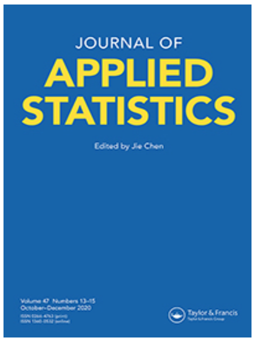

Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Editorial to special issue V WCDANM 2018

M. Stehlík, L. M. Grilo & P. K. Jordanova

To cite this article: M. Stehlík, L. M. Grilo & P. K. Jordanova (2020) Editorial to special issue V WCDANM 2018, Journal of Applied Statistics, 47:13-15, 2289-2298, DOI: 10.1080/02664763.2020.1818489

To link to this article: https://doi.org/10.1080/02664763.2020.1818489

Published online: 14 Sep 2020.

JOURNAL OF APPLIED STATISTICS
2020, VOL. 47, NOS. 13–15, 2289–2298
https://doi.org/10.1080/02664763.2020.1818489

Taylor & Francis
Taylor & Francis Group

EDITORIAL

Check for updates

Editorial to special issue V WCDANM 2018

M. Stehlík[a,b], L. M. Grilo[c,d,e] and P. K. Jordanova[f]

[a]Linz Institute of Technology & Department of Applied Statistics, J. Kepler University in Linz, Linz, Austria; [b]Instituto de Estadística, Universidad de Valparaíso, Valparaíso, Chile; [c]Instituto Politécnico de Tomar (IPT) & Universidade Aberta, Tomar, Tomar, Portugal; [d]Centro de Matemática e Aplicações (CMA), FCT, Universidade Nova de Lisboa, Lisboa, Portugal; [e]Centro de Investigação em Cidades Inteligentes (Ci2), IPT & CIICESI / ESTG – P. Porto, Porto, Portugal; [f]Faculty of Mathematics and Informatics, Konstantin Preslavsky University of Shumen, Shumen, Bulgaria

The special issue Advances in Computational Data Analysis of the *Journal of Applied Statistics* (JAS), Taylor & Francis, contains mainly papers that were presented in the fifth Annual Workshop of Computational Data Analysis and Numerical Methods (V WCDANM), which took place on 11–12 May 2018, at the Polytechnic Institute of Porto, Portugal. The organizing committee of V WCDANM – 2018, with the support of the Polytechnic Institute of Tomar and the University of Évora, developed a program that includes prominent keynote speakers and a high scientific level of oral and poster sessions, with participants from Portugal and abroad. Theoretical and applied works in different research fields were presented, namely in health and social sciences, environmental science, economics and engineering (some involving data science, data mining, big data and machine learning). A considerable number of manuscripts were submitted to this special issue and more than 30 papers, after carefully reviewed by referees, were accepted and are distributed in three issues of JAS Volume 47. The selected papers offer readers the opportunity to access different statistical approaches, as well as to view a wide range of application areas. These research works provide the appropriate framework and background for real-life problems and also they reflect a comprehensive view of different statistical fields, promoting links with a variety of related disciplines, exploring computational issues and presenting some future research trends.

Branching processes

Bulgarian probabilistic school has long traditions in studying branching processes. Tchorbadjieff and Mayster [36] start with a very good summary of the contemporary results in this area. The authors study a linear birth-death process starting with a random number of particles. At first, they consider the cases when the initial distribution follows different standard probabilistic distributions – Negative Binomial and its closest Geometric, Poisson or Pólya-Aeppli distributions. They proved analytically and numerically that the random initial conditions can not change the critical parameter of branching mechanism, but they

CONTACT M. Stehlík mlnstehlik@gmail.com, milan.stehlik@jku.at, milan.stehlik@uv.cl

modify the extinction probability by the time. Finally, the possibility for general solution for more complex initial conditions is demonstrated by implementing a numerical solution developed in a previous study on linear birth-death process initiated after some particular cases of Pólya urns trails.

Clinical trials and related topics

Flournoy and Oron [15] reveal and analyze the important fact that dynamics of design lead to bias, which is in particular interesting from the point of view of Dose finding. Such bias can happen by many adaptive design strategies. These results have also been discussed in an Invited Session for the JSM 2019 'Optimal designs for modeling asymmetries in big data'.

Fermín and Lévy-Véhel [13] model stochastic dynamic of drug concentration in the case of multiple intravenous and poor patient adherence behavior. In order to describe the result of this situation the authors use Piecewise-Deterministic Markov Process framework. They investigate the distribution of the limit concentration and its variability. The case with two different doses is thoroughly investigated and compared with deterministic model by means of the simulation study.

Dependent and censored data

Economou *et al.* [11] describe problems with non-random sample setup and describe cases when it is reasonable to replace them with a sample of randomly weighted data from a specific weighted distribution. The standard Maximum Likelihood equation for estimation of the parameters is correspondingly replaced by the Maximum Likelihood equation with weighted probability density function, which can in some particular cases outperform the standard Maximum Likelihood approach.

Many of the real-life phenomena contain dependent data which comes from relatively short multivariate time series of counts. Due to the small sample size the asymptotic statistical theory does not provide useful results for analyzing such data. Two particular cases of these processes when the dependence structure is based on the binomial thinning operator and bivariate moving average processes are considered in [35]. The main probabilistic small sample properties are investigated in particular cases, when the bivariate cross-correlated innovations are Poisson or Negative binomial. The results could be expresses also in the terms of mixed or compound distributions.

Another important question which this issue addresses is the problem of working with missing data. One of the particular cases when one can analyze them via probabilistic approaches is the case when some of the experimental units are removed during the experiment. Such a model is Type-II censored competing risk model with two competing risk factors. It is studied in [1]. The authors propose EM-algorithm to compute the solution of the Maximum Likelihood system of equations for estimation of the parameters of Chen's distribution. They illustrate cases when bootstrap techniques outperform the expected Fisher's information matrix approaches and also consider Bayesian estimation techniques.

Extremes and tails

Good estimation of the parameters of rare events is one of the greatest statistical challenges of our century. The sizes of damages and losses or the intervals between them are usually modeled via probability distributions whose corresponding cumulative distribution function has regularly varying right tail. However, small errors in estimation of the corresponding index of regular variation cause big differences in estimation of the quantiles outside the range of the data. Hill estimator of this index is very well-known in the scientific literature given that the data are independent identically distributed (i.i.d.). It is based on the average of the logarithms of order statistics. A natural question is: Why the average should be the best mean for estimation of the regular variation parameter? One natural alternative, which enabled construction of t-Hill estimator introduced in [12], was replacement of the arithmetic mean by the harmonic mean. It was also demonstrated that for small values of the tail parameter t-Hill estimator outperforms Hill's estimator. The next step is done simultaneously by parametrization of such estimator, and it was introduced independently almost simultaneously by [7] (MOP; the harmonic mean with mean of order p), [6] and [27]. In [28] MOP is investigated for the values of p for which the estimator has smallest variance and bias. They deal with the semi-parametric estimation of the EVI, for heavy tails. A recent class of EVI-estimators, based on the Lehmer's mean-of-order p, L_p which generalizes the arithmetic mean, is considered. An asymptotic comparison at optimal levels performed in previous works has revealed the competitiveness of this class of EVI-estimators. A large-scale Monte-Carlo simulation study for finite simulated samples has been performed. A bootstrap adaptive choice of $(k; p)$, where k is the number of upper order statistics used in the estimation, and a second algorithm based on a stability criterion is computationally studied and applied to simulated and real data.

In real data analysis it is often difficult to assert that the sample are i.i.d. observations, e.g. in time series analysis. Therefore, the next natural step in extreme values theory is to develop analogous technique, when there exists some form of dependence between the observations. In these cases the exceedances of some high threshold usually appear in clusters. Therefore, the cluster size distribution is important to be studied. Its main characteristics is the extremal index. Gomes and Neves [17] compare some of the methods for estimation of the extremal index in a strictly stationary sequence, and describes their advantages and disadvantages. The authors propose a new method for defining blocks. They define a new threshold and corresponding new block maxima estimator.

Quantiles play crucial role in data transformations. Normalization transformations have recently been applied widely in machine learning, particularly in data preprocessing. However, the classical methods that can be adapted to cross-validation are not always effective. Peterson and Cavanaugh [29] introduce Ordered Quantile (ORQ) normalization, a one-to-one transformation that is designed to consistently and effectively transform a vector from an arbitrary distribution into a vector that follows a normal (Gaussian) distribution. In the absence of ties, ORQ normalization is guaranteed to produce normally distributed transformed data. Furthermore, if a normalizing function exists for a generating distribution, ORQ will approximate this function. Once trained, an ORQ transformation can be readily and effectively applied to new data. Authors compare the effectiveness of the ORQ technique with other popular normalization methods in a simulation study where the true data generating distributions are known. It turned out that ORQ normalization is the only method that works consistently and effectively, regardless of the underlying distribution. They built `bestNormalize` to evaluate the normalization efficiency of many candidate transformations; the package is freely available via the Comprehensive R Archive Network.

A correct modelization of the distribution of insurance losses is crucial in the insurance industry. Tomarchio and Punzo [37] present a model class for modeling insurance losses which enables to model extreme losses in a convenient way. This distribution is generally highly positively skewed, unimodal hump-shaped, and with a heavy right tail. Compound models are a profitable way to accommodate situations in which some of the probability masses are shifted to the tails of the distribution. Therefore, in this work, a general approach to compound unimodal hump-shaped distributions with a mixing dichotomous distribution is introduced. A two-parameter unimodal hump-shaped distribution, defined on a positive support, is considered and reparametrized with respect to the mode and to another parameter related to the distribution variability. The compound is performed by scaling the latter parameter by means of a dichotomous mixing distribution that governs the tail behavior of the resulting model. The proposed model can also allow for automatic detection of typical and atypical losses via a simple procedure based on maximum a posteriori probabilities. Unimodal gamma and log-normal are considered as examples of unimodal hump-shaped distributions.

The questions about the estimation of the extremal index are important when we need to estimate quantiles outside the range of the data. However, more frequently in practice we need to model the data within their range. In such cases it is better to determine values which do not correspond to the model assumed and to analyze them separately. It is well-known that such values are called 'outliers'. There is no unique procedure for their characterization in the scientific literature. Ref. [23] is devoted to this topic. The authors define a new parametric approach for determination of outliers. Then, they work with a sample of independent observations of a normal random variable (r.v.), compute the corresponding process capability indexes, analyze their behavior in presence of outliers, and propose new robust parametric estimators of the parameters of the observed distribution. They are based on method of moments and on the maximum likelihood estimation procedure. The question about detection of outliers in high-dimensional case is even much more difficult. A step forward in this direction is done in [20]. The author proposes an extension of the affine invariant version of the classification rule inclusively for classifying functional data. The main rule is based on cumulative distribution function of multivariate rank functions for high-dimensional data.

The infinite number of probability distributions requires the study of different indexes for comparisons between them. Touré *et al.* [38] restrict to the set of distributions with finite variance which is a subset of the set of all light-tailed distributions. The authors use plug-in principle and obtain estimators of relative dispersion and the relative variation indexes with respect to the relative variability of a nonnegative natural exponential family of distributions. They provide the strong consistency and asymptotic normality results.

Forecasting and LASSO

Silva and Alonso [34] develop and used different methodologies to predict the time series of the monthly overnight stays in the North Region of Portugal. The authors compared the forecasts given by neural networks, singular spectrum analysis, the seasonal naive method, seasonal autoregressive integrated moving average models and exponential smoothing models. Accordingly, the two forecasting accuracy measures that used the singular spectrum analysis showed the best results.

To analyze the datasets containing matrix and vector variables, Li *et al.* [19] proposed a novel matrix regression model which combines fused LASSO and nuclear norm penalty. They designed an algorithm to make the low-rank and fused LASSO matrix regression model practically feasible. They also proposed the linearized alternating direction method of multipliers and establish its global convergency. A numerical experiments to demonstrate the efficiency of the proposed method was carried out.

Giacomazzo and Kamarianakis [16] present a Bayesian three-step model-building procedure for parsimonious estimation of Threshold-Autoregressive (TAR) models, designed for location- day- and horizon-specific forecasting of traffic occupancy. In addition to forecasting, the proposed specification and model-building scheme, may assist in determining location-specific congestion thresholds and associations between traffic dynamics observed in different regions of a network. Empirical results applied to data from a traffic forecasting competition illustrate the efficacy of the proposed procedures in obtaining interpretable models and in producing satisfactory point and density forecasts at multiple horizons.

Information selection criteria

The existing procedures for information-theoretic model selection (such as the Akaike Information Criterion and related information criteria) do not provide explicit and uniform control over error rates for the choice between models, a key feature of classical hypothesis testing. Cullan *et al.* [9] shows how to extend notions of Type-I and Type-II error to more than two models without requiring a null. They presented the Error Control for Information Criteria (ECIC) method, a bootstrap approach to controlling Type-I error using Difference of Goodness of Fit (DGOF) distributions, and applied ECIC to empirical and simulated data in time series and regression contexts to illustrate its value for parametric Neyman–Pearson classification.

Mixed models

To model dengue occurrences in the Goiás state, Brazil (where the tropical climate favors the proliferation of the main transmitting vector of this disease), Oliveira *et al.* [25] apply generalized linear mixed modeling to the data from multiple governmental bodies, which describes the number of dengue cases and meteorological conditions in 20 cities. The number of reported dengue cases is estimated using meteorological variables as fixed effects, and city and year data are included in the model as random effects. The results confirm that precipitation, minimum temperature, and relative air humidity contribute to the increase of dengue cases number, while year and city location are determining factors.

Nunes *et al.* [24] extend the theory of analysis of variance, considering mixed effects models, for the situations where the sample sizes may not be previously known and it is more appropriate consider them as realizations of random variables. The authors assumed that the occurrences of observations correspond to a counting process and the sample sizes have Poisson distribution. Application of the proposed approach to a study of cancer patients is given.

Models, estimation and testing

Arnold and Arvanitis [3] introduce an important class of bivariate pseudo-exponential distributions as a bivariate conditionally specified distribution. A variation of this conditioning regime is introduced, and its characteristics are contrasted with the original, provided series of works of Jerzy and Lidia Filus, starting with [14]. Per-capita Gross Domestic Product (GDP) is a measure of a nation's total annual production of goods and services, divided by its population. Two variations of both the original and the new conditioning regime are applied to GDP and infant mortality data across nations and territories.

Carter and Cavanaugh [8] consider the problem of variable selection within the linear model framework. Taking a hypothesis testing approach, where models can only be tested in pairs with one model nested within another, the authors develop a clever algorithm for deciding which variables are to be included in the selected model. The paper provides extensive background into the model selection problem, covering the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and likelihood ratio testing. The Sufficiently Improved Fitting Term (SIFT) Procedure introduced in the paper is based on multiple likelihood ratio tests with an adjustment for the multiple testing. Adjusting for multiplicities is often overlooked in hypothesis testing, particularly when testing is used for variable selection, so the authors are providing a nice contribution to the statistical inference literature.

Antunes *et al.* [2] use a well-known property of Cumulant Generating Function to estimate the first four order cumulants, using Least Squares Estimators. In the case of Additive Models, Empirical Best Linear Unbiased Predictors are also obtained. Pairs of independent and identically distributed models associated to the treatments of a base design are used to obtain unbiased estimators for the fourth-order cumulants.

Arrué *et al.* [4] derive a new type of Birnbaum–Saunders model based on the modified skew-normal distribution, which resulted to be widely flexible in skewness and kurtosis (controlled by a shape parameter). A simulation study was developed to correct the bias of the corresponding estimators, thus, improve the inferential performance. In order to show the good agreement between these new types of model, it was adjusted to life-time data related to fatigue.

A new approach based on linear programming is proposed by [31] to estimate voter transitions among parties (or candidates) between two elections. This approach presents two important innovations: it explicitly deals with new entries and exits in the election census without assuming unrealistic hypotheses; explores the information contained in the model residuals. The method is illustrated estimating the vote transfer matrix between the first and second rounds of the 2017 French presidential election and measuring its level of uncertainty. An R-function has been developed for this procedure and is available in this paper.

Witkovský [39] uses the inversion formulas of Gil-Pelaez and the trapezoidal rule, to invert the characteristic function. Moreover, the explicit characteristic function of the corrected Bartlett's test statistic together with the computationally fast and efficient implementation of the approach based on numerical inversion of this characteristic function, suggested for evaluating the exact null distribution used for testing homogeneity of variances in several normal populations, with possibly unequal sample sizes. The author developed a Matlab package, base on the previous techniques.

Riedle *et al.* [30] consider pairwise comparisons of fitted models based on Kullback–Leibler Discrepancy using bootstrap samples. Additionally, they present an interesting connection between their proposed method and the likelihood ratio test p-value which provides a new way of thinking the likelihood ratio test p-value.

Oliveira *et al.* [26] present an application of prime base factorial fixed effects models to situations where the number of replicates for each treatment is not known in advance. It was assumed that the number of replicates by treatment corresponds to a realization of a random variable with geometric distribution. The results obtained through the simulation studies show the relevance of this approach. In particular, a balanced factorial model was considered.

Marques and Coelho [21] present a precise approximation method for computing the exact null distribution of the likelihood ratio test (LRT) statistic for testing equality of covariance matrices of several normal populations with complicated block structure (which can include the sphericity structure, the circular structure, the compound symmetry structure, and/or the unspecified structure of specific diagonal blocks). As mentioned by the authors, this type of hypotheses may be further generalized, e.g. for testing the equality of some of the diagonal blocks or simultaneously testing the equality of mean vectors of the populations, but for a matter of simplicity this was not included in the paper. Since the likelihood ratio statistic is bounded between 0 and 1 its distribution is completely specified by its moments. The suggested approximation is based on deriving the exact and the approximate characteristic function of the (negative) log-transformed LRT statistic (derived from the moments of the LRT null distribution) and subsequent application of the near-exact approximation based on using a mixture of generalized integer gamma (GIG) distributions and/or a mixture of generalized near-integer gamma (GNIG) distributions, a methodology originally suggested and developed by the authors. The precision of the suggested approximation method is assessed for several specific situations (with numerical studies and simulations) together with an analysis of real data set in order to illustrate the applicability of the suggested test.

Nikitin and Ragozin [22] propose two new goodness-of-fit tests for the logistic distribution with arbitrary location parameter, based on a random shift type characterizations. The authors derived limiting distributions of the test statistics under the null hypothesis, and calculated local Bahadur efficiencies against some common close alternatives. New tests are much more efficient than those previously proposed.

Barbu *et al.* [5] introduce a general class of distributions for independently but not identically distributed random variables, which is claimed to be closed under extrema. The parameter which generates the class is assumed to be time-varying, which underlines relation to parameter dependence models of type [14]. The maximum likelihood estimation is discussed with real and simulated data illustrations.

In [10] a logistic regression model is applied to credit scoring data from a given Portuguese financial institution to evaluate the default risk of consumer loans. The authors applied a logistic regression model to predict the default credit risk using data from consumer loans in a banking institution in Portugal. It was found that the risk of default increases with the loan spread, loan term and age of the customer, but decreases if the customer owns more credit cards. Clients receiving the salary in the same banking institution of the loan have less chances of default and clients in the lowest income tax echelon have more propensity to default.

A linear mixed model whose variance-covariance matrix is a linear combination of known pairwise orthogonal projection matrices that add to the identity matrix, is a model with orthogonal block structure (OBS). OBS have estimators with good behavior for estimable vectors and variance components, moreover it may be interesting that the least squares estimators give the best linear unbiased estimators, for estimable vectors. Santos *et al.* [32] achieved that, requiring commutativity between the orthogonal projection matrix, on the space spanned by the mean vector, and the orthogonal projection matrices involved in the expression of the variance-covariance matrix. They present a commutativity condition, resorting to a special class of matrices, named U-matrices.

Sharifipanah *et al.* [33] propose an extension of the uni/bimodal two-piece skew-normal (TN) distribution [18], in order to overcome its main drawback relative to the symmetrical behavior on both sides from the origin of the probability density function (pdf). The transformation used by the authors seems to preserve the bimodality of the pdf and introduces an asymmetric behavior on the two sides around the zero of the pdf.

The COVID-19 period interfaced this Editorial and some of refereeing processes has been slowed down. The inspired start of Editorial was written in May 2018 in Basílica de Nossa Senhora do Rosário de Fátima, in Fátima, Portugal. The last words on 27 August 2020 in Enns, Austria.

Acknowledgments

We acknowledge the great support, patience, professional advises and practical help from the Editor-in-Chief Jie Chen who enabled us to conceptualize these special issues. We are also indebted to Komal Patel, Justin Robinson, and Rahul Unni for their editorial assistance. Milan Stehlik's special thanks to Silvia for her support of my editorial work on these special issues during 2018–2020. This work was also supported by the Bulgarian National Science Funds under the bilateral projects Bulgaria – Austria, 2016–2019, Feasible statistical modeling for extremes in ecology and finance, Contract number 01/8, 23/08/2017 and WTZ Project BG 09/2017.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was also supported by the Bulgarian National Science Funds under the bilateral projects Bulgaria – Austria, 2016–2019, Feasible statistical modeling for extremes in ecology and finance, Contract number 01/8, 23/08/2017 and WTZ Project BG 09/2017.

References

[1] E.A. Ahmed, Z.A. Alhussain, M.M. Salah, H.H. Ahmed, and M.S. Eliwa, Inference of progressively type-II censored competing risks data from chen distribution with an application, J. Appl. Stat. 47 (2020), pp. 2492–2524.

[2] P. Antunes, S.S. Ferreira, D. Ferreira, C. Nunes, and J.T. Mexia, *Estimation in additive models and ANOVA-like applications*, J. Appl. Stat. 47 (2020), pp. 2374–2383.

[3] B.C. Arnold and M.A. Arvanitis, *On bivariate pseudo-exponential distributions*, J. Appl. Stat. 47 (2019), pp. 2299–2311.

[4] J. Arrué, R.B. Arellano-Valle, H.W. Gómez, and V. Leiva, *On a new type of Birnbaum-Saunders models and its inference and application to fatigue data*, J. Appl. Stat. 47 (2019), pp. 2690–2710.

[5] V.S. Barbu, A. Karagrigoriou, and A. Makrides, *Statistical inference for a general class of distributions with time-varying parameters*, J. Appl. Stat. 47 (2020), pp. 2354–2373.

[6] J. Beran, D. Schell, and M. Stehlík, *The harmonic moment tail index estimator: asymptotic distribution and robustness*, Ann. Inst. Stat. Math. 66 (2014), pp. 193–220.

[7] M.F. Brilhante, M.I. Gomes, and D. Pestana, *A simple generalization of the Hill estimator*, Comput. Stat. Data Anal. 57 (2013), pp. 518–535.

[8] K.D. Carter and J.E. Cavanaugh, *Best-subset model selection based on multitudinal assessments of likelihood improvements*, J. Appl. Stat. 47 (2019), pp. 2384–2420.

[9] M. Cullan, S. Lidgard, and B. Sterner, *Controlling the error probabilities of model selection information criteria using bootstrapping*, J. Appl. Stat. 47 (2019), pp. 2565–2581.

[10] E.C. Silva, I.C. Lopes, A. Correia, and S. Faria, *A logistic regression model for consumer default risk*, J. Appl. Stat. 47 (2020), pp. 2879–2894.

[11] P. Economou, G. Tzavelas, and A. Batsidis, *Robust inference under r-size-biased sampling without replacement from finite population*, J. Appl. Stat. 47 (2020), pp. 2808–2824.

[12] Z. Fabián and M. Stehlík, *On robust and distribution sensitive Hill like method*, IFAS Res. Report 43, 2009.

[13] L.J. Fermín and J. Lévy-Véhel, *Variability and singularity arising from a piecewise-deterministic Markov process applied to model poor patient compliance in the multi-IV case*, J. Appl. Stat. 47 (2020), pp. 2525–2545.

[14] J.K. Filus and L.Z. Filus, *On some new classes of multivariate probability distributions*, Pak. J. Stat. 22 (2006), pp. 21–42.

[15] N. Flournoy and A.P. Oron, *Bias induced by adaptive dose-finding designs*, J. Appl. Stat. 47 (2019), pp. 2431–2442.

[16] M. Giacomazzo and Y. Kamarianakis, *Bayesian estimation of subset threshold autoregressions: short-term forecasting of traffic occupancy*, J. Appl. Stat. 47 (2020), pp. 2658–2689.

[17] D.P. Gomes and M.M. Neves, *Extremal index blocks estimator: the threshold and the block size choice*, J. Appl. Stat. 47 (2020), pp. 2846–2861.

[18] H.J. Kim, *On a class of two-piece skew-normal distributions*, Statistics 39 (2005), pp. 537–553.

[19] M. Li, Q. Guo, W.J. Zhai, and B.Z. Chen, *The linearized alternating direction method of multipliers for low-rank and fused LASSO matrix regression model*, J. Appl. Stat. 47 (2020), pp. 2623–2640.

[20] O.S. Makinde, *On rank distribution classifiers for high-dimensional data*, J. Appl. Stat. 47 (2020), pp. 2895–2911.

[21] F.J. Marques and C.A. Coelho, *Testing simultaneously different covariance block diagonal structures – the multi-sample case*, J. Appl. Stat. 47 (2020), pp. 2765–2784.

[22] Ya.Yu. Nikitin and I.A. Ragozin, *Goodness-of-fit tests for the logistic location family*, J. Appl. Stat. 47 (2020), pp. 2610–2622.

[23] M.J. Nooghabi, *Process capability indices in normal distribution with the presence of outliers*, J. Appl. Stat. 47 (2020), pp. 2443–2478.

[24] C. Nunes, E. Moreira, S.S. Ferreira, D. Ferreira, and J.T. Mexia, *Considering the sample sizes as truncated Poisson random variables in mixed effects models*, J. Appl. Stat. 47 (2019), pp. 2641–2657.

[25] A.N. Oliveira, R. Menezes, S. Faria, and P. Afonso, *Mixed-effects modelling for crossed and nested data: an analysis of dengue fever in the state of Goiás, Brazil*, J. Appl. Stat. 47 (2020), pp. 2912–2926.

[26] S. Oliveira, C. Nunes, E. Moreira, M. Fonseca, and J.T. Mexia, *Balanced prime basis factorial fixed effects model with random number of observations*, J. Appl. Stat. 47 (2019), pp. 2737–2748.

[27] V. Paulaskas and M. Vaičiulis, *On the improvement of Hill and some others estimators*, Lith. Math. J. 53 (2013), pp. 336–355.

[28] H. Penalva, M.I. Gomes, F. Caeiro, and M.M. Neves, *Lehmer's mean-of-order-p extreme value index estimation: a simulation study and applications*, J. Appl. Stat. 47 (2019), pp. 2825–2845.
[29] R.A. Peterson and J.E. Cavanaugh, *Ordered quantile normalization: a semiparametric transformation built for the cross-validation era*, J. Appl. Stat. 47 (2019), pp. 2312–2327.
[30] B. Riedle, A.A. Neath, and J.E. Cavanaugh, *Reconceptualizing the p-value from a likelihood ratio test: a probabilistic pairwise comparison of models based on Kullback-Leibler discrepancy measures*, J. Appl. Stat. 47 (2020), pp. 2582–2609.
[31] R. Romero, J.M. Pavía, J. Martín, and G. Romero, *Assessing uncertainty of voter transitions estimated from aggregated data. Application to the 2017 French presidential election*, J. Appl. Stat. 47 (2020), pp. 2711–2736.
[32] C. Santos, C. Nunes, C. Dias, and J.T. Mexia, *Models with commutative orthogonal block structure: a general condition for commutativity*, J. Appl. Stat. 47 (2020), pp. 2421–2430.
[33] N. Sharifipanah, R. Chinipardaz, and G.A. Parham, A new class of weighted bimodal distribution with application to gamma-ray burst duration data, J. Appl. Stat. 47 (2020), pp. 2785–2807.
[34] I. Silva and H. Alonso, *New developments in the forecasting of monthly overnight stays in the North Region of Portugal*, J. Appl. Stat. 47 (2020), pp. 2927–2940.
[35] I. Silva, M.E. Silva, and C. Torres, *Inference for bivariate integer-valued moving average models based on binomial thinning operation*, J. Appl. Stat. 47 (2020), pp. 2546–2564.
[36] A. Tchorbadjieff and P. Mayster, *Models induced from critical birth–death process with random initial conditions*, J. Appl. Stat. 47 (2020), pp. 2862–2878.
[37] S.D. Tomarchio and A. Punzo, *Dichotomous unimodal compound models: application to the distribution of insurance losses*, J. Appl. Stat. 47 (2020), pp. 2328–2353.
[38] A.Y. Touré, S. Dossou-Gbété, and C.C. Kokonendji, *Asymptotic normality of the test statistics for the unified relative dispersion and relative variation indexes*, J. Appl. Stat. 47 (2020), pp. 2479–2491.
[39] V. Witkovský, *Computing the exact distribution of the Bartlett's test statistic by numerical inversion of its characteristic function*, J. Appl. Stat. 47 (2019), pp. 2749–2764.