

Book review:

**ADVANCES IN DATA ANALYSIS.
THEORY AND APPLICATIONS TO RELIABILITY
AND INFERENCE, DATA MINING,
BIOINFORMATICS, LIFETIME DATA, AND NEURAL
NETWORKS**

by

Christos H. Skiadas, Editor

The reviewed book contains the main part of the invited papers presented at the Twelfth International Conference on Applied Stochastic Models and Data Analysis (ASMDA), which took place in Chania, Crete, Greece, May 29–June 1, 2007. The aim of ASMDA conferences is to stimulate discussion about stochastic modelling, data analysis and their real-life applications.

Therefore, both theoretical and practical papers are included in the reviewed volume – apart from theoretical developments, new results focusing on real-life problems are presented. As noted by editors “an important objective was to select material presenting new methods for solving these problems by analyzing relevant data and leading to the advancement of related fields”.

The book contains twenty nine chapters organized in eight parts, entitled Data Mining and Text Mining, Information Theory and Statistical Applications, Asymptotic Behaviour of Stochastic Processes and Random Fields, Bioinformatics and Markov Chains, Life Table Data, Survival Analysis and Risk in Household Insurance, Neural Networks and Self-Organizing Maps, Parametric and Nonparametric Statistics, and Statistical Theory and Methods.

In the first part, Data Mining and Text Mining, there are four chapters. They are dedicated to bootstrap resampling used to assess the stability of the configurations issued from principal axes methods in case of textual analysis, double projection analysis for lexical tables, adoption of the extended multiple factor analysis to multilingual data of results obtained from surveys, and analysis of number of frequent patterns in random databases.

In the second part, Information Theory and Statistical Applications, there are four chapters. The first chapter is an introduction, focusing on history and concepts of Statistical Information Theory (SIT), like entropy and divergence measures. The following three papers were presented in the Special Session during ASMDA 2007 and they describe model selection criterion – the Divergence

Information Criterion (DIC) with a simulation study to illustrate the comparison of DIC with other criteria such as AIC and BIC, influence or leverage of points in the fitted value of a generalized linear model for ordinal multinomial data with application of minimum phi-divergence estimator and generalized Hessian (hat) matrix as diagnostic tool for measuring the influence, and extensions of the classical divergence measures for application to study of graduating mortality rates.

In the third part, Asymptotic Behaviour of Stochastic Processes and Random Fields, there are seven chapters. The first chapter is an introduction, concentrating on short description of the results, methods, and applications of other chapters. The following papers are devoted to the proof of invariance principle for a branching process in a random environment in case its population reaches a high level, proof of convergence of one-dimensional distribution of the customers' number in a queueing system with a large number of channels to the Gaussian distribution, optimality in the sense of cost approach and sensitivity analysis of class of discrete-time stochastic insurance models, new variant of central limit theorem for random fields defined in \mathbb{R}^d with discussion of some stochastic models in radiobiology, study of dependent random systems indexed by transitive graphs with proof of central limit theorem for such systems, and a study of a symmetric continuous time branching random walk on a d -dimensional lattice with results for asymptotic behaviour of the survival probability of particles and conditional limit theorems for the population size.

In the fourth part, Bioinformatics and Markov Chains, there are two chapters. They are dedicated to the exact method, capable of accounting for both nonstationarity and fragmentary sequences for pattern statistics with application to a simulated and a real set of sequences and a study of stochastic matrix for discrete-time homogenous Markov Chains.

In the fifth part, Life Table Data, Survival Analysis and Risk in Household Insurance, there are three chapters. The papers focus on analysis of Gompertz-type probability density functions and comparison of these functions to a first passage time density function with application to mortality data, approaches to testing homogeneity in a finite mixture model via modified ratio test (MLRT) and D-tests with comparison to the ADDS test, and Bayesian hierarchical model for the Gaussian Markov random field driven nonhomogenous spatial Poisson process of claims counts applied for spatial dependence structure for the counts of certain type of claims occurring in the household insurance in Hungary.

In the sixth part, Neural Networks and Self-Organizing Maps, there are three chapters. They are dedicated to theoretical results related to the Fuzzy Cognitive Network (FCN) development based on theorems specifying the conditions for the uniqueness of solutions for the FCN concept values with case application studies, application of self-organizing maps to analysis of spectral data with a proposal of dimension reduction for providing synthetic representation of the map to the most relevant variables, and a study of an adaptive neural network with a fuzzy inference system (ANFIS) for a wind energy production forecasting

model applied to a real-life case with comparison to autoregressive (AR) and autoregressive moving average (ARMA) models.

In the seventh part, Parametric and Nonparametric Statistics, there are two chapters. The papers describe nonparametric hypothesis testing for making decisions whether the baseline distributions of several sequential k -out-of- n systems are equal with study of asymptotic distributions of the test statistics and calibration issues for adjusting p -values in statistical tests for the large sample size in normal model with known and unknown variance.

In the eighth part, Statistical Theory and Methods, there are four chapters. The papers present implementation of the forward search to fitting the Pareto II distribution to firm size data with possible economic interpretations, estimation of an extremely high risk level for Value-at-Risk (VaR) measure via Extreme Value Theory (EVT) and bootstrap techniques, estimation of the mean and variance models of the responses in case of correlation between multiple responses and heteroscedasticity of the observations with application to an electron beam welding experiment, and a procedure for detecting a systematic change in parameter for a sequence of binomial variables.

The book is useful for graduate students, researchers and practitioners. Taking into account the wide range of the results presented, the theory itself and variety of real-life applications provided in this book, it should be recommended for readers interested in various flavours of data analysis.

Maciej Romaniuk

Christos H. Skiadas, editor, *Advances in Data Analysis. Theory and Applications to Reliability and Inference, Data Mining, Bioinformatics, Lifetime Data, and Neural Networks*. SIT – Statistics for Industry and Technology. Birkhäuser Verlag, Basel–Boston–Berlin, 2010. XXIV + 364 pages. ISBN 978-0-8176-4798-8. Price: 109.95 EUR (hardcover).

