Contents

Contributions to the Special Volume of the XVI School on Regression Models (“Escola de Modelos de Regressão”) in honour of Bent Jørgensen

Preface ................................................................................................................................. 1

G. M. CORDEIRO, R. LABOURIAU and D. A. BOTTER
An introduction to Bent Jørgensen’s ideas ........................................................................ 2

P. V. C. PEREIRA, I. T. S. PREVIDELLI and A. C. DAVISON
Practical issues with modeling extreme Brazilian rainfall ............................................ 21

T. M. MAGALHÃES, D. A. BOTTER and M. C. SANDOVAL
A general expression for second-order covariance matrices—an application to dispersion models .......................................................................................................................... 37

W. S. KENDAL
Self-organized criticality of aggregated animals attributed to Tweedie convergence .......... 50

H. SAULO, J. LEÃO, J. NOBRE and N. BALAKRISHNAN
A class of asymmetric regression models for left-censored data .................................. 62

F. F. DO NASCIMENTO and A. S. ASSUNÇÃO
Regression models for change point data in extremes ................................................. 85

W. S. KENDAL
The snap, crackle and pop of solar flares explained ..................................................... 101

H. SAULO, J. LEÃO, V. LEIVA, R. VILA and V. TOMAZELLA
A bivariate fatigue-life regression model and its application to fracture of metallic tools .... 119

Articles

M. SAHA, S. DEY, A. S. YADAV and S. ALI
Confidence intervals of the index $C_{pk}$ for normally distributed quality characteristics using classical and Bayesian methods of estimation .................................................................. 138

E. G. COSTA, C. D. PAULINO and J. M. SINGER
Sample size for estimating organism concentration in ballast water: A Bayesian approach ... 158

F. N. DEMARQUI and V. D. MAYRINK
Yang and Prentice model with piecewise exponential baseline distribution for modeling lifetime data with crossing survival curves ......................................................................................... 172

L. H. M. MORITA, V. L. D. TOMAZELLA, P. L. RAMOS, P. H. FERREIRA and F. LOUZADA
The random deterioration rate model with measurement error based on the inverse Gaussian distribution .................................................................................................................. 187
Preface
An introduction to Bent Jørgensen’s ideas

Gauss M. Cordeiro\textsuperscript{a}, Rodrigo Labouriau\textsuperscript{b} and Denise A. Botter\textsuperscript{c}

\textsuperscript{a}Universidade Federal de Pernambuco
\textsuperscript{b}Aarhus University
\textsuperscript{c}Universidade de São Paulo

Abstract. We briefly expose some key aspects of the theory and use of dispersion models, for which Bent Jørgensen played a crucial role as a driving force and an inspiration source. Starting with the general notion of dispersion models, built using minimalistic mathematical assumptions, we specialize in two classes of families of distributions with different statistical flavors: exponential dispersion and proper dispersion models. The construction of dispersion models involves the solution of integral equations that are, in general, untractable. These difficulties disappear when more mathematical structure is assumed: it reduces to the calculation of a moment generating function or of a Riemann–Stieltjes integral for the exponential dispersion and the proper dispersion models, respectively. A new technique for constructing dispersion models based on characteristic functions is introduced turning the integral equations above into a tractable convolution equation and yielding examples of dispersion models that are neither proper dispersion nor exponential dispersion models. A corollary is that the cardinality of regular and non-regular dispersion models are both large.

Some selected applications are discussed including exponential families non-linear models (for which generalized linear models are particular cases) and several models for clustered and dependent data based on a latent Lévy process.

References


Key words and phrases. Dispersion models, exponential dispersion models, exponential family, non-linear models, proper dispersion models, saddlepoint approximations.


Practical issues with modeling extreme Brazilian rainfall

Paulo V. C. Pereira\textsuperscript{a}, Isolde T. S. Previdelli\textsuperscript{a} and Anthony C. Davison\textsuperscript{b}

\textsuperscript{a}Universidade Estadual de Maringá
\textsuperscript{b}École Polytechnique Fédérale de Lausanne

Abstract. Accurately quantifying extreme rainfall is important for the design of hydraulic structures, for flood mapping and zoning and for disaster management. In order to produce maps of estimates of 25-year rainfall return levels in Brazil, we selected 893 shorter and 104 longer rainfall time series from the Agência Nacional de Águas (ANA), and applied the framework of extreme value theory. Care was needed to reduce the impact of poor data. Estimates of the shape parameter of the extreme-value model fitted to rainfall data are typically biased, so we discuss an empirical correction that takes into account not only the sample-size bias, but also a so-called penultimate approximation that we use to inform a Bayesian spatial latent variable model for the annual rainfall maxima. This model accounts for subtle patterns of spatial variation in the data and provides plausible return level estimates.

References


\textit{Key words and phrases}. Bayesian hierarchical model, Data quality, penultimate approximation, return level map, small-sample bias, statistics of extremes.


A general expression for second-order covariance matrices—an application to dispersion models

Tiago M. Magalhães\textsuperscript{a}, Denise A. Botter\textsuperscript{b} and Mônica C. Sandoval\textsuperscript{b}

\textsuperscript{a}Federal University of Juiz de Fora
\textsuperscript{b}University of São Paulo

Abstract. We present a general expression that allows the calculation of both the $n^{-2}$ asymptotic covariance matrices of the maximum likelihood estimator (MLE) and the first-order bias corrected MLE, where $n$ is the sample size. The formula is presented in a matrix notation which has numerical advantages since it requires only simple operations on matrices and vectors. The usefulness of the formula is to construct better Wald statistics. We apply our findings to dispersion models and develop simulation studies which show that modification in the Wald statistic effectively removes size distortions of the type I error probability with no power loss. For illustrative purposes, a real data application is considered to support our theoretical results.

References


Key words and phrases. Bias estimator, covariance matrix, dispersion models, Wald test.


Self-organized criticality of aggregated animals attributed to Tweedie convergence

Wayne S. Kendal
University of Ottawa

In memoriam: Prof. Bent Jørgensen (1954–2015), who contributed the fundamentals on which this study was based.

Abstract. Ecologists have had an ongoing interest in a variance to mean power law that governs the clustering of individuals of animal and plant species. This same power law has been reported from disparate biological, physical and mathematical systems, and also characterizes a family of statistical distributions known as the Tweedie exponential dispersion models. Its widespread appearance can be explained by fundamental statistical convergence effects on random data that cause this, and related, power laws to emerge and provide mechanistic insight into its origin, as well as the origin of $1/f$ noise, multifractality and other phenomena attributable to self-organized criticality. A meta-analysis of ecological field data was conducted here to examine how such statistical convergence might affect the power law. These findings provided conjectural insight into a form of self-organized criticality, driven and modulated by the statistical convergence of random data, which could underlie the power law’s emergence.

References


Key words and phrases. Emergent phenomena, critical phenomena, phase transitions, Taylor’s power law, weak convergence.


A class of asymmetric regression models for left-censored data

Helton Saulo\textsuperscript{a}, Jeremias Leão\textsuperscript{b}, Juvêncio Nobre\textsuperscript{c} and Narayanaswamy Balakrishnan\textsuperscript{d}

\textsuperscript{a}University of Brasília
\textsuperscript{b}Federal University of Amazonas
\textsuperscript{c}Federal University of Ceará
\textsuperscript{d}McMaster University

Abstract. A common assumption in the standard tobit model is the normality for the error distribution. However, asymmetry and bimodality may be present and alternative tobit models must be used in such cases. In this paper, we propose a tobit model based on the class of log-symmetric distributions, which includes as special cases heavy/light tailed distributions and bimodal distributions. We implement a likelihood-based approach for parameter estimation and consider a type of residual. We then discuss the problem of performing hypothesis tests within the proposed class by using the likelihood ratio and gradient statistics, which are particularly convenient for tobit models, as they do not require the information matrix. An elaborate Monte Carlo study is carried out for evaluating the performance of the maximum likelihood estimates, the likelihood ratio and gradient tests and the empirical distribution of the residuals. Finally, we illustrate the proposed methodology with the use of a real data set.

References


Key words and phrases. Log-symmetric distributions, likelihood ratio test, gradient test, R software, tobit models.


Regression models for change point data in extremes

Fernando Ferraz do Nascimento\textsuperscript{a} and Alan da Silva Assunção\textsuperscript{b}

\textsuperscript{a}Universidade Federal do Piauí
\textsuperscript{b}Universidade Federal do Ceará

Abstract. Many extreme events are characterized by being always susceptible to outside influences that will modify their behavior at some point in time. The change point tool has been used in statistical models to detect when these changes occur. This paper presents a model based on a Bayesian approach that describes the behavior of extreme data regarding river quota, which may present more than one change point. In each one of the regimes, the GEV distribution is adjusted and each GEV parameter of each regime is written in function of presence of covariates. In the applications proposed here, the results showed that the model was able to accurately estimate the actual amount of change points in the series, and also showed that it was extremely important to consider them in the analysis, since it was verified that after the change of regime, the levels of return have changed considerably. The results were also able to show which months the occurrence of an extreme event is greater.

References


\textit{Key words and phrases.} Extreme events, change points, regression structure, Bayesian approach, river quota.


The snap, crackle and pop of solar flares explained

Wayne S. Kendal
University of Ottawa

In memoriam: Prof. Bent Jørgensen (1954–2015), who contributed the fundamentals on which this study was based.

Abstract. The irregular fluctuations of solar flare emissions, as determined from terrestrial neutron monitors, remains poorly understood. These records empirically revealed a temporally-related variance to mean power law, 1/f noise and a non-Gaussian distribution, all features indicative of self-organized criticality, a theory of how deterministic dynamical systems can spontaneously evolve to unstable states that express erratic changes. The non-Gaussian distribution found here approximated a Tweedie compound Poisson exponential dispersion model, a statistical distribution characterized by a variance to mean power law that itself can imply 1/f noise. Tweedie exponential dispersion models serve a primary role in statistical theory as foci for weak convergence for a wide range of random distributions, a role which supports an alternative conjecture to explain the solar flare fluctuations as being based on random processes rather than a deterministic system.

References


Key words and phrases. Solar cosmic rays, emergent property, Taylor’s power law, multifractals, weak convergence.


A bivariate fatigue-life regression model and its application to fracture of metallic tools

Helton Saulo\textsuperscript{a}, Jeremias Leão\textsuperscript{b}, Víctor Leiva\textsuperscript{c}, Roberto Vila\textsuperscript{a} and Vera Tomazella\textsuperscript{d}

\textsuperscript{a}Universidade de Brasília
\textsuperscript{b}Universidade Federal do Amazonas
\textsuperscript{c}Pontificia Universidad Católica de Valparaíso
\textsuperscript{d}Universidade Federal de São Carlos

Abstract. The Birnbaum–Saunders distribution has been widely used to model reliability and fatigue data. In this paper, we propose a regression of generalized linear models type based on a new bivariate Birnbaum–Saunders distribution. This is parameterized in terms of its means and allows data to be described in their original scale. We estimate the model parameters and carry out inference with the maximum likelihood method. A case study with real-world reliability data is conducted for motivating our investigation, illustrating the potential applications of the proposed results. We obtain a predictive model which can be a useful addition to the tool-kit of diverse practitioners, reliability engineers, applied statisticians, and data scientists.

References


Key words and phrases. Bivariate Birnbaum–Saunders distribution, fatigue data, maximum likelihood method, predictive models, \texttt{R} software.


Confidence intervals of the index $C_{pk}$ for normally distributed quality characteristics using classical and Bayesian methods of estimation

Mahendra Saha$^a$, Sanku Dey$^b$, Abhimanyu Singh Yadav$^c$ and Sajid Ali$^d$

$^a$Central University of Rajasthan
$^b$St. Anthony’s College
$^c$Banaras Hindu University
$^d$Quaid-I-Azam University

Abstract. One of the indicators for evaluating the capability of a process potential and performance in an effective way is the process capability index (PCI). It is of great significance to quality control engineers as it quantifies the relation between the actual performance of the process and the pre-set specifications of the product. Most of the traditional PCIs performed well when process follows the normal behaviour. In this article, we consider a process capability index, $C_{pk}$, suggested by Kane (Journal of Quality Technology 18 (1986) 41–52) which can be used for normal random variables. The objective of this article is three fold: First, we address different methods of estimation of the process capability index $C_{pk}$ from frequentist approaches for the normal distribution. We briefly describe different frequentist approaches, namely, maximum likelihood estimators, least squares and weighted least squares estimators, maximum product of spacings estimators, Cramèr–von-Mises estimators, Anderson–Darling estimators and Right-Tail Anderson–Darling estimators and compare them in terms of their mean squared errors using extensive numerical simulations. Second, we compare three parametric bootstrap confidence intervals (BCIs) namely, standard bootstrap, percentile bootstrap and bias-corrected percentile bootstrap. Third, we consider Bayesian estimation under squared error loss function using normal prior for location parameter and inverse gamma for scale parameter for the considered model. Monte Carlo simulation study has been carried out to compare the performances of the classical BCIs and highest posterior density (HPD) credible intervals of $C_{pk}$ in terms of average widths and coverage probabilities. Finally, two real data sets have been analyzed for illustrative purposes.

References


Key words and phrases. Process capability index, normal distribution, different classical methods of estimation, bootstrap confidence intervals, sensitivity analysis.


Sample size for estimating organism concentration in ballast water: A Bayesian approach

Eliardo G. Costa\textsuperscript{a}, Carlos Daniel Paulino\textsuperscript{b} and Julio M. Singer\textsuperscript{c}

\textsuperscript{a}Universidade Federal do Rio Grande do Norte
\textsuperscript{b}CEAUL, Universidade de Lisboa
\textsuperscript{c}Universidade de São Paulo

Abstract. Estimation of microorganism concentration in ballast water tanks is important to evaluate and possibly to prevent the introduction of invasive species in stable ecosystems. For such purpose, the number of organisms in ballast water aliquots must be counted and used to estimate their concentration with some precision requirement. Poisson and negative binomial models have been employed to describe the organism distribution in the tank, but determination of sample sizes required to generate estimates with pre-specified precision is still not well established. A Bayesian approach is a flexible alternative to accommodate adequate models that account for the heterogeneous distribution of the organisms and may provide a sequential way of enhancing the estimation procedure by updating the prior distribution along the ballast water discharging process. We adopt such an approach to compute sample sizes required to construct credible intervals obtained via two optimality criteria that have not been employed in this context. Such intervals may be used in the decision with respect to compliance with the D-2 standard of the Ballast Water Management Convention. We also conduct a simulation study to verify whether the credible intervals obtained with the proposed sample sizes satisfy the precision criteria.

References


Key words and phrases. Average coverage criterion, average length criterion, Poisson/gamma distribution, negative binomial/Pearson Type VI distribution.


Yang and Prentice model with piecewise exponential baseline distribution for modeling lifetime data with crossing survival curves

Fábio N. Demarqui and Vinícius D. Mayrink
Universidade Federal de Minas Gerais

Abstract. Proportional hazards (PH), proportional odds (PO) and accelerated failure time (AFT) models have been widely used to deal with survival data in different fields of knowledge. Despite their popularity, such models are not suitable to handle survival data with crossing survival curves. Yang and Prentice (2005) proposed a semiparametric two-sample approach, denoted here as the YP model, allowing the analysis of crossing survival curves and including the PH and PO configurations as particular cases. In a general regression setting, the present work proposes a fully likelihood-based approach to fit the YP model. The main idea is to model the baseline hazard via the piecewise exponential (PE) distribution. The approach shares the flexibility of the semiparametric models and the tractability of the parametric representations. An extensive simulation study is developed to evaluate the performance of the proposed model. We demonstrate how useful is the new method through the analysis of survival times related to patients enrolled in a cancer clinical trial. Finally, an \texttt{R} package called \texttt{YPPE} was developed to fit the proposed model. The simulation results indicate that our model performs well for moderate sample sizes in the general regression setting. A superior performance is also observed with respect to the original YP model designed for the two-sample scenario.

References


\textit{Key words and phrases.} Survival analysis, short-term and long-term hazard ratios, semiparametric modeling, maximum likelihood estimation.


The random deterioration rate model with measurement error based on the inverse Gaussian distribution

Lia H. M. Morita\textsuperscript{a}, Vera L. D. Tomazella\textsuperscript{b}, Pedro L. Ramos\textsuperscript{c}, Paulo H. Ferreira\textsuperscript{c} and Francisco Louzada\textsuperscript{c}

\textsuperscript{a}Federal University of Mato Grosso  \\
\textsuperscript{b}Federal University of São Carlos  \\
\textsuperscript{c}University of São Paulo

Abstract. In this paper, we introduce the random deterioration rate model with measurement error in order to incorporate the variability among different components. The motivation behind the random variable model is to capture the randomness in the individual differences across the population. This model incorporates only sample uncertainty of the degradation, and no temporal variability is included. The measurement error models appear to overcome this problem. The random rate analysis is based on repeated measurements of failure sizes generated by a degradation process over time in a components population. Some characteristics of the random deterioration rate model based on the inverse Gaussian distribution and subject to measurement error, are examined. We carry out simulation studies to (i) assess the performance of the maximum likelihood estimates obtained through the Gaussian quadrature along with Quasi-Newton optimization method; and (ii) examine the effects of model misspecification on the model selection criteria’s performance, as well as on the lifetime prediction’s accuracy and precision. The potentiality of the proposed model is illustrated through two real data sets.

References


\textit{Key words and phrases}. Degradation analysis, inverse Gaussian distribution, LASER data, locomotive wheels data, random deterioration rate.


AIMS AND SCOPE

The Brazilian Journal of Probability and Statistics aims to publish high quality research papers in applied probability, applied statistics, computational statistics, mathematical statistics, probability theory and stochastic processes.

More specifically, the following types of contributions will be considered:

(i) Original articles dealing with methodological developments, comparison of competing techniques or their computational aspects;
(ii) Original articles developing theoretical results;
(iii) Articles that contain novel applications of existing methodologies to practical problems. For these papers the focus is in the importance and originality of the applied problem, as well as, applications of the best available methodologies to solve it.
(iv) Survey articles containing a thorough coverage of topics of broad interest to probability and statistics. The journal will occasionally publish book reviews, invited papers and essays on the teaching of statistics.

GENERAL INFORMATION


Submissions: Manuscripts for Brazilian Journal of Probability and Statistics should be submitted online. Authors may access the Electronic Journals Management System (EJMS) at http://www.e-publications.org/ims/submission.


Permissions Policy. Authorization to photocopy items for internal or personal use is granted by the Institute of Mathematical Statistics. For multiple copies or reprint permission, contact The Copyright Clearance Center, 222 Rosewood Drive, Danvers, Massachusetts 01923. Telephone (978) 750-8400. http://www.copyright.com. If the permission is not found at the Copyright Clearance Center, please contact the IMS Business Office: ims@imstat.org.

Correspondence. Mail concerning membership, subscriptions, nonreceipt claims, copyright permissions, advertising or back issues should be sent to the IMS Dues and Subscription Office, 9650 Rockville Pike, Suite L 2510, Bethesda, Maryland 20814-3998. Mail concerning submissions or editorial content should be sent to the Editor at nancy@ime.unicamp.br. Mail concerning the production of this journal should be sent to: Geri Mattson at bjps@mattsonpublishing.com.

Individual and Organizational Memberships: http://www.imstat.org/individual-membership/

Individual and General Subscriptions: http://www.imstat.org/orders/


The Brazilian Journal of Probability and Statistics is an IMS supported journal: http://www.imstat.org/journals-and-publications/ims-supported-journals/