wiskunde en voorbeelden.
- K. Chaloner, 'Residual analysis and outliers in Bayesian hierarchical models'. Merkwaardig gebeuren waarin een subjectieve kansverdeling wordt gegeven die weergeeft of een waarneeming al dan niet een outlier is. Mij lijkt dat we hier voorzichtig mee moeten zijn.
- A.P. Dawid, 'The island problem: coherent use of identification evidence'. Een zeer aardige nauwkeurige analyse van het eiland probleem ($N+1$ inwoners van wie 1 een moordenaar moet zijn, en enkele karakteristieken zijn bekend). Dit is een belangrijk praktisch vraagstuk voor de rechtspraak, een gebied waar ook Lindley's interesse al geruime tijd naar uit gaat.
- S. French, 'Utility: probability's younger twin?'. Weinig schokkend stukje over beslissingstheorie. Het boek 'Decision theory: An Introduction to the Mathematics of Rationality' (1986) van dezelfde auteur is veel boeiender.
- E.I. George, U.E. Makov en A.F.M. Smith, 'Fully Bayesian hierarchical analysis for exponential families via Monte Carlo computation'. Sluit naadloos aan bij veel recent werk waarbij een voorbeeld besproken wordt met het rekenwerk via een Monte Carlo methode.
- M. Goldstein, 'Revising exchangeable beliefs: subjectivist foundations for the inductive argument'.

Leest U dit artikel minimaal 3 keer. Ik moet oppassen met mijn oordeel omdat ik in deze niet neutraal ben, maar ik kan U zeggen dat deze prachtige theorie daadwerkelijk gebruikt kan worden om enorm omvangrijke praktische problemen aan te pakken (hierover zullen 2 artikelen van Goldstein e.a. verschijnen in 'Bayesian Statistics 5'). Voor mij het absolute hoogtepunt, voor Dennis Lindley wellicht iets minder omdat het aansluit bij het meer subjectieve werk van De Finetti en als zodanig buiten het standaard Bayesiaanse kader valt. (Voor een reactie van Lindley op gerelateerd werk leze U zijn stukje getiteld 'Simplicity' in RSS News, April 1995.)

- B.M. Hill, 'On Steinian shrinkage estimators: the finite/infinite problem and formalism in probability and statistics'. Als zoveel werk van Hill ook nu weer een bijzonder boeiend fundamenteel artikel.
- J.B. Kadane, 'Bayesian decision theory and the legal structure'. Het gebruik van nutsfuncties in de rechtspraak. Het lijkt zo mooi en logisch, maar of het ooit een aanvaardbare praktijk zou kunnen zijn?
- F. Lad en J. Deely, 'Experimental design from a subjective utilitarian viewpoint'. Dit is heel mooi, en ten eerste aan te bevelen voor een ieder die geïnteresseerd is in proefopzetten en eens wat verder wil kijken dan Fisher's werk. De vraag blijft hoe de ideeën gebruikt kunnen worden voor grote problemen.
- T. Leonard en J.S.J. Hsu, 'The Bayesian analysis of categorical data - a selective review'. Een nuttig literatuur overzicht, beginnend bij een fundamenteel artikel van Lindley uit 1964.
- A. O'Hagan en H. Le, 'Conflicting information and a class of bivariate heavy-tailed distributions'. Interessant, maar er lijkt weinig rechtvaardiging te zijn voor deze 'robuuste' modellen. Wellicht zijn er andere (betere?) mogelijkheden om met tegenstrijdige informatie om te gaan.
- G. Parmigiani en D.A. Berry, 'Applications of Lindley information measure to the design of clinical experiments'. Aardig, maar niet bijzonder, stukje over Lindley's informatie maat en clinical trials.
- W.L. Smith, 'On two classic theorems involving the characteristic function'. Nieuwe bewijzen voor Lindberg's central limit theorem en Cramér's beroemde theorem met betrekking tot de normale verdeling. Niets Bayesiaans, maar zeer aardige wiskunde, geïnspireerd door Lindley's vroege (begin jaren '50) onderwijs verrichtingen op het gebied van de kanstheorie.
- M. West, P. Müller en M.D. Escobar, 'Hierarchical priors and mixture models, with application in regression and density estimation'. Dit is een toonbeeld van modern Bayesiaans werk, met alles er op en er aan, vooral indrukwekkend rekenwerk (Dirichlet mixture models en Gibbs sampling) en mooie plaatjes.

Een probleem met zo'n uitgave is dat het eigenlijk natuurlijk geen boek is, maar het is een waardevolle collectie artikelen voor in de wiskunde bibliotheek. Er zijn tijdschriften die in een volledig jaar niet zoveel aardige werken bij elkaar publiceren, dus de prijs valt ook wel mee. Het is alleen jammer dat de meester zelf de artikelen niet bespreekt, gezien de aard van zijn bijdragen aan discussie artikelen zou dat leuk geweest zijn. In het eerste hoofdstuk lezen we nog dat Dennis Lindley, zich aansluitend bij De Finetti, ervan uitgaat dat in het jaar 2020 de gehele statistische

wereld bekeerd zal zijn tot de Bayesiaanse methoden. Hoezeer ik hier ook aan twijfel, het zou een aardig vriendenboek ter gelegenheid van zijn honderdste verjaardag op kunnen leveren.

F.P.A. Coolen

Department of Mathematical Sciences,
University of Durham, England

N.L. Johnson, S. Kotz en A.W. Kemp

Univariate Discrete Distributions

John Wiley and Sons, New York, 1992, 2e ed., xx + 565 pag., ISBN 0-471-54897-7

De boeken van Johnson en Kotz (1969, 1970, 1972) over statistische verdelingen zijn duidelijke en uitgebreide naslagwerken. De boeken bevatten zeer veel informatie: veel resultaten en veel formules worden gegeven zodat gemakkelijk het gewenste resultaat, met de bijbehorende referentie, gevonden kan worden. Gedetailleerde bewijzen laten ze achterwege. Dit boek doet dat, zoals de titel al aangeeft, op het gebied van univariate discrete distributies.

Dit is de tweede editie van het boek; de reden hiervoor is, volgens de auteurs, dat er in dit vakgebied **veel** ontwikkelingen geweest zijn en er sinds 1984 (*Dictionary and Classified Bibliography of Statistical Distributions in Scientific work*) geen uitgebreide overzichten meer verschenen zijn. De grote vooruitgang in de computertechnologie heeft een belangrijke rol gehad in de ontwikkeling van dit vakgebied. Enerzijds werden approximaties steeds minder belangrijk en anderzijds kwam er de mogelijkheid om snel veel random getallen te genereren. Dit is dan ook een belangrijk verschil tussen de twee edities: minder aandacht voor approximaties en (veel) meer aandacht voor het genereren van random getallen. Andere interessante verschillen met de eerdere uitgave zijn:

- 1) er wordt nu alleen aandacht besteed aan univariate verdelingen (in de eerste editie stond nog een stuk over multivariate verdelingen);
- 2) er wordt meer aandacht besteed aan Bayesiaanse statistiek, met name met betrekking tot de binomiale en de Poisson verdeling;
- 3) er is een nieuw hoofdstuk bijgekomen (hoofdstuk 11) dat allerlei verdelingen behandelt die niet passen in de overige hoofdstukken.

De opbouw van het boek is als volgt. In het eerste hoofdstuk wordt de notatie en regelmatig terugkerende wiskunde behandeld. Het is prettig dat ze duidelijk aangeven wanneer er in de literatuur verschillende notaties gebruikt worden voor hetzelfde begrip en wanneer er dezelfde notatie gebruikt wordt voor verschillende begrippen. Er wordt in dit hoofdstuk ook een introductie gegeven op het genereren van random getallen voor de diverse distributies.

In hoofdstuk twee worden families van discrete distributies besproken. In de daarop volgende

hoofdstukken worden de binomiale, Poisson, negatief binomiale, hypergeometrische en logaritmische verdelingen besproken (respectievelijk hoofdstuk 3 t/m 7). De hoofdstukken hebben alle een gelijke opbouw. Eerst worden de algemene definities gegeven en wordt de historie besproken. Daarna worden de momenten en de eigenschappen van de verdeling behandeld. Approximaties, tabellen en berekening van kansen komen daarna aan de orde. Tot slot worden in de overige paragrafen verschillende onderwerpen behandeld, zoals schatten, toepassingen, afgeknotte verdelingen en gerelateerde verdelingen. Bij de binomiale verdeling wordt de 'order statistic' nog behandeld.

Hoofdstuk 8 gaat over samengestelde verdelingen. In de overige drie hoofdstukken worden allerlei verdelingen behandeld, zoals 'generalized (stopped-sum) distributions' en 'matching, occupancy, and runs distributions'.

Samenvattend: het boek geeft zeer gedetailleerd de facetten van univariate distributies weer. Het boek is echt een naslagwerk: er wordt beknopt uitleg gegeven en het bevat zeer veel informatie. Bij elk van de onderwerpen geldt namelijk dat er veel resultaten, met bijbehorende referenties en zonder bewijs, worden gegeven.

Referenties:

Johnson N.L. en S. Kotz (1969), *Discrete Distributions*, John Wiley & Sons, New York

Johnson N.L. en S. Kotz (1970), *Continuous Univariate Distributions*, John Wiley & Sons, NY

Johnson N.L. en S. Kotz (1972), *Continuous Multivariate Distributions*, John Wiley & Sons, NY

Hiek van der Scheer,
Vagroep Econometrie,
Rijkuniversiteit Groningen

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

Hieronder volgt een overzicht van de recente uitgaven die sinds de vorige aflevering van *Kwantitatieve Methoden* bij de redactie zijn binnengekomen. Titels waarvoor het symbool π 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. Indien het boek dan nog niet vergeven is, krijgt de kandidaat het boek gratis thuisgezonden. Hiervoor dient uiteraard wel binnen een half jaar een recensie te worden teruggezonden. Van enkele titels, bijvoorbeeld de uitgaven van het ILO, zijn bij ons slechts aankondigingen binnengekomen. Indien voor een dergelijke titel belangstelling bestaat, dan kan een recensie-exemplaar worden opgevraagd. In dat geval zal de recensent zijn of haar boek iets later ontvangen.

- π **Per K. Andersen, Ornulf Borgan, Richard Gill en Niels Keiding**
Statistical models based on counting processes,
 Springer series in statistics, Springer Verlag, Berlin, 1995, 770 pp., ISBN 0-387-94519-9, DM 58.00

- π **H.S. Bear,**
A primer of Lebesgue integration
 Academic Press, London, 1995, xi + 163 pp., ISBN 0-12-083970-9

- π **M. Ishaaq Bhatti,**
Testing regression models based on sample survey data
 Avebury, Aldershot UK, 1995, xvi + 200 pp., ISBN 1-85628-642-8, £ 35.00

- π **Patrick Billingsley,**
Probability and measure (3d ed.),
 Wiley series in probability and mathematical statistics, John Wiley & Sons, New York, 1995, xii + 593 pp., ISBN 0-471-00710-2, £ 49.95

- π **A. Bronstein en C. Thomas**
European labour courts: international and European labour standards in labour courts
 ILO publications, Geneve, 1995, ISBN 92-2-108012-9

- π **D. Edwards,**
Introduction to graphical modelling,
 Springer texts in statistics, Springer Verlag, Berlin, 1995, 290 pp., ISBN 0-387-94483-4, DM 68.00
 (incl. diskette)

- π **R.A. Epstein,**
The theory of gambling and statistical logic,

Academic Press, London, 1995, xvi + 450 pp., ISBN 0-12-240761-X

▣ **Ronald N. Forthofer en Eun Sul Lee**

Introduction to biostatistics: a guide to design, analysis and discovery

Academic Press, San Diego, 1995, xvii + 567 pp., ISBN 0-12-262270-7

▣ **Peter Hall,**

The bootstrap and Edgeworth expansion,

Springer ser. in statistics, Springer Verl. Berlin, 1995, xiii+352 pp., ISBN 0-387-94508-3, DM 48.00

▣ **J. den Heeten en A.O. Verweij**

Identificatie en registratie van ethnische herkomst, een handleiding voor beleid

VNG Uitgeverij, Den Haag, 1995, 68 pp., ISBN 90-322-1121-8, f 16.50

▣ **T.P. Hutchinson,**

The common sense for a first course in statistics

Rumsby Scientific Publ. Sydney, 1995, ii + 41 pp., £ 2.00

▣ **ILO**

Women workers: an annotated bibliography '83-'94

ILO Publications, Geneve, 1995, ISBN 92-2-109201-1,

▣ **ILO**

World labour report,

ILO Publications, Geneve, 1995, ISBN 92-2-109447-2

▣ **Roger D. Leitch,**

Reliability analysis for engineers: an introduction

Oxford science publications, New York, 1995, xv + 230 pp., ISBN 0-19-856371-X, £ 18.95

▣ **Prem S. Mann,**

Introductory Statistics (2nd ed.)

John Wiley & Sons, New York, 1995, xxii + 800 pp., ISBN 0-471-31009-3, £ 19.95

▣ **E. Marubini en M. G. Valsecchi**

Analysing survival data from clinical trials and observational studies

John Wiley & Sons, Chichester, 1995, xvi + 414 pp., ISBN 0-471-93987-0, £ 39.95

▣ **Anne Trebilcock,**

European labour courts: remedies and sanctions in industrial action

ILO Publications, Geneve, 1995, ISBN 92-2-109185-6

Oproep voor boekbespekers

De redactie van Kwantitatieve Methoden is steeds op zoek naar mensen die bereid zijn een pas verschenen boek te lezen met het doel daarover een oordeel te geven. Zo'n recensie verschaft de leden van de VVS en overige lezers van Kwantitatieve Methoden inzicht in de kwaliteiten van het boek en verschaft nuttige informatie bij een overwogen aanschaf. Voor de schrijvers biedt een kritisch oordeel van het publiek waarvoor het boek geschreven een mogelijkheid tot verdere verbetering. De uitgever heeft uiteraard meer baat bij een lovende recensie. Als tegenprestatie mogen de recensenten de besproken boeken behouden.

Behalve de algemene boeken over statistiek of besliskunde, waar veel mensen zinnige dingen over kunnen zeggen, behandelen veel boeken een specialistisch onderwerp. Als wij, soms op de gok, iemand een dergelijk boek toesturen, blijkt een enkele keer dat 'het boek precies aansluit bij het onderzoek' van de recensent. 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 succeskans van het toesturen van een boek te vergroten, willen wij graag in contact komen met zoveel mogelijk collega statistici, kansrekenaars en besliskundigen. 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 in de toekomst wellicht een gloednieuw boek op uw onderzoeksgebied binnenkomt hebben wij een grotere kans op een perfecte *match*.

Ik ben graag bereid een boek op mijn interessegebied te bespreken voor de leden van de VVS

naam

adres

telefoon/e-mail

interessegebieden

(nauwkeurig

omschrijven)

