Review

Reviewed Work(s): Statistical Applications Using Fuzzy Sets by Kenneth G. Manton, Max A. Woodbury and H. Dennis Tolley

Review by: Shelby J. Haberman

Source: *Journal of the American Statistical Association*, Vol. 90, No. 431 (Sep., 1995), pp. 1131-1133

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: http://www.jstor.org/stable/2291362

Accessed: 02-12-2017 18:40 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

*American Statistical Association, Taylor & Francis, Ltd.* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

This content downloaded from 128.59.222.107 on Sat, 02 Dec 2017 18:40:29 UTC
All use subject to http://about.jstor.org/terms

selection and ranking theory, a preliminary study of the work of Gibbons, Olkin, and Sobel (1977) would be highly desirable.

Pinyuen CHEN
*Syracuse University*

REFERENCES

Bechhofer, R. E. (1954), "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations With Unknown Variances," *The Annals of Mathematical Statistics,* 25, 16–39.

Bechhofer, R. E., and Sobel, M. (1954), "A Single-Sample Multiple-Decision Procedure for Ranking Variances of Normal Populations," *The Annals of Mathematical Statistics,* 25, 273–289.

Gibbons, J. D., Olkin, I., and Sobel, M. (1977), *Selecting and Ordering Populations: A New Statistical Methodology,* New York: John Wiley.

Gupta, S. S. (1956), "On a Decision Rule for a Problem in Ranking Means," Ph.D. dissertation, University of North Carolina, Chapel Hill.

Gupta, S. S., and Panchapakesan, S. (1979), *Multiple Decision Procedures,* New York: John Wiley.

## Statistical Models for Ordinal Variables.

Clifford C. CLOGG and Edward S. SHIHADEH. Thousand Oaks, CA: Sage Publications, 1994. xiii + 192 pp. $32.95.

Ordinal variables are ubiquitous in the social sciences. They are used whenever there is an intention to quantify concepts for which the ideas of more or less, greater or smaller, stronger or weaker, and the like are relevant, but no measuring instrument of the kind found in the physical sciences is available. Usually, the number of ordered categories in a particular unit of response is on the order of 4 or 5, and in elementary analyses, the analogy to physical measurement is taken literally, with the successive categories simply assigned successive integers. To overcome the doubtful assumptions of equal units, equal discrimination at the category boundaries, and a nonfinite number of categories implied in these elementary analyses, statistical models have been formalized for the classification of entities. In the last two-and-a-half decades or so, these formalizations have burgeoned, in part because of the development of log-linear models for discrete data and in part because computing associated with model fitting is no longer a problem.

*Statistical Models for Ordinal Variables* adds to the literature in this important field, focusing on the applications of relatively recent developments in model construction and emphasizing the analysis of the relationships among two ordered discrete variables. The book is one in a series on advanced quantitative techniques in the social sciences, and it fits this classification perfectly. As stated by the series editor, the book synthesizes two approaches to the study of association: Pearson's approach to extending correlations to discrete variables and Yule's direct approach to properties of association in contingency tables. The level of statistics required is not trivial, but neither is it prohibitive; a general degree that emphasizes mathematics, an interest in the social sciences, and a good collaborative study group would serve excellently.

There are seven chapters in this book of almost 200 pages. The first chapter sets the framework for the subsequent chapters with a thorough commentary on the $2 \times 2$ contingency table, and the last chapter deals with the regression of ordinal, discrete dependent variables on explanatory variables. The device used for synthesizing the material, at present scattered in various statistical and social science journals, is a series of analyses of a manageable number of real data sets using the models that are developed. Following the successively more sophisticated analysis of these data sets is an excellent pedagogical approach; it shows the relationships among models and ensures that each data set does not seem to have its own idiosyncratic method of analysis.

Although this book was not written to be a textbook, its rich analyses and interpretation of the models means that it could be set readily as a text in a graduate course on quantitative methods for the social sciences. Although there are no formal problems or solutions, excellent exercises are suggested throughout the book. Completing these exercises, together with learning how to implement one or two standard software packages in order to confirm the statistical results presented, would be an excellent course of study. The book mentions existing software packages that could be used, but it does not—and I think quite correctly—go into the details of how to use them.

The mathematical material is written efficiently and clearly. A highlight of the book, though, is the fluent and deceptively elegant way in which the interpretations of the mathematical relationships are written. Although clear in context on the first reading, many sentences and paragraphs can be studied in more and more depth on successive readings. I illustrate only one such topic here, but there are many others. Very early in the second chapter, one finds "the unit point or scale of the scores thus affects the size of the interaction effect. . . . We shall find later that it is very important to be aware of unit points implicit in a set of scores" (p. 25). Later in the same chapter the point is made that "measurement should always be taken seriously because the inferences or summaries that we obtain are always dependent on how variables have been defined. When analyzing categorical data (or categorical variables), it is also the case that the number and kinds of categories used must be reckoned with" (p. 32). On first reading, these assertions seem relatively straightforward, but then in the final chapter the authors point out that "the adjacent category model is affected by collapsing, at least in the form given above, and so choice of categories should be taken seriously if this modeling approach is used" (p. 149; italics in original). Even the simplest of models that involves the log odds of adjacent categories does not permit the collapsing of frequencies in adjacent categories ($k + 1$) and $k$ if the data fit the model. On further reflection, these relationships seem related in some fundamental way, and indeed they are. In both cases the effect of the precision is paramount; in indices of association, the level of precision affects the level of association that can be detected, and the level of precision is incorporated in the model involving the log odds of adjacent categories.

Although collapsing categories is not permitted if the data fit the model, in certain situations the categories are not operating with the intended precision, and thus categories should be collapsed to reveal the actual, rather than the intended, precision of the ordered categories. Not only is care needed before collapsing categories, but checks are possible on whether or not the operational ordering of the categories is as intended. The authors give examples where they infer that the intended ordering has not been successful. The appreciation of these relationships is important for social scientists, not only for analyzing data after they are collected, but for structuring the categories for collecting data in the first place.

This set of related results gain even greater importance and pause for reflection when it is recognized that another popular model considered—that involving the log odds of complementary cumulative categories—has the property that adjacent categories can be collapsed. Understanding the implications of this major difference between these models—perhaps a surprising difference—is the kind of issue that can be pursued at both technical and philosophical levels from results in this book. For example, if the model involving the log odds of adjacent categories reflects precision of the data, does the model involving the log odds of complementary cumulative categories not reflect this precision? If so, for which kinds of data is each model relevant, and is it more than a matter of taste which of these two classes of models are used for any data set?

*Statistical Models for Ordinal Variables* sets up many issues of this kind for the diligent student to explore. In conjunction with the generally excellent quality of the presentation, this is why I recommend the book strongly for graduate courses involving quantitative methods for the social and related sciences.

David ANDRICH
*Murdoch University*

## Statistical Applications Using Fuzzy Sets.

Kenneth G. MANTON, Max A. WOODBURY, and H. Dennis TOLLEY. New York: John Wiley, 1994. xi + 312 pp. $59.95.

Manton, Woodbury, and Tolley (MWT) have worked since the 1970s to apply classification by grade of membership (GoM) to analysis of data in the health sciences. This classification approach is described and used in a substantial number of papers and book chapters in which one or more of these investigators serves as a coauthor. Generally, these works have been oriented more toward researchers in the health sciences than toward the audience of professional statisticians. Indeed, as far as I can ascertain, the authors do not have articles on classification by grade of membership in the main statistical journals of the United States and Great Britain. This book, part of the Wiley Series in Probability and Mathematical Statistics, thus provides the professional statistician with an introduction to a statistical technique that has been applied for some time but has not been heavily treated in the statistical literature.

To describe the classification technique used by MWT, it is appropriate to consider the basic GoM model examined in Chapters 1–3. Here $I$ individuals numbered from 1 to $I$ are sampled from a population, and for each individual $i$, $J$ polytomous responses are observed. Responses are numbered from 1 to $J$, and response $j$ has integer values from 1 to $M_j \geq 2$. Response $j$ of individual $i$ may be denoted by the random variable $X_{ij}$. MWT then postulate that there are $K \geq 2$ classes for each individual $i$, and that each individual can be regarded as having partial memberships in these classes. Thus each individual $i$ may be characterized by unknown nonnegative random GoM scores $g_{ik}$ for integers $k$ from 1 to $K$ such that $\sum_i g_{ik} = 1$. If for some integer $p$ from 1 to $K$, $g_{ip}$ is 1 and $g_{ik}$ is 0 for $k \neq p$, then individual $i$

This content downloaded from 128.59.222.107 on Sat, 02 Dec 2017 18:40:29 UTC
All use subject to http://about.jstor.org/terms

belongs exclusively to class $p$. On the other hand, if for distinct integers $p$ and $q$ from 1 to $K$, $g_{ip} = g_{iq} = \frac{1}{2}$, then individual $i$ is regarded as half characterized by class $p$ and half characterized by class $q$. For each individual $i$, the score vector $\mathbf{g}_i$ has coordinates $g_{ik}$ for $1 \le k \le K$. It is assumed that the $\mathbf{g}_i$ are iid. As in latent-class analysis, the local independence condition is then imposed that given the $\mathbf{g}_i$ for $1 \le i \le I$, the responses $X_{ij}$ are mutually independent. It is further assumed that for each response $j$ and class $k$ correspond probabilities $\lambda_{jkm}$, $1 \le m \le M_j$, so that each $\lambda_{jkm}$ is nonnegative and $\sum_m \lambda_{jkm} = 1$. Given $\mathbf{g}_i$, it is assumed that the probability is $\sum_k g_{ik}\lambda_{jkm}$ that response $X_{ij}$ is $m$.

The GoM model may be regarded as a generalization of a latent-class model, for one obtains a latent-class model with $K$ latent classes if the probability is 1 that $\max_k g_{ik} = 1$. On the other hand, the GoM model may be regarded as a special case of latent-class models studied by Goodman (1974a,b) and Haberman (1977a; 1979, ch. 10; 1988), among others, in which linear constraints are imposed on latent probabilities and on conditional probabilities of manifest variables given latent variables. For each individual $i$, consider $J$ latent variables $Z_{ij}$ with integer values from 1 to $K$. As is typically the case in latent-class analysis, let the latent vectors $\mathbf{Z}_i = (Z_{ij}: 1 \le j \le J)$ be mutually independent, and let the $X_{ij}$ be conditionally independent given the $\mathbf{Z}_i$. To yield the special latent-class model under study, for each individual $i$, let the $Z_{ij}$ be exchangeable random variables and let $X_{ij}$ and $Z_{ij}$ be conditionally independent given $Z_{ia}$, $a \ne j$. Thus if $\mathbf{K}$ is the set of vectors $(k(j): 1 \le j \le J)$ with $1 \le k(j) \le K$ for $1 \le j \le J$, $\mu(\mathbf{k})$ is the probability that $\mathbf{Z}_i = \mathbf{k}$ for $\mathbf{k}$ in $\mathbf{K}$, $P$ is the set of permutations of the integers 1 to $J$, $\pi$ is in $P$, the vector $\mathbf{k} = (k(j): 1 \le j \le J)$ is in $\mathbf{K}$, and $\mathbf{k}(\pi) = (k(\pi(j)): 1 \le j \le J)$, then $\mu(\mathbf{k}) = \mu(\mathbf{k}(\pi))$. In this fashion, there are no more than $f = (J + K - 1)!/[J!(K - 1)!]$ distinct values of $\mu(\mathbf{k})$ for $\mathbf{k}$ in $\mathbf{K}$ (Feller 1968, p. 38). Let the conditional probability be $\lambda_{jkm}$ that $X_{ij} = m$ given that $Z_j = k$. Let $1 \le m(j) \le M_j$ for $1 \le j \le J$, $\mathbf{m} = (m(j): 1 \le j \le J)$, $\lambda = (\lambda_{jkm}: 1 \le m \le M_j,\ 1 \le j \le J,\ 1 \le k \le K)$, and $\mu = (\mu(k): k \in K)$. Then the sum $p(m, \mu, \lambda)$ over $\mathbf{k}$ in $\mathbf{K}$ of the products $\mu(\mathbf{k}) \prod_j \lambda_{jkm(j)}$ is the observed probability that $\mathbf{X}_i = (X_{ij}: 1 \le j \le J)$.

The probability that $\mathbf{X}_i = \mathbf{m}$ is also $p(\mathbf{m}, \mu, \lambda)$ under a GoM model in which $\mu(\mathbf{k})$ is the expectation of $\prod_j g_{ik(j)}$ for $1 \le i \le I$. Note that $\mu = (\mu(\mathbf{k}): \mathbf{k} \in \mathbf{K})$ then satisfies the requirements for the distribution of $\mathbf{Z}_i$, for $\sum_{\mathbf{k} \in \mathbf{K}} \mu(\mathbf{k}) = 1$, $\mu(\mathbf{k}) \ge 0$ for $\mathbf{k}$ in $\mathbf{K}$, and $\mu(\mathbf{k}) = \mu(\mathbf{k}(\pi))$ whenever $\mathbf{k}$ is in $\mathbf{K}$ and $\pi$ is in $P$. The GoM model is slightly less general than the latent-class model with exchangeable variables due to constraints on the values that moments may assume. For example, if $K$ is 2 and $J$ is 2, then the constraint that $g_{i2} = 1 - g_{i1}$ leads via the Schwarz inequality to the constraint that $\mu((1, 2)) \le [\mu((1, 1))\mu((2, 2))]^{1/2}$. Following Karlin and Studden (1966, chap. 3), the set $Q$ of vectors $\mu$ compatible with the GoM model is nonempty, closed, and convex. The vector $(\prod_j x_{k(j)}: \mathbf{k} \in \mathbf{K})$ is in $Q$ if $x_k \ge 0$ for $\mathbf{k}$ in $\mathbf{K}$ and $\sum_k x_k = 1$. Given standard completeness results for multinomial distributions, one may show that $Q$ has a nonempty interior.

For practical use of the GoM model, parameters must be estimated and model validity must be assessed. If identifiability conditions are met and the latent-class model with exchangeable latent variables is valid, then (as in Goodman 1974a,b and Haberman 1977a; 1979, ch. 10; 1988) parameter estimation is readily accomplished by use of the likelihood function $L(\mu, \lambda) = \prod_i p(\mathbf{X}_i, \mu, \lambda)$, at least if the model holds and the sample size $I$ is large relative to the number $d = K \sum_j (M_j - 1) + f$ of independent parameters. One may use available numerical procedures to compute maximum likelihood estimates $\hat{\mu}$ for $\mu$ and $\hat{\lambda}$ for $\lambda$ that are consistent and asymptotically normal and to estimate asymptotic standard deviations of parameter estimates. Although identifiability conditions are relatively complicated, a necessary condition for identifiability is that $d$ not exceed $\prod_j M_j - 1$. For example, if there are five dichotomous responses, then $K$ must be 2 for identifiability to be possible. In estimation it is simplest to ignore the requirement of the GoM model that $\mu$ be in $Q$. If $\mu$ is in the interior of $Q$, then as $I$ becomes large, the probability approaches 1 that $\hat{\mu}$ is in $Q$. Given results concerning latent-class models with restrictions, it is also feasible to test model validity via residual analysis and chi-squared tests.

The use of latent-class analysis for the GoM model does have limitations in practice due to the rapid increase in the number $d$ of parameters encountered as $K$ increases. For example, consider $J = 15$, $K = 6$, and $M_j = 2$ for each $j$. In this case $f$ is 15,594. Given such an $f$, it is very difficult to achieve a sample size $I$ large enough so that asymptotic results provide appropriate approximations. Nonetheless, it is quite conceivable that large-sample approximations may be appropriate for realistic but large sample sizes if $J = 10$, $K = 3$, $M_j = 2$ for each $j$, and $f = 96$.

The suggested analysis of the GoM model considered by the reviewer is not the method of analysis actually adopted by MWT. The authors do contemplate an estimation procedure based on work of Kiefer and Wolfowitz (1956) that leads to maximization of $L(\mu, \lambda)$, but this approach is not actually used by the authors in their numerical work, and appropriate citations are not provided to the literature on restricted latent-class models. Instead, the authors use a conditional approach in which inferences are conditional on the $\mathbf{g}_i$ for $1 \le i \le I$. For $g = (\mathbf{g}_i: 1 \le i \le I)$, this conditional approach involves maximization of the conditional likelihood function

$$L_c(g, \lambda) = \prod_i \prod_j \left( \sum_k g_{ik}\lambda_{jkX(i,j)} \right).$$

Note that in this approach, the $\mathbf{g}_i$ are treated as unknown parameters for each individual $i$. The likelihood function $L_c$ is a special case of a likelihood function considered by Haberman (1977a) in the examination of product models derived by indirect observation. Functional iterations algorithms considered there may be used to maximize $L_c(g, \lambda)$. An alternative approach of Goodman (1974a,b) is readily adapted to the problem under study. The authors' approach is essentially that of Goodman (1974a,b).

Unfortunately, the large-sample theory discussed by Haberman (1977a) does not apply to the situation at hand because the number of $g_i$ under study becomes large as $I$ becomes large. This difficulty is a fundamental one. The underlying problem was discussed by Kiefer and Wolfowitz (1956) and goes back to work of Neyman and Scott (1948). It is to be expected in this class of problems that as the sample size $I$ increases, estimates of $\mathbf{g}_i$ do not approach $\mathbf{g}_i$ and estimates of $\lambda$ do not approach $\lambda$. As Haberman (1977b) noted, it is conceivable that consistency and asymptotic normality of estimates can be achieved for large-sample arguments in which both $I$ and $J$ become large and $K$ remains constant. But I must caution that rigorous demonstration of consistency, asymptotic normality, or inconsistency appears to be very difficult in this example. It should be further emphasized that even if some desirable properties of estimates can be found, use of chi-squared statistics for hypothesis tests involves even more difficulties. In practice, the authors consider use of their estimation approach with relatively large $I$ and with moderately large $J$; for instance, in Table 3.2, $I = 3{,}349$, $J = 33$, $M_j = 2$ for each $j$, and $K = 6$. Analyses based on the authors' approach needs much more justification than is provided. Such justification should involve both rigorous proofs of results and extensive simulations.

A further difficulty with the methodology developed in Chapters 1–3 is the lack of procedures for verification of model validity. Given the work involved in use of the GoM model and given the numerous questions considering validity of the methods used, it would at least be comforting to feel that the model was valid in some real cases. The authors tend to content themselves with verification that their model tells a plausible story; however, stories are not tests of validity. There is an attempt to compare results with latent-class analysis, but the comparison uses nonstandard methods of estimation in latent-class analysis. Once again, rigorous justification is needed both by proof and by simulation.

The methodology in Chapters 1–3 does not provide procedures for assessment of accuracy of parameter estimates. Neither large-sample theory nor resampling approaches are used. This lack of accuracy assessment is a serious weakness—it is important to know something concerning the reproducibility of results.

As evident in Chapters 4–8, numerous variations on the standard GoM model are available. In Chapter 4, the GoM model is adopted to Poisson rates. As is commonly true in generalizations from the multinomial to the Poisson case, the issues of estimation and model validity in the Poisson case are essentially those of the multinomial case. In Chapter 5, the GoM model is applied to longitudinal data. Further problems of model validity arise in this chapter due to problems in application of local independence with longitudinal data.

In Chapter 6, the GoM model is modified so that the probability that $X_{ij} = m$ is equal to $\sum_k g_{ik}\lambda_{jkm} / \sum_k g_{ik} \sum_p \lambda_{jkp}$ and the constraint that $\sum_m \lambda_{jkm} = 1$ is removed. Once again, the methodological issues are similar to those previously encountered. In Chapter 7 GoM techniques from Chapter 5 are applied to forecasting survival. In Chapter 8, the GoM model of Chapters 1 to 3 is modified to permit comparisons of groups.

I find it very difficult to recommend the book. The fundamental issue is that the methodology used has not been shown to be appropriate for the problem and the models used have not been shown to be appropriate for the data. These problems are not the only ones encountered. The authors often do not cite available literature. For example, the latent-class literature of Goodman (1974a,b) and Haberman (1977a; 1979, chap. 10; 1988) and the issues discussed by Neyman and Scott (1948) and Kiefer (1956) go unmentioned. Even when literature is cited, the reference is often incorrectly provided or the contents of the reference are misconstrued. For example, in the references for Chapter 3, the reference to a paper coauthored by Tolley and Manton does not cite the correct year. In Section 3.3.2, the authors cite Wu's (1983) discussion of the EM algorithm to provide evidence for the superiority of their algorithm over the EM algorithm, but they fail to show any reason for any difference in convergence properties if the EM algorithm is used in their problems rather than their rather similar cyclical

This content downloaded from 128.59.222.107 on Sat, 02 Dec 2017 18:40:29 UTC
All use subject to http://about.jstor.org/terms

algorithm. It is also difficult to see from the authors' discussion that Wu's paper implies that all limit points of the EM algorithm are stationary points in the problem at hand.

Formulas are often incorrect, although it is generally possible to reconstruct the correct expression after some effort. In addition, the book is very difficult to understand. The mathematical techniques used are often more complex than needed. For example, in the discussion of identifiability in Chapter 3, reference to standard results concerning moments and completeness would save much effort. The authors discuss latent-class analysis with an emphasis on assignment of sample members to latent class to facilitate estimation, even though such procedures are rarely used and are not well understood, either theoretically or numerically. The examples are interesting and clearly relevant to the authors' published empirical work; however, not enough detail is provided to permit readers to verify results themselves. It would be very helpful to present very simple examples so that methodology could be applied in a clear fashion and results studied. Including an author index also would be helpful.

*Statistical Applications Using Fuzzy Sets* is not readily useful as a text, given its lack of exercises and the difficulty of the writing. A further problem is that the reader should be quite familiar with mathematical statistics, mathematical analysis, and linear algebra. Given the lack of rigor, such requirements are somewhat unreasonable.

The professional statistician may find this book useful if it is necessary to know what GoM methodology is about. It is to be hoped that future research by the authors will result in a rigorous evaluation of methodology for classification by grade of membership.

Shelby J. HABERMAN
*Northwestern University*

REFERENCES

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, Vol. I, New York: John Wiley.

Goodman, L. A. (1974a), "The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. I. A Modified Latent Structure Approach," *American Journal of Sociology*, 79, 1179–1259.

——— (1974b), "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models," *Biometrika*, 61, 215–231.

Haberman, S. J. (1977a), "Product Models for Frequency Tables Involving Indirect Observation," *The Annals of Statistics*, 5, 1124–1147.

——— (1977b), "Maximum Likelihood Estimates in Exponential Response Models," *The Annals of Statistics*, 5, 815–841.

——— (1979), *Analysis of Qualitative Data. II. New Developments*, New York: Academic Press.

——— (1988), "A Stabilized Newton-Raphson Algorithm for Log-Linear Models for Frequency Tables Derived by Indirect Observation," *Sociological Methodology*, 18, 193–211.

Karlin, S., and Studden, W. (1966), *Tchebycheff Systems, With Applications to Analysis and Statistics*, New York: Wiley Interscience.

Kiefer, J., and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Nuisance Parameters," *The Annals of Mathematical Statistics*, 27, 887–906.

Neyman, J., and Scott, E. L. (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

Wu, C. F. J. (1983), "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, 11, 95–103.

## Networks and Chaos: Statistical and Probabilistic Aspects.

O. E. BARNDORFF-NIELSEN, J. L. JENSEN, and W. S. KENDALL. London: Chapman and Hall, 1993. xvii + 307 pp. $59.95.

In 1992 a tutorial, "Séminaire Européen de Statistique," was held at Sandbjerg/Aarhus University. This volume contains revised versions of seven of the papers presented there. Four of these are major reviews; two each on neural networks and chaos. The book provides an excellent approach to these subjects. The other three chapters are shorter and more specialized surveys of topics to do with networks.

The first paper, 39 densely-packed pages by Shun-Ichi Amari, reviews "Mathematical methods of neurocomputing." More than half of the references are to Amari's own work; this is appropriate, as Amari has been very influential in developing this field. The main topics are statistical neurodynamics, which studies dynamical aspects of neural networks, a new geometrical theory that treats ensembles of neural networks, and some basic theory of learning.

The second paper, 84 pages by B. D. Ripley, gives a magnificent survey of "Statistical aspects of neural networks." This chapter is alone easily worth the price of the book and will surely become the standard introductory reference. Ripley reviews the history of the subject, describes applications, and compares these methods with ones using classical statistical methods. He points out that "in one sense neural networks are little more than nonlinear regression and allied optimization methods. However, they do have a methodology of their own . . . their success is a warning to statisticians who have worked in a simply structured linear world for too long."

The next two papers are "Statistical aspects of chaos: a review" by V. Isham and "Chaotic dynamical systems with a view towards statistics: a review" by J. L. Jensen. Isham's paper starts with a gentle introduction to the field and then surveys methods for estimating dimension (of which there are several kinds) and the more detailed structure of chaotic systems. Jensen's review surveys the mathematical theory of chaos in some detail. He says very little about statistics.

The remaining three papers are "A tutorial on queueing networks," by S. Asmussen, "River networks: a brief guide to the literature for statisticians and probabilists," by O. E. Barndorff-Nielsen, and "Random graphical networks," by G. Grimmett. These are all straightforward expositions of basic concepts.

I recommend the book, mainly because of Ripley's paper, to everyone who wants to find out what these fields are about. If future "Séminaires Européen" sustain the level attained here, they will be most welcome.

Colin MALLOWS
*AT&T Bell Labs*

## Laws of Small Numbers: Extremes and Rare Events.

Michael FALK, Jurg HÜSLER, and Rolf-Dieter REISS. Boston: Birkhäuser Boston, 1994. xi + 282 pp. $49.50.

This monograph is based on a series of lectures given by the authors—three well-known specialists on extreme value theory—for a DMV seminar (a winterschool of the German Mathematical Association). The book gives a short introduction to topics on classical extreme value theory but mainly emphasizes recent results and new directions in the field, in particular the authors' own research.

The title suggests that Poisson approximations to the binomial law are the basic tools. But Poisson approximation is mainly understood in a modern sense as approximation of point processes by Poisson random measures. This is what the authors call a "functional law of small numbers." Point process techniques are the bread and butter of contemporary extreme value theory; see, for example, the standard monographs by Leadbetter, Lindgren, and Rootzén (1983), Resnick (1987), and Reiss (1989). The reader should have some background on point processes (see Cox and Isham 1980, Daley and Vere-Jones 1988, Karr 1986, or Resnick 1987 for the necessary tools) and on classical extreme value theory. Thus the book aims at the graduate student, the mathematical researcher, and everybody interested in a recent account of extreme value theory.

Point process techniques are also used for "conditional curve estimation," which means estimating the regression curve $P(X \leq x \mid Y = y)$. The authors show convincingly that point processes are the right tool for dealing with these objects in a similar way as extremes of a sequence of random variables; however, extremes are the book's main purpose. This includes the classical theory and rates of convergence in limit theorems for extremes (the authors make extensive use of the Hellinger distance as appropriate means), multivariate maxima, and extremes.

Almost half of the book is dedicated to extremes for non-identically distributed sequences (i.e., general independent, stationary, and Gaussian sequences). A survey of the statistical methods (e.g., for estimating the extremal index and the parameters of the generalized Pareto distribution) is given.

The book is accompanied by a diskette, XTREMES, by S. Hassmann, R.-D. Reiss, and M. Thomas, providing an entertaining and educational statistical software system (for IBM compatible PC's under MS-DOS). This software seems to be unique. It is easy to handle and can be recommended for classroom purposes to illustrate basic distributions, fundamental notions, and the quality of statistical procedures in extreme value theory. Using the software is well documented in a chapter and an appendix.

It is my impression that books on extreme value theory must be very technical, and this monograph is no exception. At least in some of the more specialized chapters the reader will be confronted with some very technical assumptions and proofs. But this does not reduce the book's general value as a compendium of modern extreme value theory with introductions, surveys and long reference lists to different topics.

In summary, *Laws of Small Numbers: Extremes and Rare Events* is a very recommendable book for those interested in regression, extremes, point processes, dependence in probability theory, and statistics.

Thomas MIKOSCH
*University of Groningen*

This content downloaded from 128.59.222.107 on Sat, 02 Dec 2017 18:40:29 UTC
All use subject to http://about.jstor.org/terms