ISSN 1645-6726

REVSTAT

Statistical Journal

vol. 21 - n. 4 - October 2023

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT-Statistical Journal, vol.21, n. 4 (October 2023)

vol.1, 2003- . - Lisbon : Statistics Portugal, 2003- .

Continues: Revista de Estatística = ISSN 0873-4275.

ISSN 1645-6726 ; e-ISSN 2183-0371

Editorial Board (2019-2023)

Publisher – *Statistics Portugal*

Layout-Graphic Design – *Carlos Perpétuo* | **Cover Design*** – *Helena Nogueira*

Edition - 130 copies | **Legal Deposit Registration** - 191915/03 | **Price** [VAT included] - € 9,00

**image*: stain glass window by Abel Manta (1888-1982)

INDEX

Abstracted/indexed in: *Journal Citation Reports™ – JCR (Clarivate); DOAJ – Directory of Open Access Journals; Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; SCImago Journal & Country Rank; Scopus®.*

REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 447–468
https://doi.org/10.57805/revstat.v21i4.425

A Note on the Stochastic EM Algorithm Based on Left Truncated Right Censored Data from Burr XII Distribution

Author: Debanjan Mitra
– Operations Management, Quantitative Methods, and Information Systems, Indian Institute of Management Udaipur, India
debanjan.mitra@iimu.ac.in

Received: June 2021 Revised: January 2022 Accepted: January 2022

Abstract:

- The Burr XII distribution is a flexible model for failure-time data. A very general and commonly observed structure for failure-time data involves left truncation and right censoring. In this article, modelling of left truncated right censored failure-time data by the Burr XII distribution is discussed. The steps of the stochastic expectation maximization algorithm, which is a useful technique of estimation for incomplete data structures, are developed to estimate the model parameters of the Burr XII distribution. The Newton-Raphson method, which is a direct method of obtaining maximum likelihood estimates by optimizing the observed likelihood is also used. The two methods of inference are assessed and compared through a Monte Carlo simulation study. Discussions of the inferential methods are extended to the cases of a three-parameter Burr XII model, and a covariate-included model. An illustrative example based on real data is provided. Finally, an application of the inferential results to a prediction issue is discussed with an illustration.

Keywords:

- *lifetime distribution; censoring; truncation; stochastic expectation maximization algorithm; prediction.*

AMS Subject Classification:

- 62F10, 62H10.

1. INTRODUCTION

The Burr system of distributions was proposed by Burr, for modelling a wide variety of data observed in real life [8]. Subsequently, among this system of distributions, the Burr XII distribution received special attention by researchers, see Tadikamalla [22] and the references therein. The Burr XII distribution is given by the cumulative distribution function (CDF)

$$F_Y(y;\alpha,\beta) = 1-(1+y^{\beta})^{-\alpha}, \quad y>0, \tag{1.1}$$

with corresponding probability density function (PDF)

$$f_Y(y;\alpha,\beta) = \alpha\beta y^{\beta-1}(1+y^{\beta})^{-(\alpha+1)}, \quad y>0. \tag{1.2}$$

with $\alpha > 0$ and $\beta > 0$, both of which are shape parameters. The Burr XII distribution gets its flexibility through the shape parameters β and α. Figures 1 and 2 display the PDF of the Burr XII distribution for different values of β (keeping α fixed), and α (keeping β fixed), respectively. Note that for $\beta \leq 1$, the distribution is L-shaped, while it is unimodal for $\beta > 1$, as observed by Beirlant *et al.* [6].

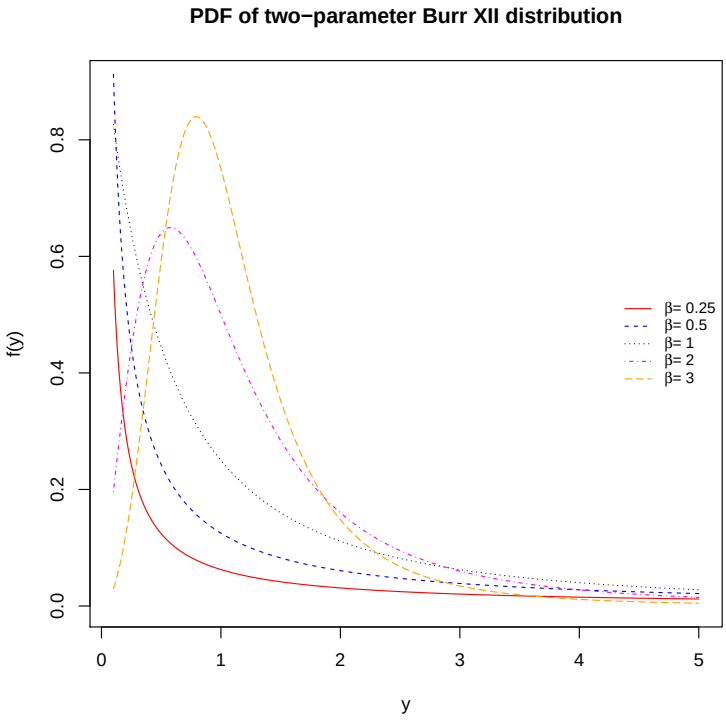


Figure 1: Density function of the Burr XII model for different values of β when $\alpha = 1$.

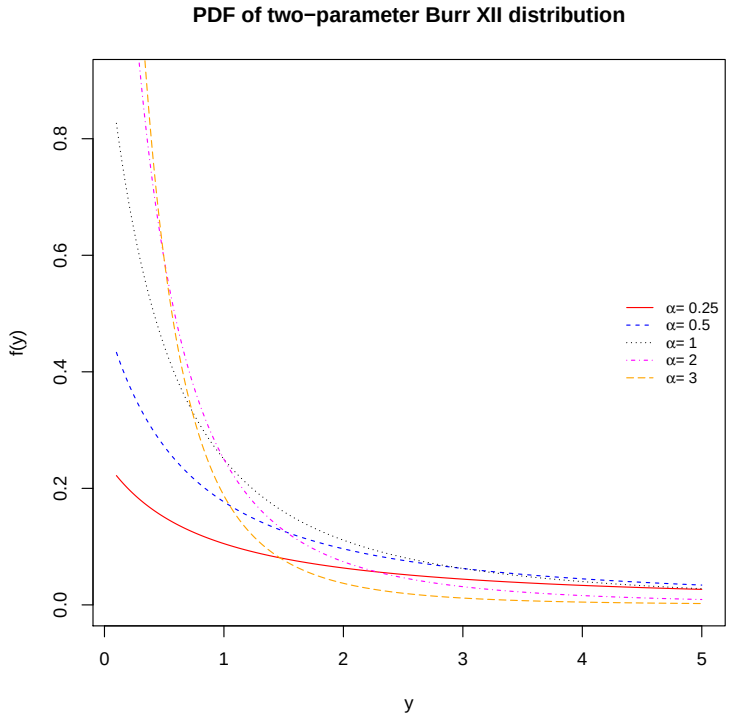


Figure 2: Density function of the Burr XII model for different values of α when $\beta = 1$.

Zimmer *et al.* [24] advocated for the use of the Burr XII distribution as an alternative to lognormal and Weibull distributions, and mentioned the many advantages this model has. The log-logistic distribution, which is another important lifetime model, is a special case of the Burr XII distribution; this also is a motivation to use the Burr XII distribution to model failure-time data [24].

Despite its great flexibility, however, the Burr XII distribution was relatively less used in survival and reliability studies, especially compared to the well-known models like Weibull, gamma etc. Recently, some researchers have used the Burr XII model in the context of failure-time data. For example, Soliman [21] modelled progressively type-II censored data by the Burr XII distribution. Silva *et al.* [19] and Silva *et al.* [20] discussed regression models for the Burr XII distribution based on censored data.

Left truncated right censored (LTRC) data are commonly observed in studies involving lifetimes of experimental units [14]. For example, in many reliability and survival experiments, the main event of interest is the failure of experimental units. Due to practical time constraints on sample collection in such experiments, the observed samples are often either left truncated, or right censored, or both. In medical studies, for example, groups of subjects are often followed over time for observing the occurrence of certain disease or event such as death. LTRC data arise naturally in situations of this type. Another example of LTRC data may be found in oraganisational or social science studies where start-up businesses are observed over a time-window during which they may fail.

As left truncation and right censoring are quite commonly observed features among data arising out of survival and reliability studies, it is of natural importance to develop

inferential methods for the Burr XII distribution based on LTRC data, especially as due to its flexible nature the Burr XII model has been posed as a general purpose model for failure-time data by Zimmer *et al.* [24]. To use the Burr XII distribution as a general purpose failure-time model, it is important to develop inferential methods for the model based on LTRC data which is one of the most common and general structures among incomplete data formats in lifetime studies. However, so far, no researcher has attempted modelling LTRC data by using the Burr XII distribution.

In this article, we discuss modelling LTRC failure-time data by the Burr XII distribution in detail. First, we consider the two-parameter version of the Burr XII distribution, as it is the more frequently used version. The stochastic expectation maximization (St-EM) algorithm has emerged as a stable, efficient, and convenient method for parameter estimation for incomplete data problems. For estimating the parameters of the Burr XII model, we develop the steps of the St-EM algorithm based on LTRC data. We discuss two approaches for constructing confidence intervals, one of them being based on an adaptation of the missing information principle of Louis [15], and the other being based on parametric bootstrap approach. For comparison purposes, we also use the Newton-Raphson (NR) method which is a direct approach to obtain maximum likelihood estimates by optimizing the observed likelihood function. Through detailed Monte Carlo simulations, we study the performance of the proposed methods of inferences. Further, we extend our discussion of inferential methods to the cases for a covariate-included model, and the three-parameter Burr XII distribution. These are the main contributions of these paper.

The article is organized as follows. A brief introduction to LTRC data is provided in Section 2. The St-EM algorithm for the two-parameter Burr XII model based on LTRC data is discussed in detail in Section 3; both point and interval estimation procedures are presented. The direct method of obtaining MLEs is presented in this section too. This section also contains a discussion of inferential methods for a covariate-included model. The detailed results of the numerical experiments are presented in Section 4. Then, in Section 5, discussion of the St-EM algorithm is extended to a three-parameter Burr XII model with a scale parameter in addition to the two shape parameters α and β. This three-parameter model is also used in failure-time data modelling [24]. An application of the inferential methods in predicting the expected number of failures in a future time interval is presented in Section 6. Along with an estimate of the expected number of failures in a future time interval, we provide asymptotic confidence intervals for this expected number of failures. This is of direct practical relevance, as in many situations like maintenance, the researcher may want to have an estimate of the expected number of future failures during a certain time period. In Section 7, a numerical illustration based on a real data is provided. Finally, mentioning some future directions of research in this area, the paper is concluded with some remarks in Section 8.

2. LEFT TRUNCATED RIGHT CENSORED DATA

Hong *et al.* [13] analyzed LTRC data obtained from an electrical industry in the US. Following the setup used by Hong *et al.* [13], Balakrishnan and Mitra [1, 2, 3, 4, 5] discussed the EM algorithm based on LTRC data for some commonly used failure-time models such as lognormal, Weibull, gamma, and generalized gamma; see also Mitra *et al.* [16].

Consider a life-test involving n industrial units. Let T denote the underlying failure-time variable. Let L and R denote points of left truncation and right censoring, respectively; that is, we suppose that the study starts at time point L and continues till time point R. Some units start operating before L, while some start after L. No information are available for units that fail before L, making the data left truncated. Some units may not have failed when the study ends at R, and those units become right censored at R. A unit that starts operating before L, has to live through a threshold time, say κ_L, before its failure become an observable event. We call κ_L the left truncation time. An indicator variable ν indicates whether a unit is left truncated or not; for a left truncated unit ν is 0, otherwise it is 1. Note that failures are observable only in the window from L to R. As a result, for each operating unit, there is a time κ_R depending on the starting point of the unit, such that the unit is right censored if $T > \kappa_R$. As different units may have different starting points, values of κ_L and κ_R may be differ from unit to unit. Thus in effect, for each unit we can define the observed lifetime as $Y = \text{Min}(T, \kappa_R)$, provided $Y > \kappa_L$. Let δ denote an indicator variable for censoring; δ is 0 for a right censored unit, and 1 otherwise.

For subsequent formulation of the problem, let S_1 and S_2 denote index sets for untruncated and truncated units, respectively, that is,

$$S_1 = \{i : \nu_i = 1\}, \quad \text{and} \quad S_2 = \{i : \nu_i = 0\},$$

where ν_i is the truncation indicator for the $i-$th unit, $i = 1, ..., n$. Incorporating the censoring indicator δ, we define the index sets

$$S_{11} = \{i : i \in S_1, \delta_i = 1\}, \quad S_{10} = \{i : i \in S_1, \delta_i = 0\},$$

$$S_{21} = \{i : i \in S_2, \delta_i = 1\}, \quad S_{20} = \{i : i \in S_2, \delta_i = 0\}.$$

We further define S_{cen}:

$$S_{cen} = S_{10} \cup S_{20}.$$

We assume that the underlying lifetime T follows the Burr XII distribution with parameters α and β, i.e., $T \sim Burr(\alpha, \beta)$.

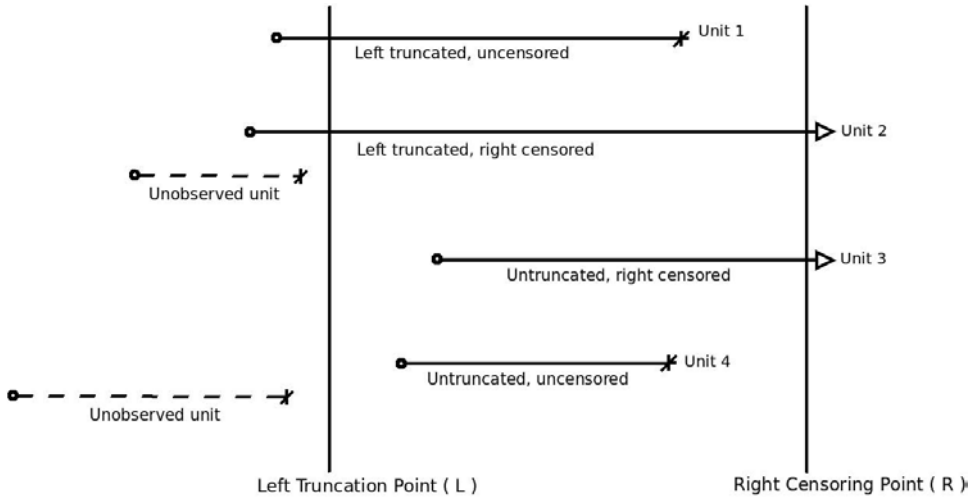


Figure 3: Illustration of LTRC Data.

In Figure 3, we present an illustration of the structure of LTRC data we consider here. We would also like to point out that this is a very general structure that can accommodate units with different combinations of truncation and censoring: left truncated and right censored, left truncated and uncensored, untruncated and right censored, and untruncated and uncensored. This enhances the scope of this model greatly, to be applied to a wide array of observational studies involving failure-times.

3. INFERENCE VIA THE STOCHASTIC EM ALGORITHM

The St-EM algorithm has emerged as a strong tool for analyzing incomplete data. Compared to the traditional EM algorithm, the St-EM algorithm has some distinct advantages. For example, in the EM algorithm, one needs to analytically calculate the conditional expectation of the complete data log-likelihood given the observed data and the current parameter values. Analytical calculation of this conditional expectation may be very difficult, or even intractable, for complex problems. However, in St-EM algorithm, one does not require the analytic calculation of the conditional expectations unlike the EM algorithm. Moreover, in the EM algorithm, the sequence of estimated parameters may get trapped in saddle points depending on the nature of the likelihood surface. But in the St-EM algorithm, due to its stochastic nature, one does not encounter such a problem [23].

The St-EM algorithm has been used for various incomplete data problems in statistical literature; see [9], [23], for example. The asymptotic properties of the St-EM algorithm have been explored by Nielsen [17], among others. Bordes and Chauveau [7] and Ng and Ye [18] recommended the use of the St-EM algorithm for LTRC data.

In St-EM algorithm, for each censored failure-time, a randomly drawn observation from an appropriate conditional distribution is obtained given the observed data and the current value of the parameter. By replacing all censored failure times by such randomly drawn observations, a pseudo-complete dataset is obtained and the pseudo-complete likelihood function is constructed. Then, the pseudo-complete likelihood is optimized to obtain updated parameter estimates. The whole process is then iterated large number of times, to get a sequence of estimates corresponding to each stage of the algorithm. Finally, after discarding some initial values of the estimates for burn-in, the remaining values are averaged to obtain the final estimates.

Note that corresponding to the underlying failure-time variable T, the observed data can be written as

$$\boldsymbol{t} = \boldsymbol{\Upsilon} \cup \boldsymbol{\Gamma},$$

where $\boldsymbol{\Upsilon} = \{t_i : \delta_i = 1\}$ and $\boldsymbol{\Gamma} = \{t_i : \delta_i = 0\}$ contain the observed and right censored failure-times, respectively. For each unit in $\boldsymbol{\Gamma}$, we generate a random observation from the conditional distribution

$$\begin{aligned} f_{T_i|T_i>y_i}(t_i|t_i > y_i; \boldsymbol{\theta}) &= \frac{f_T(t_i;\boldsymbol{\theta})}{1 - F_T(y_i;\boldsymbol{\theta})} \\ &= \alpha\beta t_i^{\beta-1}(1+t_i^{\beta})^{-(\alpha+1)}(1+y_i^{\beta})^{\alpha}, \quad t_i > y_i, \end{aligned} \tag{3.1}$$

where $\boldsymbol{\theta} = (\alpha, \beta)$, and y_i is the censored failure-time. By replacing the censored failure-times

by these randomly drawn observations, we obtain the pseudo-complete data

$$\boldsymbol{t}_{PC} = \boldsymbol{\Upsilon} \cup \boldsymbol{\Gamma}_{PC}.$$

In the EM algorithm, the complete data likelihood is constructed considering the situation where there would be no incompleteness in the data. In case of the St-EM algorithm, having imputed the censored lifetimes by randomly generated observations from the above conditional distributions, the pseudo-complete data $\boldsymbol{t}_{PC}$ will now be used in a similar fashion, i.e., as if there was no censoring in the data. For a unit that belongs to the untruncated group, the contribution to the likelihood would be $f_T(t_i;\boldsymbol{\theta})$; and for a unit that belongs to the left truncated group, the contribution would be $\frac{f_T(t_i;\boldsymbol{\theta})}{1-F_T(\kappa_{Li};\boldsymbol{\theta})}$. Therefore, by using the pseudo-complete data, the pseudo-complete likelihood is constructed as

$$\begin{aligned} L_{PC}(\boldsymbol{\theta}) &= \prod_{i\in S_1}\{f_T(t_i;\boldsymbol{\theta})\} \times \prod_{i\in S_2}\left\{\frac{f_T(t_i;\boldsymbol{\theta})}{1-F_T(\kappa_{Li};\boldsymbol{\theta})}\right\} \\ &= \prod_{i=1}^{n}\{\alpha\beta t_i^{\beta-1}(1+t_i^{\beta})^{-(\alpha+1)}\} \times \prod_{i\in S_2}\{(1+\kappa_{Li}^{\beta})^{\alpha}\}. \end{aligned} \tag{3.2}$$

The pseudo-complete log-likelihood function, given by

$$\begin{aligned} \log L_{PC}(\boldsymbol{\theta}) &= n(\log\alpha + \log\beta) + \sum_{i=1}^{n}\Big[(\beta-1)\log t_i - (\alpha+1)\log(1+t_i^{\beta})\Big] \\ &\quad + \alpha\sum_{i\in S_2}\log(1+\kappa_{Li}^{\beta}), \end{aligned} \tag{3.3}$$

which we shall denote by $Q_{PC}(\boldsymbol{\theta})$, essentially serves as the pseudo-Q function in this setup, where the Q-function in the traditional EM algorithm is defined as

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = E_{\boldsymbol{\theta}^{(k)}}[\log L_C(\boldsymbol{\theta})|\boldsymbol{\Gamma}],$$

with $\log L_C(\boldsymbol{\theta})$ as the complete data log-likelihood, and $\boldsymbol{\theta}^{(k)}$ as the available value of the parameter vector at the k-th stage of iteration.

To optimize $Q_{PC}(\boldsymbol{\theta})$, for fixed β, equating the first derivative of $Q_{PC}(\boldsymbol{\theta})$ with respect to α to zero we obtain

$$\alpha = \frac{n}{\sum_{i=1}^{n}\log(1+t_i^{\beta}) - \sum_{i\in S_2}\log(1+\kappa_{Li}^{\beta})} = \alpha(\beta). \tag{3.4}$$

Substituting (3.4) in (3.3), we obtain the pseudo-profile log-likelihood in β as

$$\begin{aligned} p_{PC}(\beta) &= n\log\beta + \beta\sum_{i=1}^{n}\log t_i - n\log\left\{\sum_{i=1}^{n}\log(1+t_i^{\beta}) - \sum_{i\in S_2}\log(1+\kappa_{Li}^{\beta})\right\} \\ &\quad - \sum_{i=1}^{n}\log(1+t_i^{\beta}). \end{aligned} \tag{3.5}$$

Note that maximizing $p_{PC}(\beta)$ in (3.5) is a one-dimensional optimization problem, and can be achieved by using any routine optimizer of a statistical software, for example, the maxNR() function in the "maxLik" package [12] available in R software.

The following algorithm implements the St-EM algorithm for obtaining estimates of the model parameters.

ALGORITHM 1: At the k-th stage of the algorithm:

Stochastic Expectation (St-E) step:

STEP 1: The available parameter value is $\boldsymbol{\theta}^{(k)} = (\alpha^{(k)}, \beta^{(k)})$;

STEP 2: Replace each unit in $\boldsymbol{\Gamma}$ by generating observations from $f_{T_i|T_i>y_i}(t_i|t_i > y_i; \boldsymbol{\theta}^{(k)})$ to obtain pseudo-complete data $\boldsymbol{t}_{PC}$;

STEP 3: With $\boldsymbol{t}_{PC}$ thus obtained in Step 2, construct $Q_{PC}(\boldsymbol{\theta}^{(k)})$ following (3.3);

Maximization (M) step:

STEP 4: Choose an initial value $\beta_{init}^{(k)}$;

STEP 5: Optimize $p_{PC}(\beta)$ in (3.5) to get $\widehat{\beta^{(k+1)}}$ subject to a tolerance level;

STEP 6: Using (3.4), calculate $\widehat{\alpha^{(k+1)}} = \alpha(\widehat{\beta^{(k+1)}})$;

STEP 7: With the updated estimate $\boldsymbol{\theta}^{(k+1)} = (\alpha^{(k+1)}, \beta^{(k+1)})$, go back to Step 2.

These steps are iterated N times to get a sequence of estimates $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$, ..., $\boldsymbol{\theta}^{(N)}$. After discarding first M of these estimates for burn-in, the remaining ones are averaged to get estimates of α and β. As mentioned in Ye and Ng [23], sufficiently large values of N and M must be chosen for very complex data, while the values as 1000 and 100, respectively may be good enough for most problems.

The algorithm starts with an initial value for the parameter vector $\boldsymbol{\theta}^{(0)} = (\alpha^{(0)}, \beta^{(0)})$. For such optimization problems to numerically estimate the parameters, moments estimates may be used as the initial values, provided the moments of the concerned distribution exist and are easily available, for example, in closed form expressions. For the Burr XII distribution, however, moments estimates based on left truncated data are not available in closed form. A practical solution to the problem of selecting initial values for the parameters α and β in this case would be to use a two-dimensional grid search approach. However, it may be noted here that use of a two-dimensional grid search approach followed by the St-EM algorithm will be computationally costly. In this work, for given sets of true values of the parameters α and β, we have tried different arbitrarily chosen initial values for the St-EM algorithm. And we have noticed that the St-EM algorithm as described above is reasonably robust to the choice of initial values. That is, the final estimates obtained from the algorithm by using different choices of initial values are quite close. In this connection, it may be noted here that the direct method of optimization, for example, the Newton-Raphson method, is heavily dependent on the choice of initial parameters in general.

3.1. Regression

In many applications, failure-times of experimental units depend on covariates. For example, failure-time of electrical machines may depend on temperature and humidity of the place of operation; failure-time of patients may depend on their respective demographic conditions etc. In view of this, it is of interest to consider a failure-time model that can accommodate relevant covariate information.

Regression models for Burr XII distribution have been considered by some authors. Beirlant *et al.* [6], in the context of a financial application involving portfolio segmentation, discussed different strategies for accommodating covariate information in the Burr XII model through the shape and scale parameters. Regression model for the log-Burr XII distribution was considered by Silva *et al.* [19], where covariate information was modelled as a linear function of the location parameter of the log-transformed model.

In this paper, we use an approach suggested in Beirlant *et al.* [6]. However, for parameter estimation, instead of using the observed likelihood based estimation approach which may not be computationally stable for complex models, we indicate the use of stochastic EM algorithm which is more reliable for its convergence.

We allow the shape parameter β in (1.1) to vary with covariates. Thus, when the $\boldsymbol{x}$ represent the vector of covariates, we assume the model

$$\beta(\boldsymbol{x}) = \exp(\boldsymbol{\gamma}'\boldsymbol{x}), \tag{3.6}$$

where $\boldsymbol{\gamma}$ is the vector of regression parameters. Under this assumption, our model for the failure-time variable Y becomes

$$Y_i|\boldsymbol{x}_i \sim Burr(\alpha, \beta_i), \quad \text{with} \quad \beta_i = \exp(\boldsymbol{\gamma}'\boldsymbol{x}_i), \quad i = 1, ..., n. \tag{3.7}$$

The conditional distributions for generating observations are given in this case by

$$\begin{aligned} f_{T_i|T_i>y_i}(t_i|t_i > y_i; \alpha, \boldsymbol{\gamma}, \boldsymbol{x}_i) \;&=\; \alpha \exp(\boldsymbol{\gamma}'\boldsymbol{x}_i) t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)-1} (1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})^{-(\alpha+1)} \\ &\quad \times (1 + y_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})^{\alpha}, \qquad t_i > y_i. \end{aligned} \tag{3.8}$$

Corresponding to the censored failure-times, given the covariates, observations are generated from (3.8). Based on the pseudo-complete data, the pseudo-complete likelihood function is

$$\begin{aligned} L_{PC}(\alpha, \boldsymbol{\gamma}) \;&=\; \prod_{i=1}^{n} \left\{ \alpha \exp(\boldsymbol{\gamma}'\boldsymbol{x}_i) t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)-1} (1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})^{-(\alpha+1)} \right\} \\ &\quad \times \prod_{i \in S_2} \left\{ (1 + \kappa_{Li}^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})^{\alpha} \right\}. \end{aligned} \tag{3.9}$$

The corresponding pseudo-complete log-likelihood function is given by

$$\begin{aligned} \log L_{PC}(\alpha, \boldsymbol{\gamma}) \;&=\; n \log \alpha + \sum_{i=1}^{n} \boldsymbol{\gamma}'\boldsymbol{x}_i + \sum_{i=1}^{n} (\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i) - 1) \log t_i \\ &\quad - (\alpha + 1) \sum_{i=1}^{n} \log(1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) + \alpha \sum_{i \in S_2} \log(1 + \kappa_{Li}^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}), \end{aligned} \tag{3.10}$$

which is also the pseudo-Q function, denoted by $Q_{PC}(\alpha, \boldsymbol{\gamma})$. Equating the first derivative of (3.10) with respect to α to zero, we have

$$\alpha = \frac{n}{\sum_{i=1}^{n} \log(1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) - \sum_{i \in S_2} \log(1 + \kappa_{Li}^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})} = \alpha(\boldsymbol{\gamma}). \tag{3.11}$$

Substituting (3.11) in (3.10), the profile-likelihood in $\boldsymbol{\gamma}$ is obtained as

$$\begin{aligned} p_{PC}(\boldsymbol{\gamma}) \;=\; & \sum_{i=1}^{n} \boldsymbol{\gamma}'\boldsymbol{x}_i + \sum_{i=1}^{n} \exp(\boldsymbol{\gamma}'\boldsymbol{x}_i) \log t_i - \sum_{i=1}^{n} \log(1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) \\ & - n \log \left\{ \sum_{i=1}^{n} \log(1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) - \sum_{i \in S_2} \log(1 + \kappa_{Li}^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) \right\}. \end{aligned} \tag{3.12}$$

The profile log-likelihood in the regression parameters $\boldsymbol{\gamma}$ can be maximized first using some numerical approach such as Newton-Raphson or Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Then, the α can be estimated using (3.11). An algorithm similar to Algorithm 1 can be easily constructed for this purpose.

3.2. Asymptotic confidence intervals

For obtaining asymptotic confidence intervals for the parameters, we use the missing information principle of Louis [15] that says

$$\text{Observed Information} = \text{Complete Information} - \text{Missing Information}. \tag{3.13}$$

For the traditional EM algorithm, Louis' principle is used to obtain the asymptotic variances of the estimates. For the St-EM algorithm also, an adaptation of the Louis' principle is possible, see Ye and Ng [23]. Mitra *et al.* [16] used the same approach in connection to the Lehmann family of distributions. The approach of Ye and Ng [23] is as follows.

Let $S(\boldsymbol{\theta}, \boldsymbol{t}_{PC})$ and $H(\boldsymbol{\theta}, \boldsymbol{t}_{PC})$ denote the first, and negative of the second derivatives of $Q_{PC}(\boldsymbol{\theta})$ given in (3.3) with respect to $\boldsymbol{\theta}$. Then, by the missing information principle, following Ye and Ng [23], the observed information matrix is given by

$$I(\boldsymbol{\theta}) = E[H(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}] - E[S^2(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}] + \{E[S(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}]\}^2. \tag{3.14}$$

For evaluating $I(\boldsymbol{\theta})$ in (3.14) for a given a LTRC data, multiple samples $\boldsymbol{\Gamma}_{PC}^{(m)}$, $m = 1, ..., M$ are imputed corresponding to the censored data $\boldsymbol{\Gamma} = \{i : \delta_i = 0\}$, thus obtaining multiple pseudo-complete datasets $\boldsymbol{t}_{PC}^{(m)}$, $m = 1, ..., M$. Then, $I(\boldsymbol{\theta})$ is estimated as

$$\widehat{I(\boldsymbol{\theta})} = \frac{1}{M} \sum_{m=1}^{M} H(\boldsymbol{\theta}, \boldsymbol{t}_{PC}^{(m)}) - \frac{1}{M} \sum_{m=1}^{M} [S(\boldsymbol{\theta}, \boldsymbol{t}_{PC}^{(m)})]^2 + \left[\frac{1}{M} \sum_{m=1}^{M} S(\boldsymbol{\theta}, \boldsymbol{t}_{PC}^{(m)}) \right]^2 \Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}. \tag{3.15}$$

Finally, the asymptotic variance-covariance matrix of the estimates is obtained by inverting $\widehat{I(\boldsymbol{\theta})}$, and the asymptotic confidence intervals for the parameters can be constructed using the asymptotic variances.

A second approach we use here is based on the bootstrap procedure. Bootstrap confidence intervals [11] are widely used in statistical literature. These intervals are particularly of interest for LTRC data, as the presence of truncation and censoring often tend to bias the estimates of parameters. We use the following algorithm to obtain parametric bootstrap confidence intervals. Here, SP is the starting point, and TP is the termination point of units.

Algorithm 2:

Step 1: Based on the given LTRC data, obtain the estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta})$;

Step 2: Construct empirical distribution of SPs for left truncated units;

Step 3: Construct empirical distribution of SPs for untruncated units;

Step 4: To get a bootstrap sample (preserving proportion of truncation):

- Step 4.1: Sample SPs for truncated units from empirical distribution of Step 2;
- Step 4.2: Sample SPs for untruncated units from empirical distribution of Step 3;
- Step 4.3: Generate failure-times from Burr$(\widehat{\alpha}, \widehat{\beta})$;
- Step 4.4: Add failure-times to SPs, to obtain corresponding TPs;
- Step 4.5: Determine censoring status of units according to their TPs;

Step 5: For this bootstrap sample, obtain bootstrap estimate $\widehat{\boldsymbol{\theta}}^* = (\widehat{\alpha}^*, \widehat{\beta}^*)$;

Step 6: Repeat Steps 4 and 5 B times, to obtain $\widehat{\boldsymbol{\theta}}_1^*, \widehat{\boldsymbol{\theta}}_2^*, ..., \widehat{\boldsymbol{\theta}}_B^*$.

The value of B, i.e., the number of bootstrap samples, should be sufficient to stabilize the estimated bootstrap bias and variance of $\widehat{\boldsymbol{\theta}}$. For constructing parametric bootstrap confidence intervals based on LTRC data from the Burr XII distribution, we recommend using $B \geq 200$.

A $100(1-\delta)\%$ parametric bootstrap confidence interval for α is then given by

$$\Big(\widehat{\alpha} - b_\alpha - z_{\delta/2}\sqrt{v_\alpha},\ \widehat{\alpha} - b_\alpha + z_{\delta/2}\sqrt{v_\alpha}\Big),$$

where b_α and v_α are the bootstrap bias and bootstrap variance, respectively. Here, z_δ is the upper δ-percentile point of standard normal distribution. The $100(1-\delta)\%$ parametric bootstrap confidence intervals for β are constructed in a similar way.

3.3. Direct optimization of observed likelihood

Considering the four different types of units with different combinations of left truncation and right censoring as mentioned in Figure 3 in Section 2, the observed likelihood for LTRC data is given by

$$\begin{aligned} L(\boldsymbol{\theta}|DATA) \;=\; & \prod_{i\in S_1} \big\{f(t_i;\boldsymbol{\theta})\big\}^{\delta_i} \big\{1-F(t_i;\boldsymbol{\theta})\big\}^{1-\delta_i} \\ & \times \prod_{i\in S_0} \left\{\frac{f(t_i;\boldsymbol{\theta})}{1-F(\kappa_{Li};\boldsymbol{\theta})}\right\}^{\delta_i} \left\{\frac{1-F(t_i;\boldsymbol{\theta})}{1-F(\kappa_{Li};\boldsymbol{\theta})}\right\}^{1-\delta_i}. \end{aligned} \tag{3.16}$$

By plugging in the PDF and CDF of the Burr XII distribution in (3.16), we get the specific likelihood for the Burr XII distribution based on LTRC data. The likelihood (or the corredponding log-likelihood) function may then be maximized using routine functions in statistical software. The performance of the estimates obtained by the St-EM algorithm, and those obtained by the direct method of optimization based on observed likelihood can then be compared through Monte Carlo simulations. In this paper, we have used the "maxLik" package in R software for maximizing the observed likelihood provided in (3.16). In particular, we have used the Newton-Raphson method for direct numerical optimization; the Newton-Raphson method may be employed by using the maxNR() routine available in the maxLik package. Details of the numerical results are presented in the next Section.

4. NUMERICAL EXPERIMENTS

The methods of inference are assessed through Monte Carlo simulations using the R software. For simulating LTRC data, the following process is followed. We consider lifetime data at the yearly scale. Left truncation and right censoring points are fixed at the years 2000 and 2004, respectively, without loss of generality. That is, in connection to the notations used in Section 2, we set $L = 2000$, and $R = 2004$. The total sample size n is fixed; here, we consider several values of n, namely, 50, 100, 200, 300, and 500.

First, a truncation percentage p $(0 \leq p \leq 1)$ is specified; this implies that in the sample, there will be np left truncated units, and $n(1-p)$ untruncated units. For this simulation study, p is chosen as 20, and 30. Two arbitrary sets of years as installation points (IPs), say W_{LT} and W_{NT}, respectively, are chosen; W_{LT} corresponds to the left truncated group (i.e., less than 2000), and W_{NT} corresponds to the untruncated group (i.e., more than or equal to 2000), are taken as reference frames from sampling. Then, through equal probability sampling, two sets of samples of IPs are generated from W_{LT} and W_{NT} according to the pre-specified truncation percentage; these samples represent the left truncated and the untruncated groups, respectively. For example, for $n = 100$, and $p = 20$, a sample of 20 IPs is taken from W_{LT}, and a sample of 80 IPs is taken from W_{NT}.

For generating lifetimes from the Burr XII distributions, two sets of values for the model parameters are used. A LTRC dataset of size n is generated as follows. Corresponding to each IP ω_i, with $\omega_i \in W$, $i = 1, ..., n$, where $W = W_{LT} \bigcup W_{NT}$, a failure-time y_i is generated from the Burr XII distribution in (1.2), and is added to ω_i to obtain the respective termination point (TP). Left truncation and right censoring, corresponding to $L = 2000$ and $R = 2004$, are incorporated into the generated data through the following mechanism. For $i \in W_{LT}$, if $\omega_i + y_i < 2000$ for a unit, that unit is completely discarded, and is replaced by a new set of values for ω_i, and y_i; this process of discarding continues until for that unit we have $\omega_i + y_i > 2000$. This ensures that all units have to cross a threshold to be included in the study, as required by left truncation. For a unit $i \in W$, if $\omega_i + y_i > 2004$, it is a right-censored unit; otherwise, it is not censored. The chosen values of the parameters of the Burr XII distribution ensures that there are enough censored units.

The bias, and mean squared error (MSE) of the point estimates for different simulation parameter settings are reported in Tables 1–6. The coverage probability and average length of the asymptotic confidence intervals for the model parameters are also reported in these tables.

Table 1: Performance of point and interval estimates for truncation percentage 20, corresponding to true parameter value $(\alpha, \beta) = (2, 0.5)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.051	0.123	0.037	0.121	0.960	1.298	0.960	1.335
	β	0.014	0.004	0.012	0.004	0.954	0.253	0.956	0.262
100	α	0.028	0.054	0.021	0.053	0.942	0.907	0.950	0.915
	β	0.005	0.002	0.004	0.002	0.950	0.177	0.946	0.179
200	α	0.009	0.025	0.005	0.025	0.958	0.633	0.958	0.639
	β	0.003	0.001	0.002	0.001	0.958	0.124	0.952	0.125
300	α	0.013	0.018	0.011	0.018	0.952	0.516	0.946	0.519
	β	0.004	0.001	0.004	0.001	0.956	0.102	0.956	0.102
500	α	0.002	0.011	0.001	0.011	0.932	0.396	0.948	0.400
	β	0.001	0.000	0.001	0.000	0.954	0.078	0.954	0.079

Table 2: Performance of point and interval estimates for truncation percentage 30, corresponding to true parameter value $(\alpha, \beta) = (2, 0.5)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.026	0.103	0.012	0.102	0.966	1.313	0.964	1.342
	β	0.017	0.005	0.015	0.005	0.956	0.263	0.956	0.273
100	α	0.013	0.054	0.006	0.053	0.956	0.919	0.950	0.931
	β	0.009	0.002	0.008	0.002	0.962	0.183	0.960	0.186
200	α	0.018	0.030	0.015	0.030	0.940	0.649	0.950	0.650
	β	0.003	0.001	0.002	0.002	0.952	0.128	0.952	0.128
300	α	0.012	0.019	0.010	0.019	0.962	0.527	0.958	0.529
	β	0.000	0.001	−0.000	0.001	0.942	0.104	0.950	0.104
500	α	0.004	0.012	0.003	0.012	0.932	0.401	0.942	0.411
	β	0.000	0.000	0.000	0.000	0.934	0.080	0.934	0.081

Table 3: Performance of point and interval estimates for truncation percentage 50, corresponding to true parameter value $(\alpha, \beta) = (2, 0.5)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.027	0.125	0.015	0.124	0.956	1.410	0.956	1.436
	β	0.015	0.005	0.013	0.005	0.968	0.287	0.968	0.304
100	α	0.026	0.074	0.019	0.073	0.930	0.989	0.934	1.000
	β	0.006	0.003	0.005	0.003	0.948	0.199	0.962	0.204
200	α	0.014	0.030	0.010	0.030	0.946	0.693	0.952	0.701
	β	0.001	0.001	0.001	0.001	0.956	0.139	0.956	0.141
300	α	0.012	0.020	0.010	0.020	0.946	0.563	0.952	0.569
	β	0.001	0.001	0.001	0.001	0.970	0.114	0.964	0.114
500	α	0.008	0.012	0.006	0.012	0.960	0.432	0.966	0.441
	β	0.001	0.001	0.001	0.001	0.962	0.088	0.958	0.088

Table 4: Performance of point and interval estimates for truncation percentage 20, corresponding to true parameter value $(\alpha, \beta) = (3, 1)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.080	0.231	0.076	0.231	0.946	1.764	0.955	1.875
	β	0.024	0.012	0.023	0.012	0.957	0.425	0.955	0.441
100	α	0.032	0.098	0.029	0.098	0.958	1.225	0.952	1.265
	β	0.014	0.006	0.014	0.006	0.946	0.298	0.946	0.305
200	α	0.015	0.052	0.014	0.052	0.952	0.862	0.950	0.875
	β	0.004	0.003	0.004	0.003	0.948	0.209	0.944	0.211
300	α	0.014	0.030	0.013	0.030	0.962	0.703	0.964	0.705
	β	0.006	0.002	0.006	0.002	0.944	0.171	0.952	0.172
500	α	−0.005	0.016	−0.005	0.016	0.960	0.541	0.964	0.543
	β	0.002	0.001	0.002	0.001	0.958	0.132	0.954	0.132

Table 5: Performance of point and interval estimates for truncation percentage 30, corresponding to true parameter value $(\alpha, \beta) = (3, 1)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.110	0.290	0.106	0.209	0.947	1.772	0.937	1.866
	β	0.021	0.012	0.020	0.012	0.968	0.428	0.958	0.446
100	α	−0.002	0.105	−0.004	0.105	0.944	1.209	0.942	1.239
	β	0.010	0.006	0.009	0.006	0.964	0.301	0.958	0.308
200	α	0.013	0.046	0.012	0.046	0.964	0.858	0.966	0.869
	β	0.005	0.003	0.005	0.003	0.942	0.211	0.944	0.214
300	α	−0.004	0.035	−0.005	0.035	0.946	0.697	0.946	0.701
	β	0.004	0.002	0.004	0.002	0.956	0.173	0.948	0.174
500	α	0.008	0.017	0.008	0.017	0.962	0.542	0.968	0.542
	β	0.002	0.001	0.002	0.001	0.958	0.133	0.956	0.133

Table 6: Performance of point and interval estimates for truncation percentage 50, corresponding to true parameter value $(\alpha, \beta) = (3, 1)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.053	0.234	0.050	0.234	0.946	1.810	0.950	1.873
	β	0.027	0.015	0.026	0.015	0.948	0.465	0.954	0.491
100	α	0.027	0.112	0.026	0.112	0.944	1.269	0.948	1.284
	β	0.008	0.007	0.007	0.007	0.950	0.323	0.958	0.332
200	α	−0.004	0.057	−0.004	0.057	0.944	0.889	0.928	0.891
	β	0.013	0.003	0.013	0.003	0.964	0.230	0.970	0.232
300	α	0.009	0.030	0.009	0.030	0.942	0.729	0.944	0.728
	β	0.002	0.002	0.002	0.002	0.944	0.185	0.940	0.187
500	α	0.005	0.020	0.005	0.020	0.940	0.564	0.944	0.562
	β	0.004	0.001	0.004	0.001	0.946	0.144	0.944	0.145

The Monte Carlo estimate of coverage probability of an asymptotic confidence interval corresponding to a nominal level of confidence (say, 95%) is the proportion of times the asymptotic confidence interval includes the true parameter value out of the total number of Monte Carlo runs of the experiment. The average length of an asymptotic confidence interval is the mean length of the interval, averaged over the lengths obtained in the Monte Carlo runs.

From Tables 1–6, we notice that the estimates obtained by the St-EM algorithm are quite efficient in general, with respect to their bias and MSE. As one would expect, with increase in sample size, the bias and MSE of the estimates reduce. Truncation percentage does not seem to have a significant effect on the point estimates, as the bias and MSE values do not change much with change in truncation percentage.

It may also be of interest to compare the results of the St-EM algorithm with that of the Newton-Raphson method. It is observed from Tables 1–6 that for parameter α, the biases corresponding to the St-EM algorithm and the Newton-Raphson method are to some extent different for smaller sample sizes (i.e., $n = 50$ and 100). However, with increase in sample size (i.e., for $n = 200, 300, 500$), the biases become very close. For parameter β, the biases of the estimates corresponding to the two methods are close for all simulation settings considered here. Finally, the MSE of the estimates of both α and β are always quite close for the two methods.

Tables 1–6 also report coverage probabilities (CP) and average lengths (AL) for asymptotic 95% confidence intervals. The coverage probability and average length are two important criteria for assessing the performance of confidence intervals. For a confidence interval to be reasonable, its coverage probability should be close to the nominal confidence level, and its average length should not be large. It may be noted that the coverage probabilities corresponding to the missing information principle are always very close to the nominal level. The coverage probabilities corresponding to the parametric bootstrap are also close to the nominal level. With respect to average length of the intervals, both methods perform closely. It is also observed that with increase in sample size, though their average lengths reduce as expected, but the confidence intervals are able to retain the coverage probability close to the nominal level.

5. THE THREE PARAMETER BURR XII DISTRIBUTION

Considering a scale parameter λ along with the two shape parameters α and β, the PDF of the three parameter Burr XII distribution is given by (see [24])

$$f_Y(y;\lambda,\alpha,\beta) = \frac{\alpha\beta}{\lambda}\left(\frac{y}{\lambda}\right)^{\beta-1}\left(1+\left(\frac{y}{\lambda}\right)^{\beta}\right)^{-(\alpha+1)}, \quad y>0, \quad \lambda,\alpha,\beta>0.$$

For the St-EM algorithm, the conditional distributions for generating random observations is given by

(5.1)

$$f_{T_i|T_i>y_i}(t_i|t_i>y_i;\lambda,\alpha,\beta) = \frac{\alpha\beta}{\lambda}\left(\frac{t_i}{\lambda}\right)^{\beta-1}\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right)^{-(\alpha+1)}\left(1+\left(\frac{y_i}{\lambda}\right)^{\beta}\right)^{\alpha}, \quad t_i>y_i.$$

After replacing the right censored failure-times by randomly generated observations from (5.1), the pseudo-Q function is obtained as

$$\begin{aligned} Q_{PC}(\boldsymbol{\theta}) \;=\; & n(\log\alpha + \log\beta - \beta\log\lambda) + (\beta-1)\sum_{i=1}^{n}\log t_i - \sum_{i=1}^{n}\log\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right) \\ & - \alpha\left[\sum_{i=1}^{n}\log\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right) - \sum_{i\in S_2}\log\left(1+\left(\frac{\kappa_{Li}}{\lambda}\right)^{\beta}\right)\right], \end{aligned} \tag{5.2}$$

with $\boldsymbol{\theta} = (\lambda, \alpha, \beta)$. For fixed λ and β, equating the first derivative of $Q_{PC}(\boldsymbol{\theta})$ with respect to α to zero, we obtain

$$\alpha = \frac{n}{W_{PC}(\lambda,\beta)}, \tag{5.3}$$

where

$$W_{PC}(\lambda,\beta) = \sum_{i=1}^{n}\log\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right) - \sum_{i\in S_2}\log\left(1+\left(\frac{\kappa_{Li}}{\lambda}\right)^{\beta}\right).$$

Substituting (5.3) in (5.2), the profile log-likelihood in λ and β is obtained as

$$p_{PC}(\lambda,\beta) = n(\log\beta - \beta\log\lambda - \log W_{PC}(\lambda,\beta)) + (\beta-1)\sum_{i=1}^{n}\log t_i - \sum_{i=1}^{n}\log\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right), \tag{5.4}$$

which can then be maximized by a routine two-parameter optimizer.

Starting with an initial value for the parameter vector as $\boldsymbol{\theta}^{(0)} = (\lambda^{(0)}, \alpha^{(0)}, \beta^{(0)})$, to the choice of which the St-EM algorithm is quite robust, the following are the steps of the St-EM algorithm for the three-parameter Burr XII distribution based on LTRC data.

ALGORITHM 3: At the k-th stage of the algorithm:

Stochastic Expectation (St-E) step:

STEP 1: The available parameter value is $\boldsymbol{\theta}^{(k)} = (\lambda^{(k)}, \alpha^{(k)}, \beta^{(k)})$;

STEP 2: Replace each unit in $\boldsymbol{\Gamma}$ by generating observations from $f_{T_i|T_i>y_i}(t_i|t_i>y_i;\boldsymbol{\theta}^{(k)})$ to obtain pseudo-complete data $\boldsymbol{t}_{PC}$;

STEP 3: With $\boldsymbol{t}_{PC}$ from Step 2, construct $Q_{PC}(\boldsymbol{\theta}^{(k)})$ following (5.2);

Maximization (M) step:

STEP 4: Choose initial values $\lambda_{init}^{(k)}$ and $\beta_{init}^{(k)}$ based on the pseudo-complete data $\boldsymbol{t}_{PC}$;

STEP 5: Optimize $p_{PC}(\lambda,\beta)$ in (5.4) to get $\widehat{\lambda^{(k+1)}}$ and $\widehat{\beta^{(k+1)}}$ subject to a tolerance level;

STEP 6: Using (5.3), calculate $\widehat{\alpha^{(k+1)}} = \frac{n}{W_{PC}(\lambda^{(k+1)},\beta^{(k+1)})}$;

STEP 7: With updated estimate $\boldsymbol{\theta}^{(k+1)} = (\lambda^{(k+1)}, \alpha^{(k+1)}, \beta^{(k+1)})$, go back to Step 2.

From a sequence of estimates $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, ..., \boldsymbol{\theta}^{(N)}$, after discarding first M estimates for burn-in, the remaining ones are averaged to get estimates $\widehat{\lambda}$, $\widehat{\alpha}$, and $\widehat{\beta}$. The confidence intervals for the model parameters may be obtained by similar processes as described in Section 3 for the two-parameter Burr XII distribution.

It may be mentioned here that in case of LTRC data from the three-parameter Burr XII distribution, based on some limited simulations, we have noticed an indication that the St-EM algorithm performs better compared to the direct optimization of observed likelihood, in terms of bias and MSE of the estimates. However, implementation of the St-EM algorithm in this case is very challenging due to its computational cost; the running time of the St-EM algorithm for the three-parameter case is significantly longer than that of the direct optimization method.

6. PREDICTION OF EXPECTED NUMBER OF FAILURES IN A FUTURE INTERVAL

Consider the right censored units with the i-th unit having right censored lifetime y_i, $i \in S_{cen}$. Consider a future interval $(\tau_1, \tau_2]$ with $t_{max} < \tau_1$, where $t_{max} = \text{Max}\{t_i; i \in S_{cen}\}$. The probability that the i-th unit fails in this interval $(\tau_1, \tau_2]$ is given by

$$\pi_i = P(\tau_1 < T_i \leq \tau_2 | T_i > y_i) = \frac{S(\tau_1; \boldsymbol{\theta}) - S(\tau_2; \boldsymbol{\theta})}{S(y_i; \boldsymbol{\theta})}, \tag{6.1}$$

where $S(t) = P(T > t)$ is the survival function of the underlying failure-time variable T. Note that the expression for this probability remains same regardless of the truncation status of the i-th unit, $i \in S_{cen}$. We are interested in obtaining the expected number of failures in the future interval $(\tau_1, \tau_2]$.

Let us define random variables U_i, $i \in S_{cen}$, such that

$$U_i = \begin{cases} 1, & \text{if } i\text{-th item fails in } (\tau_1, \tau_2] \\ 0, & \text{otherwise.} \end{cases}$$

Note that $E[U_i] = P(U_i = 1) = \pi_i$. We want to obtain the expected number of failures in the future interval (τ_1, τ_2), given by

$$\zeta = E\left[\sum_{i \in S_{cen}} U_i\right] = \sum_{i \in S_{cen}} \pi_i.$$

Now, using the expression for π_i in (6.1), we obtain

$$\begin{aligned} \zeta &= \{(1 + \tau_1^{\beta})^{-\alpha} - (1 + \tau_2^{\beta})^{\alpha}\} \sum_{i \in S_{cen}} (1 + y_i^{\beta})^{\alpha} \\ &= h(\boldsymbol{\theta}) \quad \text{(say)}. \end{aligned} \tag{6.2}$$

Clearly, an estimate $\widehat{\zeta}$ of the expected number of failures ζ can be obtained by simply plugging-in the estimated parameters in (6.2). It is also possible to provide an asymptotic confidence interval for ζ by a straightforward application of the delta-method, by using the asymptotic

normality and the delta-method. That is, using the fact that $\sqrt{n}(\widehat{\zeta} - \zeta) \xrightarrow{D} N(0, \text{Var}(\widehat{\zeta}))$, where the variance can be estimated as

$$\widehat{\text{Var}(\widehat{\zeta})} = \left(\left(\frac{\partial h}{\partial \alpha}\right)^2 \text{Var}(\widehat{\alpha}) + 2\left(\frac{\partial h}{\partial \alpha}\right)\left(\frac{\partial h}{\partial \beta}\right) \text{Cov}(\widehat{\alpha}, \widehat{\beta}) + \left(\frac{\partial h}{\partial \beta}\right)^2 \text{Var}(\widehat{\beta}) \right) \Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}. \tag{6.3}$$

Finally, using the estimated variance, an asymptotic $100(1-\gamma)\%$ confidence interval for ζ can be easily obtained.

7. ILLUSTRATIVE DATA ANALYSIS

The Channing House data involves lives of residents of a retirement centre in Palo Alto, California. The dataset contains lifetimes of residents of the centre since it started operations in 1965 till July, 1975. A person had to be at least 60 years of age to be a resident of the centre; this fact incorporated left truncation in the data. In fact, due to this restriction on the entry of individuals to the centre, the entire data (i.e., 100% of the observations) is left truncated according to the notion of left truncation followed in here.

Some individuals died as residents of the centre, while some other were still alive when the collection ended in July, 1975. This incorporated right censoring in the data. The dataset contains lives of total 462 residents. Out of 462, the number of observed failures is only 176, and the rest of the units are right censored. A summary of the dataset is presented in Table 7.

Table 7: Summary of Channing House Data.

Group	Total number	Right censoring	Mean lifetime (Years)	SD lifetimes (Years)
Male	97	52.58%	82.63	6.14
Female	365	64.38%	82.04	6.15
Combined	462	61.90%	82.17	6.15

Before analyzing the dataset, we change the origin and scale of this data by subtracting 720 from each of the lifetimes (and left truncation times), and by dividing them by 200; this change of origin and scale of the data will not impact the inferential results in any way. We assume that the underlying failure-time variable follow a two-parameter Burr XII distribution. The results of point and interval estimation by the proposed methods are given in Table 8. The estimated parameters can then be used in further analyses, for example, in predicting future failures as described below.

Table 8: Point and interval estimates of model parameters for Channing House Data.

Parameter	Point Estimate	Interval Estimate	
		MI	BB
α	0.508	(0.354, 0.662)	(0.354, 0.655)
β	3.976	(2.915, 5.038)	(2.764, 5.002)

As the estimates of the parameters turn out to be $\widehat{\alpha} = 0.508$ and $\widehat{\beta} = 3.976$, using these, we can obtain the asymptotic variance-covariance matrix by using the missing information principle as described in Section 3.2 as

$$\begin{pmatrix} \mathrm{Var}(\widehat{\alpha}) & \widehat{\mathrm{Cov}(\alpha, \beta)} \\ & \mathrm{Var}(\widehat{\beta}) \end{pmatrix} = \begin{pmatrix} 0.0061 & -0.0369 \\ & 0.2934 \end{pmatrix}.$$

The maximum value among the transformed right censored lifetimes is 2.435. Suppose, we are interested in predicting the expected number of failures in the interval $(2.5, 2.7]$. By plugging-in the estimates $\widehat{\alpha}$ and $\widehat{\beta}$ in (6.2), we get the expected number of failures in the above interval as $\widehat{\zeta} = 13.4000$. Finally, upon estimating the variance of $\widehat{\zeta}$ by (6.3) to be 0.0490, a 95% confidence interval for the expected number of failures in this future interval is obtained as $(12.9660, 13.8339)$.

Suppose it is of interest to select an appropriate model for this dataset among many candidate models. One way to achieve this would be fit different models to the dataset, and then to choose the model for which the value of the maximized log-likelihood, evaluated at the MLE, is the largest. Naturally, the distributions which are frequently used to model lifetime data would be the candidate models. As suggested by a reviewer, here, we consider Weibull, Gompertz, and Lomax distributions as the candidate models, along with the Burr XII model. It may be mentioned here that Weibull, Gompertz, and Lomax distributions belong to a family of distributions known as the Lehmann family of distributions [16].

Table 9 gives the results of the model selection. We fit the Burr XII model to the Channing House data by using the St-EM algorithm; we also fit Weibull, Gompertz, and Lomax distributions to the data by using St-EM algorithm. Then, we evaluate the log-likelihood functions corresponding to the four distributions at the respective MLEs, the log-likelihood being constructed by using the LTRC data structure. It may be mentioned here that since all the models considered here have same number of parameters, the process of using the maximized log-likelihood is essentially equivalent to using the Akaike's information criterion (AIC) for model selection.

Table 9: Maximized log-likelihood for different models.

Model	Distribution Function	Maximized log-likelihood
Weibull	$F_W(t;\lambda,\alpha) = 1 - e^{-\alpha t^{\lambda}}, \quad t > 0$	-155.9704
Gompertz	$F_G(t;\lambda,\alpha) = 1 - e^{-\alpha(e^{\lambda t}-1)}, \quad t > 0$	-152.9099
Lomax	$F_L(t;\lambda,\alpha) = 1 - \left[\frac{1}{1+\lambda t}\right]^{\alpha}, \quad t > 0$	-189.7542
Burr XII	$F_B(t;\alpha,\beta) = 1 - (1+t^{\beta})^{-\alpha}, \quad t > 0$	-181.7247

It turns out that the maximized log-likelihood is the largest for the Gompertz distribution based on this data. Therefore, the Gompertz distribution turns out to be the most suitable model for the Channing House data by the above criterion.

8. CONCLUSION AND FUTURE WORK

In this article, statistical inferential procedures for the Burr XII distribution based on LTRC data are discused. the two- and the three-parameter Burr XII models are considered. Detailed steps of the stochastic EM algorithm based on LTRC data are developed for obtaining point estimates of the model parameters. Two methods for constructing asymptotic confidence intervals of the parameters are discussed: one by using the missing information principle, and the other by using a parametric bootstrap approach. A method for including covariates in the Burr XII model in this setup is also discussed. An application of the estimated parameters in predictive number of failures in a future interval is presented.

From the numerical results of a detailed Monte Carlo simulation study, it is observed that the stochastic EM algorithm performs reasonably well in estimating the model parameters. The approaches for constructing confidence intervals also perform satisfactorily, as the coverage probabilities of the confidence intervals remain always close to the nominal confidence level of 95%. It is also observed that the performance of the St-EM algorithm is close to that of the Newton-Raphson method.

While parametric inference can generate accurate results when the assumptions regarding the underlying distribution of data are appropriate, it may be of interest to verify whether the distributional assumptions are reasonable or not. In view of this, it will be of interest to develop a test for goodness of fit for the Burr XII distribution based on LTRC data.

Another problem of interest would be to study Bayesian inference for the Burr XII distribution based on LTRC data. The Bayesian methods can provide significant information regarding a model, especially when the prior assumptions are appropriate. In particular, the Bayesian methods may outperform classical inferential methods when the sample size is not very large, provided meaningful prior assumptions are made. However, the most critical task of performing Bayesian inference would be the elicitation of the prior distributions.

ACKNOWLEDGMENTS

The author expresses his sincere thanks to the two anonymous reviewers who made many constructive comments that helped to signficantly improve an earlier version of this manuscript.

REFERENCES

[1] Balakrishnan, N. and Mitra, D. (2011). Likelihood inference for log-normal data with left truncation and right censoring with an illustration, *Journal of Statistical Planning and Inference*, **141**, 3536–3553.

[2] Balakrishnan, N. and Mitra, D. (2012). Left truncated and right censored Weibull data and likelihood inference with an illustration, *Computational Statistics and Data Analysis*, **56**, 4011–4025.

[3] Balakrishnan, N. and Mitra, D. (2013). Likelihood inference based on left truncated and right censored data from a gamma distribution, *IEEE Transactions on Reliability*, **62**, 679–688.

[4] Balakrishnan, N. and Mitra, D. (2014a). Some further issues concerning likelihood inference for left truncated and right censored lognormal data, *Communications in Statistics – Simulation and Computation*, **43**, 400–416.

[5] Balakrishnan, N. and Mitra, D. (2014b). EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination (with discussions), *South African Statistical Journal*, **48**, 125–204.

[6] Beirlant, J.; Goegebeur, Y.; Verlaak, R. and Vynckier, P. (1998). Burr regression and portfolio segmentation, *Insurance: Mathematics and Economics*, **23**, 231–250.

[7] Bordes, L. and Chauveau, D. (2014). Comments: EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination, *South African Statistical Journal*, **48**, 197–200.

[8] Burr, I.W. (1942). Cumulative frequency functions, *Annals of Mathematical Statistics*, **13**, 215–232.

[9] Celeux, G.; Chauveau, D. and Diebolt, J. (1996). Stochastic versions of the em algorithm: an experimental study in the mixture case, *Journal of Statistical Computation and Simulation*, **55**, 287–314.

[10] Dempster, A.P.; Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

[11] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

[12] Henningsen, A. and Toomet, O. (2011). maxLik: a package for maximum likelihood estimation in R, *Computational Statistics*, **26**, 443–458.

[13] Hong, Y.; Meeker, W.Q. and McCalley, J.D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data, *The Annals of Applied Statistics*, **3**, 857–879.

[14] Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, Second Edition, Springer-Verlag, New York.

[15] Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society, Series B*, **44**, 226–233.

[16] Mitra, D.; Kundu, D. and Balakrishnan, N. (2021). Likelihood analysis and stochastic EM algorithm for left truncated right censored data and associated model selection from the Lehmann family of life distributions, *Japanese Journal of Statistics and Data Science*, **4**, 1019–1048.

[17] Nielsen, S.F. (2000). The stochastic EM algorithm: Estimation and asymptotic results, *Bernoulli*, **6**, 457–489.

[18] Ng, H.K.T. and Ye, Z. (2014). Comments: EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination, *South African Statistical Journal*, **48**, 177–180.

[19] Silva, G.O.; Edwin, M.M.O.; Cancho, V.G. and Barreto, M.L. (2008). Log-Burr XII regression models with censored data, *Computational Statistics & Data Analysis*, **52**, 3820–3842.

[20] Silva, G.O.; Edwin, M.M.O. and Paula, G.A. (2011). Residuals for log-Burr XII regression models in survival analysis, *Journal of Applied Statistics*, **38**, 1435–1445.

[21] Soliman, A.A. (2005). Estimation of parameters of life from progressively censored data using Burr-XII model, *IEEE Transaction on Reliability*, **54**, 34–42.

[22] Tadikamalla, P.R. (1980). A look at the Burr and related distributions, *International Statistical Review*, **48**, 337–344.

[23] Ye, Z. and Ng, H.K.T. (2014). On analysis of incomplete field failure data, *The Annals of Applied Statistics*, **8**, 1713–1727.

[24] Zimmer, W.J.; Keats, J.B. and Wang, F.K. (1998). The Burr XII distribution in reliability analysis, *Journal of Quality Technology*, **30**, 386–394.

REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 469–490
https://doi.org/10.57805/revstat.v21i4.430

Random Environment Integer-Valued Autoregressive Process with Discrete Laplace Marginal Distributions

Authors: Bogdan A. Pirković ✉
– Department of Mathematics and Informatics, Faculty of Science, University of Kragujevac, Kragujevac, Serbia
bogdanpirkovic@gmail.com

Miroslav M. Ristić
– Department of Mathematics, Faculty of Sciences and Mathematics, University of Niš, Niš, Serbia
miristic72@gmail.com

Aleksandar S. Nastić
– Department of Mathematics, Faculty of Sciences and Mathematics, University of Niš, Niš, Serbia
anastic78@gmail.com

Received: June 2021 Revised: February 2022 Accepted: February 2022

Abstract:

- A new random environment integer-valued autoregressive process of order 1 with discrete Laplace marginal distributions and with r states (abbrev. $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$) is introduced. It is shown that this process is distributed as a difference of two independent generalized random environment integer-valued autoregressive processes, when their orders are equal to 1. Other distributional and correlation properties of the $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ process are discussed. Strongly consistent Yule-Walker estimates are defined. The method of moments is implemented for different cases of simulated samples. Finally, the proposed model is applied to real-life data and the obtained results show its effectiveness.

Keywords:

- *random environment;* $\mathrm{INAR}(1)$, $\mathrm{rDLINAR}_1(\mathcal{M},\mathcal{A})$*;* $\mathrm{DLINAR}(1)$*; discrete Laplace distribution.*

AMS Subject Classification:

- 62M10.

✉ Corresponding author.

1. INTRODUCTION

The major breakthrough concerning integer-valued autoregressive (INAR) processes was made, independently of each other, by [16] and [1]. In the period that followed, many generalizations and modifications of these models were published, and many authors were trying to create their own models in order to better describe the data obtained from some natural processes. Some of them introduced new thinning operators dependent of a single parameter, as it was done by [12], [22, 23] and [13]. More recently, a thinning operator with two parameters has appeared in [15]. On contrary to this, others researchers discussed marginal distributions, as given in [2] and [3]. In recent years, authors have been trying to model some specific count data. Thus, data sets with excess zeros and excess ones are modeled in [20], while a model for modeling heavily-tailed count data is proposed in [21]. Most of the introduced models were applied to non-negative data, although in many real-life situations there are processes which may consist of integer values including both positive and negative numbers. A step forward in this direction was made by [9], who introduced a true integer-valued process, defined in distribution as a difference of two non-negative, independent INAR processes with Poisson marginal distributions. In the same way, [17] introduced an INAR model with discrete Laplace marginal distributions (DLINAR(1)), defined in distribution as a difference of two non-negative, independent INAR processes with the same geometric marginal distributions. Some generalizations of this idea emerged in the work of [4] and [7]. Lately, [6] and [14] have also come up with innovative ideas for creating models on $\mathbb{Z}$. Although each of these models deserves attention, the DLINAR(1) model introduced by [17] is of particular importance for this paper.

All the processes mentioned here are stationary, since stationarity simplifies the calculation when describing the model and determining the estimates of model parameters. But, we can say that stationary processes are rigid, since some of their properties are conserved in time. Nevertheless, the real data are not usually like that. The first non-stationary model appeared in [18] and is flexible towards the environment conditions changes. Namely, quantitative properties of phenomena from nature depend on environment, so it's logical, as well, to assume the distribution to depend on environment. It is supposed that environment conditions can be divided into r different types, called states. Each state is associated with a fixed distribution, and an element of the process has the distribution of its state. Various authors have tried to generalize or modify this idea in recent years. They assumed that the order of the model, or even the thinning parameter value, are also determined by the environment state at a particular moment, as it was done by [11] and [19].

Although the DLINAR(1) model successfully estimates the data it was tested on, due to its property of stationarity, the model shows substantial difficulties in adjusting to the elements that deviate significantly from zero. In particular, the model is struggling to estimate the highest and the lowest peaks, with a large difference between real values and their estimated values. This fact leaves room for model improvement. The main goal of this article is to make the DLINAR(1) model more flexible, using the idea given in [18]. In other words, the goal is to construct the DLINAR(1) process dependent on the environment states. Nevertheless, the idea mentioned in [18] cannot be fully taken over and certain adaptations have to be made. So, in Section 2 of this article, the construction of such a process, which overcomes problems mentioned above, is given alongside with its main distributional properties.

Section 3 provides the k-step ahead conditional expectation and a correlation structure. Yule-Walker (YW) estimates of the parameters of the defined model are given in Section 4. In Section 5, the quality of YW estimates is examined on simulated data. Section 6 deals with forecasting and provides a criterion to compare the prediction results between different models. An application of the introduced model to some real-life data is presented in Section 7, and results are compared for different models.

2. CONSTRUCTION OF THE PROCESS

As mentioned in the previous section, the first attempt to increase the flexibility of the DLINAR(1) process followed the idea given in [18]. An attempt to construct such an improved process brought some difficulties, because the newly acquired process had the same shape of one-step ahead conditional expectation as it was the case with DLINAR(1) process. To avoid this issue, the flexibility of the DLINAR(1) process is improved using the concept given in [11], although in a bit simpler form. Namely, it is assumed that information about the environment state is not only carried by the marginal distribution parameter, but it can also be expressed through the thinning parameter value. In other words, we assume that the value of the marginal distribution parameter and the value of the thinning parameter, both in moment n, depend on environment state in the same moment. A new INAR process with discrete Laplace marginal distributions, that meets the aforementioned assumptions, is defined in this section and some of its properties are discussed.

In order to make the reading of the manuscript easygoing, definitions of RrNGINAR$(\mathcal{M},\mathcal{A},\mathcal{P})$ and DLINAR(1) processes are given, since the paper relies heavily on those. As mentioned in [11], we call $\{X_n(z_n)\}$ the RrNGINAR$(\mathcal{M},\mathcal{A},\mathcal{P})$ process if its element $X_n(z_n)$ at moment $n\in\mathbb{N}$ is determined by the recursive relation

$$X_n=\begin{cases}\alpha_{z_n}*X_{n-1}(z_{n-1})+\varepsilon_n(z_n,z_n-1) & \text{w.p. } \phi_{1,P_n}^{z_n},\\ \alpha_{z_n}*X_{n-2}(z_{n-2})+\varepsilon_n(z_n,z_n-2) & \text{w.p. } \phi_{2,P_n}^{z_n},\\ \vdots & \\ \alpha_{z_n}*X_{n-P_n}(z_{n-P_n})+\varepsilon_n(z_n,z_n-P_n) & \text{w.p. } \phi_{P_n,P_n}^{z_n},\end{cases}$$

where $\{z_n\}_{n=1}^{\infty}$ is the realization of the random environment process $\{Z_n\}_{n=1}^{\infty}$ (which is a Markov chain) whose elements take values in $E_r=\{1,...,r\}$, $r\in\mathbb{N}$, for r being the number of different environment states. Probabilities $\phi_{i,P_n}^{z_n}$, $i=1,2,...,P_n$, are all in $[0,1]$ and $\sum_{i=1}^{P_n}\phi_{i,P_n}^{z_n}=1$. In addition, "$\alpha*$", $\alpha\in(0,1)$, denotes the negative binomial thinning operator defined as $\alpha*X=\sum_{i=1}^{X}U_i$. Such defined thinning operator assigns to each integer-valued random variable X the sum of X independent random variables having the same geometric distribution with the mean α. Sets $\mathcal{M}=\{\mu_1,...,\mu_r\}$, $\mathcal{A}=\{\alpha_1,...,\alpha_r\}$, $\mathcal{P}=\{p_1,...,p_r\}$ contain parameter values of the model, μ_{z_n} is the mean of the marginal geometric distribution of $X_n(z_n)$, α_{z_n} is the thinning parameter value and p_{z_n} represents the maximal value that the order P_n may take for a fixed state $z_n\in\{1,...,r\}$.

Now, let us define the thinning operator "$\alpha\odot$" as it was done in [17]. Let Y be a random variable with discrete Laplace distribution DL$(\mu/(1+\mu))$, $\mu>0$, with probability

mass function given by

$$P(Y=y)=\frac{1}{1+2\mu}\left(\frac{\mu}{1+\mu}\right)^{|y|},\quad y=0,\pm1,\pm2,...,$$

and let $X^{(1)}$ and $X^{(2)}$ be two independent random variables with the same Geom $(\mu/(1+\mu))$ distribution. In that case, operator "$\alpha\odot$" is defined as

$$\alpha\odot Y|Y\stackrel{d}{=}\left(\alpha*X^{(1)}-\alpha*X^{(2)}\right)|\left(X^{(1)}-X^{(2)}\right), \tag{2.1}$$

where "$\alpha*$", $\alpha\in(0,1)$, represents the negative binomial thinning operator. In addition, the counting sequences involved in $\alpha*X^{(1)}$ and $\alpha*X^{(2)}$ are mutually independent and independent of random variables $X^{(1)}$ and $X^{(2)}$.

Using this newly defined thinning operator, [17] defined the DLINAR(1) process in the following way:

$$Y_n=\alpha\odot Y_{n-1}+e_n,\ n\in\mathbb{N},$$

where $\{Y_n\}$ represents a discrete Laplace distributed process, while $\{e_n\}$ is an innovation sequence of independent and identically distributed (i.i.d.) random variables, such that e_n and Y_{n-l} are mutually independent for all $l>0$.

For the purpose of better understanding the content which follows, it is convenient to introduce here as well the skewed discrete Laplace distribution $\mathrm{SDL}(\mu/(1+\mu),\nu/(1+\nu))$, $\mu>0,\ \nu>0$, with probability mass function given by

$$P(Y=y)=\begin{cases}\dfrac{1}{1+\mu+\nu}\left(\dfrac{\mu}{1+\mu}\right)^{y}, & y\geq 0,\\[2ex] \dfrac{1}{1+\mu+\nu}\left(\dfrac{\nu}{1+\nu}\right)^{-y}, & y<0.\end{cases}$$

Following notations given in the definition of $\mathrm{RrNGINAR}(\mathcal{M},\mathcal{A},\mathcal{P})$, let $E_r=\{1,2,...,r\}$ be the set of all possible environment states, where $r\in\mathbb{N}$, and let $\{z_n\},\ n\in\mathbb{N}_0$, be a realization of the r states random environment process $\{Z_n\}$. For $i,j\in E_r$, let $\{e_n(i,j),\ n\in\mathbb{N}\}$ be the sequences of i.i.d. random variables. Notation $Y_n(z_n)$ will be used to tag an element of a new process, where z_n represents the realized value of the random environment process in moment $n\geq 0$. Regarding this, let us introduce the following notations:

$$\begin{aligned} Y_n(Z_n) &= \sum_{z=1}^{r} Y_n(z)I_{\{Z_n=z\}},\\ e_n(Z_{n-1},Z_n) &= \sum_{z_1=1}^{r}\sum_{z_2=1}^{r} e_n(z_1,z_2)I_{\{Z_{n-1}=z_1,Z_n=z_2\}},\\ \alpha_{Z_n} &= \sum_{z=1}^{r}\alpha_z I_{\{Z_n=z\}}, \end{aligned}$$

whereby $I_{\{Z_n=z\}}$ represents an indicator random variable associated with the event $Z_n=z$.

Before introducing the definition of a new process that will be in the focus of this research, it is necessary to define a random environment INAR process based on the thinning operator "$\alpha\odot$", with variable marginal distribution and inconstant thinning parameter value.

Definition 2.1. Let $\{Z_n\}$ be a random environment process with r possible states from the set $E_r = \{1, 2, ..., r\}$, $r \in \mathbb{N}$. Let $\mathcal{M} = \{\mu_1, \mu_2, ..., \mu_r\}$ and $\mathcal{A} = \{\alpha_1, \alpha_2, ..., \alpha_r\}$, with $\mu_i > 0$ and $\alpha_i \in (0, 1)$, for all $i \in E_r$. We say that $\{Y_n(Z_n)\}$ is a random environment INAR process of order 1 based on the thinning operator "$\alpha \odot$", with r states, distribution parameters set $\mathcal{M}$ and thinning parameters set $\mathcal{A}$ ($\mathrm{RrINAR}_1(\mathcal{M}, \mathcal{A})$), if the random variable $Y_n(Z_n)$ is defined for $n \geq 1$ as

$$Y_n(Z_n) = \alpha_{Z_n} \odot Y_{n-1}(Z_{n-1}) + e_n(Z_{n-1}, Z_n), \tag{2.2}$$

where "$\alpha_{Z_n} \odot$" is defined by (2.1) and the following conditions are satisfied:

1. For fixed $i, j \in E_r$, the sequence $\{e_n(i, j)\}_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables;
2. The sequences of random variables $\{Z_n\}, \{e_n(1, 1)\}, \{e_n(1, 2)\}, ..., \{e_n(r, r)\}$ are mutually independent;
3. Random variables Z_m and $e_m(i, j)$ are independent of $Y_n(l)$ for all $n < m$ and all $i, j, l \in E_r$.

It is convenient now to define a random environment INAR process of order 1 with discrete Laplace marginals. In order to simplify the process, we can assume we know a realization $\{z_n\}$ of the random environment process $\{Z_n\}$. This assumption is plausible, since the estimate of $\{z_n\}$ can be easily obtained by applying the appropriate clustering procedure.

Definition 2.2. Let $\{z_n\}$ be a realization of the random environment process $\{Z_n\}$ with r possible states from the set $E_r = \{1, 2, ..., r\}$, $r \in \mathbb{N}$, and let $\mathcal{M} = \{\mu_1, \mu_2, ..., \mu_r\}$ and $\mathcal{A} = \{\alpha_1, \alpha_2, ..., \alpha_r\}$, with $\mu_i > 0$ and $\alpha_i \in (0, 1)$, for all $i \in E_r$. We say that $\{Y_n(z_n)\}$ is a random environment discrete Laplace INAR process of order 1 with r states, distribution parameters set $\mathcal{M}$ and thinning parameters set $\mathcal{A}$ ($\mathrm{RrDLINAR}_1(\mathcal{M}, \mathcal{A})$), if the random variable $Y_n(z_n)$ satisfies

$$Y_n(z_n) = \alpha_{z_n} \odot Y_{n-1}(z_{n-1}) + e_n(z_{n-1}, z_n), \tag{2.3}$$

for $n \geq 1$, where conditions 1-3 from Definition 2.1 are satisfied and the random variable $Y_n(z_n)$ has $\mathrm{DL}(\mu_{z_n}/(1 + \mu_{z_n}))$ distribution, for all $n \in \mathbb{N}_0$.

The introduced process is fully determined if the distributions of random variables $e_n(i, j)$ are known for all $n \geq 1$ and all $i, j \in E_r$. The following theorem reveals distributions of these random variables.

Theorem 2.1. *Let $\{Y_n(z_n)\}$ be a RrDLINAR$_1(\mathcal{M}, \mathcal{A})$ process. Let us suppose that $z_n = j$ and $z_{n-1} = k$ for some k and $j \in E_r$. If $0 < \alpha_j \leq \frac{\mu_j}{1+\max_{i\in E_r} \mu_i}$, then the distribution of the random variable $e_n(k, j)$ can be written as a mixture of discrete Laplace and skewed discrete Laplace distributed random variables in the following form:*

$$e_n(k, j) \stackrel{d}{=} \begin{cases} DL\left(\frac{\mu_j}{1+\mu_j}\right), & w.p.\ \left(1 - \frac{\alpha_j\mu_k}{\mu_j - \alpha_j}\right)^2, \\ SDL\left(\frac{\mu_j}{1+\mu_j}, \frac{\alpha_j}{1+\alpha_j}\right), & w.p.\ \frac{\alpha_j\mu_k}{\mu_j - \alpha_j}\left(1 - \frac{\alpha_j\mu_k}{\mu_j - \alpha_j}\right), \\ SDL\left(\frac{\alpha_j}{1+\alpha_j}, \frac{\mu_j}{1+\mu_j},\right), & w.p.\ \frac{\alpha_j\mu_k}{\mu_j - \alpha_j}\left(1 - \frac{\alpha_j\mu_k}{\mu_j - \alpha_j}\right), \\ DL\left(\frac{\alpha_j}{1+\alpha_j}\right), & w.p.\ \left(\frac{\alpha_j\mu_k}{\mu_j - \alpha_j}\right)^2. \end{cases} \tag{2.4}$$

Proof: Let $\varphi_{e_n(k,j)}(t)$ represents the characteristic function of the random variable $e_n(k,j)$. Based on the definition and properties of the process and the assumption that $z_{n-1}=k$ and $z_n=j$, it holds

$$\varphi_{e_n(k,j)}(t)=\frac{\varphi_{Y_n(j)}(t)}{\varphi_{\alpha_j\odot Y_{n-1}(k)}(t)}.$$

As can be seen in [17],

$$\begin{aligned}\varphi_{Y_n(j)}(t) &= \frac{1}{(1+\mu_j-\mu_je^{it})(1+\mu_j-\mu_je^{-it})},\\ \varphi_{\alpha_j\odot Y_{n-1}(k)}(t) &= \frac{(1+\alpha_j-\alpha_je^{it})(1+\alpha_j-\alpha_je^{-it})}{(1+\alpha_j(1+\mu_k)-\alpha_j(1+\mu_k)e^{it})(1+\alpha_j(1+\mu_k)-\alpha_j(1+\mu_k)e^{-it})}.\end{aligned}$$

Using these facts, we obtain that

$$\begin{aligned}\varphi_{e_n(k,j)}(t) &= \frac{[1+\alpha_j(1+\mu_k)-\alpha_j(1+\mu_k)e^{it}][1+\alpha_j(1+\mu_k)-\alpha_j(1+\mu_k)e^{-it}]}{(1+\alpha_j-\alpha_je^{it})(1+\alpha_j-\alpha_je^{-it})(1+\mu_j-\mu_je^{it})(1+\mu_j-\mu_je^{-it})}\\ &= \frac{A}{(1+\alpha_j-\alpha_je^{it})(1+\alpha_j-\alpha_je^{-it})}+\frac{B}{(1+\alpha_j-\alpha_je^{it})(1+\mu_j-\mu_je^{-it})}\\ &\quad+\frac{C}{(1+\mu_j-\mu_je^{it})(1+\mu_j-\mu_je^{-it})}+\frac{D}{(1+\mu_j-\mu_je^{it})(1+\alpha_j-\alpha_je^{-it})}.\end{aligned}$$

By solving the system

$$\begin{aligned}AM^2+BKM+CK^2+DMK &= (1+\alpha_j(1+\mu_k))^2,\\ AM\mu_j+BK\mu_j+CK\alpha_j+DM\alpha_j &= \alpha_j(1+\mu_k)(1+\alpha_j(1+\mu_k)),\\ AM\mu_j+BM\alpha_j+CK\alpha_j+DK\mu_j &= \alpha_j(1+\mu_k)(1+\alpha_j(1+\mu_k)),\\ A\mu_j^2+B\alpha_j\mu_j+C\alpha_j^2+D\alpha_j\mu_j &= (\alpha_j(1+\mu_k))^2,\end{aligned}$$

where $K=1+\alpha_j$ and $M=1+\mu_j$, we obtain that

$$A=\left(\frac{\alpha_j\mu_k}{\mu_j-\alpha_j}\right)^2,\ B=D=\frac{\alpha_j\mu_k}{\mu_j-\alpha_j}\left(1-\frac{\alpha_j\mu_k}{\mu_j-\alpha_j}\right),\ C=\left(1-\frac{\alpha_j\mu_k}{\mu_j-\alpha_j}\right)^2.$$

Knowing that the characteristic functions of random variables with $\mathrm{DL}\left(\frac{\mu_j}{1+\mu_j}\right)$, $\mathrm{SDL}\left(\frac{\mu_j}{1+\mu_j},\frac{\alpha_j}{1+\alpha_j}\right)$, $\mathrm{SDL}\left(\frac{\alpha_j}{1+\alpha_j},\frac{\mu_j}{1+\mu_j}\right)$ and $\mathrm{DL}\left(\frac{\alpha_j}{1+\alpha_j}\right)$ distributions are of the form

$$\begin{aligned}\varphi_1(t) &= \frac{1}{(1+\mu_j-\mu_je^{it})(1+\mu_j-\mu_je^{-it})},\quad \varphi_2(t)=\frac{1}{(1+\mu_j-\mu_je^{it})(1+\alpha_j-\alpha_je^{-it})},\\ \varphi_3(t) &= \frac{1}{(1+\alpha_j-\alpha_je^{it})(1+\mu_j-\mu_je^{-it})},\quad \varphi_4(t)=\frac{1}{(1+\alpha_j-\alpha_je^{it})(1+\alpha_j-\alpha_je^{-it})}\end{aligned}$$

respectively, it becomes obvious that (2.4) holds.

It is left to provide that A,B,C and D are probabilities, i.e. that $A+B+C+D=1$ and all of them belong to $[0,1]$. First condition is easily confirmable. To provide the second one, it is enough to confirm that $0\le\frac{\alpha_j\mu_k}{\mu_j-\alpha_j}\le 1$. By solving this double inequality, we get $\alpha_j\le\frac{\mu_j}{1+\mu_k}$. Since this condition must hold for an arbitrary k and j, and $\alpha_j\in(0,1)$, we have that $0<\alpha_j\le\frac{\mu_j}{1+\max_{i\in E_r}\mu_i}$. This fact completes the proof. □

According to the previous theorem and the fact that discrete Laplace and skewed discrete Laplace distributed random variables can be represented as a difference of two random variables with geometric distributions, it is possible to make an interesting conclusion.

Corollary 2.1. *If* $0<\alpha_j\leq\frac{\mu_j}{1+\max_{i\in E_r}\mu_i}$, *then* $e_n(i,j)\overset{d}{=}\varepsilon_n(i,j)-\eta_n(i,j)$, *where* $\varepsilon_n(i,j)$ *and* $\eta_n(i,j)$ *are two i.i.d. random variables with the distribution given as*

$$\begin{cases} Geom\left(\dfrac{\mu_j}{1+\mu_j}\right), & w.p.\ \left(1-\dfrac{\alpha_j\mu_i}{\mu_j-\alpha_j}\right), \\ Geom\left(\dfrac{\alpha_j}{1+\alpha_j}\right), & w.p.\ \dfrac{\alpha_j\mu_i}{\mu_j-\alpha_j}. \end{cases} \tag{2.5}$$

Presenting the distribution of the innovation time series $\{e_n(i,j)\},\ n\in\mathbb{N}$, in this shape may simplify the calculation of many properties of the process itself, as can be seen in the following corollary.

Corollary 2.2. *Let us suppose that* $z_n=j$ *and* $z_{n-1}=i$, *for some* i *and* $j\in E_r$. *Then, we have:*

$$\begin{aligned} E(e_n(i,j)) &= 0, \\ \mathrm{Var}(e_n(i,j)) &= 2(\mu_j(1+\mu_j)-\alpha_j\mu_i(1+2\alpha_j+\alpha_j\mu_i)). \end{aligned}$$

Proof: The proof of the first equality is trivial, given that $\varepsilon_n(i,j)$ and $\eta_n(i,j)$ have the same distribution.
Bearing in mind the shape of the distribution of $\varepsilon_n(i,j)$ and $\eta_n(i,j)$ and using properties of the probability generating function (p.g.f.), it is easy to prove that

$$\begin{aligned} \mathrm{Var}(\eta_n(i,j))=\mathrm{Var}(\varepsilon_n(i,j)) &= \Phi''_{\varepsilon_n(i,j)}(1)+\Phi'_{\varepsilon_n(i,j)}(1)-\left[\Phi'_{\varepsilon_n(i,j)}(1)\right]^2 \\ &= \mu_j(1+\mu_j)-\alpha_j\mu_i(1+2\alpha_j+\alpha_j\mu_i). \end{aligned}$$

Now, it is obvious that

$$\mathrm{Var}(e_n(i,j))=\mathrm{Var}(\varepsilon_n(i,j))+\mathrm{Var}(\eta_n(i,j))=2(\mu_j(1+\mu_j)-\alpha_j\mu_i(1+2\alpha_j+\alpha_j\mu_i)).$$

□

Remark 2.1. Let us highlight here two interesting facts:

- For $z_n=j$ and $z_{n-1}=i$, $\varepsilon_n(i,j)$ and $\eta_n(i,j)$ have the same distribution as an innovation process given in RrNGINAR$(\mathcal{M},\mathcal{A},\mathcal{P})$ model (see [11]);
- For $j=i$, the distribution of the innovation process $\{e_n(i,j)\}$ coincides with the distribution of the innovation process of the DLINAR(1) model (see [17]).

3. PROPERTIES OF THE PROCESS

In this section, the most important properties of the $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ model will be derived and analyzed. It is interesting to notice that many properties can be derived by observing $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ process as a difference of two independent $\mathrm{RrNGINAR}(\mathcal{M},\mathcal{A},\mathcal{P})$ processes, in case of $\mathcal{P}=\{1\}$.

To that purpose, for given sets $\mathcal{M},\mathcal{A}$ and $\mathcal{P}=\{1\}$, let us define two $\mathrm{RrNGINAR}(\mathcal{M},\mathcal{A},\mathcal{P})$ time series

$$\begin{aligned} X_n^{(1)}(z_n) &= \alpha_{z_n} * X_{n-1}^{(1)}(z_{n-1}) + \varepsilon_n(z_{n-1},z_n), \quad n\geq 1,\\ X_n^{(2)}(z_n) &= \alpha_{z_n} * X_{n-1}^{(2)}(z_{n-1}) + \eta_n(z_{n-1},z_n), \quad n\geq 1. \end{aligned}$$

Time series $\{X_n^{(1)}(z_n)\}$ and $\{X_n^{(2)}(z_n)\}$ are mutually independent and, for fixed $z_n=j$, $X_n^{(1)}(j)$ and $X_n^{(2)}(j)$ have the same $\mathrm{Geom}\Big(\frac{\mu_j}{1+\mu_j}\Big)$, $\mu_j\in\mathcal{M}$, distribution. Also, for fixed values $z_n=j$ and $z_{n-1}=i$, $\{\varepsilon_n(i,j)\}$ and $\{\eta_n(i,j)\}$ are two mutually independent time series with the same marginal distribution given in Corollary 2.1. Based on the definition of the $\mathrm{RrNGINAR}(\mathcal{M},\mathcal{A},\mathcal{P})$ process, $X_{n-l}^{(1)}(k)$ and $\varepsilon_n(i,j)$, as well as $X_{n-l}^{(2)}(k)$ and $\eta_n(i,j)$, are mutually independent for all $l\geq 1$ and for all $i,j,i\in E_r$.

Let $Y_n(z_n)$ be a $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ process with $\mathrm{DL}\Big(\frac{\mu_j}{1+\mu_j}\Big)$ marginals, given $z_n=j$. Now, using Corollary 2.1 and Corollary 2.1 from [17], we have

$$\begin{aligned} X_n^{(1)}(z_n) - X_n^{(2)}(z_n) &= \Big(\alpha_{z_n} * X_{n-1}^{(1)}(z_{n-1}) - \alpha_{z_n} * X_{n-1}^{(2)}(z_{n-1})\Big)\\ &\quad + \big(\varepsilon_n(z_{n-1},z_n) - \eta_n(z_{n-1},z_n)\big)\\ &\overset{d}{=} \alpha_{z_n}\odot Y_{n-1}(z_{n-1}) + e_n(z_{n-1},z_n) = Y_n(z_n). \end{aligned} \tag{3.1}$$

Now, it is easy to prove that $E(Y_n(z_n))=0$ and

$$\mathrm{Var}(Y_n(z_n)) = 2\,\mathrm{Var}\Big(X_n^{(1)}(z_n)\Big) = 2\mu_{z_n}(1+\mu_{z_n}).$$

One important property holds for $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ process. Namely, according to the Theorem 2.4 given in [17],

$$\alpha\odot Y \overset{d}{=} \mathrm{sgn}(Y)(\alpha * |Y|) + \sum_{j=1}^{\min\left(X^{(1)},X^{(2)}\right)} D_j,$$

whereby the following conditions are satisfied:

a) $Y\sim \mathrm{DL}\Big(\frac{\mu}{1+\mu}\Big)$, $X^{(1)}\sim \mathrm{Geom}\Big(\frac{\mu}{1+\mu}\Big)$, $X^{(2)}\sim \mathrm{Geom}\Big(\frac{\mu}{1+\mu}\Big)$;

b) $D_j\sim \mathrm{DL}\Big(\frac{\alpha}{1+\alpha}\Big)$;

c) random variables $Y, X^{(1)}, X^{(2)}, D_j$, $j\geq 1$, and random variables involved in "$\alpha*$" are independent.

For $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ process, the following result holds.

Theorem 3.1. *The RrDLINAR$_1(\mathcal{M},\mathcal{A})$ process $\{Y_n(z_n)\}$ is a Markov process.*

Proof: Let us define sets A and B as $A=\{Y_s(z_s)=y_s,\ s=0,1,...,n-2\}$ and $B=A\cup\{Y_{n-1}(z_{n-1})=y_{n-1}\}$. According to the property of "$\alpha\odot$" mentioned above, it holds

$$\alpha_{z_n}\odot Y_{n-1}(z_{n-1}) = \operatorname{sgn}(Y_{n-1}(z_{n-1}))(\alpha_{z_n}*|Y_{n-1}(z_{n-1})|) + \sum_{j=1}^{\min\left(X_{n-1}^{(1)}(z_{n-1}),X_{n-1}^{(2)}(z_{n-1})\right)} D_j(z_n),$$

whereby $X_{n-1}^{(1)}(z_{n-1})$ and $X_{n-1}^{(2)}(z_{n-1})$ have the same $\text{Geom}\left(\frac{\mu_{z_{n-1}}}{1+\mu_{z_{n-1}}}\right)$ distribution and $D_j(z_n)$ has the $\text{DL}\left(\frac{\alpha_{z_n}}{1+\alpha_{z_n}}\right)$ distribution. Now, we have

$$P(Y_n(z_n)=y_n|B) = P\Bigg(\operatorname{sgn}(Y_{n-1}(z_{n-1}))(\alpha_{z_n}*|Y_{n-1}(z_{n-1})|) + \sum_{j=1}^{\min\left(X_{n-1}^{(1)}(z_{n-1}),X_{n-1}^{(2)}(z_{n-1})\right)} D_j(z_n)+e_n(z_{n-1},z_n)=y_n\Big|B\Bigg).$$

Bearing in mind property (c) mentioned above and condition (3) of the Definition 2.1, it becomes obvious that

$$\begin{aligned}P(Y_n(z_n)=y_n|B) &= \sum_{j=-\infty}^{+\infty} P(\operatorname{sgn}(Y_{n-1}(z_{n-1}))(\alpha_{z_n}*|Y_{n-1}(z_{n-1})|)=j|B)\\ &\quad\times P\Bigg(\sum_{j=1}^{\min\left(X_{n-1}^{(1)}(z_{n-1}),X_{n-1}^{(2)}(z_{n-1})\right)} D_j(z_n)+e_n(z_{n-1},z_n)=y_n-j\Bigg)\\ &= \sum_{j=-\infty}^{+\infty}\binom{|y_{n-1}|+|j|-1}{|j|}\frac{\alpha_{z_n}^{|j|}}{(1+\alpha_{z_n})^{|y_{n-1}|+|j|}}\\ &\quad\times P\Bigg(\sum_{j=1}^{\min\left(X_{n-1}^{(1)}(z_{n-1}),X_{n-1}^{(2)}(z_{n-1})\right)} D_j(z_n)+e_n(z_{n-1},z_n)=y_n-j\Bigg).\end{aligned}$$

As the last expression depends only on y_{n-1}, it is obvious that RrDLINAR$_1(\mathcal{M},\mathcal{A})$ is a Markov process. □

3.1. The k-step ahead conditional expectation

Theorem 3.2. *Let $\{Y_n(z_n)\}$ be a RrDLINAR$_1(\mathcal{M},\mathcal{A})$ process. Then for $k\geq 1$,*

$$E(Y_{n+k}(z_{n+k})|Y_n(z_n)) = \left(\prod_{j=1}^{k}\alpha_{z_{n+j}}\right)Y_n(z_n). \tag{3.2}$$

Proof: The proof will be derived by induction. Let $k = 1$. Using Theorem 2.3 from [17], we have

$$\begin{aligned} E(Y_{n+1}(z_{n+1})|Y_n(z_n)) &= E(\alpha_{z_{n+1}} \odot Y_n(z_n)|Y_n(z_n)) + E(e_{n+1}(z_n, z_{n+1})) \\ &= \alpha_{z_{n+1}} Y_n(z_n). \end{aligned} \tag{3.3}$$

Suppose the equality (3.2) holds for $k < m$. Bearing in mind the Markov property of the RrDLINAR$_1(\mathcal{M}, \mathcal{A})$ process, we will prove that (3.2) holds for $k = m$ as well. Namely,

$$\begin{aligned} E(Y_{n+m}(z_{n+m})|Y_n(z_n)) &= E\big[E(Y_{n+m}(z_{n+m})|Y_{n+m-1}(z_{n+m-1}), ..., Y_n(z_n))|Y_n(z_n)\big] \\ &= E\big[E(Y_{n+m}(z_{n+m})|Y_{n+m-1}(z_{n+m-1}))|Y_n(z_n)\big] \\ &= E(\alpha_{z_{n+m}} Y_{n+m-1}(z_{n+m-1})|Y_n(z_n)) \\ &= \alpha_{z_{n+m}} \left(\prod_{j=1}^{m-1} \alpha_{z_{n+j}} \right) Y_n(z_n) \\ &= \left(\prod_{j=1}^{m} \alpha_{z_{n+j}} \right) Y_n(z_n). \end{aligned}$$

This completes the proof of this theorem. □

3.2. Correlation structure

Theorem 3.3. *A RrDLINAR$_1(\mathcal{M}, \mathcal{A})$ process $\{Y_n(z_n)\}$ given by (2.3) is the correlated process with*

$$\operatorname{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) = \begin{cases} \left(\prod_{j=0}^{k-1} \alpha_{z_{n-j}} \right) \sqrt{\dfrac{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}{\mu_{z_n}(1+\mu_{z_n})}}, & k \geq 0, \\ \left(\prod_{j=1}^{-k} \alpha_{z_{n+j}} \right) \sqrt{\dfrac{\mu_{z_n}(1+\mu_{z_n})}{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}}, & k < 0. \end{cases} \tag{3.4}$$

Proof: Since $\{Y_n(z_n)\}$ is a process with the k-step ahead conditional expectation of the form $E(Y_{n+k}(z_{n+k})|Y_n(z_n)) = \left(\prod_{j=1}^{k} \alpha_{z_{n+j}}\right) Y_n(z_n)$, unconditional expectation $E(Y_n(z_n)) = 0$ and finite variance $\operatorname{Var}(Y_n(z_n)) = 2\mu_{z_n}(1 + \mu_{z_n})$, for $k \geq 0$ it becomes easy to obtain

$$\begin{aligned} \operatorname{Cov}(Y_n(z_n), Y_{n-k}(z_{n-k})) &= \operatorname{Cov}(E(Y_n(z_n)|Y_{n-k}(z_{n-k})), Y_{n-k}(z_{n-k})) \\ &= \operatorname{Cov}\left(\left(\prod_{j=0}^{k-1} \alpha_{z_{n-j}} \right) Y_{n-k}(z_{n-k}), Y_{n-k}(z_{n-k}) \right) \\ &= \left(\prod_{j=0}^{k-1} \alpha_{z_{n-j}} \right) \operatorname{Var}(Y_{n-k}(z_{n-k})) \\ &= 2 \left(\prod_{j=0}^{k-1} \alpha_{z_{n-j}} \right) \mu_{z_{n-k}}(1 + \mu_{z_{n-k}}), \end{aligned}$$

whence we have

$$\begin{aligned}\mathrm{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) &= \frac{2\left(\prod_{j=0}^{k-1}\alpha_{z_{n-j}}\right)\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}{\sqrt{2\mu_{z_n}(1+\mu_{z_n})2\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}} \\ &= \left(\prod_{j=0}^{k-1}\alpha_{z_{n-j}}\right)\sqrt{\frac{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}{\mu_{z_n}(1+\mu_{z_n})}}.\end{aligned}$$

Similar to this, for $k<0$ we obtain

$$\begin{aligned}\mathrm{Cov}(Y_n(z_n), Y_{n-k}(z_{n-k})) &= E(Y_n(z_n)\cdot Y_{n-k}(z_{n-k})) \\ &= E[E(Y_n(z_n)Y_{n-k}(z_{n-k})|Y_n(z_n))] \\ &= E\left(Y_n(z_n)\left(\prod_{j=1}^{-k}\alpha_{z_{n+j}}\right)Y_n(z_n)\right) \\ &= \left(\prod_{j=1}^{-k}\alpha_{z_{n+j}}\right)\mathrm{Var}(Y_n(z_n)) \\ &= \left(\prod_{j=1}^{-k}\alpha_{z_{n+j}}\right)2\mu_{z_n}(1+\mu_{z_n}),\end{aligned}$$

whence we have

$$\mathrm{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) = \left(\prod_{j=1}^{-k}\alpha_{z_{n+j}}\right)\sqrt{\frac{\mu_{z_n}(1+\mu_{z_n})}{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}}.$$

□

Remark 3.1. If $z_n = z_{n-1} = \cdots = z_{n-k} = j$, then it holds that $\mathrm{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) = \alpha_j^{|k|}$, which matches with correlation function of the DLINAR(1) process.

Bearing in mind the equality (3.4) and the facts that $\mu_{z_{n-k}} > 0$, $\mu_{z_n} > 0$ and $\alpha_{z_{n-j}} > 0$ for all $j = 0, 1, ..., k-1$, it is obvious that in case of $k \geq 0$, $\mathrm{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) > 0$. Let us prove now the validity of the relation $\mathrm{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) < 1$.

For all $j = 0, 1, ..., k-1$, $\alpha_{z_{n-j}} \leq \frac{\mu_{z_{n-j}}}{1+\max_{i\in E_r}\mu_{z_i}}$, so, obviously

$$\alpha_{z_{n-j}} \leq \frac{\mu_{z_{n-j}}}{1+\mu_{z_{n-j-1}}} < \frac{\mu_{z_{n-j}}}{\mu_{z_{n-j-1}}} < \frac{1+\mu_{z_{n-j}}}{\mu_{z_{n-j-1}}}.$$

Then,

$$\alpha_{z_{n-j}}^2 < \frac{\mu_{z_{n-j}}}{1+\mu_{z_{n-j-1}}}\cdot\frac{1+\mu_{z_{n-j}}}{\mu_{z_{n-j-1}}},$$

so we can conclude that $\alpha_{z_{n-j}} < \sqrt{\frac{\mu_{z_{n-j}}(1+\mu_{z_{n-j}})}{\mu_{z_{n-j-1}}(1+\mu_{z_{n-j-1}})}}$, and further, that

$$\begin{aligned}\prod_{j=0}^{k-1}\alpha_{z_{n-j}} &< \sqrt{\frac{\mu_{z_n}(1+\mu_{z_n})}{\mu_{z_{n-1}}(1+\mu_{z_{n-1}})}}\sqrt{\frac{\mu_{z_{n-1}}(1+\mu_{z_{n-1}})}{\mu_{z_{n-2}}(1+\mu_{z_{n-2}})}}\cdots\sqrt{\frac{\mu_{z_{n-k+1}}(1+\mu_{z_{n-k+1}})}{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}} \\ &= \sqrt{\frac{\mu_{z_n}(1+\mu_{z_n})}{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}}.\end{aligned}$$

Finally, it holds

$$\mathrm{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) < \sqrt{\frac{\mu_{z_n}(1+\mu_{z_n})}{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}}\sqrt{\frac{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}{\mu_{z_n}(1+\mu_{z_n})}} = 1.$$

Similarly, it can be shown that $0 < \mathrm{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) < 1$, for $k < 0$.

4. YULE-WALKER ESTIMATION

In this section, the YW estimators will be provided and their strong consistency will be proven. To that purpose, we will use procedure similar to the one described in [18].

Thus, let $Y_1(z_1), Y_2(z_2), ..., Y_N(z_N)$ be a sample of the RrDLINAR$_1(\mathcal{M}, \mathcal{A})$ process $\{Y_n(z_n)\}$ of size N. The main idea of the procedure described in [18] is to estimate parameters μ_k and α_k only by using elements corresponding to the state k. Thus, let us divide the initial sample into r subsamples S_k, $k = 1, 2, ..., r$, where S_k is a subsample which contains all the elements corresponding to the state k and doesn't contain elements corresponding to any other state. This division can be performed in the following way:

$$\begin{aligned} I_k &= \{i \in \{1, 2, ..., N\} | z_i = k\}, \quad k \in \{1, 2, ..., r\}, \\ \bigcup_{k=1}^{r} I_k &= \{1, 2, ..., N\}, \quad |I_k| = n_k, \quad n_1 + n_2 + \cdots + n_r = N, \\ S_k &= \{Y_{k_1}(k), Y_{k_2}(k), ..., Y_{k_{n_k}}(k)\}, \quad k_j \in I_k, \quad k_j < k_{j+1}, \quad \forall j \in \{1, 2, ..., n_k - 1\}. \end{aligned}$$

In more detail, each S_k, $k = 1, 2, ..., r$, represents a disjoint union of subsamples $S_{k,1}, S_{k,2}, ..., S_{k,i_k}$, which we call 'maximal' subsamples. For an arbitrary subsample $S_{k,l}$, $l = 1, 2, ..., i_k$, we can find natural numbers m_l and n_l, $m_l < n_l$, such that $z_{m_l} \neq k$, $z_{m_l+1} = z_{m_l+2} = \cdots = z_{n_l} = k$, $z_{n_l+1} \neq k$. In that case, the subsample $S_{k,l} = \{Y_{m_l+1}(z_{m_l+1}), Y_{m_l+2}(z_{m_l+2}), ..., Y_{n_l}(z_{n_l})\}$ corresponds to the state k and is maximal in the sense that it cannot be expanded neither to the left nor right side in the way that all of its elements correspond to the state k. Now, each of those maximal subsamples $S_{k,l}$, $l = 1, 2, ..., i_k$ may be observed as a sample of some DLINAR(1) process with the marginal distribution parameter μ_k. Let us introduce the following notation: $J_{k,l} = \{i \in \{1, 2, ..., N\} | Y_i(z_i) \in S_{k,l}\}$, $|J_{k,l}| = n_{k,l}$ for all $l = 1, 2, ..., i_k$ and $n_{k,1} + n_{k,2} + \cdots + n_{k,i_k} = n_k$. As shown in [17], the DLINAR(1) process is stationary and ergodic, and the corresponding sample variance and the first-order sample covariance are strongly consistent estimates of the variance and the first-order covariance of the process. Finally, in case of subsample $S_{k,l}$, these estimators are of the form

$$\widehat{\gamma}_{0,l}^{(k)} = \frac{1}{n_{k,l}} \sum_{i \in J_{k,l}} Y_i^2(k) \quad \text{and} \quad \widehat{\gamma}_{1,l}^{(k)} = \frac{1}{n_{k,l}} \sum_{i, i+1 \in J_{k,l}} Y_i(k) Y_{i+1}(k).$$

Let us define now the corresponding estimators without taking maximal subsamples into account.

Definition 4.1. Estimators obtained from the subsample S_k corresponding to the state k are defined as

$$\widehat{\gamma}_0^{(k)} = \frac{1}{n_k} \sum_{i \in I_k} Y_i^2(k), \quad \widehat{\gamma}_1^{(k)} = \frac{1}{n_k} \sum_{i, i+1 \in I_k} Y_i(k) Y_{i+1}(k). \tag{4.1}$$

Theorem 4.1. *Estimators* $\widehat{\gamma}_0^{(k)}$ *and* $\widehat{\gamma}_1^{(k)}$ *from Definition 4.1 are strongly consistent.*

Proof: This theorem shall be proven in a similar way as it was done by [18]. First of all, the strong consistency property for $\widehat{\gamma}_0^{(k)}$ shall be proven. Because $\widehat{\gamma}_{0,l}^{(k)}$ is strongly consistent for all $l \in \{1, 2, ..., i_k\}$ it holds that $\widehat{\gamma}_{0,l}^{(k)} \to \gamma_0^k$, $n_{k,l} \to \infty$ everywhere except on the set $\Omega_{k,l}$, where $P(\Omega_{k,l}) = 0$. Now, it holds that

$$\widehat{\gamma}_0^{(k)} = \frac{1}{n_k} \sum_{i \in I_k} Y_i^2(k) = \frac{1}{n_k} \sum_{l=1}^{i_k} \sum_{i \in J_{k,l}} Y_i^2(k) = \sum_{l=1}^{i_k} \frac{n_{k,l}}{n_k} \frac{1}{n_{k,l}} \sum_{i \in J_{k,l}} Y_i^2(k) = \sum_{l=1}^{i_k} \frac{n_{k,l}}{n_k} \widehat{\gamma}_{0,l}^{(k)}.$$

Let $n_k \to \infty$. Following the technique introduced by [18], it is easy to show that

$$\lim_{n_k \to \infty} \widehat{\gamma}_0^{(k)} = \lim_{n_k \to \infty} \sum_{l=1}^{d} \frac{n_{k,l}}{n_k} \widehat{\gamma}_{0,l}^{(k)},$$

where $n_{k,l}$, $l = 1, 2, ..., d$, represent those maximal sample sizes which approach infinity when n_k does so. Thus, we have

$$\begin{aligned} \lim_{n_k \to \infty} \widehat{\gamma}_0^{(k)} &= \lim_{n_{k,l} \to \infty,\ \forall l \in \{1,2,...,d\}} \sum_{l=1}^{d} \frac{n_{k,l}}{n_k} \widehat{\gamma}_{0,l}^{(k)} \\ &= \gamma_0^{(k)} \lim_{n_{k,l} \to \infty,\ \forall l \in \{1,2,...,d\}} \sum_{l=1}^{d} \frac{n_{k,l}}{n_k} = \gamma_0^{(k)}. \end{aligned} \tag{4.2}$$

We mentioned earlier that $\lim_{n_{k,l} \to \infty} \widehat{\gamma}_{0,l}^{(k)} = \gamma_0^{(k)}$ everywhere except on the set $\Omega_{k,l}$, where $P(\Omega_{k,l}) = 0$. Thus, the equality (4.2) holds everywhere except on the set $\Omega_k = \bigcup_{l=1}^{d} \Omega_{k,l}$, where

$$P(\Omega_k) = P\left(\bigcup_{l=1}^{d} \Omega_{k,l} \right) \leq \sum_{l=1}^{d} P(\Omega_{k,l}) = 0.$$

From the non-negativity of the probability, it follows that $P(\Omega_k) = 0$. Hence, $\widehat{\gamma}_0^{(k)}$ is a strongly consistent estimator of the variance $\gamma_0^{(k)}$.

Proof for $\widehat{\gamma}_1^{(k)}$ is analogous to the one proposed for $\widehat{\gamma}_0^{(k)}$. □

It has remained to estimate parameters μ_k and α_k on the subsample S_k. According to Theorem 3.3 from [17], it holds

$$\gamma_0^{(k)} = 2\mu_k(1 + \mu_k), \quad \gamma_1^{(k)} = 2\alpha_k\mu_k(1 + \mu_k).$$

It follows that

$$\widehat{\mu}_k^{YW} = -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 2\widehat{\gamma}_0^{(k)}}, \quad \widehat{\alpha}_k^{YW} = \frac{\widehat{\gamma}_1^{(k)}}{\widehat{\gamma}_0^{(k)}}.$$

Now, it is obvious that $f(x) = -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 2x}$, $x \geq 0$ is a continuous function. According to Proposition 6.3.4 provided by [5], we conclude that $\widehat{\mu}_k^{YW}$ is strongly consistent. The strong consistency of $\widehat{\alpha}_k^{YW}$ follows from the Theorem 4.1 given in [8].

5. MODEL SIMULATIONS

We now focus on the YW estimating procedures on simulated data series, in order to clarify the utility of the observed non-stationary modeling and to justify the quality of the presented estimation method. Both processes, the random environment process $\{Z_n\}$ and the $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ process $\{Y_n(z_n)\}$ have been simultaneously simulated in 100 replicates, each of size 10000. Sequences $\{Y_n(z_n)\}$ are simulated using the fact that the newly defined process is distributed the same as a difference between two independent $\mathrm{RrNGINAR}(\mathcal{M},\mathcal{A},\mathcal{P})$ processes $\{X_n^{(1)}(z_n)\}$ and $\{X_n^{(2)}(z_n)\}$, with $\mathcal{P}=\{1\}$. Thus, first we simulated $\{X_n^{(1)}(z_n)\}$, and independently of it, $\{X_n^{(2)}(z_n)\}$ and derive $\{Y_n(z_n)\}$ as $Y_n(z_n)=X_n^{(1)}(z_n)-X_n^{(2)}(z_n),\ n\geq 1$. Considering the number of possible random states, we have decided to discuss, in our opinion, two of the most plausible cases in practice, as follows.

5.1. The case of two states

Here we have assumed that the random environment process is performed in two different states. Also, we have considered two different combinations of the model parameter values. Bearing in mind that $\alpha_j \in \left(0, \frac{\mu_j}{1+\max_{i\in E_r}\mu_i}\right]$, first of all we have analyzed the case when parameters $\alpha_j,\ j=1,2$, were close to their upper limits. So, we have used the following true values: $\mu=(1,3)$ and $\alpha=(0.25,0.7)$. It has been assumed that the choice of the initial random state is fair, which led us to have $p_{vec}=(0.5,0.5)$. It has remained for us to set the random environment process transition probability matrix. In order to preserve the simulated $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ process in one state as long as possible, we have chosen to prefer the present state of the random environment process, i.e. probabilities that the random environment process changes its state are significantly smaller. Thus, $p_{mat}=\begin{bmatrix}0.6 & 0.4\\ 0.2 & 0.8\end{bmatrix}$.

In the second case, we have chosen parameters $\alpha_j,\ j=1,2$, to be smaller than the midpoints of intervals $\left(0, \frac{\mu_j}{1+\max_{i\in E_r}\mu_i}\right],\ j=1,2$. In order not to shrink these intervals too much, parameters $\mu_j,\ j=1,2$, must have relatively close values. In that manner, we have chosen the following true parameter values: $\mu=(2,3)$ and $\alpha=(0.2,0.3)$. An initial state is nearly fair, due to the value of its distribution $p_{vec}=(0.45,0.55)$, and the random states transition probabilities are given as $p_{mat}=\begin{bmatrix}0.7 & 0.3\\ 0.3 & 0.7\end{bmatrix}$.

5.2. The case of three states

Similarly, as aforementioned, in the case of three different random states we have also considered two different combinations of the true parameter values. And yet again, the case when parameters α_j, $j=1,2,3$, approach their upper limits has been analyzed in the first place. The following true parameter values have been used: $\mu=(1,2,5)$ and $\alpha=(0.1,0.25,0.7)$.

An initial state is nearly fair, due to the value of its distribution $p_{vec} = (0.3, 0.4, 0.3)$, and the random environment process transition probability matrix is $p_{mat} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$.

At the very end, the second case represents the simulation when parameters α_j, $j = 1, 2, 3$, are smaller than the midpoints of corresponding intervals $\left(0, \frac{\mu_j}{1+\max_{i \in E_r} \mu_i}\right]$, $j = 1, 2, 3$. Thus, we have chosen the following true parameter values: $\mu = (2, 3, 5)$ and $\alpha = (0.1, 0.2, 0.4)$. An initial state is fair, due to the value of its distribution $p_{vec} = (0.33, 0.34, 0.33)$ and the random environment process transition probability matrix is $p_{mat} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$.

5.3. Estimation results

For both presented cases, $r = 2$ and $r = 3$, we estimate parameters μ_j and α_j, $j = 1, 2, ..., r$, of the RrDLINAR$_1(\mathcal{M}, \mathcal{A})$ model using the YW method. Compared to the DLINAR(1) model, the newly defined model has a greater number of unknown parameters, which leads to better fitting to the data, because it is more "flexible". However, since each state has its own parameters that can be estimated only based on the part of the sample corresponding to that state, it is expected to need a bigger sample for the same precision of the estimation.

Table 1: The case of 2 states.

a) True values $\mu = (1, 3)$, $\alpha = (0.25, 0.7)$, $p_{vec} = (0.5, 0.5)$, $p_{mat} = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$.

N_1	$\widehat{\mu}_1^{YW}$	$\widehat{\mu}_2^{YW}$	$\widehat{\alpha}_1^{YW}$	$\widehat{\alpha}_2^{YW}$
200	0.967	2.865	0.229	0.662
St. errors	(0.164)	(0.498)	(0.136)	(0.114)
500	0.980	2.943	0.236	0.686
St. errors	(0.113)	(0.350)	(0.092)	(0.068)
1000	0.984	3.003	0.241	0.700
St. errors	(0.082)	(0.238)	(0.066)	(0.052)
5000	0.997	3.002	0.249	0.700
St. errors	(0.038)	(0.117)	(0.033)	(0.023)
10000	0.990	2.998	0.250	0.700
St. errors	(0.025)	(0.087)	(0.024)	(0.015)

b) True values $\mu = (2, 3)$, $\alpha = (0.2, 0.3)$, $p_{vec} = (0.45, 0.55)$, $p_{mat} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$.

N_1	$\widehat{\mu}_1^{YW}$	$\widehat{\mu}_2^{YW}$	$\widehat{\alpha}_1^{YW}$	$\widehat{\alpha}_2^{YW}$
200	2.052	3.033	0.210	0.291
St. errors	(0.274)	(0.368)	(0.114)	(0.121)
500	2.026	2.992	0.197	0.294
St. errors	(0.166)	(0.256)	(0.080)	(0.073)
1000	2.024	2.992	0.202	0.298
St. errors	(0.108)	(0.209)	(0.057)	(0.056)
5000	2.000	3.002	0.198	0.298
St. errors	(0.055)	(0.082)	(0.027)	(0.023)
10000	2.000	3.002	0.198	0.298
St. errors	(0.041)	(0.056)	(0.017)	(0.017)

Thus, parameter estimates are derived for subsamples of sizes $200, 500, 1000, 5000$ and 10000. In each mentioned case, 100 simulated data series are used, and the corresponding standard errors are calculated. With the sample size increment, all estimates are convergent with the standard errors decreasing towards 0. It is also visible that the standard errors for μ_j are smaller for smaller values of μ_j. These results are presented in Table 1 and Table 2, corresponding to the cases of two and three random states, respectively. Transition probabilities are not estimated this time, since those are not parameters of the $\mathrm{RrDLINAR}_1(\mathcal{M}, \mathcal{A})$ model itself.

Table 2: The case of 3 states.

a) True values $\mu = (1, 2, 5)$, $\alpha = (0.1, 0.25, 0.7)$, $p_{vec} = (0.3, 0.4, 0.3)$, $p_{mat} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$.

N_1	$\widehat{\mu}_1^{YW}$	$\widehat{\mu}_2^{YW}$	$\widehat{\mu}_3^{YW}$	$\widehat{\alpha}_1^{YW}$	$\widehat{\alpha}_2^{YW}$	$\widehat{\alpha}_3^{YW}$
200	1.030	2.103	4.957	0.115	0.263	0.695
St. errors	(0.202)	(0.407)	(0.945)	(0.130)	(0.130)	(0.141)
500	1.007	2.038	5.051	0.107	0.259	0.705
St. errors	(0.133)	(0.250)	(0.696)	(0.085)	(0.092)	(0.104)
1000	0.993	2.037	5.018	0.104	0.259	0.705
St. errors	(0.090)	(0.174)	(0.521)	(0.062)	(0.061)	(0.078)
5000	1.000	2.010	5.009	0.103	0.250	0.701
St. errors	(0.035)	(0.073)	(0.216)	(0.030)	(0.030)	(0.034)
10000	1.000	2.004	4.995	0.103	0.250	0.700
St. errors	(0.025)	(0.053)	(0.164)	(0.020)	(0.020)	(0.026)

b) True values $\mu = (2, 3, 5)$, $\alpha = (0.1, 0.2, 0.4)$, $p_{vec} = (0.33, 0.34, 0.33)$, $p_{mat} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$.

N_1	$\widehat{\mu}_1^{YW}$	$\widehat{\mu}_2^{YW}$	$\widehat{\mu}_3^{YW}$	$\widehat{\alpha}_1^{YW}$	$\widehat{\alpha}_2^{YW}$	$\widehat{\alpha}_3^{YW}$
200	1.978	3.016	4.937	0.110	0.181	0.390
St. errors	(0.359)	(0.652)	(0.737)	(0.122)	(0.141)	(0.119)
500	2.016	3.016	4.946	0.103	0.207	0.390
St. errors	(0.216)	(0.366)	(0.471)	(0.083)	(0.121)	(0.083)
1000	1.991	2.994	4.954	0.103	0.201	0.393
St. errors	(0.157)	(0.248)	(0.312)	(0.060)	(0.080)	(0.054)
5000	1.991	3.006	4.989	0.099	0.199	0.396
St. errors	(0.078)	(0.120)	(0.146)	(0.028)	(0.036)	(0.027)
10000	1.991	3.006	4.991	0.099	0.201	0.399
St. errors	(0.048)	(0.087)	(0.112)	(0.018)	(0.023)	(0.018)

6. FORECASTING

Accuracy of forecasting in real-life data analysis is as important as evaluating the fit of the model. We introduce here a criterion suitable to compare the prediction results of different models. However, before we introduce the criterion itself, the forecasting procedure will be described in brief. At the beginning, the data sample of size $N = n_1 + n_2$ is divided into two sets, the training set and the prediction set. The training set contains the first n_1 sample elements and the prediction set contains the last n_2 elements of the sample.

We use training set to estimate model parameters and evaluate the fitting quality. In order to evaluate the forecasting accuracy, we generate m sequences of predictions from the estimated model parameters, each of size n_2. Using the proposed criterion, we compare generated predictions with the prediction set and finally determine the accuracy of forecasting.

We decided to use the forecasting log-score criterion (FLSC) already introduced in [14]. The criterion represents a modification of the log-score criterion (LSC) proposed by [10]. The FLSC criterion is given by the formula

$$\text{FLSC} = \sum_{h=1}^{n_2} \log \hat{p}_{n_1+h}(x_{n_1+h}),$$

where $h = 1, 2, ..., n_2$ and $\hat{p}_{n_1+h}(x_{n_1+h})$ represents the estimated probability of correctly predicting the value x_{n_1+h} from the prediction set, that is,

$$\hat{p}_{n_1+h}(x_{n_1+h}) = \frac{\text{number of correct predictions}}{m}.$$

Hihger values of the FLSC imply better forecasting.

7. APPLICATION

Regarding the application to the real-life data, we want to examine whether there is any progress compared to other INAR models that deal with both positive and negative values. For that purpose, we consider the number of motor vehicle thefts reported on a monthly basis to police stations number 1608 and 2811, in Pittsburgh, Pennsylvania, USA, between January 1990. and December 2001. The data were collected by the City of Pittsburgh Bureau of Police and reported under the FBI Uniform Crime Report. The differences between motor vehicle thefts on a monthly basis reported to these two police stations are calculated and provided in Table 3.

Table 3: Differences between motor vehicle thefts reported on a monthly basis to police stations number 1608 and 2811.

12	−1	2	3	8	−2	−3	4	4	6	5	5	5	4	4
5	4	5	4	0	1	0	1	2	3	−6	0	−1	−1	1
0	2	−1	0	1	−4	−5	−13	−4	−4	−5	−4	−6	−5	−8
−5	−5	−4	−4	−6	−5	0	1	−3	3	0	1	−2	0	0
−3	−1	−3	−3	−1	3	1	−1	0	0	−1	−1	2	1	1
1	3	0	2	1	0	0	2	1	1	−2	−2	−1	0	1
0	0	−3	0	1	−2	0	−2	2	−2	−3	2	2	2	3
2	1	−2	0	0	2	3	−3	0	−2	3	3	1	0	0
2	3	1	0	−3	−2	1	−3	−3	−3	2	3	−2	−2	1
3	1	2	0	3	2	3	2	−3						

Based on the sample size of $N = 144$, the sample mean of the differences between motor vehicle thefts in police stations number 1608 and 2811 is $\overline{y} = 0.048$, which proves the fact that the mean values of the number of motor vehicle thefts in both stations are approximately the same.

This condition is crucial here, since it had been claimed earlier that both processes $\{X_n^{(1)}(z_n)\}$ and $\{X_n^{(2)}(z_n)\}$ must have the same distribution.

The usual first step in standard INAR modeling is to obtain the plots of autocorrelation and partial autocorrelation functions. Those are given in Figure 1 and successfully justify the usage of the INAR(1) modeling. Bearing that in mind, we decide to compare our RrDLINAR$_1(\mathcal{M}, \mathcal{A})$ model on the training set with TINAR(1) (introduced by [9]), DLINAR(1) (proposed by [17]) and STINAR(1) (given in [4]). All mentioned models involve two i.i.d. latent AR components and are therefore suitable to compete with the RrDLINAR$_1(\mathcal{M}, \mathcal{A})$ model. As the criteria of the model validity, we take the following goodness-of-fit statistics: the root mean square error (RMSE), the mean absolute error (MAE) and the median absolute error (MdAE). Here, errors are defined as differences between observed values and corresponding predictions. In general, the model that provides better fitting to the data is expected to show lower values of these statistics.

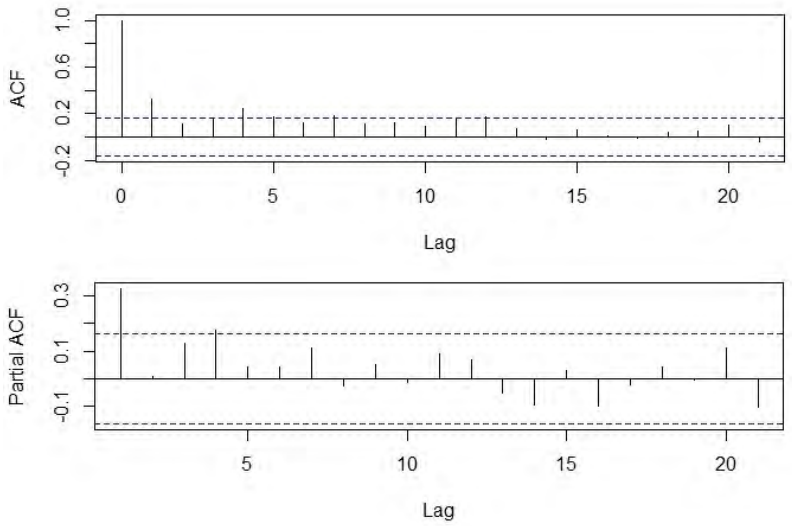


Figure 1: Autocorrelations and partial autocorrelations.

In the next step, the observed data values are clustered. This is how we actually obtain realized values of the corresponding random environment process. For all $n \in \{1, 2, ..., 144\}$, if the theft difference in the n-th month is in the i-th cluster (where $i \in E_r$ and $E_r = \{1, 2, ..., r\}$ is a set of possible random states), then $z_n = i$ is a n-th realized value of the corresponding r states random environment process. In that way, realizations z_n are determined. In this particular case, we decided to divide the theft difference realizations into two clusters, using the K-means clustering technique. These clustering results are given in Figure 2. Based on this figure, the decision to divide the theft difference realizations into two clusters is justified. As can be noticed, the differences that do not deviate too much from zero are located in the first cluster (triangles). The second cluster mainly consists of realizations from two time intervals (August 1990–July 1991 and December 1993–March 1995). These realizations deviate significantly from zero (circles), indicating that changes in environment state may have occurred. With each additional increase in the number of clusters, at least one cluster with very few realizations in it is created. This leads to frequent state changes, which is ruinous to any model in a random environment, including RrDLINAR$_1(\mathcal{M}, \mathcal{A})$.

Hence, $\mathrm{RrDLINAR}_1(\mathcal{M},\mathcal{A})$ models with more than 2 environment states should not be considered for this data series.

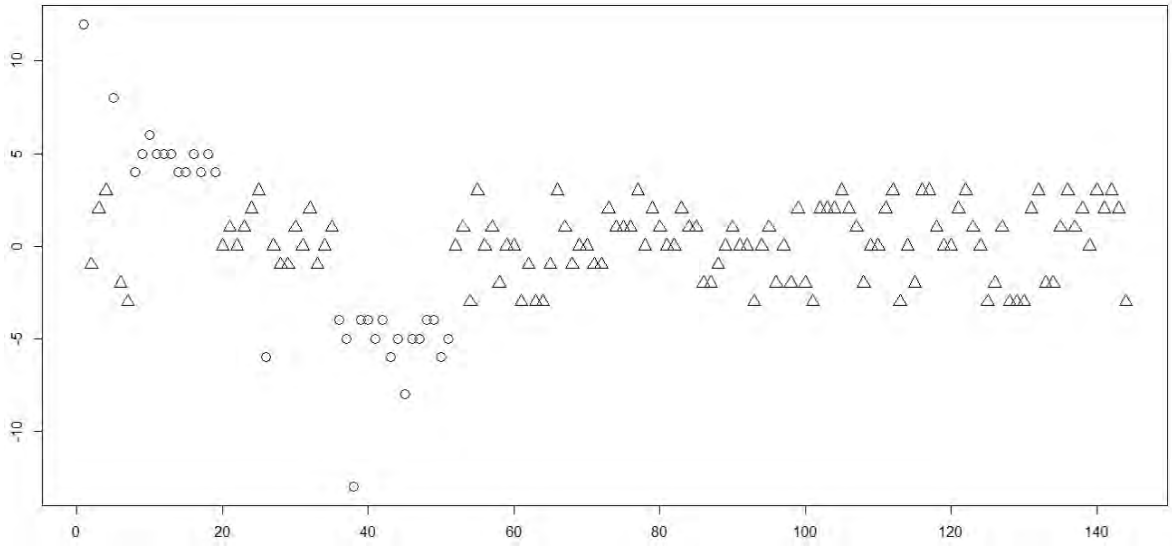


Figure 2: Clustering results obtained by applying the K-means technique to the motor vehicle theft differences.

The data set was divided into two parts, the training set with the first 120 sample elements and the prediction set with the last 24 sample elements. Based on the training set, we are able to construct the $\mathrm{R2DLINAR}_1(\mathcal{M},\mathcal{A})$ process $\{Y_n(z_n)\}$, of which arbitrary element $Y_n(z_n)$ has the $\mathrm{DL}\left(\frac{\widehat{\mu}_{z_n}}{1+\widehat{\mu}_{z_n}}\right)$ distribution, where $\widehat{\mu}_1,\widehat{\mu}_2$ and $\widehat{\alpha}_1,\widehat{\alpha}_2$ are previously obtained from the training set using corresponding YW estimates.

Furthermore, we perform on training set the fitting quality comparison of all applied models by calculating RMSE, MAE and MdAE for each particular model. These values, together with the Yule-Walker parameter estimates, are presented in Table 4. Obviously, the $\mathrm{R2DLINAR}_1(\mathcal{M},\mathcal{A})$ model based on two states random environment process showed the best performance when fitting this kind of real-life data sequences, providing the smallest values of all goodness-of-fit statistics. Regarding other competitors, STINAR(1) performed slightly better. However, even for this model, goodness-of-fit statistics are significantly higher then corresponding counterparts calculated for the $\mathrm{R2DLINAR}_1(\mathcal{M},\mathcal{A})$ model. One more fact is noticeable. Compared to the parameter α_2, parameter α_1 from $\mathrm{R2DLINAR}_1(\mathcal{M},\mathcal{A})$ model is more similar to the parameter α from the other three models. Bearing in mind the form of the one-step ahead conditional expectation given with (3.3), it can be concluded that all proposed models are suppose to give relatively similar approximations of data points from the first cluster. However, in the second cluster, the approximations should differ dramatically, which would explain the difference in RMSE, MAE and MdAE values.

In Figure 3, the realization of the theft difference process, as well as the predicted values of $\mathrm{R2DLINAR}_1(\mathcal{M},\mathcal{A})$ and STINAR(1) models, are shown. The TINAR(1) and DLINAR(1) model predictions are omitted here due to their similarity to STINAR(1) predictions.

Table 4: YW parameter estimates and goodness-of-fit statistics RMSE, MAE and MdAE for INAR(1) modeling of the theft difference.

Model	YW	RMSE	MAE	MdAE
TINAR(1)	$\widehat{\alpha} = 0.334$ $\widehat{\lambda} = 1.818$	2.691	2.139	2.010
DLINAR(1)	$\widehat{\alpha} = 0.332$ $\widehat{\mu} = 1.882$	2.690	2.135	2.001
STINAR(1)	$\widehat{\alpha} = 0.338$ $\widehat{\mu}_1 = 2.008$ $\widehat{\mu}_2 = 1.992$	2.681	2.131	1.993
R2DLINAR$_1(\mathcal{M}, \mathcal{A})$	$\widehat{\alpha}_1 = 0.190$ $\widehat{\alpha}_2 = 0.809$ $\widehat{\mu}_1 = 0.816$ $\widehat{\mu}_2 = 3.649$	2.188	1.665	1.001

As expected, both models (R2DLINAR$_1(\mathcal{M}, \mathcal{A})$ and STINAR(1)) approximate well the values which are not that far from zero. A large difference in quality fitting is noticeable in realizations that correspond to another state, i.e., that deviate significantly from zero. In this case, the newly defined R2DLINAR$_1(\mathcal{M}, \mathcal{A})$ model shows much better ability to adjust to the real-life realizations, which leads to better fitting. For high deviations, the difference between the predicted values of the R2DLINAR$_1(\mathcal{M}, \mathcal{A})$ and STINAR(1) models is larger. This adaptability is certainly a consequence of the non-stationary nature of the RrDLINAR$_1(\mathcal{M}, \mathcal{A})$. It is important to add that the trajectory of the R2DLINAR$_1(\mathcal{M}, \mathcal{A})$ model generally follows the trajectory of the realized process.

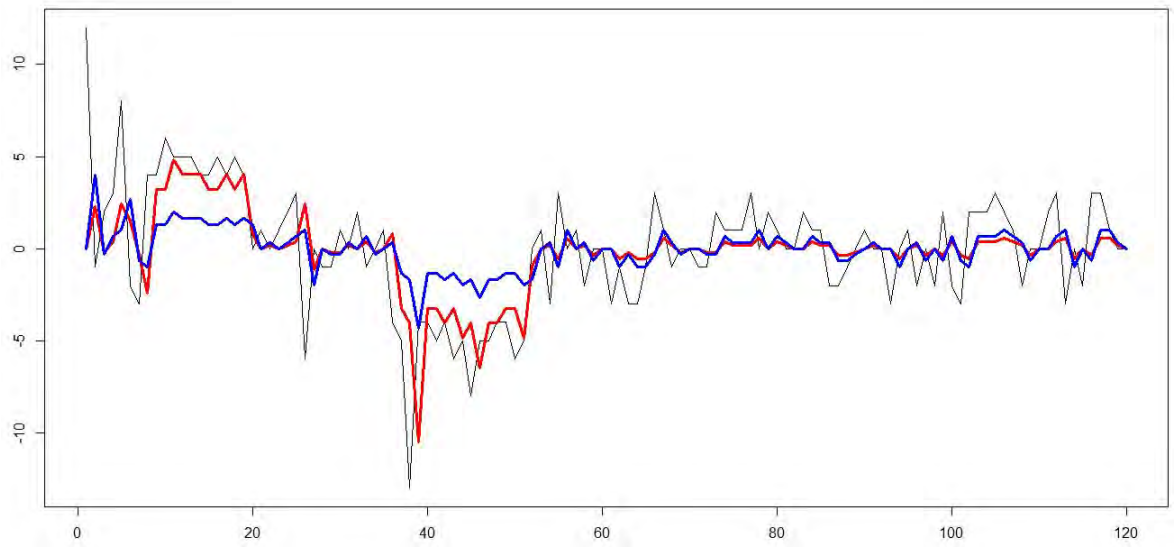


Figure 3: Black line – realization of the theft difference process;
red line – R2DLINAR$_1(\mathcal{M}, \mathcal{A})$ predicted values;
blue line – STINAR(1) predicted values.

Finally, we focus our attention on forecasting. To neutralize the impact of randomness on forecasting procedure, 10000 different sequences of size 24 are generated for each model, whereby corresponding model parameters are based on the training set. We compare these generated sequences of predictions with the prediction set and calculate FLSC. Table 5 shows results of the FLSC for all considered models. According to this table, the $\text{R2DLINAR}_1(\mathcal{M}, \mathcal{A})$ model has the largest FLSC among all considered models and thus, it provides the most accurate forecasting.

Table 5: FLSC for the prediction set of the theft difference.

	TINAR(1)	DLINAR(1)	STINAR(1)	$\text{R2DLINAR}_1(\mathcal{M}, \mathcal{A})$
FLSC	−65.712	−63.982	−64.151	−63.136

8. CONCLUSION

In this paper, we introduced a random environment integer-valued autoregressive process with discrete Laplace marginal distributions, $\text{RrDLINAR}_1(\mathcal{M}, \mathcal{A})$. Since the construction of this process was inspired by the work of [17], some of its features have been obtained in a similar way. Besides the definition of the $\text{RrDLINAR}_1(\mathcal{M}, \mathcal{A})$ process, we presented the full characterization of the process including its k-step ahead conditional expectation, correlation properties and the innovation process distribution. Parameter estimation was carried out using the method of moments and the strong consistency was proven. The YW estimates quality has been verified using subsamples of different sizes of 100 simulated data series, each of length 10000. At the very end, an application of the introduced model on real-life data series has been presented.

Further research might be performed in two directions. First of all, the model itself can be generalized to an order higher than 1, following the technique used in [11]. As for the second direction, an idea presented in this paper might be applied to the INAR process with asymmetric discrete Laplace marginal distribution, introduced by [7].

ACKNOWLEDGMENTS

We sincerely thank two reviewers for critically reading the manuscript and suggesting substantial improvements.

REFERENCES

[1] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis*, **8**, 261–275.

[2] Alzaid, A.A. and Al-Osh, M.A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions, *Annals of the Institute of Statistical Mathematics*, **45**, 223–232.

[3] Bakouch, H.S. and Ristić, M.M. (2010). Zero truncated Poisson integer-valued AR(1) model, *Metrika*, **72**(2), 265–280.

[4] Barreto-Souza, W. and Bourguignon, M. (2015). A skew INAR(1) process on Z, *AStA Advances in Statistical Analysis*, **99**(2), 189–208.

[5] Brockwell, P.J. and Davis, R.A. (1987). *Time Series – Theory and Methods*, Springer, New York.

[6] Cui, Y.; Li, Q. and Zhu, F. (2021). Modeling Z-valued time series based on new versions of the Skellam INGARCH model, *Brazilian Journal of Probability and Statistics*, **35**, 293–314.

[7] Djordjević, M.S. (2017). An extension on INAR models with discrete Laplace marginal distributions, *Communications in Statistics – Theory and Methods*, **46**(12), 5896–5913.

[8] Du, J.G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model, *Journal of Time Series Analysis*, **12**, 129–142.

[9] Freeland, R.K. (2010). True integer value time series, *AStA Advances in Statistical Analysis*, **94**, 217–229.

[10] Gelman, A.; Hwang, J. and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models, *Statistics and Computing*, **24**, 997–1016.

[11] Laketa, P.N.; Nastić, A.S. and Ristić, M.M. (2018). Generalized random environment INAR models of higher order, *Mediterranean Journal of Mathematics*, **15**, 9–30.

[12] Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process, *Journal of Time Series Analysis*, **19**, 439–455.

[13] Liu, Z.; Li, Q. and Zhu, F. (2020). Random environment binomial thinning integer-valued autoregressive process with Poisson or geometric marginal, *Brazilian Journal of Probability and Statistics*, **34**(2), 251–272.

[14] Liu, Z.; Li, Q. and Zhu, F. (2021). Semiparametric integer-valued autoregressive models on $\mathbb{Z}$, *Canadian Journal of Statistics*, **49**(4), 1317–1337.

[15] Liu, Z. and Zhu, F. (2021). A new extension of thinning-based integer-valued autoregressive models for count data, *Entropy*, **23**(1), 62.

[16] McKenzie, E. (1985). Some simple models for discrete variate time series, *Water Resources Bulletin*, **21**, 645–650.

[17] Nastić, A.S.; Ristić, M.M. and Djordjević, M.S. (2016). An INAR model with discrete Laplace marginal distribution, *Brazilian Journal of Probability and Statistics*, **30**(1), 107–126.

[18] Nastić, A.S.; Laketa, P.N. and Ristić, M.M. (2016). Random environment integer-valued autoregressive process, *Journal of Time Series Analysis*, **37**, 267–287.

[19] Nastić, A.S.; Laketa, P.N. and Ristić, M.M. (2019). Random environment INAR models of higher order, *RevStat – Statistical Journal*, **17**(1), 35–65.

[20] Qi, X.; Li, Q. and Zhu, F. (2019). Modeling time series of count with excess zeros and ones based on INAR(1) model with zero-and-one inflated Poisson innovations, *Journal of Computational and Applied Mathematics*, **346**, 572–590.

[21] Qian, L.; Li, Q. and Zhu, F. (2020). Modelling heavy-tailedness in count time series, *Applied Mathematical Modelling*, **82**, 766–784.

[22] Zheng, H.; Basawa, I.V. and Datta, S. (2006). Inference for pth-order random coefficient integer-valued autoregressive processes, *Journal of Time Series Analysis*, **27**, 411–440.

[23] Zheng, H.; Basawa, I.V. and Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes, *Journal of Statistical Planning and Inference*, **137**, 212–229.

REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 491–507
https://doi.org/10.57805/revstat.v21i4.431

Bias Reduction of Maximum Likelihood Estimates for an Asymmetric Class of Power Models with Applications

Authors: Yolanda M. Gómez
– Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile
ygomez@ubiobio.cl

Bruno Santos
– School of Mathematics Statistics and Actuarial Science, University of Kent, Canterbury, England
b.santos@kent.ac.uk

Diego I. Gallardo
– Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile
dgallardo@ubiobio.cl

Osvaldo Venegas ✉
– Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco, Chile
ovenegas@uct.cl

Héctor W. Gómez
– Departamento de Estadística y Ciencias de Datos, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile
hector.gomez@uantof.cl

Received: November 2020 Revised: February 2022 Accepted: February 2022

Abstract:

- In this paper we study some methods to reduce the bias for maximum likelihood estimation in the general class of alpha power models, specifically for the shape parameter. We find the modified maximum likelihood estimator using Firth's method and we show that this estimator is the uniformly minimum variance unbiased estimator (UMVUE) in this class. We consider three special cases of this class, namely the exponentiated exponential (EE), the power half-normal and the power piecewise exponential models. We compare the bias in simulation studies and find that the modified method is definitely superior, especially for small sample sizes, in both the bias and the root mean squared error. We illustrate our modified estimator in four real data set examples, in each of which the modified estimates better explain the variability.

Keywords:

- *UMVUE; Firth's method; exponentiated exponential model; power half-normal model.*

AMS Subject Classification:

- 62E20, 62F10.

✉ Corresponding author.

1. INTRODUCTION

The power (P) model for a probability density function (pdf), say f, was discussed by Lehmann [18] for the case $\alpha \in \mathbb{N}$ and later by Durrans [5] for $\alpha \in \mathbb{R}^+$, the shape parameter. If a random variable Z is distributed according to this class of distributions, we use the notation $Z \sim P_f(\alpha)$.

Its probability density function (pdf) is given by

$$g(z;\alpha) = \alpha f(z) F(z)^{\alpha-1}, \qquad z \in \mathcal{Z}, \tag{1.1}$$

where $\mathcal{Z}$ is the sample space defined for f and F is the cumulative distribution function (cdf) related to f. Durrans [5] considers $f = \phi(\cdot)$, the density function of the normal distribution with location and scale parameters, which we will refer to as power normal (PN) distribution. The case when f is symmetrical is discussed by Gupta and Gupta [10], including some fundamental properties of this family; Pewsey *et al.* [38] show that the information matrix is not singular when the symmetrical case is recovered ($\alpha = 1$).

Based on this representation, there are some extensions in the literature and we refer to Table 1 for a few references.

Table 1: Some extensions for the Power family.

Authors	**Distribution**	F
Gupta and Kundu [11, 12]	EE	Exponential
Mudholkar *et al.* [32, 33]	Exponentiated Weibull (WE)	Weibull
Gupta *et al.* [9]	Exponentiated-Pareto (EP)	Pareto
Nadarajah [35]	Exponentiated Gumbel (EG)	Gumbel
Kakde and Shirke [14]	Exponentiated lognormal (ELN)	Lognormal
Nadarajah and Gupta [36]	Exponentiated gamma (EG)	Gamma
Martínez-Flórez *et al.* [25]	Skew-normal alpha power (SNAP)	Skew-normal
Martínez-Flórez *et al.* [26]	Power Birnbaum-Saunders (PBS)	Birbaum-Saunders
Gómez and Bolfarine [7]	Power half-normal (PHN)	Half-normal
Zhao and Kim [42]	Power t (PT)	t-Student
Gómez *et al.* [8]	Power piecewise exponential (PPE)	Piecewise exponential

Another important property of the $P_f(\alpha)$ class is its interpretability for $\alpha \in \mathbb{N}$. In this case, the $P_f(\alpha)$ model can be interpreted as the distribution of the maximum of $X_1, X_2, ..., X_\alpha$, where the X_i's are independent, identically distributed random variables from $X \sim f$.

A similar interpretation is given by Durrans [5] for the extended case $\alpha \in \mathbb{R}^+$ using fractional order statistics.

Other extensions related to this model are presented in Martínez-Flórez *et al.* [23], introducing a multivariate version of the model; Martínez-Flórez *et al.* [24], performing applications in regression models; Martínez-Flórez *et al.* [27], studying the exponential transformation of the model; in [28], studying a version of the doubly censored model with inflation in a regression context. In one of the references in Table 1, Gupta and Kundu [12] reported a simulation study for the EE model in which an overestimation problem for the shape parameter in small sample sizes was observed. Based on this, we propose a bias correction methodology which should be useful not only for this distribution, but for the whole P family. To motivate our discussion, we start our presentation with two special cases from this class of models. Later, in the simulation studies, we add another case; and in the applications we also add another member of this family. This is done to ensure that our bias correction method works with different distributions, not just certain carefully selected distributions.

The first case is the EE model with shape and rate parameters α and λ respectively (which we denote as $\mathrm{EE}(\alpha, \lambda)$). The second case is the PHN model with shape and scale parameters α and σ (which we denote as $\mathrm{PHN}(\alpha, \sigma)$).

For $Z \sim \mathrm{EE}(\alpha, \lambda)$, the pdf is

$$g(z; \alpha, \lambda) = \alpha\lambda e^{-\lambda z}[1 - e^{-\lambda z}]^{\alpha-1}, \qquad z > 0, \tag{1.2}$$

and for $Z \sim \mathrm{PHN}(\alpha, \sigma)$, the pdf is given by

$$g(z; \alpha, \sigma) = \frac{2\alpha}{\sigma}\phi\Big(\frac{z}{\sigma}\Big)\Big[2\Phi\Big(\frac{z}{\sigma}\Big) - 1\Big]^{\alpha-1}, \qquad z > 0, \tag{1.3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and the cdf for the standard normal distribution.

When dealing with these cases, it is important to study the associated bias in parameter estimation, if we use maximum likelihood estimation methods for instance. Although the unbiased property of these estimators is well known in asymptotic condition, we need to be careful when using this estimation method for small sample sizes. For applications of recent mean bias reduction methodology in different contexts, see for instance Kosmidis *et al.* [15], Melo *et al.* [31], Maity *et al.* [22], Magalhães *et al.* [20] and Mazucheli *et al.* [30]. An alternative method is the median bias reduction methodology recently proposed in Pagui *et al.* [37] and applied in different contexts in Kyriakou *et al.* [16] and Ioannis *et al.* [13], for example.

For the $P_f(\alpha)$ class of distributions, a simple simulation study can be set up to identify some weaknesses of the maximum likelihood estimators (MLE's) for different values of each parameter and different sample sizes, considering the cases of the exponentiated exponential and the power-half normal models. In Figures 1 and 2 we report the estimated bias based on 10,000 replicates for the EE and PHN models for the MLEs.

Note that in both models, the observed average bias of the estimator of α is considerably greater (in relative terms) when compared with the average bias for the other parameters. This fact motivates the study of a method for reducing the mean bias for the MLE of α in the general class of model defined in (1.1), which can be applied to any member of the class.

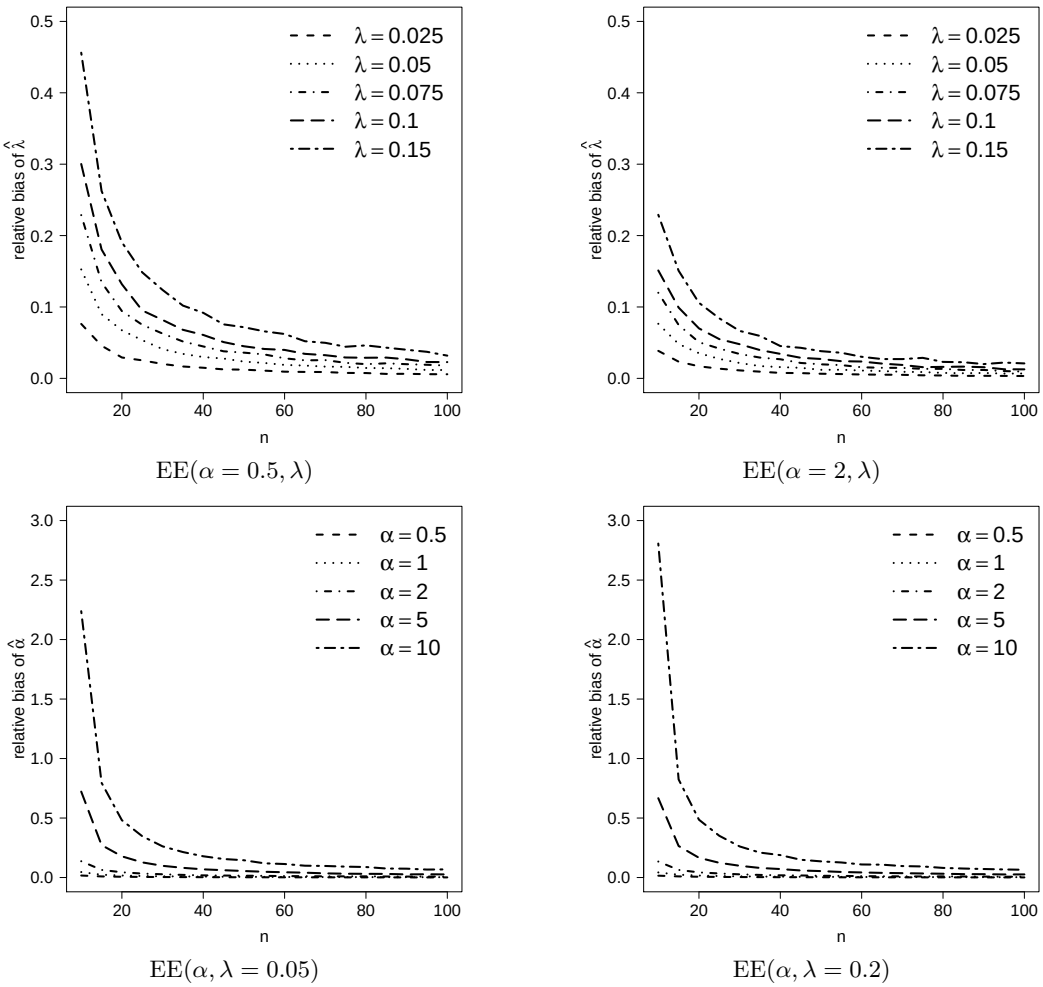


Figure 1: Estimated bias for the MLE of λ and α in the EE(α, λ) under different scenarios based on 10,000 replicates.

Initially, considering only the shape parameter α, it is possible not only to study the bias, but also to define an unbiased estimator, namely the UMVUE. This result is also related to the method proposed by Firth [6] for prevention of bias in maximum likelihood estimation. Moreover, when there are more parameters in this class of distributions, the approach used in Sartori [41] for skew normal models provides a convenient scheme to focus our attention on the bias for α, while also maximizing the likelihood for the other parameters; it has been applied more recently in Arrué *et al.* [1] and Magalhães *et al.* [21] in the study of bias for skew-normal, modified skew-normal and Marshall-Olkin models respectively.

This paper is organized as follows. Section 2 defines the bias for the shape parameter and presents the UMVUE for α in the general class of power models. Section 3 discusses the method to prevent bias to the shape parameter, while also obtaining maximum likelihood estimates for the other parameters, where we describe an iterative algorithm to find these estimates. Section 4 shows a simulation study, where we consider the cases not only of the exponentiated exponential and the power-half models, but also the power piecewise exponential, to illustrate the superior performance of the modified estimator. In Section 5, we highlight the improvements provided by the methods proposed here with three applications, which are known in the literature for this type of data.

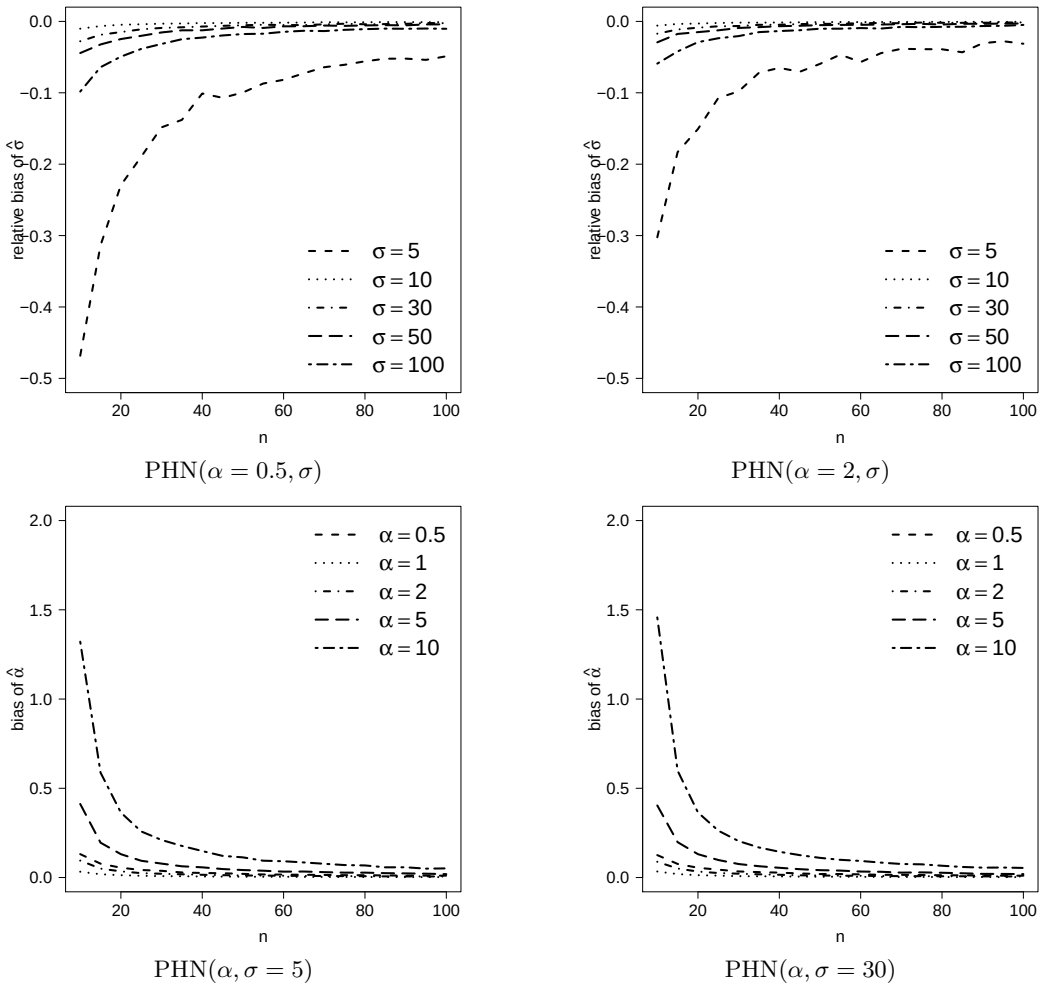


Figure 2: Estimated bias for the MLE of σ and α in the PHN(α, σ) under different scenarios based on 10,000 replicates.

2. CASE I: $F(\cdot)$ IS FREE OF PARAMETERS

The likelihood for a random sample $\boldsymbol{z} = (z_1, z_2, ..., z_n)$ from $P_f(\alpha)$ is given by

$$L(\alpha) = \prod_{i=1}^{n} f(z_i) \times \alpha^n \exp\left\{(\alpha - 1)\left(\sum_{i=1}^{n} \log F(z_i)\right)\right\}. \tag{2.1}$$

Theorem 2.1. *For the $P_f(\alpha)$ model, $T(\boldsymbol{z}) = -\sum_{i=1}^{n} \log F(z_i)$ is a complete statistic.*

Proof: Note that the likelihood function in (2.1) can be broken down as

$$L(\alpha) = \underbrace{\prod_{i=1}^{n} f(z_i)}_{h(\boldsymbol{z})} \times \underbrace{\alpha^n \exp\left\{-(\alpha - 1)\left(-\sum_{i=1}^{n} \log F(z_i)\right)\right\}}_{g_\alpha\left(-\sum_{i=1}^{n} \log F(z_i)\right)}.$$

According to the Neyman factorization theorem, $T = T(\boldsymbol{z}) = -\sum_{i=1}^{n} \log F(z_i)$ is a sufficient statistic for α. It is possible to verify that $-\log F(Z_i) \sim E(\alpha)$, where $E(\alpha)$ denotes the exponential distribution with rate α. As the Z_i's are independent, then $T \sim G(n, \alpha)$, with $G(a, b)$ the gamma distribution with shape and rate parameters a and b respectively. Let g be a measurable function in the interval $(0, \infty)$. Therefore,

$$\mathbb{E}(g(T)) = 0 \Leftrightarrow \int_0^{\infty} g(t) t^{n-1} e^{-\alpha t} dt = 0.$$

As $\alpha > 0$ and $n \geq 1$, then $t^{n-1}e^{-\alpha t} > 0$, $\forall t > 0$. Thus, $\mathbb{E}(g(T)) = 0 \Leftrightarrow g(T) = 0$, implying that T is a complete statistic. □

On the other hand, the log-likelihood function is given by

$$\ell(\alpha) = n \log \alpha + \sum_{i=1}^{n} [\log f(z_i) + (\alpha - 1) \log F(z_i)].$$

It is direct that the MLE of α is given by

$$\widehat{\alpha} = \frac{n}{-\sum_{i=1}^{n} \log F(z_i)}.$$

Theorem 2.2. *$\widehat{\alpha}$ is a biased estimator for α.*

Proof: As $-\sum_{i=1}^{n} \log F(Z_i) \sim G(n, \alpha)$, we have that

$$\mathbb{E}(\widehat{\alpha}) = \frac{n\alpha}{n-1}, \qquad n > 1.$$ □

Remark 2.1. Note that $\text{bias}(\widehat{\alpha}) = \alpha/(n-1)$, so that the bias can be "too large" when α is increased and the sample size is small. Clearly, for $n \to \infty$, $\widehat{\alpha}$ is unbiased.

Theorem 2.3. *$\widehat{\alpha}_M = (n-1)/(-\sum_{i=1}^{n} \log F(z_i))$ is the UMVUE for α.*

Proof: It is clear that $\widehat{\alpha}_M$ is an unbiased estimator for α. As $\widehat{\alpha}_M$ depends on a complete statistic, by the Lehmann-Scheffé theorem $\widehat{\alpha}_M$ is the UMVUE for α. □

2.1. Connection with the Firth method

A popular method to reduce the bias of an estimator is the Firth method [6]. For the univariate case, the method consists in modifying the score function, say $S(\alpha)$, by

$$S_M(\alpha) = S(\alpha) + M(\alpha),$$

where $M(\alpha) = \frac{1}{2}I(\alpha)^{-1}(\nu_{\alpha,\alpha,\alpha} + \nu_{\alpha,\alpha\alpha})$, $I(\alpha)$ the information matrix for the model, $\nu_{\alpha,\alpha,\alpha} = \mathbb{E}\left[\left(\frac{\partial\ell(\alpha)}{\partial\alpha}\right)^3\right]$ and $\nu_{\alpha,\alpha\alpha} = \mathbb{E}\left[\frac{\partial\ell(\alpha)}{\partial\alpha}\frac{\partial^2\ell(\alpha)}{\partial\alpha^2}\right]$. The solution of the modified score equation $S_M(\alpha) = 0$ produces the modified MLE, say $\widehat{\alpha}_M$. Firth [6] shows that the bias of $\widehat{\alpha}_M$ is reduced from $O(n^{-1})$ to $O(n^{-2})$ when compared with the ordinary MLE. Moreover, the asymptotic distribution of $\widehat{\alpha}_M$ coincides with that of $\widehat{\alpha}$, i.e.

$$\sqrt{n}(\widehat{\alpha}_M - \alpha) \to N\big(0, I(\alpha)^{-1}\big), \quad \text{as } n \to \infty.$$

Note that for the $P_f(\alpha)$ model

$$\frac{\partial\ell(\alpha)}{\partial\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\log F(z_i) \qquad \text{and} \qquad \frac{\partial^2\ell(\alpha)}{\partial\alpha^2} = -\frac{n^2}{\alpha^2}.$$

As $-\sum_{i=1}^{n}\log F(Z_i) \sim G(n,\alpha)$, it can be verified that

$$I(\alpha) = \frac{n^2}{\alpha^2}, \quad \nu_{\alpha,\alpha,\alpha} = -\frac{2n}{\alpha^3} \quad \text{and} \quad \nu_{\alpha,\alpha\alpha} = 0.$$

Then, for the $P_f(\alpha)$ model we have that $M(\alpha) = -\alpha^{-1}$. Therefore, $S_M(\alpha) = \frac{n-1}{\alpha} + \sum_{i=1}^{n}\log F(z_i)$. Solving $S_M(\alpha) = 0$, we obtain newly

$$\widehat{\alpha}_M = (n-1)/\left(-\sum_{i=1}^{n}\log F(z_i)\right),$$

as the solution.

3. CASE II: $F(\cdot)$ DEPENDS OF $\boldsymbol{\psi}$, A VECTOR OF PARAMETERS

Consider that $F(\cdot)$ is indexed by a vector of parameters $\boldsymbol{\psi}$. In this case, the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\psi}, \alpha)$ is given by

$$\ell(\boldsymbol{\theta}) = n\log\alpha + \sum_{i=1}^{n}\Big[\log f(z_i;\boldsymbol{\psi}) + (\alpha-1)\log F(z_i;\boldsymbol{\psi})\Big].$$

Our proposal is to consider the bias correction only for α and not for $\boldsymbol{\psi}$. This is justifiable in some models such as EE and PHN because the bias for α is considerable in small and median sample sizes and lower for the components of $\boldsymbol{\psi}$ as presented in Figures 1 and 2.
Following the approach used in Sartori [41], we first compute the constrained MLE $\widehat{\boldsymbol{\psi}}(\alpha)$ for fixed α, and then we apply Firth's method to the profile score function of α, which produces the modified estimator

$$\widehat{\alpha}_M = \frac{n-1}{-\sum_{i=1}^{n}\log F(x_i;\widehat{\boldsymbol{\psi}})}.$$

In short, the estimation procedure can be described as:

- Step 0: Choose an initial value for $\boldsymbol{\theta} = (\boldsymbol{\psi}, \alpha)$, say $\widehat{\boldsymbol{\theta}}^{(0)}$. A possible value can be $\widehat{\boldsymbol{\theta}}^{(0)} = (\widehat{\boldsymbol{\psi}}^{(0)}, 1)$, where $\widehat{\boldsymbol{\psi}}^{(0)}$ is the MLE for $\boldsymbol{\psi}$ considering that $X_1, ..., X_n$ are iid from $F(\cdot;\boldsymbol{\psi})$.

- Step 1: For $k = 1, 2, ...$, choose $\widehat{\boldsymbol{\psi}}^{(k)}$ as the vector that maximizes

$$\ell_p(\boldsymbol{\psi}; \widehat{\alpha}^{(k-1)})$$

in relation to $\boldsymbol{\psi}$.

- Step 2: For $k = 1, 2, ...$, do

$$\widehat{\alpha}_M^{(k)} = \frac{n-1}{-\sum_{i=1}^{n} \log F\left(x_i; \widehat{\boldsymbol{\psi}}^{(k)}\right)}.$$

Although we apply the bias correction only for α, we will see in the next section that this procedure also provides better estimates for $\boldsymbol{\psi}$.

Remark 3.1. Given the MLE of $\boldsymbol{\psi}$, say, $\widehat{\boldsymbol{\psi}}$ and considering Remark 2.1, we can compute the corrective method of Cox–Snell for α. This estimator will be denoted as $\widehat{\alpha}_C$ and is given by

$$\widehat{\alpha}_C = \frac{n-1}{-\sum_{i=1}^{n} \log F(z_i; \widehat{\boldsymbol{\psi}})}.$$

Note that in this procedure $\widehat{\boldsymbol{\psi}}$ is not recomputed and matched directly with the MLE estimator. However, to avoid confusion in the simulation study, we consider the notation $\widehat{\boldsymbol{\theta}}_C = (\widehat{\boldsymbol{\psi}}_C, \widehat{\alpha}_C)$ to refer to the estimators obtained by this method.

4. SIMULATION STUDY

In this section, we illustrate the method discussed in Section 3 for the EE, PHN and PPE models (see Gómez *et al.* [8]). All the computational programs were developed in R Core Team [39] and are available upon request. Random samples for those distributions can be obtained using the inverse transformation method, considering that the inverse of the cdf for the basal models are implemented in R. We consider sample sizes ranging from 10 to 100, taking one sample for every 5 units. For the EE model, we consider all combinations among the sets $\mathcal{A} = \{0.25, 0.5, 1, 2, 5, 10\}$ and $\mathcal{L} = \{0.1, 0.5, 2\}$ for α and λ, respectively. In a similar manner, for the PHN model we consider all combinations among the sets $\mathcal{A} = \{0.5, 2, 5, 10\}$ and $\mathcal{S} = \{5, 30, 50\}$ for α and σ, respectively. For the PPE models we choose a different way to select the parameters. We consider the case $L = 2$, which includes three parameters for the model. For a given time partition a and α, we take $\lambda_1 = -(1/a)\log(1 - 0.6^{1/\alpha})$, which guarantees that each observation belongs to the intervals $(0, a)$ and (a, ∞) with probabilities 0.6 and 0.4, respectively. We consider α in $\mathcal{A} = \{0.5, 2, 5, 10\}$, a in $\{6, 10\}$ and $\lambda_2 = 1$ for all combinations.

We consider 10,000 replicates for each combination between n, the sample size, α and λ, σ or (λ_1, λ_2) (depending on the model). In each replication we compute the ordinary MLEs, the estimators considering the Cox–Snell corrective method, and the proposed modified MLEs. For each scenario, we present the relative bias and the relative root mean squared error $\sqrt{MSE}$. In Figure 3 we can find the bias, and in Figure 4 we see the $\sqrt{MSE}$ for one

case in the EE distribution. The remaining combinations for the EE, PHN and PPE models are presented as supplementary material.

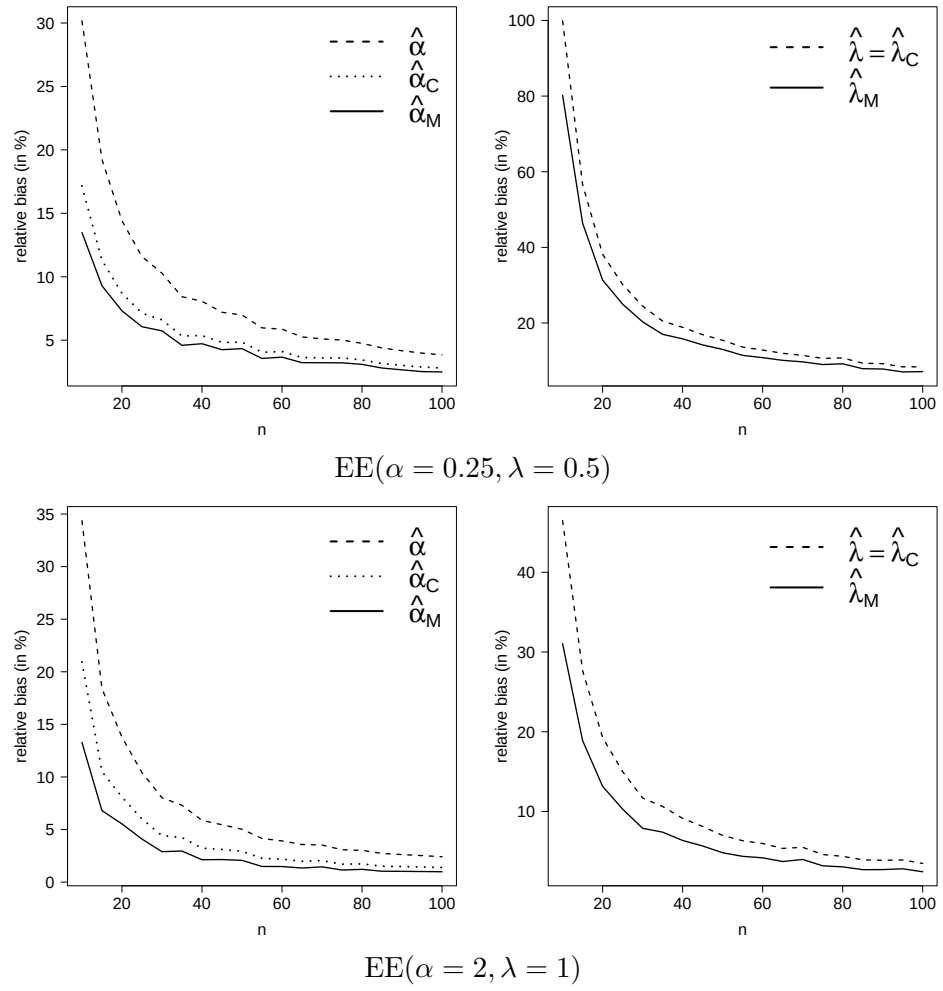
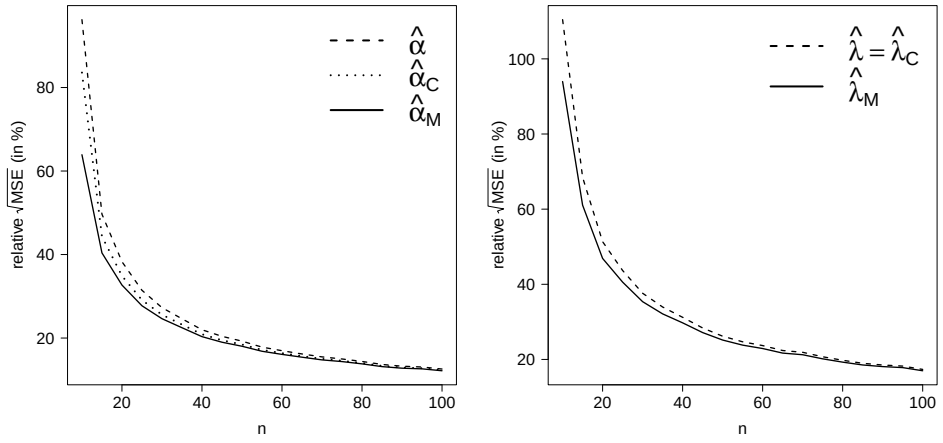


Figure 3: Estimated bias for the MLE and the modified MLE in the EE(α, λ) model under different scenarios based on 10,000 replicates.

Note that the bias of $\widehat{\alpha}_M$ is reduced considerably when compared with $\widehat{\alpha}$ and $\widehat{\alpha}_C$ in the three models, EE, PHN and PPE, especially for small sample sizes (say $n \leq 20$). Specifically, for the EE and PHN distributions in all the cases considered for $n = 10$ (the smallest sample size), the bias reduction is at least 10% when $\widehat{\alpha}$ is compared with $\widehat{\alpha}_M$ and at least 5% when is compared $\widehat{\alpha}_C$ with $\widehat{\alpha}_M$. For the PPE distribution in all the cases considered for $n = 10$, the bias reduction is at least 40% when $\widehat{\alpha}$ is compared with $\widehat{\alpha}_M$ and at least 30% when $\widehat{\alpha}_C$ is compared with $\widehat{\alpha}_M$. In all the models, the difference is even greater when the true value of α is increased. On the other hand and as expected, this difference is reduced when the sample size is increased.



$\mathrm{EE}(\alpha = 2, \lambda = 0.05)$

Figure 4: Estimated root MSE for the MLE and the modified MLE in the $\mathrm{EE}(\alpha = 0.5, \lambda = 0.05)$ model under different scenarios based on 10,000 replicates.

The components of the vector $\boldsymbol{\psi}$ also are benefited in terms of bias, although the bias reduction is only proposed for α. For the EE model in all the cases considered and $n = 10$, the bias reduction is at least 10% when $\widehat{\lambda}$ or $\widehat{\lambda}_C$ is compared with $\widehat{\lambda}_M$. For the PHN and PPE models, in all the cases considered the bias reduction for λ and (λ_1, λ_2) exists, but is marginal.

Additionally, $\sqrt{MSE}$ related to $\widehat{\alpha}_M$ is also lower when compared with $\widehat{\alpha}$ and $\widehat{\alpha}_C$. For the EE and PHN distributions, in all the cases considered for $n = 10$ the $\sqrt{MSE}$ is reduced by at least around 5% when $\widehat{\alpha}$ or $\widehat{\alpha}_C$ is compared with $\widehat{\alpha}_M$, whereas for the PPE model in all the cases considered for $n = 10$ this reduction is at least around 20%. On the other hand, for the EE and PPE models in all the cases considered for $n = 10$, the $\sqrt{MSE}$ is reduced by at least around 10% when the modified estimator is compared with the traditional estimator or the Cox–Snell estimator; whereas for the PHN distribution, in all the considered models the reduction for $\sqrt{MSE}$ is marginal. Again, in all the models, the difference is even greater when the true value for α is increased and as expected, the difference between the different estimators is reduced when the sample size is increased.

These simulation results are encouraging since they show, for these three particular members of the class of power models, that even though we focus this bias prevention method on the shape parameter, α, we still observe better bias results for the other parameters.

5. APPLICATIONS

In this section, we illustrate the methods in three real data sets for the EE, PPE and PN models. All data sets are already known in the literature and we wish to compare the performance of the modified MLE against the ordinary MLE and also using the Cox–Snell correction method. We examine not only how well they both fit the data, but also the bias, which could be estimated through bootstrap. Additional applications for the PHN and PBS models are presented as supplementary material.

5.1. Illustration 1

In this first application, we consider data on the number of million revolutions before failure for each of 23 ball bearings in a life test. More details about the data are presented in Lawless [17]. This data set was analyzed in Gupta and Kundu [12] using the EE model. The estimates considering the ordinary MLE's and the modified MLEs for this model are presented in Table 2, with respective s.e. and confidence intervals.

Table 2: MLE and modified MLE for the EE model in the ball bearings data set.

Parameter	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\lambda}_M$	$\widehat{\alpha}_M$
Estimate	0.0323	5.2832	0.0302	4.5379
s.e.	0.0064	2.0492	0.0060	1.6566
95% C.I.	(0.0197 ; 0.0449)	(1.2667 ; 9.2998)	(0.0206 ; 0.0440)	(2.0363 ; 8.5302)

Note that the confidence intervals when we consider the modified MLEs are more accurate for both parameters, since they have a smaller length compared to the confidence intervals obtained with the estimates of the ordinary MLEs. The histogram of the data and the estimated density functions of both estimates are presented in Figure 5a.

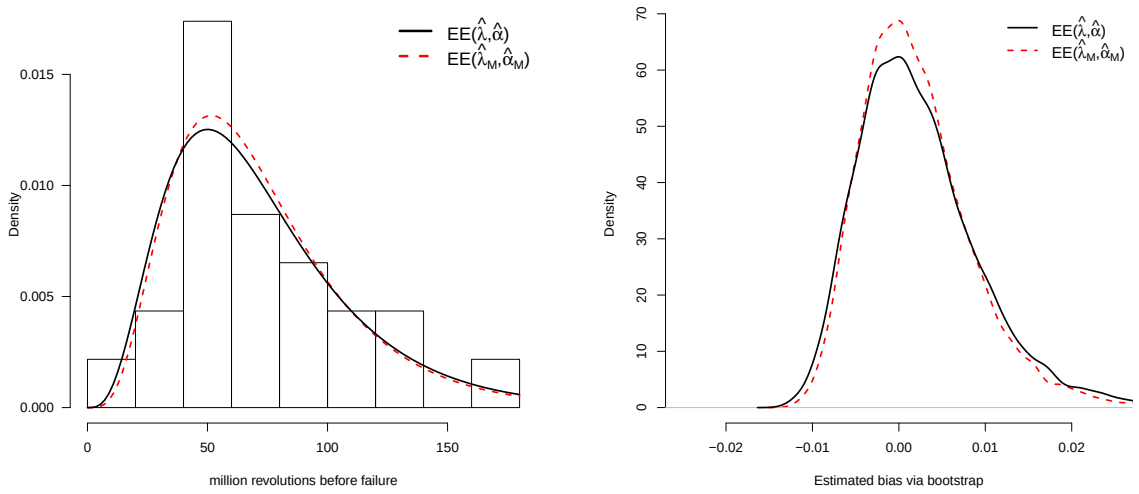


Figure 5: Comparisons for the ordinary and modified MLEs in the ball bearings data set. Left panel: Histogram for EE model and respective density estimate for each proposal. Right panel: Estimated bias via 10,000 bootstrap samples for each estimator.

The estimated density for the modified MLEs presents a better fit with the data, if we note that this density better represents the peak of values around 50 million revolutions before failure. Furthermore, if we take 10,000 bootstrap samples to estimate the bias, we have Figure 5b, where we are able to compare these two estimators empirically. The ordinary MLE

present an estimated bias, via bootstrap, equal to 1.4774, while this value for the modified bias is 1.0626. The bootstrap standard errors are equal to 3.8744 and 2.7440 for the ordinary and modified MLE, respectively. These results are confirmed by the analysis of Figure 5b, where the estimated density for the modified MLE is clearly more concentrated around zero, further evidence of the superior performance of the modified method.

5.2. Illustration 2

This data set is available in Murthy *et al.* [34] (data set 6.1 in Section 6.6.5). The data set represents the failure time of 20 components. We propose to analyze this data set based on the PPE model with $L = 2$ in order to illustrate the advantage of our methodology to reduce the bias for parameters. Table 3 shows the ordinary, corrected and modified MLEs for the PPE distribution in this data set. The main differences between the three methods are given in the estimates for λ_1 and α. We also highlight that the standard errors are lower for the modified MLEs, which also provides more accurate confidence intervals.

Table 3: Ordinary, corrected and modified MLE for the PPE model in failure time data set.

Parameter	Estimate	s.e.	95% C.I.
$\widehat{\lambda}_1$	0.8766	0.2643	(0.4855 ; 1.5829)
$\widehat{\lambda}_2$	3.8663	1.1347	(2.1751 ; 6.8725)
$\widehat{\alpha}$	5.1751	2.6060	(1.9287 ; 13.8856)
$\widehat{\lambda}_{1C}$	0.8766	0.2477	(0.4967 ; 1.5472)
$\widehat{\lambda}_{2C}$	3.8663	1.1394	(2.1759 ; 6.8699)
$\widehat{\alpha}_C$	4.9163	2.2601	(1.8781 ; 12.8693)
$\widehat{\lambda}_{1M}$	0.7600	0.2366	(0.3893 ; 1.4836)
$\widehat{\lambda}_{2M}$	3.8379	1.1303	(2.1542 ; 6.8376)
$\widehat{\alpha}_M$	4.0096	1.8634	(1.4397 ; 11.1671)

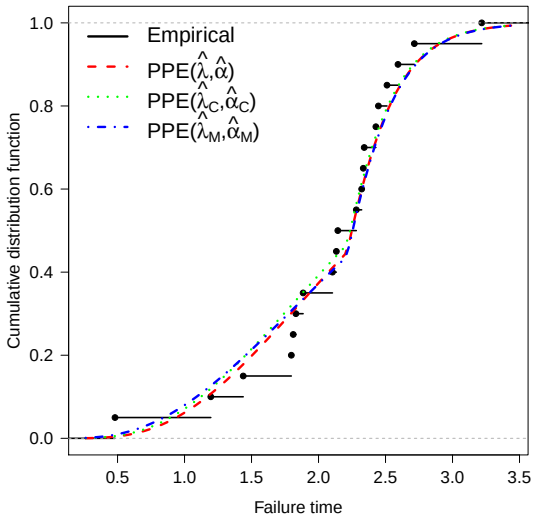


Figure 6: Cumulative distribution function for failure time data set using PPE model for ordinary MLE, corrected MLE and modified MLE.

Figure 6 shows the empirical cdf and the estimated cdf for the three methods. The main difference between the curves is given before the median of the distribution (approximately 2.2 units). Finally, Figure 7 shows the estimated distribution for the bias of the estimators of λ_1, λ_2 and α for the three methods, which are computed based on 10,000 bootstrap samples. Again, the main differences are given for the estimators for λ_1 and α. For this last term, the estimators provided by the ordinary and corrected MLEs have an evident and considerable bias, in contrast to the modified MLE where the bias is negligible.

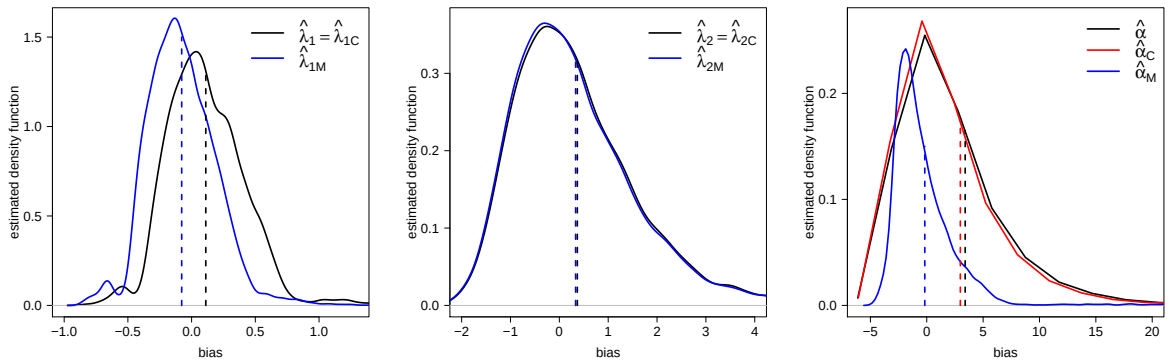


Figure 7: Estimated distribution for the bias for MLEs, corrected MLEs and modified MLEs based on 10,000 bootstrap samples. Dashed lines represents the respective average bias.

5.3. Illustration 3

This data set is related to 3,848 observations of the variable "density" in the data available at `http://lib.stat.cmu.edu/datasets/pollen.data` and was analyzed by Pewsey *et al.* [38] using the PN distribution. Although the PN distribution was not considered in the simulation studies, we decided to include this example using the PN model in order to demonstrate the effectiveness of our modified MLE for the parameters in a class of location-scale (μ and σ, respectively) within the power models. Evidently, as the sample size is large, it is to be expected that the ordinary MLE and their corrections will be closer. However, in order to illustrate our proposal, we considered a subsample of $n = 30$ from the original data. Table 4 presents the results. Note that the estimates for α are closer for the ordinary MLE and the Cox–Snell corrective method, but differ strongly from the modified MLE. The impact of this can be assessed by the huge reduction in the bias of $\widehat{\alpha}_M$ in comparison with the bias of $\widehat{\alpha}$ and $\widehat{\alpha}_C$. In addition, Figure 8 shows the histogram with the estimated density function based on the PN model for the three estimation methods.

Table 4: Ordinary, corrected, modified MLE and estimated bias based on 1,000 non-parametric bootstrap samples for the PN model in the pollen data set.

Parameter	Estimate	s.e.	95% C.I.	bias
$\widehat{\mu}$	−16.3238	36.7854	(−53.1092 ; 20.4616)	10.6265
$\widehat{\sigma}$	6.8207	6.5279	(0.2928 ; 13.3486)	−2.8944
$\widehat{\alpha}$	115.1611	970.8617	(0.0000 ; 1086.0228)	265.4226
$\widehat{\mu}_C$	−16.3238	5.3937	(−21.7175 ; −10.9301)	10.6265
$\widehat{\sigma}_C$	6.8207	1.3255	(5.4952 ; 8.1462)	−2.8944
$\widehat{\alpha}_C$	111.3105	117.7705	(0.0000 ; 229.0813)	256.5461
$\widehat{\mu}_M$	4.7456	1.6875	(3.0581 ; 6.4331)	1.1611
$\widehat{\sigma}_M$	1.7058	0.5620	(1.1438 ; 2.2678)	−0.4343
$\widehat{\alpha}_M$	0.1838	0.1674	(0.0163 ; 0.3512)	−0.0601

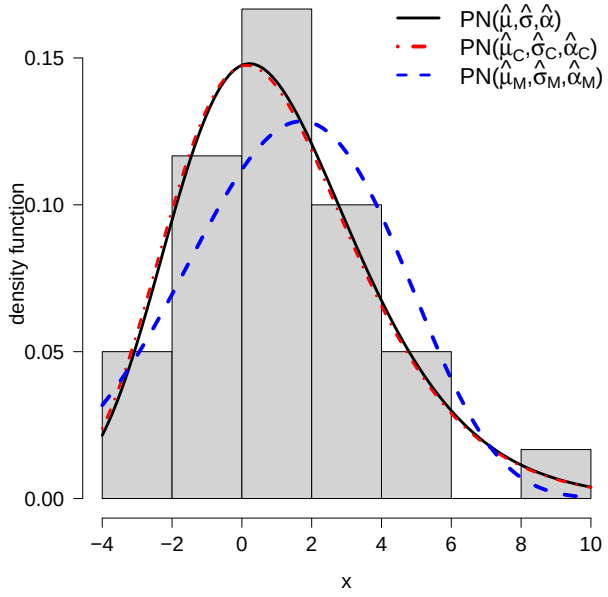


Figure 8: Estimated pdf for pollen data set using PN model for ordinary MLE, corrected MLE and modified MLE.

6. CONCLUSIONS

In this paper, we considered the problem of estimation of the shape parameter in the general class of power models; we recognize empirically the difficulties in this task, especially for small sample sizes. For the case where we have only α to estimate, we present the unbiased estimator as a function of a complete statistic for this class of models, obtaining the UMVUE for α. We discuss the connection of these results with the bias prevention method proposed by Firth [6]. We also propose an estimation method for the case when there are more parameters, limiting the bias correction to α.

Although our results are valid for all members of the general class of power models, we selected some members of this class of models, namely EE, PN, PPE, PHN and PBS, in order to demonstrate and compare the results between the ordinary MLE and the modified MLE proposed in this paper. The simulation studies confirm the bias reduction for the shape parameter, but they also show that there is an improvement related to bias for the other parameters involved, for each distribution considered here. According to our simulation results, the improvements are not only related to bias, as we also noticed lower root mean squared errors when we use the modified estimator. Although we do not consider families of bimodal distributions belonging to this family (as presented in Bolfarine *et al.* [4]), we see no reason why the method should not work in families of this type.

We illustrate our findings with three known data sets from the literature, for each distribution. Although this is a large number of examples, we thought it was important to make sure our method was tested with different members of this class, and not just some selected cases. We show that our modified estimator gives a better fit with the data in each case, and also we estimate the bias via bootstrap, validating that our proposal performs better.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous referees for their comments and/or suggestions, which undoubtedly improved the presentation of our article.

The research of O. Venegas was supported by Vicerrectoría de Investigación y Postgrado de la Universidad Católica de Temuco, Projecto interno FEQUIP 2019-INRN-03. Héctor W. Gómez's research has been partially supported by PUENTE UA.

REFERENCES

[1] Arrué, J.; Arellano-Valle, R.B. and Gómez, H.W. (2016). Bias reduction of maximum likelihood estimates for a modified skew normal distribution, *Journal of Statistical Computation and Simulation*, **86**, 2967–2984.

[2] Birnbaum, Z.W. and Saunders, S.C. (1969a). A new family of fife distributions, *Journal of Applied Probability*, **6**, 319–327.

[3] Birnbaum, Z.W. and Saunders, S.C. (1969b). Estimation for a family of life distributions with applications to fatigue, *Journal of Applied Probability*, **6**, 328–347.

[4] Bolfarine, H.; Martínez-Flórez, G. and Salinas, H.S. (2018). Bimodal symmetric-asymmetric power-normal families, *Communications in Statistics – Theory and Methods*, **47**, 259–276.

[5] Durrans, S.R. (1992). Distributions of fractional order statistics in hydrology, *Water Resources Research*, **28**, 1649–1655.

[6] Firth, D. (1993). Bias reduction of maximum likelihood estimates, *Biometrika*, **80**, 27–38.

[7] Gómez, Y.M. and Bolfarine, H. (2015). Likelihood-based inference for the power half-normal distribution, *Journal of Statistical Theory and Applications*, **14**, 383–398.

[8] Gómez, Y.M.; Gallardo, D.I. and Arnold, B.C. (2017). The power piecewise exponential model, *Journal of Statistical Computation and Simulation*, **88**, 825–840.

[9] Gupta, R.C.; Gupta, P.L. and Gupta, R.D. (1998). Modeling failure time data by Lehmann alternatives, *Communications in Statistics – Theory and Methods*, **27**, 887–904.

[10] Gupta, R.D. and Gupta, R.C. (2008). Analyzing skewed data by power normal model, *Test*, **17**, 197–210.

[11] Gupta, R.D. and Kundu, D. (1999). Generalized exponential distributions, *Australian & New Zealand Journal of Statistics*, **41**, 173–188.

[12] Gupta, R.D. and Kundu, D. (2001). Exponentiated exponential family: An alternative to Gamma and Weibull distributions, *Biometrical Journal*, **43**, 117–130.

[13] Ioannis, K.; Pagui, E.C.K. and Sartori, N. (2020). Mean and median bias reduction in generalized linear models, *Statistics and Computing*, **30**, 43–59.

[14] Kakde, C.S. and Shirle, D.T. (2006). On exponentiated lognormal distribution, *International Journal of Agricultural and Statistical Sciencies*, **2**, 319–326.

[15] Kosmidis, I. and Firth, D. (2009). Bias reduction in exponential family nonlinear models, *Biometrika,* **96**, 793–804.

[16] Kyriakou, S.; Ioannis, K. and Sartori, N. (2019). Median bias reduction in random-effects meta-analysis and meta-regression, *Statistical Methods in Medical Research*, **28**, 1622–1636.

[17] Lawless, E.L. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.

[18] Lehmann, E.L. (1953). The power of rank tests, *Annals of Mathematical Statistics*, **24**, 23–43.

[19] Lemonte, A. (2007) Improved statistical inference for the two-parameter Birnbaum-Saunders distribution, *Computational Statistics & Data Analysis*, **51**, 4656–4681.

[20] Magalhães, T.M.; Gallardo, D.I. and Bourguignon, M. (2021). Improved point estimation for inverse gamma regression models, *Journal of Statistical Computation and Simulation*, **91**, 2444–2456.

[21] Magalhães, T.M.; Gómez, Y.M.; Gallardo, D.I. and Venegas, O. (2020). Bias reduction for the Marshall-Olkin extended amily of distributions with application to an airplane's air aonditioning aystem and arecipitation data, *Symmetry*, **12**(5), 851.

[22] Maity, A.K.; Pradhan, V. and Das, U. (2019). Bias reduction in logistic regression with missing responses when the missing data mechanism is nonignorable, *The American Statistician*, **73**, 340–349.

[23] Martínez-Flórez, G.; Arnold, B.C.; Bolfarine, H. and Gómez, H.W. (2013a). The multivariate alpha-power model, *Journal of Statiscal Planning and Inference*, **143**, 1236–1247.

[24] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2013b). Asymmetric regression models with limited responses with an application to antibody response to vaccine, *Biometrical Journal*, **55**, 156–172.

[25] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2013c). Skew-normal alpha-power model, *Statistics: A Journal of Theoretical and Applied Statistics*, **48**, 1414–1428.

[26] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2014a). An alpha-power extension for the Birnbaum-Saunders distribution, *Statistics: A Journal of Theoretical and Applied Statistics*, **48**, 896–912.

[27] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2014b). The log alpha-power asymmetric distribution with application to air pollution, *Environmetrics*, **25**, 44–56.

[28] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2015). Doubly censored power-normal regression models with inflation, *Test*, **24**, 265–286.

[29] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2017). The log-linear Birnbaum-Saunders power model, *Methodology and Computing in Applied Probability*, **19**, 913–933.

[30] Mazucheli, J.; Menezes, A.F.B.; Alqallaf, F. and Ghitany, M.E. (2021). Bias-corrected Maximum Likelihood estimators of the parameters of the Unit-Weibull distribution, *Austrian Journal of Statistics*, **50**, 41–53.

[31] Melo, T.F.N.; Ferrari, S.L.P. and Patriota, A.G. (2018). Improved estimation in a general multivariate elliptical model, *Brazilian Journal of Probability and Statistics*, **32**, 44–68.

[32] Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate date, *IEEE Transactions on Reliability*, **42**, 299–302.

[33] Mudholkar, G.S.; Srivastava, D.K. and Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus motor failure data, *Technometrics*, **37**, 436–445.

[34] Murthy, D.P.; Xie, M. and Jiang, R. (2004). *Weibull Models*, John Wiley & Sons, New York.

[35] Nadarajah, S. (2005). The exponentiated Gumbel distribution with climate application, *Environmetrics*, **17**, 13–23.

[36] Nadarajah, S. and Gupta, A.K. (2007). The exponentiated gamma distribution with application to drought data, *Calcutta Statistical Association Bulletin*, **59**, 29–54.

[37] Pagui, K.; Salvan, E.C. and Sartori, N. (2017). Median bias reduction of maximum likelihood estimates, *Biometrika*, **104**, 923–938.

[38] Pewsey, A.; Gómez, H.W. and Bolfarine, H. (2012). Likelihood-based inference for power distributions, *Test*, **21**, 775–789.

[39] R Core Team. (2021). *R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing*, ISBN: 3-900051-07-0, version 4.1.1

[40] Rieck, J.R. and Nedelman, J.R. (1991). A log-linear model for the Birnbaum-Saunders distribution, *Technometrics*, **33**, 51–60.

[41] Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions, *Journal of Statistical Planning and Inference*, **136**, 4259–4275.

[42] Zhao, J. and Kim, H.M. (2016). Power t distribution, *Communications for Statistical Applications and Method*, **23**, 321–334.

REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 509–533
https://doi.org/10.57805/revstat.v21i4.440

Estimation, Prediction and Life Testing Plan for the Exponentiated Gumbel Type-II Progressive Censored Data

Authors: Kousik Maiti
– Department of Mathematics, National Institute of Technology Rourkela, Rourkela-769008, India
kousikulu@gmail.com

Suchandan Kayal ✉
– Department of Mathematics, National Institute of Technology Rourkela, Rourkela-769008, India
kayals@nitrkl.ac.in ; suchandan.kayal@gmail.com

Received: August 2020 Revised: February 2022 Accepted: February 2022

Abstract:

- This article accentuates the estimation and prediction of a three-parameter exponentiated Gumbel type-II (EGT-II) distribution when the data are progressively type-II (PT-II) censored. We obtain maximum likelihood (ML) estimates using expectation maximization (EM) and stochastic expectation maximization (StEM) algorithms. The existence and uniqueness of the ML estimates are discussed. We construct bootstrap confidence intervals. The Bayes estimates are derived with respect to a general entropy loss function. We adopt Lindley's approximation, importance sampling and Metropolis-Hastings (MH) methods. The highest posterior density credible interval is computed based on MH algorithm. Bayesian predictors and associated Bayesian predictive interval estimates are obtained. A real life data set is considered for the purpose of illustration. Finally, we propose different criteria for comparison of different sampling schemes in order to obtain the optimal sampling scheme.

Keywords:

- *EM algorithm; stochastic EM algorithm; Lindley's approximation; importance sampling; MH algorithm; optimal censoring.*

AMS Subject Classification:

- 62F10, 62F12, 62F15, 62F40, 62N02.

✉ Corresponding author.

1. INTRODUCTION

In lifetime and reliability studies, an experimenter may not have complete information of the failure times for each and every experimental units. Due to various reasons, it is sometimes required to remove few units from an experiment and as a result, one gets censored data set. There are two most common censoring schemes: (i) type-I and (ii) type-II. The type-I censoring is censored at fixed time, whereas the type-II censoring is censored at a fixed number. These two censoring schemes can not handle the situations, in which we need to remove units at various stages of a test. The removal of experimental units can be done in the progressive censoring scheme. In this paper, we consider PT-II censoring scheme for estimation and prediction for an EGT-II distribution. The PT-II censored scheme is described below. Let n units be placed in a life test. It is pre-decided by the experimenter that m number of failures will be observed. At the time of first failure, we assume that Φ_1 of the remaining $n-1$ surviving units are randomly withdrawn from the experiment. Further, Φ_2 of the remaining $n-\Phi_1-2$ units are removed from the on-going experiment. This procedure continues till the occurrence of m-th failure. We remove all the remaining surviving units $\Phi_m = n-m-\Phi_1-\cdots-\Phi_{m-1}$, when the m-th failure takes place. In PT-II censoring scheme, we denote the m observed failure times as $x_{1:m:n}, ..., x_{m:m:n}$. For simplicity, we use $x_i = x_{i:m:n}$, for $i = 1, ..., m$.

The most popular lifetime models are those with monotone hazard rates (gamma, Weibull), which reflect a wear out or a work hardening behaviour of the population under study. However, there are many other situations, in which the failure pattern is somehow different. When studying the life-cycle of an industrial product or the entire life-span of a biological entity, the three-phase behaviour of the failure rate is likely to be observed. For example, consider a high failure rate in infancy which decreases to a certain level, where it remains fixed for some time, and then increases from a point onwards due to ageing. Thus, in this case, a model having bathtub-shaped hazard rate will be appropriate to study the population's survival capacity. Further, there are also some situations, in which the failure pattern looks like upside-down bathtub. The distributions with upside-down bathtub-shaped hazard rate function is often associated with overload of a component or a subsystem. Intuitively, a lifetime distribution with upside-down shaped hazard rate would suggest a hard stress on the components, leading to fast ageing processes for a part of them but leading to a decreasing failure rate for the surviving items after the stress. There are various real life applications, when the data show upside-down bathtub shape hazard rates. For example, Langlands *et al.* [10] studied cases of breast carcinoma and showed that the associated hazard rate has upside-down bathtub shape. We refer to Efron [5] for more applications of this type of hazard rate functions. The EGT-II distribution has a upside-down bathtub-shaped hazard rate function. Thus, this distribution can be useful in modelling population having bathtub-shaped failure pattern. In addition, this distribution can model various types of data as it can take various shapes (Leptokurtic, platykurtic with thick and thin tails) for various choices of the parameters.

Let X be a random variable following EGT-II distribution, with probability density and cumulative distribution functions respectively given by

$$f_{\boldsymbol{X}}(x:\alpha,\beta,\gamma) = \alpha\beta\gamma x^{-\beta-1}\exp\{-\gamma x^{-\beta}\}\Big(1-\exp\{-\gamma x^{-\beta}\}\Big)^{\alpha-1} \tag{1.1}$$

and

$$F_{\boldsymbol{X}}(x:\alpha,\beta,\gamma) \;=\; 1-\Big(1-\exp\{-\gamma x^{-\beta}\}\Big)^{\alpha}, \tag{1.2}$$

for $x>0$ and $\alpha,\beta,\gamma>0$. The constants α and β are known as the shape parameters, whereas γ is known as the scale parameter. We denote $X\sim$ EGT-II(α,β,γ) if X has the distribution function given by (1.2). The EGT-II distribution is a generalization of various well known statistical models. When $\alpha=1$, $\beta=1$, $\beta=2$ and $\gamma=1$, the EGT-II distribution reduces to the Gumbel type-II, generalized inverted exponential, inverted exponentiated Rayleigh and exponentiated Frechet distributions, respectively. The EGT-II distribution becomes Frechet distribution for $\alpha=\gamma=1$.

Several authors have considered estimation of parameters and some reliability characteristics of various lifetime distributions based on PT-II censored observations. Maiti and Kayal [13] considered estimation for the generalized Frechet distribution based on the PT-II censored data. Ghanbari *et al.* [7] studied estimation of stress-strength reliability for Marshall-Olkin distributions based on PT-II censored samples. Ren and Gui [17] explored goodness of fit test for Rayleigh distribution based on PT-II censored samples. Tarvirdizade and Nematollahi [20] proposed some inferences for the power-exponential hazard rate distribution under PT-II censored data. To the best of authors' knowledge, nobody has considered the problem of estimation of parameters of EGT-II distribution based on PT-II censored data. It is already seen that this distribution can be considered as an alternative lifetime model since it has upside-down bathtub shaped hazard rate function, which is useful in various places.

The aim of this paper is three-fold. First, we consider statistical inference of EGT-II distribution based on the PT-II censored data. The existence and uniqueness of the maximum likelihood estimates (MLEs) are investigated. Further, we obtain MLEs of the parameters. The closed-form solutions of the likelihood equations can not be obtained. Thus, we apply EM algorithm. We also use stochastic EM algorithm to compute the desired MLEs. Confidence intervals using bootstrap algorithms are obtained. The Bayes estimates are derived. It is noticed that the explicit expressions of the Bayes estimates can not be obtained. So, we use Lindley's approximation and importance sampling methods. The Metropolies-Hastings algorithm is also used for this purpose. Second, we study Bayesian prediction problem, and obtain Bayesian prediction intervals. Third, we consider optimal life testing plan for the present problem.

The paper is organized as follows. In Section 2, we present sufficient condition for the existence and uniqueness of MLEs. For the purpose of computation, two algorithms: EM and stochastic EM are used. In Section 3, we obtain observed Fisher's information matrix. The bootstrap confidence intervals are constructed in Section 4. Section 5 provides the form of Bayes estimates with respect to the entropy loss function. Since explicit expressions of the Bayes estimates do not exist, we use various approximation methods to compute the estimates in Section 6. The prediction problem has been considered in Section 7 from Bayesian point of view. Bayesian predictive intervals are also obtained. Data analysis is carried out in Section 8 based on a real life data set. In Section 9, we propose optimal PT-II censoring plan. Finally, Section 10 concludes the paper.

2. ML ESTIMATES AND THEIR COMPUTATION

In this section, first, we show that MLEs of the parameters exist and unique based on the PT-II censored sample.

2.1. Existence and uniqueness of the MLEs

Consider PT-II censored sample of size m from a sample of size n from EGT-II distribution as $\boldsymbol{X} = (X_1, ..., X_m)$. The likelihood function of α, β and γ is given by

$$L(\alpha, \beta, \gamma|\boldsymbol{x}) \propto \alpha^m \beta^m \gamma^m \prod_{i=1}^{m} x_i^{-(\beta+1)} \exp\{-\gamma x_i^{-\beta}\}\Big(1 - \exp\{-\gamma x_i^{-\beta}\}\Big)^{\alpha(\Phi_i+1)-1},$$

where $\boldsymbol{x} = (x_1, ..., x_m)$. Here, $x_1 \leq ... \leq x_m$. The log-likelihood function is denoted by $\ell(\alpha, \beta, \gamma|\boldsymbol{x}) = \ln L(\alpha, \beta, \gamma|\boldsymbol{x})$. The MLEs of α, β and γ can be obtained after solving first order partial derivatives of the log-likelihood function with respect to the parameters equal to zero, simultaneously. The normal equations are

$$\frac{\partial \ell}{\partial \alpha} = \frac{m}{\alpha} + \sum_{i=1}^{m} (1 + \Phi_i) \ln\Big(1 - \exp\{-\gamma x_i^{-\beta}\}\Big) = 0, \tag{2.1}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \frac{m}{\beta} - \gamma \sum_{i=1}^{m} \frac{(\alpha(1 + \Phi_i) - 1) x_i^{-\beta} \exp\{-\gamma x_i^{-\beta}\} \ln x_i}{1 - \exp\{-\gamma x_i^{-\beta}\}} \\ &\quad + \gamma \sum_{i=1}^{m} x_i^{-\beta} \ln x_i - \sum_{i=1}^{m} \ln x_i = 0, \end{aligned} \tag{2.2}$$

$$\frac{\partial \ell}{\partial \gamma} = \frac{m}{\gamma} + \sum_{i=1}^{m} \frac{(\alpha(1 + \Phi_i) - 1) x_i^{-\beta} \exp\{-\gamma x_i^{-\beta}\}}{1 - \exp\{-\gamma x_i^{-\beta}\}} - \sum_{i=1}^{m} x_i^{-\beta} = 0. \tag{2.3}$$

Note that the closed forms of the MLEs do not exist. So, to get approximate values of the MLEs, we use EM algorithm, which is presented in the following subsection. An important question always comes out whether the MLEs exist, and unique. To investigate this, note that the domain of $\ell(\alpha, \beta, \gamma|\boldsymbol{x})$ is $(0, \infty) \times (0, \infty) \times (0, \infty)$. So, our goal is to show that for $(\alpha, \beta, \gamma) \in (0, \infty) \times (0, \infty) \times (0, \infty)$, the function $\ell(\alpha, \beta, \gamma|\boldsymbol{x})$ has unique maximum. The second order partial derivatives of ℓ with respect to α, β and γ can be shown to be strictly negative under the following conditions:

$$\frac{\partial^2 \ell}{\partial \alpha^2} < 0, \tag{2.4}$$

$$\frac{\partial^2 \ell}{\partial \beta^2} < 0, \quad \text{if} \quad \alpha(1 + \Phi_i) > 1, \tag{2.5}$$

$$\frac{\partial^2 \ell}{\partial \gamma^2} < 0, \quad \text{if} \quad \alpha(1 + \Phi_i) > 1. \tag{2.6}$$

Therefore, ℓ is a strictly concave function with respect to one of the parameters keeping other two parameters fixed. For fixed $(\beta,\gamma),(\alpha,\gamma)$ and (α,β), we respectively have

$$\begin{aligned}
&\lim_{\alpha\to 0}\ell(\alpha,\beta,\gamma|\boldsymbol{x})=-\infty, \quad \lim_{\alpha\to\infty}\ell(\alpha,\beta,\gamma|\boldsymbol{x})=-\infty,\\
&\lim_{\beta\to 0}\ell(\alpha,\beta,\gamma|\boldsymbol{x})=-\infty, \quad \lim_{\beta\to\infty}\ell(\alpha,\beta,\gamma|\boldsymbol{x})=-\infty,\\
&\lim_{\gamma\to 0}\ell(\alpha,\beta,\gamma|\boldsymbol{x})=-\infty, \quad \lim_{\gamma\to\infty}\ell(\alpha,\beta,\gamma|\boldsymbol{x})=-\infty.
\end{aligned}$$

So, $\ell(\alpha,\beta,\gamma|\boldsymbol{x})$ is a unimodal function with respect to α,β and γ, when other two associated parameters are fixed. Now, proceeding with the similar arguments as in Dey *et al.* [4], we get the following theorem, which provides sufficient conditions for the existence and uniqueness of MLEs.

Theorem 2.1. *The MLEs of α,β and γ when $(\alpha,\beta,\gamma)\in(0,\infty)\times(0,\infty)\times(0,\infty)$ exist and unique based on the PT-II censored sample, provided $\alpha(1+\Phi_i)>1$.*

Remark 2.1. From real data set, which are presented in Section 8, we notice that the sufficient condition in Theorem 2.1 is satisfied. Thus, as stated, the MLEs of the parameters exist and are unique. The profile of the log-likelihood function of α, β and γ for the data set is depicted in Figure 2.

2.2. EM and StEM algorithms

The EM algorithm is very useful iterative process to obtain MLEs of the parameters when the data are censored. For incomplete data problems, the most attractive features of the EM algorithm relative to other optimization techniques are its simplicity and stability. Further, successive iterations of the EM algorithm are guaranteed never to decrease the likelihood function, which is not generally true of gradient methods like Newton-Raphson. Hence, in the case of the unimodal and concave likelihood function, the EM algorithm converges to the global maximizer from any starting value. Due to this, it has been widely used by various authors. One may refer to Singh and Tripathi [19] and Singh *et al.* [18] for computing MLEs of some lifetime distributions using this method. The EM algorithm is described briefly as follows. To start the EM algorithm, the likelihood function of the complete sample which have been put on a test is required. We denote the complete sample by $\boldsymbol{W}=(W_1,...,W_n)$. After conducting the test, we see that the complete sample is a combination of the observe data $\boldsymbol{X}=(X_1,...,X_m)$ and the censored data $\boldsymbol{Z}=(Z_1,...,Z_m)$. Here Z_j is a $1\times\Phi_j$ vector $(Z_{j1},...,Z_{j\Phi_j})$ for $j=1,...,m$. Then, the complete sample is $\boldsymbol{W}=(\boldsymbol{X},\boldsymbol{Z})$. The log-likelihood function of α, β and γ based on the complete sample is given by

$$\begin{aligned}
\ell_C(\alpha,\beta,\gamma|\boldsymbol{w}) \;=\; & n\ln(\alpha\beta\gamma)+\sum_{j=1}^{m}\Bigg[(\alpha-1)\Bigg(\ln(1-\exp\{-\gamma x_j^{-\beta}\})\\
&+\sum_{k=1}^{\Phi_j}\ln(1-\exp\{-\gamma z_{jk}^{-\beta}\})\Bigg)-\gamma\Bigg(x_j^{-\beta}+\sum_{k=1}^{\Phi_j}z_{jk}^{-\beta}\Bigg)\\
&-(\beta+1)\Bigg(\ln x_j+\sum_{k=1}^{\Phi_j}\ln z_{jk}\Bigg)\Bigg].
\end{aligned}\tag{2.7}$$

Further, the pseudo log-likelihood function is obtained in E-step as

$$
\begin{aligned}
(2.8) \qquad L_{\mathrm{p}}(\alpha,\beta,\gamma) \;=\; & n\ln(\alpha\beta\gamma) + (\alpha-1)\sum_{j=1}^{m}\ln(1-\exp\{-\gamma x_j^{-\beta}\}) - \gamma\sum_{j=1}^{m}x_j^{-\beta} \\
& -(\beta+1)\sum_{j=1}^{m}\ln x_j - (\beta+1)\sum_{j=1}^{m}\Phi_j E[\ln Z_{jk}|Z_{jk}>x_j] \\
& +(\alpha-1)\sum_{j=1}^{m}\Phi_j E[\ln(1-\exp\{-\gamma Z_{jk}^{-\beta}\})|Z_{jk}>x_j].
\end{aligned}
$$

Please see the Appendix A for the expressions of the expectations, which are involved in (2.8). In M-step, we will find the values of the parameters such that the pseudo log-likelihood function is maximum. Let $(\alpha^{(k)},\beta^{(k)},\gamma^{(k)})$ be the value of (α,β,γ) obtained after k-th iteration. Mathematically, at the $(k+1)$-th iteration, $(\alpha^{(k+1)},\beta^{(k+1)},\gamma^{(k+1)})$ has to be computed by maximizing the following function based on $(\alpha^{(k)},\beta^{(k)},\gamma^{(k)})$:

$$
\begin{aligned}
(2.9) \qquad L_{\mathrm{p}}^{*}(\alpha,\beta,\gamma) \;=\; & n\ln(\alpha\beta\gamma) - (\beta+1)\sum_{j=1}^{m}\ln x_j - \gamma\sum_{j=1}^{m}x_j^{-\beta} \\
& + (\alpha-1)\sum_{j=1}^{m}\ln(1-\exp\{-\gamma x_j^{-\beta}\}) \\
& - (\beta+1)\sum_{j=1}^{m}\Phi_j E[\ln Z_{jk}|Z_{jk}>x_j,\alpha^{(k)},\beta^{(k)},\gamma^{(k)}] \\
& - \gamma\sum_{j=1}^{m}\Phi_j E[Z_{jk}^{-\beta}|Z_{jk}>x_j,\alpha^{(k)},\beta^{(k)},\gamma^{(k)}] \\
& + (\alpha-1)\sum_{j=1}^{m}\Phi_j E[\ln(1-\exp\{-\gamma Z_{jk}^{-\beta}\})|Z_{jk}>x_j,\alpha^{(k)},\beta^{(k)},\gamma^{(k)}].
\end{aligned}
$$

The normal equations are given by

$$
(2.10) \qquad \frac{n}{\alpha}+\sum_{j=1}^{m}\ln\left(1-e^{-\gamma x_j^{-\beta}}\right)+\sum_{j=1}^{m}E_3\Phi_j=0,
$$

$$
(2.11) \qquad n-(\alpha-1)\beta\gamma\sum_{j=1}^{m}\frac{x_j^{-\beta}e^{-\gamma x_j^{-\beta}}\ln x_j}{1-e^{-\gamma x_j^{-\beta}}}-\beta\left(\sum_{j=1}^{m}E_1\Phi_j-\gamma\sum_{j=1}^{m}x_j^{-\beta}\ln x_j+\sum_{j=1}^{m}\ln x_j\right)=0,
$$

$$
(2.12) \qquad \frac{n}{\gamma}+(\alpha-1)\sum_{j=1}^{m}\frac{x_j^{-\beta}e^{-\gamma x_j^{-\beta}}}{1-e^{-\gamma x_j^{-\beta}}}-\sum_{j=1}^{m}x_j^{-\beta}-\sum_{j=1}^{m}E_2\Phi_j=0,
$$

where $E_1=E[\ln Z_{jk}|Z_{jk}>x_j,\alpha^{(k)},\beta^{(k)},\gamma^{(k)}]$, $E_2=E[Z_{jk}^{-\beta}|Z_{jk}>x_j,\alpha^{(k)},\beta^{(k)},\gamma^{(k)}]$ and $E_3=E[\ln(1-\exp\{-\gamma Z_{jk}^{-\beta}\})|Z_{jk}>x_j,\alpha^{(k)},\beta^{(k)},\gamma^{(k)}]$. The $(k+1)$-th iteration values of the

unknown parameters can be obtained from

$$\alpha^{(k+1)} = -n\left[\sum_{j=1}^{m}\ln\left(1-e^{-\gamma^{(k)}x_j^{-\beta^{(k)}}}\right)+\sum_{j=1}^{m}E_3(x_j;\alpha^{(k)},\beta^{(k)},\gamma^{(k)})\Phi_j\right]^{-1}, \tag{2.13}$$

$$\begin{aligned}\beta^{(k+1)} = {} & \left(\sum_{j=1}^{m}E_1(x_j,\alpha^{(k+1)},\beta^{(k)},\gamma^{(k)})\Phi_j-\gamma^{(k)}\sum_{j=1}^{m}x_j^{-\beta^{(k)}}\ln x_j+\sum_{j=1}^{m}\ln x_j\right)^{-1}\\ & \times\left(n-(\alpha^{(k+1)}-1)\beta^{(k)}\gamma^{(k)}\sum_{j=1}^{m}\frac{x_j^{-\beta^{(k)}}\exp\{-\gamma^{(k)}x_j^{-\beta^{(k)}}\}\ln x_j}{1-\exp\{-\gamma^{(k)}x_j^{-\beta^{(k)}}\}}\right),\end{aligned} \tag{2.14}$$

$$\begin{aligned}\gamma^{(k+1)} = {} & n\left(\sum_{j=1}^{m}x_j^{-\beta^{(k+1)}}-(\alpha^{(k+1)}-1)\sum_{j=1}^{m}\frac{x_j^{-\beta^{(k+1)}}\exp\{-\gamma^{(k)}x_j^{-\beta^{(k+1)}}\}}{1-\exp\{-\gamma^{(k)}x_j^{-\beta^{(k+1)}}\}}\right.\\ & \left.+\sum_{j=1}^{m}E_2(x_j,\alpha^{(k+1)},\beta^{(k+1)},\gamma^{(k)})\Phi_j\right)^{-1}.\end{aligned} \tag{2.15}$$

Next, we present the algorithm.

Algorithm-1

Step-1: Set $k=0$. Given the starting value $(\alpha^{(0)},\beta^{(0)},\gamma^{(0)})$, we estimate the parameters α,β and γ.

Step-2: In E-step, let $(\alpha^{(k)},\beta^{(k)},\gamma^{(k)})$ be an estimate of (α,β,γ) at k-th iteration. We compute the required conditional expectations E_1,E_2 and E_3 and then substitute in (2.9).

Step-3: In M-step, we obtain $(\alpha^{(k+1)},\beta^{(k+1)},\gamma^{(k+1)})$ the updated values of the parameters at $(k+1)$-th iteration by solving Equations (2.13 – 2.15).

Step-4: If $|\alpha^{(k+1)}-\alpha^{(k)}|+|\beta^{(k+1)}-\beta^{(k)}|+|\gamma^{(k+1)}-\gamma^{(k)}|\le\epsilon$ for a given $\epsilon>0$ (small tolerance), then we stop the procedure. The latest values will be the MLEs of α,β and γ.

Step-5: If $|\alpha^{(k+1)}-\alpha^{(k)}|+|\beta^{(k+1)}-\beta^{(k)}|+|\gamma^{(k+1)}-\gamma^{(k)}|>\epsilon$, then set $k=k+1$ and go to the Step-1.

Denote the MLEs of α, β and γ by $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$, respectively.

There are various situations, where EM algorithm is difficult to implement due to difficulty in the expectation step. To overcome this, a novel technique has been introduced in the literature called Stochastic EM algorithm. It consists of replacing E-step of the EM algorithm by one iteration of a stochastic approximation procedure. We refer the reader to Nielsen *et al.* [15] for some discussions on this method. The main advantage of StEM algorithm is that it is usually less complicated and gives more appropriate results compared to EM algorithm for many problems (see Tregouet *et al.* [21]). Similar to EM algorithm, the StEM algorithm has two steps: S-step and M-step. In S-step, the missing observations $\boldsymbol{Z}$ are generated from conditional distribution given observed data $\boldsymbol{X}$. We generate Φ_i independent number of censored lifetimes z_{ij} from the condition distribution function $F_{\boldsymbol{Z}|\boldsymbol{X}}(x_j:\alpha,\beta,\gamma)$

for $j = 1, ..., m$, which is given by

$$F_{\boldsymbol{Z}|\boldsymbol{X}}(x_j : \alpha, \beta, \lambda) = \frac{F_{\boldsymbol{Z}}(z_{jk} : \alpha, \beta, \gamma) - F_{\boldsymbol{X}}(x_j : \alpha, \beta, \gamma)}{1 - F_{\boldsymbol{X}}(x_j : \alpha, \beta, \gamma)}. \tag{2.16}$$

The $\boldsymbol{Z}$ is then substituted to (2.7) to form the pseudo log-likelihood function and then this function is optimized in M-step to get $(\alpha^{(k+1)}, \beta^{(k+1)}, \gamma^{(k+1)})$ for the next iteration. These two steps are repeated until a stationary distribution is reached for each parameter. The mean of the stationary distribution is considered as an estimate of the parameters. For brevity, the details are not presented here.

3. OBSERVED FISHER'S INFORMATION MATRIX

In this section, we compute observed Fisher's information matrix, which can be used for construction of the asymptotic confidence intervals. Louis [12] derived Fisher's information matrix using the missing information based on EM algorithm. According to him, the observed information is equal to complete information minus missing information. That is,

$$I_{\boldsymbol{X}}(\alpha, \beta, \gamma) = I_{\boldsymbol{W}}(\alpha, \beta, \gamma) - I_{\boldsymbol{W}|\boldsymbol{X}}(\alpha, \beta, \gamma), \tag{3.1}$$

where $I_{\boldsymbol{X}}(\alpha, \beta, \gamma)$, $I_{\boldsymbol{W}}(\alpha, \beta, \gamma)$ and $I_{\boldsymbol{W}|\boldsymbol{X}}(\alpha, \beta, \gamma)$ are observed, complete and missing informations, respectively. Let $\ell^* = \ell_C(\boldsymbol{w}; \alpha, \beta, \gamma)$ and $a_{kl} = -E[\frac{\partial^2 \ell^*}{\partial\theta_k \partial\theta_l}]$ for $k, l = 1, 2, 3$, where $\theta_1 = \alpha$, $\theta_2 = \beta$ and $\theta_3 = \gamma$. Then, the complete information matrix $I_{\boldsymbol{W}}(\alpha, \beta, \gamma)$ is given as

$$I_{\boldsymbol{W}}(\alpha, \beta, \gamma) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \tag{3.2}$$

Further, denote $b_{kl} = -\sum_{j=1}^{m} \Phi_j E_{Z_j|X_j}[\frac{\partial^2 \ln f^*}{\partial\theta_k \partial\theta_l}]$ and $f^* = f_{Z_j|X_j}(z_j|x_j, \alpha, \beta, \gamma)$. Thus, the missing information matrix $I_{\boldsymbol{W}|\boldsymbol{X}}(\alpha, \beta, \gamma)$ is

$$I_{\boldsymbol{W}|\boldsymbol{X}}(\alpha, \beta, \gamma) = \sum_{j=1}^{m} \Phi_j I^j_{\boldsymbol{W}|\boldsymbol{X}}(\alpha, \beta, \gamma) = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}, \tag{3.3}$$

where $I^j_{\boldsymbol{W}|\boldsymbol{X}}(\alpha, \beta, \gamma)$ is missing information matrix at the j-th failure time x_j. It is given as

$$I^j_{\boldsymbol{W}|\boldsymbol{X}}(\alpha, \beta, \gamma) = -E_{Z_j|X_j} \begin{pmatrix} \frac{\partial^2 \ln f^*}{\partial\alpha^2} & \frac{\partial^2 \ln f^*}{\partial\alpha\partial\beta} & \frac{\partial^2 \ln f^*}{\partial\alpha\partial\gamma} \\ \frac{\partial^2 \ln f^*}{\partial\beta\partial\alpha} & \frac{\partial^2 \ln f^*}{\partial\beta^2} & \frac{\partial^2 \ln f^*}{\partial\beta\partial\gamma} \\ \frac{\partial^2 \ln f^*}{\partial\gamma\partial\alpha} & \frac{\partial^2 \ln f^*}{\partial\gamma\partial\beta} & \frac{\partial^2 \ln f^*}{\partial\gamma^2} \end{pmatrix}. \tag{3.4}$$

It is worthwhile to mention that the matrices in (3.2) and (3.3) are computed at $(\alpha, \beta, \gamma) = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})$. From the 3×3 order matrices given by (3.2) and (3.3), one can easily compute the observed Fisher's information matrix of the model parameters α, β and γ. We obtain the asymptotic variance covariance matrix $(\hat{M})$ for the MLEs of α, β and γ from the inverse of $I_{\boldsymbol{X}}(\alpha, \beta, \gamma)$, which is given by

$$\hat{M} = \begin{pmatrix} \text{var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\beta}) & \text{cov}(\hat{\alpha}, \hat{\gamma}) \\ \text{cov}(\hat{\alpha}, \hat{\beta}) & \text{var}(\hat{\beta}) & \text{cov}(\hat{\beta}, \hat{\gamma}) \\ \text{cov}(\hat{\alpha}, \hat{\gamma}) & \text{cov}(\hat{\beta}, \hat{\gamma}) & \text{var}(\hat{\gamma}) \end{pmatrix}.$$

The asymptotic confidence intervals of the parameters by using normal approximation (NA) to MLE, and normal approximation of the log-transformed (NL) MLE can be constructed. The derivations are omitted to maintain brevity.

4. BOOTSTRAP CONFIDENCE INTERVALS

In this section, we construct two bootstrap confidence intervals for the parameters. These are the percentile bootstrap (Boot-p) (see Efron and Tibshirani [6]) and the bootstrap-t (Boot-t) (see Hall [8]) methods. The algorithms for these methods are presented below.

Algorithm-2 (Boot-p)

Step-1: From Equations (2.1), (2.2) and (2.3), under the original data sets x_i, $i = 1, ..., m$, we obtain $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$.

Step-2: Based on the values of the estimates of the parameters, generate a bootstrap sample $\boldsymbol{x}^* = (x_1^*, ..., x_m^*)$ for a pre-specified censoring scheme. Then, compute the bootstrap estimates $\hat{\alpha}^*, \hat{\beta}^*$ and $\hat{\gamma}^*$.

Step-3: Repeat Step-2, for $N = 1000$ times to get $(\hat{\alpha}_1^*, ..., \hat{\alpha}_{1000}^*), (\hat{\beta}_1^*, ..., \hat{\beta}_{1000}^*)$ and $(\hat{\gamma}_1^*, ..., \hat{\gamma}_{1000}^*)$.

Step-4: Arrange the values obtained in Step-3 in ascending order and denote $\hat{\alpha}_{(1)}^*, ..., \hat{\alpha}_{(1000)}^*$, $\hat{\beta}_{(1)}^*, ..., \hat{\beta}_{(1000)}^*$ and $\hat{\gamma}_{(1)}^*, ..., \hat{\gamma}_{(1000)}^*$.

Then, for a specified value of σ, the $100(1-\sigma)\%$ Boot-p confidence intervals for α, β and γ are respectively given by

$$\left(\hat{\alpha}_{(N(\frac{\sigma}{2}))}, \hat{\alpha}_{(N(1-\frac{\sigma}{2}))}\right), \left(\hat{\beta}_{(N(\frac{\sigma}{2}))}, \hat{\beta}_{(N(1-\frac{\sigma}{2}))}\right) \text{ and } \left(\hat{\gamma}_{(N(\frac{\sigma}{2}))}, \hat{\gamma}_{(N(1-\frac{\sigma}{2}))}\right).$$

Algorithm-3 (Boot-t)

Step-1: In analogy to Step-1 and Step-2 as in Boot-p method, obtain bootstrap estimates of the unknown parameters.

Step-2: Compute variance-covariance matrix $I^*(\hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^*)^{-1}$. Write

$$T_{\alpha_i}^* = \frac{\hat{\alpha}_i^* - \hat{\alpha}_i}{\sqrt{\widehat{\text{Var}}(\hat{\alpha}_i^*)}}, \quad T_{\beta_i}^* = \frac{\hat{\beta}_i^* - \hat{\beta}_i}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_i^*)}} \text{ and } T_{\gamma_i}^* = \frac{\hat{\gamma}_i^* - \hat{\gamma}_i}{\sqrt{\widehat{\text{Var}}(\hat{\gamma}_i^*)}}$$

for $i = 1, ..., 1000$.

Step-3: Repeat Step-1 and Step-2, $N = 1000$ times and arrange the values in ascending order. Denote

$$T_{\alpha_{(1)}}^*, ..., T_{\alpha_{(1000)}}^*, \quad T_{\beta_{(1)}}^*, ..., T_{\beta_{(1000)}}^* \quad \text{and} \quad T_{\gamma_{(1)}}^*, ..., T_{\gamma_{(1000)}}^*.$$

Thus, for a given σ, the $100(1-\sigma)\%$ Boot-t confidence intervals for α, β and γ are respectively obtained as

$$\left(\hat{T}_{\alpha_{(N(\frac{\sigma}{2}))}}, \hat{T}_{\alpha_{(N(1-\frac{\sigma}{2}))}}\right), \quad \left(\hat{T}_{\beta_{(N(\frac{\sigma}{2}))}}, \hat{T}_{\beta_{(N(1-\frac{\sigma}{2}))}}\right) \quad \text{and} \quad \left(\hat{T}_{\gamma_{(N(\frac{\sigma}{2}))}}, \hat{T}_{\gamma_{(N(1-\frac{\sigma}{2}))}}\right).$$

5. BAYESIAN ESTIMATION

In this section, we focus on obtaining Bayes estimates of α, β and γ with respect to entropy loss function. Let δ be an estimator for the unknown parameter θ. The entropy loss function (ELF) is

$$L_{\mathrm{e}}(\theta,\delta) \;=\; \left(\frac{\delta}{\theta}\right)^{q} - q\ln\left(\frac{\delta}{\theta}\right) - 1, \quad q \neq 0. \tag{5.1}$$

This loss function is asymmetric in nature. The constant q in (5.1) stands for the magnitude and degree of symmetry. The overestimation is dangerous than the underestimation for positive values of q. When q is negative, underestimation is dangerous than the overestimation. The Bayes estimate of θ with respect to this loss function can be obtained using the following tool:

$$\hat{\theta}_{\mathrm{be}} = \left[E_{\theta}\left(\theta^{-q}|\ \boldsymbol{x}\right)\right]^{-\frac{1}{q}}, \quad q \neq 0. \tag{5.2}$$

Note that the Bayes estimate of the parameter θ under ELF reduces to the Bayes estimates with respect to the squared error loss function (SELF) when $q = -1$. For $q = -2$ and 1, it becomes Bayes estimates under the precautionary loss function (PLF) and weighted squared error loss function (WSELF). Prior distributions play an essential role for derivation of the Bayes estimators. There is no clear method on choosing priors for a particular problem. We refer to Arnold and Press [1] for more details on this. Here, we consider independent gamma prior density functions for the parameters α, β and γ as

$$g_1(\alpha : c_1, c_2) \;\propto\; \alpha^{c_1-1}\exp\{-\alpha c_2\}, \tag{5.3}$$

$$g_2(\beta : c_3, c_4) \;\propto\; \beta^{c_3-1}\exp\{-\beta c_4\}, \tag{5.4}$$

$$g_3(\gamma : c_5, c_6) \;\propto\; \gamma^{c_5-1}\exp\{-\gamma c_6\}, \tag{5.5}$$

where $\alpha, \beta, \gamma > 0$ and $c_i > 0;\ i = 1, 2, 3, 4, 5, 6$. The constants c_i's are known as the hyper-parameters. The joint prior distribution of α, β and γ is given by

$$\pi(\alpha,\beta,\gamma) \propto \alpha^{c_1-1}\beta^{c_3-1}\gamma^{c_5-1}\exp\{-(\alpha c_2 + \beta c_4 + \gamma c_6)\}. \tag{5.6}$$

Further, the joint distribution of α, β, γ and $\boldsymbol{X}$ is

$$\begin{aligned}\pi_1(\alpha,\beta,\gamma,\boldsymbol{x}) \;\propto\; & \alpha^{m+c_1-1}\beta^{m+c_3-1}\gamma^{m+c_5-1}\exp\{-(\alpha c_2 + \beta c_4 + \gamma c_6)\} \\ & \times \prod_{i=1}^{m} x_i^{-(\beta+1)}\exp\{-\gamma x_i^{-\beta}\}(1-\exp\{-\gamma x_i^{-\beta}\})^{\Phi_i+\alpha-1}.\end{aligned} \tag{5.7}$$

The posterior distribution of α, β and γ given $\boldsymbol{X} = \boldsymbol{x}$ is obtained as

$$\begin{aligned}\Pi(\alpha,\beta,\gamma|\boldsymbol{x}) \;=\; & \frac{1}{k}\alpha^{m+c_1-1}\beta^{m+c_3-1}\gamma^{m+c_5-1}\exp\{-(\alpha c_2 + \beta c_4 + \gamma c_6)\} \\ & \times \prod_{i=1}^{m} x_i^{-(\beta+1)}\exp\{-\gamma x_i^{-\beta}\}(1-\exp\{-\gamma x_i^{-\beta}\})^{\Phi_i+\alpha-1},\end{aligned} \tag{5.8}$$

where

$$\begin{aligned}k \;=\; & \int_{\alpha=0}^{\infty}\int_{\beta=0}^{\infty}\int_{\gamma=0}^{\infty}\alpha^{m+c_1-1}\beta^{m+c_3-1}\gamma^{m+c_5-1}\exp\{-(\alpha c_2 + \beta c_4 + \gamma c_6)\} \\ & \times \prod_{i=1}^{m} x_i^{-(\beta+1)}\exp\{-\gamma x_i^{-\beta}\}(1-\exp\{-\gamma x_i^{-\beta}\})^{\Phi_i+\alpha-1}\,d\alpha\,d\beta\,d\gamma.\end{aligned} \tag{5.9}$$

Thus, from Equation (5.2), the Bayes estimate of α with respect to the entropy loss function is obtained as

$$\begin{aligned}(5.10)\quad \hat{\alpha}_{\mathrm{be}} &= \Bigg[\frac{1}{k}\int_{\alpha=0}^{\infty}\int_{\beta=0}^{\infty}\int_{\gamma=0}^{\infty}\alpha^{m+c_1-q-1}\beta^{m+c_3-1}\gamma^{m+c_5-1}\exp\{-(\alpha c_2+\beta c_4+\gamma c_6)\}\\ &\quad\times\prod_{i=1}^{m}\exp\{-\gamma x_i^{-\beta}\}(1-\exp\{-\gamma x_i^{-\beta}\})^{\Phi_i+\alpha-1}x_i^{-(\beta+1)}\,d\alpha\,d\beta\,d\gamma\Bigg]^{-\frac{1}{q}},\quad q\neq 0.\end{aligned}$$

Similarly, the Bayes estimates of β and γ with respect to the entropy loss function can be obtained. We omit these for the sake of conciseness. Below, we discuss how to compute Bayes estimates using some well known techniques.

6. COMPUTING METHODS FOR BAYESIAN ESTIMATION

In the previous section, we have seen that the desired Bayes estimates can not be obtained in explicit forms. So, we consider approximation methods in this section. First, we explain Lindley's method (see Lindley [11]).

6.1. Lindley's approximation method

Let θ_1,θ_2 and θ_3 be the unknown parameters of a statistical model and $u(\boldsymbol{\theta})$ be a function of the parameters, where $\boldsymbol{\theta}=(\theta_1,\theta_2,\theta_3)$. It is known that the Bayes estimate of $u(\boldsymbol{\theta})$ is evaluated in terms of expectation, where the expectation is taken with respect to posterior distribution. Let $l(\boldsymbol{\theta}|\boldsymbol{x})$ denote the log-likelihood function and $\rho(\boldsymbol{\theta})$ is the logarithm of the joint prior distribution of θ_1,θ_2 and θ_3. From the Lindley's approximation technique, we obtain (*see* Lindley [11])

$$\begin{aligned}(6.1)\quad \hat{\delta}_{\mathrm{be}}(\boldsymbol{x}) &\approx u(\hat{\boldsymbol{\theta}})+W(\hat{\boldsymbol{\theta}})+\rho_1(\hat{\boldsymbol{\theta}})W_{123}+\rho_2(\hat{\boldsymbol{\theta}})W_{213}+\rho_3(\hat{\boldsymbol{\theta}})W_{321}\\ &\quad+0.5\Big[\ell^*_{300}V_{123}+\ell^*_{030}V_{213}+\ell^*_{003}V_{321}+2\ell^*_{111}(E_{123}+E_{213}+E_{312})\\ &\quad+\ell^*_{210}C_{123}+\ell^*_{201}C_{132}+\ell^*_{120}C_{213}+\ell^*_{102}C_{312}+\ell^*_{021}C_{231}+\ell^*_{012}C_{321}\Big],\end{aligned}$$

where $W(\hat{\boldsymbol{\theta}})=\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}u_{ij}(\hat{\boldsymbol{\theta}})\tau_{ij}(\hat{\boldsymbol{\theta}})$, $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$, $\ell^*_{ijk}=\frac{\partial^3\ell(\boldsymbol{\theta}|\boldsymbol{x})}{\partial\theta_i\partial\theta_j\partial\theta_k}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ with $i,j,k=0,1,2,3$ such that $i+j+k=3$, τ_{ij} is the (i,j)-th element in the inverse matrix of $[-\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{x})}{\partial\theta_i\partial\theta_j}]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. Other unknown terms of (6.1) are given as

$$\begin{aligned}W_{ijk} &= u_i\tau_{ii}(\hat{\boldsymbol{\theta}})+u_j\tau_{ji}(\hat{\boldsymbol{\theta}})+u_k\tau_{ki}(\hat{\boldsymbol{\theta}}),\\ V_{ijk} &= \tau_{ii}(\hat{\boldsymbol{\theta}})(u_i\tau_{ii}(\hat{\boldsymbol{\theta}})+u_j\tau_{ij}(\hat{\boldsymbol{\theta}})+u_k\tau_{ik}(\hat{\boldsymbol{\theta}})),\\ E_{ijk} &= u_i(\tau_{ii}(\hat{\boldsymbol{\theta}})\tau_{jk}(\hat{\boldsymbol{\theta}})+2\tau_{ij}(\hat{\boldsymbol{\theta}})\tau_{ik}(\hat{\boldsymbol{\theta}}))\end{aligned}$$

and

$$C_{ijk} = 3u_i\tau_{ii}(\hat{\boldsymbol{\theta}})\tau_{ij}(\hat{\boldsymbol{\theta}})+u_j(\tau_{ii}(\hat{\boldsymbol{\theta}})\tau_{jj}(\hat{\boldsymbol{\theta}})+2\tau^2_{ij}(\hat{\boldsymbol{\theta}}))+u_k(\tau_{ii}(\hat{\boldsymbol{\theta}})\tau_{jk}(\hat{\boldsymbol{\theta}})+2\tau_{ij}(\hat{\boldsymbol{\theta}})\tau_{ik}(\hat{\boldsymbol{\theta}})).$$

Further, $u_{ij}(\hat{\boldsymbol{\theta}}) = \frac{\partial^2 u(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, $u_i(\hat{\boldsymbol{\theta}}) = \frac{\partial u(\boldsymbol{\theta})}{\partial\theta_i}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, $\rho_i(\boldsymbol{\theta}) = \frac{\partial\rho(\boldsymbol{\theta})}{\partial\theta_i}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, and $\rho(\boldsymbol{\theta})$ is equal to the logarithmic of the joint prior distribution of θ_1, θ_2 and θ_3, where $i, j, k = 1, 2, 3$. Now, we provide approximate Bayes estimate for the unknown parameter α with respect to the entropy loss function. In order to write the Bayes estimate of α with respect to the entropy loss function, we have $u(\alpha, \beta, \gamma) = \alpha^{-q}$, $u_1 = -q\alpha^{-(q+1)}$, $u_{11} = q(q+1)\alpha^{-(q+2)}$ and $u_2 = u_3 = u_{12} = u_{13} = u_{21} = u_{22} = u_{23} = u_{31} = u_{32} = u_{33} = 0$. Thus, from (6.1), the approximate Bayes estimate of α with respect to the entropy loss function is obtained as

$$
\begin{aligned}
\hat{\alpha}_{\text{be}}^{LI} \;=\; & \Big[\alpha^{-q} + 0.5\Big[q(q+1)\alpha^{-(q+2)}\tau_{11} - q\alpha^{-(q+1)}\Big\{\ell^*_{300}\tau_{11}^2 + \ell^*_{030}\tau_{21}\tau_{22} \\
& + \ell^*_{003}\tau_{31}\tau_{33} + 2\ell^*_{111}(\tau_{11}\tau_{23} + 2\tau_{13}\tau_{12}) + \ell^*_{120}(\tau_{11}\tau_{22} + 2\tau_{21}^2) \\
& + \ell^*_{102}(\tau_{33}\tau_{11} + 2\tau_{31}^2) + \ell^*_{021}(\tau_{22}\tau_{31} + 2\tau_{23}\tau_{21}) \\
& + \ell^*_{012}(\tau_{33}\tau_{21} + 2\tau_{32}\tau_{31}) + 2\rho_1\tau_{11} + 2\rho_2\tau_{12} + 2\rho_3\tau_{13}\Big\}\Big]\Big]^{-\frac{1}{q}}\Big|_{(\alpha,\beta,\gamma)=(\hat{\alpha},\hat{\beta},\hat{\gamma})}.
\end{aligned} \tag{6.2}
$$

Similarly, we can obtain the Bayes estimates of β and γ with respect to the entropy loss function. The expressions are omitted here.

6.2. Importance sampling method

In the subsection, we consider another approximation technique known as importance sampling method to obtain the Bayes estimates for the parameters. We rewrite the posterior distribution of α, β and γ given in (5.8) as

$$
\begin{aligned}
\Pi(\alpha, \beta, \gamma|\boldsymbol{x}) \;\propto\; & G_\beta\Big(m + c_3, c_4 + \sum_{i=1}^{m} \ln x_i\Big).G_{\gamma|\beta}\Big(m + c_5, c_6 + \sum_{i=1}^{m} x_i^{-\beta}\Big) \\
& \times G_{\alpha|\beta,\gamma}\Big(m + c_1, c_2 - \sum_{i=1}^{m} \ln(1 - \exp\{-\gamma x_i^{-\beta}\})\Big).\psi(\alpha, \beta, \gamma),
\end{aligned} \tag{6.3}
$$

where

$$
\begin{aligned}
\psi(\alpha, \beta, \gamma) \;=\; & \frac{(c_4 + \sum_{i=1}^{m} \ln x_i)^{-(m+c_3)}}{(c_2 - \sum_{i=1}^{m} \ln(1 - \exp\{-\gamma x_i^{-\beta}\}))^{(m+c_1)}} \\
& \times \frac{\exp\{\sum_{i=1}^{m}(\Phi_i + 1)\ln(1 - \exp\{-\gamma x_i^{-\beta}\})\}}{(c_6 + \sum_{i=1}^{m} x_i^{-\beta})^{m+c_5}\exp\{\sum_{i=1}^{m} \ln x_i\}}.
\end{aligned} \tag{6.4}
$$

Below, we present the steps which will be used for the implementation of importance sampling technique.

Algorithm-4

Step-1: Generate β from $G_\beta(m + c_3, c_4 + \sum_{i=1}^{m} \ln x_i)$, that is, from a gamma distribution with shape parameter $(m + c_3)$ and scale parameter $(c_4 + \sum_{i=1}^{m} \ln x_i))$.

Step-2: For a given β as obtained in Step-1, we generate γ from $G_{\gamma|\beta}(m + c_5,\ c_6 + \sum_{i=1}^{m} x_i^{-\beta})$.

Step-3: For β and γ as generated in Step-1 and Step-2, we will generate parameter α from $G_{\alpha|\beta,\gamma}(m+c_1, c_2 - \sum_{i=1}^{m} \ln(1-\exp\{-\gamma x_i^{-\beta}\}))$.

Step-4: Repeat Steps-1,2 and 3, $N = 1000$ times to obtain $(\alpha_1, \beta_1, \gamma_1), ..., (\alpha_N, \beta_N, \gamma_N)$.

Finally, the Bayes estimate of a parametric function $g(\alpha, \beta, \gamma)$ with respect to entropy loss function is obtained as

$$\hat{g}_{\text{be}}^{IS}(\alpha,\beta,\gamma) = \left[\frac{\sum_{i=1}^{N} g(\alpha_i,\beta_i,\gamma_i)^{-q}\psi(\alpha_i,\beta_i,\gamma_i)}{\sum_{i=1}^{N}\psi(\alpha_i,\beta_i,\gamma_i)}\right]^{-\frac{1}{q}}. \tag{6.5}$$

Substituting $q = -1, 1$ and $q = -2$ in the above expression, we obtain Bayes estimates with respect to the SELF, WSELF and PLF, respectively. Further, to get the Bayes estimates of α, β and γ, one needs to respectively replace α, β and γ in place of $g(\alpha, \beta, \gamma)$ in (6.5).

6.3. Metropolis-Hastings algorithm

In this subsection, we use an alternative method to get Bayes estimates of α, β and γ using Gibbs sampling method and Metropolis-Hastings algorithm. The MH algorithm is also used for the construction of credible intervals. After analysing the posterior distribution given by (5.8), the marginal posterior distribution of α given β, γ and $\boldsymbol{x}$ is obtained as

$$\Pi_1(\alpha|\beta,\gamma,\boldsymbol{x}) \propto G\left(m+c_1, \left(c_1 - \sum_{i=1}^{m}\ln(1-\exp\{-\gamma x_i^{-\beta}\})\right)\right). \tag{6.6}$$

Similarly, the marginal posterior distributions of β given α, γ and $\boldsymbol{x}$; and γ given α, β and $\boldsymbol{x}$ can be obtained as

$$\Pi_2(\beta|\alpha,\gamma,\boldsymbol{x}) \propto \beta^{m+c_3-1}\exp\{-\beta c_4\}\prod_{i=1}^{m}\frac{x_i^{-\beta}\exp\{-\gamma x_i^{-\beta}\}}{(1-\exp\{-\gamma x_i^{-\beta}\})} \tag{6.7}$$

and

$$\Pi_3(\gamma|\alpha,\beta,\boldsymbol{x}) \propto \gamma^{m+c_5-1}\exp\{-\gamma c_6\}\prod_{i=1}^{m}\frac{\exp\{-\gamma x_i^{-\beta}\}}{(1-\exp\{-\gamma x_i^{-\beta}\})}, \tag{6.8}$$

respectively. Note that the marginal posterior distribution in (6.6) is gamma distribution. But, other two marginal posterior distributions in (6.7) and (6.8) do not follow any know models. Thus, one has to generate random samples for β and γ from the normal proposal distribution. The following algorithm is useful for the generation of the posterior samples.

Algorithm-5

Step-1: Set an initial value $(\alpha^{(0)}, \beta^{(0)}, \gamma^{(0)})$ and set $j = 1$.

Step-2: Generate β^* and γ^* from the proposal distributions $N(\beta^{(j-1)}, \text{var}(\beta))$ and $N(\gamma^{(j-1)}, \text{var}(\gamma))$, respectively. Then, generate α^* from $G(m+c_1, (c_1 - \sum_{i=1}^{m}\ln(1-\exp\{-\gamma^{(j-1)}x_i^{-\beta^{(j-1)}}\})))$.

Step-3: Compute

$$\omega_\beta = \min\left\{1, \frac{\Pi_2(\beta^*|\alpha^{(j)}, \gamma^{(j)}, \boldsymbol{x})}{\Pi_2(\beta^{(j-1)}|\alpha^{(j-1)}, \gamma^{(j-1)}, \boldsymbol{x})}\right\} \tag{6.9}$$

$$\text{and} \quad \omega_\gamma = \min\left\{1, \frac{\Pi_3(\gamma^*|\alpha^{(j)}, \beta^{(j)}, \boldsymbol{x})}{\Pi_3(\gamma^{(j-1)}|\alpha^{(j-1)}, \beta^{(j-1)}, \boldsymbol{x})}\right\}.$$

Step-4: Generate samples u_2 and u_3 from uniform distribution $U(0,1)$.

Step-5: If $u_2 \leq \omega_\beta$ and $u_3 \leq \omega_\gamma$ then $\beta^{(j)} \longleftarrow \beta^*$, else $\beta^{(j)} \longleftarrow \beta^{(j-1)}$ and $\gamma^{(j)} \longleftarrow \gamma^*$, else $\gamma^{(j)} \longleftarrow \gamma^{(j-1)}$, respectively. Further, set $j = j + 1$.

Step-6: Repeat Steps $(3-5)$, $N = 1000$ times to obtain MCMC samples. These are denoted as $(\alpha^{(1)}, \beta^{(1)}, \gamma^{(1)}), ..., (\alpha^{(N)}, \beta^{(N)}, \gamma^{(N)})$.

Now, the Bayes estimate of α with respect to entropy loss function based on MCMC samples is given by

$$\hat{\alpha}_{be}^{MH} = \left[\frac{1}{N}\sum_{j=1}^{N}\left(\alpha^{(j)}\right)^{-q}\right]^{-\frac{1}{q}}. \tag{6.10}$$

Similarly, the Bayes estimates of β and γ under entropy loss function can be obtained. Next, we compute HPD credible intervals of α, β and γ by using the method due to Chen and Shao [3]. Here, we use MH algorithm to generate samples from the posterior density. After that we arrange $\hat{\alpha}^{(j)}, \hat{\beta}^{(j)}$ and $\hat{\gamma}^{(j)}$ in ascending order, and denote $\hat{\alpha}^{(1)}, ..., \hat{\alpha}^{(N)}$, $\hat{\beta}^{(1)}, ..., \hat{\beta}^{(N)}$ and $\hat{\gamma}^{(1)}, ..., \hat{\gamma}^{(N)}$, respectively. Thus, the $100(1-\sigma)\%$ credible intervals for α, β and γ are respectively given by

$$\left(\hat{\alpha}^{(N(\frac{\sigma}{2}))}, \hat{\alpha}^{(N(1-\frac{\sigma}{2}))}\right), \quad \left(\hat{\beta}^{(N(\frac{\sigma}{2}))}, \hat{\beta}^{(N(1-\frac{\sigma}{2}))}\right) \quad \text{and} \quad \left(\hat{\gamma}^{(N(\frac{\sigma}{2}))}, \hat{\gamma}^{(N(1-\frac{\sigma}{2}))}\right).$$

6.4. Computation of hyper-parameters

Here, we briefly discuss the procedure how to calculate the hyper-parameters when informative priors are known to us. The hyper-parameters are c_1, c_2, c_3, c_4, c_5 and c_6. These are obtained from gamma prior distributions as given in Section 5. Suppose r samples are available from the EGT-II distribution. The MLEs of the parameters α, β and γ are $\hat{\alpha}^j, \hat{\beta}^j$ and $\hat{\gamma}^j$ for $j = 1, ..., r$, respectively for each of these r number of samples. Note that these hyper-parameter values are evaluated from the past data set. First, we calculate hyper-parameters c_1 and c_2. The mean and variance of the gamma prior of α are c_1/c_2 and c_1/c_2^2, respectively. Further, the mean and variance of the MLEs of α for r samples are $\frac{1}{r}\sum_{j=1}^{r}\hat{\alpha}^j$ and $\frac{1}{r-1}\sum_{j=1}^{r}(\hat{\alpha}^j - \frac{1}{r}\sum_{j=1}^{r}\hat{\alpha}^j)^2$, respectively. Therefore, $\frac{c_1}{c_2} = \frac{1}{r}\sum_{j=1}^{r}\hat{\alpha}^j$ and $\frac{c_1}{c_2^2} = \frac{1}{r-1}\sum_{j=1}^{r}(\hat{\alpha}^j - \frac{1}{r}\sum_{j=1}^{r}\hat{\alpha}^j)^2$. Solving these equations, we get

$$c_1 = \frac{(\frac{1}{r}\sum_{j=1}^{r}\hat{\alpha}^j)^2}{\frac{1}{r-1}\sum_{j=1}^{r}(\hat{\alpha}^j - \frac{1}{r}\sum_{j=1}^{r}\hat{\alpha}^j)^2}$$

$$\text{and} \quad c_2 = \frac{\frac{1}{r}\sum_{j=1}^{r}\hat{\alpha}^j}{\frac{1}{r-1}\sum_{j=1}^{r}(\hat{\alpha}^j - \frac{1}{r}\sum_{j=1}^{r}\hat{\alpha}^j)^2}. \tag{6.11}$$

Similarly, other hyper-parameters c_3, c_4 and c_5, c_6 can be obtained from (6.11) replacing $\hat{\beta}^j$ and $\hat{\gamma}^j$ in place of $\hat{\alpha}^j$, respectively.

7. BAYESIAN PREDICTION AND INTERVAL ESTIMATION

In this section, we discuss Bayesian prediction for the future observations depending upon the PT-II censored sample. We assume that the sample is taken from the EGT-II distribution. We also compute the corresponding prediction intervals. Many authors have studied prediction problems related to Bayesian prediction and interval estimation. We refer to Bdair *et al.* [2] and Maiti and Kayal [13] for some references. We illustrate the one-sample prediction problem. Suppose n independent life testing units are put in an experiment. Let $\boldsymbol{x} = (x_1, ..., x_m)$ be the observed PT-II censored sample. Further, assume that the censoring scheme is taken as $\boldsymbol{\Phi} = (\Phi_1, ..., \Phi_m)$. Let $y_i = (y_{i1}, ..., y_{i\Phi_i})$ represent the ordered lifetimes of the units which are censored at the i-th failure x_i. Our goal is to predict the future observations based on $\boldsymbol{x}$. We assume that these are $y = (y_{ip}\ ; i = 1, ..., m; p = 1, ..., \Phi_i)$. The conditional density of y under the given information can be obtained as

$$
\begin{aligned}
f_1(y|\boldsymbol{x}, \alpha, \beta, \gamma) \;=\; & \alpha\beta\gamma p \binom{\Phi_i}{p} \sum_{k=0}^{p-1} (-1)^{p-k-1} \binom{p-1}{k} y^{-(\beta+1)} \\
& \times \exp\{-\gamma y^{-\beta}\}(1 - \exp\{-\gamma y^{-\beta}\})^{\alpha(\Phi_i - k)-1} \\
& \times (1 - \exp\{-\gamma x_i^{-\beta}\})^{\alpha(k-\Phi_i)}, \qquad y > x_i.
\end{aligned} \tag{7.1}
$$

The conditional distribution function is

$$
\begin{aligned}
F_1(y|\boldsymbol{x}, \alpha, \beta, \gamma) \;=\; & p \binom{\Phi_i}{p} \sum_{k=0}^{p-1} \frac{(-1)^{p-k-1}}{\Phi_i - k} \binom{p-1}{k} \\
& \times \left[1 - (1 - \exp\{-\gamma x_i^{-\beta}\})^{\alpha(k-\Phi_i)} (1 - \exp\{-\gamma y^{-\beta}\})^{\alpha(\Phi_i - k)}\right].
\end{aligned} \tag{7.2}
$$

Notice that the posterior predictive density and the distribution functions are respectively given by

$$
f_1^*(y|\boldsymbol{x}) = \int_0^\infty \int_0^\infty \int_0^\infty f_1(y|\boldsymbol{x}, \alpha, \beta, \gamma) \Pi(\alpha, \beta, \gamma|\boldsymbol{x})\ d\alpha\, d\beta\, d\gamma \tag{7.3}
$$

and

$$
F_1^*(y|\boldsymbol{x}) = \int_0^\infty \int_0^\infty \int_0^\infty F_1(y|\boldsymbol{x}, \alpha, \beta, \gamma) \Pi(\alpha, \beta, \gamma|\boldsymbol{x})\ d\alpha\, d\beta\, d\gamma. \tag{7.4}
$$

Thus, the Bayesian predictive estimate of y under the entropy loss function is obtained as

$$
\hat{y}_{\text{be}} \;=\; [E(P_1(\alpha, \beta, \gamma)|\boldsymbol{x})]^{-\frac{1}{q}},
$$

where

$$
P_1(\alpha, \beta, \gamma) = \int_{x_i}^\infty y^{-q} f_1(y|\boldsymbol{x}, \alpha, \beta, \gamma)\, dy.
$$

Note that the above integrals can not be evaluated analytically. Therefore, we have to adopt numerical technique for the computation of the predictive estimates. In this purpose, we use importance sampling method which is mentioned in Subsection 6.2. Equation (7.5) can be computed using the importance sampling method as

$$
\hat{y}_{\text{be}}^{BP} = \left[\frac{\sum_{i=1}^{1000} P_1(\alpha_i, \beta_i, \gamma_i)\psi(\alpha_i, \beta_i, \gamma_i)}{\sum_{i=1}^{1000} \psi(\alpha_i, \beta_i, \gamma_i)}\right]^{-1/q}. \tag{7.5}
$$

Now, we obtain the Bayesian predictive interval (BPI). The prior predictive survival function is obtained as

$$S_1(t|\boldsymbol{x},\alpha,\beta,\gamma)=\frac{P(y>t|\boldsymbol{x},\alpha,\beta,\gamma)}{P(y>x_i|\boldsymbol{x},\alpha,\beta,\gamma)}=\frac{\int_t^\infty f_1(u|\boldsymbol{x},\alpha,\beta,\gamma)\,du}{\int_{x_i}^\infty f_1(u|\boldsymbol{x},\alpha,\beta,\gamma)\,du}.$$

The posterior survival function is

$$S_1^*(t|\boldsymbol{x})=\int_0^\infty\int_0^\infty\int_0^\infty S_1(t|\boldsymbol{x},\alpha,\beta,\gamma)\Pi(\alpha,\beta,\gamma|\boldsymbol{x})\;d\alpha\,d\beta\,d\gamma. \tag{7.6}$$

Using (7.6), we obtain two-sided $100(1-\sigma)\%$ equal-tail symmetric predictive interval (L,U) by solving the following non-linear equations

$$S_1^*(L|\boldsymbol{x})=1-\frac{\sigma}{2}\quad\text{and}\quad S_1^*(U|\boldsymbol{x})=\frac{\sigma}{2}. \tag{7.7}$$

For the algorithm to obtain L and U from the above equations, we refer to Singh and Tripathi [19].

8. REAL DATA ANALYSIS

In this section, we analyze a real life data set to illustrate our established results. We consider real life data set representing the window strength in a life test. The data set was provided by Ed Fuller of the NICT Ceramics Division in December 1993. It contains polished window strength data. The data set was introduced by Pepi [16]. The data set is presented below:

18.83	20.8	21.657	23.03	23.23	24.05	24.321	25.5	25.52
25.8	26.69	26.77	26.78	27.05	27.67	29.9	31.11	33.2
33.73	33.76	33.89	34.76	35.75	35.91	36.98	37.08	37.09
39.58	44.045	45.29	45.381					

For the purpose of goodness of fit test, we consider various methods such as Bayesian information criterion (BIC), Akaikes-information criterion (AIC), the associated second-order information criterion (AICc), negative log-likelihood criterion and Kolmogorov-Smirnov (KS) statistic. Five distributions such as exponential (Exp), half-logistic (HL), inverse Weibull (InWE), Weibull (WE) and EGT-II distributions. The values of the MLEs and the five goodness of fit test statistics are presented in Table 1. It is observed that the values of test statistics corresponding to the EGT-II distribution are smaller comparing to the other distributions. Thus, it can be assumed that the given data set follows EGT-II distribution.

Next, we consider the PT-II censoring sample and two different censoring schemes (CS) as CS-I and CS-II with the failure sample size $m=20$ in Table 2. The CS-I is progressive type-II censoring and CS-II is conventional type-II censoring schemes.

In Table 3, we present the values of the proposed estimates of α,β and γ for different censoring schemes. Note that CS-III represents for the case of the complete sample. We assume $c_1=2,c_2=c_3=4,c_4=3,c_5=2$ and $c_6=4$ while computing the Bayes estimates.

Table 1: The MLE, BIC, AICc, AIC, negative log-likelihood and KS values for the real data set.

Method	Parameter		Exp	HL	InWE	WE	EGT-II
MLE	Shape	α					55.68475
		β			17.18068	4.63630	1.11743
	Scale	γ	0.03247	0.04961	0.58803	33.67241	198.992
BIC			277.9629	266.8936	260.3572	218.8458	218.6415
AICc			276.6668	265.5976	257.9178	215.4064	215.2285
AIC			276.5289	265.4596	257.4892	214.9779	214.3396
$-$InL			137.2645	131.7298	126.7446	105.4889	104.1698
KS			0.45878	0.44230	0.47472	0.15257	0.13645

Table 2: PT-II censored data and censoring schemes for the real data set.

(n, m)								
(31, 20)	x_i		18.83	20.80	21.657	24.05	24.321	25.8	26.78
			27.05	27.67	29.9	33.73	33.89	34.76	35.75
			35.91	36.98	37.08	37.09	39.58	45.381	
	Φ_i	(CS-I)	2	0	0	0	0	0	0
			0	0	2	2	0	0	0
			0	5	0	0	0	0	
		(CS-II)	0	0	0	0	0	0	0
			0	0	0	0	0	0	0
			0	0	0	0	0	11	

The Bayes estimates with respect to the ELF are computed for two distinct values of q, say -0.5 and 0.5, which are denoted by $\hat{(\cdot)}_{\text{be}}^{\text{EN}}|{-0.5}$ and $\hat{(\cdot)}_{\text{be}}^{\text{EN}}|0.5$, respectively. Further, under the squared error, weighted squared error and precautionary loss functions, the Bayes estimates are presented, which are respectively denoted by $\hat{(\cdot)}_{\text{be}}^{\text{SE}}$, $\hat{(\cdot)}_{\text{be}}^{\text{WS}}$ and $\hat{(\cdot)}_{\text{be}}^{\text{PL}}$. We use $\hat{(\cdot)}_{\text{EM}}$ and $\hat{(\cdot)}_{\text{StEM}}$ for the MLEs by the EM and StEM algorithms, respectively. The sixth column presents three different methods such as Lindley's approximation (LI), importance sampling (IS) and Metropolies-Hastings algorithm. In Table 4, the 95% various confidence and credible intervals for α, β and γ are presented. These are the asymptotic (asy) confidence intervals based on the NA to MLE and NL, the bootstrap (t and p) confidence intervals and the HPD credible intervals. Table 5 reports one-sample predictive observations and 95% predictive interval estimates of the lifetime of first two units at i-th failure. The following points can be pointed out from Tables 3, 4 and 5:

- From Table 3, we notice that the estimated values of the parameters obtained based on MH algorithm are smaller compared to that obtained using LI and IS methods. The Lindley's method provides largest Bayes estimates with respect to WSELF. For PLF, we get largest Bayes estimates when IS method is used. Under the ELF with $q = -0.5$, MH method yields largest estimates. It is also observed that the estimated values for $q = 0.5$ are always smaller than that for $q = -0.5$.

Table 3: Estimates of the parameters α, β and γ for the real data set.

(n,m) Schemes		$\hat{(\cdot)}_{\text{EM}}$ $\hat{(\cdot)}_{\text{StEM}}$		$\hat{(\cdot)}_{\text{be}}^{\text{SE}}$	$\hat{(\cdot)}_{\text{be}}^{\text{WS}}$	$\hat{(\cdot)}_{\text{be}}^{\text{PL}}$	$\hat{(\cdot)}_{\text{be}}^{\text{EN}}\vert{-}0.5$	$\hat{(\cdot)}_{\text{be}}^{\text{EN}}\vert 0.5$
$(31,20)$ (CS-I)	α	92.76551	LI	91.16425	92.34685	91.94697	91.46521	91.13645
		92.64304	MH	90.26894	90.13469	91.23567	92.2641	91.56770
			IS	93.10641	94.34077	96.11584	92.36499	91.82408
	β	0.96694	LI	0.94365	0.99643	0.94895	0.93482	0.91142
		0.99004	MH	0.90876	0.90397	0.93499	0.96315	0.94315
			IS	0.92465	0.95157	0.95708	0.93145	0.91349
	γ	147.2729	LI	144.26052	146.48510	143.05582	143.89825	143.63215
		147.5308	MH	141.68496	140.99953	142.53064	143.10289	142.94891
			IS	144.70046	147.83406	148.31648	144.32213	141.70526
$(31,20)$ (CS-II)	α	29.74175	LI	28.16496	29.03008	27.69961	27.16431	26.92482
		29.31584	MH	26.64515	26.31584	27.00948	27.67598	27.20806
			IS	28.16845	31.06412	31.47601	28.10648	26.16004
	β	0.78636	LI	0.73168	0.765461	0.74524	0.72886	0.71145
		0.79857	MH	0.68065	0.641328	0.69546	0.70094	0.69088
			IS	0.71094	0.71643	0.71948	0.68315	0.66081
	γ	66.33405	LI	65.10594	65.84694	63.40869	64.28256	63.81425
		66.94850	MH	61.16764	60.46131	63.16512	63.84247	61.23548
			IS	63.16185	67.99107	68.57093	66.91354	64.09728
$(31,31)$ (CS-III)	α	55.68475	LI	54.06889	54.76121	53.16428	53.81254	53.46942
		55.40823	MH	51.68434	51.20809	53.46849	55.64813	51.65741
			IS	53.64794	54.58215	54.61348	52.16310	50.82622
	β	1.11743	LI	1.10144	1.12465	1.10412	1.09526	1.09034
		1.13526	MH	1.07164	1.04316	1.08797	1.08964	1.05152
			IS	1.11364	1.15049	1.15942	1.11310	1.10034
	γ	198.992	LI	194.56894	197.16421	196.46852	192.84542	192.08806
		199.35641	MH	191.47964	191.08871	194.94568	195.81774	192.67880
			IS	195.39486	196.03513	196.37460	193.81345	191.10130

- Table 4 shows that among the asymptotic intervals, estimates obtained via NA method performs better than that obtained using NL method. Here, performance has been measured in terms of the length. In boot type intervals, Boot-t provides better confidence interval estimates than Boot-p method. Considering all the five methods together, it is observed that the HPD method outperforms others. Further, the lengths of the confidence and credible intervals decrease when effective sample size increases. The length of the interval estimates under CS-I is smaller than that under CS-II. Also, the lengths in the scheme CS-III is smaller compared to the other schemes. When progressive type-II censoring and type-II censoring plans are compared, the progressive type-II plan provides better result.

- From Table 5, we see that the values of the predictive estimates and prediction lengths increase as i and p increase. Further, when the effective sample size (m) increases, the predictive estimate values and predictive interval lengths decrease. The PT-II plan provides smaller length of the interval estimates compared to the type-II scheme.

Table 4: 95% confidence and credible intervals of α, β and γ for the real data set.

(n, m) Schemes	Methods		α	β	γ
$(31, 20)$ (CS-I)	Asy	NA	(87.96352, 97.39403)	(0.50153, 2.85852)	(138.13641, 159.88704)
		NL	(88.31042, 97.82276)	(0.61582, 3.12609)	(135.99003, 158.96112)
	Boot	t	(85.10587, 96.93401)	(0.68148, 3.44237)	(137.46815, 161.96911)
		p	(88.73142, 99.93547)	(0.70824, 3.64435)	(134.69740, 161.75805)
	HPD		(89.50718, 98.16640)	(0.83540, 2.71149)	(138.74876, 158.76592)
$(31, 20)$ (CS-II)	Asy	NA	(24.16824, 33.83525)	(0.46815, 3.12882)	(23.16784, 45.18696)
		NL	(25.34025, 35.07839)	(0.51672, 3.33375)	(25.43152, 48.55777)
	Boot	t	(25.76253, 36.60352)	(0.54157, 3.57718)	(22.08264, 45.74319)
		p	(23.15687, 34.76890)	(0.55481, 3.71807)	(23.56840, 47.79902)
	HPD		(25.84508, 34.94702)	(0.76109, 2.86216)	(25.94206, 47.15132)
$(31, 31)$ (CS-III)	Asy	NA	(47.16494, 55.78248)	(0.81995, 2.68102)	(189.18240, 205.44600)
		NL	(49.15205, 57.99300)	(0.86454, 2.76610)	(183.49482, 201.24983)
	Boot	t	(46.85641, 56.57569)	(0.84099, 2.79355)	(185.10064, 204.99083)
		p	(47.52215, 57.84276)	(0.84672, 2.89734)	(185.76185, 206.85296)
	HPD		(48.77806, 55.97009)	(0.92187, 2.21199)	(188.64287, 202.75298)

Table 5: One-sample predicted values and 95% prediction intervals for future observations for the real data set.

(n, m) Scheme	i	p	$\hat{(\cdot)}_{\text{be}}^{\text{SE}}$	$\hat{(\cdot)}_{\text{be}}^{\text{WS}}$	$\hat{(\cdot)}_{\text{be}}^{\text{PL}}$	$\hat{(\cdot)}_{\text{be}}^{\text{EN}}\|{-0.5}$	$\hat{(\cdot)}_{\text{be}}^{\text{EN}}\|0.5$	Interval
$(31, 20)$ (CS-I)	1	1	0.09415	0.11180	0.13157	0.11097	0.10894	(0.00769, 0.13465)
		2	0.13642	0.13758	0.14064	0.12771	0.11756	(0.02482, 0.17559)
	10	1	0.74288	0.81256	0.88157	0.80526	0.79157	(0.60408, 1.02431)
		2	0.77051	0.78109	0.78698	0.75281	0.75101	(0.66826, 1.14794)
$(31, 20)$ (CS-II)	1	1	0.13065	0.16848	0.17582	0.14033	0.11005	(0.07534, 0.26626)
		2	0.14359	0.18157	0.18278	0.18072	0.16121	(0.09077, 0.31233)
	10	1	0.70204	0.76241	0.76587	0.76112	0.71089	(0.48262, 1.08883)
		2	0.74885	0.79485	0.79948	0.74158	0.71826	(0.53170, 1.16397)
$(31, 31)$ (CS-III)	1	1	0.02465	0.03110	0.03345	0.03197	0.03128	(0.00894, 0.04656)
		2	0.06004	0.06422	0.06784	0.05682	0.04997	(0.01348, 0.09229)
	10	1	0.70909	0.71184	0.71648	0.70482	0.67158	(0.51582, 0.77149)
		2	0.72187	0.75001	0.75389	0.72807	0.72554	(0.62877, 0.91890)

Figure 1(a) presents the histogram and fitted probability density plots of five models based on real data set. From the graphs, we visualize that the EGT-II distribution covers the maximum area of the data set comparing to other distributions. The scaled total time on test (TTT) plot reveal that the hazard rate function of the fitted distribution is upside-down bathtub in Figure 1(b). The profile of the log-likelihood function of α, β and γ for real data set is shown in Figure 2(a, b, c).

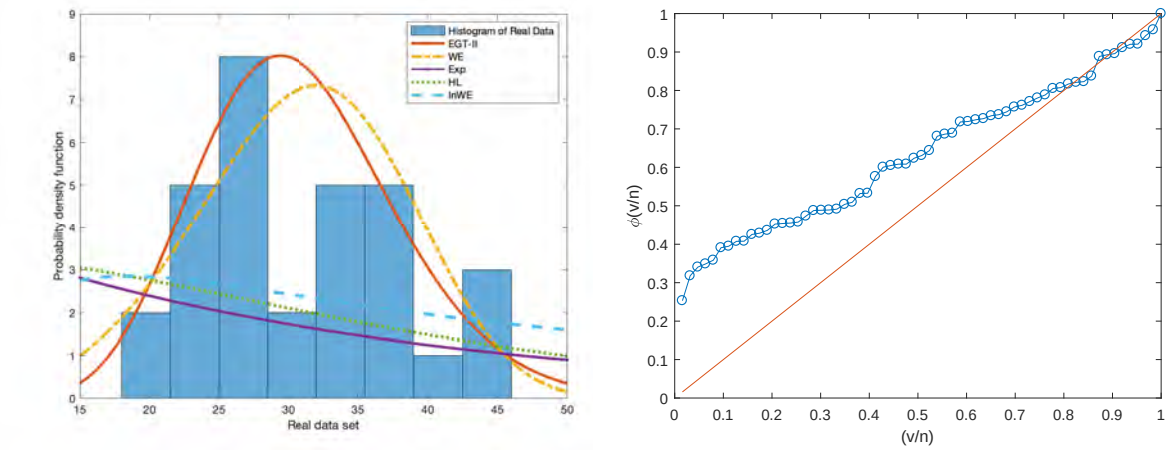


Figure 1: The first figure (in left) is the plots of the histogram and probability density functions of the fitted EGT-II, WE, Exp, HL, InWE models for the real data set. The second figure (in right) is for the scaled TTT plot.

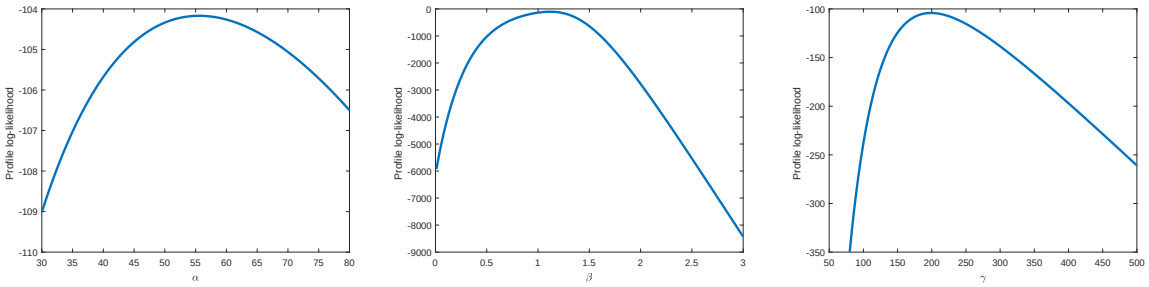


Figure 2: The profile log-likelihood plots for α (left), β (middle) and γ (right) for real data set.

9. OPTIMAL PT-II CENSORING SCHEME

In this section, we obtain optimum progressive censoring scheme from different censoring schemes for which the value of the chosen criterion is minimum. At first, we need to define a criterion. Define

$$Cr_1(\Phi) = E_{\mathrm{D}}\{V_{\mathrm{Pos}(\Phi)}(\ln T_p)\},$$

where $V_{\mathrm{Pos}(\Phi)}(\ln T_p)$ is the posterior variance of $\ln T_p$, $\Phi = (\Phi_1, ..., \Phi_m)$ is the censoring scheme and E_{D} is the expectation with respect to the data set. Further, T_p is the p-th quantile of the EGT-II distribution, which is given by

$$T_p = \left[-\left(\frac{1}{\gamma}\right)\ln\left(1-(1-p)^{\frac{1}{\alpha}}\right)\right]^{-\left(\frac{1}{\beta}\right)}. \tag{9.1}$$

Note that the total number of possible censoring schemes, given by $\binom{n-1}{m-1}$ is finite and large for fixed n and m. For example, when for $n = 30$ and $m = 20$, it is equal to 20030010, which is quite large. We say that a scheme $\Phi^{(1)} = (\Phi_1^{(1)}, ..., \Phi_m^{(1)})$ is better than another scheme $\Phi^{(2)} = (\Phi_1^{(2)}, ..., \Phi_m^{(2)})$, if $\Phi^{(1)}$ gives more information about the parameters than $\Phi^{(2)}$.

Mathematically, this is equivalent to $Cr_1(\Phi^{(1)}) < Cr_1(\Phi^{(2)})$. We refer to Kundu and Pradhan [9] and Singh and Tripathi [19] for more discussions in this direction. It is easy to see that the explicit expressions of the criterion are hard to obtain. Therefore, we use Lindley's approximation method. In the criterion, we compute approximated value of $V_{\mathrm{Pos}(\Phi)}(\ln(T_p))$. We know that

$$V_{\mathrm{Pos}(\Phi)}(\ln T_p) = E_{\mathrm{Pos}(\Phi)}[\ln T_p]^2 - (E_{\mathrm{Pos}(\Phi)}[\ln T_p])^2. \tag{9.2}$$

For simulation purpose, we generate all the parameters α, β and γ from $Gamma(5,5)$ distribution. To evaluate both terms in the RHS of (9.2), we apply Lindley's approximation method, which is explained in Section 6.1. To approximate $E_{\mathrm{Pos}(\Phi)}[\ln T_p]^2$, we have $u(\alpha,\beta,\gamma) = (\ln T_p)^2$. Further,

$$\begin{aligned}
u_1 &= -\frac{2(1-p)^{\frac{1}{\alpha}}\ln(1-p)\ln T_p}{\alpha^2\beta\gamma\Big((1-p)^{\frac{1}{\alpha}}-1\Big)\exp\{-\beta\ln T_p\}}, \quad u_2 = -\frac{2u(\alpha,\beta,\gamma)}{\beta},\\
u_3 &= \frac{2\ln T_p}{\beta\gamma}, \quad u_{12} = u_{21} = \frac{4(1-p)^{\frac{1}{\alpha}}\ln(1-p)\ln T_p}{\alpha^2\beta^2\gamma\Big((1-p)^{\frac{1}{\alpha}}-1\Big)\exp\{-\beta\ln T_p\}},\\
u_{31} &= u_{13} = -\frac{2(1-p)^{\frac{1}{\alpha}}\ln(1-p)}{\alpha^2\beta^2\gamma^2\Big((1-p)^{\frac{1}{\alpha}}-1\Big)\exp\{-\beta\ln T_p\}},\\
u_{32} &= u_{32} = \frac{4\ln T_p}{\beta^2\gamma}, \quad u_{22} = \frac{6u(\alpha,\beta,\gamma)}{\beta^2}, \quad u_{33} = \frac{2(1-\beta\ln T_p)}{\beta^2\gamma^2},\\
u_{11} &= -\frac{2(1-p)^{\frac{1}{\alpha}}\ln(1-p)}{\alpha^4\beta^2\gamma^2\Big((1-p)^{\frac{1}{\alpha}}-1\Big)^2(\exp\{-\beta\ln T_p\})^2}\\
&\quad\times\Big((1-p)^{\frac{1}{\alpha}}\Big((2\alpha\gamma\exp\{-\beta\ln T_p\})+\ln(1-p)\Big)(-\beta\ln T_p)+\ln(1-p)\Big)\\
&\quad-\beta\gamma\exp\{-\beta\ln T_p\}\Big(2\alpha+\ln(1-p)\Big)\ln T_p.
\end{aligned}$$

Other terms in (6.1) are same. In this way, $E_{\mathrm{Pos}(\Phi)}[\ln T_p]^2$ can be approximated. Similarly, to compute $E_{\mathrm{Pos}(\Phi)}[\ln T_p]$, we have $u(\alpha,\beta,\gamma) = \ln T_p$. Furthermore,

$$\begin{aligned}
u_1 &= -\frac{(1-p)^{\frac{1}{\alpha}}\ln(1-p)}{\alpha^2\beta\gamma\Big((1-p)^{\frac{1}{\alpha}}-1\Big)\exp\{-\beta\ln T_p\}}, \quad u_2 = -\frac{\ln T_p}{\beta}, \quad u_3 = \frac{1}{\beta\gamma},\\
u_{11} &= \frac{\Big((1-p)^{\frac{1}{\alpha}}\ln(1-p)-\gamma\exp\{-\beta\ln T_p\}\Big(\ln(1-p)-2\alpha\Big((1-p)^{\frac{1}{\alpha}}-1\Big)\Big)\Big)}{\alpha^4\beta\gamma^2\Big((1-p)^{\frac{1}{\alpha}}-1\Big)^2(\exp\{-\beta\ln T_p\})^2}\\
&\quad\times(1-p)^{\frac{1}{\alpha}}\ln(1-p), \quad u_{31} = u_{13} = 0,\\
u_{22} &= \frac{2\ln T_p}{\beta^2}, \quad u_{33} = -\frac{1}{\beta\gamma^2}, \quad u_{12} = u_{21} = -\frac{(1-p)^{\frac{1}{\alpha}}\ln(1-p)}{\alpha^2\beta^2\gamma\Big((1-p)^{\frac{1}{\alpha}}-1\Big)\exp\{-\beta\ln T_p\}}.
\end{aligned}$$

From Table 6, we observe that $\Phi^{(3)}$ plan gives maximum information compared to other plans, when $p = 0.25$ and $(n,m) = (25,15)$. So, plan $\Phi^{(3)}$ is optimal. Similarly, when $(n,m) = (25,15)$, $\Phi^{(1)}$ and $\Phi^{(2)}$ plans are optimal for $p = 0.5, 0.9$ and $p = 0.75$, respectively. In each censoring scheme, p increases, then the value of criterion increases. Next, for $(n,m) = (25,20)$, the plans $\Phi^{(2)}, \Phi^{(3)}, \Phi^{(1)}$ and $\Phi^{(3)}$ are optimal for $p = 0.25, 0.5, 0.75$ and $p = 0.9$, respectively.

Table 6: The values of $Cr_1(\Phi)$ for different censoring schemes Φ.

(n,m)	Φ	$(\Phi_1,...,\Phi_m)$	$p=0.25$	$p=0.5$	$p=0.75$	$p=0.9$
$(25,15)$	$\Phi^{(1)}$	(10,0,0,0,0,0,0,0,0,0,0,0,0,0,0)	0.74826	0.76445	0.90050	0.92418
	$\Phi^{(2)}$	(5,0,0,0,0,0,0,0,0,0,0,0,0,0,5)	0.85085	0.86170	0.87642	0.98054
	$\Phi^{(3)}$	(1,1,1,1,0,1,1,0,0,0,0,1,1,1,1)	0.66523	0.81064	0.91135	0.97068
	Type-II	(0,0,0,0,0,0,0,0,0,0,0,0,0,0,10)	1.09417	1.14158	1.15333	1.20081
$(25,20)$	$\Phi^{(1)}$	(5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)	0.55135	0.61053	0.67182	0.72992
	$\Phi^{(2)}$	(2,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,1)	0.39408	0.51581	0.71540	0.76204
	$\Phi^{(3)}$	(1,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,1)	0.45826	0.47643	0.69471	0.70648
	Type-II	(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,5)	0.63105	0.68846	0.72283	0.78524

10. CONCLUDING REMARKS

In this paper, we studied the problem of estimation and prediction when the lifetime data follow EGT-II distribution under the constraint that the sample is progressively censored. First, we proved that the MLEs exist and are unique. Further, it was seen that the closed form expressions of the MLEs do not exist. Thus, we used EM algorithm. The process of EM algorithm is little complicated since it requires integrations which need to be computed numerically. So, we next used stochastic version of the EM algorithm for the purpose of computation of the MLEs. In numerical study, it has been noticed that the performance of the stochastic EM algorithm is better than that of the EM algorithm. The observed Fisher's information matrix was also calculated. This is useful for obtaining the asymptotic confidence intervals. In addition, we used Boot-t and p algorithms for the computation of the confidence intervals. Bayes estimates were derived. Like the MLEs, the explicit forms of the Bayes estimates are difficult to obtain. Thus, we adopted three approximation techniques: (i) Lindley's approximation method, (ii) Importance sampling method and (iii) Metropolis-Hastings algorithm. The HPD credible intervals were also proposed. In data analysis, it was seen that the HPD credible intervals outperform others. The discussions on the elicitation of the hyper-parameters have been presented. Next, we presented the prediction problem. Here, we obtained Bayes prediction estimates and the associated Bayesian predictive interval estimates. Finally, we proposed the use of a criteria for the comparison of different sampling schemes, and then, pointed out the optimal sampling scheme for the given criterion.

A. APPENDIX

Theorem A.1. *The conditional distribution of z_{jk} for $k = 1, ..., R_j$ given $X_1 = x_1, \ ..., \ X_j = x_j$ has the form*

$$f_{Z|X}(z_j|X_1 = x_1, ..., X_j = x_j) \ = \ f_{Z|X}(z_j|X_j = x_j)$$
$$= \begin{cases} \dfrac{f(z_j : \alpha, \beta, \gamma)}{1 - F(x_j : \alpha, \beta, \gamma)}, & z_j > x_j \\ 0, & \textit{elsewhere.} \end{cases}$$

Proof: The proof is straightforward. For details, see Ng *et al.* [14]. □

Using Theorem A.1, we can write

$$E[\ln Z_{jk}|Z_{jk} > x_j, \alpha, \beta, \gamma] \ =$$
$$= \frac{\alpha\beta\gamma}{1 - F_X(x_j : \alpha, \beta, \gamma)} \int_{x_j}^{\infty} t^{-\beta-1} \exp\{-\gamma t^{-\beta}\} \left(1 - \exp\{-\gamma t^{-\beta}\}\right)^{\alpha-1} \ln t \ dt$$
$$= \frac{\alpha}{\beta\Big(1 - \exp\{-\gamma x_j^{-\beta}\}\Big)^{\alpha}} \int_{1-\exp\{-\gamma x_j^{-\beta}\}}^{0} u^{\alpha-1} \ln\left(\frac{\ln(u-1)}{\gamma}\right) du,$$

$$E[Z_{jk}^{-\beta}|Z_{jk} > x_j, \alpha, \beta, \gamma] \ =$$
$$= \frac{\alpha\beta\gamma}{1 - F_X(x_j : \alpha, \beta, \gamma)} \int_{x_j}^{\infty} t^{-2\beta-1} \exp\{-\gamma t^{-\beta}\} \left(1 - \exp\{-\gamma t^{-\beta}\}\right)^{\alpha-1} dt$$
$$= \frac{\alpha}{\gamma(1 - \exp\{-\gamma t^{-\beta}\})^{\alpha}} \int_{1-\exp\{-\gamma x_j^{-\beta}\}}^{0} u^{\alpha-1} \ln(1-u) \ du,$$

$$E[\ln(1 - \exp\{-\gamma Z_{jk}^{-\beta}\})|Z_{jk} > x_j, \alpha, \beta, \gamma] \ =$$
$$= \frac{\alpha\beta\gamma}{1 - F_X(x_j : \alpha, \beta, \gamma)} \int_{x_j}^{\infty} t^{-\beta-1} \exp\{-\gamma t^{-\beta}\} \left(1 - \exp\{-\gamma t^{-\beta}\}\right)^{\alpha-1} \ln(1 - \exp\{-\gamma t^{-\beta}\}) \ dt$$
$$= \frac{\alpha}{\Big(1 - \exp\{-\gamma x_j^{-\beta}\}\Big)^{\alpha}} \int_{0}^{1-\exp\{-\gamma x_j^{-\beta}\}} u^{\alpha-1} \ln u \ du.$$

ACKNOWLEDGMENTS

The authors like to thank the Editor-in-chief, an associate editor and two anonymous referees for careful reading and useful comments, which have led this improved version. One of the authors Kousik Maiti thanks the Ministry of Education (MoE), Government of India for the financial assistantship received to carry out this research work.

REFERENCES

[1] Arnold, B.C. and Press, S.J. (1983). Bayesian inference for Pareto populations, *Journal of Econometrics*, **21**(3), 287–306.

[2] Bdair, O.; Awwad, R.; Abufoudeh, G. and Naser, M. (2020). Estimation and prediction for flexible weibull distribution based on progressive type ii censored data, *Communications in Mathematics and Statistics*, **8**(3), 255–277.

[3] Chen, M.H. and Shao, Q.M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals, *Journal of Computational and Graphical Statistics*, **8**(1), 69–92.

[4] Dey, T.; Dey, S. and Kundu, D. (2016). On progressively type-II censored two-parameter Rayleigh distribution, *Communications in Statistics – Simulation and Computation*, **45**(2), 438–455.

[5] Efron, B. (1988). Nonconventional methods of multivariate statistical analysis, in Russian, translation from English of articles collection, *Financy i Statistika, Moscow*, **263**.

[6] Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science*, **1**(1), 54–75.

[7] Ghanbari, S.; Rezaei Roknabadi, A. and Salehi, M. (2022). Estimation of stress–strength reliability for Marshall–Olkin distributions based on progressively Type-II censored samples, *Journal of Applied Statistics*, **49**(8), 1913–1934.

[8] Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals, *The Annals of Statistics*, **16**(3), 927–953.

[9] Kundu, D. and Pradhan, B. (2009). Bayesian inference and life testing plans for generalized exponential distribution, *Science in China Series A: Mathematics*, **52**(6), 1373–1388.

[10] Langlands, A.O.; Pocock, S.J.; Kerr, G.R. and Gore, S.M. (1979). Long-term survival of patients with breast cancer: a study of the curability of the disease, *Br med J*, **2**(6200), 1247–1251.

[11] Lindley, D.V. (1980). Approximate Bayesian methods, *Trabajos de Estadistica Y de Investigacion Operativa*, **31**(1), 223–245.

[12] Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 226–233.

[13] Maiti, K. and Kayal, S. (2019). Estimation for the generalized Fréchet distribution under progressive censoring scheme, *International Journal of System Assurance Engineering and Management*, **10**(5), 1276–1301.

[14] Ng, H.; Chan, P. and Balakrishnan, N. (2002). Estimation of parameters from progressively censored data using EM algorithm, *Computational Statistics & Data Analysis*, **39**(4), 371–386.

[15] Nielsen, S.F. *et al.* (2000). The stochastic EM algorithm: estimation and asymptotic results, *Bernoulli*, **6**(3), 457–489.

[16] Pepi, J.W. (1994). Fail-safe design of an all BK-7 glass aircraft window, *Window and Dome Technologies and Materials IV*, Vol. 2286, International Society for Optics and Photonics, pp. 431–443.

[17] Ren, J. and Gui, W. (2021). Goodness-of-fit test for Rayleigh distribution based on progressively type-II censored sample, *Communications in Statistics – Theory and Methods*, **50**(16), 3851–3874.

[18] Singh, S.; Belaghi, R.A. and Asl, M.N. (2019). Estimation and prediction using classical and Bayesian approaches for Burr III model under progressive type-I hybrid censoring, *International Journal of System Assurance Engineering and Management*, **10**(4), 746–764.

[19] Singh, S. and Tripathi, Y.M. (2018). Estimating the parameters of an inverse Weibull distribution under progressive type-I interval censoring, *Statistical Papers*, **59**(1), 21–56.

[20] Tarvirdizade, B. and Nematollahi, N. (2021). Inference for the power-exponential hazard rate distribution under progressive type-II censored data, *Journal of Statistics and Management Systems*, **24**(6), 1169–1212.

[21] Tregouet, D.; Escolano, S.; Tiret, L.; Mallet, A. and Golmard, J. (2004). A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm, *Annals of Human Genetics*, **68**(2), 165–177.

REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 535–556
https://doi.org/10.57805/revstat.v21i4.429

New Members of the Johnson Family of Probability Distributions: Properties and Application

Author: Piotr Sulewski
– Institute of Exact and Technical Sciences, Pomeranian University, Poland
piotr.sulewski@apsl.edu.pl

Received: June 2021 Revised: February 2022 Accepted: February 2022

Abstract:

- In this paper we introduce new non-mixed bimodal distributions belonging to the Johnson family of distributions (JFD) named SC and SD, where SC is a special case of SD. The SD is the only distribution among the JFD that can be both unimodal and bimodal distribution. Properties of the SD are methodically studied. The SC and SD are compared with the seven (except the normal distribution) members of the JFD for flexibility and applicability. In order to test for flexibility, a special measure called skewness-kurtosis-square is defined. The best dispersion of points with coordinates (skewness, kurtozis) occurs for the SD and for the well-known SU and SB. Two real datasets were used to test the applicability. EN turned out to be better than its competitors in terms of information criteria and results of three goodness-of-fit tests.

Keywords:

- *normal distribution; flexibility of distribution; departure from normality.*

AMS Subject Classification:

- 60E05, 65C20.

1. INTRODUCTION

Routinely statisticians start on analysis of data estimating PDF (commonly called histogram) or plotting empirical CDF on an appropriate probability paper. It depends on the sample size they possess. It may happen that a solid evidence emerges from the obtained figures suggesting bimodality of the population distribution as it is shown in Figure 1.

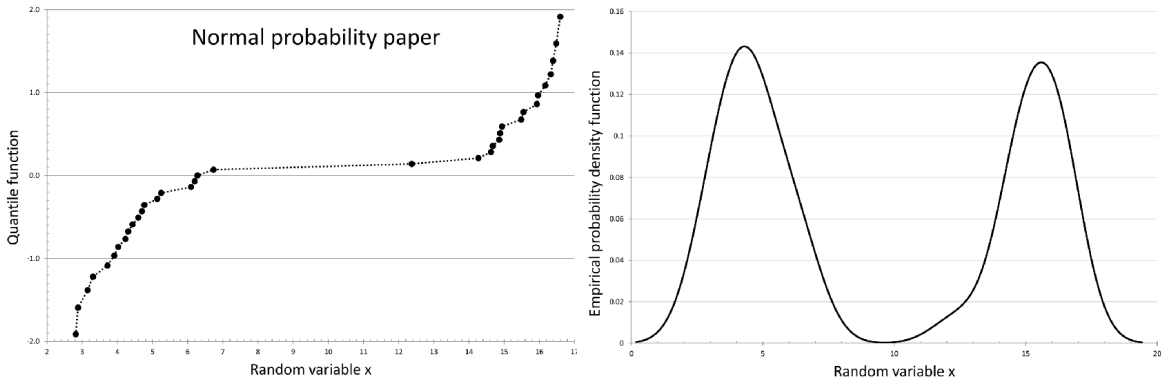


Figure 1: Bimodality of the population distribution.

In such a situation the statisticians as a rule employ the mixed (also called compound) theoretical distribution that has a general form:

$$f_m(x;\omega,\Theta_1,\sigma_1,\Theta_2,\sigma_2) = \omega f_1(x;\Theta_1,\sigma_1) + (1-\omega) f_2(x;\Theta_2,\sigma_2). \tag{1.1}$$

Let us denote such a distribution as mixed bimodal distribution (MBD). In (1.1) ω is the fraction parameter whereas Θ_1, σ_1 and Θ_2, σ_2 are pairs of location — scale or scale — shape parameters depending on sorts of distributions being mixed.

The MBD can be made bimodal and fitted to data of the sort we say about. However, employing MBD the statisticians unambiguously state that the population is nonhomogeneous. Wide applicability of the MBD comes from its clarity and interpretability of parameters. Nevertheless, it is hard to believe that non-homogeneity is a sole cause of distribution bimodality. It is hard to believe because many factors other than intentional or unintentional mixing sample items play a role in shaping population distribution. Thus, we see that a vital necessity arises to develop non-mixed bimodal distribution (nMBD) arises. What makes our task more difficult is that parameters of such distribution should be relatively clearly interpretable. In order that nMBD be a worthy challenger to MBD. However, we are fully aware that the parameters in question will never be so clearly interpretable as parameters of the MBD are, which is easy to explain. The MBD comes into being due to one factor, which causes mixing in particular proportion items belonging to two different subpopulations. In contrast nMBD comes into being due to many factors. And effects of "activity" of all these many factors have to be expressed also by means of only five parameters as estimators of parameters consume information.

Let us recall the Johnson family of probability distributions (JFD). All the distributions that belong to JFD have the following general form

$$F(x) = \Phi\left[c + \rho\varphi\left(\frac{x-a}{b}\right); 0, 1\right], \tag{1.2}$$

where $\varphi(x)$ can be any non-decreasing function of x, $\Phi(x; u, v)$ is CDF of $N(u, v)$. Literature related to JFD is numerous (see e.g. [14], [15], [3]).

Before defining the new members of the JFD, it is worth taking a look at the distributions that belong to this family.

The normal (N) distribution with location parameter $a \in R$ and scale parameter $b > 0$ is defined as

$$F_N(x; a, b) = \Phi\left[\varphi\left(\frac{x-a}{b}\right); 0, 1\right]. \tag{1.3}$$

We obtain (1.3) from (1.2) by considering $c = 0$, $\delta = 1$, $\varphi(y) = y$ and $y = (x-a)/b$.

The Birnbaum–Saunders (BS) distribution with location parameter $a \in R$, scale parameter $b > 0$ and shape parameter $\alpha > 0$, is defined as [7]

$$F_{BS}(x; \alpha, a, b) = \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{x-a}{b}} - \sqrt{\frac{b}{x-a}}\right); 0, 1\right] \quad (x > a). \tag{1.4}$$

We obtain (1.4) from (1.2) by considering $c = 0$, $\delta = 1/\alpha$, $\varphi(y) = \sqrt{y} - \sqrt{1/y}$ and $y = (x-a)/b$.

The generalization of the Birnbaum–Saunders (GBS) distribution with location parameter $a \in R$, scale parameter $b > 0$ and shape parameters $\alpha > 0$, $\beta > 0$, is defined as [18]

$$F_{GBS}(x; \alpha, a, b, \beta) = \Phi\left[\frac{1}{\alpha}\left(\left(\frac{x-a}{\beta}\right)^{\beta} - \left(\frac{\beta}{x-a}\right)^{\beta}\right); 0, 1\right] \quad (x > a). \tag{1.5}$$

The BS distribution is a special case of the GBS distribution for $\beta = 0.5$. We obtain (1.5) from (1.2) by considering $c = 0$, $\delta = 1/\alpha$, $\varphi(y) = y^{\beta} - y^{-\beta}$ and $y = (x-a)/b$.

The Four-Parameter BS (FBS) distribution with location parameter $a \in R$, scale parameter $b > 0$, shape parameter $\delta > 0$ and non-centrality parameter $c \in R$, is given by [2]

$$F_{FBS}(x; c, \delta, a, b) = \Phi\left[c + \delta\left(\sqrt{\frac{x-a}{b}} - \sqrt{\frac{b}{x-a}}\right); 0, 1\right] \quad (x > a). \tag{1.6}$$

Formula (1.4) is a special case of (1.6) for $c = 0$, $\delta = 1/\alpha$. We obtain (1.6) from (1.2) by considering $\varphi(y) = \sqrt{y} - \sqrt{1/y}$ and $y = (x-a)/b$.

The sinh-normal (SN) distribution with the location parameter $a \in R$, the scale parameter $b > 0$ and the shape parameter $\alpha > 0$, is given by [19]

$$F_{SN}(x; \alpha, a, b) = \Phi\left[\frac{2}{\alpha}\sinh\left(\frac{x-a}{b}\right); 0, 1\right]. \tag{1.7}$$

This distribution is symmetric about the location parameter $a \in R$. We obtain (1.7) from (1.2) by considering $c = 0$, $\delta = 2/\alpha$, $\varphi(y) = \sinh(y)$ and $y = (x - a)/b$.

The lognormal or SL distribution with location parameter $a \in R$, scale parameter $b > 0$ and shape parameters $c_1 \in R$, $\delta > 0$, is defined as [14]

$$F_{SL}(x; c_1, \delta, a, b) = \Phi\left[c_1 + \delta \ln\left(\frac{x-a}{b}\right); 0, 1\right] \quad (x > a). \tag{1.8}$$

Formula (1.8) can be written using three parameters, namely:

$$F_{SL}(x; c_1, \delta, a, b) = \Phi[c + \delta \ln(x - a); 0, 1] \quad (x > a), \tag{1.9}$$

where $c = c_1 - \delta \ln(b)$, $c \in R$. We obtain (1.9) from (1.2) by considering $b = 1$, $\varphi(y) = \ln(y)$ and $y = (x - a)$. Please notice that the lognormal distribution with the CDF [11] $\breve{F}_{SL}(x; e_1, e_2) = \Phi\left[\frac{\ln(x)-e_1}{e_2}; 0, 1\right]$ $(x > 0)$ widely used in practice can be treated as a special case of (1.8) when $a = 0$, $\delta = 1/e_2$ and $c = -e_1/e_2$.

The SB distribution with the location parameter $a \in R$, the scale parameter $b > 0$ and the shape parameters $c \in R$, $\delta > 0$, is defined as [14]

$$F_{SB}(x; c, \delta, a, b) = \Phi\left[c + \delta \ln\left(\frac{x-a}{b+a-x}\right); 0, 1\right] \quad (a < x < a + b). \tag{1.10}$$

We obtain (1.10) from (1.2) by considering $\varphi(y) = \ln(y) - \ln(1 - y)$ and $y = (x - a)/b$. Let $a = 0$, $b = 1$, $\delta = 1/e_2$ and $c = -e_1/e_2$, then we obtain the special case of (1.10) widely used in practice defined as

$$\breve{F}_{SB}(x; e_1, e_2) = \Phi\left[\frac{\ln\left(\frac{x}{1-x}\right) - e_1}{e_2}; 0, 1\right].$$

The SU distribution with the location parameter $a \in R$, the scale parameter $b > 0$ and the shape parameters $c \in R$, $\delta > 0$, is defined as [14]

$$F_{SU}(x; c, \delta, a, b) = \Phi\left[c + \delta \operatorname{asinh}\left(\frac{x-a}{b}\right); 0, 1\right]. \tag{1.11}$$

We obtain (1.11) from (1.2) by considering $\varphi(y) = \operatorname{asinh}(y)$ and $y = (x - a)/b$.

This paper introduces two new members of the JFD, namely SC and SD. In the SU distribution Johnson employed $\operatorname{asinh}(x) = \ln\left(x + \sqrt{1 + x^2}\right)$. In the SC and SD distributions we will employ $\sinh(x) = \frac{\exp(x)-\exp(-x)}{2}$.

The SC distribution with location parameter $a \in R$, the scale parameter $b > 0$ and shape parameter $c \in R$, is defined as

$$F_{SC}(x; c, a, b) = \Phi\left[c + 2 \sinh\left(\frac{x-a}{b}\right); 0, 1\right]. \tag{1.12}$$

Please notice that ρ parameter appearing in (1.2) has been in (1.12) replaced with a constant equal to 2. This constant compensates denominator in definition of the $\sinh(x)$ function. For $c = 0$ in (1.12) and $\alpha = 1$ in (1.7), the SN distribution is equivalent to the SC distribution. We obtain (1.12) from (1.2) by considering $\delta = 2$, $\varphi(y) = \sinh(y)$ and $y = (x - a)/b$.

The SD distribution with multipurpose parameters $a_1, a_2 \in R$, $b_1, b_2 > 0$ and semi-fraction parameter $c > 0$ (see Figure 2), is defined as

$$F_{SD}(x; c, a_1, b_1, a_2, b_2) = \Phi\left[c - \exp\left(\frac{a_1 - x}{b_1}\right) + \exp\left(\frac{x - a_2}{b_2}\right); 0, 1\right]. \tag{1.13}$$

The SD distribution is obtained from SC by adding the second exponential function. The CDF $F_{SD}(x; c, a_1, b_1, a_1, b_1)$ is equal to the CDF $F_{SC}(x; c, a_1, b_1)$, so the SC is a special case of the SD.

Although SD cannot be acknowledged as a special case of (1.2), it seems reasonable to treat SD as a member of the JFD, which can be justified by appearing as a generalization or extension of an element of the SN distribution family obviously belonging to the JFD family. The mentioned element is the SC distribution. The F_{SD} involves two exponential components that can be independently movable on the x axis. Owing to this we are able to obtain bimodal distribution provided we locate the components sufficiently far from one another on the x axis as it is exemplified in Figure 2 (left). This figure shows examples of CDFs of the SD distribution plotted on the Normal probability paper. The reader is prompted to compare Figure 2 (left) with Figure 1 (left). Figure 2 (right) shows examples of PDFs of the SD distribution with exemplifying a role of c parameter. No doubt, c parameter can be called the semi-fraction parameter.

Let

$$P(x; c, a_1, b_1, a_2, b_2) = c - \exp\left(\frac{a_1 - x}{b_1}\right) + \exp\left(\frac{x - a_2}{b_2}\right).$$

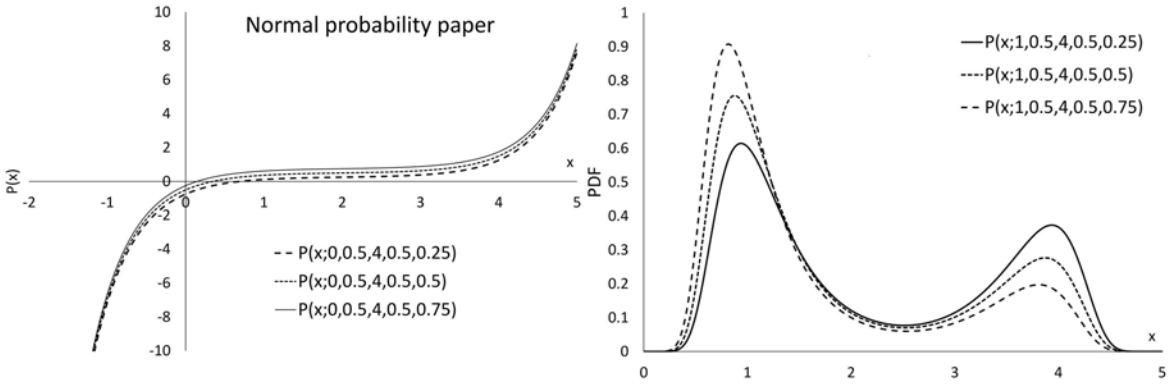


Figure 2: Bimodality in the SD distribution.

If we subject a random variable to a linear transformation, the skewness and kurtosis retain their values. This fact was also confirmed by a simulation study. To simplify the study of the skewness and kurtosis of the SD distribution, let us standardize a random variable x: $z = \frac{x - a_1}{b_1} \Rightarrow x = b_1 z + a_1$. As a result of simple transformation the CDF (1.13) has the form

$$F_{SD}(x; c, a_0, b_0) = \Phi\left[c - \exp(-z) + \exp\left(\frac{x - a_0}{b_0}\right); 0, 1\right],$$

where $a_0 = \frac{a_2 - a_1}{b_1}$, $b_0 = \frac{b_2}{b_1}$.

Let (γ_1, γ_2) be coordinate of a point described by skewness and kurtosis, respectively. The normal distribution (ND) is characterized by only one point $(0, 3)$, obviously. For every distribution from the JFD (except the ND), values of (γ_1, γ_2) are calculated for 10^4 randomly determined values of parameters influencing skewness and kurtosis in the Malakhov area (MA) $10 \geq \gamma_2 \geq \gamma_1^2 + 1$ (Table 1). In the MA $\gamma_2 \in [1, 10]$, then $\gamma_1 \in [-3, 3]$ (see Figure 3). The parameter value ranges are selected to maximize MA filling according to the SKS measure (1.14). To make the calculations more reliable (without the so called outliers) the normalization conditions were checked.

Table 1: Ranges of parameter values influencing skewness and kurtosis as well as skewness and kurtosis values for the JDF in the Malakhov area $10 \geq \gamma_2 \geq \gamma_1^2 + 1$.

JFD	Parameter ranges	Skewness range	Kurtosis range
BS	$\alpha \in (0, 0.46)$	$\gamma_1 \in (0, 2.22)$	$\gamma_2 \in (3, 5.92)$
GBS	$\alpha \in (0, 7),\ \beta \in (0, 1.75)$	$\gamma_1 \in (0.07, 2.38)$	$\gamma_2 \in (2.03, 10)$
FBS	$c \in (-6.25, 6.25),\ \delta \in (0, 2.75)$	$\gamma_1 \in (0.44, 2.23)$	$\gamma_2 \in (3.27, 10)$
SN	$\alpha \in (0.1, 180.4)$	$\gamma_1 = 0$	$\gamma_2 \in (1.15, 3)$
SL	$\delta \in (1.88, 100)$	$\gamma_1 \in (0.03, 1.89)$	$\gamma_2 \in (3, 9.98)$
SB	$c \in (-3.35, 3.39),\ \delta \in (0.1, 1.2)$	$\gamma_1 \in (-2.84, 2.91)$	$\gamma_2 \in (1.13, 10)$
SU	$c \in (-2.05, 2.05),\ \delta \in (1.31, 1.9)$	$\gamma_1 \in (-1.8, 1.79)$	$\gamma_2 \in (4.76, 10)$
SC	$c \in (-89.94, 89.97)$	$\gamma_1 \in (-0.69, 0.69)$	$\gamma_2 \in (2.52, 3.90)$
SD	$c \in (-4.1, 4.1),\ a_0 \in (-4.3, 4.3),$ $b_0 \in (0.1, 0.9)$	$\gamma_1 \in (-2.79, 2.46)$	$\gamma_2 \in (1.26, 10)$

Figure 3 presents sets of points (γ_1, γ_2) and the MP $\gamma_2 = \gamma_1^2 + 1$ in the MA $10 \geq \gamma_2 \geq \gamma_1^2 + 1$ related to the SB, SU, SC, SD distributions. The SD and SU distributions are the best filling the MA. The SD distribution has common areas of skewness and kurtosis with the SB and SU distributions. Sets of points (γ_1, γ_2) and the MP $\gamma_2 = \gamma_1^2 + 1$ in the MA $10 \geq \gamma_2 \geq \gamma_1^2 + 1$ related to the BS, GBS, FBS, SL distributions are presented in the supplementary material.

In addition to visual assessment, the skewness-kurtosis-square (SKS) measure [22] is used to compare the flexibility of distributions. Colored circles of diameter and coordinates of their centers determined by skewness γ_1 and kurtosis γ_2 are placed within the MA that is described by inequality $\gamma_2 \geq \gamma_1^2 + 1$ [17]. Then colored area fraction is calculated. Squares of sides equal to η seem a reasonable alternative to circles since they simplify calculation of the total colored area. Obviously, when some squares overlap, only one is taken into account. The SKS measure is given by [22]

$$SKS = \frac{SI}{ST}, \tag{1.14}$$

where ST denotes a total number of squares within the MA, SI — a number of squares to which the point (γ_1, γ_2) has fallen. The SKS measure takes values in $[0, 1]$. The maximum value denotes a perfect dispersal of points (γ_1, γ_2) in the MA. The R codes for calculating the SKS measure are presented in the supplementary material.

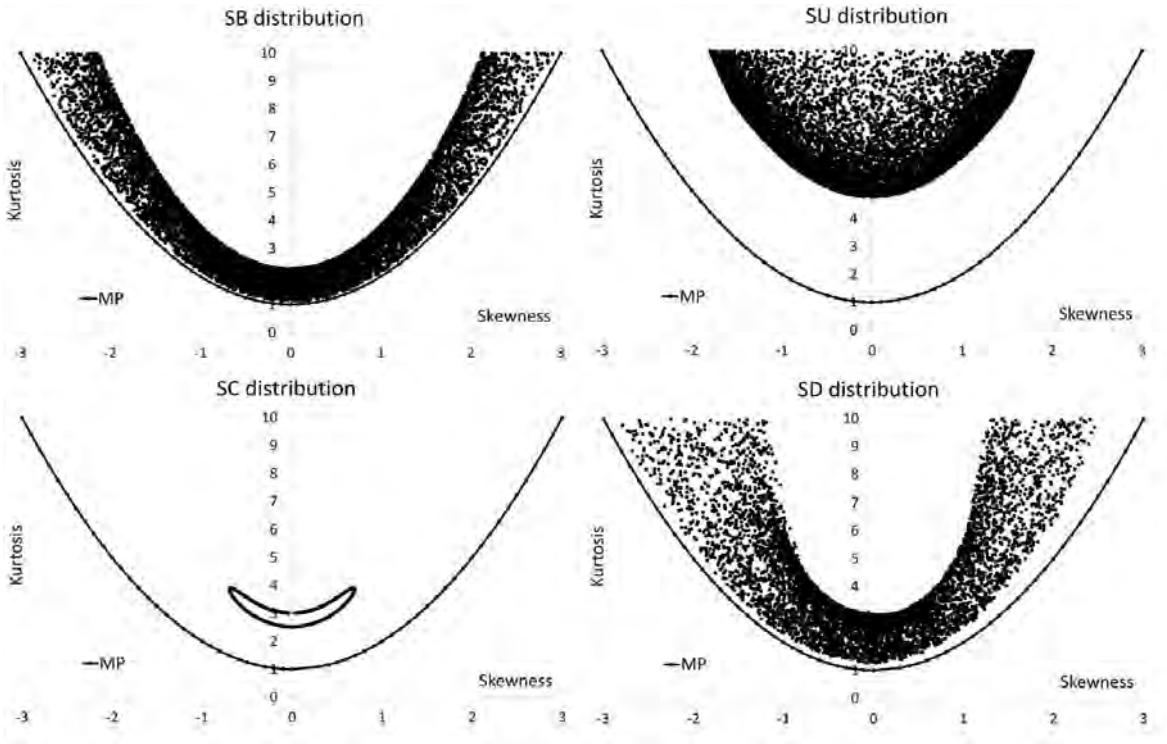


Figure 3: Skewness and kurtosis for the SB, SU, SC, SD distributions.

Table 2 presents values of SKS measures (1.14) obtained for square side $\eta = 0.05, 0.1, 0.15, 0.20$. The best dispersion of points (γ_1, γ_2), taking into account the accuracy expressed by η, occurs for the SD, SU and SB distribution (see bold).

Table 2: SKS measure values for JFD in the MA $10 \geq \gamma_2 \geq \gamma_1^2 + 1$ for square side η.

JFD	$\eta = 0.05$	$\eta = 0.1$	$\eta = 0.15$	$\eta = 0.2$
BS	0.0070	0.0140	0.0203	0.0271
GBS	0.0778	0.1051	0.1277	0.1333
FBS	0.0561	0.0694	0.0794	0.0906
SL	0.0111	0.0237	0.0346	0.0458
SB	**0.2410**	0.3401	0.3813	0.4052
SU	0.2375	0.3433	0.3693	0.3740
SC	0.0092	0.0175	0.0274	0.0323
SD	0.2185	**0.4102**	**0.4994**	**0.5469**

New distributions, modelled on the SL, SB, SU distributions, was named as SC and SD distributions. The SC is a special cases of the SD, so the remainder of the paper is devoted to the SD distribution. The lognormal distribution is defined with the log function and the SD distribution is defined with the exp function, therefore the SD distribution is also called the expnormal (EN) distribution.

This paper is organized as follows. Section 2 presents properties of the SD distribution. The unknown parameters are estimated in Section 3 and entropies are calculated in Section 4. Examples are presented in Section 5. Section 6 deals with conclusions. Due to the size of the paper, the selected figures and tables as well as the main R codes have been transferred to the supplementary material.

2. MAIN PROPERTIES OF INTRODUCED DISTRIBUTION

2.1. Distribution and density function

Definition 2.1. The distribution of the random variable X with PDF given by

$$f(x;a_1,b_1,a_2,b_2,c)=\left(\frac{1}{b_1}e^{-z_1(x)}+\frac{1}{b_2}e^{z_2(x)}\right)\phi[c-\exp(-z_1(x))+\exp(z_2(x));0,1], \tag{2.1}$$

where $\phi(x;u,v)$ is PDF of $N(u,v)$, $z_1(x)=\frac{x-a_1}{b_1}$ and $z_2(x)=\frac{x-a_2}{b_2}$, is called the expnormal (EN) distribution. In (2.1) $a_1,a_2\in R$ are position parameters, $b_1,b_2>0$ are scale parameters and $c\in R$ is the semi-fraction parameter (see Figure 2). For these parameter values, the main argument of ϕ in (2.1) is an increasing function, hence

$$\int_{-\infty}^{+\infty} f(x;a_1,b_1,a_2,b_2,c)=1.$$

PDF of the EN distribution is calculated using the R function dEN (see supplementary material).

If $a_1=a_2,\ b_1=b_2,\ c=0$, then $EN(a_1,b_1,a_2,b_2,c)$ is very similar to the $N\left(a_1,\frac{b_1}{2}\right)$. According to the similarity measure between two distributions defined in [23], we have for $a_1\in R$, $b_1>0$.

$$\int_{-\infty}^{+\infty}\min\left[f(x;a_1,b_1.a_1,b_1,0),\phi\left(x;a_1,\frac{b_1}{2}\right)\right]=0.966. \tag{2.2}$$

Thus the $EN(0,2,0,2,0)$ is similar to the $N(0,1)$ in 96.6%. The distribution with multipurpose parameter $a_1,\ b_1,\ a_2,\ b_2=b_1$ is symmetrical for $c=0$ (see Table 4 and Figure 4, series D1,D2). If $X\sim EN(a_1,\ b_1,\ a_2,\ b_2=b_1,\ c=0)$ then $E(X)=\frac{a_1+a_2}{2}$. In this case the modes are at the same height. The mean value formula is also confirmed by numerical methods. The $EN(a_1,b_1,a_2,b_2,\ c>0)$ is positively skewed (Figure 4, series A1, A2, E1, E2) and the $EN(a_1,b_1,a_2,b_2,\ c\le 0)$ is negatively skewed (Figure 4, series B1, B2, F1, F2). The EN distribution can be unimodal (Figure 4, series A1, B1, D1, E1, F1) and bimodal (Figure 4, series A2, B2, D2, E2, F2). See Table 4 for more information.

Table 3 presents the division of distributions by their skewness and excess kurtosis [22]. The ND obviously does not belong to this family. Selecting appropriate parameter values of the EN distribution, we can obtain skewness and excess kurtosis values belonging to the analyzed groups A1–B2 and D1–F2 (Table 4).

Table 3: Groups of distributions according to their skewness and excess kurtosis [23]. Denote: * unimodal distribution, ** bimodal distribution.

Group	Skewness	Ex. kurtosis	Group	Skewness	Ex. kurtosis
A1*	positive	positive	D1*	zero	negative
A2**	positive	positive	D2**	zero	negative
B1*	negative	positive	E1*	positive	negative
B2**	negative	positive	E2**	positive	negative
C1*	zero	positive	F1*	negative	negative
C2**	zero	positive	F2**	negative	negative

Table 4: The $EN(a_1, b_1, a_2, b_1, 0)$ distribution with parameter values for groups A1–B2 and D1–F2.

a_1	b_1	a_2	b_2	c	Skewness	Ex. kurtosis	Group
0	1	1	1.25	1	0.740	0.268	A1
−1	1	3	1	1	1.239	0.608	A2
1	2	0	1	0	−0.527	0.151	B1
−4	0.5	1	1	−1	−1.298	0.334	B2
0	2	0	2	0	0	−0.479	D1
0	0.5	1	0.5	0	0	−1.024	D2
0	1	1	1	1	0.584	−0.13	E1
−1	1	3	1	0.5	0.601	−0.961	E2
0	1	1	1	−1	−0.584	−0.13	F1
−1	1	3	1	−0.5	−0.601	−0.961	F2

Figure 4 plots the PDF of the $EN(a_1, b_1, a_2, b_2, c)$ for groups of parameters presented in Table 4.

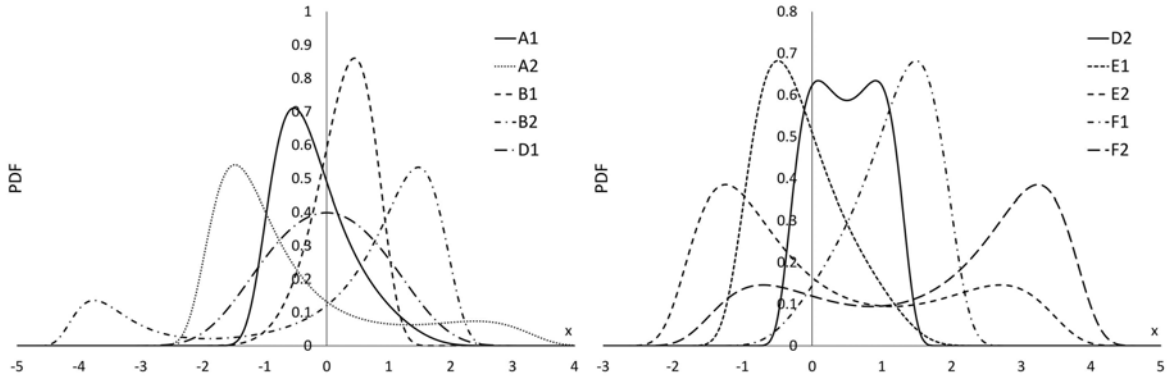


Figure 4: PDF of the $EN(a_1, b_1, a_2, b_1, 0)$ for groups from Table 4.

Theorem 2.1. *Let* $X \sim EN(a_1, b_1, a_2, b_2, c)$*, then the CDF of* X *is given by*

$$F(x; a_1, b_1, a_2, b_2, c) = \Phi\left[c - \exp\left(-\frac{x - a_1}{b_1}\right) + \exp\left(\frac{x - a_2}{b_2}\right); 0, 1\right]. \tag{2.3}$$

Proof: Obtaining (2.3) based on (2.1) is trivial. □

CDF of the EN distribution is calculated using the R function pEN (see supplementary material).

Figure 5 (left) plots the CDF of the $EN(a_1, b_1, a_2, b_2, c)$ for groups A1, A2, B1, B2. The CDF of the $EN(a_1, b_1 > 0, a_2, b_2 > 0, c)$ on the normal Q-Q plot is monotonically increasing curve (Figure 5, right).

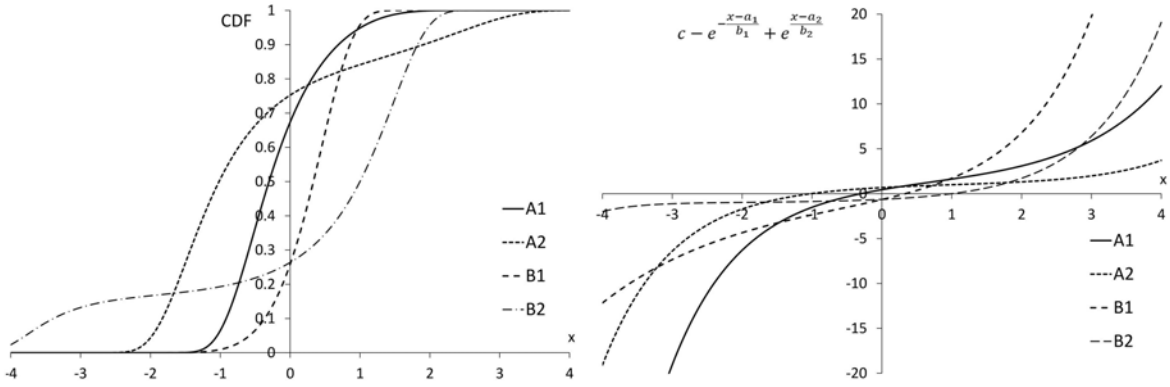


Figure 5: CDF of the $EN(a_1, b_1, a_2, b_1, 0)$ (left) and the normal Q-Q plot (right).

Theorem 2.2. *The $EN(a_1, b_1, a_2, b_2, c)$ with the PDF given by (14) is identifiable in a parameter space $v = (a_1, b_1, a_2, b_2, c)$.*

Proof: Let $v_1 = (a_{11}, b_{11}, a_{21}, b_{21}, c_1)$ and $v_2 = (a_{12}, b_{12}, a_{22}, b_{22}, c_2)$. Let us suppose that $f_{v_1}(x) = f_{v_2}(x)$ for all x. This condition based on (2.3) implies that

$$\Phi\left[c_1 - \exp\left(-\frac{x-a_{11}}{b_{11}}\right) + \exp\left(\frac{x-a_{21}}{b_{21}}\right); 0, 1\right] =$$
$$= \Phi\left[c_2 - \exp\left(-\frac{x-a_{12}}{b_{12}}\right) + \exp\left(\frac{x-a_{22}}{b_{22}}\right); 0, 1\right].$$

The function Φ is an increasing function which implies that

$$c_1 - \exp\left(-\frac{x-a_{11}}{b_{11}}\right) + \exp\left(\frac{x-a_{21}}{b_{21}}\right) = c_2 - \exp\left(-\frac{x-a_{12}}{b_{12}}\right) + \exp\left(\frac{x-a_{22}}{b_{22}}\right)$$

or

$$c_1 - c_2 + \exp\left(-\frac{x-a_{12}}{b_{12}}\right) - \exp\left(-\frac{x-a_{11}}{b_{11}}\right) + \exp\left(\frac{x-a_{21}}{b_{21}}\right) - \exp\left(\frac{x-a_{22}}{b_{22}}\right) = 0.$$

As a result of simple transformation $a_{11} = a_{12}$, $b_{11} = b_{12}$, $a_{21} = a_{22}$, $b_{21} = b_{22}$, $c_1 = c_2$. □

2.2. Hazard function

Proposition 2.1. *Let $X \sim EN(a_1, b_1, a_2, b_2, c)$. The hazard function associated with the EN distribution is*

$$h(x) = \frac{\left(\frac{1}{b_1}e^{-\frac{x-a_1}{b_1}} + \frac{1}{b_2}e^{\frac{x-a_2}{b_2}}\right)\phi\left[c - \exp\left(-\frac{x-a_1}{b_1}\right) + \exp\left(\frac{x-a_2}{b_2}\right); 0, 1\right]}{1 - \Phi\left[c - \exp\left(-\frac{x-a_1}{b_1}\right) + \exp\left(\frac{x-a_2}{b_2}\right); 0, 1\right]}. \tag{2.4}$$

The limits of the EN hazard function as $x \to -\infty$ and $x \to \infty$ are respectively 0 and ∞ (Figure 6).

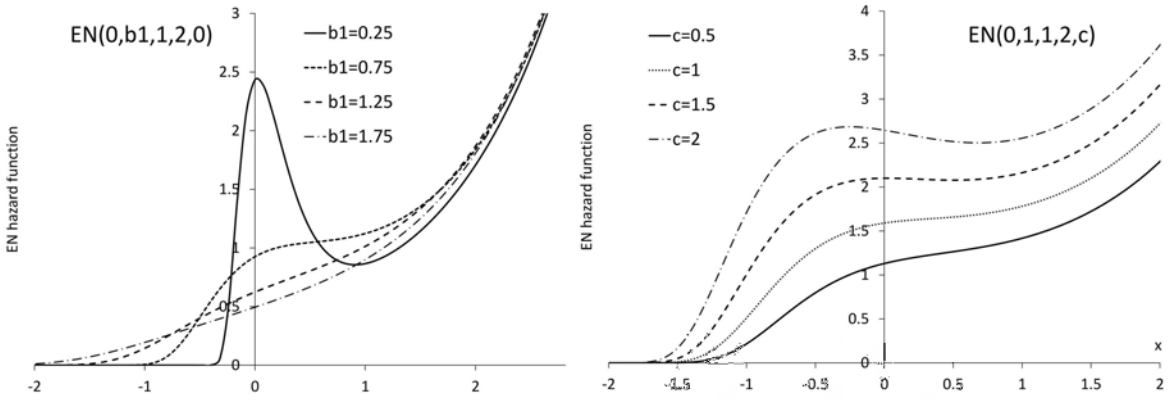


Figure 6: The EN hazard function for various values of parameters.

2.3. Quantiles

Proposition 2.2. *Let* $X \sim EN(a_1, b_1, a_2, b_2, c)$. *The p-th* $(0 < p < 1)$ *quantiles are the solution of the following equation*

$$c - \exp\left(-\frac{x_p - a_1}{b_1}\right) + \exp\left(\frac{x_p - a_2}{b_2}\right) - \Phi^{-1}(p) = 0.$$

The value of x_p is obtained by the numerical method, e.g. using the R software. Quantile function of the EN distribution is calculated using the R function qEN (see supplementary material).

2.4. Moments and moment generating function

Proposition 2.3. *Let* $X \sim EN(a_1, b_1, a_2, b_2, c)$. *The k-th,* $k \in Z$ *non-central moments from (14) are given by*

$$\alpha_k = \int_{-\infty}^{+\infty} x^k \left(\frac{1}{b_1} e^{-z_1} + \frac{1}{b_2} e^{z_2}\right) \phi[c - \exp(-z_1) + \exp(z_2); 0, 1], \tag{2.5}$$

where $z_1 = \frac{x-a_1}{b_1}$ *and* $z_2 = \frac{x-a_2}{b_2}$, $\phi(x; a, b)$ *is PDF of* $N(a, b)$

Thus the variance μ_2, skewness γ_1 and kurtosis γ_2 of the EN distribution are defined as

$$\mu_2 = \alpha_2 - \alpha_1^2, \quad \gamma_1 = \frac{\alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3}{\mu_2^{1.5}}, \quad \gamma_2 = \frac{\alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4}{\mu_2^2}.$$

Table 5 provides the mode x_{mod}, mean α_1, variance μ_2, skewness γ_1 and kurtosis γ_2 of the EN distribution for various parameter combinations.

Table 5: Mode, mean, variance, skewness and kurtosis of the $EN(a_1, b_1, a_2, b_2, c)$.

a_1	b_1	a_2	b_2	**c**	x_{mod}	α_1	μ_2	γ_1	γ_2
0	2	0	2	0	0	0	0.837	0	2.521
1					0.5	0.5	0.538	0	2.643
2					1	1	0.34	0	2.745
3					1.5	1.5	0.212	0	2.826
0	0.5	1	1	0	−0.005	0.444	0.334	0.194	2.532
	1				0.5	0.5	0.474	0	2.245
	1.5				0.95	0.52	0.622	−0.216	2.455
	2				1.077	0.523	0.775	−0.469	2.812
−2	1	−2	1	0	−2	−2	0.209	0	2.521
		−1			−1.5	−1.5	0.474	0	2.245
		1			−1.946,0.946	−0.5	1.791	0	1.751
		2			1.981	0	3.009	0	1.578
0	0.5	0.5	0.25	1	−0.215	−0.025	0.096	0.019	2.156
			0.5		−0.245	−0.08	0.096	0.056	2.87
			0.75		−0.281	−0.114	0.101	0.089	3.686
			1		−0.308	−0.137	0.107	0.12	4.583
0	1	1	2	0.5	−0.377	0.072	0.597	0.534	3.594
				1	−0.615	−0.274	0.427	0.48	4.583
				1.5	−0.815	−0.561	0.293	0.367	5.503
				2	−0.988	−0.798	0.199	0.252	6.006
−2	2	2	1	−1	2.508	1.721	1.633	−2.662	5.457
				0	−1.833,1.99	0.226	3.357	−1.557	2.014
				1	−2.949,1.273	−1.604	3.253	1.832	2.271
				2	−3.762	−3.097	1.642	2.352	5.387

Table 5 shows that the PDF of EN distribution may be unimodal or bimodal. The EN is a symmetric distribution for $c = 0$ and $b_1 = b_2$. If $c > 0$ or $c = 0$ and $b_1 < b_2$, then the EN distribution is positively skewed. If $c < 0$ or $c = 0$ and $b_1 > b_2$ — negatively skewed.

Equidispersion occurs when the variance is equal to the mean ([1]). Overdispersion is a situation in which the variance exceeds the mean, underdispersion is the opposite. The mean of the $EN(a_1, b_1, a_2, b_2, 0)$ — as mentioned earlier — equals $\frac{a_1+a_2}{2}$, so the $EN(a_1, b_1, a_2 \leq -a_1, b_1, 0)$ has underdispersion property. Figure 7 shows the regions in which the $EN(a_1, b_1, 0, 1, 2)$ and $EN(a_1, b_1, 0, 2, 1)$ distributions are overdispersed and underdispersed for selected parameter values. The regions for the $EN(a_1, b_1, 1, 1, -2)$ and $EN(a_1, b_1, 1, 1, 0)$ as well as for the $EN(0, b_1, 0, 1, c)$ and $EN(0, b_1, 0, 2, c)$ are presented in the supplementary material. The curve connects the points where the distribution is equidispersed. It is interesting to point out that the relationship between a_1 and b_1 in the $EN(a_1, b_1, 0, b_2, c > 0)$ remains linear for $b_2 = 1$, $c = 2$ and $b_2 = 2$, $c = 1$ (see Figure 7).

Proposition 2.4. *The moment generating function (MGF) of the EN distribution, based on (2.1), is given by*

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} \left(\frac{1}{b_1} e^{-z_1} + \frac{1}{b_2} e^{z_2} \right) \phi[c - \exp(-z_1) + \exp(z_2); 0, 1], \tag{2.6}$$

where $z_1 = \frac{x-a_1}{b_1}$ *and* $z_2 = \frac{x-a_2}{b_2}$.

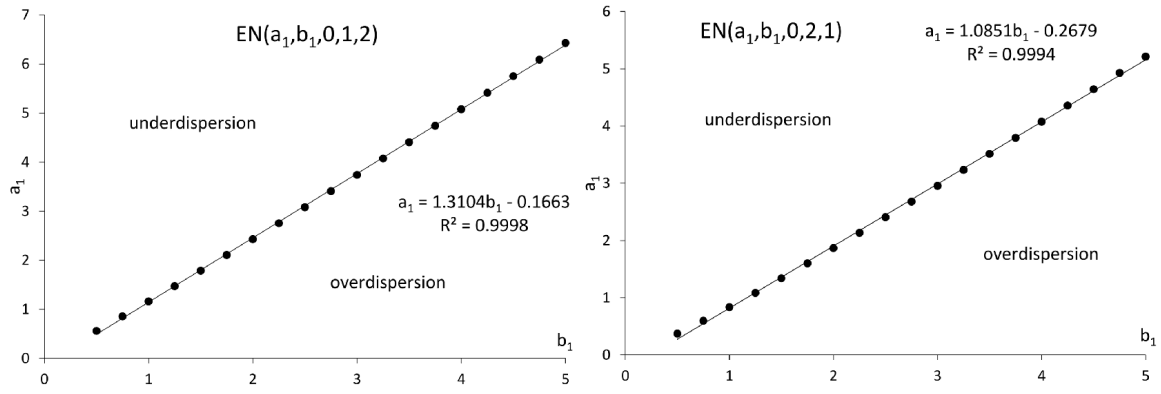


Figure 7: Dispersion regions for the $EN(a_1, b_1, 0, 1, 2)$ and $EN(a_1, b_1, 0, 2, 1)$.

2.5. Moments of order statistics

Proposition 2.5. *Let the random variable $X_{i,n}$ be the i-th order statistic $X_{1.n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ in a sample of size n from the $EN(a_1, b_1, a_2, b_2, c)$. The PDF of $X_{k,n}$ is given by*

$$f_{i.n}(x; *) = \frac{n!}{(i-1)!(n-i)!} f(x; *) F(x; *)^{i-1} [1 - F(x; *)]^{n-i},$$

where $* = (a_1, b_1, a_2, b_2, c)$, and $f(x; *)$, $F(x; *)$ are respectively given by (2.1) and (2.3).

Figure 8 plots the PDF of $X_{i,20}$ for some parameter values of the EN distribution. The k-th moment of the i-th order statistic $X_{k,n}$ is defined as

$$E\left(X_{i.n}^k\right) = \int_{-\infty}^{+\infty} x^k f_{i,n}(x).$$

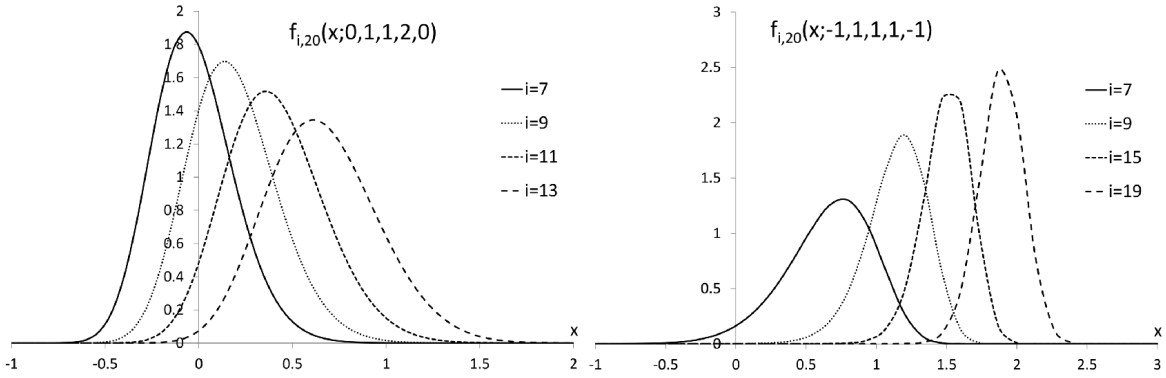


Figure 8: The PDF of the $X_{i,20}$ of the EN distribution.

2.6. Random numbers generator

Proposition 2.6. *Let* $X \sim EN(a_1, b_1, a_2, b_2, c)$, $R \sim Uniform(0,1)$. *The formula for generating* X *value, using the quantile function qEN of the EN distribution, is given by*

$$X = qEN(R; a_1, b_1, a_2, b_2, c).$$

The R codes for generating n values of X in increasing order are in the supplementary material as function rEN .

3. ESTIMATION PROCEDURES

Let $x_1^*, x_2^*, ..., x_n^*$ be a random sample of size n from the $EN(a_1, b_1, a_2, b_2, c)$. Our aim is to estimate the unknown parameter vector $\Theta = (a_1, b_1, a_2, b_2, c)$. The log-likelihood function based on (2.1) is given by

$$l(\Theta) = \sum_{i=1}^{n} \ln\left(\frac{1}{b_1} e^{-z_{1i}^*} + \frac{1}{b_2} e^{z_{2i}^*}\right) + \sum_{i=1}^{n} \ln\left[\phi\left(c - e^{-z_{1i}^*} + e^{z_{2i}^*}\right)\right], \tag{3.1}$$

where $z_{1i}^* = \frac{x_i^* - a_1}{b_1}, z_{2i}^* = \frac{x_i^* - a_2}{b_2}$. Solving the system of five complicated nonlinear equations in the form

$$\frac{dl(\Theta)}{da_1} = 0, \frac{dl(\Theta)}{db_1} = 0, \frac{dl(\Theta)}{da_2} = 0, \frac{dl(\Theta)}{db_2} = 0, \frac{dl(\Theta)}{dc} = 0$$

is not possible analytically. We had better maximize the log-likelihood function (3.1) in mathematical computing environments such as Excel, R and Mathcad. The MLEs of parameters a_1, b_1, a_2, b_2, c were calculated in R software using "optim" function.

The ordinary least square estimators (OLSEs) can be obtained by minimizing

$$O(\Theta) = \sum_{i=1}^{n} \left[F(x_i; a_1, b_1, a_2, b_2, c) - \frac{i}{n+1}\right]^2,$$

where $F(x; \Theta)$ is the CDF of the EN distribution (2.3).

The weighted least square estimators (WLSEs) can be obtained by minimizing

$$W(\Theta) = (n+1)^2(n+2) \sum_{i=1}^{n} \frac{1}{i(n-i+1)} \left[F(x_i; a_1, b_1, a_2, b_2, c) - \frac{i}{n+1}\right]^2,$$

where $F(x; \Theta)$ is the CDF of the EN distribution (2.3).

A simulation study is conducted to assess the properties of the MLEs, OLSEs, WLSEs of the parameter vector $\Theta = (a_1, b_1, a_2, b_2, c)$ using sample sizes of 50, 500 and 1000. In each case, 10^4 samples from the EN distribution with the specified parameters are drawn (see Figure 9).

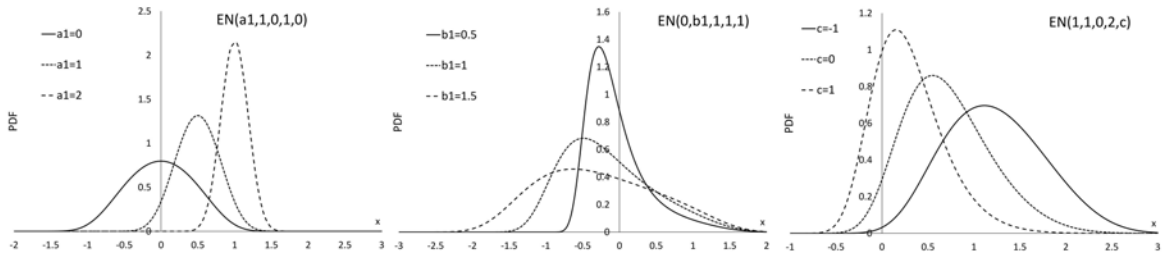


Figure 9: PDF of the EN distribution used in the estimation procedures (EPs).

The biases and the root mean squared errors (RMSEs) of the MLEs, OLSEs, WLSEs for the $EN(a_1, 1, 0, 1, 0)$ are presented in Table 6. The biases and the root mean squared errors (RMSEs) of the MLEs, OLSEs, WLSEs for the $EN(0, b_1, 1, 1, 1)$ and $EN(1, 1, 0, 2, c)$ are presented in the supplementary material.

Table 6: Biases and RMSEs of the MLEs (denoted as 1), OLSEs (denoted as 2), WLSEs (denoted as 3) for the $EN(a_1, 1, 0, 1, 0)$.

a_1	EP	n	Bias					RMSE				
			$\widehat{a}_1$	$\widehat{b}_1$	$\widehat{a}_2$	$\widehat{b}_2$	$\widehat{c}$	$\widehat{a}_1$	$\widehat{b}_1$	$\widehat{a}_2$	$\widehat{b}_2$	$\widehat{c}$
0	1	50	0.53	0.16	−0.65	0.15	−0.01	2.48	1.35	2.99	1.44	1.82
	2		0.56	0.47	−0.76	0.44	−0.12	1.62	1.38	1.97	1.42	1.15
	3		0.77	0.57	−1.06	0.57	−0.18	2.04	1.62	2.50	1.51	1.33
	1	500	0.10	0.03	−0.10	0.02	0.01	0.61	0.33	0.65	0.35	1.00
	2		0.24	0.14	−0.26	0.14	0.00	0.76	0.50	0.87	0.55	0.49
	3		0.14	0.09	−0.16	0.09	−0.01	0.57	0.37	0.69	0.42	0.44
	1	1e3	0.07	0.02	−0.05	0.01	0.03	0.47	0.24	0.47	0.25	0.84
	2		0.15	0.09	−0.15	0.09	0.01	0.52	0.36	0.58	0.39	0.37
	3		0.07	0.04	−0.06	0.04	0.01	0.33	0.23	0.37	0.25	0.34
1	1	50	0.41	0.13	−0.46	0.07	0.06	2.38	1.37	2.79	1.42	1.87
	2		0.28	0.24	−0.37	0.25	−0.05	1.15	0.94	1.43	1.14	1.07
	3		0.51	0.34	−0.62	0.34	−0.08	1.53	1.07	1.79	1.21	1.36
	1	500	0.29	0.10	−0.29	0.10	0.03	1.29	0.60	1.35	0.63	1.15
	2		0.15	0.07	−0.14	0.06	0.02	0.67	0.40	0.72	0.41	0.58
	3		0.22	0.09	−0.20	0.08	0.04	0.83	0.45	0.87	0.47	0.69
	1	1e3	0.19	0.07	−0.14	0.05	0.06	0.86	0.41	0.85	0.42	0.91
	2		0.10	0.05	−0.10	0.04	0.01	0.52	0.31	0.58	0.32	0.48
	3		0.16	0.07	−0.13	0.06	0.04	0.64	0.34	0.65	0.36	0.57
2	1	50	−0.02	−0.05	0.08	−0.08	0.23	1.63	1.32	1.72	0.97	2.34
	2		0.09	0.19	−0.13	0.26	0.04	0.93	0.84	1.43	1.60	1.45
	3		0.16	0.20	−0.20	0.24	−0.02	1.05	0.92	1.26	1.11	1.53
	1	500	0.11	0.01	−0.09	0.01	0.02	1.07	0.47	0.97	0.44	1.86
	2		0.05	0.03	−0.03	0.04	0.08	0.35	0.20	0.37	0.25	0.84
	3		0.06	0.03	−0.04	0.03	0.05	0.50	0.25	0.48	0.25	0.81
	1	1e3	0.08	0.01	−0.06	0.01	0.02	0.88	0.38	0.75	0.34	1.63
	2		0.03	0.02	−0.02	0.02	0.05	0.26	0.14	0.27	0.17	0.61
	3		0.04	0.02	−0.02	0.01	0.03	0.41	0.20	0.39	0.20	0.65

We observe in Table 6 that the estimates approach true values and RMSEs decrease as the sample size increases implying the consistency of the estimates. For $EN(0,1,0,1,0)$ and $EN(1,1,0,1,0)$ biases are the smallest for $\widehat{c}$ and the greatest for $\widehat{a}_2$ as well as RMSEs are the smallest for $\widehat{b}_1$ and the greatest for $\widehat{a}_2$ (see Table 6). The smallest biases are for maximum likelihood estimate (MLE) related to the $EN(0,1,0,1,0)$.

To examine the accuracy of the coverage probability of the asymptotic confidence intervals (CIs) using MLEe, another simulation study was performed with 10^4 samples using sample sizes of $50, 100, 250, 500$ and 1000. The study focused on the parameters a_1, b_1, a_2, b_2, c and samples were drawn from the $EN(0,1,1,1.25,1)$ (see Table 4). The coverage probabilities of the obtained 95% CIs for $a_1 = 0$, $b_1 = 1$, $a_2 = 1$, $b_2 = 1.25$, $c = 1$ reported in Table 7 are very close to the nominal level. The results suggest that the obtained standard errors and hence the asymptotic CIs are reliable.

Table 7: Coverage probabilities for the standard asymptotic 95% CIs.

Sample size n	a_1	b_1	a_2	b_2	c
50	0.9531	0.9521	0.9495	0.9496	0.9500
100	0.9511	0.9517	0.9422	0.9495	0.9455
250	0.9484	0.9507	0.9513	0.9529	0.9495
500	0.9509	0.9522	0.9519	0.9543	0.9522
1000	0.9449	0.9461	0.9495	0.9499	0.9472

4. SHANNON, RENYI AND TSALLIS ENTROPIES

Let $f(x, a_1, b_1, a_2, b_2, c)$ be a PDF of the EN distribution (2.1). The Shannon entropy of the EN distribution is given by [26]

$$S(a_1, b_1, a_2, b_2, c) = -\int_{-\infty}^{+\infty} f(x; a_1, b_1.a_1, b_1, c) \ln f(x; a_1, b_1.a_1, b_1, c) dx.$$

The Renyi entropy of order α for the EN distribution is defined as [21]

$$R_\alpha(a_1, b_1, a_2, b_2, c) = \frac{1}{1-\alpha} \ln\left(\int_{-\infty}^{+\infty} f(x; a_1, b_1.a_1, b_1, c)^\alpha dx\right) \quad (\alpha > 0, \alpha \neq 1).$$

The Tsallis entropy of order α for the EN distribution has the form [29]

$$T_\alpha(a_1, b_1, a_2, b_2, c) = \frac{1}{\alpha - 1}\int_{-\infty}^{+\infty} f(x; a_1, b_1.a_1, b_1, c)^\alpha dx - 1 \quad (\alpha > 0, \alpha \neq 1).$$

Renyi and Tsallis entropies converge to the Shannon entropy. Table 8 presents values of the Shannon, Renyi and Tsallis entropies for parameter values from groups A1–B2 and D1–F2 (see Table 4).

Table 8: Shannon (S), Renyi (R_α) and Tsallis (T_α) entropies. Groups of parameter values A1–B2, D1–F2.

Group	S	R_α			T_α		
		$\alpha=0.5$	$\alpha=2$	$\alpha=3$	$\alpha=0.5$	$\alpha=2$	$\alpha=3$
A1	0.89	1.06	0.727	0.65	−4.39	−0.52	−0.86
A2	1.43	1.64	1.17	1.02	−5.53	−0.69	−0.94
B1	0.65	0.82	0.50	0.43	−4.02	−0.39	−0.79
B2	1.47	1.69	1.19	1.04	−5.66	−0.70	−0.94
D1	1.32	1.45	1.21	1.15	−5.14	−0.70	−0.95
D2	0.65	0.71	0.59	0.57	−3.86	−0.45	−0.84
E1	0.89	1.03	0.75	0.68	−4.35	−0.53	−0.87
E2	1.66	1.76	1.51	1.40	−5.83	−0.78	−0.97
F1	0.89	1.03	0.75	0.68	−4.35	−0.53	−0.87
F2	1.66	1.76	1.51	1.40	−5.83	−0.78	−0.97

5. APPLICATION

The aim of this Section is to demonstrate the flexibility and applicability of the EN distribution. This section is composed of two real data examples. As mentioned in Introduction, the EN distribution is bimodal, so the analyzed real data are also bimodal. In papers devoted to probability distributions, Johnson distributions such as SB and SU are used very rarely in the examples, perhaps because of their unimodality. The other models selected for comparison with the new proposal are:

a) compound normal (CN) with PDF:

$$f(x; a_1, b_1, a_2, b_2, c) = \omega\phi(x; a_1, b_1) + (1-\omega)\phi(x; a_2, b_2);$$

b) compound Gumbel (CG) with PDF:

$$f_G(x; a, b) = \frac{1}{b}\exp\left[\frac{a-x}{b} - \exp\left(\frac{a-x}{b}\right)\right],$$

$$f(x; a_1, b_1, a_2, b_2, c) = \omega f_G(x; a_1, b_1) + (1-\omega) f_G(x; a_2, b_2);$$

c) two-piece power normal (TPPN) [22] with PDF:

$$\sigma = \sigma_1 I(x < \theta) + \sigma_2 I(x \geq \theta),$$

$$f(x; \theta, \sigma_1, \sigma_2, c) = \frac{c}{\sigma\sqrt{2\pi}}\left|\frac{x-\theta}{\sigma}\right|^{c-1}\exp\left[-0.5\left|\frac{x-\theta}{\sigma}\right|^{2c}\right];$$

d) bimodal skew-symmetric normal (BSSN) [12] with PDF:

$$f(x; \theta_1, \theta_2, c, d) = \frac{2c^{1.5}\left[d + (x-\theta_2)^2\right]\exp\left[-c(x-\theta_1)^2\right]}{\sqrt{\pi}\left[1 + 2c\left[d + (\theta_2-\theta_1)^2\right]\right]};$$

e) flexible generalized skew-normal of order 3 (FGSN) [16] with PDF:

$$u = \frac{x-a}{b},$$

$$f(x;a,b,\alpha_0,\alpha_1) = \frac{2}{b}\phi(u;0,1)\Phi\big(\alpha_0 u + \alpha_1 u^3;0,1\big);$$

f) bimodal asymmetric power-normal (BAPN) [8] with PDF:

$$u = \frac{x-\theta}{\sigma},$$

$$f(x;\alpha,\beta,\theta,\sigma) = \frac{\alpha 2^{\alpha}}{2^{\alpha}-1}\phi(u;0,1)\Phi(u;0,1)^{\alpha-1}\Phi(\beta u;0,1);$$

g) normal distribution with plasticizing component (NDPC) [24] with PDF:

$$u = \frac{x-a_2}{b_2}, \quad f_{pc}(x;a_2,b_2,c) = \frac{c}{b_2}|u|^{c-1}\phi(|u|^c;0,1),$$

$$f(x;a_1,b_1,a_2,b_2,c,\omega) = \omega\phi(x;a_1,b_1) + (1-\omega)f_{pc}(x;a_2,b_2,c).$$

The estimation of the model parameters is carried out by the maximum likelihood method. To avoid local maxima of the logarithmic likelihood function, the optimization routine is run 100 times with several different starting values that are widely scattered in the parameter space.

Table 9 presents the MLEs, confidence interval (CI), log-likelihood function l, AIC, BIC and HQIC for the first data sets. Models are sorted by AIC values.

Following the bootstrap method proposed in [5], [4] and [20], we used the obtained estimates $\widehat{\Theta}$ (Table 9) to derive the 95% bootstrap CIs for the parameters of distributions. We generated 10^4 samples of size n from the given distribution with values of the parameters equal to $\widehat{\Theta}$. For each obtained sample, we obtained the MLEs $\widehat{\Theta}_i^*\big(i = 1,2,...,10^4\big)$ using the true values of estimates as starting values for the maximum likelihood estimation. For the 95% bootstrap CIs, we took the 250-th and 9750-th ordered estimates.

Table 10 shows p-values (sorted by p-value of the KS test) for mentioned GoFTs calculated as follows. First, we obtain the values of the Kolmogorov–Smirnov (KS), Anderson-Darling (AD) and Cramer-von Mises (CvM) test statistics (denoted ST) for true values of parameters $\widehat{\Theta}$ based on the sample $x_1, x_2, ..., x_n$. In the next step we simulate 10^4 samples $x_1^{'}, x_2^{'}, ..., x_n^{'}$ from the given distribution with true values of parameters $\widehat{\Theta}$. For each sample, we calculate the values of the KS, AD and CvM test statistics (denoted ST^S). Finally, the p-value is calculated as $p \approx \#\big\{i : ST_i^S \geq ST\big\}10^{-4}$.

5.1. Example 1

The first real data present waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA ([13]).

The data consist of 272 observations of the variable "eruptions numeric Eruption time in mins" and are available in the R software with code faithful[1].

As shown in Table 9 the EN model is definitely the best in terms of the $-l$, AIC, BIC and HQIC values. The AIC ranking is the same as the BIC and HQIC rankings. The EN model is definitely distinguished by the p-values (see Table 10). The p-value ranking for the KS test is, with only one exception, the same as the p-value rankings for the AD and CvM tests. The information criteria ranking is not the same as the p-value ones. It is worth noting that the rankings are similar for most models, with the biggest difference in the rankings for the TPPN model.

Table 9: Results of estimation. Information criteria. Example 1.

Model		$\widehat{\Theta}$	95%CI	$-l$	AIC	BIC	HQIC
EN	$\widehat{a}_1$	-1.453	$[-1.486, -1.416]$	224.331	456.663	471.086	462.453
	$\widehat{b}_1$	0.185	$[0.147, 0.224]$				
	$\widehat{a}_2$	0.820	$[0.729, 0.896]$				
	$\widehat{b}_2$	0.481	$[0.405, 0.563]$				
	$\widehat{c}$	-0.427	$[-0.602, -0.264]$				
NDPC	$\widehat{a}_1$	0.508	$[0.353, 0.611]$	227.238	466.476	488.111	475.161
	$\widehat{b}_1$	0.375	$[0.275, 0.444]$				
	$\widehat{a}_2$	-0.173	$[-0.210, -0.137]$				
	$\widehat{b}_2$	1.219	$[1.182, 1.256]$				
	$\widehat{c}$	4.795	$[4.186, 5.786]$				
	$\widehat{\omega}$	0.342	$[0.235, 0.432]$				
CN	$\widehat{a}_1$	0.688	$[0.631, 0.745]$	240.394	490.788	508.817	498.026
	$\widehat{b}_1$	0.383	$[0.341, 0.423]$				
	$\widehat{a}_2$	-1.287	$[-1.328, -1.245]$				
	$\widehat{b}_2$	0.206	$[0.175, 0.237]$				
	$\widehat{\omega}$	0.652	$[0.597, 0.706]$				
TPPN	$\widehat{\theta}_1$	-0.454	$[-0.537, -0.370]$	244.651	497.301	511.724	503.092
	$\widehat{\sigma}_1$	0.921	$[0.835, 1.007]$				
	$\widehat{\sigma}_2$	1.357	$[1.267, 1.448]$				
	$\widehat{c}$	3.166	$[2.891, 3.549]$				
CG	$\widehat{a}_1$	-1.367	$[-1.405, -1.307]$	250.318	510.636	528.665	517.874
	$\widehat{b}_1$	0.180	$[0.145, 0.213]$				
	$\widehat{a}_2$	0.532	$[0.456, 0.604]$				
	$\widehat{b}_2$	0.362	$[0.218, 0.411]$				
	$\widehat{\omega}$	0.362	$[0.305, 0.427]$				
FGSN	$\widehat{a}$	0.191	$[0.153, 0.236]$	271.813	551.626	566.049	557.416
	$\widehat{b}$	1.016	$[0.930, 1.102]$				
	$\widehat{\alpha}_0$	4.148	$[3.389, 5.351]$				
	$\widehat{\alpha}_1$	-3.406	$[-4.942, -2.460]$				
BSSN	$\widehat{\theta}_1$	-0.212	$[-0.265, -0.155]$	277.255	562.509	576.932	568.300
	$\widehat{\theta}_2$	1.402	$[1.279, 1.625]$				
	$\widehat{c}$	-0.323	$[-0.372, -0.265]$				
	$\widehat{d}$	0.003	$[-0.045, 0.021]$				
BAPN	$\widehat{\alpha}$	16.160	$[14.243, 18.776]$	464.240	936.479	953.675	943.203
	$\widehat{\beta}$	0.048	$[-0.011, 0.108]$				
	$\widehat{\theta}$	-0.070	$[-0.090, -0.040]$				
	$\widehat{\sigma}$	0.543	$[0.520, 0.565]$				

Table 10: The KS, AD and CvM tests. Example 1.

Model	KS test		AD test		CvM test	
	TS	p-value	TS	p-value	TS	p-value
EN	0.0306	0.935	0.2845	0.9546	0.0414	0.9316
CN	0.049	0.4644	1.063	0.322	0.124	0.4741
NDPC	0.0514	0.4108	1.1111	0.3066	0.1724	0.3336
CG	0.0639	0.1858	2.129	0.0801	0.2636	0.1775
BSSN	0.0751	0.0814	3.896	0.0118	0.5454	0.0302
FGSN	0.0832	0.0404	3.289	0.0195	0.4366	0.0537
BAPN	0.1331	0.0001	8.7874	0.0002	1.0307	0.0019
TPPN	0.1495	0	7.201	0	1.516	0

Concluding, the EN model fits better than the other models analyzed in this case.

The second real data present Intercountry Life-Cycle Savings Data ([27], [6]). A detailed analysis of this example done identically to Example 1 is presented in the supplementary material.

6. CONCLUSIONS

Heterogeneity is not the only one cause of population distribution's bimodality. The population distribution is shaped by many factors. Therefore, the aim of the paper was to introduce into a family of the mixed bimodal distributions two distant relatives more. The relatives in question are distant since they are not of mixture form. So, they was denoted as non-mixed bimodal distributions. It is author's duty to give potential user of non-mixed bimodal distributions warning. Parameters of non-mixed bimodal distributions are not so clearly interpretable as parameters of mixed bimodal distributions are. Interpretability complication may, in turn, complicate conclusions when statistical reasoning procedure involves non-mixed bimodal distributions.

As a result of considerations presented in this paper two probability distributions denoted SC and SD came into existence. The distributions are members of the Johnson family of distribution. The SC and SD were tested in great depth, first for flexibility then for applicability.

In order to test for flexibility the Malachov plot was applied. The Malachov plot is a rectangular coordinate system with skewness (γ_1) as the abscissa and kurtozis (γ_2) as the ordinate. Points located below Malachov parabola $\gamma_2 = \gamma_1^2 + 1$ are related to obtainable γ_1/γ_2 combinations. The more flexible distribution is the wider points are scattered on the Malachov plot. In this paper the above fact served as a basis for definition of numerical flexibility measure being a fraction of an area "occupied" by particular distribution. The skewness-kurtosis-square measure was denoted SKS. Points are dimensionless entities, for a purpose of SKS measure, they were replaced with micro-squares. The best dispersion of points (γ_1, γ_2) occurs for the SD, SU and SB distribution.

After having flexibility testing completed the EN distribution was tested for applicability. For the purpose of applicability testing two real data sets were used. Empirical pdf's estimated from these data sets display bimodality. The EN had seven competitors with respect to applicability. These were already existing distributions that all have a property of bimodality. The competition consisted in fitting distributions to the data sets. Two types of rankings were performed. First the EN and its competitors were ranked with respect to information criteria. The criteria were AIC, BIC and HQIC ones. Then the EN and its competitors were ranked with respect to results of goodness-of-fit tests. The results were measured with p-values. The goodness-of-fit test involved in rankings were Kolmogorov-Smirnow, Anderson-Darling and Crmaer-von Mises ones. Altogether there were three information criteria rankings and three p-value rankings performed. It is interesting that all three information criteria rankings gave quite the same results. What makes a matter of rankings more interesting is that all three p-value rankings gave quite the same results too! So, one can say about one joint information criteria ranking and one joint p-value ranking. These rankings considerably differed from each other. In its essence this fact is not even strange since criteria differ considerably too. It is of special interest that the EN ranks high in all the rankings.

The content of the paper shows that the EN (including SC) as a new member of the Johnson family of distributions and simultaneously as a new distribution from the non-mixed bimodal distribution category, is a competitive model that deserves to be added to the existing distributions in modeling data.

ACKNOWLEDGMENTS

The author is thankful to the two Referees whose helpful comments and suggestions contributed to improve the quality of the paper.

REFERENCES

[1] Alzaatreh A.; Famoye, F. and Lee, C. (2014). The gamma-normal distribution: properties and application, *Computational Statistics & Data Analysi*, **69**, 67–80.

[2] Athayde, E.; Azevedo, C.; Leiva, V. and Sanhueza, A. (2012). About Birnbaum–Saunders distributions based on the Johnson system, *Communications in Statistics – Theory and Methods*, **41**(11), 2061–2079.

[3] Babak, V.; Eremenko, V. and Zaporozhets, A. (2019). Research of diagnostic parameters of composite materials using Johnson distribution, *Int. J. Comput*, **18**(4), 483–494.

[4] Balakrishnan, N. and Ristić, M.M. (2016). Multivariate families of gamma-generated distributions with finite or infinite support above or below the diagonal, *Journal of Multivariate Analysis*, **143**, 194–207.

[5] Balakrishna, N. and Shiji, K. (2014). On a class of bivariate exponential distributions, *Statistics & Probability Letter*, **85**, 153–180.

[6] Belsley, D.A.; Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics*, Wiley, New York.

[7] Birnbaum, Z.W. and Saunders, S.C. (1969). A new family of life distributions, *Journal of Applied Probability*, **6**(2), 319–327.

[8] Bolfarine, H.; Martínez-Flórez, G. and Salinas, H.S. (2018). Bimodal symmetric-asymmetric power-normal families, *Communications in Statistics – Theory and Methods*, **47**(2), 259–276.

[9] Castro R. (2013). Goodness-of-fit (GoF) Tests. Available at `http://www.win.tue.nl/~rmcastro/AppStat2013/files/lectures23.pdf`

[10] Díaz-García, J.A. and Domínguez-Molina, J.R. (2005). Generalised Birnbaum–Saunders and sinh-normal distributions, *Comunicación Técnica*, **I**, 5.

[11] Gaddum, J.H. (1945). *Lognormal distributions*, **1**(56), 463–466.

[12] Hassan, M.Y. and El-Bassiouni, M.Y. (2016). Bimodal skew-symmetric normal distribution, *Communications in Statistics – Theory and Methods*, **45**(5), 1527–1541.

[13] Härdle, W. (1991). *Smoothing Techniques with Implementation in S*, Springer, New York.

[14] Johnson, N.L. (1949). System of frequency curves generated by methods of translation, *Biometrika*, **36**, 149–176.

[15] Lu, Y.; Ramirez, O.A.; Rejesus, R.M.; Knight, T.O. and Sherrick, B.J. (2008). Empirically evaluating the flexibility of the Johnson family of distributions: A crop insurance application, *Agricultural and Resource Economics Review*, **37**, 79–91.

[16] Ma, Y. (2004). Flexible class of skew-symmetric distributions, *Journal of Statistics*, **31**(3), 456–468.

[17] Malakhov, A.N. (1978). *A Cumulant Analysis of Random Non-Gaussian Processes and Their Transformations*, Soviet Radio, Moscow (in Russian).

[18] Owen, W.J. (2004). Another look at the Birnbaum–Saunders distribution. Available at `https://www.researchgate.net/publication/267833513_Another_Look_at_the_Birnbaum_Saunders_Distribution`

[19] Rieck, J.R. and Nedelman, J.R. (1991). A log-linear model for the Birnbaum–Saunders distribution, *Technometrics*, **33**(1), 51–60.

[20] Ristić, S.; Popovic, M.M.B.V.; Zografos K. and Balakrishnan N. (2018). Discrimination among bivariate beta-generated distributions, *Statistics*, **52**(2), 303–320.

[21] Renyi, A. (1961). On measures of entropy and information, *Proc. Berekeley Symposium, Statist. Probability*, **1**, 547–561.

[22] Sulewski, P. (2019). Two-Piece Power Normal Distribution, *Communications in Statistics – Theory and Method*, DOI:10.1080/03610926.2019.1674871.

[23] Sulewski, P. (2019). Modified Lilliefors goodness-of-fit test for normality, *Communications in Statistics – Theory and Method*, DOI:10.1080/03610918. 2019.1664580.

[24] Sulewski, P. (2020). Normal distribution with plasticizing component, *Communications in Statistics – Theory and Method*, DOI:10.1080/03610926.2020.1837881.

[25] Stute W. (1993). Bootstrap based goodness-of-fit-tests, *Metrika*, **40**, 243–256.

[26] Shannon, C.E. (1948). A mathematical theory of communication, *Bell System Technical Journa*, **27**, 379–423.

[27] Sterling, A. (1977). Unpublished BS Thesis, Massachusetts Institute of Technology.

[28] Tabass, M.S.; Borzadaran, G.M. and Amini, M. (2016). Renyi entropy in continuous case is not the limit of discrete case, *Mathematical Sciences and Applications e-Notes*, **4**(1), 113–117.

[29] Tsallis, C. (1988). Possible generalizations of Boltzmann-Gibbs statistics, *Journal of Statistical Physics*, **52**, 479–487.

REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 557–572
https://doi.org/10.57805/revstat.v21i4.435

Exponential-Gaussian Distribution and Associated Time Series Models

Authors: K.U. Nitha ✉
– Department of Statistics, Farook College (Autonomous), Kozhikode, Kerala, India
nithajilesh@gmail.com

S.D. Krishnarani
– Department of Statistics, University of Calicut, Malappuram, Kerala-673635, India
krishnaranisd@gmail.com

Received: July 2020 Revised: February 2022 Accepted: February 2022

Abstract:

- Exponential-Gaussian distribution has already appeared in the literature and it is widely used in many fields. In this paper, we study its application in time series through a model-based approach. An autoregressive process of order one with exponential-Gaussian distribution as marginals is introduced. Structural aspects of the innovation sequence is derived and analytical properties of the process are studied. Estimation of the parameters is done and the application is established through an illustration with real data.

Keywords:

- *exponential distribution; Gaussian distribution; autoregressive process; stationarity; convolution models.*

✉ Corresponding author.

1. INTRODUCTION

Convolution based models are introduced in the literature as a combination of two random variables. It is used in various fields like physics, engineering, biological studies, etc. Radovic *et al.* [22] studied breakdown voltage distribution in neon using the convolution of distributions. For applications of convolution models, one may refer to Burbeck and Luce [6], Rosso *et al.* [23] and Golubev [11]. Fajriyah [10] noted that beta convolutions and beta convoluted normal can be used in microarray experiments due to the presence of some non-biological noises. Plancade *et al.* [20] introduced gamma-normal convolution to model the background correction of the Illumina BeadArrays. Exponential-lognormal convolution is found to be a good fit for microarray data. Several convolution models based on different underlying distributions such as the exponential-gamma, the normal-gamma and the exponential-normal were studied in the past and different estimation methods of the parameters have been discussed and illustrated with real life data sets. Chen *et al.* [7] used exponential-gamma distribution and its highly skewed behavior for the improved estimation of the detection of differently expressed genes.

For the modeling and analysis of symmetrical and tailed peaks in the data, the exponential-Gaussian distribution has been used by several researchers in the past. The exponential-Gaussian distribution discussed by Xie *et al.* [26], is a convolution distribution for the observed gene expression intensities by assuming that the true signal intensities are exponentially distributed and the noise intensity is normally distributed. Ding *et al.* [9] used the exponential-normal convolution model to correct the background of the Illumina platform by using Markov chain Monte Carlo simulation. With the name exponentially modified Gaussian, the convolution of exponential and Gaussian distribution has been found as a good model in modeling chromatographic peaks as seen in Naish and Hartwell [19]. This distribution is used to model residuals in Ament *et al.* [2]. Application of this distribution in flow injection analysis, quantitation of chromatographic peaks etc. is explained in Jeansonne and Foley [14]. A recent work on exponential-Gaussian distribution is also seen in Jehan *et al.* [13].

Time series analysis using autoregressive models having non-normality assumptions had been an interesting area of researchers of all times. See Lawrance [17], Lawrance and Lewis [16], Popovici [21] and Billard [5] for the details of stationary autoregressive models under the assumption that the marginal distribution is exponential, and refer Sim [24] for gamma distributed marginals. The convolution distribution is relatively less explored in time series data analysis. In the regression context, one may refer to Gori and Rioul [12], where they estimated a linear bound in the presence of outliers under the assumption that the noise is exponential-Gaussian distributed. Also, it is of interest to study time series models developed under the assumption that data is exponential-Gaussian distributed. In this paper, we study the first order autoregressive time series models having exponential-Gaussian as marginals. The paper is systematically organized into various sections as follows.

In Section 2, we consider the probability density function (pdf) of the exponential-Gaussian distribution and bring out its analytical properties. The autoregressive process of order 1(AR(1)) with exponential-Gaussian distribution as marginals is introduced and the distribution of the innovation random variable is identified in Section 3. Important properties

of the proposed model are derived in Section 4. The parameters involved in the proposed model are estimated using different methods and the performance of the same are verified using a simulation study in Section 5. Section 6 is devoted to the analysis of real data of GDP growth rate using the proposed model.

2. EXPONENTIAL-GAUSSIAN DISTRIBUTION

Let $U \in \mathbb{R}$ and $V \in \mathbb{R}^+$ be two independent and continuous random variables with pdfs $g(\cdot)$ and $h(\cdot)$ respectively. Then the pdf of the random variable $X = U + V$ is

$$f(x) = (g * h)(x) = \int_0^{+\infty} g(x-v)h(v)dv. \tag{2.1}$$

With particular choice of U as Gaussian with parameters μ and σ and V as exponential with mean λ in (2.1), the pdf of X is

$$f(x) = \frac{1}{\lambda\sigma\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{(x-\mu-v)^2}{2\sigma^2}} e^{-\frac{v}{\lambda}} dv.$$

Using the erfc$(\cdot)$ function, Naish and Hartwell [19] expressed the above integral in a more convenient form as

$$f(x) = \frac{1}{2\lambda} e^{\frac{1}{\lambda}\left(\frac{\sigma^2}{2\lambda}+\mu-x\right)} \operatorname{erfc}\left(\frac{1}{\sqrt{2}\sigma}\left(\frac{\sigma^2}{\lambda}+\mu-x\right)\right), \tag{2.2}$$
$$-\infty < x < \infty, \lambda > 0, \mu \in \mathbb{R}, \sigma > 0,$$

where

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt.$$

We denote the exponential-Gaussian random variable having pdf (2.2) as EG(λ, μ, σ). A striking feature of such a construction is that, the resultant distribution is capable of capturing the skewed behaviour of the data. X being the sum of independent normal and exponential random variables, it is obvious that

$$E(X) = \mu + \lambda, \tag{2.3}$$

$$\operatorname{Var}(X) = \sigma^2 + \lambda^2, \tag{2.4}$$

$$\text{Skewness(X)} = \frac{2\lambda^3}{(\sigma^2+\lambda^2)^{3/2}} \tag{2.5}$$

and

$$\text{Kurtosis(X)} = 3\frac{(\sigma^4 + 2\lambda^2\sigma^2 + 3\lambda^4)}{(\sigma^2+\lambda^2)^2}. \tag{2.6}$$

When $\lambda \to 0$, the exponential-Gaussian becomes a Normal distribution with skewness zero and kurtosis value 3.

The characteristic function of the EG(λ, μ, σ) is given by

$$\phi_X(t) = \frac{e^{i\mu t - \frac{1}{2}t^2\sigma^2}}{1-\lambda it}. \tag{2.7}$$

The shape of the pdf of EG(λ, μ, σ) random variable for various values of the parameter λ, by taking $\mu = 0$ and $\sigma = 1$, is depicted in Figure 1. The shape of EG(λ, μ, σ) is determined by the value of $k = \frac{\sigma}{\lambda}$. When $k \to 0$ the EG(λ, μ, σ) density function will be very close to the exponential density, and when k is very large, the distribution is close to the Gaussian distribution. The density plots reveal an apparent similarity in shape, but the peakedness increases significantly and becomes heavy tailed as λ increases.

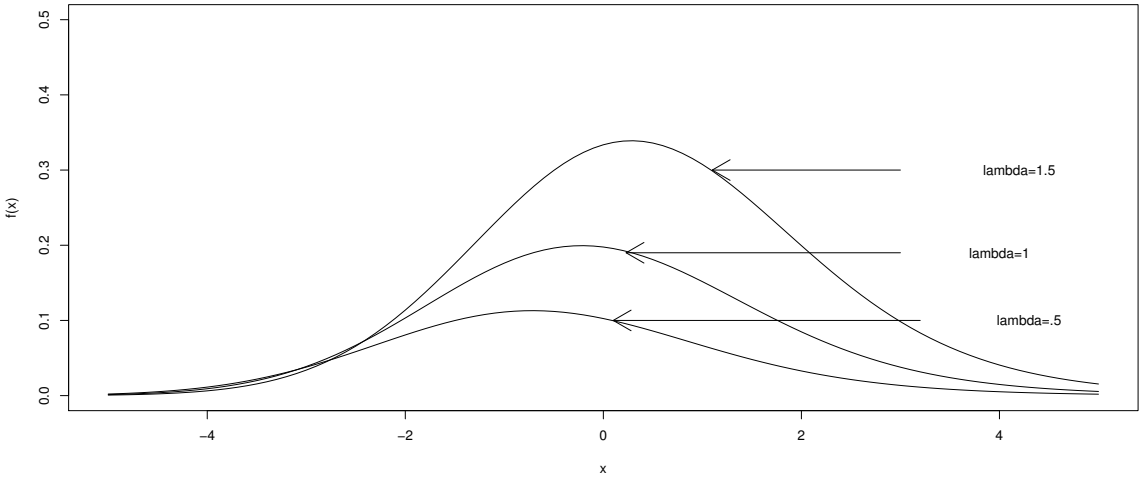


Figure 1: Shape of the density function of exponential-Gaussian for the different values of $\lambda \in \{0.5, 1, 1.5\}$, $\mu = 0$, $\sigma = 1$.

To discuss its application in time series, we propose an AR(1) process with EG(λ, μ, σ) distribution as marginals in the next section.

3. AR(1) MODEL WITH EXPONENTIAL-GAUSSIAN AS MARGINAL

Let $\{X_n\}$ be a first order autoregressive process having the linear structure

$$X_n = aX_{n-1} + \epsilon_n, \ |a| < 1. \tag{3.1}$$

Assume that $\{X_n\}$ is a stationary process with exponential-Gaussian distribution as marginals and $\{\epsilon_n\}$ is a sequence of independent and identically distributed (i.i.d) random variables independent of $\{X_t\}$, where $t < n$.
Since X_n's are stationary, by using the characteristic function of X_n we can write

$$\phi_{\epsilon_n}(t) = \frac{\phi_X(t)}{\phi_X(at)}. \tag{3.2}$$

Since X_n is following EG(λ, μ, σ) distribution, substituting (2.7) we obtain

$$\phi_{\epsilon_n}(t) = e^{it\mu(1-a) - \frac{\sigma^2 t^2 (1-a^2)}{2}} \left[\frac{1 - \lambda i a t}{1 - \lambda i t}\right]. \tag{3.3}$$

Using the expression of $\phi_{\epsilon_n}(t)$, we can represent the random variable $\{X_n\}$ as

$$X_n = aX_{n-1} + \begin{cases} Z_n, & \text{with probability } a, \\ W_n, & \text{with probability } 1-a, \end{cases} \tag{3.4}$$

where $Z_n \sim N(\mu(1-a), \sigma\sqrt{1-a^2})$ and $W_n \sim \mathrm{EG}(\lambda, \mu(1-a), \sigma\sqrt{(1-a^2)})$. Alternatively (3.3) can be expressed as

$$\phi_{\epsilon_n}(t) = e^{it\mu(1-a) - \frac{\sigma^2 t^2 (1-a^2)}{2}} \left[a + (1-a)\frac{1}{1-\lambda it} \right]. \tag{3.5}$$

Further it may be noted that, using the tailed exponential random variable of Littlejohn [18], X_n may be written as

$$X_n = aX_{n-1} + Y_{1n} + Y_{2n}, \tag{3.6}$$

where $Y_{1n} \sim N(\mu(1-a), \sigma\sqrt{1-a^2})$ and

$$Y_{2n} = \begin{cases} 0, & \text{with probability } a, \\ Exp(\lambda) & \text{with probability } 1-a, \end{cases} \tag{3.7}$$

and $Exp(\lambda)$ is the exponential distributed random variable with mean λ.

Now we define the first order exponential-Gaussian autoregressive process (EGAR(1)) as given below.

Definition 3.1. A Markovian sequence $\{X_n\}$ defined according to (3.1), is said to be an exponential-Gaussian autoregressive process of order 1 (EGAR(1)) with $\mathrm{EG}(\lambda, \mu, \sigma)$ distribution as marginals if and only if $\{\epsilon_n\}$ admits the following representation

$$\epsilon_n = \begin{cases} Z_n, & \text{with probability } a, \\ W_n, & \text{with probability } 1-a, \end{cases} \tag{3.8}$$

where $Z_n \sim N(\mu(1-a), \sigma\sqrt{1-a^2})$ and $W_n \sim \mathrm{EG}(\lambda, \mu(1-a), \sigma\sqrt{(1-a^2)})$.

From (3.8), the pdf of ϵ_n can be written as

$$f_{\epsilon_n}(x) = a f_{Z_n}(x) + (1-a) f_{W_n}(x),$$

where

$$f_{Z_n}(x) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{1-a^2}} e^{\frac{-1}{2(1-a^2)}\left(\frac{x-\mu(1-a)}{\sigma}\right)^2},$$

$$f_{W_n}(x) = \frac{1}{2\lambda} e^{\frac{1}{\lambda}\left(\frac{\sigma^2(1-a^2)}{2\lambda} + \mu(1-a) - x\right)} \operatorname{erfc}\left(\frac{1}{\sqrt{2(1-a^2)}\sigma} \left(\frac{\sigma^2(1-a^2)}{\lambda} + \mu(1-a) - x \right) \right).$$

In the next section, we shall bring together the important properties of EGAR(1).

4. PROPERTIES OF THE EGAR(1) PROCESS

Proposition 4.1. *If X_0 is distributed arbitrarily, the Markovian process (3.1) is again exponential-Gaussian distributed asymptotically.*

Proof: We can rewrite $X_n = aX_{n-1} + \epsilon_n$, as

$$X_n = a^n X_0 + \textstyle\sum_{k=0}^{n-1} a^k \epsilon_{n-k}.$$

Consequently, the characteristic function is

$$\begin{aligned}\phi_{X_n}(t) &= \phi_{X_0}(a^n t) \prod_{k=0}^{n-1} \phi_\epsilon(a^k t) \\ &= \phi_{X_0}(a^n t) \left[exp\left(i\mu(1-a) \sum_{k=0}^{n-1} a^k t - \frac{1}{2}\sigma^2(1-a^2) \sum_{k=0}^{n-1} a^{2k} t^2 \right) \right] \prod_{k=0}^{n-1} \frac{1 - ia^{k+1}t\lambda}{1 - ia^k t\lambda}.\end{aligned}$$

As $n \to \infty$,

$$\phi_{X_n}(t) \to e^{i\mu t - \frac{\sigma^2 t^2}{2}} \left[\frac{1}{1 - it\lambda} \right], \tag{4.1}$$

implying that X_n is asymptotically EG(λ, μ, σ) distributed. □

Proposition 4.2. *For the EGAR(1) process, the $(k+1)$ step ahead conditional mean is given by*

$$E(X_{n+k}|X_{n-1} = x_{n-1}) = a^{k+1}x_{n-1} + (1 - a^{k+1})(\lambda + \mu). \tag{4.2}$$

Proof: Using (3.1), we have

$$X_{n+k} = a^{k+1}X_{n-1} + a^k\epsilon_n + a^{k-1}\epsilon_{n+1} + \cdots + \epsilon_{n+k}. \tag{4.3}$$

By taking expectation conditionally on $X_{n-1} = x_{n-1}$ on both sides, we obtain the desired result. □

Remark 4.1. When $k \to \infty$,

$$E(X_{n+k}|X_{n-1} = x_{n-1}) \to \lambda + \mu, \tag{4.4}$$

which is the unconditional mean of the process.

Proposition 4.3. *For the EGAR(1) process, the $(k+1)$ step ahead conditional variance is given by*

$$\mathrm{Var}(X_{n+k}|X_{n-1} = x_{n-1}) = (1 - a^{2(k+1)})(\sigma^2 + \lambda^2). \tag{4.5}$$

Remark 4.2. As $k \to \infty$,

$$\mathrm{Var}(X_{n+k}|X_{n-1} = x_{n-1}) \to (\sigma^2 + \lambda^2). \tag{4.6}$$

Proposition 4.4. *EGAR(1) process is not time-reversible.*

Proof: The joint characteristic function of (X_n, X_{n+1}) is

$$\begin{aligned}
\phi_{X_n,X_{n+1}}(t) &= E\left(e^{it_1X_n+it_2X_{n+1}}\right) \\
&= E\left(e^{(it_1X_n+it_2(aX_n+\epsilon_{n+1})}\right) \\
&= \phi_{X_n}(t_1 + at_2)\phi_{\epsilon_{n+1}}(t_2) \\
&= e^{i\mu(t_1+t_2)-\frac{\sigma^2}{2}(t_1^2+t_2^2+2at_1t_2)}\frac{1 - ia\lambda t_2}{(1 - i\lambda t_2)(1 - i\lambda(t_1 + at_2))},
\end{aligned}$$

which is not symmetric in t_1 and t_2. So the process EGAR(1) is not time reversible. □

Remark 4.3. From the model defined in (3.1),

$$E(X_n|X_{n-1} = x) = ax + (1 - a)(\lambda + \mu).$$

Therefore, we can see that regression in the forward direction is linear and the conditional variance is constant.

Following the steps in Lawrance [17], the joint moment generating function (m.g.f) of (X_n, X_{n+1}) is

$$M_{X_n,X_{n+1}}(t_1, t_2) = \frac{M_X(t_1 + at_2)M_X(t_2)}{M_X(at_2)}. \tag{4.7}$$

Differentiating this with respect to t_1 and setting $t_1 = 0, t_2 = t$,

$$\begin{aligned}
E(e^{tX_{n+1}}E(X_n|X_{n+1})) &= \frac{M'_X(at)M_X(t)}{M_X(at)} \\
&= M'_X(at)M_\epsilon(t) \\
&= e^{t\mu+\frac{\sigma^2t^2}{2}}\left[\frac{\lambda a + (1 - \lambda at)(\mu a + a^2\sigma^2 t)}{(1 - \lambda at)(1 - \lambda t)}\right]. \qquad (4.8)
\end{aligned}$$

Also differentiating (4.7) with respect to t_2 and setting $t_2 = 0, t_1 = 0$, we get $E(X_n) = \lambda + \mu$.

Proposition 4.5. *The characteristic function of the partial sums* $S_r = X_n + X_{n+1} + \cdots + X_{n+r-1}$ *is*

$$\begin{aligned}
\phi_{S_r}(t) \;=\; & \left[exp\left(i\mu\frac{1 - a^r}{1 - a}t - \frac{\sigma^2}{2}\left(\frac{1 - a^r}{1 - a}\right)^2 t^2\right)\right]\frac{1}{1 - \lambda i\left(\frac{1-a^r}{1-a}\right)t} \\
& \cdot \prod_{j=1}^{r-1}\left[exp\left(i\mu(1 - a^{r-j})t - \frac{\sigma^2}{2}(1 - a^2)\left(\frac{1 - a^{r-j}}{1 - a}\right)^2 t^2\right)\right]\frac{1 - a\lambda i\left(\frac{1-a^{r-j}}{1-a}\right)t}{1 - \lambda i\left(\frac{1-a^{r-j}}{1-a}\right)t}.
\end{aligned}$$

Proof:

$$\begin{aligned} S_r &= X_n + X_{n+1} + \cdots + X_{n+r-1} \\ &= \sum_{j=0}^{r-1} a^j X_n + \sum_{j=0}^{r-2} a^j \epsilon_{n+1} + \sum_{j=0}^{r-3} a^j \epsilon_{n+2} + \cdots + \epsilon_{n+r-1} \\ &= X_n \left(\frac{1-a^r}{1-a}\right) + \sum_{j=1}^{r-1} \epsilon_{n+j} \left(\frac{1-a^{r-j}}{1-a}\right), \end{aligned}$$

$$\begin{aligned} \phi_{S_r}(t) &= \phi_{X_n}\left(\frac{1-a^r}{1-a}t\right) \prod_{j=1}^{r-1} \phi_\epsilon \left(\frac{1-a^{r-j}}{1-a}t\right) \\ &= \left[exp\left(i\mu \frac{1-a^r}{1-a}t - \frac{\sigma^2}{2}\left(\frac{1-a^r}{1-a}\right)^2 t^2\right)\right] \frac{1}{1-\lambda i\left(\frac{1-a^r}{1-a}\right)t} \\ &\prod_{j=1}^{r-1}\left[exp\left(i\mu(1-a^{r-j})t - \frac{\sigma^2}{2}(1-a^2)\left(\frac{1-a^{r-j}}{1-a}\right)^2 t^2\right)\right] \frac{1-a\lambda i\left(\frac{1-a^{r-j}}{1-a}\right)t}{1-\lambda i\left(\frac{1-a^{r-j}}{1-a}\right)t}. \end{aligned}$$

On inverting the above expression of the characteristic function of S_r, one may obtain its distribution. □

5. ESTIMATION

In this section we will discuss the estimation of the parameters. The parameters involved in the process are μ, a, σ and λ. Let $(X_1, ..., X_n)$ be the realizations from the EGAR(1) process. Method of moments, conditional least square method, and Gaussian estimation method are discussed in the following sections. A simulation study is also conducted.

5.1. Estimation using the Method of Moments

Using (2.3), (2.4), and (2.5), we can identify the estimates for the parameters μ, σ and λ under the method of moments estimation. The autoregressive parameter a can be estimated by the sample autocorrelation function (ACF), that is $\hat{a} = corr(X_n, X_{n-1})$. Other moment estimates are given by

$$\hat{\mu} = m - s\left(\frac{\gamma}{2}\right)^{1/3}, \tag{5.1}$$

$$\hat{\sigma^2} = s^2\left[1 - \left(\frac{\gamma}{2}\right)^{2/3}\right] \tag{5.2}$$

and

$$\hat{\lambda} = s\left(\frac{\gamma}{2}\right)^{1/3}, \tag{5.3}$$

where m is the sample mean, s is the sample standard deviation and γ is the skewness.

It may be noted that explicit expression for the mean and variance of the above estimators are not available.

5.2. Conditional Least Square Estimation

The conditional least square estimates of the parameters are obtained by minimizing the conditional sum of squares function

$$D_n(a,\mu,\sigma,\lambda) = \sum_{i=1}^{n}(x_i - E(X_i|X_{i-1}=x_{i-1}))^2. \tag{5.4}$$

From the linearity of the regression of EGAR(1) process, we have

$$E(X_i|X_{i-1}=x) = ax_{i-1} + (1-a)(\lambda+\mu). \tag{5.5}$$

Therefore, (5.4) can be written as

$$D_n(a,\mu,\sigma,\lambda) = \sum_{i=1}^{n}[x_i - ax_{i-1} - (1-a)(\lambda+\mu)]^2. \tag{5.6}$$

Solving the normal equations obtained from (5.6) we obtain estimates of a and μ in terms of $\hat{\lambda}$ as

$$\hat{a} = \frac{n\sum x_i x_{i-1} - \sum x_i \sum x_{i-1}}{n\sum x_{i-1}^2 - (\sum x_i)^2}, \tag{5.7}$$

$$\hat{\mu} = \frac{\sum x_i - \hat{a}\sum x_{i-1}}{n(1-\hat{a})} - \hat{\lambda}. \tag{5.8}$$

Estimates of σ and λ can be identified numerically through other methods, like maximizing the conditional likelihood function, and also by making use of (5.7) and (5.8). The conditional likelihood function is given by

$$\begin{aligned} L(x;a,\mu,\sigma,\lambda) &= \left[\prod_{i=1}^{n} f_{X_i|X_{i-1}}(x_i|x_{i-1})\right] f_{X_0}(x_0) \\ &= \left[\prod_{i=1}^{n} f_{X_i|X_{i-1}}(x_i|x_{i-1})\right] \frac{1}{2\lambda} e^{\frac{1}{\lambda}\left(\frac{\sigma^2}{2\lambda}+\mu-x_0\right)} \\ &\quad \cdot \operatorname{erfc}\left(\frac{1}{\sqrt{2}\sigma}\left(\frac{\sigma^2}{\lambda}+\mu-x_0\right)\right), \end{aligned}$$

where

$$\begin{aligned} f_{X_i|X_{i-1}}(x_i|x_{i-1}) &= a f_{Z_n}(x_i - ax_{i-1}) + (1-a) f_{W_n}(x_i - ax_{i-1}) \\ &= a\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma\sqrt{1-a^2}} e^{\frac{-1}{2}\frac{(x_i - ax_{i-1} - \mu(1-a))^2}{\sigma^2(1-a^2)}} \\ &\quad + (1-a)\frac{1}{2\lambda} e^{\frac{1}{\lambda}\left(\frac{\sigma^2(1-a^2)}{2\lambda}+\mu(1-a)-x_i+ax_{i-1}\right)} \\ &\quad \cdot \operatorname{erfc}\left(\frac{1}{\sqrt{2(1-a^2)}\sigma}\left(\frac{\sigma^2(1-a^2)}{\lambda}+\mu(1-a)-x_i+ax_{i-1}\right)\right). \end{aligned}$$

Since EGAR(1) is a stationary process and the moments are finite, using the regularity conditions of Klimko and Nelson [15], it is verified that the conditional least square estimators

obtained are consistent and asymptotically normal. That is, $\sqrt{n}\Big[(\hat{a},\hat{\lambda})' - (a,\lambda)'\Big] \to N(0,\Sigma)$ where $N(0,\Sigma)$ is a bivariate normal distribution with mean 0 and dispersion matrix

$$\Sigma = \begin{bmatrix} (1-a^2) & 0 \\ 0 & \frac{1+a}{1-a}\lambda^2 \end{bmatrix}.$$

5.3. Gaussian Estimation Method

Whittle [25] introduced this method by taking, Gaussian likelihood function as the baseline distribution for the estimation. Later, Crowder [8] used this method of estimation for the analysis of correlated binomial data. Al-Nachawati *et al.* [1] and Alwasel *et al.* [3] used the same estimation procedure in the context of first order autoregressive process. Although this method has an approximate nature, this gives a good estimation to our model and also the possibility to estimate all the parameters in the model. The conditional maximum likelihood function is given by

$$L = f(x_1)\prod_{t=2}^{n} f(x_t|x_{t-1}). \tag{5.9}$$

Here $f(x_t|x_{t-1})$ and $f(x_1)$ are the conditional and marginal probability function of $X_t|X_{t-1}$ and X_t, respectively. We assume Gaussian pdf for $f(x_1)$ and $f(x_t|x_{t-1})$ with conditional mean and conditional variance as the parameters. Then the log-likelihood function can be written as

$$log(L) = n\, log\frac{1}{\sqrt{2\pi}} - \frac{1}{2}\sum_{t=2}^{n}\left(log(\sigma^2_{x_{t-1}}) + \frac{(x_t - m_{x_{t-1}})^2}{\sigma^2_{x_{t-1}}}\right), \tag{5.10}$$

where $m_{x_{t-1}} = E(X_t|X_{t-1}) = ax_{t-1} + (1-a)(\lambda+\mu)$ and $\sigma^2_{x_{t-1}} = \mathrm{Var}(X_t|X_{t-1}) = (1-a^2)(\lambda^2+\sigma^2)$. So, the Gaussian log-likelihood function corresponding to $EGAR(1)$ process becomes

$$log(L) = n\, log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{t=2}^{n}\left[log\big((1-a^2)(\lambda^2+\sigma^2)\big) + \frac{(x_t - ax_{t-1} - (1-a)(\lambda+\mu))^2}{(1-a^2)(\lambda^2+\sigma^2)}\right]. \tag{5.11}$$

The Gaussian estimators are, thus, obtained by maximising the above non linear equation. But explicit expressions as the solution for the parameters a, λ, μ and σ are not available. Therefore, we have used numerical methods for identifying the value for these parameters. We use the `nlminb()` function in R with the Nelder–Mead method for this purpose. Crowder [8] pointed out that under Gaussian method of estimation of the parameter θ, $\sqrt{n}(\hat{\theta}-\theta)$ is asymptotically normally distributed with mean zero and asymptotic variance $[J(\theta)]^{-1}$, where $J(\theta)$ is the conditional expected information matrix. An approximation of the same can be done using the observed conditional information matrix, see Bakouch and Popovic [4].
To check the performance of the estimates, we have conducted a simulation study and the mean square error (MSE) is used for the comparison purpose.

5.4. Simulation Study

For checking the validity of the model, we simulated 100 samples of sizes 100, 500, 1000, 5000 and 10000 for different values of the parameters. The values considered are:

(**1**) $a = 0.1,\ \lambda = 1,\ \mu = 2$ and $\sigma = 1$;

(**2**) $a = 0.2,\ \lambda = 3,\ \mu = 5$ and $\sigma = 4$;

(**3**) $a = 0.5,\ \lambda = 8,\ \mu = 10$ and $\sigma = 8$.

We obtained the estimates of the parameters and corresponding MSE values and the results are presented in Table 1. It can be seen from the table that the Gaussian estimators are very close to the true value of the parameters and it shows better performance when the sample size increases. Further it may be noted that the MSE values decrease as the sample size increases for different parameter values.

Table 1: Estimated values of a, λ, μ and σ corresponding mean squared error (MSE).

True values are: $a = 0.1,\ \lambda = 1,\ \mu = 2,\ \sigma = 1$								
Sample size	$\widehat{a}$	$\widehat{\lambda}$	$\widehat{\mu}$	$\widehat{\sigma}$	MSE($\widehat{a}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{\mu}$)	MSE($\widehat{\sigma}$)
100	0.1371	0.9158	1.8221	1.1812	0.0613	0.1514	0.1034	0.1588
500	0.1158	0.9821	1.9948	1.1609	0.0227	0.0863	0.0949	0.1368
1000	0.1117	0.9969	1.9951	1.2321	0.0184	0.0546	0.0602	0.1211
5000	0.1041	1.0028	1.9934	1.0860	0.0124	0.0299	0.0287	0.0947
10000	0.1030	1.0003	2.0002	1.0636	0.0055	0.0269	0.0272	0.0782
True values are: $a = 0.2,\ \lambda = 3,\ \mu = 5,\ \sigma = 4$								
Sample size	$\widehat{a}$	$\widehat{\lambda}$	$\widehat{\mu}$	$\widehat{\sigma}$	MSE($\widehat{a}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{\mu}$)	MSE($\widehat{\sigma}$)
100	0.1895	3.1918	5.1041	3.7119	0.0629	0.4664	0.2149	0.5053
500	0.1908	2.9903	4.9784	3.7372	0.0369	0.3011	0.1347	0.4146
1000	0.1946	2.9960	4.9863	3.7753	0.0299	0.1769	0.0738	0.3238
5000	0.1980	2.9986	4.9975	3.8110	0.0261	0.1369	0.0573	0.3009
10000	0.2004	3.0001	4.9993	3.8410	0.0095	0.0605	0.0393	0.2363
True values are: $a = 0.5,\ \lambda = 8,\ \mu = 10,\ \sigma = 8$								
Sample size	$\widehat{a}$	$\widehat{\lambda}$	$\widehat{\mu}$	$\widehat{\sigma}$	MSE($\widehat{a}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{\mu}$)	MSE($\widehat{\sigma}$)
100	0.4842	7.8667	10.1034	7.5170	0.0631	0.8189	0.5678	1.0678
500	0.4951	7.9766	9.9780	7.6179	0.0383	0.6861	0.4468	0.9906
1000	0.4973	7.9827	9.9863	7.8915	0.0288	0.4723	0.3221	0.8955
5000	0.4985	7.9930	9.9945	7.8955	0.0120	0.2179	0.1620	0.8915
10000	0.5001	7.9989	10.0005	7.9301	0.0080	0.1339	0.1059	0.8207

Next we shall look into the sample path behaviour of the EGAR(1) process. We simulated 500 observations from the proposed process by taking $a = 0.5,\ \mu = 0,\ \sigma = 1$ and $\lambda = 0.2, 0.4, 0.6, 0.8$, and the same is plotted in Figure 2. The sample path clearly shows that the simulated data is stationary.

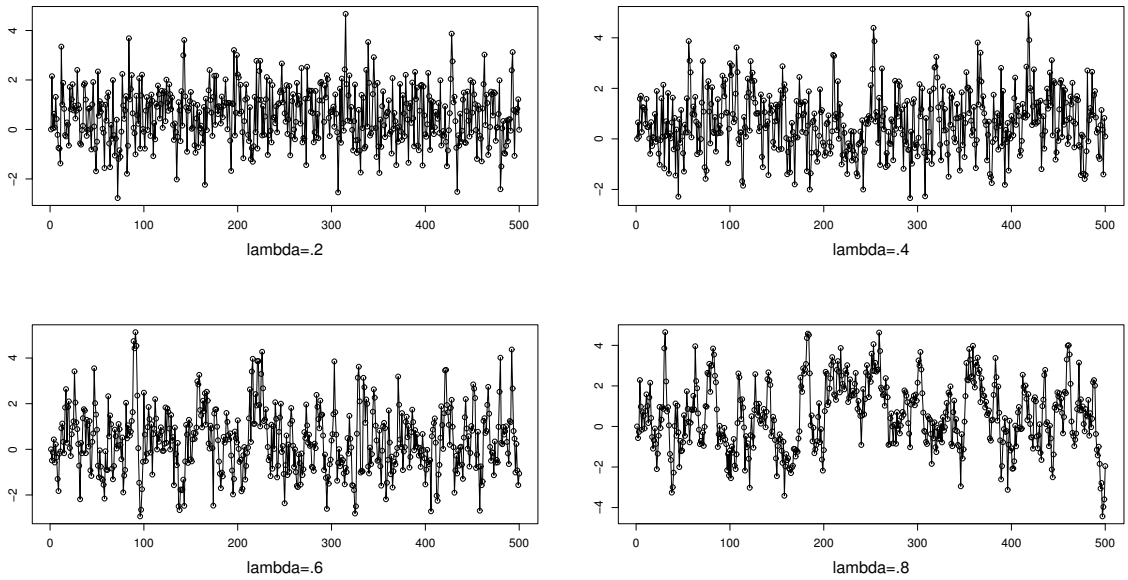


Figure 2: Sample path for the values of lambda=0.2, 0.4, 0.6, 0.8, μ=0 and σ=1.

6. DATA ANALYSIS

For establishing the applicability of the model, we considered the US GDP growth rate data for the period from 1961 to 2018 which is available in `https://data.worldbank.org`. The exponential-Gaussian distribution is found to be a suitable distribution for the data and the fitted density is ploted in Figure 3. We performed the Kolmogorov–Smirnov test of goodness of fit for the data set to check the adequacy of the exponential-Gaussian distribution and obtained the p-value as $0.28 > 0.05$. Significantly high p-value indicates the acceptance of the hypothesis that EG is a good fit to the data.

Figure 3: Fitted Exponential-Gaussian distribution for the GDP data.

Hence we tried to model the data using the proposed EGAR(1) model. The time series plot, the plots of the ACF and the partial autocorrelation function (PACF) of the data are presented in Figure 4. We can find that the ACF and PACF is significant only at lag 1.

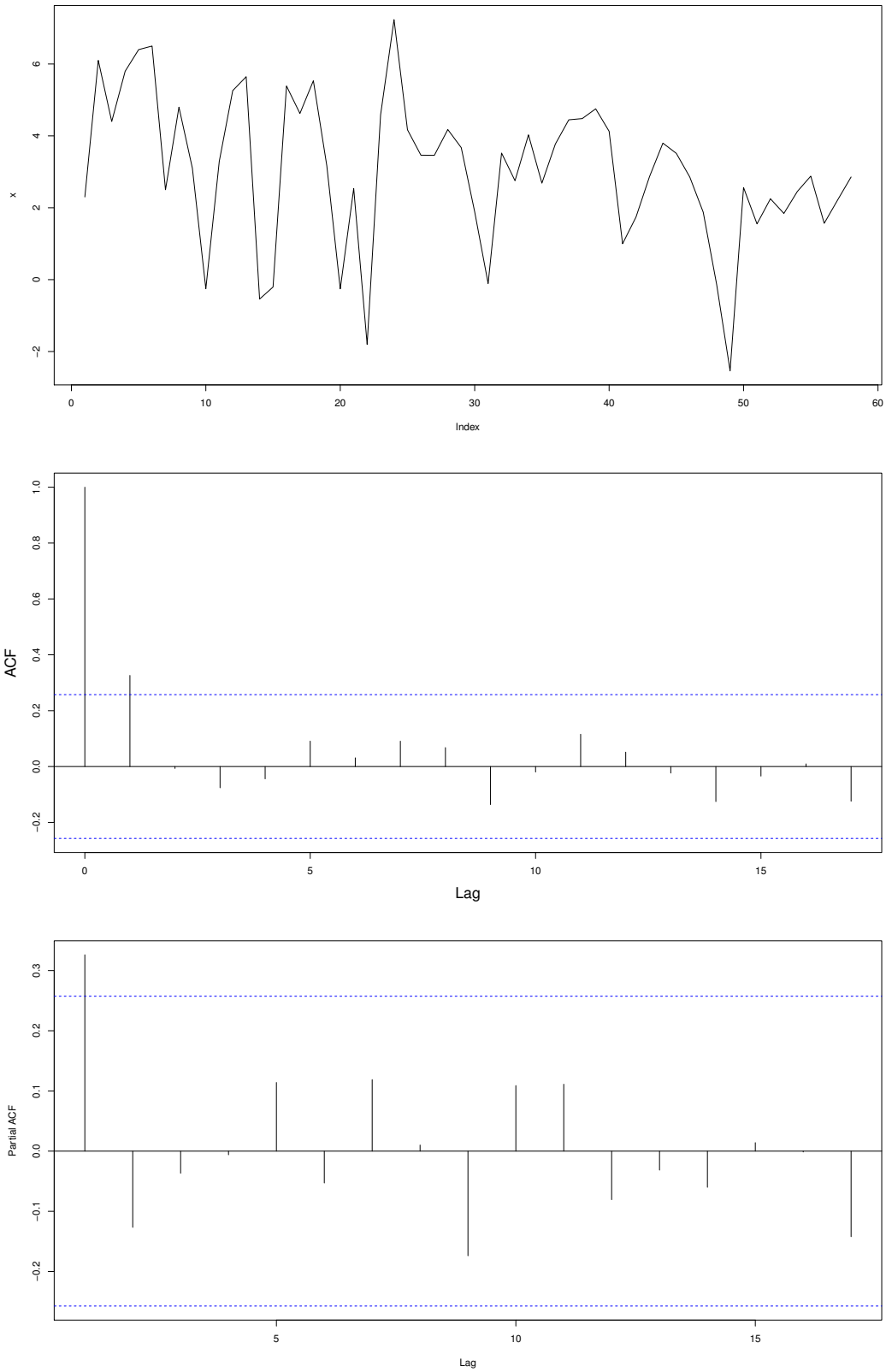


Figure 4: GDP data, ACF, PACF plots.

Therefore, we use AR(1) model for this data. The Gaussian estimation discussed in Section 5.3 is performed and the values of the parameters of EGAR(1) process are obtained as $\widehat{a} = 0.33$, $\widehat{\lambda} = 1.88$, $\widehat{\mu} = 1.18$ and $\widehat{\sigma} = 0.19$. Residual analysis has been carried out and the model adequacy has been checked. The p-value of Ljung–Box test is $0.79 > 0.05$, accepting the null hypothesis that the residuals are white noise. Also, the ACF and PACF of the residual are within the limits as represented in Figure 5. We have calculated the standard errors of the estimates using the Hessian matrix and got the standard errors of a as 0.05, μ as 0.08 and λ as 0.08.

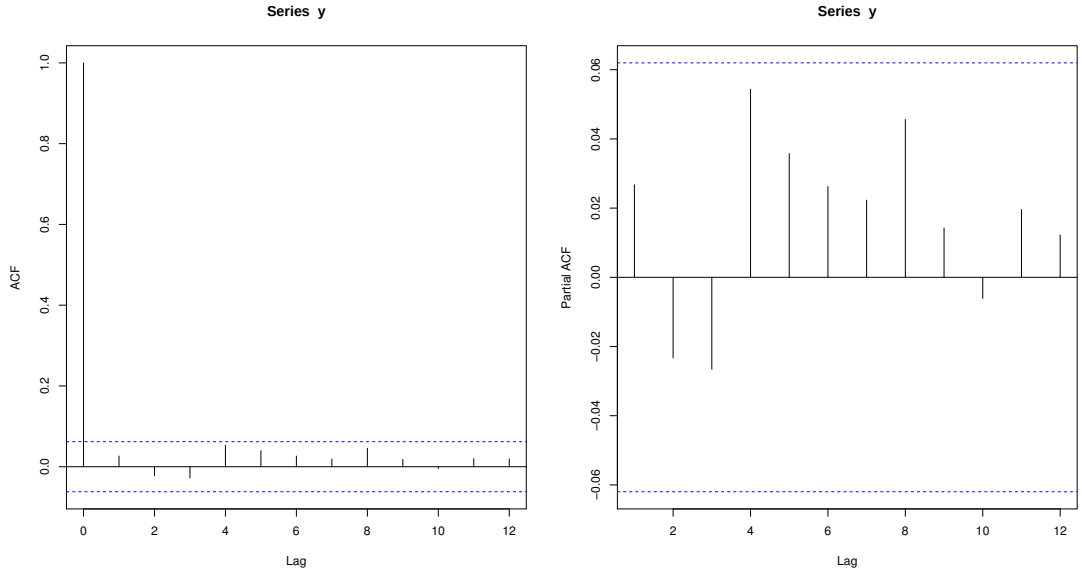


Figure 5: ACF and PACF plots of residuals.

Due to the evaluation of the likelihood function outside the given range of the parameter values, the standard error of σ is not evaluated correctly. We have predicted the GDP values for the next years and plotted them in Figure 6. In particular, note that the predicted value of GDP growth rate in the year 2019 was 2.32, where the actual value is 2.161.

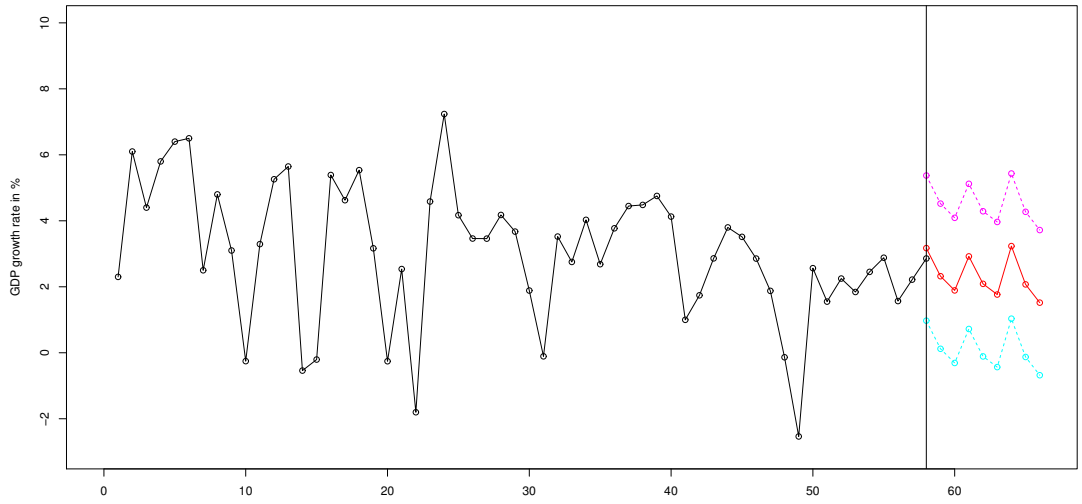


Figure 6: Prediction for the GDP data.

7. CONCLUSION

In this paper, we studied exponential-Gaussian distribution as a suitable model for data having symmetry or heavy tailed behaviour. The time series application of the EG distribution has been explored using an AR(1) process. The estimation of the model parameters and the problem of fitting of the model to real time series data and simulated data are perused. Application of the similar convolution model like Lindley-Gaussian, Gamma-Gaussian etc. are themes for future works. It may be interesting to investigate non-linear time series models and stochastic volatility models based on EG distribution.

ACKNOWLEDGMENTS

The authors express their sincere thanks to the editor-in-chief and anonymous reviewers for their constructive comments which improved the quality of the paper. Also we would like to acknowledge DST, India for the financial assistance through the FIST programme.

REFERENCES

[1] Al-Nachawati, H.; Alwasel, I. and Alzaid, A.A. (1997). Estimating the parameters of the generalized Poisson AR(1) process, *Journal of Statistical Computation and Simulation*, **56**, 337–352.

[2] Ament, S.; Gregoire, J. and Gomes, C. (2019). Exponentially-Modified Gaussian mixture model: Applications in Spectroscopy, *ArXiv preprint arXiv:1902.05601 (arxiv.org)*.

[3] Alwasel, I.; Alzaid, A. and Al-Nachawati, H. (1998). Estimating the parameters of the binomial autoregressive process of order one, *Applied Mathematics and Computation*, **95**, 193–204.

[4] Bakouch, H.S. and Popovic, B.V. (2016). Lindley first-order autoregressive model with applications, *Communications in Statistics – Theory and Methods*, **45**(17), 1–20.

[5] Billard, L. (2014). Exponential autoregressive models, *Wiley StatsRef: Statistics Reference Online*, John Wiley & Sons Ltd.

[6] Burbeck, S.L. and Luce, R.D. (1982). Evidence from auditory simple reaction times for both change and level detectors, *Perception and Psychophysics*, **32**(2), 117–133.

[7] Chen, M.; Xie, Y. and Story, M. (2011). An exponential-gamma convolution model for background correction of Illumina BeadArray data, *Communications in Statistics – Theory and Methods*, **40**(17), 3055–3069.

[8] Crowder, M. (1985). Gaussian estimation for correlated binomial data, *Journal of Royal Statistical Society*, **47**(2), 229–237.

[9] Ding, L.H.; Xie, Y.; Park, S.; Xiao, G. and Story, M.D. (2008). Enhanced identification and biological validation of differential gene expression via illumina hole genome expression arise through the use of the model based background correction methodology, *Nucleic Acids Research*, **36**, 10:e58.

[10] Fajriyah, R. (2016). A Study of convolution models for background correction of BeadArrays, *Austrian Journal of Statistics*, **45**, 15–33.

[11] Golubev, A. (2010). Exponentially modified Gaussian(EMG) relevance to distributions related to cell proliferation and differentiation, *Journal of Theoretical Biology*, **262**, 257–266.

[12] Gori, J. and Rioul, O. (2019). Regression to a linear lower bound with outliers an exponentially modified Gaussian noise model, *Eusipco 2019 27th European Signal Processing Conference*, Sep. 2019, A Coruna, Spain.

[13] Jehan, A.; Sajid A. and Ismail, S. (2020). Monitoring schedule time using exponentially modified Gaussian distribution, *Quality Technology & Quantitative Management*, **17**(4), 448–469.

[14] Jeansonne, M.S. and Foley, J.P. (1991). Review of the exponentially modified Gaussian (EMG) Function Since 1993, *Journal of Chromatographic Science*, **29**(6), 258–266.

[15] Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes, *The Annals of Statistics*, **6**(3), 629–642.

[16] Lawrance, A.J. and Lewis, P.A.W. (1981). A new autoregressive time Series model in exponential variables (NEAR(1)), *Advances in Applied Probability*, **13**(04), 826–845.

[17] Lawrance, A.J. (1978). Some autoregressive models for point processes, *Colloquia Mathematica Societatis Janos Bolyai*, **24**, 257–275.

[18] Littlejohn, R.P. (1994). A reversibility relationship for two markovian time series models with stationary exponential tailed distribution, *Journal of Applied Probability*, **31**, 575–581.

[19] Naish, P.J. and Hartwell, S. (1988). Exponentially modified Gaussian functions – a good model for chromatographic peaks in isocratic HPLC, *Chromatographia*, **26**, 285–296.

[20] Plancade, S.Y.; Rozenholc, Y. and Lund, E. (2012). Generalization of the normal-exponential model: exploration of a more accurate parametrisation for the signal distribution on Illumina BeadArrays, *BMC Bioinfomatics*, **78**, 493–506.

[21] Popovici, G. (2010). On three classes of time series involving exponential distribution, *Mathematical Reports*, **12**(1), 45–57.

[22] Radovic, M.K.; Maluckov, C.A.; Karamarkovic, J.P.; Rancev, S.A. and Mitic, S.D. (2014). Convolution based model of breakdown voltage distributions in neon at 1.33 mbar with corona appearance in pre-breakdown regime, *Romanian Reports in Physics*, **66**(2), 472–480.

[23] Rosso, L.; Cathryn, S.B.; James, M.G.; Azeem, S.; Patricia, M.P.; Federico, E.T. and Eric, O.A. (2009). A new model for prediction of drug distribution in tumor and normal tissues: Pharmacokinetics of temozolomide in glioma patients, *American Association for Cancer Research*, 120–127.

[24] Sim, C.H. (1994). Modelling non-normal first-order autoregressive time series, *Journal of Forecasting*, **13**, 369–381.

[25] Whittle, P. (1961). Gaussian estimation in stationary time series, *Bulletin of the International Statistical Institute*, **39**, 105–130.

[26] Xie, Y.; Wang, X. and Story, M. (2009). Statistical methods of background correction for Illumina BeadArray data, *Bioinformatics*, **25**(6), 751–757.

REVSTAT-Statistical Journal

Aims and Scope

The aim of REVSTAT-Statistical Journal is to publish articles of high scientific content, developing Statistical Science focused on innovative theory, methods, and applications in different areas of knowledge. Important survey/review contributing to Probability and Statistics advancement is also welcome.

Background

Statistics Portugal started in 1996 the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, a quarterly publication whose goal was the publication of papers containing original research results, and application studies, namely in the economic, social and demographic fields. Statistics Portugal was aware of how vital statistical culture is in understanding most phenomena in the present-day world, and of its responsibilities in disseminating statistical knowledge.

In 1998 it was decided to publish papers in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work. At the time, the editorial board was mainly composed by Portuguese university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal. In 2001, the *Revista de Estatística* published a three volumes special issue containing extended abstracts of the invited and contributed papers presented at the 23rd European Meeting of Statisticians (EMS). During the EMS 2001, its editor-in-chief invited several international participants to join the editorial staff.

In 2003 the name changed to REVSTAT-Statistical Journal, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

Editorial policy

REVSTAT-Statistical Journal is an open access peer-reviewed journal published quarterly, in English, by Statistics Portugal.

The editorial policy of REVSTAT is mainly placed on the originality and importance of the research. The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage revstat.ine.pt based in Open Journal System (OJS). The only working language allowed is English. Authors intending to submit any work must register, login and follow the guidelines.

There are no fees for publishing accepted manuscripts that will be made available in open access.

All articles consistent with REVSTAT aims and scope will undergo scientific evaluation by at least two reviewers, one from the Editorial Board and another external. Authors can suggest an editor or reviewer who is expert on the paper subject providing her/his complete information, namely: name, affiliation, email and, if possible, personal URL or ORCID number.

All published works are Open Access (CC BY 4.0) which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Also, in the context of archiving policy, REVSTAT is a *blue* journal welcoming authors to deposit their works in other scientific repositories regarding the use of the published edition and providing its source.

Journal prints may be ordered at expenses of the author(s), and prior to publication.

Abstract and Indexing services

REVSTAT-Statistical Journal is covered *by Journal Citation Reports - JCR (Clarivate); DOAJ-Directory Of Open Access Journals; Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; Scimago Journal & Country Rank; Scopus*

Author guidelines

The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage https://revstat.ine.pt/ based in Open Journal System (OJS). Authors intending to submit any work must *register, login* and follow the indications choosing *Submissions*.

REVSTAT - Statistical Journal adopts the COPE guidelines on publication ethics.

Work presentation

- the only working language is English;
- the first page should include the name, ORCID iD (optional), Institution, country, and mail-address of the author(s);
- a summary of fewer than one hundred words, followed by a maximum of six keywords and the MSC 2020 subject classification should be included also in the first page;
- manuscripts should be typed only in black, in double-spacing, with a left margin of at least 3 cm, with numbered lines, and a maximum of 25 pages;
- the title should be with no more than 120 characters (with spaces);
- figures must be a minimum of 300dpi and will be reproduced online as in the original work, however, authors should take into account that the printed version is always in black and grey tones;
- authors are encouraged to submit articles using LaTeX which macros are available at *REVSTAT style*;
- citations in text should be included in the text by name and year in parentheses, as in the following examples: § article title in lowercase (Author 1980); § This

theorem was proved later by AuthorB and AuthorC (1990); § This subject has been widely addressed (AuthorA 1990; AuthorB et al. 1995; AuthorA and AuthorB 1998).

- references should be listed in alphabetical order of the author's scientific surname at the end of the article;
- acknowledgments of people, grants or funds should be placed in a short section before the References title page. Note that religious beliefs, ethnic background, citizenship and political orientations of the author(s) are not allowed in the text;
- authors are welcome to suggest one of the Editors or Associate Editors or yet other reviewer expert on the subject providing a complete information, namely: name, affiliation, email and personal URL or ORCID number in the Comments for the Editor (submission form).

Accepted papers

After final revision and acceptance of an article for publication, authors are requested to provide the corresponding LaTeX file, as in REVSTAT style.

Supplementary files may be included and submitted separately in .tiff, .gif, .jpg, .png .eps, .ps or .pdf format. These supplementary files may be published online along with an article, containing data, programming code, extra figures, or extra proofs, etc; however, REVSTAT is not responsible for any supporting information supplied by the author(s).

Copyright Notice

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information.

According to REVSTAT's *archiving policy*, after assigning the copyright form, authors may cite and use limited excerpts (figures, tables, etc.) of their works accepted/published in REVSTAT in other publications and may deposit only the published edition in scientific repositories providing its source as REVSTAT while the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.