From the Editor

The March issue of *Statistics in Transition new series* presents readers with a set of eleven articles written by twenty-four authors from nine countries (in order of appearance): USA, Poland, Nepal, Indie, Nigeria, Indonesia, Iran, Algeria, and Saudi Arabia. This volume opens with an *invited article* by Piotr Kokoszka, Mengting Lin, HaonanWang, and Stephen Hayne.

Invited paper

In the first paper, *Statistical risk quantification of two-directional internet traffic flows*, Piotr Kokoszka, Mengting Lin, HaonanWang, and Stephen Hayne discuss recent developments in statistical methodology for the quantification of risk of source-destination pairs in an internet network within the framework of functional data analysis and copula modeling. It was summarized in the form of computational algorithms that use bidirectional source-destination packet counts as input. The usefulness of the proposed approach was evaluated by an application to real internet traffic flows and via a simulation study. The performance, and relative performance, of the two algorithms using simulated data that has certain features of real data sets, but also certain characteristics that are known targets, were assessed.

Research articles

Henryk Gurgul's and Robert Syrek's article *Mutual information between Polish subindexes – the use of copula entropy around the time of the COVID-19 pandemic* demonstrates application of the copula theory to describe the dependence structure between variables, while the information theory provides the tools necessary to measure the uncertainty associated with these variables. What both theories have in common is copula entropy, which is strictly related to mutual information. The findings of this study of the dependence of the (sub)indexes of the Polish stock market during the pandemic period seem to be useful not only to investors in Poland, but in other countries as well, especially in Central Europe, in making investment decisions. The results of calculating the interdependencies between WIG, sectoral indexes and among sectoral indexes of the Polish economy using copula entropy and Pearson's correlation are quite different. The next article, by **Arjun Kumar Gaire** and **Yogendra Bahadur Gurung**, entitled *Skew Log-logistic distribution: properties and application*, presents a novel threeparameter skew-log-logistic distribution. It starts from the development of a new random variable based on Azzalini and Capitanio's (2013) proposition, including various statistical properties of this distribution. A maximum likelihood method for estimating the distribution's parameters is employed. The density function exhibits unimodality with heavy right tails, while the hazard function exhibits rapid increase, unimodality, and slow decrease, resulting in a right-skewed curve. Furthermore, four real datasets are utilized to assess the applicability of this new distribution. The AIC and BIC criteria are employed to assess the goodness of fit, revealing that the new distribution offers greater flexibility compared to the baseline distribution.

In the paper *A* chain ratio-type exponential estimator for population mean in double sampling, by Nirupama Sahoo and Sananda Kumar Jhankar, an efficient ratio-type exponential estimator for estimating the population mean by incorporating two auxiliary variables in two-phase (double) sampling is proposed. The bias and the mean square error of the presented estimator have been obtained up to the first order of approximation. The new estimator offers more precision in comparison to other competing estimators, theoretically as well as empirically, by considering a known value of some population parameter.

Abimibola V. Oladugba's and Oluwagbenga T. Babatunde's paper, *Improved* calibration estimation of population mean in stratified sampling using two auxiliary variables, discusses possibilities to improve the standard estimator of the population mean in a stratified sampling through calibration estimation approach using two auxiliary variables. A simulation study was carried out to evaluate the performance and efficiency of an estimator with respect to three estimators proposed in the literature for estimating the population mean in a stratified sampling (using two auxiliary variables). The proposed estimator has the least absolute relative bias and mean square error for all the cases under consideration. The results showed that the new estimator proved to be more efficient than the three existing estimators considered.

Agnieszka Palma and Dorota Kałuża-Kopias in the paper Inter-voivodship migration in Poland in the 2000–2020 period based on Markov chain analysis deal with the scale and directions of inter-voivodship migration in Poland in selected years of the 2000–2020 period. The study focused on permanent residence migration and aimed to identify areas of migration attractiveness and migration catchment voivodships. The application of the Markov Chain allowed for evaluation of the population flow between voivodships. The results of the study indicate that the most favourable situation remains in the Mazowieckie voivodship, which is considered the most attractive area for people from other regions of the country – mainly from the

Lubelskie, Podlaskie, and Łódzkie voivodships, and to a lesser extent from the Świętokrzyskie and Warmińsko-Mazurskie voivodships. The approach used allowed for determining the properties of the transition probability matrix as well as stationary probability in order to characterise the mechanism of inter-voivodship migrations in the years 2000, 2010 and 2020.

In the next article, *Implementation of K-Nearest Neighbor using the oversampling technique on mixed data for the classification of household welfare status*, Nur **Mutmainnah Djafar** and Achmad Fauzan took up the task to classify the household welfare status in Kulon Progo using the K-Nearest Neighbor (KNN) method. Since imbalance was found between the poor and non-poor categories, an oversampling technique was employed. Imbalanced data affect classification, especially when it comes to predicting the results of the classification. The following oversampling techniques were employed: Random Oversampling (RO), the Adaptive Synthetic (ADASYN) and the Synthetic Minority Oversampling Technique (SMOTE). It was found that, of the three techniques, RO was the most efficient with k = 5, which yielded the best performance in terms of sensitivity, specificity, the G-mean. Therefore, it can be concluded that the classification model performed well enough to classify household welfare status, especially among the poor (minority group).

The paper **On Bayesian inference of reliability parameter in Burr-type XII model based on imprecise data: a survey on fuzzy modelling** by **Iman Makhdoom** and **Abbas Pak** examines the classical and Bayesian inference procedures for the BT XII distribution parameters, including the corresponding reliability parameter when the available data are described regarding fuzzy numbers. In this context, the authors considered three priors as noninformative prior, i.e. $a_1 = b_1 = a_2 = b_2 = 0$, less informative prior, i.e. $a_1 = b_1 = a_2 = b_2 = 0.01$, and informative prior, i.e. $a_1 = b_1 = a_2 =$ $b_2 = 4$. Considering the criterion MSE for all methods, with increasing *n*, the estimates are improved. The performance of the Bayes estimates with assumptions of noninformative prior and less informative prior regarding AVs and MSEs is almost identical. The simulation study for all methods shows that the estimate of *R* is satisfactory, even for samples with sizes small and moderate. Using the NR or EM algorithms for the computation of MLEs gives similar estimation results.

Joanna Kisielińska's article presents *Estimation of quantiles with the exact bootstrap method*. The estimation uses the bootstrap method in the so-called exact. Three bootstrap estimators were used: two of them based on one order statistic, and the third on a linear combination of two order statistics (for an integer). It has been shown that there is no general form of the distribution of the exact bootstrap estimator based on two order statistics. However, it is possible to calculate such a distribution – the algorithm that performs such a task is presented. The bootstrap confidence intervals

were constructed using the exact percentile method. It has been shown that if the estimator is based on a single order statistic, it is known in advance which elements of the primary sample are the limits of the confidence intervals (so there is no need to resample). The intervals determined by the exact percentile method were compared with those constructed using other methods.

Research Communicates and Letters

Thara Belhamra, Halim Zeghdoudi, and Vinoth Raman analyse *Reliability for Zeghdoudi distribution with an outlier, fuzzy reliability and application*. This study focuses on estimating reliability P[Y < X], where *Y* has a Zeghdoudi distribution with parameter *a*, *X* has a Zeghdoudi distribution with one outlier present and parameter *c*, and the remaining (n - 1) random variables are from a Zeghdoudi distribution with parameter *b*, in order for *X* and *Y* to be independent. Several findings of a simulation study and the maximum likelihood estimate of *R* are provided. Some results related to fuzzy dependability were also presented. In order to demonstrate the adaptability of the Zeghdoudi distribution, the authors use real data on the survival times (in days) of 72 Algerian people who were infected with coronaviruses, and then compared the outcomes with those of other distributions. Studies have been done on the maximum likelihood estimator for *R* and fuzzy dependability.

Piyush Kant Rai's and **Sweta Singh's** paper, *Composite estimators for domain estimation and sensitivity performance interval of their weights* presents some composite estimators based on various combinations of two different existing estimators. To account for the absence of optimum weights, the sensitivity performance intervals for weights with respect to the proposed composite estimators were obtained and the sensible values of the involved weights have been determined. The aim of this procedure was to confine the superiority for different composite combinations – i.e. simple direct vs. direct ratio, simple direct vs. synthetic ratio and direct ratio vs. synthetic ratio composite estimators – as compared to the existing estimators. It was concluded that the composite estimators for the weights lie in the sensitivity performance intervals that are less varying in terms of MSE. The outcomes of the study will be useful to develop efficient composite estimators for the estimation domain in general, and for small area estimation in particular.

Włodzimierz Okrasa Editor

© Włodzimierz Okrasa. Article available under the CC BY-SA 4.0 licence 💽 💓 💿