

RESEÑAS / BOOK REVIEWS

AN INTRODUCTION TO CATEGORICAL DATA ANALYSIS

A. Agresti
Wiley.
ISBN: 0-471-11338-7
Xi-290 p

This book gives a systematic introduction to Categorical Data Analysis. The text is self contained and most of the models are accompanied by a worked real life example. The author had in mind the needs of practitioners. A list of proposed problems is provided at the end of each chapter and an Appendix gives an account of the features of some popular statistical packages.

Chapter 1 presents a concise introduction to the inference with categorical data. Chapter 2 describes Contingency Tables and discusses the inferences based on them. Some well known examples illustrate their use. Generalized Linear Models are discussed in Chapter 4. Models for binary responses (logistic regression, loglinear models, etc). A special feature is the discussion of the fitness of a Generalized Linear Model. Six tables illustrate results of the examined example.

In Chapters 5, 6 & 8 the particular models are deeply analyzed. Logistic regression is the theme of Chapter 5, Loglinear Modeling is treated in Chapter 6 and Logit models in Chapter 8. Chapter 7 discusses the model building, selection and diagnostic.

Chapter 9 introduces methods for the analysis of dependent data. Chapter 10 gives a historical sketch of the development of Categorical Data Analysis.

This book is more readable for students and specialists without a formation in Mathematical Statistics than the other book of the author "Categorical Data Analysis". It meets its purpose of raising and stimulating the interest of practitioners of statistics and students. I recommend it as a useful source of information for any person dealing with applications. It is also valuable as a means of instruction.

A PRACTICAL GUIDE TO HEAVY TAILS. STATISTICAL TECHNIQUES AND APPLICATIONS

J. Robert, R. Adler, R., E. Feldman and S. Murad. Taqqu. [editors], 1998

xvi-533

ISBN 0-8176-3951-9

ISBN 3-7643-3951-9 [Basel]

Birkhauser, Boston

The existence of Heavy Tailed distributions and processes appears as a problem to be solved in many applications. This book seeks to spread light on it. The contributors are well known and have published different papers on the topic. A comprehensive and lucid exposition of the features of the problem is made. The importance of the theme is endured by the support of grants from the Israel Science Foundation, US-Israel Binational Science Foundation, the Office of Naval Research and the National Science of Foundation of Usa.

The need of working with the available data is present in all the applications. To avoid the use of 'outliers' is a common solution for granting that the available statistical models work correctly. It is a crazy solution but to disregard the 'bad data' is unfortunately a too popular solution in the experimental sciences. This volume is devoted to the discussion of methods that cope with the existence of a Heavy Tailed [HT] Distribution [HTD].

The first part presents contributions that deal with the intrinsically heavy tailed distributions arising in communication and finance problems. They are two major areas where applied mathematics are playing an increasing role. Crovella-Taqqu-Bestravros discuss on the distributions that arise in the work with the number of files sizes in Word Wide Web.

Willinger-Paxson-Taqqu complete this study with the results of experiments conducted on 'Local area networks' and 'Wide area networks'. They are analyzed through structural models. Financial problems are analyzed in the following 3 papers. Muller-Dacorogna-Pictet established that HTD arises in foreing exchange and interbank markets. The tail of financial assets are estimated. Mittnik-Rachev-Paolella provide an insight in the stable probability schemes. Conditional models are estimated using ARMA and ARMA-GARCH.

Density functions are approximated by fast Fourier transforms. The stable paretian model appears as a realistic model for optimal valuation, capital asset pricing model, and arbitrage pricing. Theory of Risk management is presented by Bassi-Embrechts-Kafetzaki. The estimation of quantiles appears as and adequate solution in the use of different instruments.

The proposed methods are based on the use of Likelihood, Goodness of fit Bayesian and graphical techniques. Their conclusions are supported by numerous tables and graphics.

The Time Series [TS] part is more theoretically oriented. Adler-Feldman-Gallagher concentrate on the use of ARMA models. Box-Jenkins technique is the kernel of the method used for identification, estimation and diagnostic checking. Simulation studies illustrate the behavior of the methods. The next paper, Clader-Davies, is devoted to the estimation of parameters in this setting while Taqqu-Teverovsky's paper tackled the estimation of long range dependence, taking into account if the TS is finite or not. Resnick's paper deals with the challenge posed by the un-convergence of the sample autocorrelation coefficient.

Simulation experiments illustrate the authors remarks. The next paper analyzes how statistical methods must be modified for dealing with TS when the noises have HTD's, Mikosh. The potentialities of a Bayesian approach is the subject of the paper of Ravishanker-Chou. The behavior of some Gibb's samplers is discussed.

Part III, HT Estimation, contains two papers. The first one deals with the study of the use of classic estimators and robust resampling plans, Picet-Darogna-Muller, while the other Kogon-Williams, copes with the use of estimation based characteristic functions under the hypothesis that the distribution parameters are stable.

In part IV are presented applications to regression of some ideas of the third part. Le Page-Podgorsky-Rysnar-White use robust resampling when the errors have a HTD and McCulloch assumes that the disturbances are symmetrically stable and derived that the Maximum Likelihood [ML] Estimators are asymptotically normal distributed. Illustrations of the performance of the proposal are given in both papers.

Part V is devoted to signal processing. The paper of Tsakalidas-Nikias deals with the estimation of parameters when additive interferences are present. A ML estimator procedure is used and simulation experiments validate the proposals. Tsikrintzis develops a similar model for fading signals for a α -stable class constructed Papers of Goldie-Kluppelberg (subexponential distributions), Rosinsky (Structure of Stationary Stable processes) and Samorodnitsky (Tail behavior of some shot noise processes) are placed in Part VI. They study the model structures.

The last part [Numerical Procedures] has four papers. They deal with stable distribution problems. McCulloch derived approximations which permit to develop statistical analysis.

In another paper, coauthored with D.B. Panton, rates are tabulated for maximally skewed stable distributions for various values of α . The first paper of J.P. Nolan presents the state of the art of multilevel stable distributions problems. In a second paper he presents a computational package for computing the parameters of univariate stable distributions.

The book is very well organized. It gives a coverage of the subject of HTD and is suitable for obtaining an overview of the latest results in the numerical treatment of statistical problems when HTD are present.

HANDBOOK OF STATISTICAL MODELING FOR THE SOCIAL AND BEHAVIORAL SCIENCES

C. Arminger, C. Clogg and M. Sobel [editors], 1995
xiv-592
Plenum, New York
ISBN 0-306-44805-x

This book presents, in ten chapters, the roots of different methods of wide use in applications. The name of the book is somewhat misleading because its use is broader than the expressed. The discussed models and approaches to the analysis of data are of wide use in biomedicine, ecology and other non behavioral sciences. The readers should have a good mathematical statistics basis as well as some experience on applications. The chapters are self contained and proofs are not included. A huge list of references permit to obtain the nowadays knowledge of the exposed theme. The authors are well known. The set of subjects, though not exhaustive, contains the discussion of problems of common interest for statisticians that aim to use modern statistical procedures.

The first Chapter tackles the crucial problem of cause-effect, which is the generator of causal inference. M. Sobel presents a historical and philosophical sketch. The probabilistic point of view is discussed and the statistical approaches are presented. Particularly the discussion on the effect of a manipulated variable and the model based approach is very illustrating of the inferential errors that can be generated by the incorrect use of statistical tools.

R. Little and N. Schwanker wrote on Missing Data. The chapter gives an insight of the general problem, not only for sampling and design of experiments but for other particular problems as Probit models. The proposal of Bayesian Monte Carlo methods and the corresponding discussion is very on line with the modern approach to inference. I missed a discussion on the possibility of the use of Reference Analysis within the Bayesian modeling.

The third chapter is almost a book, 107 pages. G. Arminger discussed in it one of the big challenges to regression modeling in applications. He focused on the problem of fixing the relationship between sets of independent and dependent variables. The concept of 'mean structures' is used for coping with the general problem. The diminish of the set of needed hypothesis for fixing a model permits to cope with the general problems and to have another look to regression problems such as diagnostics, censoring, generalized estimating equations.

The next chapter, M. Brown and G. Arminger, completes the mean structure inference analyzed in the previous one. The role of the existence of latent variables is included. The presentation of estimations and computational aspects of structure models which include mean and covariance is the kernel of the chapter.

M. Sobel wrote a chapter on the Analysis of Contingency Tables which provides an almost complete account of the subject. It is an adequate overview of the problem. I make a similar evaluation of the chapter "Latent class models" developed by the late C.C. Clogg. This is a problem that is not as well known as Contingency Tables. Therefore the provided overview is very important.

The next two chapters deal with the use of Panels. The first one is devoted to the analysis when we deal with metric data, Cheng Hsiao, and the second one with the case of qualitative data. A. Hamerle-G. Ronning. Both themes are very important in longitudinal analysis. The usefulness of the use of the same sample on several occasions is discussed. The subsequent chapter, T. Petersen, deals with the same problem but considering a dynamical model when coping the survival models. In this case the respondent's state can change. Several illustration are made using real life data.

The last chapter is due to N.T. Longford and studies the problem of the effect of using multistage clustering instead of simple sampling design in the validity of data analysis procedures. These effects are discussed and general estimating criteria for evaluating them is made through multilevel modeling.

The discussion of the existing statistical packages takes a place in the different chapters.

I highly recommend it for specialists of statistics dealing with applications and for the work in seminars of advanced students.

Carlos Bouza
Universidad de La Habana