ISSN 1645-6726

REVSTAT

Statistical Journal

vol. 21 - n. 3 - July 2023

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT-Statistical Journal, vol.21, n. 3 (July 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 3, July 2023, 303–319
https://doi.org/10.57805/revstat.v21i3.360

On an Induced Distribution and its Statistical Properties

Authors: Brijesh P. Singh
– Department of Statistics, Institute of Science, Banaras Hindu University, India
brijesh@bhu.ac.in

Utpal Dhar Das ✉
– Department of Mathematics, School of Basic Sciences and Research, Sharda University, India
utpal.statmath@gmail.com

Received: December 2020 Revised: November 2021 Accepted: November 2021

Abstract:

- In this study an attempt has been made to propose a way to develop new distribution. For this purpose, we need only idea about distribution function. Some important statistical properties of the new distribution like moments, cumulants, hazard and survival function has been derived. The Rényi entropy, Shannon entropy has been obtained. Also ML estimate of parameter of the distribution is obtained, that is not closed form. Therefore, numerical technique is used to estimate the parameter. Some real data sets are used to check the suitability of the proposed distribution over some other existing one parameter lifetime distributions. The various diagnostic tools such as -2LL, AIC, BIC and K-S test shows that the proposed distribution provides better fit than other distributions for the considered data sets.

Keywords:

- *induced distribution; Bonferroni and Gini index; entropy; generating function; hazard function; MLE; MRLF; order statistics.*

AMS Subject Classification:

- 49A05, 78B26.

✉ Corresponding author.

1. INTRODUCTION

Almost all applied sciences including, biomedical science, engineering, finance, demography, environmental and agricultural sciences, there is a need of statistical analysis and modeling of the data. A number of continuous distributions for modeling lifetime data have been introduced in statistical literature such as Exponential, Lindley, Gamma, Lognormal and Weibull. Among these Gamma and Lognormal distributions are less popular because their survival functions cannot be expressed in closed forms and both require numerical integration. Researchers in probability distribution theory often use a probability distributions based on either their mathematical simplicity or because of their flexibility. Several parametric models are used in the analysis of lifetime data and in the problems associated with the modeling of the failure process. The Exponential distribution is often used to model the time interval between successive random events but Gamma and Weibull distribution is the most widely used model for lifetime distribution due to its flexibility. The exponential distribution is a particular case of the Gamma and Weibull distribution. In order to increase the suitability of the well-known distributions, many authors have proposed different transformations to generate new distributions, it has been an increased interest in defining new generators for univariate continuous distributions by introducing one or more additional shape parameter(s) to the baseline model. This improves the goodness-of-fit of the proposed generated distribution.

In the context of increasing flexibility in distribution, many generalization or transformation methods are available in the literature based on baseline distribution. Ghitany *et al.* [6] developed a two-parameter weighted Lindley distribution and discussed its applications to survival data. Zakerzadeh and Dolati [26] obtained a generalized Lindley distribution and discussed its various properties and applications. Shaw and Buckley [23] proposed a new transformation method by adding one extra parameter and Kumaraswamy [9] gives another method of proposing new distribution by taking baseline distribution. A families of distributions for the median of a random sample drawn from an arbitrary lifetime distribution is introduced by Abd-Elrahman [1]. Since its failure rate function is monotonically increasing with finite limit for this they generalize distribution by making transformation $X = \left(\frac{Y-\delta}{\theta}\right)^{\lambda}$, the parameter δ is a threshold parameter, θ and λ are the scale and the shape parameters, respectively. Gupta *et al.* [7] proposed an exponentiated type distribution by adding one more shape parameter. A new generalization of Lindley distribution, i.e. SSD distribution, appear in Singh *et al.* [25]. In very recent compounded exponential lindley distribution (CEL) has been studied by Singh *et al.* [24]. A new class of distribution by adding two additional shape parameters is found (see Cordeiro *et al.* [4]). Also some well-known generators are the beta-G by Eugene *et al.* [5], gamma-G by Zografos and Balakrishnan [27], the Zografos-Balakrishnan-G family by Nadarajah *et al.* [13].

2. GENESIS OF THE DISTRIBUTION

In this study, an attempt has been made to develop a new continuous distribution using concept discussed by Gupta and Kirmani [8]. Let X be a continuous random variable with the cumulative distribution function (cdf) $F(x)$ and expectation $E(X)$. It is worthwhile to mention here that the $E(X)$ can be defined in terms of cdf of any distribution as follows:

$$E(X) = \int_0^\infty [1 - F(x)]dx.$$

Let us have, for positive x,

$$\int_0^\infty [1 - F(x)]dx = \lim_{k\to\infty} \int_0^k 1.[1 - F(x)]dx.$$

Now, integrating by parts, we have

$$\lim_{k\to\infty} [\{1 - F(k)\}k] + \lim_{k\to\infty} \int_0^k xf(x)dx, \quad \text{where} \quad \frac{d}{dx}[F(x)] = f(x). \tag{2.1}$$

Since $F(\infty) = 1$, $\lim_{k\to\infty} [\{1 - F(k)\}k] = 0$, then

$$\int_0^\infty [1 - F(x)]dx = \lim_{k\to\infty} \int_0^k xf(x)dx = \int_0^\infty xf(x)dx = E(X). \tag{2.2}$$

Keeping the above concept into mind we define a pdf $g^*(x)$ as

$$g^*(x) = \frac{1 - F(x)}{E(X)}, \qquad x > 0. \tag{2.3}$$

If $g^*(x)$ is a pdf then its integration over the range should be equal to 1. Now we have

$$\int_0^\infty g^*(x)dx = \int_0^\infty \frac{[1 - F(x)]}{E(X)}dx = \frac{1}{E(X)} \int_0^\infty [1 - F(x)]dx = \frac{E(X)}{E(X)} = 1.$$

Therefore the generated pdf using the above transformation technique will be a valid pdf. This $g^*(x)$ may be called an induced or equilibrium distribution. Actually this distribution is a particular case of weighted distribution defined by Patil and Rao [14]. According to the Patil and Rao [14], if $f(x;\theta)$ be the probability distribution function of random variable X and the unknown parameter θ the weighted distribution is defined as:

$$f^*(x;\theta) = \frac{w(x)f(x;\theta)}{E[w(x)]}, \qquad x \in \mathbb{R}, \quad \theta > 0,$$

where $w(x)$ is the weight function, and $f(x;\theta)$ is the base line distribution. We know that $1 - F(x) = S(x) = \frac{f(x)}{h(x)}$, i.e. if we take $w(x) = h(x)^{-1}$, we can get the induced distribution defined above in equation number (2.3). This distribution is well connected to its parent distribution and many of the statistical properties can be easily studied.

2.1. Proposed distribution

We consider cdf of one parameter Lindley distribution and using the idea of induced distribution given in the equation (2.3), the pdf and cdf of transformed distribution is given in equations (2.4) and (2.5) respectively:

$$f(x;\theta) = \frac{\theta}{\theta+2}(1+\theta+\theta x)e^{-\theta x}, \tag{2.4}$$

$$F(x;\theta) = 1 - \left[1 + \frac{\theta x}{\theta+2}\right]e^{-\theta x}, \qquad x > 0, \quad \theta > 0. \tag{2.5}$$

In fact this distribution is Garima distribution and already discussed by Shanker [20], which is a mixture of Exponential (θ) and Gamma $(2, \theta)$ distribution with mixing proportion $\frac{\theta+1}{\theta+2}$. Also he discussed its various statistical properties.

Therefore in this paper, we consider cdf $F(x)$ of Garima distribution as a base line distribution and try to develop a new distribution. The pdf and cdf of the new distribution is as follows:

$$g(x;\theta) = \frac{\theta}{\theta+3}(2+\theta+\theta x)e^{-\theta x}, \qquad x > 0, \quad \theta > 0, \tag{2.6}$$

and the corresponding cdf is

$$G(x;\theta) = 1 - \left[1 + \frac{\theta x}{\theta+3}\right]e^{-\theta x}, \qquad x > 0, \quad \theta > 0. \tag{2.7}$$

The above distribution is similar to the base line distribution and develop using concept of induced distribution thus named as induced Garima (i-Garima) distribution. This distribution can also be consider as second order induced Lindley distribution. The cdf of i-Garima distribution is displayed in Figure (1).

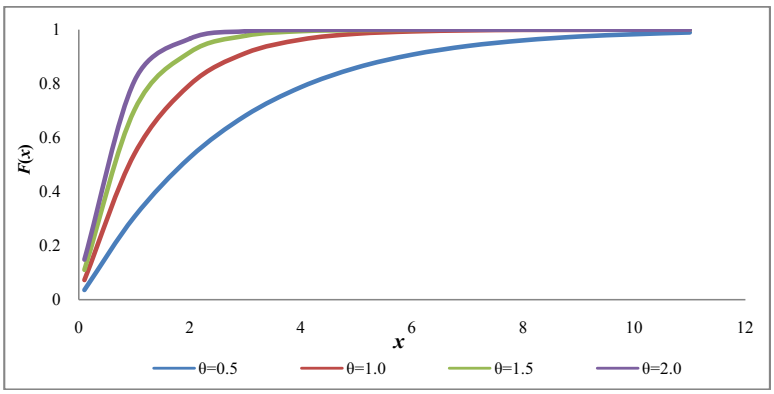

Figure 1: Cumulative distribution function of i-Garima distribution.

The proposed distribution, i.e. i-Garima distribution, can be easily expressed as a mixture of Exponential (θ) and Gamma $(2, \theta)$ as

$$f(x;\theta) = pg_1(x) + (1-p)g_2(x), \tag{2.8}$$

where $p = \frac{\theta+2}{\theta+3}$, $g_1(x) = \theta e^{-\theta x}$, and $g_2(x) = \theta^2 x e^{-\theta x}$.

3. PROPERTIES

The r-th order moments about origin is given by

$$E(X^r) = \int_0^\infty x^r g(x)dx = \frac{\theta}{\theta+3}\int_0^\infty x^r e^{-\theta x}(2+\theta+\theta x)dx.$$

Hence,

$$\mu'_r = \frac{r!}{\theta^r}\frac{(\theta+r+3)}{(\theta+3)}, \qquad r = 1, 2, 3, ... \tag{3.1}$$

First four moments about origin are obtained as:

$$\mu'_1 = \frac{1}{\theta}\frac{(\theta+4)}{(\theta+3)}, \qquad \mu'_2 = \frac{2}{\theta^2}\frac{(\theta+5)}{(\theta+3)}, \qquad \mu'_3 = \frac{6}{\theta^3}\frac{(\theta+6)}{(\theta+3)}, \qquad \mu'_4 = \frac{24}{\theta^4}\frac{(\theta+7)}{(\theta+3)}.$$

Using the above expression we get the four moments about mean, i.e. central moments of the proposed distribution are given by

$$\mu_1 = \frac{\theta+4}{\theta(\theta+3)}, \qquad \mu_2 = \frac{\theta^2+8\theta+14}{\theta^2(\theta+3)^2},$$
$$\mu_3 = \frac{2\big(\theta^3+12\theta^2+42\theta+46\big)}{\theta^3(\theta+3)^3}, \qquad \mu_4 = \frac{3\big(3\theta^4+48\theta^3+260\theta^2+592\theta+488\big)}{\theta^4(\theta+3)^4}.$$

The coefficient of variation (CV), coefficient of skewness $\sqrt{\beta_1}$, coefficient of kurtosis β_2 and index of dispersion γ of proposed distribution are obtained as:

$$CV = \frac{\sigma}{\mu_1} = \frac{\sqrt{\theta^2+8\theta+14}}{\theta+4}, \qquad \sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{2\big(\theta^3+12\theta^2+42\theta+46\big)}{(\theta^2+8\theta+14)^{\frac{3}{2}}},$$
$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\big(3\theta^4+48\theta^3+260\theta^2+592\theta+488\big)}{(\theta^2+8\theta+14)^2}, \qquad \gamma = \frac{\mu_2}{\mu_1} = \frac{\big(\theta^2+8\theta+14\big)}{\theta(\theta+3)(\theta+4)}.$$

The coefficient of variation (CV), index of dispersion (γ), coefficient of skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) are calculated for different values of θ. Coefficient of variation (CV) is observed less than 1 for all values of θ. Coefficient of skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) are found more than 1 and 3 respectively for different values of θ, therefore the proposed distribution is positively skewed and leptokurtic. The index of dispersion (γ) shows that the proposed distribution is under-dispersed as well as over-dispersed. It is observed that for θ= 1.1474, the value of γ is 1. For $\theta > 1.1474$, the distribution is under-dispersed and for $\theta < 1.1474$, it is over-dispersed. The graph for CV, γ, $\sqrt{\beta_1}$ and β_2 for different values of θ are shown in Figure 2.

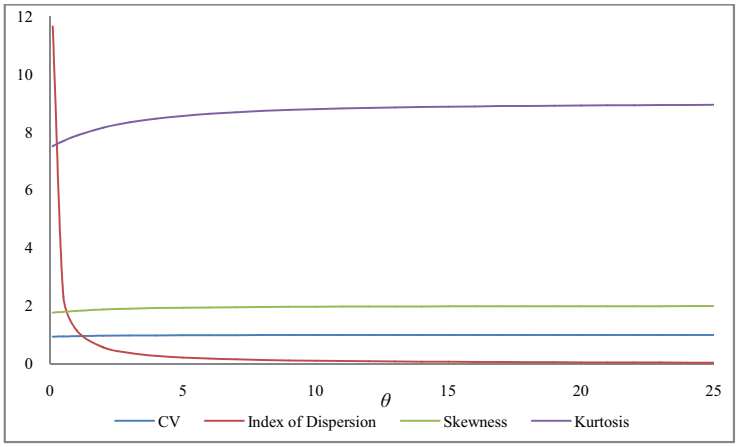

Figure 2: Graph of the CV, γ, β_1 and β_2 for different values of θ.

3.1. Generating functions

The moment generating function $M_x(t)$, characteristic function $\Phi_x(t)$ and cumulant generating function $\kappa_x(t)$ of proposed distribution are given by:

$$M_x(t) = \left[1 - \frac{(2+\theta)t}{(3+\theta)\theta}\right]\left(1 - \frac{t}{\theta}\right)^{-2}, \qquad \left|\frac{t}{\theta}\right| < 1, \tag{3.2}$$

$$\Phi_x(t) = \left[1 - \frac{(2+\theta)it}{(3+\theta)\theta}\right]\left(1 - \frac{it}{\theta}\right)^{-2}, \qquad i = \sqrt{-1}, \tag{3.3}$$

$$\kappa_x(t) = \log\left(1 - \frac{(2+\theta)it}{(3+\theta)\theta}\right) - 2\log\left(1 - \frac{it}{\theta}\right). \tag{3.4}$$

By series expansion of $\log(1-x) = -\sum_{r=0}^{\infty} \frac{x^r}{r}$, we get

$$\begin{aligned}\kappa_x(t) &= -\sum_{r=0}^{\infty}\left(\frac{(2+\theta)}{(3+\theta)\theta}\right)^r \frac{(it)^r}{r} + 2\sum_{r=0}^{\infty}\frac{\left(\frac{it}{\theta}\right)^r}{r} \\ &= 2\sum_{r=0}^{\infty}\frac{(r-1)!}{\theta^r}\frac{(it)^r}{r!} - \sum_{r=0}^{\infty}(r-1)!\left[\frac{\theta+2}{\theta(\theta+3)}\right]^r\frac{(it)^r}{r!}.\end{aligned}$$

Hence r-th cumulant of i-Garima distribution is given by

$$\begin{aligned}\kappa_r &= \text{coefficient of } \frac{(it)^r}{r!} \text{ in } \kappa_x(t) \\ &= 2\frac{(r-1)!}{\theta^r} - \frac{(r-1)!(\theta+2)^r}{[\theta(\theta+3)]^r}, \qquad r = 1, 2, 3, \ldots\end{aligned}$$

From the above equation we have four moments, that are the same as obtained earlier by equation (3.1):

$$\mu_1 = \kappa_1 = \frac{\theta+4}{\theta(\theta+3)}, \qquad \mu_2 = \kappa_2 = \frac{\theta^2+8\theta+14}{\theta^2(\theta+3)^2},$$

$$\mu_3 = \kappa_3 = \frac{2\left(\theta^3+12\theta^2+42\theta+46\right)}{\theta^3(\theta+3)^3}, \qquad \mu_4 = \kappa_4 + 3\kappa_2^2 = \frac{3\left(3\theta^4+48\theta^3+260\theta^2+592\theta+488\right)}{\theta^4(\theta+3)^4}.$$

3.2. Hazard rate and mean residual life function

Let X be a random variable with pdf $g(x)$ and cdf $G(x)$. The hazard function is given as

$$h(x) = \lim_{\Delta x \to \infty} \frac{P(X < x + \Delta x | X > x)}{\Delta x} = \frac{g(x;\theta)}{1 - G(x;\theta)}. \tag{3.5}$$

After using pdf and cdf of i-Garima distribution in above expression we get the hazard rate function $h(x)$ of i-Garima distribution as

$$h(x) = \frac{\theta(2 + \theta + \theta x)}{(3 + \theta + \theta x)}, \tag{3.6}$$

taking limit as $x \to 0$ in (3.6), we get

$$\lim_{x \to 0} h(x) = \lim_{x \to 0} \theta \left[1 - \frac{1}{(3 + \theta + \theta x)}\right] = \theta \left[1 - \frac{1}{(3 + \theta)}\right] > 0, \qquad \theta \in \mathbb{R}^+,$$

and for $x \to \infty$ we get

$$\lim_{x \to \infty} h(x) = \lim_{x \to \infty} \theta \left[1 - \frac{1}{(3 + \theta + \theta x)}\right] = \theta > 0, \qquad \theta \in \mathbb{R}^+.$$

Hence, $h(x) > 0$ for $x > 0$, $\theta > 0$. Therefore, $h(x)$ is an increasing function. The figure of hazard function is displayed in the Figure (3).

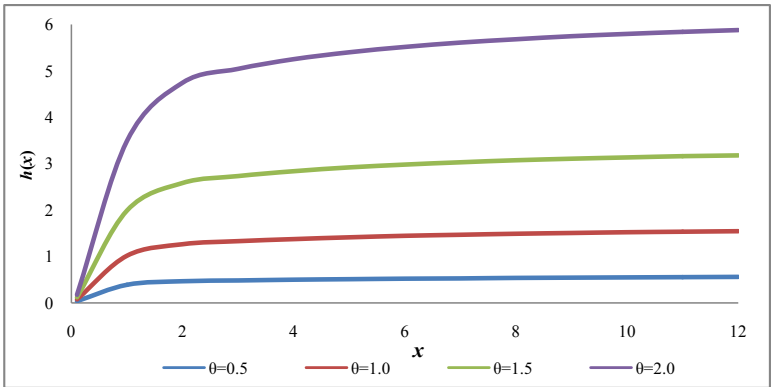

Figure 3: Hazard function of i-Garima distribution.

Now the mean residual life function (MRLF) is given as (3.7). We know that if a component of age t, the remaining lifetime after age t will be random. The expected value of the random life time is called the mean residual life and the mathematical form is known as MRLF. This may be more relevant than the hazard rate function in the study of repairable or replacement time. The MRLF provide idea about the entire residual life distribution or life expectancy, whereas the hazard rate is related only to the risk of immediate failure.

We have

$$m(x) = E[X - x|X > x] = \frac{1}{1 - G(x;\theta)} \int_x^\infty [1 - G(t;\theta)]dt,$$

$$m(x) = \frac{(4 + \theta + \theta x)}{\theta(3 + \theta + \theta x)}. \tag{3.7}$$

If $x = 0$, we get, $m(0) = \frac{\theta+4}{\theta(\theta+3)}$ which is $E(X)$ of the proposed distribution and also $m(x)$ is decreasing function for all $x > 0$ and $\theta > 0$. The graph of MRLF of i-Garima distribution is given in the Figure (4), which is decreasing type.

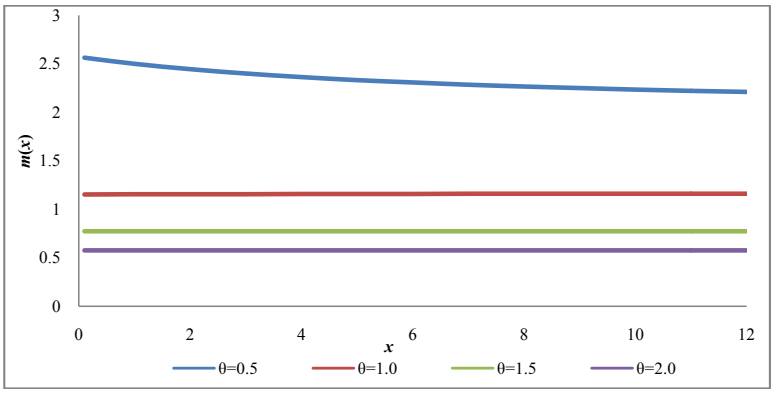

Figure 4: Mean residual life function (MRLF) of i-Garima distribution.

3.3. Quantile function

Theorem 3.1. *If $X \sim i$-Garima(θ), then Quantile function of X is defined as*

$$Q(p) = -1 - \frac{3}{\theta} - \frac{1}{\theta} W_{-1}\Big(-(1-p)(\theta+3)e^{-(-\theta+3)}\Big),$$

where $p \in (0,1)$ and W_{-1} is the negative branch of the Lambert W function.

Proof: Let

$$Q(p) = F^{-1}(p), \qquad p \in (0,1).$$

The quantile function, say $q(p)$, defined by $G(Q(p)) = p$ is the root of the equation

$$1 - \left(1 + \frac{\theta Q(p)}{\theta + 3}\right) e^{-\theta Q(p)} = p,$$

$$[3 + \theta + \theta Q(p)]e^{-\theta Q(p)} = (1-p)(\theta+3).$$

Multiplying both sides by $-e^{-(\theta+3)}$ we get

$$-[3+\theta+\theta Q(p)]e^{-(3+\theta+\theta Q(p))} = -(1-p)(\theta+3)e^{-(3+\theta)}.$$

Now $(3+\theta+\theta Q(p)) > 1$, $\forall\theta > 0$, $Q(p) > 0$. By applying W-function defined as the solution of the equation $w(z)e^{W(z)} = z$, the above equation can be written as

$$W_{-1}\Big(-(1-p)(\theta+3)e^{-(-\theta+3)}\Big) = -(3+\theta+\theta Q(p)),$$

where and $W_{-1}(\cdot)$ is the negative branch of the Lambert W function and we get the required result:

$$Q(p) = -1 - \frac{3}{\theta} - \frac{1}{\theta}W_{-1}\Big(-(1-p)(\theta+3)e^{-(-\theta+3)}\Big). \tag{3.8}$$

□

3.4. Stochastic ordering

Stochastic ordering of a continuous random variable is an important tool to judging their comparative behaviour. A random variable X is said to be smaller than a random variable Y, when:

(**i**) Stochastic order $X \leq_{st} Y$ if $F_X(x) \geq F_Y(x)$ for all x;

(**ii**) Hazard rate order $X \leq_{hr} Y$ if $h_X(x) \geq h_Y(x)$ for all x;

(**iii**) Mean residual life order $X \leq_{mrl} Y$ if $m_X(x) \geq m_Y(x)$ for all x;

(**iv**) Likelihood ratio order $X \leq_{lr} Y$ if $\frac{f_X(x)}{f_Y(x)}$ decreases in x.

The following results by Shaked and Shanthikumar [16] are well known for introducing stochastic ordering of distributions:

$$\begin{array}{ccccc} X \leq_{lr} Y & \Longrightarrow & X \leq_{hr} Y & \Longrightarrow & X \leq_{mrl} Y \\ & & \Downarrow & & \\ & & X \leq_{st} Y. & & \end{array}$$

With the help of the following theorem we claim that i-Garima distribution is ordered with respect to strongest likelihood ratio ordering.

Theorem 3.2. *Let $X \sim i$-Garima(θ_1) distribution and $Y \sim i$-Garima(θ_2) distribution. If $\theta_1 > \theta_2$ then $X \leq_{lr} Y$ and therefore $X \leq_{hr} Y$, $X \leq_{mrl} Y$ and $X \leq_{st} Y$.*

Proof: We have

$$\frac{f_X(x)}{f_Y(x)} = \frac{\theta_1(\theta_2+3)}{\theta_2(\theta_1+3)}\left(\frac{2+\theta_1+\theta_1 x}{2+\theta_2+\theta_2 x}\right)e^{-(\theta_1-\theta_2)x}, \qquad x > 0.$$

Now taking log both sides we get

$$\log\left[\frac{f_X(x)}{f_Y(x)}\right] = \log\left[\frac{\theta_1(\theta_2+3)}{\theta_2(\theta_1+3)}\right]\log\left[\frac{2+\theta_1+\theta_1 x}{2+\theta_2+\theta_2 x}\right] - (\theta_1-\theta_2)x.$$

By differentiating both sides we get

$$\frac{d}{dx}\log\left[\frac{f_X(x)}{f_Y(x)}\right] = \frac{2(\theta_1-\theta_2)}{(2+\theta_1+\theta_1 x)(2+\theta_2+\theta_2 x)} - (\theta_1-\theta_2).$$

Thus, for $\theta_1 > \theta_2$, $\frac{d}{dx}\log\left[\frac{f_X(x)}{f_Y(x)}\right] < 0$. This means that $X \leq_{lr} Y$ and hence $X \leq_{hr} Y$, $X \leq_{mrl} Y$ and $X \leq_{st} Y$. □

3.5. Order statistics

Let $X_1, X_2, ..., X_m$ be a random sample of size m from i-Garima distribution and also let $X_{(1)}, X_{(2)}, ..., X_{(m)}$ be the corresponding order statistics. The pdf and cdf of r-th order statistics, say $Y = X_{(r)}$, are given by

$$\begin{aligned} f_{(r:m)}(y) &= \frac{m!}{(r-1)!(m-r)!}F^{r-1}(y)[1-F(y)]^{m-r}f(y) \\ &= \frac{m!}{(r-1)!(m-r)!}\sum_{l=0}^{m-r}\binom{m-r}{l}(-1)^l F^{r+l-1}(y)f(y) \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} F_{(r:m)}(y) &= \sum_{j=r}^{m}\binom{m}{j}F^j(y)[1-F(y)]^{m-j} \\ &= \sum_{j=r}^{m}\sum_{l=0}^{m-j}\binom{m}{j}\binom{m-j}{l}(-1)^l F^{j+l}(y) \end{aligned} \tag{3.10}$$

respectively, for $r = 1(1)\, m$.

Based on equations (3.9) and (3.10) the pdf and cdf of r-th order statistics of i-Garima distribution is given in equations (3.11) and (3.12):

$$f_{(r:m)}(y) = \frac{m!\theta(3+\theta+\theta x)e^{-\theta x}}{(\theta+3)(r-1)!(m-r)!}\sum_{l=0}^{m-r}\binom{m-r}{l}\left[1-\frac{\theta x+(\theta+3)}{(\theta+3)}e^{-\theta x}\right]^{r+l-1} \tag{3.11}$$

and

$$F_{(r:m)}(y) = \sum_{j=r}^{m}\sum_{l=0}^{m-j}\binom{m}{j}\binom{m-j}{l}\left[1-\frac{\theta x+(\theta+3)}{(\theta+3)}e^{-\theta x}\right]^{j+l}. \tag{3.12}$$

3.6. Bonferroni and Lorenz curves

Let the random variable X is non-negative with a continuous and twice differentiable cumulative function. The Bonferroni [3] curve of the random variable X is defined as

$$B(p) = \frac{1}{p\mu}\int_0^q xg(x)dx = \frac{1}{p\mu}\left[\int_0^\infty xg(x)dx - \int_q^\infty xg(x)dx\right] = \frac{1}{p\mu}\left[\mu - \int_q^\infty xg(x)dx\right] \tag{3.13}$$

and the Lorenz curve (see Lorenz [12]) is defined by

$$L(p) = \frac{1}{\mu}\int_0^q xg(x)dx = \frac{1}{\mu}\left[\int_0^\infty xg(x)dx - \int_q^\infty xg(x)dx\right] = \frac{1}{\mu}\left[\mu - \int_q^\infty xg(x)dx\right] \tag{3.14}$$

where $q = G^{-1}(p)$ and $\mu = E(X)$, $p \in (0, 1]$.

The Gini index is given by

$$G = 1 - \frac{1}{\mu}\int_0^\infty (1 - G(x))^2 dx = \frac{1}{\mu}\int_0^\infty G(x)(1 - G(x))dx. \tag{3.15}$$

The Bonferroni, Lorenz curve and Gini index have application not only in economics to study income and poverty, but also in other fields like reliability, population studies, insurance, and medicine. Using the equations (3.13), (3.14) and (3.15) we get the Bonferroni curve, Lorenz curve and the Gini index as:

$$B(p) = \frac{1}{p}\left[1 - \frac{\{\theta^2 q^2 + (\theta^2 + 4\theta)q + (\theta + 4)\}e^{-\theta q}}{\theta + 4}\right], \tag{3.16}$$

$$L(p) = 1 - \frac{\{\theta^2 q^2 + (\theta^2 + 4\theta)q + (\theta + 4)\}e^{-\theta q}}{\theta + 4}, \tag{3.17}$$

$$G = \frac{2\theta^2 + 16\theta + 29}{4(\theta + 3)(\theta + 4)}. \tag{3.18}$$

4. ENTROPIES

Entropy, measures the variation in uncertainties associated with a random variable of a probability distributions. Rényi's and Shannon entropy are widely used to understand the uncertainty involved in random variables.

4.1. Rényi entropy

If X is a continuous random variable having probability density function $g(\cdot)$, then the Rényi Entropy (see Rényi [15]) is defined as

$$e(\eta) = \frac{1}{1 - \eta}\log\left[\int_0^\infty g^\eta(x)dx\right], \tag{4.1}$$

where $\eta > 0$ and $\eta = 0$.

The Rényi entropy for the i-Garima distribution is defined as

$$e(\eta) = \frac{1}{1-\eta}\log\left[\int_0^\infty \left(\frac{\theta}{\theta+3}\right)^\eta (2+\theta+\theta x)^\eta e^{-\eta\theta x}dx\right]$$

$$= \frac{1}{1-\eta}\log\left[\int_0^\infty \frac{\theta^\eta(\theta+2)^\eta}{(\theta+3)^\eta}\left(1+\frac{\theta x}{\theta+2}\right)^\eta e^{-\eta\theta x}dx\right]. \tag{4.2}$$

Now from the above equation (4.2), applying binomial expansion $(1+x)^n = \sum_{k=0}^{n}\binom{n}{k}x^k$, we get

$$\frac{1}{1-\eta}\log\left[\int_0^\infty \frac{\theta^\eta(\theta+2)^\eta}{(\theta+3)^\eta}\sum_{j=0}^{\eta}\binom{\eta}{j}\left(\frac{\theta x}{\theta+2}\right)^j e^{-\eta\theta x}dx\right],$$

$$\text{i.e.}\quad \frac{1}{1-\eta}\log\left[\sum_{j=0}^{\eta}\binom{\eta}{j}\frac{\theta^{\eta+j}(\theta+2)^{\eta-j}}{(\theta+3)^\eta}\int_0^\infty x^j e^{-\eta\theta x}dx\right]. \tag{4.3}$$

After solving equation (4.3), we get the required results in equation (4.4):

$$= \frac{1}{1-\eta}\log\left[\sum_{j=0}^{\eta}\binom{\eta}{j}\frac{\theta^{\eta-1}(\theta+2)^{\eta-j}}{(\theta+3)^\eta}\frac{\Gamma(j+1)}{(\eta)^{j+1}}\right], \tag{4.4}$$

since $\int_0^\infty x^{n-1}e^{-\theta x}dx = \frac{\Gamma(n)}{\theta^n}$.

4.2. Shannon entropy

The Shannon entropy (see Shannon [22]) of i-Garima distribution is given as

$$\Omega = E(-\log x) = -\int_0^\infty \log(f(x))f(x)dx$$

$$= -\log\left(\frac{\theta}{\theta+3}\right)\int_0^\infty f(x)dx - \int_0^\infty \log(2+\theta+\theta x)f(x)dx + \int_0^\infty \theta x f(x)dx$$

$$= -\log\left(\frac{\theta}{\theta+3}\right) - \log(\theta+2) - \int_0^\infty \log\left(1+\frac{\theta x}{\theta+2}\right)f(x)dx + \theta E(x). \tag{4.5}$$

Here, $E(X) = \frac{\theta+4}{\theta(\theta+3)}$, mean of the distribution. Applying $\log(1+x) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^n}{n}$, $|x|<1$, in equation (4.5), we get

$$= -\log\left(\frac{\theta(\theta+2)}{\theta+3}\right) + \left(\frac{\theta+4}{\theta+3}\right) - \frac{\theta}{\theta+3}\int_0^\infty \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\left(\frac{\theta x}{\theta+2}\right)^k (2+\theta+\theta x)e^{-\theta x}dx$$

$$= -\log\left(\frac{\theta(\theta+2)}{\theta+3}\right) + \left(\frac{\theta+4}{\theta+3}\right) - \frac{\theta}{\theta+3}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\left(\frac{\theta}{\theta+2}\right)^k \int_0^\infty x^k(2+\theta+\theta x)e^{-\theta x}dx.$$

After the simplification above, we obtained Shannon entropy as

$$\Omega = \left(\frac{\theta+4}{\theta+3}\right) - \log\left(\frac{\theta(\theta+2)}{\theta+3}\right) - \frac{1}{\theta+3}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\frac{k!(\theta+k+3)}{(\theta+2)^k}. \tag{4.6}$$

5. STRESS-STRENGTH RELIABILITY

Stress-strength model describes the life of a system of component having a random strength X and random stress Y. If stress is more than strength, the system of component fails immediately. The measure of system reliability $R = P(Y < X)$ is also known as stress-strength parameter. It is used in engineering science such as deterioration of any structures, motors, static fatigue of ceramic components and aging of concrete pressure vessels.

Let X and Y be independently distributed, with $X \sim i$-Garima(θ_1) and $Y \sim i$-Garima(θ_2). The CDF F_1 of X and pdf f_2 of Y are obtained from equations (2.7) and (2.6), respectively. Then stress-strength reliability R is obtained as

$$\begin{aligned} R &= P(Y < X) = \int_0^{\infty} P(Y < X | X = x) f_x(X) dx = \int_0^{\infty} f(x;\theta_1) F(x;\theta_2) dx \\ &= 1 - \frac{\theta_1\left[(\theta_1\theta_2 + 3\theta_1 + 2\theta_2 + 6)(\theta_1+\theta_2)^2 + (2\theta_1\theta_2 + 3\theta_1 + 2\theta_2)(\theta_1+\theta_2) + 2\theta_1\theta_2\right]}{(\theta_1+3)(\theta_2+3)(\theta_1+\theta_2)^3}. \end{aligned} \tag{5.1}$$

6. MAXIMUM LIKELIHOOD ESTIMATION

Let $(x_1, x_2, ..., x_n)$ be a random sample from $X \sim i$-Garima(θ). The likelihood function, L, is obtained as

$$L = \left(\frac{\theta}{\theta+3}\right)^n \prod_{i=1}^{n}(2+\theta+\theta x_i) e^{-\theta\sum_{i=1}^{n} x_i}. \tag{6.1}$$

Taking log both sides of equation (6.1) we get

$$\log L = n\log\left(\frac{\theta}{\theta+3}\right) + \sum_{i=1}^{n}\log(2+\theta+\theta x_i) - \theta\sum_{i=1}^{n} x_i. \tag{6.2}$$

Now differentiating both sides of (6.2) by θ we get

$$\frac{d(\log L)}{d\theta} = \frac{3n}{\theta^2+3\theta} + \sum_{i=1}^{n}\frac{1+x_i}{2+\theta+\theta x_i} - n\bar{x} = 0, \tag{6.3}$$

where $\bar{x}$ is the sample mean. The maximum likelihood estimate $(\hat{\theta})$ of θ is the solution of the equation (6.3). Since this is a non-linear equation, thus we solve this by numerical method.

7. EMPIRICAL ILLUSTRATIONS AND GOODNESS OF FIT

In this section, we present applications of the proposed distribution and their competent models for two real data sets to illustrate their potentiality. We estimate the unknown parameters of the model by the maximum likelihood estimation (MLE) using Newton–Raphson method. First data is about vinyl chloride obtained from clean up gradient monitoring wells in mg/l, provided by Bhaumik *et al.* [2], and second data set represents completed remission times (in months) of a random sample of 128 bladder cancer patients reported in Lee and Wang [10]. The summary measures of the two data sets are given below in Table 1.

Table 1: Summary measures of two data sets.

Datasets	n	mean	sd	median	skewness	kurtosis	min	max
1st data set	34	1.953	1.879	1.150	1.604	5.005	0.10	8.000
2nd data set	128	9.209	10.40	6.280	3.399	19.39	0.08	79.05

Table 1 reveals that both data sets are positively skewed and leptokurtic. First data set is under-dispersed however second data set is over-dispersed. We applied the i-Garima distributions for the above data sets and compared the results with some other competent distributions (see Lindley [11], Shanker [17, 18, 19, 20, 21]).

The goodness of fit of the i-Garima distribution has been explained for two real data sets using $-2LL$ ($-2\log$ likelihood), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and K-S Statistic (Kolmogorov-Smirnov Statistic). The estimate of these have been computed and shown in Tables 2 and 3, respectively. Smaller values of the AIC and BIC indicates better model fittings. The formulae for computing AIC, BIC, and K-S Statistics are as follows:

$$AIC = -2LL + 2k, \qquad BIC = -2LL + k\log n, \qquad D = \sup_x |F_n(x) - F_0(x)|,$$

where k = the number of parameters, n = the sample size, and $F_n(x)$ = empirical distribution function.

Table 2 and 3 reveals that i-Garima distribution provides closer fit for both data sets as it has lower -2LL, AIC, BIC, K-S values and higher p-values corresponding to K-S statistics than the other competitor models. Therefor, the proposed distribution i-Garima will consider as a potential alternative in modeling life time data and can be recommended for modelling data from engineering, medical, biological science and other applied sciences.

Table 2: MLE's, −2LL, AIC, BIC, K-S and p-values of the fitted distributions for the vinyl chloride dataset given by Bhaumik *et al.* [2].

Distribution	Estimate	−2LL	AIC	BIC	K-S	p-value
i-Garima	0.674	111.18	113.18	114.71	0.1039	0.8567
Garima	0.723	111.50	113.50	115.03	0.1135	0.7731
Aradhana	1.133	116.06	118.06	119.59	0.1695	0.2826
Sujatha	1.146	115.54	117.54	119.07	0.1640	0.3196
Akash	1.166	115.15	117.15	118.68	0.1564	0.3762
Shanker	0.853	112.91	114.91	116.44	0.1308	0.6062
Lindley	0.199	112.61	114.61	116.13	0.1326	0.5881

Table 3: MLE's, −2LL, AIC, BIC, K-S and p-values of the fitted distributions for the bladder cancer patients data given by Lee and Wang [10].

Distribution	Estimate	−2LL	AIC	BIC	K-S	p-value
i-Garima	0.143	825.57	827.57	830.42	0.0768	0.4374
Garima	0.158	826.49	828.49	831.34	0.0873	0.2835
Aradhana	0.295	868.28	870.28	873.13	0.1713	0.0011
Sujatha	0.303	873.22	875.22	878.08	0.1792	0.0005
Akash	0.315	881.04	883.04	885.89	0.1904	0.0002
Shanker	0.214	841.68	843.68	846.53	0.1243	0.0382
Lindley	0.199	833.79	835.79	838.64	0.1114	0.0832

8. CONCLUSION

Better modeling of the survival data is a major concern for statisticians and applied researchers. As a consequence, a significant progress has been made towards the extension of lifetime models and their application to various data sets. The present study suggests a technique for developing new probability distribution. A Single parameter distribution named i-Garima, is suggested and investigated in this study. Different statistical properties have been derived and studied for the proposed model. Moments about origin and mean have been obtained. The nature of pdf, cdf, hazard rate function and mean residual life function have been measured. The expression of stress-strength reliability is obtained, we can calculate system reliability when stress and strength parameter is known. Bonferroni, Lorenz curves and Gini index of the i-Garima are also measured. Maximum likelihood estimator of the model parameter is derived and obtained through Newton-Raphson method. The Rényi and Shannon entropies, order statistics and stochastic ordering are derived. An application of i-Garima distribution is given using two real lifetime data sets to show the suitability and the goodness of fit. Although the second data set have some censored cases but here we use only completed cases for the analysis. i-Garima provides a better fit over Garima, Aradhana, Sujatha, Akash, Shanker and Lindley distributions. It is realized that the proposed distribution in this study will consider some data sets in view of different censored mechanisms when specific interest comes into survival or reliability aspects. The article also opens a

scope for studying under Bayesian paradigm of the parameters under different loss functions. The work in this direction will perform in near future.

ACKNOWLEDGMENTS

Authors extend their appreciation to the associated editor and anonymous referees for their productive suggestions to improve the quality of manuscript.

REFERENCES

[1] Abd-Elrahman, A.M. (2013). Utilizing ordered statistics in lifetime distributions production: a new lifetime distribution and applications, *Journal of Probability and Statistical Science*, **11**(2), 153–164.

[2] Bhaumik, D.K.; Kapur, K. and Gibbons, R.D. (2009). Testing parameters of a gamma distribution for small samples, *Technometrics*, **51**(3), 326–334.

[3] Bonferroni, C.E. (1936). Statistical theory of classification and calculation of the probability, *Publication of the R Institute Superiore of the Science of Economy and Commerce of Florence*, **8**, 3–62.

[4] Cordeiro, G.M.; Castellares, F.; Montenegro, L.C. and de Castro, M. (2013). The beta generalized gamma distribution, *Statistics*, **47**(4), 888–900.

[5] Eugene, N.; Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications, *Communications in Statistics – Theory and Methods*, **31**(4), 497–512.

[6] Ghitany, M.E.; Alqallaf, F.; Al-Mutairi, D.K. and Husain, H.A. (2011). A two-parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in Simulation*, **81**(6), 1190–1201.

[7] Gupta, R.C.; Gupta, P.L. and Gupta, R.D. (1998). Modeling failure time data by Lehman alternatives, *Communications in Statistics – Theory and Methods*, **27**(4), 887–904.

[8] Gupta, R.C. and Kirmani, S.N.U.A. (1990). The role of weighted distributions in stochastic modeling, *Communications in Statistics – Theory and Methods*, **19**(9), 3147–3162.

[9] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes, *Journal of Hydrology*, **46**(1-2), 79–88.

[10] Lee, E.T. and Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis*, 3rd ed., Hoboken, NJ, Wiley.

[11] Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society. Series B (Methodological)*, 102–107.

[12] Lorenz, M.O. (1905). Methods of measuring the concentration of wealth, *Publications of the American Statistical Association*, **9**(70), 209–219.

[13] Nadarajah, S.; Cordeiro, G.M. and Ortega, E.M. (2015). The Zografos–Balakrishnan-G family of distributions: mathematical properties and applications, *Communications in Statistics – Theory and Methods*, **44**(1), 186–215.

[14] Patil, G.P. and Rao, C.R. (1977). The weighted distributions: a survey of their applications, *Applications of Statistics*, 383–405.

[15] Rényi, A. (1961). On measures of entropy and information, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1: "Contributions to the Theory of Statistics", 547–561, University of California Press.

[16] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*, Springer Science & Business Media.

[17] Shanker, R. (2015). Akash distribution and its applications, *International Journal of Probability and Statistics*, **4**(3), 65–75.

[18] Shanker, R. (2015). Shanker distribution and its applications, *International Journal of Statistics and Applications*, **5**(6), 338–348.

[19] Shanker, R. (2015). Aradhana distribution and its applications, *International Journal of Sstatistics and Applications*, **6**(1), 23–28.

[20] Shanker, R. (2016). Garima distribution and its application to model behavioral science data, *Biometrics & Biostatistics International Journal*, **4**(7), 275–281.

[21] Shanker, R. (2016). Sujatha distribution and its applications, *Statistics in Transition. New Series*, **17**(3), 391–410.

[22] Shannon, C.E. (1951). Prediction and entropy of printed English, *Bell System Technical Journal*, **30**(1), 50–64.

[23] Shaw, W.T. and Buckley, I.R. (2007). The alchemy of probability distributions: beyond Gram-Charlier & Cornish-Fisher expansions, and skew-normal or kurtotic-normal distributions, *Submitted, Feb 7*, **64**.

[24] Singh, B.P.; Dhar Das, U. and Singh, S. (2021). A compounded probability model for decreasing hazard and its inferential properties, *Reliability: Theory & Applications*, **16**(2(62)), 230–246.

[25] Singh, B.P.; Singh, S. and Dhar Das, U. (2020). A general class of new continuous mixture distribution and application, *Journal of Mathematical and Computational Science*, **11**(1), 585–602.

[26] Zakerzadeh, H. and Dolati, A. (2009). Generalized Lindley distribution, *Journal of Mathematical Extension*, **3**(2), 13–25.

[27] Zografos, K. and Balakrishnan, N. (2009). On families of beta-and generalized gamma-generated distributions and associated inference, *Statistical Methodology*, **6**(4), 344–362.

REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 321–345
https://doi.org/10.57805/revstat.v21i3.356

On Some Stationary INAR(1) Processes with Compound Poisson Distributions

Authors: Emad-Eldin A. A. Aly
– Department of Statistics and Operations Research, Kuwait University, Kuwait
emadeldin.ali@ku.edu.kw

Nadjib Bouzar ✉
– Department of Mathematical Sciences, University of Indianapolis, USA
nbouzar@uindy.edu

Received: January 2021 Revised: December 2021 Accepted: December 2021

Abstract:

- Aly and Bouzar ([2]) used the backward approach in presence of the binomial thinning operator to construct underdispersed stationary first-order autoregressive integer-valued (INAR (1)) processes. The present paper is to be seen as a continuation of their work. The focus of this paper is on the development of stationary INAR (1) processes with discrete compound Poisson innovations. We expand on some recent results obtained by several authors for these processes. A number of theoretical results are established and then used to develop stationary INAR (1) processes with compound Poisson innovations with finite mean. We apply our results to obtain in detail important distributional properties of the new models when the innovation follows the Polya–Aeppli distribution, the non-central Polya–Aeppli distribution, the negative binomial distribution, the noncentral negative binomial distribution, the Poisson–Lindley distribution, the Euler-type distribution and the Euler distribution.

Keywords:

- *integer-valued time series; the binomial thinning operator; infinite divisibility; Euler distribution.*

AMS Subject Classification:

- 62M10, 60E99.

✉ Corresponding author.

1. INTRODUCTION

Assume that X is a $\mathbb{Z}_+$-valued random variable (rv) and $\alpha \in (0,1)$. The binomial thinning operator (Steutel and van Harn ([22])) of X, denoted by $\alpha \odot X$, is defined by

$$\alpha \odot X = \sum_{i=1}^{X} Y_i, \tag{1.1}$$

where $\{Y_i\}$ is a sequence of independent identically distributed (iid) Bernoulli(α) ($Ber(\alpha)$) rv's independent of X. The operation $\odot$ acts as the analogue of the standard multiplication used in standard time series models.

The main results of this paper use the two facts below without further reference. For α and β in $(0,1)$,

$$\alpha \odot (\beta \odot X) \stackrel{d}{=} \beta \odot (\alpha \odot X) \stackrel{d}{=} (\alpha\beta) \odot X$$

and for X and Y independent $\mathbb{Z}_+$-valued rv's,

$$\alpha \odot (X+Y) \stackrel{d}{=} \alpha \odot X + \alpha \odot Y.$$

Assume that $\{\varepsilon_t\}$ is a sequence of iid $\mathbb{Z}_+$-valued rv's. A sequence $\{X_t\}$ of $\mathbb{Z}_+$-valued rv's is said to be an INAR (1) process if

$$X_t = \alpha \odot X_{t-1} + \varepsilon_t \qquad (t \geq 1), \tag{1.2}$$

where $\{\varepsilon_t\}$ is the innovation sequence and α is the coefficient of the process. The binomial thinning $\alpha \odot X_{t-1}$ in (1.2) is performed independently for each t. More precisely, we assume the existence of an array $(Y_{i,t},\ i \geq 1,\ t \geq 0)$ of iid $Ber(\alpha)$ rv's, independent of $\{\varepsilon_t\}$, such that

$$\alpha \odot X_{t-1} = \sum_{i=1}^{X_{t-1}} Y_{i,t-1}.$$

Let $\varphi_{X_t}(z)$ be the pgf of X_t of (1.2) and $\Psi(z)$ be the pgf ε_t. Then we have by (1.2)

$$\varphi_{X_t}(z) = \varphi_{X_{t-1}}(1-\alpha+\alpha z)\Psi(z).$$

If one further assumes that $\{X_t\}$ is stationary with $\varphi_X(z)$ as the pgf of its marginal distribution, then the following functional equation holds

$$\varphi_X(z) = \varphi_X(1-\alpha+\alpha z)\Psi(z). \tag{1.3}$$

It is a well known result that if $\alpha \in (0,1)$ and $\varphi_X(z)$ and $\Psi(z)$ are pgf's that satisfy (1.3), then there exists a stationary INAR (1) process $\{X_t\}$ on some probability space such that $\varphi_X(z)$ and $\Psi(z)$ are respectively the pgf of its marginal distribution and the pgf of its innovation sequence $\{\varepsilon_t\}$.

In the backward approach, one starts out with the pgf $\Psi(z)$ of the innovation sequence and solve (1.3) for the pgf $\varphi_X(\cdot)$ of the marginal distribution of the INAR (1) process. In this case

$$\varphi_X(z) = \lim_{n \longrightarrow \infty} \prod_{i=0}^{n} \Psi(1-\alpha^i+\alpha^i z),$$

provided that the limit exists and is a pgf (see [2]).

The main focus of the present paper is on the development of stationary INAR (1) models driven by (1.2) with an infinitely divisible (Compound Poisson) innovation whose mean is finite. In Section 2, we prove a number of basic results in the context of the backward approach for these models. The results of Section 2 are used in Sections 3–9 to obtain in detail key distributional properties of the marginal distributions of some important INAR (1) processes. We discuss models whose innovations follow the Polya–Aeppli distribution, the non-central Polya–Aeppli distribution, the negative binomial distribution, the noncentral negative binomial distribution, the Poisson–Lindley distribution, and the Euler-type and Euler distributions.

The above INAR (1) models are necessarily overdispersed. An example of a data set which is empirically overdispersed is presented and analyzed in [4]. This data set gives the monthly claim counts by workers in the heavy manufacturing industry who were collecting benefits due to a burn related injury. The same data set was further analyzed in [23] and [18] and shown to have an INAR (1)-like autocorrelation structure. Another example of an overdispersed data set was introduced in [11] and was further analyzed in [12]. This data set involves the number of publications produced by Ph.D. biochemists. Several examples of underdispersed data sets are reported and analyzed in [20].

In the rest of this paper we will assume that $\alpha \in (0,1)$ and $\overline{a} = 1 - a$ for $a \in (0,1)$. We will also use the notation $\mu_r^{(u)}(\kappa_r^{(u)})$ and $\mu_{[r]}^{(u)}(\kappa_{[r]}^{(u)})$ to designate the r-th moment (cumulant) and the r-th factorial moment (factorial cumulant) of the pmf $\{u_r\}$, respectively.

The backward approach rests heavily on the following important result found in [2].

Theorem 1.1. *Assume that* $\Psi'(1) < \infty$. *The function*

$$\varphi(z) = \prod_{i=0}^{\infty} \Psi(1 - \alpha^i + \alpha^i z) \tag{1.4}$$

is a pgf. Moreover, the convergence of the infinite product is uniform over the interval $[0,1]$ *and* $\varphi(z)$ *satisfies (1.3).*

2. PROCESSES WITH COMPOUND POISSON INNOVATIONS

2.1. Basic Results

We start out by specializing Theorem 1.1 to infinitely divisible distributions with finite mean. Recall that a distribution on $\mathbb{Z}_+$ is infinitely divisible if and only if it is a discrete compound Poisson distribution with pgf

$$\Psi(z) = \exp\{\lambda(H(z) - 1)\}, \tag{2.1}$$

for some $\lambda > 0$ and some unique pgf $H(z) = \sum_{r=1}^{\infty} h_r z^r$ with pmf $\{h_r\}$ and $H(0) = h_0 = 0$. We will refer to such distributions as $DCP(\lambda, H)$ distributions.

First, we need a lemma.

Lemma 2.1. *Assume that $\Psi(z)$ is the pgf of a $DCP(\lambda, H)$ distribution. Then for each $i \geq 0$, $\Psi(1-\alpha^i+\alpha^i z)$ is the pgf of a $DCP(\lambda_i', H_i)$ distribution which is described below:*

(**i**) *For every $i \geq 0$,*

$$\lambda_i' = \lambda m_i, \quad m_i = 1 - H(1-\alpha^i), \tag{2.2}$$

and

$$H_i(z) = 1 - \frac{1}{m_i}\big(1 - H(1-\alpha^i+\alpha^i z)\big). \tag{2.3}$$

(**ii**) *The pmf $\{h_r^{(i)}\}$ with pgf $H_i(z)$ is*

$$h_r^{(i)} = \frac{\alpha^{ir}}{m_i}\sum_{n=r}^{\infty}\binom{n}{r}(1-\alpha^i)^{n-r}h_n \quad (r \geq 1). \tag{2.4}$$

Note that $H_0(z) = H(z)$ and $\{h_r^{(0)}\} = \{h_r\}$.

(**iii**) *If the factorial moment generating function (fmgf) $H(1+t)$ of the pmf $\{h_r\}$ exists for $|t| < \rho_0$ for some $\rho_0 > 0$, then for every $i \geq 0$, the pmf $\{h_r^{(i)}\}$ has finite factorial moments $\{\mu_{[r]}^{(h^{(i)})}\}$ for all $r \geq 1$, and*

$$\mu_{[r]}^{(h^{(i)})} = \frac{\alpha^{ir}}{m_i}\mu_{[r]}^{(h)}. \tag{2.5}$$

Proof: By (2.1), we have $\ln\Psi(1-\alpha^i+\alpha^i z) = \lambda(H(1-\alpha^i+\alpha^i z)-1)$, $i \geq 0$, which can be rewritten as

$$\ln\Psi(1-\alpha^i+\alpha^i z) = \lambda(1-H(1-\alpha^i))\Big(\frac{H(1-\alpha^i+\alpha^i z)-H(1-\alpha^i)}{1-H(1-\alpha^i)} - 1\Big).$$

Letting m_i and λ_i' be as in (2.2), we have

$$\ln\Psi(1-\alpha^i+\alpha^i z) = \lambda_i'\Big(\frac{H(1-\alpha^i+\alpha^i z)+m_i-1}{m_i} - 1\Big),$$

which leads to (2.3). The identity $(a+b)^n = \sum_{r=0}^{n}\binom{n}{r}a^r b^{n-r}$ implies

$$H(1-\alpha^i+\alpha^i z) - H(1-\alpha^i) = \sum_{r=1}^{\infty}\Big(\sum_{n=r}^{\infty}\binom{n}{r}\alpha^{ir}(1-\alpha^i)^{n-r}h_n\Big)z^r.$$

Hence, $H_i(z)$ is the pgf of $\{h_r^{(i)}\}$ of (2.4). This establishes (i)–(ii). To prove (iii), we note that since the fmgf $H(1+t)$ of the pmf $\{h_r\}$ exists, then $\{h_r\}$ has finite factorial moments $\mu_{[r]}^{(h)}$ of all orders $r \geq 1$. It follows by equation (1.274), p. 59, in [6] and (2.3) that

$$H_i(1+t) = 1 + \frac{1}{m_i}\sum_{r=1}^{\infty}\mu_{[r]}^{(h)}\alpha^{ir}\frac{t^r}{r!} \qquad (|t| < \rho_0), \tag{2.6}$$

which in turn leads to (2.5). □

Next, we study the pgf $\varphi(\cdot)$ of (1.4) when $\Psi(z)$ is the pgf of a $DCP(\lambda, H)$ distribution.

Theorem 2.1. *Let $\varphi(\cdot)$ and $\Psi(\cdot)$ be as in (1.4). If $\Psi(z)$ is the pgf of a $DCP(\lambda, H)$ distribution with $\Psi'(1) < \infty$, then the following assertions hold:*

(**i**) *$\varphi(z)$ is the pgf of the infinite convolution of the distributions $(DCP(\lambda m_i, H_i),\ i \geq 0)$, as described in Lemma 2.1.*

(**ii**) *$\varphi(z)$ is the pgf of a $DCP(\widetilde{\lambda}, G)$ distribution, where*

$$\widetilde{\lambda} = \lambda M > 0, \quad M = \sum_{i=0}^{\infty} m_i = \sum_{i=0}^{\infty}(1 - H(1-\alpha^i)), \tag{2.7}$$

and

$$G(z) = \sum_{i=0}^{\infty} \frac{m_i}{M} H_i(z) \quad (G(0) = 0). \tag{2.8}$$

Moreover, the pmf $\{g_r\}$ with pgf $G(z)$ is the infinite countable mixture

$$g_r = \sum_{i=0}^{\infty} \frac{m_i}{M} h_r^{(i)} \quad (r \geq 1), \tag{2.9}$$

with $\left(\{h_r^{(i)}\}, i \geq 0\right)$ of (2.4) and mixing probabilities $\left(\frac{m_i}{M}, i \geq 0\right)$.

Proof: By Theorem 1.1, $\varphi(z)$ is a pgf. Part (i) follows directly from Lemma 2.1. To prove (ii), first we note $\Psi(z)$ is the pgf of an infinitely divisible distribution. Therefore, there exists a pgf $\Psi_n(z)$ such that $\Psi(z) = (\Psi_n(z))^n$ for every $n \geq 1$. Since $\Psi'(z) = n(\Psi_n(z))^{n-1}\Psi_n'(z)$ and $\Psi'(1) < \infty$, we have $\Psi_n'(1) < \infty$. Applying Theorem 1.1 to Ψ_n, it follows that $\prod_{i=0}^{\infty} \Psi_n(1 - \alpha^i + \alpha^i z)$ is a pgf. Note that

$$\varphi(z) = \prod_{i=0}^{\infty} \Psi(1 - \alpha^i + \alpha^i z) = \left\{\prod_{i=0}^{\infty} \Psi_n(1 - \alpha^i + \alpha^i z)\right\}^n \quad (n \geq 1).$$

Hence, $\varphi(z)$ is the the pgf of an infinitely divisible distribution, or a $DCP(\widetilde{\lambda}, G)$ distribution for some $\widetilde{\lambda} > 0$ and pgf $G(z)$. We have by Theorem 1.1 and (2.1)

$$\varphi(z) = \prod_{i=0}^{\infty} \Psi(1 - \alpha^i + \alpha^i z) = \exp\left\{\lambda \sum_{i=0}^{\infty}(H(1 - \alpha^i + \alpha^i z) - 1)\right\}.$$

It is clear that $\varphi'(1) < \infty$ implies $H'(1) < \infty$. Let $Q_H(z) = \frac{1-H(z)}{1-z}$ $(z \neq 1)$ be the generating function of the tail probabilities $q_r = \sum_{i=r+1}^{\infty} h_i$ of $\{h_r\}$. It follows that $1 - H(1 - \alpha^i + \alpha^i z) \leq \alpha^i H'(1)$ (recall $Q_H(1) = H'(1)$) and thus $\sum_{i=0}^{\infty}(1 - H(1 - \alpha^i + \alpha^i z))$ converges uniformly over $[0, 1]$. This implies that $M = \sum_{i=0}^{\infty} m_i < \infty$ (see (2.2)). The fact that $\widetilde{\lambda} = \lambda M$ follows by setting $z = 0$ in the equation $\lambda \sum_{i=0}^{\infty}(H(1 - \alpha^i + \alpha^i z) - 1) = \widetilde{\lambda}(G(z) - 1)$. Solving for $G(z)$ and using (2.3) leads to (2.8) and (2.9) follows from (2.4) and (2.8). □

The following result is a direct consequence of Theorem 2.1 and equation (9.43), p. 390, in [6], for infinitely divisible distributions.

Corollary 2.1. *Under the assumptions and notation of Theorem 2.1, the pmf* $\{p_r\}$ *with pgf* $\varphi(z)$ *can be derived via the recurrence formula*

$$(r+1)p_{r+1} = \lambda \sum_{j=0}^{r} (r+1-j) g_{r+1-j} p_j \quad \textit{with} \quad p_0 = e^{-\lambda M} \quad (r \geq 0). \tag{2.10}$$

Remark 2.1. A distribution on $\mathbb{Z}_+$ with pgf $\Psi(z)$ is discrete self-decomposable (DSD) (cf. Steutel and van Harn [22]) if for any $\beta \in (0,1)$,

$$\Psi(z) = \Psi(1-\beta+\beta z)\Psi_\beta(z), \tag{2.11}$$

for some pgf $\Psi_\beta(z)$. If $\Psi(z)$ is the pgf of a DSD distribution with finite mean, then $\varphi(z)$ of (1.4) is the pgf of a DSD distribution. Indeed, basic properties of infinite products and the fact that $\Psi'_\beta(1) < \infty$ lead to

$$\varphi(z) = \varphi(1-\beta+\beta z) \prod_{i=0}^{\infty} \Psi_\beta(1-\alpha^i+\alpha^i z).$$

We conclude by Theorem 1.1 applied to $\Psi_\beta(z)$ that $\prod_{i=0}^{\infty} \Psi_\beta(1-\alpha^i+\alpha^i z)$ is a pgf.

We proceed to discuss the case of INAR (1) processes with a $DCP(\lambda, H)$ innovation. We will add to results obtained in [18], [19] and [24]. These papers deal mainly with $DCP(\lambda, H)$ innovation when the compounding distribution has a pgf of the form $H(z) = \sum_{i=1}^{n} h_i z^i$, $n < \infty$. For example, on page 355 in [24], it is mentioned, quoting, "Let (X_t) be a stationary $CP_\infty - INAR(1)$ process. In general, a closed-form expression for the observations' pmf is not available". In addition, on page 624 in [19], it is mentioned that "the structural implications of Theorem 2.1 can be extended to the case of compound Poisson arrival distributions with an infinite compounding structure. The stationary distribution in this general case is again compound Poisson distributed with infinite compounding structure. However, a way to explicitly calculate the stationary distribution in this case is not known".

The next result asserts the existence of a stationary INAR (1) process whose innovation is DCP with infinite compounding structure. It is a consequence of Theorem 2.1 and the standard result on the existence of stationary INAR (1) processes recalled in the introduction. The proof is omitted.

Theorem 2.2. *Any* $DCP(\lambda, H)$ *distribution with pgf* $\Psi(z)$ *of (2.1) such that* $H'(1) < \infty$ *gives rise to a stationary INAR (1) process* $\{X_t\}$ *defined on some probability space and driven by equation (1.2). Its innovation has pgf* $\Psi(z)$ *and its marginal distribution is the* $DCP(\widetilde{\lambda}, G)$ *distribution described by (2.7)–(2.10).*

Next, we list key distributional properties of a stationary INAR (1) process $\{X_t\}$ with a $DCP(\lambda, H)$ innovation:

1. The 1-step transition probabilities of $\{X_t\}$ are given by

$$P(X_t = k | X_{t-1} = l) = \sum_{j=0}^{\min(l,k)} \binom{l}{j} \alpha^j (1-\alpha)^{l-j} f_{k-j}, \tag{2.12}$$

where

$$f_x = P(\varepsilon = x) = \begin{cases} e^{-\lambda}, & \text{if } x = 0, \\ \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} h_x^{*n}, & \text{if } x > 0, \end{cases} \tag{2.13}$$

and $\{h_x^{*n}\}$ is the n-fold convolution of the pmf $\{h_r\}$ with pgf $H(z)$. Similarly to (2.10), f_x can be obtained by the recurrence formula

$$(x+1)f_{x+1} = \lambda \sum_{j=0}^{x} (x+1-j) h_{x+1-j} f_j\,, \qquad \text{with} \quad f_0 = e^{-\lambda} \quad (x \geq 0). \tag{2.14}$$

2. The k-step-ahead version of (1.2) for $k \geq 1$ is given by

$$X_{t+k} \stackrel{d}{=} \alpha^k \odot X_t + \sum_{j=1}^{k} \alpha^{j-1} \odot \varepsilon_{t+k-j+1}. \tag{2.15}$$

Consequently, the conditional pgf of X_{t+k} given X_t satisfies

$$\varphi_{X_{t+k}|X_t}(z) = \left(1 - \alpha^k + \alpha^k z\right)^{X_t} \times \prod_{i=0}^{k-1} \Psi(1 - \alpha^i + \alpha^i z). \tag{2.16}$$

3. It follows by Lemma 2.1 and (2.16) that the conditional distribution of X_{t+k} given $X_t = n$ results from the convolution of a binomial distribution, $Bin(n, \alpha^k)$, and the distributions $(DCP(\lambda m_i, H_i),\ 0 \leq i \leq k-1)$ with characteristics (2.2)–(2.4).

4. Assume the fmgf $H(1+t)$ of the pmf $\{h_r\}$ exists for $|t| < \rho_0$ for some $\rho_0 > 0$. By Lemma 2.1-(iii), the fmgf $H_i(1+t)$ of the pmf $\{h_r^{(i)}\}$ admits the representation (2.6), for every $i \geq 0$ and $|t| < \rho_0$. Using (2.8) and a standard argument, one can show that $G(1+t) = \sum_{i=0}^{\infty} \frac{m_i}{M} H_i(1+t)$ converges uniformly in the interval $|t| \leq \rho$ for every $0 < \rho < \rho_0$. Therefore, by Weierstrass Theorem, p. 430 in [8], we have

$$G(1+t) = 1 + \sum_{r=1}^{\infty} \left[\sum_{i=0}^{\infty} \frac{m_i}{M} \mu_{[r]}^{(h^{(i)})} \right] \frac{t^r}{r!} \qquad (|t| < \rho_0),$$

which implies

$$\mu_{[r]}^{(g)} = \sum_{i=0}^{\infty} \frac{m_i}{M} \mu_{[r]}^{(h^{(i)})}. \tag{2.17}$$

By (2.5), (2.17) and equation (1.246), p. 53, in [6], the factorial moments and the moments of $\{g_r\}$ are

$$\mu_{[r]}^{(g)} = \frac{\mu_{[r]}^{(h)}}{M(1-\alpha^r)} \quad \text{and} \quad \mu_r^{(g)} = \frac{1}{M} \sum_{j=1}^{r} S(r,j) \frac{\mu_{[j]}^{(h)}}{1-\alpha^j} \qquad (r \geq 1), \tag{2.18}$$

where $\{S(r,j)\}$ are the Stirling numbers of the second kind defined as

$$S(r,j) = \frac{1}{j!} \sum_{k=0}^{j} (-1)^{j-k} \binom{j}{k} k^r \qquad (S(0,0) = 1,\ S(0,k) = S(r,0) = 0).$$

5. By (2.18), equations (9.49), p. 391, and (1.257), p. 55, in [6], the factorial cumulants and cumulants of X_t are:

$$\kappa_{[r]}^{(p)} = \frac{\lambda}{1-\alpha^r}\mu_{[r]}^{(h)} \quad \text{and} \quad \kappa_r^{(p)} = \lambda \sum_{j=1}^{r} S(r,j)\frac{\mu_{[j]}^{(h)}}{1-\alpha^j} \qquad (r \geq 1). \tag{2.19}$$

6. The first and second cumulants of a pmf are its mean and variance, respectively. The mean $\mu_1^{(p)}$ and the variance $(\sigma^{(p)})^2$ of X_t follow from the above formulas:

$$\mu_1^{(p)} = \frac{\lambda \mu_1^{(h)}}{1-\alpha} \quad \text{and} \quad (\sigma^{(p)})^2 = \frac{\lambda(\mu_2^{(h)} + \alpha\mu_1^{(h)})}{1-\alpha^2}. \tag{2.20}$$

7. The moments and factorial moments of X_t can be computed recursively by a formula in [21] for the former and equation (1.244) in [6] for the latter:

$$\mu_r^{(p)} = \sum_{i=0}^{r-1} \binom{r-1}{i} \kappa_{r-i}^{(p)} \mu_i^{(p)} \quad \text{and} \quad \mu_{[r]}^{(p)} = \sum_{j=0}^{r} s(r,j)\mu_j^{(p)}, \tag{2.21}$$

where $\{s(r,j)\}$ are the Stirling numbers of the first kind satisfying the recurrence relation

$$s(r+1,j) = s(r,j-1) - rs(r,j) \qquad (s(n,0) = 0,\ s(1,1) = 1).$$

We note that the moments and factorial moments of the marginal distributions of the INAR (1) models we introduce here are only obtainable through (2.21). Except for a couple of instances, we will make no further reference to these moments.

2.2. Processes whose innovations are convolutions of *DCP* distributions

We consider stationary INAR (1) processes whose innovation is the finite convolution of *DCP* distributions with finite means.

Let ν be a positive integer. We assume throughout the section that $(\widetilde{H}_k, 1 \leq k \leq \nu)$ is a collection of pgf's such that $\widetilde{H}_k(0) = 0$, $\widetilde{H}_k'(1) < \infty$ and $(\lambda_k, 1 \leq k \leq \nu)$ are positive constants. We denote by $\{h_r^{(k)}\}$ the pmf of $\widetilde{H}_k(z)$.

Lemma 2.2. *Let $\Psi_k(z)$ be the pgf of a $DCP(\lambda_k, \widetilde{H}_k)$ distribution, $1 \leq k \leq \nu$. The following assertions hold:*

(**i**) *The convolution of the $DCP(\lambda_k, \widetilde{H}_k)$ distributions, $1 \leq k \leq \nu$, is $DCP(\lambda, H)$, where*

$$\lambda = \sum_{k=1}^{\nu} \lambda_k \quad \text{and} \quad H(z) = \sum_{k=1}^{\nu} \frac{\lambda_k}{\lambda} \widetilde{H}_k(z). \tag{2.22}$$

(**ii**) *For each $k = 1, 2, ..., \nu$, $\Psi_k(1-\alpha^i+\alpha^i z)$ is the pgf of a $DCP(\lambda_k m_i^{(k)}, \widetilde{H}_{ki}(z))$ distribution, where $m_i^{(k)} = 1 - \widetilde{H}_k(1-\alpha^i)$ and $\widetilde{H}_{ki}(z)$ is the pgf of a pmf we denote $\{h_r^{(k,i)}\}$, with $\widetilde{H}_{ki}(0) = 0$ and $\widetilde{H}_{ki}'(1) < \infty$.*

(**iii**) $\Psi(1-\alpha^i+\alpha^i z)$ *is the pgf of a* $DCP(\lambda m_i, H_i)$ *distribution, where* $m_i = 1 - H(1-\alpha^i) = \sum_{k=1}^{\nu} \frac{\lambda_k}{\lambda} m_i^{(k)}$, *with* λ *and* H *of (2.22), and*

$$H_i(z) = \sum_{k=1}^{\nu} \frac{\lambda_k m_i^{(k)}}{\lambda m_i} \widetilde{H}_{ki}(z) \quad \text{and} \quad h_r^{(i)} = \sum_{k=1}^{\nu} \frac{\lambda_k m_i^{(k)}}{\lambda m_i} h_r^{(k,i)} \quad (r \geq 1). \tag{2.23}$$

(**iv**) *For every* $i \geq 0$, *the* $DCP(\lambda m_i, H_i)$ *distribution admits the following representation, with* $\lambda_i^{(k)} = \lambda_k m_i^{(k)}$ $(1 \leq k \leq \nu)$,

$$DCP(\lambda m_i, H_i) \sim DCP(\lambda_i^{(1)}, \widetilde{H}_{1i}) * DCP(\lambda_i^{(2)}, \widetilde{H}_{2i}) * \cdots * DCP(\lambda_i^{(\nu)}, \widetilde{H}_{\nu i}). \tag{2.24}$$

Proof: (i) is clear and (ii) follows from Lemma 2.1. For (iii), m_i follows from (2.22) by Theorem 2.1. We have by (i) $\Psi_k(1-\alpha^i+\alpha^i z) = \exp\{\lambda_k m_i^{(k)}(\widetilde{H}_{ki}(z)-1)\}$, which implies

$$\varphi(z) = \exp\Big\{\sum_{k=1}^{\nu} \lambda_k m_i^{(k)}(\widetilde{H}_{ki}(z)-1)\Big\} = \exp\Big\{\Big(\sum_{k=1}^{\nu} \lambda_k m_i^{(k)} \widetilde{H}_{ki}(z)\Big) - \lambda m_i\Big\}$$

and (2.23), as $\sum_{k=1}^{\nu} \frac{\lambda_k m_i^{(k)}}{\lambda m_i} = 1$. (iv) follows from (iii) and (2.23). □

Next, we present key distributional properties of a stationary INAR (1) with an innovation that is the convolution of DCP distributions. The proofs are omitted as the results are a direct consequence of Lemma 2.2 and Theorem 2.1.

Theorem 2.3. *Let* $\{X_t\}$ *be a stationary INAR (1) process driven by (1.2) with the* $DCP(\lambda, H)$ *innovation that results from the convolution of the* $DCP(\lambda_k, \widetilde{H}_k)$ *distributions,* $1 \leq k \leq \nu$ *(as described in Lemma 2.2). Let* $M_k = \sum_{i=0}^{\infty} m_i^{(k)}$, $1 \leq k \leq \nu$. *The following assertions hold:*

(**i**) *The marginal distribution of* $\{X_t\}$ *is the infinite convolution of the sequence of distributions* $\big(DCP(\lambda m_i, H_i), i \geq 0\big)$ *with the representation (2.24).*

(**ii**) *The marginal distribution of* $\{X_t\}$ *is* $DCP(\widetilde{\lambda}, G)$, *where*

$$M = \sum_{k=1}^{\nu} \frac{\lambda_k}{\lambda} M_k; \quad \widetilde{\lambda} = \lambda M = \sum_{k=1}^{\nu} \lambda_k M_k \tag{2.25}$$

and $G(z)$ *admits the representation (2.8).*

(**iii**) *The pmf* $\{g_r\}$ *is the infinite mixture of the pmf's* $(\{h_r^{(i)}\}, i \geq 0)$ *of (2.23) with mixing probabilities* $(\frac{m_i}{M},\ i \geq 0)$.

We discuss additional properties of the process $\{X_t\}$ of Theorem 2.3.

The 1-step transition probabilities of $\{X_t\}$ can be obtained from equations (2.12)–(2.14). By (2.16), the conditional distribution of X_{t+k} given $X_t = n$ results from the convolution of a $Bin(n, \alpha^k)$ distribution and the distributions $\big(DCP(\lambda m_i, H_i), 0 \leq i \leq k-1\big)$ of (2.24).

If we assume that for each $k = 1, 2, ..., \nu$, the fmgf $\widetilde{H}_k(1+t)$ of the pmf $\{h_r^{(k)}\}$ exists for $|t| < \rho_0^{(k)}$ for some $\rho_0^{(k)} > 0$, then it is easily seen that the fmgf $H(1+t)$ of (2.22) exists for $|t| < \min_{1\le k\le \nu} \rho_0^{(k)}$. It follows by Lemma 2.1-(iii), Theorem 2.1, and (2.18) applied to λ and $H(z)$ of (2.22) that the r-th factorial moment of $\{g_r\}$ is

$$\mu_{[r]}^{(g)} = \frac{1}{M(1-\alpha^r)} \sum_{k=1}^{\nu} \frac{\lambda_k}{\lambda} \mu_{[r]}^{(h^{(k)})}. \tag{2.26}$$

By (2.19), the factorial cumulants and the cumulants of X_t are (for $r \ge 1$)

$$\kappa_{[r]}^{(p)} = \frac{1}{1-\alpha^r} \sum_{k=1}^{\nu} \lambda_k \mu_{[r]}^{(h^{(k)})} \quad \text{and} \quad \kappa_r^{(p)} = \sum_{k=1}^{\nu} \lambda_k \left[\sum_{j=1}^{r} \frac{S(r,j)}{1-\alpha^j} \mu_{[j]}^{(h^{(k)})} \right]. \tag{2.27}$$

The mean and variance of X_t can be obtained from (2.20). We omit the details.

3. PROCESSES WITH POLYA-AEPPLI INNOVATIONS

A $\mathbb{Z}_+$-valued random variable with pgf $\Psi(z) = \exp\left(-\lambda \frac{1-z}{1-\theta z}\right)$ and pmf

$$f_r = \begin{cases} e^{-\lambda}, & \text{if } r = 0, \\ e^{-\lambda}\theta^r \sum_{j=1}^{r} \binom{r-1}{j-1} \frac{(\lambda\overline{\theta}/\theta)^j}{j!}, & \text{if } r > 0, \end{cases} \tag{3.1}$$

is said to have a Polya–Aeppli (or Poisson Geometric) distribution ($PA(\lambda, \theta)$) with parameters (λ, θ), $\lambda > 0$ and $\theta \in (0, 1)$. The $PA(\lambda, \theta)$ is $DCP(\lambda, H)$, where $H(z)$ is the pgf of the shifted geometric ($Geo_1(\theta)$) distribution with pmf $\{h_r\}$:

$$H(z) = \frac{\overline{\theta} z}{1-\theta z} \qquad \text{and} \qquad h_r = \overline{\theta}\theta^{r-1} \quad (r \ge 1). \tag{3.2}$$

Theorem 3.1. *Let $\{X_t\}$ be a stationary INAR (1) process with a $PA(\lambda, \theta)$ innovation. The following assertions hold:*

(**i**) *The sequence $\{m_i\}$ of (2.2) satisfies*

$$m_i = \frac{\alpha^i}{1-\theta(1-\alpha^i)} \quad \text{and} \quad 0 < m_i \le 1 \quad (i \ge 0).$$

(**ii**) *The pmf $\{h_r^{(i)}\}$ of (2.4), $i \ge 0$, is a $Geo_1(m_i\theta)$ distribution, and*

$$DCP(\lambda m_i, H_i) \sim PA(\lambda m_i, m_i\theta) \qquad (i \ge 0). \tag{3.3}$$

(**iii**) *The distribution of $\{X_t\}$ is the infinite convolution of the $PA(\lambda m_i, m_i\theta)$ distributions ($i \ge 0$).*

(**iv**) *The distribution of $\{X_t\}$ is $DCP(\widetilde{\lambda}, G)$, where $\widetilde{\lambda} = \lambda M$, $M = \sum_{i=0}^{\infty} m_i$, and G is the pgf of the infinite mixture of $Geo_1(m_i\theta)$ distributions with respective mixing probabilities $\frac{m_i}{M}$, $i \ge 0$.*

Proof: Part (i) and the first part of (ii) follow from Lemma 2.1, (2.4), (3.2), and the result $(1-t)^{-r-1} = \sum_{n=r}^{\infty} \binom{n}{r} t^{n-r}$. In turn, the first part of (ii) implies (3.3). Part (iii) ensues from Theorem 2.1-(i). Part (iv) is a direct consequence of Theorem 2.1. □

We state some additional properties of the process $\{X_t\}$ of Theorem 3.1.

The 1-step transition probability of $\{X_t\}$ can be computed from (2.12)–(2.14) with $P(\varepsilon = x) = f_x$ of (3.1). By (2.16) and (3.3), the conditional distribution of X_{t+k} given $X_t = n$ arises as the convolution of a $Bin(n, \alpha^k)$ distribution and the $PA(\lambda m_i, m_i\theta)$ distributions, $0 \le i \le k-1$.

The fmgf $H(1+t)$ of the $Geo_1(\theta)$ distribution with pmf $\{h_r\}$ of (3.2) exists for $|t| < \overline{\theta}/\theta$. Its power series expansion yields the factorial moments of $\{h_r\}$,

$$\mu_{[r]}^{(h)} = \frac{r!}{\theta}(\theta/\overline{\theta})^r \qquad (r \ge 1). \tag{3.4}$$

Formulas for the moments of $\{g_r\}$ and the cumulants, mean and variance of X_t are given below. They are derived from (2.18)–(2.20) and (3.4):

$$\mu_{[r]}^{(g)} = \frac{r!(\theta/\overline{\theta})^r}{M\theta(1-\alpha^r)} \quad \text{and} \quad \mu_r^{(g)} = \frac{1}{M\theta}\sum_{j=1}^{r} S(r,j)\frac{j!(\theta/\overline{\theta})^j}{1-\alpha^j},$$

$$\kappa_{[r]}^{(p)} = \frac{\lambda r!(\theta/\overline{\theta})^r}{\theta(1-\alpha^r)} \quad \text{and} \quad \kappa_r^{(p)} = \frac{\lambda}{\theta}\sum_{j=1}^{r} S(r,j)\frac{j!(\theta/\overline{\theta})^j}{1-\alpha^j},$$

and

$$\mu_1^{(p)} = \frac{\lambda}{\overline{\alpha}\overline{\theta}} \quad \text{and} \quad (\sigma^{(p)})^2 = \frac{\lambda(2-\overline{\alpha}\overline{\theta})}{(1-\alpha^2)\overline{\theta}^2}.$$

Remark 3.1.

(**i**) The $PA(\lambda, 0)$ distribution is Poisson (λ) and the corresponding stationary INAR(1) process simplifies to the Poisson $(\frac{\lambda}{\alpha})$ INAR (1) process discussed in [1], [13], and [14].

(**ii**) One can extend the model discussed in this section to INAR (1) processes whose innovations are finite convolutions of Polya–Aeppli distributions. The extension can be established in fairly straightforward fashion by combining the results in this section with those in Subsection 2.2.

4. PROCESSES WITH NONCENTRAL POLYA-AEPPLI INNOVATIONS

A noncentral Polya–Aeppli distribution $(NPA(\lambda_1, \lambda_2, \theta))$ with parameters $\lambda_1, \lambda_2 > 0$ and $\theta \in (0,1)$, as introduced in [9], results from the convolution of a Poisson(λ_1) distribution and a $PA(\lambda_2, \theta)$ distribution. Its pmf is

$$f_r = \begin{cases} e^{-\lambda}, & \text{if } r = 0, \\ e^{-\lambda}\theta^r \sum_{j=1}^{r} \frac{1}{j!}\Big(\sum_{k=0}^{j}\binom{j}{k}\binom{r-j+k-1}{k-1}(\lambda_2\overline{\theta}/\theta)^k(\lambda_1/\theta)^{j-k}\Big), & \text{if } r > 0. \end{cases} \tag{4.1}$$

An $NPA(\lambda_1, \lambda_2, \theta)$ distribution is $DCP(\lambda, H)$, where $\lambda = \lambda_1 + \lambda_2$ and $H(z)$ is the pgf of a mixture of a Dirac measure δ_1 sitting at 1, i.e., $\delta_1(\{1\}) = 1$, and a $Geo_1(\theta)$ distribution, with respective mixing probabilities λ_1/λ and λ_2/λ, or

$$H(z) = \frac{\lambda_1}{\lambda} z + \frac{\lambda_2}{\lambda} \frac{\overline{\theta} z}{1 - \theta z}, \quad h_1 = \frac{\lambda_1 + \overline{\theta}\lambda_2}{\lambda} \quad \text{and} \quad h_r = \frac{\lambda_2}{\lambda} \overline{\theta} \theta^{r-1} \quad (r \geq 2). \tag{4.2}$$

Theorem 4.1. *Let $\{X_t\}$ be a stationary INAR(1) process with an $NPA(\lambda_1, \lambda_2, \theta)$ innovation. The following assertions hold:*

(**i**) *The sequence $\{m_i\}$ of (2.2) satisfies*

$$m_i = \frac{\lambda_1}{\lambda} \cdot \alpha^i + \frac{\lambda_2}{\lambda} \cdot \frac{\alpha^i}{1 - \theta(1 - \alpha^i)} \quad \text{and} \quad 0 < m_i \leq 1 \quad (i \geq 0).$$

(**ii**) *The pmf $\{h_r^{(i)}\}$ of (2.4), $i \geq 0$, is a mixture of a Dirac measure δ_1 sitting at 1 and a $Geo_1(\beta_i)$ distribution, with mixing probabilities b_{i1} and b_{i2}, where*

$$\beta_i = \frac{\theta \alpha^i}{1 - \theta(1 - \alpha^i)}, \qquad b_{i1} = \frac{\lambda_1 \alpha^i}{\lambda m_i}, \qquad b_{i2} = \frac{\lambda_2}{\lambda m_i} \frac{\alpha^i}{1 - \theta(1 - \alpha^i)},$$

$$h_1^{(i)} = 1 - b_{i2}\beta_i \quad \text{and} \quad h_r^{(i)} = b_{i2}\overline{\beta_i}\beta_i^{r-1} \quad (r \geq 2).$$

Moreover,

$$DCP(\lambda m_i, H_i) \sim NPA(\lambda_1 \alpha^i, \lambda_2 \beta_i/\theta, \beta_i) \qquad (i \geq 0). \tag{4.3}$$

(**iii**) *The marginal distribution of $\{X_t\}$ is the infinite convolution of the $NPA(\lambda_1\alpha^i, \lambda_2\beta_i/\theta, \beta_i)$ distributions ($i \geq 0$).*

(**iv**) *The marginal distribution of $\{X_t\}$ is $DCP(\widetilde{\lambda}, G)$, where $\widetilde{\lambda} = \lambda M$, $M = \frac{\lambda_1}{\lambda(1-\alpha)} + \frac{\lambda_2}{\lambda\theta} \sum_{i=0}^{\infty} \beta_i$ and G is the pgf of the infinite countable mixture of the sequence of pmf's $(\{h_r^{(i)}\}, i \geq 0)$, described in (ii) above, with respective mixing probabilities $(\frac{m_i}{M}, i \geq 0)$.*

Proof: Parts (i) and (ii) follow essentially from (3.3), (4.2), Lemma 2.2, and Theorem 2.3 (for $k = 2$). Part (iii) ensues from Theorem 2.1-(i) and part (iv) is a direct consequence of Theorem 2.1-(ii). □

We obtain additional properties of the process $\{X_t\}$ of Theorem 4.1.

The 1-step transition probability of $\{X_t\}$ is obtained from (2.12)–(2.14) with $P(\varepsilon = x) = f_x$ of (4.1). By (2.16), Lemma 2.1, and Theorem 4.1-(ii), the conditional distribution of X_{t+k} given $X_t = n$ is the convolution of a $Bin(n, \alpha^k)$ distribution and the $NPA(\lambda_1\alpha^i, \lambda_2\beta_i/\theta, \beta_i)$ distributions $(0 \leq i \leq k-1)$.

The fcmgf $H(1+t)$ of the pmf $\{h_n\}$ of (4.2) exists for $|t| < \overline{\theta}/\theta$. Its power series expansion, (2.18) and (3.4), lead to the factorial moments of $\{g_r\}$:

$$\mu_{[r]}^{(g)} = \begin{cases} \dfrac{1}{\lambda M(1-\alpha)} \big(\lambda_1 + \lambda_2/\overline{\theta}\big), & \text{if } r = 1, \\ \dfrac{1}{\lambda M(1-\alpha^r)} \big(\lambda_2 r!/\theta\big) \big(\theta/\overline{\theta}\big)^r, & \text{if } r \geq 2. \end{cases}$$

Factorial cumulants and cumulants of X_t follow from (2.19):

$$\kappa_{[r]}^{(p)} = \begin{cases} \dfrac{1}{1-\alpha}(\lambda_1 + \lambda_2/\overline{\theta}), & \text{if } r = 1, \\ \dfrac{1}{1-\alpha^r}\big(\lambda_2 r!/\theta\big)\big(\theta/\overline{\theta}\big)^r, & \text{if } r \geq 2, \end{cases}$$

and

$$\kappa_r^{(p)} = \frac{\lambda_1\overline{\theta} + \lambda_2}{\overline{\alpha}\overline{\theta}} + \frac{\lambda_2}{\theta}\sum_{j=2}^{r} S(r,j)\frac{j!\big(\theta/\overline{\theta}\big)^j}{1-\alpha^j}.$$

By (2.20), the mean and variance of X_t are

$$\mu_1^{(p)} = \frac{\lambda_1\overline{\theta} + \lambda_2}{\overline{\alpha}\overline{\theta}} \quad \text{and} \quad (\sigma^{(p)})^2 = \frac{\lambda_1\overline{\theta}^2(1+\alpha) + \lambda_2(2-\overline{\alpha}\overline{\theta})}{(1-\alpha^2)\overline{\theta}^2}.$$

5. PROCESSES WITH NEGATIVE BINOMIAL INNOVATIONS

The negative binomial (NB) distribution with parameters $s > 0$ and $\theta \in (0,1)$, denoted by $NB(s,\theta)$), has pgf and pmf

$$\Psi(z) = \left\{\frac{\overline{\theta}}{1-\theta z}\right\}^s \quad \text{and} \quad f_r = \binom{s+r-1}{r}\overline{\theta}^s\theta^r \quad (r \geq 0). \tag{5.1}$$

The $NB(s,\theta)$ distribution is $DCP(\lambda, H)$, where $\lambda = -s\ln\overline{\theta}$ and $H(z)$ is the pgf of the logarithmic distribution with pmf $\{h_r\}$ described below:

$$H(z) = \frac{\ln(1-\theta z)}{\ln\overline{\theta}} \quad \text{and} \quad h_r = -\frac{\theta^r}{n\ln\overline{\theta}}, \quad (r \geq 1). \tag{5.2}$$

Theorem 5.1. *Let $\{X_t\}$ be a stationary INAR(1) process with an $NB(s,\theta)$ innovation. The following assertions hold:*

(i) *The sequence $\{m_i\}$ of (2.2) is*

$$m_i = \frac{\ln(1-\tilde{\theta}_i)}{\ln\overline{\theta}} \quad \text{with} \quad \tilde{\theta}_i = \frac{\theta\alpha^i}{1-\theta(1-\alpha^i)} \quad (i \geq 0). \tag{5.3}$$

Note $0 < \tilde{\theta}_i \leq \theta$ and $0 < m_i \leq 1$ $(i \geq 0)$. Moreover,

$$M = \sum_{i=0}^{\infty} m_i = \frac{\ln p(\alpha,\theta)}{\ln\overline{\theta}}, \quad \text{where} \quad p(\alpha,\theta) = \prod_{i=0}^{\infty}(1-\tilde{\theta}_i).$$

(ii) *The pmf $\{h_r^{(i)}\}$ of (2.4), $i \geq 0$, is logarithmic$(\tilde{\theta}_i)$ (cf. (5.2)) and*

$$DCP(\lambda m_i, H_i) \sim NB(s,\tilde{\theta}_i) \quad (i \geq 0). \tag{5.4}$$

(iii) *The marginal distribution of $\{X_t\}$ is the infinite convolution of the $NB(s,\tilde{\theta}_i)$ distributions, $i \geq 0$.*

(iv) *The marginal distribution of $\{X_t\}$ is $DCP(\widetilde{\lambda}, G)$, where $\widetilde{\lambda} = -s\ln p(\alpha,\theta)$ and G is the pgf of an infinite countable mixture of logarithmic$(\tilde{\theta}_i)$ distributions with mixing probabilities $\left(\frac{\ln(1-\tilde{\theta}_i)}{\ln p(\alpha,\theta)},\ i \geq 0\right)$.*

Proof: By (5.2), $m_i = 1 - H(1-\alpha^i) = (\ln\overline{\theta} - \ln(1-\theta(1-\alpha^i))/\ln\overline{\theta}$, which implies (5.3), since $1-\tilde{\theta}_i = \overline{\theta}/(1-\theta(1-\alpha^i))$. Thus (i) holds. Straightforward calculations show that

$$H_i(z) = 1 - \frac{1}{m_i}\big(1 - H(1-\alpha^i+\alpha^i z)\big) = \frac{\ln(1-\tilde{\theta}_i z)}{\ln(1-\tilde{\theta}_i)},$$

where $H(z)$ is as in (5.2). This establishes the first part of (ii), which in turn implies (5.4). Clearly, (iii) follows from Theorem 2.1-(i). Part (iv) is a direct consequence of (i)–(ii) and Theorem 2.1-(ii). □

We give additional proprerties of the process $\{X_t\}$ of Theorem 5.1.

The 1-step transition probability of $\{X_t\}$ can be computed from (2.12)–(2.14) with $P(\varepsilon = x) = f_x$ of (5.1). By (2.16), Lemma 2.1, and Theorem 5.1 (i)–(ii), the conditional distribution of X_{t+k} given $X_t = n$ results from the convolution of a $Bin(n, \alpha^k)$ distribution and the $NB(s, \tilde{\theta}_i)$ distributions $(0 \le i \le k-1)$.

The fmgf $H(1+t)$ of the logarithmic(θ) distribution with pgf $H(z)$ and pmf $\{h_r\}$ of (5.2) exists for $|t| < \overline{\theta}/\theta$. The factorial moments of $\{h_r\}$ are given by (see equation 7.11, p. 305, in [6])

$$\mu^{(h)}_{[r]} = -\frac{(r-1)!\big(\theta/\overline{\theta}\big)^r}{\ln\overline{\theta}} \qquad (r \ge 1). \tag{5.5}$$

Formulas for the moments of $\{g_r\}$ and the cumulants, mean and variance of X_t are given below. They are derived from (2.18)–(2.20) and (5.5):

$$\mu^{(g)}_{[r]} = -\frac{(r-1)!\big(\theta/\overline{\theta}\big)^r}{(1-\alpha^r)\ln p(\alpha,\theta)} \quad \text{and} \quad \mu^{(g)}_r = -\frac{1}{\ln p(\alpha,\theta)}\sum_{j=1}^{r} S(r,j)\frac{(j-1)!(\theta/\overline{\theta})^j}{1-\alpha^j},$$

$$\kappa^{(p)}_{[r]} = \frac{s(r-1)!\big(\theta/\overline{\theta}\big)^r}{1-\alpha^r} \quad \text{and} \quad \kappa^{(p)}_r = s\sum_{j=1}^{r} S(r,j)\frac{(j-1)!(\theta/\overline{\theta})^j}{1-\alpha^j},$$

$$\mu^{(p)}_1 = \frac{s\theta}{\overline{\alpha}\overline{\theta}} \quad \text{and} \quad (\sigma^{(p)})^2 = \frac{s\theta(1+\alpha\overline{\theta})}{(1-\alpha^2)\overline{\theta}^2}.$$

Remark 5.1.

(**i**) Note that the special case of $s = 1$ of Theorem 5.1 covers the important special case of the unshifted geometric(θ), or $Geo_0(\theta)$, innovation. These results can be seen as extensions of some of the work in [5].

(**ii**) One can extend the model discussed in this section to INAR (1) processes whose innovations are finite convolutions of negative binomial distributions. The extension can be established in fairly straightforward fashion by combining the results in this section with those in Subsection 2.2.

6. PROCESSES WITH NONCENTRAL NEGATIVE BINOMIAL INNOVATIONS

Assume that $\theta \in (0,1)$, $s > 0$ and $\lambda_2 > 0$. Ong and Lee ([16]) introduced the noncentral NB distribution, $NNB(\lambda_2, s, \theta)$, as the mixture of $NB(v, \theta)$ distributions, where v is a value of the random variable $V = Y + s$ and Y is Poisson(λ_2). The pgf of $NNB(\lambda_2, s, \theta)$ is $\Psi(z) = \left(\frac{\overline{\theta}}{1-\theta z}\right)^s \exp\left(-\lambda_2 \frac{1-z}{1-\theta z}\right)$, and

$$f_r = \begin{cases} \overline{\theta}^s e^{-\lambda_2}, & \text{if } r = 0, \\ e^{-\lambda_2} \theta^r \overline{\theta}^s \sum_{k=0}^{r} \sum_{j=1}^{k} \binom{k-1}{j-1} \binom{s+r-k-1}{r-k} \frac{\lambda_2 (\overline{\theta}/\theta)^j}{j!}, & \text{if } r > 0. \end{cases} \tag{6.1}$$

The $NNB(\lambda_2, s, \theta)$ distribution is the convolution of an $NB(s, \theta)$ distribution and a $PA(\lambda_2, \theta)$ distribution. Hence, by Lemma 2.2 (for k=2), $NNB(\lambda_2, s, \theta) \sim DCP(\lambda, H)$, where $\lambda = \lambda_2 - s \ln \overline{\theta} > 0$ and

$$H(z) = \frac{1}{\lambda}\left(-s \ln(1-\theta z) + \lambda_2 \frac{\overline{\theta} z}{1-\theta z}\right) \quad \text{and} \quad h_r = \frac{\theta^r}{\lambda}\left(\frac{s}{r} + \lambda_2 \frac{\overline{\theta}}{\theta}\right) \quad (r \geq 1). \tag{6.2}$$

We note that $\{h_r\}$ is a mixture of a logarithmic(θ) distribution and a $Geo_1(\theta)$ distribution with respective mixing probabilities $-s \ln \overline{\theta}/\lambda$ and λ_2/λ.

Theorem 6.1. *Let $\{X_t\}$ be a stationary INAR(1) process with an $NNB(\lambda_2, s, \theta)$ innovation of (6.1)–(6.2). Let*

$$\tilde{\theta}_i = \frac{\theta \alpha^i}{1 - \theta(1-\alpha^i)} \quad \text{and} \quad p(\alpha, \theta) = \prod_{i=0}^{\infty} (1 - \tilde{\theta}_i).$$

The following assertions hold:

(i) *For $\{m_i\}$ of (2.2) we have*

$$m_i = \frac{1}{\lambda}\left(-s \ln(1-\tilde{\theta}_i) + \lambda_2 \frac{\tilde{\theta}_i}{\theta}\right) \quad \text{and} \quad M = \frac{1}{\lambda}\left(-s \ln p(\alpha, \theta) + \frac{\lambda_2}{\theta} \sum_{i=0}^{\infty} \tilde{\theta}_i\right).$$

(ii) *The pmf $\{h_r^{(i)}\}$ of (2.4), $i \geq 0$, is a mixture of a logarithmic$(\tilde{\theta}_i)$ distribution and a $Geo_1(\tilde{\theta}_i)$ distribution, with respective mixing probabilities $b_{i1} = \left(-s \ln(1-\tilde{\theta}_i)/(\lambda m_i)\right)$ and $b_{i2} = (\lambda_2 \tilde{\theta}_i)/(\lambda m_i \theta)$. Moreover,*

$$DCP(\lambda m_i, H_i) \sim NB(s, \tilde{\theta}_i) * PA\left(\lambda_2 \frac{\tilde{\theta}_i}{\theta}, \tilde{\theta}_i\right). \tag{6.3}$$

(iii) *The marginal distribution of $\{X_t\}$ is the infinite convolution of the $\big(DCP(\lambda m_i, H_i),\ i \geq 0\big)$ of (6.3).*

(iv) *The marginal distribution of $\{X_t\}$ is $DCP(\widetilde{\lambda}, G)$, where $\widetilde{\lambda} = \lambda M$ and G is the pgf of an infinite countable mixture of the sequence of pmf's $(\{h_r^{(i)}\}, i \geq 0)$ (described in (ii) above) with mixing probabilities $\left(\frac{m_i}{M},\ i \geq 0\right)$.*

Proof: The proof is similar to that of Theorem 4.1. The results follow from Lemma 2.2, Theorem 2.3 (with $k = 2$), Theorem 3.1 and Theorem 5.1. We omit the details. □

We give some additional properties of the process $\{X_t\}$ of Theorem 6.1.

The 1-step transition probability of $\{X_t\}$ can be computed from (2.12)–(2.14) with $P(\varepsilon = x) = f_x$ of (6.1). By (2.16), the conditional distribution of X_{t+k} given $X_t = n$ results from the convolution of a $Bin(n, \alpha^k)$ distribution and the distributions $\big(DCP(\lambda m_i, H_i)$, $0 \le i \le k-1\big)$ of (6.3).

As a mixture of a logarithmic(θ) distribution and a $Geo_1(\theta)$ distribution, the pmf $\{h_r\}$ of (6.2) has a finite fmgf $H(1+t)$ for $|t| < \overline{\theta}/\theta$. Therefore, the factorial moments of $\{g_r\}$ are, by (2.26), (3.4) and (5.5),

$$\mu_{[r]}^{(g)} = \frac{(r-1)!(\theta/\overline{\theta})^r}{\lambda M \theta(1-\alpha^r)}(s\theta + \lambda_2 r).$$

Combining (2.27) with the moment and cumulant formulas derived in Section 6 yields the factorial cumulants and the cumulants of X_t:

$$\kappa_{[r]}^{(p)} = \frac{(r-1)!(\theta/\overline{\theta})^r}{\theta(1-\alpha^r)}(s\theta + \lambda_2 r) \tag{6.4}$$

and

$$\kappa_r^{(p)} = \frac{1}{\theta}\sum_{j=1}^{r} S(r,j)\frac{(j-1)!(\theta/\overline{\theta})^j}{1-\alpha^j}(s\theta + \lambda_2 j). \tag{6.5}$$

By (2.20), the mean and variance of $\{X_t\}$ are

$$\mu_1^{(p)} = \frac{\lambda_2 + s\theta}{\overline{\alpha}\overline{\theta}} \quad \text{and} \quad (\sigma^{(p)})^2 = \frac{\lambda_2(2-\overline{\alpha}\overline{\theta}) + s\theta(1+\alpha\overline{\theta})}{(1-\alpha^2)\overline{\theta}}.$$

7. PROCESSES WITH POISSON-LINDLEY INNOVATIONS

In this section, we revisit the INAR (1) model with Poisson–Lindley innovation introduced in [10] (see also [17]) and expand on their results. The Poisson–Lindley distribution $(PL(\phi))$ with parameter $\phi > 0$ is the mixture of a $Geo_1(\frac{1}{1+\phi})$ distribution and an $NB(2, \frac{1}{1+\phi})$ distribution with respective mixing probabilities $\frac{\phi}{1+\phi}$ and $\frac{1}{1+\phi}$. Its pgf and pmf are

$$\Psi(z) = \frac{\phi^2}{1+\phi}\cdot\frac{2+\phi-z}{(1+\phi-z)^2} \quad \text{and} \quad f_r = \frac{\phi^2}{(1+\phi)^{r+2}}\left(1+\frac{r+1}{1+\phi}\right) \quad (r \ge 0). \tag{7.1}$$

For additional details and references on the $PL(\phi)$ distribution, we refer to [15]. A $PL(\phi)$ distribution is $DCP(\lambda, H)$ with

$$\lambda = \ln\left[\frac{(1+\phi)^3}{\phi^2(2+\phi)}\right], \qquad H(z) = 1 + \frac{1}{\lambda}\ln\left[\frac{\phi^2(2+\phi-z)}{(1+\phi)(1+\phi-z)^2}\right], \tag{7.2}$$

and

$$h_r = \frac{1}{\lambda r}\left(\frac{2}{(1+\phi)^r} - \frac{1}{(2+\phi)^r}\right) \quad (r \geq 1). \tag{7.3}$$

We introduce the Modified Poisson–Lindley distribution ($MPL(\phi, \beta)$) with parameters $\phi > 0$ and $\beta \in (0, 1]$ as the distribution of $\beta \odot X$, where $X \sim PL(\phi)$. The pgf of the $MPL(\phi, \beta))$ distribution is $\Psi(1 - \beta + \beta z)$, with $\Psi(z)$ of (7.1). Note that, $MPL(\phi, 1) \sim PL(\phi)$.

Lemma 7.1. *An $MPL(\phi, \beta)$ distribution arises as a mixture of a $Geo_1(\beta/(\beta+\phi))$ distribution and an $NB(2, \beta/(\beta+\phi))$ distribution with resp. mixing probabilities $\frac{\phi}{1+\phi}$ and $\frac{1}{1+\phi}$. Moreover, $MPL(\phi, \beta) \sim DCP(\lambda_\beta, H_\beta)$, where*

$$\lambda_\beta = \ln\left[\frac{(1+\phi)(\beta+\phi)^2}{\phi^2(1+\beta+\phi)}\right], \qquad H_\beta(z) = 1 + \frac{1}{\lambda_\beta}\ln\left[\frac{\phi^2(1+\beta+\phi-\beta z)}{(1+\phi)(\beta+\phi-\beta z)^2}\right]. \tag{7.4}$$

Moreover, the pmf $\{h_r^{(\beta)}\}$ of $H_\beta(z)$ is

$$h_r^{(\beta)} = \frac{1}{\lambda_\beta r}\left[2\left(\frac{\beta}{\beta+\phi}\right)^r - \left(\frac{\beta}{1+\beta+\phi}\right)^r\right] \qquad (r \geq 1). \tag{7.5}$$

Proof: If X is $Geo_1(1/(1+\phi))$ (resp. $NB(2, 1/(1+\phi))$, then $\beta \odot X$ is $Geo_1(\beta/(\beta+\phi))$ (resp. $NB(2, \beta/(\beta+\phi))$. By (7.1), we obtain

$$\Psi(1 - \beta + \beta z) = \frac{\phi^2}{1+\phi} \cdot \frac{1+\beta+\phi-\beta z}{(\beta+\phi-\beta z)^2}.$$

A standard argument leads to the representation

$$\Psi(1 - \beta + \beta z) = \exp\{\lambda_\beta(H_\beta - 1)\},$$

where λ_β and H_β and its pmf are as in (7.4)–(7.5). □

Theorem 7.1. *Let $\{X_t\}$ be a stationary INAR (1) process with a $PL(\phi)$ innovation with characteristics (7.1)–(7.3). The following assertions hold:*

(**i**) *For every $i \geq 0$,*

$$m_i = \frac{1}{\lambda}\ln\left[\frac{(1+\phi)(\phi+\alpha^i)^2}{\phi^2(1+\phi+\alpha^i)}\right] \qquad \text{and} \qquad M = \frac{1}{\lambda}\ln\prod_{i=0}^{\infty}(1+a_i),$$

where $a_i = \dfrac{\alpha^i(\phi^2 + 2\phi + \alpha^i\phi + \alpha^i)}{\phi^2(1+\phi+\alpha^i)}$.

(**ii**) *The pmf* $\{h_r^{(i)}\}$ *of (2.4),* $i \geq 0$, *is given in (7.5) with* $\beta = \alpha^i$ *and* $\lambda_\beta = \lambda m_i$, *and*

$$DCP(\lambda m_i, H_i) \sim MPL(\phi, \alpha^i). \tag{7.6}$$

(**iii**) *The marginal distribution of* $\{X_t\}$ *is the infinite convolution of the distributions* $\big(MPL(\phi, \alpha^i),\ i \geq 0\big)$.

(**iv**) *The marginal distribution of* $\{X_t\}$ *is* $DCP(\widetilde{\lambda}, G)$, *where* $\widetilde{\lambda} = \ln \prod_{i=0}^{\infty}(1 + a_i)$, *and* G *is the pgf of the infinite countable mixture of the pmf's* $\big(\{h_r^{(i)}\},\ i \geq 0\big)$ *with respective mixing probabilities* $\big(\frac{m_i}{M},\ i \geq 0\big)$.

Proof: (i) follows from Lemma 2.1, (7.1)–(7.2), and the formula $M = \sum_{i=0}^{\infty} m_i$. Part (ii) is a direct consequence of Lemma 9.1 by setting $\beta = \alpha^i$. Part (iii) and (iv) result from (ii) and Theorem 2.1-(ii), respectively. □

We give additional properties of the process $\{X_t\}$ of Theorem 7.1.

The 1-step transition probability of $\{X_t\}$ can be computed from (2.12)–(2.14) with $P(\varepsilon = x) = f_x$ of (7.1). By (2.16) and Theorem 7.1-(ii), the conditional distribution of X_{t+k} given $X_t = n$ results from the convolution of a $Bin(n, \alpha^k)$ distribution and the $MPL(\phi, \alpha^i)$ distributions, $0 \leq i \leq k-1$.

The fcmgf $H(1+t)$ of the pmf $\{h_r\}$ of (7.2)–(7.3) exists for $|t| < \phi/2$. Its power series expansion yields the factorial moments of $\{h_r\}$:

$$\mu_{[r]}^{(h)} = \frac{(r-1)!}{\lambda}\left(\frac{2}{\phi^r} - \frac{1}{(1+\phi)^r}\right).$$

Formulas for the factorial moment of $\{g_r\}$ and the cumulants, mean and variance of X_t are given below. They are derived from (2.18)–(2.20):

$$\mu_{[r]}^{(g)} = \frac{(r-1)!}{\lambda M(1-\alpha^r)}\left(\frac{2}{\phi^r} - \frac{1}{(1+\phi)^r}\right),$$

$$\kappa_{[r]}^{(p)} = \frac{(r-1)!}{(1-\alpha^r)}\left(\frac{2}{\phi^r} - \frac{1}{(1+\phi)^r}\right) \quad \text{and} \quad \kappa_r^{(p)} = \sum_{j=1}^{r} S(r,j)\frac{(j-1)!}{(1-\alpha^j)}\left(\frac{2}{\phi^j} - \frac{1}{(1+\phi)^j}\right),$$

and

$$\mu_1^{(p)} = \frac{2+\phi}{\overline{\alpha}\phi(1+\phi)} \qquad \text{and} \qquad (\sigma^{(p)})^2 = \frac{(1+\alpha)\phi^3 + (4+3\alpha)\phi^2 + 2(3+\alpha)\phi + 2}{(1-\alpha^2)\phi^2(1+\phi)^2}.$$

8. PROCESSES WITH EULER-TYPE INNOVATIONS

Let $l(0,1)$ be the set of sequences $\Theta = (\theta_k, k \geq 0)$ such that $\theta_k \in (0,1)$ for every $k \geq 0$ and

$$\sum_{k=0}^{\infty} \frac{\theta_k}{1-\theta_k} < \infty. \tag{8.1}$$

Define

$$S_r(\Theta) = \sum_{k=0}^{\infty} \theta_k^r \quad \text{and} \quad T_r(\Theta) = \sum_{k=0}^{\infty} \left(\frac{\theta_k}{1-\theta_k} \right)^r \qquad (r \geq 1). \tag{8.2}$$

Note that the condition (8.1) implies $S_r(\Theta) < \infty$ and $T_r(\Theta) < \infty$ for all $r \geq 1$.

A $\mathbb{Z}_+$-valued rv is said to have an Euler-type distribution ($Euler - T(\Theta)$), $\Theta \in l(0,1)$, if it is an infinite convolution of $Geo_0(\theta_k)$ rv's. Its pgf is

$$\Psi(z) = \prod_{k=0}^{\infty} \left(\frac{1-\theta_k}{1-\theta_k z} \right). \tag{8.3}$$

We gather a few basic properties of an $Euler - T(\Theta)$ distribution.

Lemma 8.1. *Let $\{q_r\}$ be the pmf of an $Euler - T(\Theta)$ for some $\Theta \in l(0,1)$. The following assertions hold:*

(**i**) *$\{q_r\}$ is the pmf of a $DCP(\lambda, H)$ with*

$$\lambda = \sum_{k=0}^{\infty} (-\ln(1-\theta_k)) \quad \text{and} \quad H(z) = \sum_{k=0}^{\infty} \frac{-\ln(1-\theta_k)}{\lambda} H_k(z), \tag{8.4}$$

where, for each $k \geq 0$, $H_k(z)$ is the pgf of a logarithmic(θ_k) distribution. The pmf $\{h_r\}$ with pgf $H(z)$ is an infinite countable mixture of logarithmic(θ_k) distributions ($k \geq 0$) with respective mixing probabilities $\left(\frac{-\ln(1-\theta_k)}{\lambda},\ k \geq 0 \right)$, or $h_r = S_r(\Theta)/(\lambda r)$, $r \geq 1$.

(**ii**) *$\{q_r\}$ satisfies the following recurrence relation:*

$$(r+1) q_{r+1} = \sum_{k=0}^{r} q_k S_{r+1-k}(\Theta) \quad \text{and} \quad q_0 = \prod_{k=0}^{\infty} (1-\theta_k). \tag{8.5}$$

(**iii**) *There exists $0 < \rho_0 \leq 1$ such that the fcmgf $H(1+t)$ of the pmf $\{h_r\}$ of part (i) is finite for $|t| < \rho_0$. Consequently, $\{h_r\}$ has finite factorial moments of all orders:*

$$\mu_{[r]}^{(h)} = \frac{(r-1)!}{\lambda} T_r(\Theta) \qquad (r \geq 1). \tag{8.6}$$

(**iv**) *$\{q_r\}$ has finite factorial cumulants of all orders:*

$$\kappa_{[r]}^{(q)} = (r-1)! T_r(\Theta) \qquad (r \geq 1). \tag{8.7}$$

Proof: Since $-\ln(1-x) \sim x$, as $x \to 0$, the two infinite series with respective positive summands $-\ln(1-\theta_k)$ and $-\ln(1-\theta_k z)$, $z \in (0,1)$, are convergent. Therefore, $\ln \Psi(z) = \sum_{k=0}^{\infty} \ln(1-\theta_k) - \sum_{j=0}^{\infty} \ln(1-\theta_k z)$. Letting λ be as in (8.4), we have

$$\ln \Psi(z) = \lambda \left(-1 + \sum_{k=0}^{\infty} \frac{-\ln(1-\theta_k)}{\lambda} \frac{\ln(1-\theta_k z)}{\ln(1-\theta_k)} \right).$$

The function $H_k(z) = \frac{\ln(1-\theta_k z)}{\ln(1-\theta_k)}$ is the pgf of a logarithmic(θ_k) for each $k \geq 0$ (see (5.2)). Therefore, $\ln \Psi(z) = \lambda(H(z) - 1)$, with $H(z)$ of (8.4). Again by (8.4), $\{h_r\}$ is an infinite countable mixture of logarithmic(θ_k) distributions with the stated mixing probabilities. We have by (8.4) and (5.2)

$$h_r = \sum_{k=0}^{\infty} \frac{-\ln(1-\theta_k)}{\lambda} \frac{\theta_k^r}{-r\ln(1-\theta_k)} \qquad (r \geq 1),$$

which establishes (i), via (8.2). Note that $q_0 = e^{-\lambda}$ and, similarly to (2.10), q_r satisfies the recurrence formula (8.5). We now prove (iii). By (8.1), there exists $k_0 > 1$ such that $\theta_k/\overline{\theta_k} < 1$ for $k \geq k_0$. Therefore, $\inf_{k\geq k_0} \overline{\theta_k}/\theta_k \geq 1$. Let $\rho_0 = \min\big(1, \min_{0\leq k<k_0} \overline{\theta_k}/\theta_k\big)$. Since $\rho_0 \leq \overline{\theta_k}/\theta_k$ for every $k \geq 0$, the fmgf $H_k(1+t)$ of the logarithmic(θ_k) distribution exists for $|t| < \rho_0$. We have by (5.5) and equation (1.274), p. 59, in [6],

$$H_k(1+t) = 1 + \sum_{r=1}^{\infty} \frac{(r-1)!\big(\theta_k/\overline{\theta_k}\big)^r}{-\ln \overline{\theta_k}} \frac{t^r}{r!} \qquad (|t| < \rho_0).$$

A standard argument shows that $H(1+t) = \sum_{k=0}^{\infty} \frac{-\ln \overline{\theta_k}}{\lambda} H_k(1+t)$ converges uniformly over the interval $|t| \leq \rho$ for every $0 < \rho < \rho_0$. By Weierstrass Theorem, p. 430, in [8], we have

$$H(1+t) = 1 + \sum_{r=1}^{\infty} \left[\sum_{k=0}^{\infty} \frac{-\ln \overline{\theta_k}}{\lambda} \frac{(r-1)!\big(\theta_k/\overline{\theta_k}\big)^r}{-\ln \overline{\theta_k}}\right] \frac{t^r}{r!} \qquad (|t| < \rho_0),$$

which implies (8.6). Finally, by equation 9.49, p. 391, in [6], we have $\kappa_{[r]}^{(q)} = \lambda \mu_{[r]}^{(h)}$ which leads to (8.7). □

One can conclude from (8.7) and (2.21) that an $Euler - T(\Theta)$ has finite moments $\{\mu_r^{(q)}\}$ of all orders, and thus finite factorial moments $\{\mu_{[r]}^{(q)}\}$ of all orders.

Theorem 8.1. *Let $\{X_t\}$ be a stationary INAR(1) process with an $Euler - T(\Theta)$ innovation for some $\Theta \in l(0,1)$. For $i, k \geq 0$, let*

$$\theta_i^{(k)} = \frac{\theta_k \alpha^i}{1 - \theta_k(1-\alpha^i)} \qquad \text{and} \qquad p_i(\alpha, \Theta) = \prod_{k=0}^{\infty} \Big(1 + \frac{\theta_k \alpha^i}{1-\theta_k}\Big). \tag{8.8}$$

The following assertions hold:

(**i**) *The sequence $\{m_i\}$ of (2.2) is*

$$m_i = \frac{1}{\lambda} \sum_{k=0}^{\infty} (-\ln(1-\theta_i^{(k)})) = \frac{1}{\lambda} \ln p_i(\alpha, \Theta) \quad (i \geq 0). \tag{8.9}$$

Note that $0 < \theta_i^{(k)} \leq \theta_k$ and $0 < m_i \leq 1$. Moreover,

$$M = \sum_{i=0}^{\infty} m_i = \frac{1}{\lambda} \ln \left[\prod_{i=0}^{\infty} p_i(\alpha, \Theta)\right]. \tag{8.10}$$

(**ii**) *The pmf $\{h_r^{(i)}\}$ of (2.4), $i \geq 0$, is an infinite countable mixture of logarithmic$(\theta_i^{(k)})$ distributions, $k \geq 0$, with mixing probabilities $\Big(\frac{-\ln(1-\theta_i^{(k)})}{p_i(\alpha,\Theta)}, k \geq 0\Big)$, and*

$$DCP(\lambda m_i, H_i) \sim Euler - T(\Theta_i), \qquad \Theta_i = (\theta_i^{(k)},\ k \geq 0). \tag{8.11}$$

(**iii**) *The marginal distribution of* $\{X_t\}$ *is the infinite convolution of the* $Euler-T(\Theta_i)$ *distributions* $(i \geq 0)$ *of (8.11).*

(**iv**) *The marginal distribution of* $\{X_t\}$ *is* $DCP(\widetilde{\lambda}, G)$, *where* $\widetilde{\lambda} = \ln\Big[\prod_{i=0}^{\infty} p_i(\alpha, \Theta)\Big]$ *and* G *is the pgf of an infinite countable mixture of the pmf's* $(h_r^{(i)}, i \geq 0)$ *of (ii) with mixing probabilities* $\Big(\ln p_i(\alpha, \Theta) \Big/ \ln\Big[\prod_{j=0}^{\infty} p_j(\alpha, \Theta)\Big], i \geq 0\Big)$.

Proof: For (i), we have by (8.4),

$$m_i \;=\; 1 - H(1-\alpha^i) \;=\; \sum_{k=0}^{\infty} \frac{-\ln(1-\theta_k)}{\lambda}(1 - H_k(1-\alpha^i)).$$

Since $H_k(z)$ is the pgf of a logarithmic(θ_k) distribution, it follows that $1 - H_k(1-\alpha^i) = \frac{\ln(1-\theta_i^{(k)})}{\ln(1-\theta_k)}$, from which we deduce the first equation in (8.9). The second equation as well as (8.10) are easily seen to hold. The convergence of the infinite products in part (i) stems from $\sum_{i=0}^{\infty}\sum_{k=0}^{\infty} \frac{\theta_k \alpha^i}{1-\theta_k} < \infty$. This leads to

$$1 - H_k(1 - \alpha^i + \alpha^i z) = \frac{\ln(1-\theta_i^{(k)})}{\ln(1-\theta_k)}(1 - H_{ki}(z)),$$

where $H_{ki}(z)$ is the pgf of a logarithmic$(\theta_i^{(k)})$. We conclude by (2.3) and (8.4)

$$H_i(z) = \sum_{k=0}^{\infty} \frac{-\ln(1-\theta_i^{(k)})}{\lambda m_i} H_{ki}(z). \tag{8.12}$$

Now, by (5.2),

$$h_r^{(i)} = \sum_{k=1}^{\infty} \frac{-\ln(1-\theta_i^{(k)})}{\lambda m_i} \frac{[\theta_i^{(k)}]^r}{-r\ln(1-\theta_i^{(k)})} = \frac{S_r(\Theta_i)}{r p_i(\alpha, \theta)}.$$

which proves the first part of (ii). Let $\Psi(z)$ be as in (8.3). By Lemma 2.1, (8.9) and (8.12), the pgf, $\Psi(1-\alpha^i+\alpha^i z)$, of $DCP(\lambda m_i, H_i)$ is shown to be

$$\Psi(1-\alpha^i+\alpha^i z) = \exp\big\{\lambda m_i (H_i(z) - 1)\big\} = \prod_{k=0}^{\infty}\Big(\frac{1-\theta_i^{(k)}}{1-\theta_i^{(k)} z}\Big).$$

It is easily seen that $\Theta_i = (\theta_i^{(k)}, k \geq 0)$ belongs to $l(0,1)$. Therefore, (8.11) holds, thus completing the proof of (ii). Part (iii) follows from (8.11) and Theorem 2.1-(i). Part (iv) is a direct consequence of (i)–(ii) and Theorem 2.1-(ii). □

We discuss additional properties of the process $\{X_t\}$ of Theorem 8.1.

The 1-step transition probability of $\{X_t\}$ can be computed from (2.12)–(2.14) where the probabilities $P(\varepsilon = x) = q_x$, $x \geq 0$, can be obtained using (8.5). By (2.16), Lemma 8.1, and Theorem 8.1 (i)–(ii), the conditional distribution of X_{t+k} given $X_t = n$ arises as the convolution of a $Bin(n, \alpha^k)$ distribution and the $Euler-T(\Theta_i)$ distributions $(0 \leq i \leq k-1)$ of (8.11).

Formulas for the moments of $\{g_r\}$ and the factorial moments, mean and variance of X_t are obtained from (2.18)–(2.20) and (8.6):

$$\mu_{[r]}^{(g)} = \frac{(r-1)!}{\lambda M(1-\alpha^r)} T_r(\Theta) \qquad \text{and} \qquad \mu_r^{(g)} = \frac{1}{\lambda M} \sum_{j=1}^{r} S(r,j) \frac{(j-1)!}{(1-\alpha^j)} T_j(\Theta),$$

$$\kappa_{[r]}^{(p)} = \frac{(r-1)!}{(1-\alpha^r)} T_r(\Theta) \qquad \text{and} \qquad \kappa_r^{(p)} = \sum_{j=1}^{r} S(r,j) \frac{(j-1)!}{(1-\alpha^j)} T_j(\Theta).,$$

and

$$\mu_1^{(p)} = \frac{T_1(\Theta)}{1-\alpha} \qquad \text{and} \qquad (\sigma^{(p)})^2 = \frac{(1+\alpha)T_1(\Theta) + T_2(\Theta)}{1-\alpha^2}.$$

9. PROCESSES WITH EULER INNOVATIONS

The Euler distribution ($Euler(\eta, q)$) introduced by Benkherouf and Bather ([3]) (see [6]) is an $Euler - T(\Theta)$ distribution with $\Theta = (\eta q^k, k \geq 0)$ for $0 < \eta < 1$ and $0 < q < 1$. An application of the ratio test shows that indeed $\Theta \in l(0,1)$. We also note that $S_r(\Theta) = \frac{\eta^r}{1-q^r}$, $r \geq 1$. We use the notation $T_r(\eta, q)$ in lieu of $T_r(\Theta)$.

We recall a few basic properties of the $Euler(\eta, q)$ distribution (cf., for example, [7]). Its pmf $\{q_x\}$ is

$$q_0 = \prod_{j=0}^{\infty}(1-\eta q^j) \qquad \text{and} \qquad q_x = \frac{\eta^x}{\prod_{l=1}^{x}(1-q^l)} q_0 \quad (x \geq 1). \tag{9.1}$$

Its mean and variance are

$$\mu = \sum_{x=0}^{\infty} \frac{\eta q^x}{1-\eta q^x} \qquad \text{and} \qquad \sigma^2 = \sum_{x=0}^{\infty} \frac{\eta q^x}{(1-\eta q^x)^2}.$$

The following result is known. We refer to Lemma 8.1 for convenience.

The $Euler(\eta, q)$ distribution is $DCP(\lambda, H)$ with $\lambda = -\ln\left(\prod_{k=0}^{\infty}(1-\eta q^k)\right)$ and $H(z)$ is the pgf of an infinite countable mixture of logarithmic(ηq^k) distributions, $k \geq 0$, with respective mixing probabilities $\left(\frac{-\ln(1-\eta q^k)}{\lambda}, k \geq 0\right)$. Its pmf is $h_r = \eta^k/(\lambda k(1-q^k))$, $r \geq 1$.

The main result of the section is stated without proof as it is a particular case of Theorem 8.1.

Theorem 9.1. *Let $\{X_t\}$ be a stationary INAR (1) process with an $Euler(\eta, q)$ innovation for some η, $q \in (0,1)$. For $i, k \geq 0$, let*

$$\theta_i^{(k)} = \frac{\eta q^k \alpha^i}{1-\eta q^k(1-\alpha^i)} \qquad \text{and} \qquad p_i(\alpha, \eta, q) = \prod_{k=0}^{\infty}\left(1 + \frac{\eta q^k \alpha^i}{1-\eta q^k}\right). \tag{9.2}$$

The following assertions hold:

(**i**) *The sequence* $\{m_i\}$ *of (2.2) and* $M = \sum_{i=0}^{\infty} m_i$ *are as follows:*

$$m_i = \frac{1}{\lambda} \ln p_i(\alpha, \eta, q) \qquad \text{and} \qquad M = \frac{1}{\lambda} \ln \left[\prod_{i=0}^{\infty} p_i(\alpha, \eta, q) \right]. \tag{9.3}$$

Note that $0 < \theta_i^{(k)} \le \eta q^k$ *and* $0 < m_i \le 1$ $(i \ge 0)$.

(**ii**) *The pmf* $\{h_r^{(i)}\}$ *of (2.4),* $i \ge 0$, *is an infinite countable mixture of logarithmic*$(\theta_i^{(k)})$ *distributions,* $k \ge 0$, *with mixing probabilities* $\left(\frac{-\ln(1-\theta_i^{(k)})}{p_i(\alpha,\eta,q)},\ k \ge 0\right)$, *and*

$$DCP(\lambda m_i, H_i) \sim Euler - T(\Theta_i), \qquad \Theta_i = (\theta_i^{(k)},\ k \ge 0). \tag{9.4}$$

(**iii**) *The marginal distribution of* $\{X_t\}$ *is the infinite convolution of the* $Euler - T(\Theta_i)$ *distributions* $(i \ge 0)$ *of (9.4).*

(**iv**) *The marginal distribution of* $\{X_t\}$ *is* $DCP(\widetilde{\lambda}, G)$, *where* $\widetilde{\lambda} = \ln\left[\prod_{i=0}^{\infty} p_i(\alpha, \Theta)\right]$ *and* G *is the pgf of an infinite countable mixture of the pmf's* $(h_r^{(i)},\ i \ge 0)$ *of (ii) with mixing probabilities* $\left(\ln p_i(\alpha, \eta, q) / \ln\left[\prod_{j=0}^{\infty} p_j(\alpha, \eta, q)\right],\ i \ge 0\right)$.

Additional properties of the process $\{X_t\}$ of Theorem 9.1 are given next.

The 1-step transition probability of $\{X_t\}$ can be computed from (2.12)–(2.14) where the probabilities $P(\varepsilon = x) = q_x$, $x \ge 0$, are as in (9.1). By (2.16) and Theorem 8.1 (i)–(ii), the conditional distribution of X_{t+k} given $X_t = n$ arises as from the convolution of a $Bin(n, \alpha^k)$ distribution and the $Euler - T(\Theta_i)$ distributions $(0 \le i \le k-1)$ of (9.4).

Formulas for the moments of $\{g_r\}$ and the factorial moments, mean and variance of X_t are as follows:

$$\mu_{[r]}^{(g)} = \frac{(r-1)!}{\lambda M(1-\alpha^r)} T_r(\alpha, \eta, q) \qquad \text{and} \qquad \mu_r^{(g)} = \frac{1}{\lambda M} \sum_{j=1}^{r} S(r, j) \frac{(j-1)!}{(1-\alpha^j)} T_j(\alpha, \eta, q),$$

$$\kappa_{[r]}^{(p)} = \frac{(r-1)!}{(1-\alpha^r)} T_r(\alpha, \eta, q) \qquad \text{and} \qquad \kappa_r^{(p)} = \sum_{j=1}^{r} S(r, j) \frac{(j-1)!}{(1-\alpha^j)} T_j(\alpha, \eta, q),$$

and

$$\mu_1^{(p)} = \frac{T_1(\alpha, \eta, q)}{1-\alpha} \qquad \text{and} \qquad (\sigma^{(p)})^2 = \frac{(1+\alpha) T_1(\alpha, \eta, q) + T_2(\alpha, \eta, q)}{1-\alpha^2}.$$

ACKNOWLEDGMENTS

The authors are very grateful to the referee for comments and suggestions that have significantly improved the presentation of the paper.

REFERENCES

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

[2] Aly, E.-E.A.A. and Bouzar, N. (2020). Stationary underdispersed INAR (1) models based on the backward approach, *arXiv:2103.10471* [**math.ST**].

[3] Benkherouf, L. and Bather, J.A. (1988). Oil exploration: Sequential decisions in the face of uncertainty, *Journal of Applied Probability*, **25**, 529–543.

[4] Freeland, R.K. (1998). Statistical analysis of discrete time series with applications to the analysis of workers compensation claims data, *Ph.D. Thesis*, University of British Columbia, Canada.

[5] Jazi, M.A.; Jones, G. and Lai, C.-D. (2012). Integer-valued AR(1) with geometric innovations, *Journal of the Iranian Statistical Society*, **11**, 173–190.

[6] Johnson, L.N.; Kemp, A.W. and Kotz, S. (2005). *Univariate Discrete Distributions*, Third Ed., John Wiley & Sons Inc.

[7] Kemp, A.W. (1992). Heine-Euler extensions of the Poisson distribution, *Communications in Statistics – Theory and Methods*, **21**, 571–588.

[8] Knopp, K. (1990). *Theory and Applications of Infinite Series*, Dover, New York.

[9] Lazarova, M. and Minkova, L. (2018). *Non-central Polya–Aeppli distribution.* In "Proceedings of the 44th International Conference on Applications of Mathematics in Engineering and Economics" (V. Pasheva, N. Popivanov and G. Venkov, Eds.), AIP Conf. Proc. 2048), AIP Publishing. `https://doi.org/10.1063/1.5082018`

[10] Livio, T.; Khan, N.M.; Bourguignon, M. and Bakouch, H.S. (2018). An INAR(1) model with Poisson–Lindley innovations, *Economics Bulletin*, **38**, 1505–1513.

[11] Long, J.S. (1990). The origins of sex differences in science, *Social Forces*, **68**, 1297–1316.

[12] Long, J.S. and Freese, J. (2014). *Regression Models for Categorical Dependent Variables Using Stata*, Third Edition, A Stata Press Publication, StataCorp LP, College Station, Texas.

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

[14] McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts, *Advances in Applied Probability*, **20**, 822–835.

[15] Mohammadpour, M.; Bakouch, H.S. and Shirozhan, M. (2018). Poisson–Lindley INAR(1) model with applications, *Brazilian Journal of Probabability and Statistics*, **32**, 262–280.

[16] Ong, S.H. and Lee, P.A. (1979). The non-central negative binomial distribution, *Biometrical Journal*, **21**, 611–628.

[17] Rostami, A.; Mahmoudi, E. and Roozegar, R. (2018). A new integer valued AR(1) process with Poisson–Lindely innovations, *arXiv:1802.00994* [**stat.AP**].

[18] Schweer, S. and Weiss, C.H. (2014). Compound Poisson INAR(1processes: Stochastic properties and testing for overdispersion, *Computational Statistics and Data Analysis*, **77**, 267–284.

[19] Schweer, S. and Wichelhaus, C. (2015). Queueing systems of INAR(1) processes with compound Poisson arrivals, *Stochastic Models*, **31**, 618–635.

[20] Sellers, K.F. and Morris, D.S. (2017). Underdispersion models: models that are under the radar, *Communications in Statistics – Theory and Methods*, **46**, 12075–12086.

[21] Smith, P.J. (1995). A recursive formulation of the old problem of obtaining moment from cumulants and vice versa, *American Statistician*, **49**, 217–218.

[22] Steutel, F.W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *Annals of Probability*, **7**, 893–899.

[23] Weiss, C.H. (2009). Modelling time series of counts with overdispersion, *Statistical Methods and Applications*, **18**, 507–519.

[24] Weiss, C.H. and Puig, P. (2015). *The Marginal Distribution of Compound Poisson INAR(1) Processes*. In "Stochastic Models, Statistics and Their Applications" (A. Steland, E. Rafajlowicz and K. Szajowski, Eds.), Springer Proceedings in Mathematics & Statistics, Springer, **122**, 351–359.

REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 347–366
https://doi.org/10.57805/revstat.v21i3.406

The Effects of Ranking Error Models on Mean Estimators Based on Ranked Set Sampling

Authors: Sami Akdeniz
– The Graduate School of Natural and Applied Sciences, Dokuz Eylul University, Izmir, Turkey
samiakdeniz19@gmail.com

Tugba Ozkal Yildiz ✉
– Department of Statistics, Dokuz Eylul University, Izmir, Turkey
tugba.ozkal@deu.edu.tr

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

Abstract:

- Ranked Set Sampling (RSS) is a sampling method commonly used in recent years. This sampling method is especially useful for studies in medicine, agriculture, forestry and ecology. In this study, the widely used ranking error models in RSS literature are investigated. This study is aimed to explore the effects of ranking error models on the mean estimators based on RSS and some of its modified methods such as Extreme RSS (ERSS) and Percentile RSS (PRSS) for different distribution, set and cycle size in infinite population. Monte Carlo simulation study is conducted for this purpose. Additionally, the study is supported by real life data. It is observed that, RSS and some of its modified methods shows better results than Simple Random Sampling (SRS).

Keywords:

- *Ranked set sampling; ranking error models; relative efficiency; mean estimator; abalone dataset.*

AMS Subject Classification:

- 62P10, 62D99, 68U20.

✉ Corresponding author.

1. INTRODUCTION

RSS is developed as an alternative to SRS in order to estimate population parameters more efficiently where the measurement of sampling units is difficult or costly but the units are easier to rank. McIntyre [12] was the first to propose the use of RSS in the pasture research to estimate the mean amount of crops. Afterwards, Halls and Dell [11] used this method to estimate the mean weights of trees and plant leaves in pine forests located in the east of Texas. In order to compare the variances of the means obtained from RSS and SRS methods Evans [7] carried out a study on long leaf pine trees. The first mathematical theory of RSS in infinite population was developed by Takahasi and Wakimoto [24]. They also demonstrated that the estimator of the population mean obtained by RSS is unbiased and its variance is smaller than SRS when the errors in the ranking are ignored. Dell and Clutter [6] examined errors of ranking in RSS. They showed that the mean estimator of RSS is an unbiased estimator of population mean when ranking is imperfect. David and Levine [5] conducted a study to determine the effects of the errors in the ranking in RSS. The concept of concomitant variable for RSS which is an effective way to increase the accuracy of ranking was proposed by Stokes [22]. This variable should be highly correlated with the variable of interest. Also, Stokes [23] suggested RSS based variance estimator which is asmptotically unbiased and more efficient compared to SRS based variance estimator. In order to review other results and examples for RSS see these studies, Patil *et al.* [20] and Al-Omari and Bouza [1]. Also detailed information regarding theoretical and applicational studies based on RSS can be found in Chen *et al.* [4].

Ranking of the units in a set is made on the basis of the visual judgement of the researcher or a concomitant variable which has a strong correlation with the variable of interest. These ranking methods are defined as ranking error models. There are many studies in the literature that are focused on the modelling of ranking errors. Primarily, Dell and Clutter [6] developed a model including a term of random error for the observations. Later, Bohn and Wolfe [3] proposed a ranking error model based on the expected value of the difference between two order statistics. Fligner and MacEachern [9] used the principle of monotone likelihood ratio to model the ranking information in RSS. New class of models is presented for imperfect rankings, in a study carried out by Frey [10]. A calibration model is developed by Ozturk [17] to reduce the errors in the ranking for RSS. Besides, Ozturk [18] suggested inference techniques for ranked set sample data in the presence of judgement ranking errors. Alexandridis and Ozturk [2] developed robust statistical inference against imperfect ranking in a ranked set sample data obtained from a family of discrete distributions. By taking the ranking errors in RSS into account, Ozturk [19] obtained non-parametric maximum likelihood estimators.

The motivation of this study is to see the effects of ranking error models on the mean estimators of RSS and some of its modified methods and compare them with the mean estimator of SRS. For this purpose, a simulation study is conducted. In addition, an abalone data set is used to support the results of the simulation study.

This study consists of six sections. The first section includes the aim of the study and literarure review on RSS. The second section contains methodological background and detailed information about RSS, extreme RSS (ERSS) and percentile RSS (PRSS).

Ranking error models in RSS literature such as visual ranked set sampling (VRSS) and concomitant ranked set sampling (CRSS) are defined in section 3. In addition, a Monte Carlo simulation study is conducted to determine the effects of ranking error models on the MSE of the mean estimators based on RSS and some of its modified methods. Besides, abalone data set is used for comparing the results obtained from the simulation study in section 5. The final section contains the conclusions.

2. RANKED SET SAMPLING AND SOME OF ITS MODIFIED METHODS

2.1. Ranked Set Sampling

In recent years, RSS is a commonly used sampling method in literature. RSS was introduced by McIntyre [12] as an alternative sampling method to SRS in order to estimate the population parameters more efficiently. It is useful and preferable method due to several important factors. Set size and the relative costs of various operations such as sampling, ranking and measurement are the most important ones among these factors. Also RSS provides advantages due to its features such as the ability to work with finite or infinite populations and it does not require to measure all units in the selected sample in RSS.

There are two important parameters in RSS. These are the set size and the number of cycles which are denoted by n and m, respectively. The set size in RSS usually ranges from 2 to 5. Also, there are many studies available in the literature in which more sets are used. On the other hand, there is no limit for the number of cycles. RSS procedure is applied in 5 steps which are described as below:

1. Select a sample of size n^2 from the population of interest using SRS.
2. Divide this randomly chosen sample of size n^2 into n sets with size n.
3. Rank the units within each set via cost effective and straightforward measurement. This ranking can be made by using visual ranking method, a concomitant variable or other methods.
4. Select the smallest ranked unit from the first set, the second smallest ranked unit from the second set and the n-th smallest ranked unit from the n-th set for actual measurement of units.
5. This process is repeated m times, until maintaining the required sample size.

The following expression represents the RSS procedure for one cycle:

$$\begin{bmatrix} \boldsymbol{X_{1[1:n]}} \leq X_{1[2:n]} \leq X_{1[3:n]} \leq \cdots \leq X_{1[n:n]} \\ X_{2[1:n]} \leq \boldsymbol{X_{2[2:n]}} \leq X_{2[3:n]} \leq \cdots \leq X_{2[n:n]} \\ X_{3[1:n]} \leq X_{3[2:n]} \leq \boldsymbol{X_{3[3:n]}} \leq \cdots \leq X_{3[n:n]} \\ \vdots \\ X_{n[1:n]} \leq X_{n[2:n]} \leq X_{n[3:n]} \leq \cdots \leq \boldsymbol{X_{n[n:n]}} \end{bmatrix}.$$

Here, $X_{(i[k:n]j)}$ represents the unit which has the rank of k in the i-th set and j-th cycle where $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$. The obtained ranked set sample for n set and m cycle

can be shown as

$$\begin{bmatrix} X_{[1]1} & X_{[2]1} & X_{[3]1} & \cdots & X_{[i]1} \\ X_{[1]2} & X_{[2]2} & X_{[3]2} & \cdots & X_{[i]2} \\ X_{[1]3} & X_{[2]3} & X_{[3]3} & \cdots & X_{[i]3} \\ \vdots & \vdots & & \vdots & \\ X_{[1]j} & X_{[2]j} & X_{[3]j} & \cdots & X_{[i]j} \end{bmatrix},$$

where $X_{[i]j}$ denotes the i-th ranked observation in the j-th cycle for i and j changing from 1 to n and m, respectively. The sets in RSS are random samples that are elements of the i-th set $X_{[i]1}, X_{[i]2}, ..., X_{[i]j}$ and each set has the same distribution function $F(x;\theta)$ and same probability density function $f(x;\theta)$, where $i = 1, 2, 3, ..., n$ and $j = 1, 2, 3, ..., m$.

The sample mean estimator of the population mean for RSS can be shown as

$$\bar{X}_{\text{RSS}} = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} X_{[i]j}\,. \tag{2.1}$$

Also the variance of the mean estimator for RSS can be shown as

$$Var(\bar{X}_{\text{RSS}}) = \frac{\sigma_x^2}{mn} \left[1 - \sum_{i=1}^{n} \frac{(E(X_{[i]j}) - \mu_x)^2}{n\sigma_x^2} \right], \tag{2.2}$$

where, μ_x and σ_x^2 are the mean and the variance of the population of interest, respectively.

2.2. Extreme Ranked Set Sampling

ERSS is developed by Samawi *et al.* [21] to estimate the population parameters more efficiently than SRS with the same number of units by only using the minimum and maximum ranked units for n when it is even and, the median ranked unit when it is odd.

For example, when $n = 6$, extreme ranked set sample is given below:

$$\begin{bmatrix} \boldsymbol{X_{1[1:6]}} \le X_{1[2:6]} \le X_{1[3:6]} \le X_{1[4:6]} \le X_{1[5:6]} \le X_{1[6:6]} \\ \boldsymbol{X_{2[1:6]}} \le X_{2[2:6]} \le X_{2[3:6]} \le X_{2[4:6]} \le X_{2[5:6]} \le X_{2[6:6]} \\ \boldsymbol{X_{3[1:6]}} \le X_{3[2:6]} \le X_{3[3:6]} \le X_{3[4:6]} \le X_{3[5:6]} \le X_{3[6:6]} \\ X_{4[1:6]} \le X_{4[2:6]} \le X_{4[3:6]} \le X_{4[4:6]} \le X_{4[5:6]} \le \boldsymbol{X_{4[6:6]}} \\ X_{5[1:6]} \le X_{5[2:6]} \le X_{5[3:6]} \le X_{5[4:6]} \le X_{5[5:6]} \le \boldsymbol{X_{5[6:6]}} \\ X_{6[1:6]} \le X_{6[2:6]} \le X_{6[3:6]} \le X_{6[4:6]} \le X_{6[5:6]} \le \boldsymbol{X_{6[6:6]}} \end{bmatrix}.$$

Since the set size $n = 6$ is even, the actual measurement of units is made over the smallest ranked units ($X_{1[1:6]}$, $X_{2[1:6]}$, $X_{3[1:6]}$) from the first three sets and the largest ranked units ($X_{4[6:6]}$, $X_{5[6:6]}$, $X_{6[6:6]}$) from the last three sets, where $X_{i[m:n]}$ represents the m-th ranked unit in the i-th set for $i = 1, 2, ..., n$, $m = 1, 2, ..., n$. On the other hand, an example

for odd set size, $n = 7$, is given below:

$$\begin{bmatrix} \boldsymbol{X_{1[1:7]}} \leq X_{1[2:7]} \leq X_{1[3:7]} \leq X_{1[4:7]} \leq X_{1[5:7]} \leq X_{1[6:7]} \leq X_{1[7:7]} \\ \boldsymbol{X_{2[1:7]}} \leq X_{2[2:7]} \leq X_{2[3:7]} \leq X_{2[4:7]} \leq X_{2[5:7]} \leq X_{2[6:7]} \leq X_{2[7:7]} \\ \boldsymbol{X_{3[1:7]}} \leq X_{3[2:7]} \leq X_{3[3:7]} \leq X_{3[4:7]} \leq X_{3[5:7]} \leq X_{3[6:7]} \leq X_{3[7:7]} \\ X_{4[1:7]} \leq X_{4[2:7]} \leq X_{4[3:7]} \leq X_{4[4:7]} \leq X_{4[5:7]} \leq X_{4[6:7]} \leq \boldsymbol{X_{4[7:7]}} \\ X_{5[1:7]} \leq X_{5[2:7]} \leq X_{5[3:7]} \leq X_{5[4:7]} \leq X_{5[5:7]} \leq X_{5[6:7]} \leq \boldsymbol{X_{5[7:7]}} \\ X_{6[1:7]} \leq X_{6[2:7]} \leq X_{6[3:7]} \leq X_{6[4:7]} \leq X_{6[5:7]} \leq X_{6[6:7]} \leq \boldsymbol{X_{6[7:7]}} \\ X_{7[1:7]} \leq X_{7[2:7]} \leq X_{7[3:7]} \leq \boldsymbol{X_{7[4:7]}} \leq X_{7[5:7]} \leq X_{7[6:7]} \leq X_{7[7:7]} \end{bmatrix}.$$

In this case, the actual measurement of units is made over the smallest ranked units ($X_{1[1:7]}$, $X_{2[1:7]}$, $X_{3[1:7]}$) from the first three sets and the largest ranked units ($X_{4[7:7]}$, $X_{5[7:7]}$, $X_{6[7:7]}$) from the following three sets. In addition, the fourth ranked unit ($X_{7[4:7]}$) is selected from the remaining set for the measurement where $X_{i[m:n]}$ represents the m-th ranked unit in the i-th set for $i = 1, 2, ..., n$, $m = 1, 2, ..., n$. For this case, the last unit corresponds to the median value of the last set in the sample.

For even set size, the mean estimator of ERSS is given by

$$\bar{X}_{\text{ERSS}} = \frac{1}{n}\left[\sum_{i=1}^{n/2} X_{2i-1[1:n]} + \sum_{i=1}^{n/2} X_{2i[n:n]}\right]. \tag{2.3}$$

Also, the variance of the mean estimator based on ERSS is given by

$$Var(\bar{X}_{\text{ERSS}}) = \frac{1}{n^2}\left[\sum_{i=1}^{n/2} Var(X_{2i-1[1:n]}) + \sum_{i=1}^{n/2} Var(X_{2i[n:n]})\right]. \tag{2.4}$$

For odd set size, the mean estimator of ERSS is given by

$$\bar{X}_{\text{ERSS}} = \frac{1}{n}\left[\sum_{i=1}^{(n-1)/2} X_{2i-1[1:n]} + \sum_{i=1}^{(n-1)/2} X_{2i[n:n]} + X_{n[(n-1/2):n]}\right]. \tag{2.5}$$

Also, the variance of the mean estimator based on ERSS is given by

$$Var(\bar{X}_{\text{ERSS}}) = \frac{1}{n^2}\left[\sum_{i=1}^{(n-1)/2} Var(X_{2i-1[1:n]}) + \sum_{i=1}^{(n-1)/2} Var(X_{2i[n:n]}) + Var(X_{n[((n+1)/2):n]})\right]. \tag{2.6}$$

2.3. Percentile Ranked Set Sampling

PRSS is suggested by Muttlak [13] to estimate the population parameters more efficiently than SRS with the same number of units by only using the $[p(n+1)]$-th and $[q(n+1)]$-th ranked units for n when it is even and, the median ranked unit when it is odd.

In this sampling method, p is denoted as the percentile value and takes value between 0 and 1, $(0 < p < 1)$. On the other hand, $q = 1 - p$ and $[p(n+1)]$ and $[q(n+1)]$ are rounded to the nearest integer. PRSS procedures are presented in the following examples.

Let the set size be $n = 6$, $p = 0.35$ and $q = 0.65$. Following the ranking process of units, the second ranked units from the first half of the sets ($X_{1[2:6]}$, $X_{2[2:6]}$, $X_{3[2:6]}$) and the fifth ranked units from the following three sets ($X_{4[5:6]}$, $X_{5[5:6]}$, $X_{6[5:6]}$) are selected:

$$\begin{bmatrix} X_{1[1:6]} \leq \boldsymbol{X_{1[2:6]}} \leq X_{1[3:6]} \leq X_{1[4:6]} \leq X_{1[5:6]} \leq X_{1[6:6]} \\ X_{2[1:6]} \leq \boldsymbol{X_{2[2:6]}} \leq X_{2[3:6]} \leq X_{2[4:6]} \leq X_{2[5:6]} \leq X_{2[6:6]} \\ X_{3[1:6]} \leq \boldsymbol{X_{3[2:6]}} \leq X_{3[3:6]} \leq X_{3[4:6]} \leq X_{3[5:6]} \leq X_{3[6:6]} \\ X_{4[1:6]} \leq X_{4[2:6]} \leq X_{4[3:6]} \leq X_{4[4:6]} \leq \boldsymbol{X_{4[5:6]}} \leq X_{4[6:6]} \\ X_{5[1:6]} \leq X_{5[2:6]} \leq X_{5[3:6]} \leq X_{5[4:6]} \leq \boldsymbol{X_{5[5:6]}} \leq X_{5[6:6]} \\ X_{6[1:6]} \leq X_{6[2:6]} \leq X_{6[3:6]} \leq X_{6[4:6]} \leq \boldsymbol{X_{6[5:6]}} \leq X_{6[6:6]} \end{bmatrix}.$$

This time, let $n = 7$, $p = 0.4$ and $q = 0.6$. Following the ranking process of units, the third ranked units from the first three sets ($X_{1[3:7]}$, $X_{2[3:7]}$, $X_{3[3:7]}$), the fifth ranked units from the following three sets ($X_{4[5:7]}$, $X_{5[5:7]}$, $X_{6[5:7]}$) and the median unit ($X_{7[4:7]}$) of the last set are selected:

$$\begin{bmatrix} X_{1[1:7]} \leq X_{1[2:7]} \leq \boldsymbol{X_{1[3:7]}} \leq X_{1[4:7]} \leq X_{1[5:7]} \leq X_{1[6:7]} \leq X_{1[7:7]} \\ X_{2[1:7]} \leq X_{2[2:7]} \leq \boldsymbol{X_{2[3:7]}} \leq X_{2[4:7]} \leq X_{2[5:7]} \leq X_{2[6:7]} \leq X_{2[7:7]} \\ X_{3[1:7]} \leq X_{3[2:7]} \leq \boldsymbol{X_{3[3:7]}} \leq X_{3[4:7]} \leq X_{3[5:7]} \leq X_{3[6:7]} \leq X_{3[7:7]} \\ X_{4[1:7]} \leq X_{4[2:7]} \leq X_{4[3:7]} \leq X_{4[4:7]} \leq \boldsymbol{X_{4[5:7]}} \leq X_{4[6:7]} \leq X_{4[7:7]} \\ X_{5[1:7]} \leq X_{5[2:7]} \leq X_{5[3:7]} \leq X_{5[4:7]} \leq \boldsymbol{X_{5[5:7]}} \leq X_{5[6:7]} \leq X_{5[7:7]} \\ X_{6[1:7]} \leq X_{6[2:7]} \leq X_{6[3:7]} \leq X_{6[4:7]} \leq \boldsymbol{X_{6[5:7]}} \leq X_{6[6:7]} \leq X_{6[7:7]} \\ X_{7[1:7]} \leq X_{7[2:7]} \leq X_{7[3:7]} \leq \boldsymbol{X_{7[4:7]}} \leq X_{7[5:7]} \leq X_{7[6:7]} \leq X_{7[7:7]} \end{bmatrix}.$$

For even set size, the mean estimator of PRSS is obtained as

$$\bar{X}_{\mathrm{PRSS}} = \frac{1}{n}\left[\sum_{i=1}^{n/2} X_{i[a:n]} + \sum_{i=(n/2)+1}^{n} X_{i[b:n]}\right]. \tag{2.7}$$

Also, the variance of the mean estimator based on PRSS is obtained as

$$Var(\bar{X}_{\mathrm{PRSS}}) = \frac{1}{n^2}\left[\sum_{i=1}^{n/2} Var(X_{i[a:n]}) + \sum_{i=(n/2)+1}^{n} Var(X_{i[b:n]})\right]. \tag{2.8}$$

For odd set size, the mean estimator of PRSS is obtained as

$$\bar{X}_{\mathrm{PRSS}} = \frac{1}{n}\left[\sum_{i=1}^{(n-1/2)} X_{i[a:n]} + \sum_{i=((n-1)/2)+1}^{n-1} X_{i[b:n]} + X_{i[((n-1)/2):n]}\right). \tag{2.9}$$

Also, the variance of the mean estimator based on PRSS is obtained as

$$Var(\bar{X}_{\mathrm{PRSS}}) = \frac{1}{n^2}\left[\sum_{i=1}^{(n-1)/2} Var(X_{i[a:n]}) \sum_{i=((n-1)/2)+1}^{n-1} Var(X_{i[b:n]}) + Var(X_{i[(n+1/2):n]})\right], \tag{2.10}$$

where $a = [p(n+1)]$ and $b = [q(n+1)]$.

3. RANKING ERROR MODELS

3.1. Visual Ranked Set Sampling

Visual judgement ranking is firstly noted by McIntyre [12] to estimate the mean amount of products. This ranking method is a subjective ranking method since the ranking of units in the set is based on the personal judgement of the researcher. The reliability of visual ranking depends on the knowledge and experience of the researcher based on the subject of study and also on the materials used to rank the units.

Modelling the i-th visual score V_i was suggested by Dell and Clutter [6]. This model is given as follows:

$$V_i = X_i + \tau_i\,, \tag{3.1}$$

where

V_i : i-th visual judgement order statistic,

X_i : i-th true order statistic,

τ_i : i-th random error term where $\tau_i \sim \text{iid}(0, \sigma_\tau^2)$ and X_i's are mutually independent of τ_i's.

In RSS, visual ranking process can be defined as follows:

1. Generate $V_i = X_i + \tau_i$ with $\tau_i \sim \text{iid}(0, \sigma_\tau^2)$ where X_i's and τ_i's are mutually independent.

2. Rank the visual scores ($V_1, V_2, ..., V_n$) from the lowest to the highest.

3. In the last step select the sampling unit corresponding to the r-th visual score (V_r) and measure the $X_{[r]}$ value for this unit.

This method is called Visual Ranked Set Sampling (VRSS). The correlation between visual judgement order statistic (V) and true order statistic (X) is computed by the following equation proposed by Nahhas *et al.* [14, 15]:

$$\rho_{xv} = \frac{\sigma_x}{\sqrt{\sigma_x^2 + \sigma_\tau^2}}\,. \tag{3.2}$$

3.2. Concomitant Ranked Set Sampling

In RSS, another method used to rank the units in the set is concomitant variable (Y) based ranking which is suggested by Stokes [22]. The concomitant variable (Y) is a variable that has a high correlation with the variable of interest (X). The accuracy of the ranking is increased by using this variable. As an example, to estimate the mean weight of a certain number of fish belonging to a population, a researcher may use a concomitant variable, such as fish size, which has a high correlation with the fish weight.

David and Levine [5] were the first to study concomitant variable (Y). Detailed information and some limiting assumptions for concomitant variable (Y) were developed by Stokes [22] in order to determine its effects on RSS. These assumptions are given as follows:

- There is a linear relationship between concomitant variable (Y) and the variable of interest (X).
- Standardized concomitant variable (Y) and the standardized variable of interest (X) have identical distribution.

Concomitant based ranking can be modelled as

$$X_i = \mu_x + \frac{\rho_{xy}\sigma_x}{\sigma_y}(Y_i - \mu_y) + \tau_i\,, \tag{3.3}$$

where

μ_x: the mean of the variable of interest (X),
σ_x: the standard deviation of the variable of interest (X),
μ_y: the mean of the concomitant variable (Y),
σ_y: the standard deviation of the concomitant variable (Y),
ρ_{xy}: the correlation between the variable of interest (X) and concomitant variable (Y),
X_i: the i-th observation on the variable of interest (X),
Y_i: the i-th observation on the concomitant variable (Y),
τ_i: i-th random error term.

The random error term is independent identically distributed (iid) with mean 0 and variance σ_τ^2 and τ_i's and Y_i's are mutually independent. The stepwise period of ranking the units in the set with respect to the concomitant variable is given below:

1. Generate Equation (3.3) where τ_i's and Y_i's are mutually independent.

2. The Y_i's are ranked from the lowest to the highest to obtain the Y_i order statictics $Y_1 \leq \cdots \leq Y_n$.

3. Select the r-th correctly ranked order statistic Y_r and measure the r-th true order statistic $X = X_r$ from the sampling unit.

This method is defined as Concomitant Ranked Set Sampling (CRSS) method.

4. A MONTE CARLO SIMULATION STUDY

Our basic goal in this simulation study is to investigate the effects of ranking error models on the mean estimators based on RSS, ERSS and PRSS. For this reason, bias and MSE of the mean estimators are computed and compared with MSE of mean estimator based on SRS for different set and cycle sizes, distributions and ranking error models such as VRSS and CRSS in infinite population. The simulation study is performed via R Project with 10000 repetitions. In the simulation study:

- The population of the variable of interest (X) and concomitant variable (Y) are generated from $N(0,1)$ (symmetric), $Uniform(0,1)$ (symmetric), $Exp(1)$ (right skewed) and $Gamma(4,2)$ (right skewed) distributions with size (N) 10000.
- Set sizes (n) are determined to be 2, 3, 4 and 5. Also cycle sizes (m) are determined to be 5 and 10.
- Four sampling methods are used. These sampling methods are SRS, RSS, ERSS and PRSS. (In this study, p and q value for PRSS is determined as 0.4 and 0.6, respectively. $p = 0.4$ and $q = 0.6$ values were used in the simulation study since they offer the best results for PRSS.)
- For CRSS, the correlation values between the variable of interest (X) and concomitant variable (Y) ρ_{xy} are determined as 0.95, 0.75, 0.50 and 0.25. (The same values for ρ_{xv} and ρ_{xy} were used in the simulation study.)
- For VRSS, the random error term $\tau_i \sim N(0, \sigma_\tau^2)$. For the distributions used in the simulation study, the ρ_{xv} values corresponding to σ_τ^2 were calculated by Equation (3.2). These values are given in the table below.

Table 1: The values of σ_τ^2 corresponding to ρ_{xv} for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$.

ρ_{xv}	σ_τ^2			
	$N(0,1)$	$Uniform(0,1)$	$Exp(1)$	$Gamma(4,2)$
0.95	0.108	0.009	0.108	1.7285
0.75	0.778	0.0649	0.778	12.4444
0.50	3	0.25	3	48
0.25	15	1.25	15	240

The bias and mean squared error (MSE) of an estimator $\hat{\theta}$ of a parameter θ formulas given below are used in the simulation study:

$$Bias(\hat{\theta}) = \hat{\theta} - \theta, \tag{4.1}$$

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2. \tag{4.2}$$

Note that, θ represents the population mean (μ) and $\hat{\theta}$ represents the mean estimators of population mean based on SRS ($\bar{X}_{\text{SRS}}$), RSS ($\bar{X}_{\text{RSS}}$), ERSS ($\bar{X}_{\text{ERSS}}$) and PRSS ($\bar{X}_{\text{PRSS}}$), respectively. The performance of the mean estimators of RSS, ERSS and PRSS is compared with respect to SRS in terms of relative efficiency criteria. The relative efficiency formulas given below are used:

$$RE_1(\bar{X}_{\text{RSS}}, \bar{X}_{\text{SRS}}) = \frac{MSE(\bar{X}_{\text{SRS}})}{MSE(\bar{X}_{\text{RSS}})}, \tag{4.3}$$

$$RE_2(\bar{X}_{\text{ERSS}}, \bar{X}_{\text{SRS}}) = \frac{MSE(\bar{X}_{\text{SRS}})}{MSE(\bar{X}_{\text{ERSS}})}, \tag{4.4}$$

$$RE_3(\bar{X}_{\text{PRSS}}, \bar{X}_{\text{SRS}}) = \frac{MSE(\bar{X}_{\text{SRS}})}{MSE(\bar{X}_{\text{PRSS}})}. \tag{4.5}$$

The comparisons of the mean estimators are constructed in terms of bias, mean squared error and relative efficiency for different correlation levels, variances of the random error term, set and cycle sizes. The results of the simulation study with 10000 repetitions are presented in tables.

Table 2 and Table 3 show bias values of mean estimators in VRSS and CRSS. The results indicate that:

- For symmetric distributions, the bias values obtained from mean estimators of RSS, ERSS and PRSS are close to 0. This means the mean estimators of RSS, ERSS and PRSS are unbiased estimators of population mean for symmetric distributions.
- For right skewed distributions, the bias values obtained from the mean estimator of RSS are close to symmetric distributions. On the other hand, for right skewed distributions, the bias values obtained from mean estimators of ERSS and PRSS are far from 0 when the set size increases. This means the mean estimators of ERSS and PRSS are biased estimators of population mean when the set size increases.

Table 2: Bias values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in VRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xv}=0.95$			$\rho_{xv}=0.75$			$\rho_{xv}=0.50$			$\rho_{xv}=0.25$		
			RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS
$N(0,1)$	5	2	−0.001	−0.001	0.003	−0.001	0.000	−0.006	−0.003	0.002	0.000	0.002	−0.003	0.000
		3	0.003	−0.001	0.002	0.000	−0.003	−0.004	0.000	0.000	0.001	−0.003	0.000	0.000
		4	0.001	0.003	−0.001	0.001	−0.001	0.002	−0.001	−0.001	−0.001	0.002	0.002	−0.004
		5	0.000	0.000	0.001	0.000	0.001	−0.001	−0.004	−0.001	−0.001	−0.004	−0.002	0.002
	10	2	−0.003	0.000	0.003	−0.003	0.002	−0.001	0.004	0.001	−0.004	−0.002	0.001	0.001
		3	−0.002	−0.002	0.000	0.001	−0.001	0.001	0.000	−0.003	0.000	0.000	0.001	0.001
		4	0.000	0.000	0.002	0.000	−0.001	0.000	0.002	0.001	0.000	0.002	−0.002	0.000
		5	0.000	0.001	0.001	0.000	0.001	−0.001	0.000	−0.002	−0.001	0.001	0.001	0.000
$Uniform(0,1)$	5	2	−0.002	0.000	−0.001	−0.001	0.000	0.001	0.000	0.000	0.001	0.000	0.000	−0.001
		3	0.000	0.000	0.000	−0.001	0.000	0.001	0.000	0.000	0.000	0.001	0.001	0.001
		4	0.000	0.000	−0.001	0.000	−0.001	−0.001	0.001	0.000	0.000	0.001	0.001	−0.001
		5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	10	2	0.000	0.000	0.000	−0.002	0.000	−0.001	0.001	0.000	0.001	0.000	0.000	−0.001
		3	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001
		4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	0.000
		5	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	−0.001	0.000	0.000	0.000	0.000
$Exp(1)$	5	2	−0.004	0.006	0.005	−0.002	−0.002	0.006	0.005	0.004	−0.001	−0.001	−0.003	0.004
		3	0.000	0.002	−0.178	−0.003	0.001	−0.149	−0.002	−0.003	−0.094	−0.002	−0.001	−0.033
		4	−0.002	0.179	−0.177	−0.001	0.150	−0.154	0.001	0.092	−0.092	−0.001	0.032	−0.028
		5	−0.004	0.159	−0.160	−0.001	0.136	−0.139	0.001	0.082	−0.087	0.002	0.029	−0.024
	10	2	0.003	0.001	−0.001	0.000	−0.002	0.003	0.001	0.002	0.002	0.000	−0.001	−0.001
		3	−0.001	−0.001	−0.176	0.000	−0.002	−0.152	−0.001	0.000	−0.091	0.001	0.000	−0.028
		4	−0.001	0.177	−0.178	0.000	0.150	−0.153	0.000	0.092	−0.091	0.000	0.032	−0.027
		5	0.000	0.158	−0.159	0.000	0.137	−0.137	0.000	0.084	−0.082	0.000	0.027	−0.026
$Gamma(4,2)$	5	2	0.001	0.005	−0.032	0.000	−0.025	0.004	−0.016	−0.010	0.013	−0.018	−0.012	0.005
		3	0.005	−0.003	−0.399	0.000	−0.002	−0.371	0.006	−0.005	−0.239	0.014	0.009	−0.064
		4	0.006	0.401	−0.392	0.005	0.359	−0.375	0.014	0.230	−0.205	0.001	0.062	−0.061
		5	−0.003	0.347	−0.350	0.005	0.320	−0.319	−0.012	0.205	−0.201	−0.003	0.055	−0.064
	10	2	0.009	−0.008	0.011	−0.009	−0.003	−0.002	0.002	−0.006	0.011	0.016	−0.008	−0.005
		3	0.009	0.002	−0.380	0.007	−0.001	−0.365	0.005	−0.002	−0.222	−0.008	−0.005	−0.060
		4	0.002	0.387	−0.390	0.005	0.354	−0.371	−0.006	0.227	−0.221	0.003	0.060	−0.064
		5	−0.003	0.350	−0.349	−0.001	0.329	−0.325	−0.002	0.186	−0.194	0.001	0.062	−0.053

Table 3: Bias values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in CRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xy}=0.95$			$\rho_{xy}=0.75$			$\rho_{xy}=0.50$			$\rho_{xy}=0.25$		
			RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS
$N(0,1)$	5	2	0.001	0.003	0.001	0.004	−0.004	0.000	−0.006	−0.005	0.003	−0.001	0.001	0.002
		3	0.000	0.001	−0.006	0.001	−0.001	0.005	−0.004	−0.003	−0.002	−0.002	0.002	−0.002
		4	0.002	−0.003	−0.003	0.002	−0.001	0.000	0.000	0.001	0.005	0.001	0.001	0.004
		5	0.001	0.005	−0.002	−0.002	0.004	0.000	−0.001	0.000	−0.004	−0.002	0.005	−0.001
	10	2	0.000	−0.002	0.000	0.003	0.001	−0.001	−0.003	−0.002	0.001	0.000	0.002	−0.007
		3	0.001	0.001	0.003	0.003	0.001	−0.001	0.000	0.001	−0.005	−0.003	0.000	−0.004
		4	0.000	0.001	0.002	0.003	−0.001	−0.002	0.001	0.003	0.001	0.000	0.000	0.001
		5	0.001	0.003	0.005	0.000	0.003	0.004	0.000	0.008	−0.002	0.000	0.005	0.003
$Uniform(0,1)$	5	2	−0.002	0.000	0.000	0.001	0.000	0.000	−0.001	0.001	0.000	0.000	0.000	−0.001
		3	0.000	−0.001	0.001	−0.001	−0.001	0.001	0.001	0.000	0.000	0.000	0.000	−0.001
		4	0.000	0.001	−0.001	−0.001	−0.001	−0.001	0.000	−0.001	0.001	−0.001	0.000	0.000
		5	0.000	0.000	0.000	0.000	0.001	−0.001	0.001	0.000	−0.001	0.000	0.000	0.002
	10	2	−0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	0.000	−0.002	−0.001
		3	0.000	0.001	−0.001	0.000	0.000	0.001	0.000	0.000	0.000	−0.001	0.000	−0.002
		4	0.000	−0.001	0.000	0.000	0.000	0.002	0.000	−0.001	0.000	0.000	0.001	−0.001
		5	0.000	0.000	0.000	−0.001	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000
$Exp(1)$	5	2	−0.002	0.001	0.001	0.001	0.001	0.004	−0.001	−0.001	−0.002	−0.002	0.000	−0.001
		3	0.002	0.003	−0.152	0.002	0.002	−0.094	0.001	0.002	−0.046	−0.003	0.002	0.001
		4	0.002	0.152	−0.149	−0.003	0.094	−0.099	0.001	0.045	−0.039	−0.004	0.006	0.001
		5	0.000	0.134	−0.135	0.000	0.086	−0.081	−0.001	0.029	−0.032	−0.001	0.012	0.002
	10	2	−0.003	−0.002	0.001	0.000	−0.001	0.003	−0.001	0.000	0.001	−0.002	0.003	0.000
		3	0.001	0.000	−0.148	−0.002	0.002	−0.096	0.000	0.002	−0.045	0.001	0.002	−0.002
		4	0.000	0.151	−0.153	0.002	0.045	−0.088	−0.001	0.040	−0.047	0.001	0.007	−0.002
		5	0.000	0.143	−0.133	0.000	0.029	−0.079	0.002	0.034	−0.038	−0.001	0.009	0.000
$Gamma(4,2)$	5	2	0.005	−0.004	−0.001	0.010	0.010	0.011	0.003	0.003	0.005	−0.018	0.002	0.016
		3	0.010	0.003	−0.324	0.003	0.008	−0.193	0.001	0.002	−0.110	0.007	0.005	−0.054
		4	0.004	0.320	−0.330	0.008	0.203	−0.235	0.007	0.104	−0.043	−0.003	−0.004	−0.023
		5	0.002	0.278	−0.285	0.006	0.180	−0.182	0.003	0.064	−0.068	0.015	0.010	−0.013
	10	2	−0.005	0.009	−0.003	−0.014	−0.007	0.005	−0.013	0.006	0.001	−0.004	0.009	−0.006
		3	0.002	−0.001	−0.346	−0.007	0.006	−0.205	0.008	−0.015	−0.095	−0.009	0.011	−0.014
		4	−0.004	0.335	−0.334	0.003	0.209	−0.199	−0.007	0.079	−0.098	0.005	0.014	−0.030
		5	0.004	0.275	−0.301	−0.005	0.195	−0.185	−0.006	0.086	−0.081	−0.004	0.022	−0.014

Table 4 and Table 5 show MSE values of the mean estimators in VRSS and CRSS. The results indicate that:

- Based on Table 4 and Table 5, the smallest and the highest MSE values were obtained from $Uniform(0,1)$ and $Gamma(4,2)$, respectively.
- MSE values obtained from $N(0,1)$ are less than $Exp(1)$.

Table 4: MSE values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in VRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xv}=0.95$			$\rho_{xv}=0.75$			$\rho_{xv}=0.50$			$\rho_{xv}=0.25$		
			RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS
$N(0,1)$	5	2	0.072	0.072	0.071	0.085	0.082	0.082	0.092	0.092	0.091	0.096	0.099	0.099
		3	0.038	0.038	0.033	0.048	0.049	0.046	0.058	0.059	0.057	0.065	0.064	0.066
		4	0.025	0.028	0.021	0.034	0.036	0.033	0.044	0.044	0.041	0.048	0.049	0.049
		5	0.017	0.019	0.015	0.026	0.027	0.024	0.033	0.034	0.033	0.039	0.039	0.039
	10	2	0.035	0.036	0.036	0.041	0.041	0.041	0.045	0.046	0.046	0.048	0.048	0.048
		3	0.019	0.019	0.017	0.024	0.024	0.023	0.029	0.030	0.028	0.032	0.033	0.032
		4	0.012	0.014	0.011	0.017	0.018	0.016	0.021	0.021	0.021	0.024	0.024	0.024
		5	0.008	0.010	0.007	0.013	0.013	0.012	0.017	0.017	0.016	0.019	0.019	0.019
$Uniform(0,1)$	5	2	0.006	0.006	0.006	0.007	0.007	0.007	0.008	0.008	0.008	0.008	0.008	0.008
		3	0.003	0.003	0.004	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005
		4	0.002	0.001	0.002	0.003	0.002	0.003	0.004	0.003	0.004	0.004	0.004	0.004
		5	0.001	0.001	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003
	10	2	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.004	0.004
		3	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003
		4	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		5	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002
$Exp(1)$	5	2	0.077	0.077	0.079	0.084	0.085	0.087	0.095	0.095	0.092	0.096	0.096	0.100
		3	0.045	0.044	0.058	0.052	0.052	0.057	0.061	0.058	0.054	0.063	0.065	0.060
		4	0.029	0.071	0.049	0.037	0.072	0.048	0.044	0.063	0.042	0.048	0.055	0.045
		5	0.020	0.053	0.039	0.028	0.054	0.038	0.034	0.047	0.034	0.038	0.043	0.036
	10	2	0.040	0.039	0.039	0.043	0.043	0.044	0.046	0.047	0.046	0.049	0.050	0.048
		3	0.021	0.022	0.045	0.026	0.026	0.041	0.030	0.030	0.031	0.033	0.033	0.030
		4	0.014	0.052	0.041	0.018	0.047	0.036	0.022	0.035	0.025	0.024	0.027	0.023
		5	0.010	0.039	0.032	0.014	0.037	0.029	0.017	0.028	0.020	0.019	0.021	0.019
$Gamma(4,2)$	5	2	1.166	1.181	1.161	1.348	1.360	1.370	1.494	1.456	1.480	1.576	1.572	1.578
		3	0.629	0.622	0.661	0.802	0.803	0.834	0.942	0.952	0.911	1.066	1.035	1.010
		4	0.412	0.672	0.480	0.556	0.750	0.634	0.692	0.801	0.672	0.758	0.784	0.742
		5	0.291	0.460	0.350	0.416	0.579	0.463	0.544	0.618	0.536	0.606	0.624	0.605
	10	2	0.599	0.583	0.590	0.659	0.670	0.664	0.737	0.738	0.761	0.786	0.778	0.770
		3	0.321	0.319	0.403	0.401	0.402	0.481	0.479	0.475	0.477	0.511	0.515	0.503
		4	0.207	0.390	0.315	0.274	0.448	0.372	0.347	0.438	0.360	0.385	0.402	0.372
		5	0.145	0.299	0.238	0.213	0.344	0.291	0.269	0.324	0.279	0.308	0.326	0.302

Table 5: MSE values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in CRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xy}=0.95$			$\rho_{xy}=0.75$			$\rho_{xy}=0.50$			$\rho_{xy}=0.25$		
			RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS
$N(0,1)$	5	2	0.073	0.071	0.072	0.081	0.082	0.081	0.091	0.090	0.091	0.101	0.096	0.097
		3	0.038	0.039	0.032	0.047	0.050	0.045	0.059	0.059	0.056	0.064	0.064	0.063
		4	0.024	0.026	0.021	0.034	0.036	0.032	0.043	0.042	0.041	0.050	0.048	0.047
		5	0.016	0.020	0.014	0.026	0.026	0.024	0.033	0.033	0.032	0.038	0.037	0.037
	10	2	0.035	0.036	0.032	0.040	0.039	0.040	0.045	0.046	0.045	0.050	0.048	0.049
		3	0.018	0.018	0.016	0.024	0.023	0.022	0.029	0.028	0.028	0.032	0.032	0.030
		4	0.011	0.013	0.010	0.016	0.018	0.015	0.021	0.022	0.020	0.024	0.024	0.023
		5	0.008	0.009	0.007	0.013	0.013	0.011	0.016	0.017	0.016	0.018	0.018	0.018
$Uniform(0,1)$	5	2	0.006	0.006	0.006	0.007	0.007	0.007	0.008	0.007	0.008	0.008	0.008	0.008
		3	0.003	0.003	0.004	0.004	0.004	0.004	0.005	0.005	0.005	0.006	0.005	0.005
		4	0.002	0.002	0.002	0.003	0.003	0.003	0.004	0.003	0.004	0.004	0.004	0.004
		5	0.001	0.001	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003
	10	2	0.003	0.003	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		3	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003
		4	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		5	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002
$Exp(1)$	5	2	0.074	0.076	0.075	0.088	0.085	0.089	0.092	0.087	0.088	0.091	0.101	0.007
		3	0.044	0.041	0.049	0.052	0.052	0.051	0.065	0.061	0.054	0.065	0.066	0.005
		4	0.028	0.061	0.040	0.035	0.054	0.038	0.044	0.051	0.042	0.004	0.004	0.003
		5	0.019	0.045	0.031	0.027	0.041	0.029	0.034	0.036	0.033	0.003	0.003	0.003
	10	2	0.037	0.038	0.039	0.042	0.041	0.042	0.045	0.048	0.047	0.004	0.004	0.004
		3	0.021	0.021	0.035	0.024	0.025	0.030	0.030	0.030	0.028	0.002	0.002	0.002
		4	0.013	0.042	0.032	0.018	0.030	0.021	0.022	0.024	0.021	0.002	0.002	0.002
		5	0.010	0.035	0.024	0.013	0.025	0.017	0.018	0.021	0.017	0.001	0.001	0.001
$Gamma(4,2)$	5	2	1.173	1.138	1.165	1.363	1.317	1.313	1.453	1.490	1.415	1.579	1.564	1.624
		3	0.638	0.630	0.618	0.804	0.775	0.768	0.940	0.950	0.901	1.037	1.062	1.054
		4	0.413	0.587	0.418	0.555	0.660	0.549	0.700	0.709	0.677	0.776	0.742	0.737
		5	0.290	0.408	0.312	0.423	0.483	0.425	0.525	0.562	0.543	0.589	0.586	0.617
	10	2	0.590	0.571	0.605	0.681	0.633	0.667	0.719	0.724	0.751	0.788	0.782	0.796
		3	0.304	0.308	0.375	0.396	0.399	0.400	0.468	0.478	0.465	0.508	0.510	0.498
		4	0.203	0.363	0.277	0.277	0.343	0.292	0.343	0.365	0.330	0.391	0.395	0.363
		5	0.145	0.246	0.210	0.209	0.270	0.226	0.276	0.287	0.277	0.302	0.316	0.308

Table 6 and Table 7 show RE values of mean estimators in VRSS and CRSS. RE values obtained from the simulation study which are greater than 1 mean that RSS, ERSS or PRSS are more efficient than SRS:

- RE values obtained from symmetric distributions give better results than right skewed distributions.

Table 6: RE values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in VRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xv}=0.95$			$\rho_{xv}=0.75$			$\rho_{xv}=0.50$			$\rho_{xv}=0.25$		
			RE_1	RE_2	RE_3	RE_1	RE_2	RE_3	RE_1	RE_2	RE_3	RE_1	RE_2	RE_3
$N(0,1)$	5	2	1.422	1.387	1.412	1.146	1.223	1.236	1.077	1.101	1.119	1.044	1.010	1.020
		3	1.746	1.754	1.975	1.356	1.398	1.464	1.165	1.146	1.178	1.036	1.026	1.006
		4	1.989	1.819	2.381	1.453	1.401	1.484	1.118	1.131	1.194	1.027	1.036	1.011
		5	2.323	2.028	2.646	1.587	1.461	1.626	1.214	1.195	1.197	1.006	1.020	1.009
	10	2	1.417	1.386	1.389	1.212	1.185	1.225	1.153	1.079	1.076	1.057	1.033	1.012
		3	1.750	1.764	1.940	1.377	1.385	1.453	1.137	1.109	1.228	1.024	1.005	1.020
		4	2.086	1.881	2.361	1.448	1.407	1.622	1.162	1.191	1.173	1.062	1.052	1.038
		5	2.379	2.111	2.673	1.509	1.471	1.589	1.184	1.152	1.218	1.064	1.028	1.011
$Uniform(0,1)$	5	2	1.417	1.394	1.455	1.257	1.228	1.232	1.099	1.104	1.101	1.009	1.027	1.032
		3	1.830	1.879	1.521	1.458	1.429	1.193	1.140	1.160	1.081	1.050	1.044	1.029
		4	2.181	2.928	1.852	1.524	1.876	1.314	1.164	1.303	1.105	1.055	1.052	0.990
		5	2.511	3.119	2.112	1.636	1.922	1.419	1.246	1.325	1.134	1.067	1.036	1.037
	10	2	1.477	1.447	1.411	1.242	1.224	1.240	1.081	1.060	1.062	1.016	0.992	1.031
		3	1.857	1.785	1.492	1.426	1.442	1.157	1.128	1.176	1.072	1.072	1.015	1.013
		4	2.218	2.801	1.755	1.530	1.879	1.318	1.162	1.257	1.045	1.046	1.085	1.050
		5	2.474	3.067	2.052	1.670	2.000	1.417	1.187	1.275	1.152	1.044	1.007	1.009
$Exp(1)$	5	2	1.296	1.229	1.278	1.188	1.190	1.168	1.031	1.074	1.121	1.071	1.017	1.005
		3	1.539	1.550	1.140	1.267	1.306	1.203	1.111	1.136	1.298	1.066	1.008	1.126
		4	1.692	0.701	1.022	1.384	0.696	1.044	1.134	0.793	1.193	1.060	0.910	1.159
		5	1.953	0.778	1.066	1.422	0.757	1.068	1.158	0.839	1.193	1.036	0.946	1.114
	10	2	1.266	1.272	1.270	1.174	1.135	1.188	1.068	1.076	1.101	1.022	1.028	1.058
		3	1.596	1.594	0.737	1.221	1.279	0.811	1.119	1.106	1.081	1.019	1.001	1.088
		4	1.780	0.479	0.621	1.370	0.537	0.688	1.142	0.735	0.994	1.010	0.917	1.078
		5	1.968	0.515	0.625	1.430	0.548	0.694	1.228	0.724	0.994	1.027	0.962	1.088
$Gamma(4,2)$	5	2	1.391	1.344	1.371	1.186	1.179	1.152	1.073	1.107	1.067	1.046	1.038	1.006
		3	1.718	1.718	1.595	1.307	1.342	1.306	1.141	1.098	1.198	0.984	1.016	1.066
		4	1.953	1.188	1.624	1.469	1.075	1.255	1.154	1.004	1.208	1.036	1.036	1.091
		5	2.211	1.420	1.840	1.555	1.094	1.365	1.179	1.035	1.194	1.048	1.039	1.096
	10	2	1.344	1.384	1.355	1.191	1.186	1.222	1.119	1.065	1.043	1.046	1.020	1.076
		3	1.630	1.666	1.339	1.352	1.342	1.084	1.126	1.110	1.102	1.041	1.013	1.057
		4	1.923	1.019	1.273	1.456	0.884	1.078	1.137	0.930	1.105	1.004	0.999	1.067
		5	2.221	1.071	1.328	1.486	0.942	1.097	1.207	1.012	1.129	1.026	0.991	1.044

Table 7: RE values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in CRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xy}=0.95$			$\rho_{xy}=0.75$			$\rho_{xy}=0.50$			$\rho_{xy}=0.25$		
			RE_1	RE_2	RE_3	RE_1	RE_2	RE_3	RE_1	RE_2	RE_3	RE_1	RE_2	RE_3
$N(0,1)$	5	2	1.351	1.398	1.361	1.188	1.236	1.180	1.115	1.060	1.072	1.024	1.020	0.993
		3	1.739	1.728	2.028	1.401	1.360	1.449	1.118	1.147	1.144	1.061	0.998	1.044
		4	2.050	1.890	2.302	1.452	1.388	1.511	1.166	1.185	1.220	1.048	1.043	1.041
		5	2.395	2.074	2.676	1.509	1.562	1.620	1.194	1.154	1.206	1.040	1.033	1.064
	10	2	1.405	1.402	1.511	1.213	1.222	1.224	1.082	1.062	1.107	0.999	1.043	1.038
		3	1.740	1.726	1.962	1.345	1.382	1.430	1.147	1.149	1.152	1.043	1.019	1.061
		4	2.127	1.882	2.372	1.558	1.317	1.588	1.188	1.138	1.223	1.029	0.995	1.041
		5	2.349	2.062	2.735	1.559	1.529	1.686	1.182	1.137	1.216	1.065	1.068	1.041
$Uniform(0,1)$	5	2	1.474	1.408	1.442	1.193	1.251	1.208	1.079	1.106	1.093	1.022	1.000	1.047
		3	1.804	1.789	1.509	1.388	1.386	1.233	1.116	1.163	1.085	0.994	1.035	1.022
		4	2.202	2.670	1.844	1.497	1.648	1.340	1.186	1.206	1.101	1.043	1.025	1.029
		5	2.470	2.921	2.153	1.559	1.698	1.374	1.168	1.254	1.120	1.053	1.056	1.015
	10	2	1.447	1.414	1.416	1.214	1.197	1.225	1.097	1.094	1.102	1.008	1.022	1.022
		3	1.752	1.866	1.545	1.393	1.369	1.275	1.108	1.092	1.104	1.055	1.021	0.992
		4	2.177	2.755	1.802	1.417	1.601	1.308	1.132	1.214	1.117	1.031	1.044	1.067
		5	2.513	2.923	2.182	1.569	1.693	1.415	1.208	1.199	1.158	1.041	1.022	0.995
$Exp(1)$	5	2	1.323	1.284	1.297	1.167	1.186	1.182	1.099	1.077	1.102	1.017	1.010	1.040
		3	1.526	1.501	1.347	1.288	1.293	1.387	1.018	1.110	1.221	1.034	1.027	1.009
		4	1.705	0.802	1.286	1.362	0.944	1.312	1.101	1.009	1.197	1.028	1.011	1.017
		5	1.912	0.882	1.271	1.381	0.986	1.324	1.112	1.079	1.182	1.021	0.991	1.050
	10	2	1.384	1.281	1.232	1.165	1.177	1.145	1.097	1.064	1.040	1.023	0.998	0.998
		3	1.539	1.551	0.912	1.321	1.224	1.077	1.097	1.102	1.138	0.997	1.070	1.042
		4	1.801	0.586	0.754	1.313	0.810	1.150	1.166	0.970	1.087	1.039	1.028	0.996
		5	1.939	0.614	0.818	1.480	0.826	1.077	1.143	0.979	1.162	1.054	0.961	1.023
$Gamma(4,2)$	5	2	1.370	1.364	1.374	1.180	1.198	1.231	1.121	1.063	1.053	1.022	0.999	0.967
		3	1.681	1.688	1.773	1.300	1.357	1.371	1.122	1.120	1.147	1.024	1.007	1.053
		4	1.926	1.373	1.907	1.427	1.259	1.471	1.180	1.097	1.116	1.018	1.075	1.052
		5	2.201	1.530	2.040	1.481	1.284	1.505	1.204	1.138	1.175	1.043	1.062	1.068
	10	2	1.336	1.366	1.338	1.190	1.227	1.198	1.086	1.089	1.066	1.041	1.017	0.988
		3	1.741	1.729	1.442	1.358	1.371	1.330	1.103	1.145	1.146	1.067	1.047	1.032
		4	1.989	1.081	1.510	1.443	1.171	1.356	1.143	1.097	1.220	1.001	1.020	1.087
		5	2.165	1.260	1.573	1.532	1.204	1.380	1.111	1.104	1.199	1.047	1.000	1.016

According to the results obtained from the simulation study:

For VRSS:

- When the number of set size increases, relative efficiency increases.
- When the variance of random error term (σ_τ^2) increases, the relative efficiency decreases. On the other hand, when the correlation between visual judgement order statistic (V) and true order statistic (X) decreases, the relative efficiency decreases.

For CRSS:

- When the number of set size increases, relative efficiency increases.
- When the correlation between the variable of interest (X) and the concomitant variable (Y) increases, relative efficiency increases.

For VRSS and CRSS:

- The number of cycles didn't cause a regular increase or decrease in relative efficiency for VRSS and CRSS. For this reason, exact comment can not be made about the effect of number of cycles on relative efficiency.
- In both visual and concomitant based ranking methods MSE decreases when set size and number of cycles increase.
- MSE increases as the variance of the error term increases in visual ranking and as the correlation between the concomitant variable (Y) and the variable of interest (X) variable decreases in concomitant based ranking.
- In both visual and concomitant based ranking methods MSE values obtained from right skewed distributions are greater than the MSE values obtained from symmetric distributions.
- In both visual and concomitant based ranking methods, the bias, MSE and RE values from mean estimators based on RSS, ERSS and PRSS for symmetric distributions and right skewed distributions are similar.

5. REAL DATA APPLICATION

Abalone is a common name given to a group of small to very large sea snails, marine gastropod molluscs which are the member of Haliotidae family [8]. Age of an abalone can be determined by making some physical measurements which, in advance, include cutting and staining of the shell. After the staining process, the rings become clear and they are counted under a microscope to obtain age information. Estimating the age of abalone includes difficult, costly and time-consuming physical measurements. Therefore, it forces us to use alternative measurement techniques. A new physical measurement method which is easier than the others in estimating the age of abalone is proposed by Nash *et al.* [16]. This data set is taken from `https://archive.ics.uci.edu/ml/datasets/abalone` [25]. Abalone dataset includes 4177 samples with 9 variables. Information about these variables are given in the table below:

Table 8: Descriptions of abalone dataset.

Variable	Data Type	Measurament Unit of Data	Description
Length	Continuous	mm	Longest shell measurement
Diameter	Continuous	mm	Perpendicular to length
Height	Continuous	mm	With meat in shell
Whole weight	Continuous	gr	Whole abalone
Shucked weight	Continuous	gr	Weight of meat
Viscera weight	Continuous	gr	Gut weight (after bleeding)
Shell weight	Continuous	gr	After being died
Rings	Integer	—	+1.5 gives the age in years
Sex	Nominal	—	Male, Female and Infant

Rings variable is selected as the variable of interest (X). For concomitant based ranking, Shell weight (Y_1) and Shucked weight (Y_2) are determined as concomitant variables. The correlations between variable of interest and concomitant variables are given in table below:

Table 9: Correlations between variable of Interest (X) and concomitant variables (Y's) in abalone dataset for CRSS.

Variable of Interest (X)	Concomitant Variable (Y)	Correlations
Rings	Shell Weight	0.627
	Shucked Weight	0.420

The results obtained from abalone dataset using CRSS are given in Table 10 and Table 11, respectively.

Table 10: MSE(bias) values for CRSS based on RSS, ERSS and PRSS.

ρ_{xy}	m	n	RSS	ERSS	PRSS
0.627	5	2	0.891 (−0.029)	0.887 (−0.023)	0.909 (−0.013)
		3	0.558 (−0.019)	0.554 (−0.000)	0.543 (−0.113)
		4	0.395 (−0.004)	0.433 (0.127)	0.383 (−0.119)
		5	0.306 (−0.009)	0.329 (−0.103)	0.295 (0.099)
	10	2	0.443 (−0.002)	0.449 (0.003)	0.453 (−0.015)
		3	0.279 (0.001)	0.278 (−0.001)	0.273 (0.120)
		4	0.201 (0.001)	0.219 (0.119)	0.202 (0.116)
		5	0.152 (−0.002)	0.173 (−0.107)	0.152 (0.093)
0.420	5	2	0.979 (0.064)	0.951 (0.009)	0.981 (−0.006)
		3	0.610 (−0.008)	0.624 (−0.010)	0.751 (0.320)
		4	0.440 (−0.001)	0.518 (−0.314)	0.571 (0.309)
		5	0.346 (−0.004)	0.403 (−0.271)	0.450 (0.269)
	10	2	0.485 (−0.000)	0.467 (−0.013)	0.479 (0.003)
		3	0.305 (−0.008)	0.306 (−0.004)	0.429 (0.329)
		4	0.236 (0.003)	0.313 (0.324)	0.342 (0.319)
		5	0.176 (−0.040)	0.242 (−0.276)	0.258 (0.270)

Table 11: RE values for CRSS based on RSS, ERSS and PRSS.

ρ_{xy}	m	n	RE_1	RE_2	RE_3
0.627	5	2	1.181	1.153	1.180
		3	1.215	1.266	1.280
		4	1.298	1.192	1.384
		5	1.358	1.268	1.425
	10	2	1.203	1.150	1.118
		3	1.245	1.242	1.278
		4	1.269	1.163	1.268
		5	1.334	1.178	1.320
0.420	5	2	1.062	1.081	1.058
		3	1.135	1.086	0.900
		4	1.158	1.006	0.889
		5	1.193	1.020	0.885
	10	2	1.050	1.119	1.063
		3	1.104	1.133	0.800
		4	1.095	0.810	0.756
		5	1.138	0.855	0.822

Suppose that the ρ_{xv} values are 0.627 and 0.420, respectively. For VRSS, we need to find the value of standard deviation of the Rings variable (X). This value is $\sqrt{\sigma_x^2} = \sigma_x = \sqrt{10.395}$ = 3.224. Then, we need to find the values of σ_τ^2 corresponding to ρ_{xv}. We use Equation (3.2) to obtain the values of σ_τ^2 corresponding to ρ_{xv}. These values are given in the table below:

Table 12: The values of σ_τ^2 corresponding to ρ_{xv} for Rings variable in abalone dataset.

ρ_{xv}	σ_τ^2
0.627	16.048
0.420	48.536

Table 13: MSE(bias) values for VRSS based on RSS, ERSS and PRSS.

ρ_{xv}	m	n	RSS	ERSS	PRSS
0.627	5	2	0.921 (−0.008)	0.939 (0.001)	0.934 (−0.004)
		3	0.567 (0.011)	0.573 (−0.021)	0.537 (−0.256)
		4	0.415 (0.002)	0.560 (0.247)	0.398 (−0.247)
		5	0.321 (0.005)	0.421 (0.219)	0.318 (−0.231)
	10	2	0.463 (−0.009)	0.456 (−0.001)	0.450 (−0.001)
		3	0.287 (0.002)	0.291 (−0.004)	0.299 (0.250)
		4	0.205 (−0.001)	0.305 (0.247)	0.224 (−0.251)
		5	0.156 (−0.004)	0.235 (0.224)	0.188 (−0.231)
0.420	5	2	0.973 (−0.002)	0.986 (−0.005)	0.986 (−0.016)
		3	0.635 (−0.003)	0.639 (0.004)	0.583 (−0.152)
		4	0.480 (0.010)	0.536 (0.142)	0.441 (−0.143)
		5	0.375 (0.005)	0.423 (0.131)	0.350 (−0.122)
	10	2	0.511 (0.005)	0.496 (−0.017)	0.481 (0.007)
		3	0.328 (0.006)	0.325 (0.002)	0.310 (−0.140)
		4	0.234 (0.003)	0.279 (0.145)	0.231 (−0.138)
		5	0.179 (−0.002)	0.216 (0.127)	0.181 (−0.124)

Table 14: RE values for VRSS based on RSS, ERSS and PRSS.

ρ_{xv}	m	n	RE_1	RE_2	RE_3
0.627	5	2	1.128	1.093	1.093
		3	1.247	1.293	1.293
		4	1.286	0.954	1.312
		5	1.293	0.994	1.277
	10	2	1.108	1.144	1.178
		3	1.176	1.204	1.119
		4	1.259	0.842	1.128
		5	1.291	0.865	1.095
0.420	5	2	1.067	1.044	1.056
		3	1.067	1.088	1.201
		4	1.059	0.952	1.183
		5	1.104	0.968	1.224
	10	2	0.994	1.010	1.106
		3	1.036	1.074	1.114
		4	1.098	0.914	1.122
		5	1.152	0.971	1.145

6. CONCLUSION

In this study, we aimed to use ranking error models (VRSS and CRSS) to compare the bias and MSE of the mean estimators based on RSS and some of its modified methods such as ERSS and PRSS.

For this reason the effects of ranking errors in RSS and in some of its modified methods are examined in the simulation study. In this study, it is deduced that ranking errors may occur depending on the ranking method used. In VRSS, σ_τ^2 and ρ_{xv} change depending on the researcher's knowledge, experience and materials used in the study. The greater knowledge of researcher involved in the study and the use of more appropriate materials would yield a higher accuracy in the ranking. On the other hand, for CRSS, the accuracy of the ranking depends on the correlation between the variable of interest (X) and the concomitant variable (Y) and the distribution of (X, Y). Generally, when $\rho_{xy} \geq 0.5$, the error in the ranking decreases and the accuracy of the ranking increases. Thus, better results can be achieved by minimizing the error in the ranking. The application is performed using abalone data set in order to support the simulation study performed in the section 5. It is seen that similar results were obtained in real data application and simulation study. It is observed that, RSS and some of its modified methods such as ERSS and PRSS methods show better results than SRS.

REFERENCES

[1] Al-Omari, A.I. and Bouza, C.N. (2014). Review of ranked set sampling: modifications and applications, *Investigacion Operacional*, **35**(3), 215–241.

[2] Alexandridis, R. and Ozturk, O. (2009). Robust inference against imperfect ranking in ranked set sampling, *Communications in Statistics Theory and Methods*, **38**(12), 2126–2143.

[3] Bohn, L.L. and Wolfe, D.A. (2014). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the mannwhitney-wilcoxon statistic, *Journal of the American Statistical Association*, **89**(425), 168–176.

[4] Chen, Z.; Bai, Z. and Sinha, B. (2003). *Ranked Set Sampling: Theory and Applications*, Springer Science & Business Media, New York.

[5] David, H. and Levine, D. (1972). Ranked set sampling in the presence of judgment error, *Biometrics*, **28**, 553–555.

[6] Dell, T. and Clutter, J. (1972). Ranked set sampling theory with order statistics background, *Biometrics*, 545–555.

[7] Evans, M. (1967). Application of ranked set sampling to regeneration surveys in areas direct-seeded to longleaf pine, *Master of Forestry Dissertation*, Louisiana State University, Baton Rouge, LA.

[8] Fleming C.A. (1952). Notes on the genus Haliotis (Mollusca). A new subgenus from New Zealand and a new species from the late Cenozoic of Ohope, Bay of Plenty, *Transactions of the Royal Society of New Zealand*, **101**, 229–232.

[9] Fligner, M.A. and MacEachern, S.N. (2006). Nonparametric two-sample methods for ranked-set sample data, *Journal of the American Statistical Association*, **101**(475), 1107–1118.

[10] Frey, J.C. (2007). New imperfect rankings models for ranked set sampling, *Journal of Statistical planning and Inference*, **37**(4), 1433–1445.

[11] Halls, L.K. and Dell, T.R. (1966). Trial of ranked-set sampling for forage yields, *Forest Science*, **12**(1), 22–26.

[12] McIntyre, G. (1952). A method for unbiased selective sampling, using ranked sets, *Australian Journal of Agricultural Research*, **3**(4), 385–390.

[13] Muttlak, H.A. (2003). Modified ranked set sampling methods, *Pakistan Journal of Statistics – All Series*, **19**(3), 315–324.

[14] Nahhas, R.W.; Wolfe, D.A. and Chen, H. (2002). Ranked set sampling: cost and optimal set size, *Biometrics*, **58**(4), 964–971.

[15] Nahhas, R.W.; Wolfe, D.A. and Chen, H. (2004). Ranked set sampling: ranking error models and estimation of visual judgment error variance, *Biometrical Journal*, **46**(2), 255–263.

[16] Nash, W.J.; Sellers, T.L.; Talbot, S.R.; Cawthorn, A.J. and Ford, W.B. (1994). The population biology of abalone. Technical report. Sea Fisheries Division, Australia, *Biometrics*, **48**, 1–69.

[17] Ozturk, O. (2007). Statistical inference under a stochastic ordering constraint in ranked set sampling, *Nonparametric Statistics*, **19**(3), 131–144.

[18] Ozturk, O. (2008). Inference in the presence of ranking error in ranked set sampling, *Canadian Journal of Statistics*, **36**(4), 577–594.

[19] Ozturk, O. (2010). Nonparametric maximum-likelihood estimation of within-set ranking errors in ranked set sampling, *Journal of Nonparametric Statistics*, **22**(7), 823–840.

[20] Patil, G.; Sinha, A. and Taillie, C. (1999). Ranked set sampling: a bibliography, *Biometrical Journal*, **38**(5), 577–586.

[21] Samawi, H.M.; Ahmed, M.S. and Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling, *Environmental and Ecological Statistics*, **6**(1), 91–98.

[22] Stokes, S. (1977). Ranked set sampling with concomitant variables, *Communications in Statistics – Theory and Methods*, **6**(12), 1207–1211.

[23] Stokes, S. (1980). Estimation of variance using judgment ordered ranked set samples, *Biometrics*, 35–42.

[24] Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering, *Annals of the Institute of Statistical Mathematics*, **20**(1), 1–31.

[25] The abalone dataset. Retrieved June 1, 2019, available from:
`https://archive.ics.uci.edu/ml/datasets/abalone`

REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 367–388
https://doi.org/10.57805/revstat.v21i3.357

A Computational Approach to Confidence Intervals and Testing for Generalized Pareto Index Using the Greenwood Statistic

Authors: Marek Arendarczyk
– Mathematical Institute, University of Wroclaw, Poland
marendar@math.uni.wroc.pl

Tomasz J. Kozubowski
– Department of Mathematics and Statistics, University of Nevada Reno, USA
tkozubow@unr.edu

Anna K. Panorska ✉
– Department of Mathematics and Statistics, University of Nevada Reno, USA
ania@unr.edu

Received: April 2020 Revised: January 2022 Accepted: January 2022

Abstract:

- The generalized Pareto distributions (GPDs) play an important role in the statistics of extremes. We point various problems with the likelihood-based inference for the index parameter α of the GPDs, and develop alternative testing strategies, which do not require parameter estimation. Our test statistic is the Greenwood statistic, which probability distribution is stochastically increasing with respect to α within the GPDs. We compare the performance of our test to a test with maximum-to-sum ratio test statistic R_n. New results on the properties of the R_n are also presented, as well as recommendations for calculating the p-values and illustrative data examples.

Keywords:

- *coefficient of variation; extremes; generalized Pareto distribution; heavy tailed distribution; power law; peak-over-threshold.*

AMS Subject Classification:

- Primary: 62F03, 62F25, 65C60; Secondary: 60E15, 62G32, 62P12, 65C05.

✉ Corresponding author.

1. INTRODUCTION

Generalized Pareto distributions (GPDs), along with the generalized extreme value distributions (GEV), play a central role in the theory and applications of the statistics of extremes. Important (monographic) references include Gumbel (1958) [29], Leadbetter *et al.* (1983) [35], Castillo (1988) [9], Beirlant *et al.* (2004) [4], Embrechts *et al.* (1997) [23], Kotz and Nadarajah (2000) [33], Reiss and Thomas (2001) [44], Finkenstädt and Rootzén (2003) [24], Coles (2004) [19], Castillo *et al.* (2005) [11], de Haan and Ferreira (2006) [21], Chavez *et al.* (2016) [18], and Dey and Yan (2016) [22]. In this work, we revisit the important issue of statistical inference for the tail index α within the class of the GPD. In particular, we develop sound, simulation-enabled testing and interval estimation procedures for α with the focus on small samples.

Recall that a GP random variable X can be described through the stochastic representation

$$X \stackrel{d}{=} \frac{1}{\beta}\frac{1}{\alpha}(e^{\alpha E} - 1), \quad \alpha \in \mathbb{R}, \beta > 0, \tag{1.1}$$

where E is a standard exponential random variable, β is the scale parameter, and α is the index parameter (tail index for $\alpha > 0$). The corresponding survival function (SF) of X in (1.1) is of the form

$$S(x) = \mathbb{P}(X > x) = (1 + \alpha\beta x)^{-1/\alpha}. \tag{1.2}$$

For $\alpha > 0$ we get Pareto II (Lomax) distributions with power law tails of order α while for $\alpha = 0$, understood in the limiting sense, the variable X in (1.1) reduces to an exponential random variable with mean $1/\beta$, and the probability density function (PDF)

$$f(x) = \beta e^{-\beta x}, \text{ for } x \in \mathbb{R}_+ = (0, \infty). \tag{1.3}$$

Both, Lomax and exponential distributions are supported on the positive half-line $\mathbb{R}_+$. For $\alpha < 0$, GPDs are re-scaled beta distributions with compact support on the interval $(0, -1/(\alpha\beta))$, and include, for instance, the uniform distribution for $\alpha = -1$. The importance of this family comes from the Peak Over Threshold (POT) theory (see, e.g., Balkema and de Haan, 1974 [3]; Pickands, 1975 [42]), where the GPDs provide natural approximations for the *excess* (or *exceedence*) random variables $X = Y - d|Y > d$ for large classes of random variables Y, where d is a high threshold. This approximation property, coupled with their power-law tail behavior for $\alpha > 0$, make GPDs very relevant and commonly used in insurance mathematics, hydrology, climate science and other areas where the observations over high thresholds are of primary importance.

Our main contribution is a mathematically rigorous procedure for testing and constructing confidence intervals (CIs) for the index α within the GPD family, with the focus on small samples. Our methodology is based on the *Greenwood statistic*,

$$T_n = \frac{\sum_{i=1}^n X_i^2}{(\sum_{i=1}^n X_i)^2}, \tag{1.4}$$

where the $\{X_i\}$ are the underlying data. Since its introduction in Greenwood (1946) [28], this statistic appeared in many different contexts and application areas, and it is closely related

to several other common statistics, such as the sample coefficient of variation

$$CV_n = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}}{\overline{X}} = \sqrt{nT_n - 1}, \tag{1.5}$$

the reciprocal of CV_n known as the *Sharpe's ratio* in finance and insurance applications, the *self-normalized sum* $S_n = T_n^{-1/2}$, (see, e.g., Albrecher *et al.*, 2010 [1] and references therein), and the *student t-statistic*, $ST_n = \sqrt{(n-1)/(nT_n - 1)}$. Since tests and confidence intervals are often based on estimates of parameters, the correctness and practical execution of the estimation procedures is of primary importance for many tests, including the commonly used likelihood ratio procedure. However, for the GPDs there are serious theoretical and computational problems with the standard, likelihood-based inference for α. Indeed, in general, the maximum likelihood estimates (MLEs) for α may not exist or may not be well defined, because without artificial restrictions on the parameter space, the likelihood function is infinite along one of its boundaries. In addition, even when an MLE does exist, the question of uniqueness is still open. The problems with the likelihood function lead to practical issues with numerical calculations of the MLEs and thus implementation of standard likelihood ratio tests for α (see, e.g., Neves *et al.*, 2006 [40]). Further, we noted several errors and inaccuracies in the estimation literature, and we include a review and discussion of the selected key papers on estimation for the GPD family in Appendix A (in the Supplementary Material). The challenges of finding the MLE of α suggest a need for test procedures that do not require estimation of α. Our test based on the Greenwood statistic is an example of such a procedure.

Further, for reliable inference on small samples we need the test statistic to be stochastically monotone (increasing or decreasing) with respect to α in order to be useful for derivation of the critical regions and construction of confidence intervals (CIs) via "inversion of the test" method. Again, the Greenwood statistic satisfies this requirement, as it is stochastically increasing with respect to α within the GPD family (see Arendarczyk *et al.*, 2021 [2]).

The challenges with ML estimation and the need for stochastic order of the test statistic with respect to the parameter of interest are our main motivation for deriving a test based on the Greenwood statistic. In addition, there is a long history of using T_n in testing for exponentiality, where tests based on T_n (or related CV_n) have been shown to be *locally most powerful* within the GPD family (see, e.g., Marohn, 2000 [37]). Further, the statistic T_n comes up naturally in estimation within the GPD and Lomax (Pareto II) families, as shown in Appendix A.

While tests based on the coefficient of variation (or other statistics equivalent to T_n) have already been used for the GPD, most of them focused on testing exponentiality and had rejection regions based on the asymptotic distributions of the test statistics (see, for example, Hasofer and Wang, 1992 [30], Gomes and van Monfort, 1986 [27], Marohn, 2000 [37], Reiss and Thomas, 2001 [44]). In contrast, our approach uses the exact distribution of the test statistic, obtained by straightforward simulations, and is similar in spirit to that of Chaouche and Bacro (2004, 2006) [16, 17] and Tajvidi (2003) [49]. Chaouche and Bacro (2004, 2006) [16, 17] noticed that the population value of their statistic S used for testing exponentiality is increasing in α, and that its empirical distribution shifts to the right with increasing α. Our results formalize these observations and show that the probability distribution of the S statistic computed on a random sample from GPD is stochastically increasing in α *over the entire range* of its values. Castillo *et al.* (2014) [13] considered testing exponentiality

within GPD using a test statistic incorporating several sample CVs computed on different sets of exceedances of varying high thresholds. The power of the proposed test was compared with those of several other tests, including one based on the sample CV, for two alternatives: absolute value of student-t distribution as well as GPD with shape parameter larger than 0 (Pareto). It appeared that tests based on the CV performed best. The paper is not very clear however about the derivation of the critical values, whether these were done by simulations or using the asymptotic distribution of the test statistic. Castillo and Padilla (2015) [14] extended these ideas to the full GPD case with a similar test statistic, based on the asymptotic distribution of the sample CV for GPD samples. Castillo and Serra (2015) [15] focused on the MLEs (see also Castillo and Daoudi, 2009 [12]) and offered brief remarks about testing and interval estimation for α, but no details were provided in this regard. Tajvidi (2003) [49] considered several methods for constructing confidence intervals for the index α in the GPD family using bootstrapping, likelihood ratio (LR) test, and profiling the likelihood function, concluding that the likelihood based methods perform better than the bootstrapping in small to moderate size samples. However, in view of the considerable theoretical and computational difficulties with the MLEs of the GPD parameters, likelihood based inference may not be effective for many data sets. In summary, although substantial work was done towards testing for the GPD tail index, to date, we have not found any test for α that works well on small samples without any restrictions of the values of α.

There is also a rich body of literature on the problem of testing for a GEV domain of attraction (DoA), which is equivalent to testing for a GPD domain of attraction. The importance of this problem is well understood in the extreme value literature. In particular, testing for Gumbel DoA ($\alpha = 0$) is a common need (see, e.g., Fraga Alves and Gomes, 1996 [25]; Gomes and Alpuin, 1986 [26]). When checking domain of attraction, the estimation (or testing) is usually carried out using the excesses of the sample values over a high threshold, taken to be either a predetermined value (see, e.g., Davison and Smith, 1990 [20]) or a particular order statistic (see, e.g., Neves and Fraga Alves, 2007 [39] and the references therein). In the latter case the size of the resulting data set available for inference may be moderate to quite small. Neves and Fraga Alves (2007) [39] considered tests for GEV domain of attraction, particularly the Gumbel DoA. Both of their tests are related to the Greenwood statistic. The main result of the paper is the limiting distribution of (normalized) test statistics, assuming that $\alpha < 1/4$. Recently, Schluter and Trede (2018) [46] used one of the test statistics of Neves and Fraga Alves (2007) [39] for testing Gumbel domain of attraction ($\alpha = 0$) against heavy tailed GPD alternatives. Our results on the properties of the statistic T_n and its version computed on the exceedences (Section 2) show that T_n can be used for the DoA tests on the small samples, which are common in the problems considering exceedences.

Next, we note the tests based on the statistic R_n, involving the ratio of the maximum and the sum of the sample values,

$$R_n = \frac{\bigvee_{i=1}^n X_i}{\sum_{i=1}^n X_i}, \tag{1.6}$$

when the underlying sample $\{X_i\}$ is from a GPD or its domain of attraction (see Neves *et al.*, 2006 [40]). Our interest in this statistic stems from its properties that are similar to those of T_n, which makes R_n a major competitor of T_n in testing. As discussed in Bryson (1974) [6], a test based on R_n is the most powerful for testing exponentiality against uniformity (both special cases of GPDs). The statistic R_n is also mentioned in Chaouche and Bacro (2004) [16] in connection with testing for α within the class of GPDs, and was proposed by Neves *et al.*

(2006) [40] for testing maximum DoA. A simulation study in Neves *et al.* (2006) [40] showed that their test based on R_n, along with the test of Hasofer and Wang (1992) [30], which in essence is based on the T_n, compared very favorably in terms of power with several other tests for $\alpha = 0$ within the GPDs. However, the rejection regions of both tests were based on the asymptotic distributions of the test statistics, which may not work well for small samples. We shall revisit the test based on R_n in Section 3, where we present new results on its properties.

We selected the test based on statistic R_n to perform power comparison with our test based on the Greenwood statistic. However, instead of using the asymptotic distribution of R_n, we use essentially the same numerical procedure to compute (simulated) p-values as in our test. We selected the test based on R_n for the power comparison for the following reasons:

(**1**) A test based on R_n was shown as more powerful than several other tests (Neves *et al.*, 2006 [40]);

(**2**) The R_n statistic is stochastically increasing with respect to α which makes it appropriate for small sample testing;

(**3**) We did not find another test applicable to one- and two-sided hypotheses for α within the entire GPD family.

All other tests have (or should have) some restriction on the range of α where they are applicable.

Our paper is organized as follows. We start with Section 2, where we review the key properties of the statistic T_n. New properties of the statistic R_n are presented in Section 3. The main contribution is Section 4, containing rigorous development of tests and confidence intervals for the index α within the GPD family. This is followed by a limited power comparison between tests based on T_n and R_n in Section 5. Illustrative data examples are presented in Section 6. Proofs are collected in Section 7. Appendix T contains power tables. A review and discussion of the main works on the estimation for GPD family is presented in Appendix A. Both appendixes are available in the Supplementary Material.

2. FUNDAMENTAL PROPERTIES OF THE GREENWOOD STATISTIC T_n

Let $X_1, ..., X_n$ be a random sample from a probability distribution supported on the non-negative real line $\mathbb{R}_+ = [0, \infty)$. It is widely recognized that within the GPD family the distribution of T_n is rather complicated (even under exponentiality) with no closed form expressions for the PDF or the CDF for general n. However, its distribution is *scale-invariant* and bounded ($\frac{1}{n} \leq T_n \leq 1$ for all n), and so all the moments of the Greenwood statistic are finite, even when the distributional moments of the underlying sequence $\{X_i\}$ do not exist. The moments of T_n under exponentiality were derived by Moran (1947) [38], who also established an asymptotic (normal) distribution of T_n for general distributions of $\{X_i\}$ with finite first four moments, noting that the convergence to the limiting normal distribution is rather slow. In particular, under exponentiality, for large n, the Greenwood statistic is approximately normal with mean $2/n$ and variance $4/n^3$. Going beyond light-tail distributions, Albrecher *et al.* (2010) [1] provided exact asymptotic distributions of T_n as n goes to infinity for distributions of $\{X_i\}$ with regularly varying tail.

2.1. Stochastic ordering of T_n

Recall that if X and Y are two random variables with respective CDFs F_X and F_Y and quantile functions (QFs) $Q_X = F_X^{-1}$ and $Q_Y = F_Y^{-1}$, X is said to be stochastically smaller than Y, denoted by $X \leq_{\mathrm{st}} Y$, whenever $F_Y(x) \leq F_X(x)$ for each $x \in \mathbb{R}$. This is the ordinary stochastic order (dominance). On the other hand, X is smaller than Y in the star-shaped order, denoted by $X \leq_* Y$, whenever $Q_Y(u_1)/Q_X(u_1) \leq Q_Y(u_2)/Q_X(u_2)$ for all $u_1 \leq u_2$, so that the function $Q_Y(u)/Q_X(u)$ is non-decreasing in u. For more information on stochastic orders, see, e.g., Belzunce *et al.* (2016) [5].

An important result established in Arendarczyk *et al.* (2021) [2], which is fundamental to this work, shows that:

1. When the underlying distribution of the $\{X_i\}$ is stochastically increasing with respect to the star-shaped order $\leq_*$ then the distribution of T_n is stochastically increasing with respect to the ordinary stochastic order $\leq_{\mathrm{st}}$.
2. The GPDs given by the SF (1.2) are star-shaped ordered with respect to the parameter α.

Therefore, the Greenwood statistic T_n is stochastically increasing with respect to the parameter α within the GPD family. As discussed in Section 4, this key property of T_n plays crucial role in setting up testing and developing confidence intervals for α within this family.

2.2. Symmetry of T_n within the GPD family

In applications, the Greenwood statistic T_n and its functions are often applied to the exceedences $X_{(j)} - X_{(k)}$, where $X_{(1)} < \cdots < X_{(n)}$ are the (ascending) order statistics based on the random sample $X_1, ..., X_n$ and $j = k+1, ..., n$. This leads to the statistic

$$T_{n,k} = \frac{\sum_{i=1}^{n-k} \big(X_{(k+i)} - X_{(k)}\big)^2}{\Big[\sum_{i=1}^{n-k} \big(X_{(k+i)} - X_{(k)}\big)\Big]^2}, \quad k \in \{0, 1, ..., n-1\}, \tag{2.1}$$

where $X_{(0)} = 0$, so that $T_{n,0}$ reduces to T_n. By POT theory, when k is relatively large, then (under appropriate scaling) these $n-k$ exceedences behave as if they were $n-k$ order statistics (based on the sample of size $n-k$) from a GPD (see, e.g., Neves *et al.*, 2006 [40]). This crucial property plays a fundamental role in testing for the extreme domain of attraction (see, e.g., Marohn, 2000 [37]; Neves *et al.*, 2006 [40]; Neves and Fraga Alves, 2007 [39] and the references therein). Since the statistic $T_{n,k}$ is scale invariant, when the underlying distribution of the sample belongs to the domain of attraction of a GPD with index α given by (1.2), the distribution of the $T_{n,k}$ in (2.1) is approximately the same as that of T_{n-k} based on a sample of size $n-k$ from the GPD itself. In other words, we would have an approximate equality in distribution

$$T_{n,k} \overset{d}{\sim} \frac{\sum_{i=1}^{n-k} W_i^2}{(\sum_{i=1}^{n-k} W_i)^2}, \tag{2.2}$$

where the $\{W_i\}$ follow $GP(\alpha, \beta)$. We show below that if the data are generated by a GPD in the first place, then the the distributions of $T_{n,k}$ and T_{n-k} are *exactly* the same.

Proposition 2.1. *Let $n \in \mathbb{N}$ with $n \geq 2$ and let $k \in \{0, ..., n-1\}$. Suppose that $X_1, ..., X_n$ are IID and let $T_{n,k}$ be defined by (2.1). Let $Y_1, ..., Y_{n-k}$ be another random sample, and let T_{n-k} be the Greenwood statistic computed on the Y_i's. Suppose that both random samples are coming from a GPD with the same index α. Then, we have $T_{n,k} \stackrel{d}{=} T_{n-k}$.*

Remark 2.1. This new result complements Proposition 5 in Arendarczyk *et al.* (2021) [2], as it shows that the latter also holds with $k = 0$.

2.3. Limiting behavior of the Greenwood statistic within the GPD family

Another key property of T_n we use in this work is its limiting behavior as the parameter α approaches $\pm\infty$ within the GPD family while the sample size n stays fixed (Arendarczyk *et al.*, 2021 [2]). We include the result here for convenience of the reader.

Proposition 2.2. *Suppose that $n \in \mathbb{N}$ and $X_1, ..., X_n$ are IID and $GP(\alpha, \beta)$ distributed. Then*

$$T_n \xrightarrow{d} 1/n \text{ as } \alpha \to -\infty \text{ and } T_n \xrightarrow{d} 1 \text{ as } \alpha \to \infty. \tag{2.3}$$

In fact, the distribution of the Greenwood statistic T_n on a GPD sample changes continuously within the interval $(1/n, 1)$ as α increases within the interval $(-\infty, \infty)$.

Remark 2.2. It can be shown that as α increases within the range $(-\infty, \infty)$ then, for each $\gamma \in (0, 1)$, the $(1-\gamma) \times 100\%$ percentiles of the distribution of T_n within the GPD family continuously increase from their limiting values of $t_\gamma = 1/n$ at $\alpha = -\infty$ to $t_\gamma = 1$ at $\alpha = \infty$.

This monotone behavior of the quantiles of T_n is important for constructing confidence intervals and testing for α, as discussed in Section 4.

3. FUNDAMENTAL PROPERTIES OF THE STATISTIC R_n

Since we shall use the statistic R_n in the power comparisons in Section 5, we developed new results that facilitate R_n — based testing for α within the class of GPDs. As shown below, T_n and R_n share their key properties.

3.1. Stochastic ordering of R_n

It turns out that R_n computed on a sample from GPD is stochastically increasing with respect to α. This is due to the fact that the GPDs are star-shaped ordered with respect to α and the following new result concerning R_n, whose proof can be found in Section 7.

Theorem 3.1. *Let $\{\mathcal{P}_\theta,\ \theta \in \Theta \subset \mathbb{R}\}$ be a family of absolutely continuous probability distributions on $\mathbb{R}_+$, where for each $\theta_1 \le \theta_2$ we have $X^{(\theta_1)} \le_* X^{(\theta_2)}$, with $X^{(\theta_i)} \sim \mathcal{P}_{\theta_i}$, $i = 1, 2$. Then, for each $n \ge 2$, we have*

$$R_n^{(\theta_1)} \le_{st} R_n^{(\theta_2)} \ \text{ whenever } \theta_1 \le \theta_2,\ \theta_1, \theta_2 \in \Theta, \tag{3.1}$$

where $R_n^{(\theta)}$ is given by (1.6) with the $\{X_i\}$ having a common distribution $\mathcal{P}_\theta$.

3.2. Symmetry of R_n within the GPD family

Let $X_{(1)} \le \cdots \le X_{(n)}$ be the (ascending) order statistics based on a random sample of size n from a GPD (1.2). In analogy with $T_{n,k}$, define

$$R_{n,k} = \frac{X_{(n)} - X_{(k)}}{\sum_{i=1}^{n-k}\left(X_{(k+i)} - X_{(k)}\right)},\ \ k \in \{0, 1, ..., n-1\}, \tag{3.2}$$

where for $k = 0$ we set $X_{(0)} = 0$. This is essentially the statistic R_{n-k} evaluated on the exceedences $X_{(j)} - X_{(k)}$, with $j = k+1, ..., n$, which has been used in this form for testing the extreme DoA (see Neves *et al.*, 2006 [40]). In turn, when the statistic $R_{n-k+1,1}$ is evaluated on the set of $n-k+1$ observations $X_1, ..., X_{n-k+1}$, then we essentially get the statistic R_{n-k} computed on the $n-k$ exceedences $X_{(j)} - X_{(1)}$, with $j = 2, ..., n-k+1$. The following new result, whose proof is provided in Section7, shows that the statistics $R_{n,k}$ and $R_{n-k+1,1}$ have the same distributions.

Proposition 3.1. *Let $n \in \mathbb{N}$ with $n \ge 2$ and let $k \in \{0, 1, ..., n-1\}$. Suppose that $X_1, ..., X_n$ are IID and let $R_{n,k}$ be defined by (3.2). Let $Y_1, ..., Y_{n-k+1}$ be another random sample, and set*

$$R_{n-k+1,1} = \frac{Y_{(n-k+1)} - Y_{(1)}}{\sum_{i=1}^{n-k}\left(Y_{(1+i)} - Y_{(1)}\right)}. \tag{3.3}$$

Then, if the two samples are coming from a GPD with the same α, we have $R_{n,k} \stackrel{d}{=} R_{n-k+1,1}$.

Remark 3.1. The above result implies that if the data are generated by a GPD, then the distributions of $R_{n,k}$ and R_{n-k} are exactly the same, similarly to $T_{n,k}$ and T_{n-k} as shown in Proposition 2.1. Thus, if n and k are large and the statistic $R_{n,k}$ is evaluated on a set of order statistics $X_{(1)} \le \cdots \le X_{(n)}$ of an IID sample from a distribution in the GPD (with index α) domain of attraction then standard arguments from POT theory (see, e.g., Neves *et al.*, 2006 [40]) show that this $R_{n,k}$ behaves as if computed on a sample from the GPD (with the same α).

3.3. Limiting behavior of R_n within the GPD family

As shown below, the statistic R_n behaves very similarly to the Greenwood statistic T_n as the parameter α approaches $\pm\infty$ within the GPD family. The proof of the following new result is included in Section 7.

Proposition 3.2. *Suppose that $n \in \mathbb{N}$ and $X_1, ..., X_n$ are IID and $GP(\alpha, \beta)$ distributed with the SF (1.2). Then*

$$R_n \stackrel{d}{\to} 1/n \ \text{ as } \alpha \to -\infty \ \text{ and } \ R_n \stackrel{d}{\to} 1 \ \text{ as } \alpha \to \infty. \tag{3.4}$$

4. TESTING AND INTERVAL ESTIMATION FOR α WITHIN THE GPD FAMILY

In this section we develop exact tests and provide a rigorous derivation of confidence intervals for the parameter α within the GPD family based on the Greenwood statistic. The CIs are constructed using the standard "inversion of the test" method. In particular, our methodology is very convenient to test for exponential distribution (GPD with $\alpha = 0$) versus Pareto II distribution (GPD with $\alpha > 0$), and has essentially the same power as the likelihood ratio test developed for this special case in Kozubowski *et al.* (2009) [34].

Let $X_1, ..., X_n$ be a random sample from the GPD model (1.2). Since the test statistic T_n is scale-invariant, we can assume for convenience that $\beta = 1$. We start with one-sided tests, followed by two-sided tests, and conclude with procedures for constructing confidence intervals for α.

4.1. One-sided tests for α within the GPD family

Consider the problem of testing

$$H_0 : \alpha \leq \alpha_0 \quad \text{vs.} \quad H_1 : \alpha > \alpha_0, \tag{4.1}$$

where $\alpha \in \mathbb{R}$ is the (unknown) index of the GPD and $\alpha_0 \in \mathbb{R}$ is a known constant. We denote the corresponding partition of the parameter space by $\Omega_0 = \{\alpha : -\infty < \alpha \leq \alpha_0\}$ and $\Omega_1 = \{\alpha : \alpha > \alpha_0\}$.

Our objective is a test δ of size $\gamma \in (0, 1)$ for the hypotheses specified in (4.1). Note that when $\alpha_0 = 0$ this test is a test of a light-tail versus a heavy-tail (Pareto II) distribution within the GPD class. Let T_n be the test statistic for δ. Since the statistic T_n is stochastically increasing with respect to α, the values of α larger than α_0 will be indicated by relatively large values of T_n computed from the sample. Consider the following decision rule for the test δ: Reject H_0 when $T_n > c_n$, where c_n is such that $\mathbb{P}(T_n(\alpha_0) > c_n) = \gamma$, and $\mathbb{P}(T_n(\alpha) \in A)$ denotes the probability of the event $\{T_n \in A\}$ assuming the true value of the parameter is α. That is, the critical number c_n is the $(1 - \gamma) \times 100\%$ percentile of the distribution of T_n when $\alpha = \alpha_0$.

Proposition 4.1. *The test δ described above has size γ and is unbiased for the hypotheses specified in (4.1).*

Remark 4.1. The same decision rule δ can also be used for testing the hypotheses

$$H_0 : \alpha = \alpha_0 \quad \text{vs.} \quad H_1 : \alpha > \alpha_0, \tag{4.2}$$

with the test being unbiased as well.

Next, we consider the problem of testing

$$H_0 : \alpha \geq \alpha_0 \quad \text{vs.} \quad H_1 : \alpha < \alpha_0 \quad \text{or} \quad H_0 : \alpha = \alpha_0 \quad \text{vs.} \quad H_1 : \alpha < \alpha_0 \tag{4.3}$$

with decision rule to reject H_0 when $T_n < d_n$, where d_n is such that

$$\mathbb{P}(T_n(\alpha_0) < d_n) = \gamma. \tag{4.4}$$

These tests are also of size γ and unbiased.

Proposition 4.2. *The test δ for the hypotheses in (4.3) that rejects H_0 whenever $T_n < d_n$ with d_n such that $\mathbb{P}(T_n(\alpha_0) < d_n) = \gamma$, has size γ and is unbiased.*

Since the computation of the p-values is straightforward, we chose to implement our tests using the p-value method. We note that the p-value approach we describe is equivalent to the critical number approach. For the hypotheses in (4.1) and (4.2), the p-value is given by

$$\text{p-value} = \mathbb{P}(T_n(\alpha_0) > t_n), \tag{4.5}$$

where t_n is the observed value of the test statistic T_n. This can be easily seen from the stochasticity of T_n. Similarly, the p-value for the hypotheses in (4.3) is given by

$$\text{p-value} = \mathbb{P}(T_n(\alpha_0) < t_n). \tag{4.6}$$

In practice, one can approximate the p-values for these tests via Monte-Carlo simulation of the probabilities on the right-hand-side in (4.5) or (4.6).

4.2. Two-sided test for α

We now consider the problem of testing

$$H_0 : \alpha = \alpha_0 \quad \text{vs.} \quad H_1 : \alpha \neq \alpha_0. \tag{4.7}$$

Because of the stochastic increasing of T_n with respect to α, the critical region CR for a test δ of the hypotheses in (4.7) should consist of two sections: $CR = [1/n, C_L) \cup (C_R, 1]$. To build a test of size γ, we have a choice of the portion of γ covered by each part of the CR. In general, we can choose any $0 < r < 1$ and assign the following probabilities to the two parts of the critical region:

$$\mathbb{P}(1/n < T_n(\alpha_0) < C_L) = (1 - r)\gamma \quad \text{and} \quad \mathbb{P}(C_R < T_n(\alpha_0) < 1) = r\gamma. \tag{4.8}$$

Thus, the two critical numbers are: C_L equal to the $(1 - r)\gamma 100\%$ percentile, and C_R equal to the $(1 - r\gamma)100\%$ percentile of the distribution of T_n under the null hypothesis. To build the test, consider $R \in (1/n, 1)$ such that $\mathbb{P}(T_n(\alpha_0) < R) = 1 - r$ and $\mathbb{P}(T_n(\alpha_0) > R) = r$. We will consider two cases of the observed value t_n of the test statistic T_n in relation to R. Again, we use the p-value approach to make decisions.

Case 1: The observed value of T_n satisfies $1/n < t_n < R$. Then, $\mathbb{P}(T_n(\alpha_0) < t_n) = (1-r)\gamma_p$ for some $\gamma_p \in (0,1)$. We claim that this γ_p is actually the p-value, so that

$$\text{p-value} = \frac{\mathbb{P}(T_n(\alpha_0) < t_n)}{1-r}. \tag{4.9}$$

Indeed, if the right-hand-side in (4.9) is less than γ, then we must have $\mathbb{P}(T_n(\alpha_0) < t_n) < \gamma(1-r)$, so that by (4.8) we have $t_n < C_L$. Since the value of the test statistic is in the rejection region, the null hypothesis is rejected. On the other hand, if the right-hand-side in (4.9) is greater than γ, so that $\mathbb{P}(T_n(\alpha_0) < t_n) > \gamma(1-r)$, then we must have $t_n > C_L$. At the same time, since $t_n < R$ and $\mathbb{P}(T_n(\alpha_0) < R) = 1-r$, we clearly have $t_n < C_R$. Thus, the observed value of the test statistic is not in the CR, and we fail to reject H_0. Consequently, the p-value is indeed given by (4.9).

Remark 4.2. Many standard tests use $r = 0.5$ in a similar setting, and in practice we recommend that standard choice of r.

Case 2: The observed value of T_n satisfies $R < t_n < 1$. Then, using an argument similar to that used in Case 1, we obtain the following expression for the p-value:

$$\text{p-value} = \frac{\mathbb{P}(T_n(\alpha_0) > t_n)}{r}. \tag{4.10}$$

Again, one can easily approximate the above p-values via Monte-Carlo simulation of the probabilities on the right-hand-side in (4.9) or (4.10). Sample R-codes that return p-values described above are available from the authors upon request.

4.3. Construction of the confidence intervals

We now turn to the derivation of confidence intervals for the index α of the GPD family. We use the classical procedure of "inverting the test" to derive confidence regions, see for example Casella and Berger (2002) [8], Section 9.2.1. We start with one-sided confidence intervals, also known as upper and lower confidence bounds. First, consider the size-γ test δ for the hypotheses in (4.1). Then, a $(1-\gamma)100\%$ confidence set for α is the set of all α_0 for which the null hypothesis is *not rejected* for a given value t_n of the test statistic T_n. The null hypothesis is not rejected when the p-value given in (4.5) is greater than γ, so that

$$\mathbb{P}(T_n(\alpha_0) > t_n) > \gamma. \tag{4.11}$$

By the stochasticity of the test statistic T_n with respect to the parameter α, the set of all α_0 that satisfy this condition is an interval of the form $(\underline{\alpha}, \infty)$, where the quantity $\underline{\alpha} = \underline{\alpha}(t_n)$ satisfies the condition

$$\mathbb{P}(T_n(\underline{\alpha}) > t_n) = \gamma. \tag{4.12}$$

Note that in view of Proposition 2.2 and the remark following it, the quantity $\underline{\alpha}(t_n)$ can always be found for any $t_n \in (1/n, 1)$ and any $\gamma \in (0,1)$. In fact, $\underline{\alpha}(\cdot)$ is a well defined function on the interval $(1/n, 1)$ onto the real line $\mathbb{R}$. This discussion leads to the following result, which provides a lower confidence bound (LCB) for the parameter α within the GPD family.

Proposition 4.3. *Let t_n be the observed value of the test statistic T_n based on the random sample $X_1, ..., X_n$ from a generalized Pareto distribution $GP(\alpha_0, \beta)$. Then we have*

$$\mathbb{P}(\alpha_0 > \underline{\alpha}(T_n(\alpha_0))) = 1 - \gamma, \tag{4.13}$$

so that $(\underline{\alpha}(t_n), \infty)$ is a $(1-\gamma) \times 100\%$ LCB for the parameter α.

Next, consider a size-γ test δ for the hypotheses in (4.3). We can obtain the upper confidence bound (UCB) using similar methods to those employed to find the LCB.

Proposition 4.4. *Let t_n be the observed value of the test statistic T_n based on the random sample $X_1, ..., X_n$ from a generalized Pareto distribution $GP(\alpha_0, \beta)$. Then we have*

$$\mathbb{P}(\alpha_0 < \overline{\alpha}(T_n(\alpha_0))) = 1 - \gamma, \tag{4.14}$$

so that $(-\infty, \overline{\alpha}(t_n))$ is a $(1-\gamma) \times 100\%$ UCB for the parameter α.

Finally, we derive a two-sided $(1-\gamma)100\%$ confidence set for the parameter α by inverting the two-tail test δ for the hypotheses in (4.7). To determine the p-value, we first find the value α^* such that $\mathbb{P}(T_n(\alpha^*) < t_n) = 1 - r$. Then, by the stochasticity of T_n, whenever $\alpha_0 \geq \alpha^*$ we have $\mathbb{P}(T_n(\alpha_0) < t_n) \leq 1 - r$, so that the p-value is given by (4.9). Thus, the null hypothesis is not rejected whenever

$$\mathbb{P}(T_n(\alpha_0) < t_n) > (1 - r)\gamma. \tag{4.15}$$

Since $\mathbb{P}(T_n(\alpha^*) < t_n) = 1 - r$ and the probability on the left-hand-side of (4.15) is monotonically decreasing from $1 - r$ to zero as α_0 is increasing from α^* to ∞, we can find an $\overline{\alpha} \in (\alpha^*, \infty)$ such that

$$\mathbb{P}(T_n(\overline{\alpha}) < t_n) = (1 - r)\gamma. \tag{4.16}$$

Moreover, for all $\alpha_0 \in [\alpha^*, \overline{\alpha})$ the condition (4.15) will be fulfilled. Thus, for these values of α_0 the null hypothesis in (4.7) will not be rejected and consequently the interval $[\alpha^*, \overline{\alpha})$ is part of the confidence set. Similar analysis shows that the interval $(\underline{\alpha}, \alpha^*]$, where the quantity $\underline{\alpha}$ satisfies the condition

$$\mathbb{P}(T_n(\underline{\alpha}) < t_n) = 1 - r\gamma, \tag{4.17}$$

is part of the confidence set as well. Indeed, when $\alpha_0 \leq \alpha^*$ we have $\mathbb{P}_{\alpha_0}(T_n < t_n) \geq 1 - r$, so that the p-value is given by (4.10). Thus, the null hypothesis is not rejected whenever

$$\mathbb{P}(T_n(\alpha_0) > t_n) > r\gamma. \tag{4.18}$$

Since $\mathbb{P}(T_n(\alpha^*) > t_n) = r$ and the probability on the left-hand-side of (4.18) is monotonically increasing from zero to r as α_0 increases from $-\infty$ to α^*, we can find an $\underline{\alpha} \in (-\infty, \alpha^*)$ such that $\mathbb{P}(T_n(\underline{\alpha}) > t_n) = r\gamma$, which is equivalent to (4.17). Moreover, for all $\alpha_0 \in (\underline{\alpha}, \alpha^*]$ the condition (4.18) will be fulfilled. Thus, for these values of α_0 the null hypothesis in (4.7) will not be rejected. At the same time, the above analysis shows that for the values $\alpha_0 < \underline{\alpha}$ and $\alpha_0 > \overline{\alpha}$ the null hypothesis is rejected, so these are not part of the confidence set. In summary, the confidence set obtained by inverting the test is indeed an interval. The following result summarizes this discussion.

Proposition 4.5. *Let t_n be the observed value of the test statistic T_n based on the random sample $X_1, ..., X_n$ from a generalized Pareto distribution $GP(\alpha_0, \beta)$. Then we have*

$$\mathbb{P}(\underline{\alpha}(T_n(\alpha_0))) < \alpha_0 < \overline{\alpha}(T_n(\alpha_0))) = 1 - \gamma, \tag{4.19}$$

so that $(\underline{\alpha}(t_n), \overline{\alpha}(t_n))$ is a $(1-\gamma) \times 100\%$ confidence interval for the parameter α.

5. SIMULATION EXPERIMENTS

To assess the performance of our testing procedures discussed in Section 4, we performed two simulation experiments. First, we did power analysis of the testing procedure based on the statistic T_n in the context of testing $H_0 : \alpha = 0$ vs. $H_1 : \alpha > 0$ (exponentiality against Lomax), which is an important practical problem of detecting a power tail in the context of the GPDs (see, e.g., Kozubowski *et al.*, 2009 [34]). The results are reported in subsection 5.1. In addition, we compared the power functions of tests for α based on T_n and its major competitor R_n within the GPD family. The results are discussed in subsection 5.2.

5.1. Power analysis: Exponential vs. Pareto test

Kozubowski *et al.* (2009) [34] discussed testing exponentiality vs. Pareto distribution, finding the likelihood ratio (LR) approach to be superior (in terms of power) to several other common tests. To compare the performance of the test based on the statistic T_n with that of the LR test, we ran power analysis for the Greenwood test using the same values of α and sample sizes as those used in Kozubowski *et al.* (2009) [34]. The results are shown in Table 1 in Appendix T, which should be compared with Table 4 in Kozubowski *et al.* (2009) [34]. As it turns out, the power of the test based on T_n is very similar as that of the LR test across all values of the parameters, with the LR test having a slight edge. However, it should be noted that the test based on T_n is easier to implement, as it avoids calculating the MLE of the parameter α. In addition, it leads to a confidence *interval* for α, which is not guaranteed when inverting the LR test.

Remark 5.1. The expected loss of power as the sample size goes down is clearly visible in Table 1 of Appendix T. This fact is important for practical consideration of formal test results as well as other measures of fit. In particular, when deciding whether a data set has exponential or Pareto-type tails, it is a common practice to consider a large threshold and perform the analysis on the resulting *exceedances* of the data over the threshold. As one increases the threshold, the number of exceedances used in model fitting and testing decreases, and the test looses power. This may lead to not rejecting exponentiality, when in fact the data have Pareto tails. While the "best" choice of the threshold remains one of the most difficult albeit important problems in analysis of data from the GPD domain of attraction, there are many methods already available for threshold selection, including automatic procedures. Excellent reviews of the existing methods and new methodologies can be found in the following works and the references therein: Kiran and Srinivas (2021) [32], Schneider *et al.* (2021) [47], Silva Lomba and Fraga Alves (2020) [48], and Caeiro and Gomes (2016) [7].

5.2. Power analysis: Tests based on T_n and R_n

Below we provide the results of a limited simulation study of the power of the tests based on T_n and R_n. Note that the process of calculating the p-values is essentially the same for tests based on the R_n and T_n, and it was described in Section 4.

Before presenting the results of the comparison, let us note one slight advantage of the test based on R_n: the probability density function of the test statistic R_n has an explicit form when sampling from the exponential distribution (see, e.g., Qeadan *et al.*, 2012 [43]). In our numerical experiments, we considered four different sets of one-sided hypotheses as follows:

(**1**) $H_0 : \alpha \leq 0$ vs. $H_1 : \alpha > 0$;

(**2**) $H_0 : \alpha \geq 0$ vs. $H_1 : \alpha < 0$;

(**3**) $H_0 : \alpha \leq 1$ vs. $H_1 : \alpha > 1$;

(**4**) $H_0 : \alpha \geq 1$ vs. $H_1 : \alpha < 1$.

For the selected combinations of α and $n = 5, 10, 20, 50, 100$, we generated 1000 samples from the GPD(α, n) and tested the four sets of hypotheses using the statistics T_n and R_n. We then calculated the proportion of times the null hypothesis was rejected in each case, as an approximation of the value of the power function for that α. The results are presented in Tables 2, 3, 5, and 6 in Appendix T. While the two procedures have very similar power overall, the test based on the R_n performs slightly better when testing left-sided alternatives, and the test based on T_n performs a bit better for the right-sided alternatives. We conclude that in practice one may select R_n or T_n based on the alternative hypothesis.

6. ILLUSTRATIVE DATA EXAMPLES

We applied our test and built confidence intervals for α for two commonly used data sets. The purpose of this data analysis is checking if the results from the literature are confirmed by our test. We did not study the fit of the GPD models to the data, as that was beyond the scope of this work. The first data set contains 154 exceedances over 65 m^3/s flow threshold of river Nidd at Hunsingore Weir between 1934 and 1969. This data set was analyzed by Hosking and Wallis (1987) [31], Davison and Smith (1990) [20], Papastahopoulos and Tawn (2012) [41] and Castillo and Serra (2015) [15]. We obtained the data set "nidd.thresh" from R package "evir". The second data set contains 197 exceedances above 7s of zero-crossing hourly mean periods (in seconds) of the waves measured at Bilbao Bay in Spain. This data set was also used in the literature. Notably, it was first used by Castillo and Hadi (1997) [10], and then by Luceño (2006) [36], and Zhang and Stephens (2009) [50]. One can obtain the data "bilbao" from the R package "ercv". Since the data sets were analyzed by many researchers in the past, we only report on the results of the new analysis using the exact methods presented in this paper: test results and CIs. For computation of the p-values we used 10,000 simulations of the values of T_n from data with the distribution specified in the null hypothesis.

6.1. Analysis of the Nidd data

In search for the answer to the question whether the Nidd exceedances are Pareto or exponential, we performed our exact test on exceedances over the same threshold as those reported on p. 126 of Castillo and Serra (2015) [15], that is 65 m^3/s, 75 m^3/s, 85 m^3/s, 95 m^3/s, 100 m^3/s, 110 m^3/s, and 120 m^3/s. We tested the null hypothesis H_0: $\alpha \leq 0$ versus a one sided alternative of Pareto, H_1: $\alpha > 0$. We also computed exact 90% and 95% confidence intervals for α for exceedances over each threshold. We report the value of the number of exceedances used, the test statistic T_n, the p-value of the test, decision (P for Pareto, E for exponential), and the CIs in Table 1. Our conclusion is that the exponential model for the Nidd exceedances can not be rejected.

Table 1: The table contains results of analysis of the Nidd data. Column labeled T contains the threshold, column labeled D contains the decision.

T (m^3/s)	n	T_n	p-value	D	90% CI for α	95% CI for α
65	154	0.0165	0.0041	P	(0.0679, 0.4446)	(0.0456, 0.4916)
75	117	0.0248	0.0006	P	(0.1217, 0.6212)	(0.0941, 0.6859)
85	72	0.0349	0.0255	P	(0.0330, 0.6035)	(0.0021, 0.6999)
95	49	0.0458	0.1282	E	(−0.0634, 0.5863)	(−0.0973, 0.6927)
100	39	0.0514	0.3539	E	(−0.1886, 0.4900)	(−0.2422, 0.6007)
110	31	0.0622	0.4388	E	(−0.2658, 0.5322)	(−0.3323, 0.6533)
120	24	0.0729	0.6509	E	(−0.4920, 0.4334)	(−0.5787, 0.5682)

6.2. Analysis of Bilbao waves data

The question in the literature regarding Bilbao waves data was whether the exceedances are uniform or not. This is equivalent to testing null hypothesis of uniformity, H_0: $\alpha = -1$, versus a two sided alternative of Pareto, H_1: $\alpha \neq -1$, within the GPD family. We also computed exact 90% and 95% confidence intervals for α for exceedances over each threshold. We report the value of the number of exceedances used, test statistic T_n, p-value of the test, decision (P for Pareto, U for uniform), and the CIs in Table 2. Our conclusion is that the uniform model for the Bilbao waves' exceedances can not be rejected.

Table 2: The table contains results of analysis of the Bilbao waves data. Column labeled T contains the threshold, column labeled D contains the decision.

T (s)	n	T_n	p-value	D	90% CI for α	95% CI for α
7	179	0.0074	0.74	U	(−1.3813, −0.8014)	(−1.4534, −0.759)
7.5	154	0.0094	0.006	P	(−0.8304, −0.4208)	(−0.8788, −0.3890)
8	106	0.0135	0.054	U	(−0.9373, −0.4188)	(−1.0060, −0.3771)
8.5	69	0.0203	0.232	U	(−1.1346, −0.4274)	(−1.2420, −0.3755)
9	41	0.0333	0.632	U	(−1.4806, −0.4232)	(−1.6507, −0.3596)
9.5	17	0.0714	0.298	U	(−4.4557, −0.7056)	(−5.4287, −0.5671)

In summary, our results confirm the conclusions in Castillo and Serra (2015) [15] that the distribution of Nidd exceedances is likely exponential (for large thresholds) and the distribution of the Bilbao waves exceedances is likely uniform.

7. PROOFS

7.1. Proof of Proposition 2.1

Since for $k = 0$ the result is trivial, we shall assume that $k \geq 1$. We assume further that the two random samples are from a GPD with the same index α. By proceeding as in the proof of Proposition 5 in Arendarczyk *et al.* (2021) [2], we can express the statistic $T_{n,k}$ as

$$T_{n,k} \stackrel{d}{=} \frac{\sum_{j=1}^{n-k} \left[e^{\alpha Z_j} - 1\right]^2}{\left(\sum_{j=1}^{n-k} \left[e^{\alpha Z_j} - 1\right]\right)^2}, \tag{7.1}$$

where

$$Z_j = \frac{E_{n-k}}{n-k} + \frac{E_{n-k-1}}{n-k-1} + \cdots + \frac{E_{n-k-j+1}}{n-k-j+1}, \quad j = 1, ..., n-k, \tag{7.2}$$

and the $\{E_i\}$ are IID standard exponential variables. A similar calculation shows that the distribution of T_{n-k} coincides with that of the right-hand-side in (7.1) with the same $\{Z_j\}$. This proves the result.

7.2. Proof of Theorem 3.1

For $i = 1, 2$, let $Q_i(\cdot)$ be the quantile function of $X^{(\theta_i)}$. By standard probability transfer theorem, we have

$$R_n^{(\theta_i)} \stackrel{d}{=} \frac{Q_i(U_{(n)})}{\sum_{k=1}^{n} Q_i(U_{(k)})}, \quad n \geq 2, \quad i = 1, 2,$$

where the $\{U_{(k)}\}$ are the (ascending) standard uniform order statistics based on a sample of size n. To establish (3.1), we need to prove that $\mathbb{P}(R_n^{(\theta_1)} > x) \leq \mathbb{P}(R_n^{(\theta_2)} > x)$ for all $x > 0$. We establish this by showing that for each choice of $0 < u_1 \leq u_2 \leq \cdots \leq u_n < 1$ we have

$$r_n^{(1)}(u_1, ..., u_n) \leq r_n^{(2)}(u_1, ..., u_n), \tag{7.3}$$

where

$$r_n^{(i)}(u_1, ..., u_n) = \frac{Q_i(u_n)}{\sum_{k=1}^{n} Q_i(u_k)}, \quad n \geq 2, \quad i = 1, 2. \tag{7.4}$$

We proceed by induction to establish (7.3). First, assume that $n = 2$. Straightforward calculations show that in this case the inequality (7.3) is equivalent to

$$\frac{Q_1(u_2)}{Q_1(u_1)} \leq \frac{Q_2(u_2)}{Q_2(u_1)}, \quad 0 < u_1 \leq u_2 < 1.$$

However, the above is true by the assumption that $X^{(\theta_1)} \leq_* X^{(\theta_2)}$. Next, we assume that the inequality (7.3) holds for $n \geq 2$ and show its validity for $n + 1$ where

$$0 < u_1 \leq u_2 \leq \cdots \leq u_{n+1} < 1. \tag{7.5}$$

To see this, write

$$r_{n+1}^{(i)}(u_1,...,u_{n+1}) = H\Big(r_n^{(i)}(u_2,...,u_{n+1}), w_i(u_1,...,u_{n+1})\Big), \tag{7.6}$$

where $H(x,y) = x/(1+y)$, $x,y \in \mathbb{R}_+$, and

$$w_i(u_1,...,u_{n+1}) = \frac{Q_i(u_1)}{\sum_{k=2}^{n+1} Q_i(u_k)}, \quad i = 1,2. \tag{7.7}$$

Since by the induction step we have $r_n^{(1)}(u_2,...,u_{n+1}) \le r_n^{(2)}(u_2,...,u_{n+1})$ and the function $H(x,y)$ is increasing in x and decreasing in y, the inequality

$$r_{n+1}^{(1)}(u_1,...,u_{n+1}) \le r_{n+1}^{(2)}(u_1,...,u_{n+1}) \tag{7.8}$$

would follow by (7.6) if we could show that

$$w_2(u_1,...,u_{n+1}) \le w_1(u_1,...,u_{n+1}). \tag{7.9}$$

However, it is easy to see that the inequality in (7.9) is equivalent to

$$\sum_{k=2}^{n+1} \frac{Q_1(u_k)}{Q_1(u_1)} \le \sum_{k=2}^{n+1} \frac{Q_2(u_k)}{Q_2(u_1)},$$

which holds in view of (7.5) since we have

$$\frac{Q_1(u_k)}{Q_1(u_1)} \le \frac{Q_2(u_k)}{Q_2(u_1)}, \quad k = 2,...,n+1,$$

due to $X^{(\theta_1)} \le_* X^{(\theta_2)}$. This completes the induction argument, and the proof.

7.3. Proof of Proposition 3.1

As in the proof of Proposition 5 in Arendarczyk *et al.* (2021) [2], we express $R_{n,k}$ and $R_{n-k+1,1}$ in terms of exponential spacings using the stochastic representation (1.1). We first assume that $k > 0$ and start with $R_{n,k}$. By (1.1), we have

$$R_{n,k} \stackrel{d}{=} \frac{\frac{1}{\alpha}\big(e^{\alpha E_{n:n}} - 1\big) - \frac{1}{\alpha}\big(e^{\alpha E_{k:n}} - 1\big)}{\sum_{i=1}^{n-k}\big[\frac{1}{\alpha}\big(e^{\alpha E_{k+i:n}} - 1\big) - \frac{1}{\alpha}\big(e^{\alpha E_{k:n}} - 1\big)\big]}, \tag{7.10}$$

where the $E_{1:n} \le \cdots \le E_{n:n}$ are the order statistics based on a random sample of size n from standard exponential distribution. Further simplifications produce

$$R_{n,k} \stackrel{d}{=} \frac{e^{\alpha(E_{n:n} - E_{k:n})} - 1}{\sum_{j=1}^{n-k}\big[e^{\alpha(E_{k+j:n} - E_{k:n})} - 1\big]}. \tag{7.11}$$

We now write $E_{k+j:n} - E_{k:n} = \sum_{i=1}^{j} D_{k+i:n}$, where $D_{i:n} = E_{i:n} - E_{i-1:n}$, $i = 2,...,n$ (with $D_{1:n} = E_{1:n}$) are the associated exponential spacings. Since these are independent and exponentially distributed with parameter $n - i + 1$ (see, e.g., Rényi, 1953 [45]), we can express $R_{n,k}$ as

$$R_{n,k} \stackrel{d}{=} \frac{e^{\alpha Z_{n-k}} - 1}{\sum_{j=1}^{n-k}\big[e^{\alpha Z_j} - 1\big]}, \tag{7.12}$$

where

$$Z_j = \frac{E_{n-k}}{n-k} + \frac{E_{n-k-1}}{n-k-1} + \cdots + \frac{E_{n-k-j+1}}{n-k-j+1}, \quad j = 1, ..., n-k, \tag{7.13}$$

and the $\{E_i\}$, $i = 1, ..., n-k$, are independent standard exponential random variables. A similar approach shows that the $R_{n-k+1,1}$ in (3.3) has the same distribution as the right-hand-side in (7.12) with the $\{Z_j\}$ given by (7.13). This proves the result for $k > 0$. The case $k = 0$ can be established along the same lines.

7.4. Proof of Proposition 3.2

Since R_n does not depend on the scale parameter, we shall assume that $\beta = 1$. We start wit the limit at $-\infty$. Since $1/n \le R_n \le X_{(n)}/(nX_{(1)})$, it is enough to show that $X_{(n)}/X_{(1)} \xrightarrow{d} 1$ as $\alpha \to -\infty$. By (1.1), we have

$$\frac{X_{(n)}}{X_{(1)}} \stackrel{d}{=} \frac{e^{\alpha E_{(n)}} - 1}{e^{\alpha E_{(1)}} - 1}, \tag{7.14}$$

where the $E_{(1)} \le \cdots \le E_{(n)}$ are the order statistics based on a random sample of size n from standard exponential distribution. Since the two exponential terms on the right-hand-side in (7.14) both converge in distribution to zero as $\alpha \to -\infty$, the right-hand-side in (7.14) converges to 1 by continuous mapping and Slutsky's theorems.

Next, we consider the limit at ∞. Straightforward algebra shows that

$$\left(1 + (n-1)\frac{X_{(n-1)}}{X_{(n)}}\right)^{-1} \le R_n \le 1.$$

Thus, it is enough to show that $X_{(n-1)}/X_{(n)} \xrightarrow{d} 0$ as $\alpha \to \infty$. Again, by (1.1), we have

$$\frac{X_{(n-1)}}{X_{(n)}} \stackrel{d}{=} \frac{1 - e^{-\alpha E_{(n-1)}}}{e^{\alpha(E_{(n)} - E_{(n-1)})} - e^{-\alpha E_{(n-1)}}}, \tag{7.15}$$

where the $\{E_{(i)}\}$ are as before. It is easy to see that, as $\alpha \to \infty$, the exponential term $e^{-\alpha E_{(n-1)}}$ in the expression above converges to zero while the term $e^{\alpha(E_{(n)} - E_{(n-1)})}$ converges to ∞. Consequently, the expression in (7.15) converges to 0 as desired.

7.5. Proof of Proposition 4.1

To show the test δ has size γ, note that

$$size(\delta) = \sup_{\alpha \le \alpha_0} \mathbb{P}(T_n(\alpha) > c_n) \le \mathbb{P}(T_n(\alpha_0) > c_n) = \gamma,$$

since T_n is stochastically increasing in α. Thus, the test δ has size γ. Next, we show that δ is also unbiased. Let $\pi(\alpha)$ be the power function of δ. We shall show that the power function is at least equal to the size for all $\alpha \in \Omega_1$. Indeed, for any $\alpha \in \Omega_1$ we have

$$\pi(\alpha) = \mathbb{P}(T_n(\alpha) > c_n) \ge \mathbb{P}(T_n(\alpha_0) > c_n) = \gamma,$$

since T_n is stochastically increasing in α. This shows that δ is unbiased, as desired.

7.6. Proof of Proposition 4.3

The probability in (4.13) concerns all the values t_n of the statistic T_n for which we have $\alpha_0 > \underline{\alpha}(t_n)$. By the definition of the function $\underline{\alpha}$ and the stochasticity of T_n, this is equivalent to the condition $S(t_n|\alpha_0) > \gamma$, where $S(\cdot|\alpha_0)$ is the SF of T_n when the true value of the parameter is α_0. Equivalently, the event $\alpha_0 > \underline{\alpha}(T_n)$ in (4.13) can be stated as $F(T_n|\alpha_0) \leq 1-\gamma$, where $F(\cdot|\alpha_0)$ is the CDF of T_n when the true value of the parameter is α_0. However, the quantity $F(T_n|\alpha_0)$ has standard uniform distribution, so that

$$\mathbb{P}_{\alpha_0}(\alpha_0 > \underline{\alpha}(T_n)) = \mathbb{P}_{\alpha_0}(F(T_n|\alpha_0) \leq 1-\gamma) = 1-\gamma,$$

as desired.

7.7. Proof of Proposition 4.5

The probability in (4.19) concerns all the values t_n of the statistic T_n for which we have $\underline{\alpha}(t_n) < \alpha_0 < \overline{\alpha}(t_n)$. By the definition of the quantities $\underline{\alpha}$ and $\overline{\alpha}$ and the stochasticity of T_n, this is equivalent to the condition $\gamma - r\gamma < F(t_n|\alpha_0) < 1 - r\gamma$, where $F(\cdot|\alpha_0)$ is the CDF of T_n when the true value of the parameter is α_0. Equivalently, the event $\underline{\alpha}(T_n)) < \alpha_0 < \overline{\alpha}(T_n)$ in (4.19) can be stated as $\gamma - r\gamma < F(T_n|\alpha_0) < 1 - r\gamma$. Since the quantity $F(T_n|\alpha_0)$ is standard uniform, we have

$$\mathbb{P}_{\alpha_0}(\underline{\alpha}(T_n) < \alpha_0 < \overline{\alpha}(T_n)) = \mathbb{P}_{\alpha_0}(\gamma - r\gamma < F(T_n|\alpha_0) < 1 - r\gamma) = 1-\gamma,$$

as desired.

ACKNOWLEDGMENTS

We thank the editors and the reviewers for their comments and suggestions that helped us improve the manuscript. We also extend hearty thanks to Professor Joan del Castillo and Professor Ioannis Papasthopoulos for their kind help in obtaining the data sets. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 318984 – RARE.

REFERENCES

[1] Albrecher, H.; Ladoucette, S.A. and Teugels, J.L. (2010). Asymptotics of the sample coefficient of variation and the sample dispersion, *J. Statist. Plann. Inference*, **140**(2), 358–368.

[2] Arendarczyk, M.; Kozubowski, T.J. and Panorska, A.K. (2021). The Greenwood statistic, stochastic dominance, clustering and heavy tails, *Scand. J. Stat.*, **49**(1), 331–352.

[3] Balkema, A.A. and de Haan, L. (1974). Residual life time at high age, *Ann. Probab.*, **2**, 792–804.

[4] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J.L. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, Hoboken.

[5] Belzunce, F.; Martinez-Riquelme, C. and Mulero, J. (2016). *An Introduction to Stochastic Orders*, Academic Press, Amsterdam.

[6] Bryson, M.C. (1974). Heavy-tailed distributions: Properties and tests, *Technometrics*, **16**(1), 61–68.

[7] Caeiro, F. and Gomes, M.I. (2016). *Threshold selection in extreme value analysis*. In "Extreme Value Modeling And Risk Analysis: Methods And Applications" (D.K. Dey, and J. Yan, Eds.), pp. 69–82, Chapman and Hall/CRC Press, Boca Raton.

[8] Casella, G. and Berger, R.L. (2002). *Statistical Inference*, 2nd ed., Duxbury Press/Thomson Learning, Pacific Grove.

[9] Castillo, E. (1988). *Extreme Value Theory in Engineering*, Academic Press, San Diego.

[10] Castillo, E. and Hadi, A.S. (1997). Fitting the generalized Pareto distribution to data, *J. Amer. Statist. Assoc.*, **92**(440), 1609–1620.

[11] Castillo, E.; Hadi, A.S.; Balakrishnan, N. and Sarabia, J.M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley, Hoboken.

[12] Castillo, J. and Daoudi, J. (2009). Estimation of the generalized Pareto distribution, *Statist. Probab. Lett.*, **79**, 684–688.

[13] Castillo, J.; Daoudi, J. and Lockhart, R. (2014). Methods to distinguish polynomial and exponential tails, *Scan. J. Statist.*, **41**, 382–393.

[14] Castillo, J. and Padilla, M. (2015). Modeling extreme values by the residual coefficient of variation, arXiv: 1510.00179v1.

[15] Castillo, J. and Serra, I. (2015). Likelihood inference for generalized Pareto distribution, *Comput. Statist. Data Anal.*, **83**, 116–128.

[16] Chaouche, A. and Bacro, J.-N. (2004). A statistical test procedure for the shape parameter of a generalized Pareto distribution, *Comput. Statist. Data Anal.*, **45**, 787–803.

[17] Chaouche, A. and Bacro, J.-N. (2006). Statistical inference for the generalized Pareto distribution: Maximum likelihood revisited, *Comm. Statist. Theory Methods*, **35**(5), 785–802.

[18] Chavez, M.; Ghil, M. and Fucugauchi, U. (2016). *Extreme Events: Observations, Modeling, and Economics*, AGU/Wiley, Hoboken.

[19] Coles, S. (2004). *An introduction to Statistical Modeling of Extreme Values*, Springer, London.

[20] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *J. Roy. Statist. Soc. Ser. B*, **52**(3), 393–442.

[21] de Haan, J. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.

[22] Dey, D.K. and Yan, J. (2016). *Extreme Value Modeling and Risk Analysis: Methods and Applications*, Chapman and Hall/CRC Press, Boca Raton.

[23] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.

[24] Finkenstädt, B. and Rootzén, H. (Edit.) (2003). *Extreme Values in Finance, Telecommunications, and the Environment*, Chapman and Hall/CRC Press, Boca Raton.

[25] Fraga Alves, M.I. and Gomes, M.I. (1996). Statistical choice of extreme value domains of attraction – a comparative analysis, *Comm. Statist. Theory Methods*, **25**(4), 789–811.

[26] Gomes, M.I. and Alpuin, M.T. (1986). Inference in a multivariate generalized extreme value model: Asymptotic properties of two test statistics, *Scand. J. Stat.*, **13**(4), 291–300.

[27] Gomes, M.I. and van Montfort, M.A.J. (1986). *Exponentiality versus generalized Pareto, quick tests*. In "Proceedings of the Third International Conference on Statistical Climatology", pp. 185–195.

[28] Greenwood, M. (1946). The statistical study of infectious diseases, *J. Roy. Statist. Soc.*, **109**(2), 85–110.

[29] Gumbel, E.J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

[30] Hasofer, A.M. and Wang, Z. (1992). A test for extreme value domain of attraction, *J. Amer. Statist. Assoc.*, **87**, 171–177.

[31] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**(3), 339–349.

[32] Kiran, K.G. and Srinivas, V.V. (2021). A Mahalanobis distance-based automatic threshold selection method for Peaks Over Threshold model, *Water Resour. Research*, **57**, e2020WR027534.

[33] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London.

[34] Kozubowski, T.J.; Panorska, A.K.; Qeadan, F.; Gershunov, A. and Rominger, D. (2009). Testing exponentiality versus Pareto distribution via likelihood ratio, *Comm. Statist. Sim. Comput.*, **38**(1), 118–139.

[35] Leadbetter, M.R.; Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, Berlin.

[36] Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators, *Comput. Statist. Data Anal.*, **51**(2), 904–917.

[37] Marohn, F. (2000). Testing extreme value models, *Extremes*, **3**(4), 363–384.

[38] Moran, P.A.P. (1947). The random division of an interval, *J. Roy. Statist. Soc. Suppl.*, **9**(1), 92–98.

[39] Neves, C. and Fraga Alves, M.I. (2007). Semi-parametric approach to the Hasofer-Wang and Greenwood statistics in extremes, *Test*, **16**, 297–313.

[40] Neves, C.; Picek, J. and Fraga Alves, M.I. (2006). The contribution of the maximum to the sum of excesses for testing max-domains of attraction, *J. Statist. Plann. Inference*, **136**, 1281–1301.

[41] Papastahopoulos, I. and Tawn, J.A. (2012). Extended generalized Pareto models for tail estimation, *J. Statist. Plann. Inference*, **143**, 131–143.

[42] Pickands III, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

[43] Qeadan, F.; Kozubowski, T.J. and Panorska, A.K. (2012). The joint distribution of the sum and the maximum of n i.i.d. exponential random variables, *Comm. Statist. Theory Methods*, **41**(3), 544–569.

[44] Reiss, R.-D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology, and Other Fields*, 2nd Ed., Birkhäuser, Basel.

[45] Rényi, A. (1953). On the theory of order statistics, *Acta Math. Acad. Sci. Hungar.*, **4**, 191–232.

[46] Schluter, C. and Trede, M. (2018). Size distributions reconsidered, *Econometric Rev.*, **38**(4), 1–16.

[47] Schneider, L.F.; Krajina, A. and Krivobokova, T. (2021). Threshold selection in univariate extreme value analysis, *Extremes*, **24**, 881–913.

[48] Silva Lomba, J. and Fraga Alves, M.I. (2020). L-moments for automatic threshold selection in extreme value analysis, *Stoch. Environ. Res. Risk Assess.*, **34**, 465–491.

[49] Tajvidi, N. (2003). Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions, *Extremes*, **6**, 111–123.

[50] Zhang, J. and Stephens, M.A. (2009). A new and efficient estimation method for the generalized Pareto distribution, *Technometrics*, **51**, 316–325.

REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 389–404
https://doi.org/10.57805/revstat.v21i3.411

Assessing Homoscedasticity Graphically: Levene–Brown–Forsythe Approaches

Author: Elsayed A. H. Elamir ✉
– Management and Marketing Department, University of Bahrain, Kingdom of Bahrain
shabib@uob.edu.bh

Received: February 2021 Revised: January 2022 Accepted: January 2022

Abstract:

- The problem of homoscedasticity arises in several fields such as business, education, environments, and medicine, and common question in many statistical analyses. One of the most important tests in this direction is Levene test and its robust version Brown–Forsythe test. The goal of this paper is threefold. The first goal is to propose an expression that enable to develop a graphical way for Levene–Brown–Forsythe tests. The second goal is to derive the sampling distribution of the proposed expression as the generalized beta prime distribution. The third goal is to provide deep insight and understanding where the dispersion effects occur. Simulation study is carried out to study the level of significance and power of the proposed test in comparison with the original Levene–Brown–Forsythe tests. The results are of great values since the proposed method:
  (**a**) provides powerful visual tool and deep insight for testing homoscedasticity,
  (**b**) keeps the size and power of the test similar to Levene–Brown–Forsythe tests,
  (**c**) does not need to pairwise comparisons.

  Two applications are presented to show the utilities of the proposed method.

Keywords:

- *beta distribution; Bonferroni approximation; homogeneity of variance; nonnormality; test power; type I error.*

AMS Subject Classification:

- 62F03, 62J10.

✉ Corresponding author.

1. INTRODUCTION

It is known that the one-way analysis of variance (ANOVA) is one of the most frequently used tests to explore the differences among several treatment means; see, for example, Kutner *et al.* [15], Yigit and Mendes [28] and Nguyen *et al.* [19]. The homoscedasticity plays an important role in ANOVA test since the large deviations from the homoscedasticity can affects the results of F-test for equal means; see Fox and Weisberg [9] and Wang *et al.* [27]. The Levene test [17] and its robust extension Brown–Forsythe test [5] had been used to assess homogeneity of variances or homoscedasticity for several groups. These tests depend on transforming the ANOVA test of means into a test of variances based the absolute values of the differences between observations and a location measure (mean, trimmed mean and median). The assumption of homoscedasticity can be written as

$$H_0 : \sigma_1^2 = \cdots = \sigma_k^2$$

versus

$$H_a : \ \sigma_i^2 \neq \sigma_j^2 \qquad \text{for at least one pair } (i, j),$$

where k is the number of groups.

The assumption of homoscedasticity can also use on its own to compare the dispersion among several groups in a study. Kvamme *et al.* [16] used Levene test and Brown–Forsythe robust version of Levene test to compare the dispersion of the holes of the chalupa pots from the 3 different locations. The null hypothesis was that the dispersion or variation of each characteristic is the same in the three locations. Plourde and Watkins [22] utilized Levene's test to month-to-month price fluctuates to investigate whether the conduct of oil costs changed within the 1980s and got to be more like that of other goods, which head to have big cost vacillations, they utilized both the nonparametric Fligner–Killeen [8] test and the Brown–Forsythe modified of Levene test in an arrangement of post hoc pairs comparisons to evaluate the relative variations of the price fluctuates. Sant and Cowan [24] considered the effect of a privation of a profit by a company on the changeability of both the estimates of future profit and the real profit. They compared the profit and predicted of companies that excluded a profit amid the period 1963–1984 by comparing the fluctuations of the real or forecasted profit per share 2 years after the omission and 2 years before. They utilized Brown–Forsythe robust version of Levene test. Berger *et al.* [4] used a database of 6026 "echocardiograms" that perused by one of 3 similarly capable perusers to survey the contrasts in recurrence of many analyzes and related measurements. The numbers of "echocardiograms" examined by the pursuers (one, two, three) were 2702, 2101 and 1223, respectively. Levene's test was utilized to evaluate the variation in the measurements of many continuous characteristics. Nordstokke and Zumbo [20] had developed a nonparametric version of Levene test by pooling the observations from all sets, ranking the scores with taking ties in consideration, return the ranks into their original sets, and apply the Levene test on the ranks; for more details; see Nordstokke *et al.* [21] and Shear *et al.* [25]. In analytical methods Aslam and Khan [2] used Levene test to modify Chochran test to be applied for detecting outliers in the data. The goal of this paper is threefold. The first goal is to develop an expression that assist in plotting Levene–Brown–Forsythe tests. The second goal is to obtain the sampling distribution of the suggested expression as a beta prime distribution of the second type that can be used in creating a decision limit. The third goal is to provide deep insight and understanding where

the dispersion effects occur. Simulation study is carried out to study the level of significance and power of the proposed test in comparison with the original Levene–Brown–Forsythe tests. The results are of great value since the proposed method provides visual and deep insight where the variation occurs and does not need to post hoc pairwise comparisons. Two applications are studied to show the usage of the proposed method.

Levene–Brown–Forsythe approach is explained in Section 2. The proposed method is introduced in Section 3. The empirical type I error and test power is presented in Section 4. The usage of the proposed method in the analysis of data from two applications is described in Section 5. Section 6 is devoted for conclusion.

2. LEVENE–BROWN–FORSYTHE APPROACH

Suppose there are k groups each follows a normal distribution with means μ_i, standard deviation σ_i, n_i the number of observations in each group, and X_{ij} the response value and n the total number of observations in all groups, $i = 1, ..., k$, $j = 1, ..., n_i$. Levene [17] proposed test to assess the equality of variances for two groups or more. The test was depending on the idea of analysis of variance (ANOVA) for the absolute deviation about mean, $|X_{ij} - X_{i\cdot}|$. Levene's test is based on the classical ANOVA method that can be written as

$$W = \frac{\sum_{i=1}^{k} n_i (Z_{i\cdot} - Z_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} (Z_{ij} - Z_{i\cdot})^2/(n-k)}, \tag{2.1}$$

where k is the number of groups, n_i the number of observations in group i, $i = 1, ..., k$, $n = n_1 + \cdots + n_k$ is the total number, $Z_{ij} = \left|X_{ij} - \bar{X}_{i\cdot}\right|$ is the absolute deviation about group mean, X_{ij} is the observation for j-th case from group i, $Z_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij}$ is the mean of Z_{ij} for group i, $Z_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij}$ is the mean of all Z_{ij}.

Although Levene noticed that $\left|X_{ij} - \bar{X}_{i\cdot}\right|$ are not independent within each group, he proved that the correlation is of order $1/n_i^2$ and considered that this is small dependency within each group and would not be seriously impact the distribution of W; see Gastwirth *et al.* [11]. Therefore, the test statistic W is approximated by F-distribution with $k-1$ and $n-k$ degrees of freedom, i.e., $F(\alpha; k-1, n-k)$ where F is the quantile for F-distribution and α is prechosen level of significant. In practice it may be concluded that there is heterogeneity if $W > F(\alpha; k-1, n-k)$. Brown and Forsythe [11] proposed revised version to Levene test by using median or trimmed mean rather than mean, i.e., $Z_{ij} = \left|X_{ij} - \widetilde{X}_{i\cdot}\right|$ or $Z_{ij} = \left|X_{ij} - \check{X}_{i\cdot}\right|$, $\widetilde{X}_{i\cdot}$ median and $\check{X}_{i\cdot}$ trimmed mean, with the same approximated distribution $F(\alpha; k-1, n-k)$. Brown and Forsythe carried out simulation study that indicated that median and trimmed mean performed better in heavy-tailed symmetric and skewed distributions while mean is performed best in case of normal and moderate-tailed symmetric distribution; see Brown and Forsythe [5] and Gastwirth *et al.* [11]. Although different underlying distributions give different optimal choice for location parameter, the optimal choice based on median is a recommended one as it provides a good robustness for many types of non-normal data while hold a good power in normal and symmetric distributions; see Gastwirth *et al.* [12], Wang *et al.* [27] and Nguyen *et al.* [19].

3. THE PROPOSED METHOD

The Levene–Brown–Forsythe test can be rewritten as

$$W = \frac{\sum_{i=1}^{k} n_i (Z_{i.} - Z_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i.})^2/(n-k)} = \sum_{i=1}^{k} \frac{n_i (Z_{i.} - Z_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i.})^2/(n-k)} = \sum_{i=1}^{k} U_i. \tag{3.1}$$

Hence,

$$U_i = \frac{n_i (Z_{i.} - Z_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i.})^2/(n-k)}, \qquad i = 1, 2, ..., k. \tag{3.2}$$

This is the ratio for each between square and all treatments squares or contribution of each between squares to mean square error. Therefore, the Levene–Brown–Forsythe tests could be plotted as

$$x_{axis} = i \quad \text{versus} \quad y_{axis} = U_i \text{ with } DL, \quad \text{for } i = 1, 2, ..., k,$$

where DL is the decision limit obtained from the sampling distribution of U_i.

3.1. The sampling distribution of U_i

Under the assumptions of one-way ANOVA:

(**a**) $X_{i1}, ..., X_{kn_i}$ is a random sample of size n_i from a normal population, $i = 1, ..., k$;

(**b**) the random samples from different populations are independent;

see Johnson and Wichern [14]. Furthermore, Gastwirth *et al.* ([11], page 4) had written that "$Z_{ij} = \left|X_{ij} - \bar{X}_{i.}\right|$ are treated as independent, identically distributed, normal variables, and the usual ANOVA statistic is utilized". Since $Z_{ij} = \left|X_{ij} - \bar{X}_{i.}\right|$ is not normally distributed, the Levene's method takes usefulness of the reality that the ANOVA procedures for comparing means are robust to infraction of the assumption that the data follows a normal distribution; see Gastwirth *et al.* ([11], page 4) and Miller ([18], page 80). Therefore, if the null hypothesis of homogeneity of variance is true, hence, the sampling distribution of U_i can be derived as

$$n_i (Z_{i.} - Z_{..})^2/(k-1) \sim \sigma^2 \left(\frac{n - n_i}{n(k-1)}\right) \chi^2(1) \tag{3.3}$$

and

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i.})^2/(n-k) \sim \sigma^2 \chi^2(n-k)/(n-k). \tag{3.4}$$

Hence,

$$U_i \sim \frac{((n - n_i)/n(k-1))\chi^2(1)}{\chi^2(n-k)/(n-k)} = \frac{gamma\left(\frac{1}{2}, \frac{n(k-1)}{2(n-n_i)}\right)}{gamma\left(\frac{n-k}{2}, \frac{n-k}{2}\right)}. \tag{3.5}$$

The sampling distribution of U_i can be obtained as

$$f_{U_i}(u) = \frac{[((n-n_i)(n-k))/n(k-1)]^{-1/2}}{B\left(\frac{1}{2}, \frac{n-k}{2}\right)} \left(1 + \frac{n(k-1)}{(n-n_i)(n-k)}u\right)^{-(n-k+1)/2} u^{-1/2}, \tag{3.6}$$

where $k > 0$, $i = 1, ..., k$, and B: Beta; see Coelho and Mexia [6] and Elamir [7]. This distribution has parameters k, n_i and n and is a special type from generalized beta prime distribution with $a = 1$, $b = \frac{((n-n_i)(n-k))}{(n(k-1))}$, $p = 1/2$, $q = (n-k)/2$, $x = u$; see Coelho and Mexia [6], R Core Team [23] and GB2 package, Graf and Nedyalkova [13]. As one of the reviewers has pointed out that the distribution of U_i may also be written in terms of a scaled F-distribution. Note that U_i can be rewritten in terms of scaled F-distribution as

$$U_i \sim \frac{((n-n_i)/n(k-1))\chi^2(1)}{\chi^2(n-k)/(n-k)} = \frac{n-n_i}{n(k-1)} F(v_1 = 1, v_2 = (n-k)).$$

From Smyth ([26], page 9), the density function for scaled F-distribution ($x = (a/b)F(v_1, v_2)$) can be written as

$$f(x) = \frac{a^{v_2/2} b^{v_1/2} x^{\frac{v_1}{2}-1}}{\beta\left(\frac{v_1}{2}, \frac{v_2}{2}\right)(a+bx)^{\frac{v_1+v_2}{2}}}, \qquad x > 0.$$

The sampling distribution of U_i can be obtained from scaled F-distribution by replacing $v_1 = 1$, $v_2 = n-k$, $a = 1$, $b = (n-n_i)/(n(k-1))$.

The moments of U_i can be obtained as

$$E(U_i{}^h) = \left[\frac{(n-n_i)(n-k)}{n(k-1)}\right]^h \frac{\Gamma(0.5+h)\Gamma\left(\frac{n-k}{2}-h\right)}{\Gamma(0.5)\Gamma\left(\frac{n-k}{2}\right)}, \quad h = 1, 2, ...$$

For example,

$$E(U_i) = \left[\frac{(n-n_i)(n-k)}{n(k-1)}\right] \frac{\Gamma\left(\frac{n-k}{2}-1\right)}{2\Gamma\left(\frac{n-k}{2}\right)} = \frac{(n-n_i)(n-k)}{n(n-k-2)}$$

and

$$V(U_i) = E\left(U_i{}^2\right) - E^2(U_i) = \left[\frac{(n-n_i)(n-k)}{n(k-1)}\right]^2 \frac{3\Gamma\left(\frac{n-k}{2}-2\right)}{4\Gamma\left(\frac{n-k}{2}\right)} - \left[\frac{(n-n_i)(n-k)}{n(n-k-2)}\right]^2.$$

When sample sizes are equal in each group $n_1 = \cdots = n_k = n_e$, the sampling distribution of U_i can be simplified to

$$f_{U_i}(u) = \frac{[1/(n_e-1)]^{-1/2}}{B\left(\frac{1}{2}, \frac{k(n_e-1)}{2}\right)} \left(1 + \frac{1}{(n_e-1)}v\right)^{-(k(n_e-1)+1)/2} u^{-1/2}.$$

This distribution has parameters k and n_e. The moments for U_i can be derived as

$$E(U_i{}^h) = (n_e-1)^h \frac{\Gamma(0.5+h)\Gamma\left(\frac{k(n_e-1)}{2}-h\right)}{\Gamma(0.5)\Gamma\left(\frac{k(n_e-1)}{2}\right)};$$

see Coelho and Mexia [6].

3.2. The empirical moments of U_i

To inspect how well the beta prime distribution for U_i in different setting, a simulation study is conducted to obtain the first four empirical moments of U_i at $k = 3$ and 8, $n_i = 10$ and 25 from normal distribution, Laplace distribution (symmetric heavy-tail) and chi square distribution with 2 degrees of freedom (asymmetric heavy tail) using mean, trimmed mean (0.25) and median as a measure of location. The steps for empirical study are:

1. Select the required design for example $k = 3, n_i = 10$, normal distribution and mean as location measure;
2. Simulate data from a selected distribution with equal variance;
3. Calculate U_i, $i = 1, ..., k$, for each group;
4. Calculate the first four moments for each U_i, $i = 1, ..., k$;
5. Repeat this R times and calculate the mean for every design.

Table 1 gives the first four empirical moments for mean of U_i from normal, Laplace and chi square ($df = 2$) in addition to the theoretical value from the beta prime distribution.

Table 1: Mean of the first four empirical and theoretical (theo.) moments of mean of U_i using different setting and location measures (mean, Tri: trimmed mean (0.25) and Med: median).

k	n_i		Mean				Tri				Med			
			mean	Var.	Sk.	Ku.	mean	Var.	Sk.	Ku.	mean	Var.	Sk.	Ku.
3	10	N	0.386	0.314	3.18	19.45	0.368	0.287	3.19	20.49	0.315	0.211	3.08	21.75
		Laplace	0.443	0.396	3.24	20.02	0.377	0.302	3.44	23.51	0.34	0.237	3.22	22.35
		$\chi^2\,(df=2)$	0.708	1.043	3.24	20.53	0.539	0.675	3.73	23.61	0.364	0.276	3.81	24.62
		Theo.	0.36	0.293	3.42	23.02	0.36	0.293	3.42	23.02	0.36	0.293	3.42	23.02
	25	N	0.359	0.265	3.24	20.96	0.351	0.256	3.23	18.28	0.313	0.204	3.3	21.23
		Laplace	0.377	0.278	2.82	15.02	0.346	0.24	2.69	14.01	0.334	0.219	2.79	15.23
		$\chi^2\,(df=2)$	0.617	0.762	3.13	16.05	0.471	0.472	3.47	17.63	0.338	0.233	2.99	18.28
		Theo.	0.342	0.245	3.02	17.29	0.342	0.245	3.02	17.29	0.342	0.245	3.02	17.29
8	10	N	0.141	0.041	3.24	21.16	0.133	0.036	3.12	18.43	0.111	0.025	3.24	20.6
		Laplace	0.161	0.056	3.65	30.77	0.135	0.039	3.92	28.59	0.122	0.034	3.65	26.38
		$\chi^2\,(df=2)$	0.249	0.164	4.09	29.14	0.182	0.086	4.69	33.9	0.125	0.04	4.33	30.29
		Theo.	0.128	0.034	3.02	17.29	0.128	0.034	3.02	17.29	0.128	0.034	3.02	17.29
	25	N	0.128	0.033	2.94	15.43	0.13	0.034	2.74	16.08	0.117	0.027	2.81	15.98
		Laplace	0.137	0.039	3.11	18.29	0.126	0.033	3.11	18.49	0.121	0.03	3.1	19.12
		$\chi^2\,(df=2)$	0.221	0.114	3.56	25.74	0.164	0.065	3.72	26.1	0.123	0.033	3.55	22.76
		Theo.	0.126	0.032	2.89	15.78	0.126	0.032	2.89	15.78	0.126	0.032	2.89	15.78

This table illustrates that:

1. When the mean is the location measure, the best results (empirical is very close to theoretical) are obtained from normal distribution;
2. When the trimmed mean is the location measure, the best results (empirical is very close to theoretical) are obtained from Laplace distribution, followed by normal;
3. When the median is the location measure, the best results (empirical is very close to theoretical) are obtained from chi square distribution, followed by Laplace distribution then normal.

3.3. Decision limit

To create decision limit (DL), it must take into account k tests that required making difference between two sorts of level of significant α:

1. test-wise alpha (alpha per test $\alpha[PT]$) when working with a specific test;
2. family-wise (alpha per family or experiment alpha $\alpha[PF]$) when working with the whole experiment.

The probability of committing first error for k tests can be defined from Abdi [1] as

$$\alpha(PF) = 1 - (1 - \alpha(PT))^k. \tag{3.7}$$

Hence,

$$\alpha(PT) = 1 - (1 - \alpha(PF))^{1/k}. \tag{3.8}$$

Simpler form can be obtained using Bonferroni approximation as

$$\alpha(PT) \approx \frac{\alpha(PF)}{k}. \tag{3.9}$$

As an example, to perform $k = 8$, and the α per family $(PF) = 0.05$, based on Bonferroni approximation, the null hypothesis will be rejected its related probability is less than $\alpha(PT) \approx 0.05/8 = 0.00625$. Although the Sidak and Bonferroni corrections are closely similar, the Bonferroni correction is more conservative than Sidak and control of the expected number of type I error (Per-family error rate (PFER)) which Sidak does not. Frane [10] stated that "However, it is important to note that the Bonferroni procedure controls not only the FWER (family-wise error rate) but also the PFER (Per-family error rate (PFER))".

In addition to Bonferroni approximation, there is a good method called Benjamini–Hochberg that controls the false discovery rate (the likelihood of an incorrect rejection of a hypothesis occurs) using sequential modified Bonferroni correction for several testing rather than the family wise error rate. Benjamini and Hochberg [3] defined the false discovery rate (FDR) as the number of false discoveries in an experiment divided by the total number of discoveries in that experiment where the discovery is a test that passes one acceptance threshold. In other words, it represents one believe the result is true, but when they are accepted it is never known how many of discoveries are right or wrong. According to Benjamini and Hochberg [3], if q-value is an estimate of FDR from p-values, it may be written as $q_i = Np_i/i$, N: total p-values, p_i: i-th smallest p-value (likelihood of accepting a false result by chance), Np_i: expected value of false results if one accepts all results which have p-values of p_i or smaller, and i the number of results one accepts at i-th p-value threshold. The steps are:

(**a**) rank the p-values from all multiple hypothesis tests in an experiment;

(**b**) compute q_i;

(**c**) to ensure monotonically decreasing q-values, replace q_i with the lowest value among all lower-rank q-values that computed.

In R-software under the function "`p.adjust(p; method = " "; n = length(p))`" one of the methods is BH (Benjamini–Hochberg); see R Core Team [23]. Therefore, the decision line could be proposed by using the quantile function of beta prime distribution and the Bonferroni approximation as

$$DL = \text{qgb2}\left(1 - \frac{\alpha}{k},\ a = 1,\ b = \frac{(n - n_i)(n - k)}{n(k - 1)},\ p = 0.5,\ q = \frac{(n - k)}{2}\right).$$

Moreover, the Bonferroni approximation could be replaced by BH using R-function as follows: `p.adjust(p = 1−`α/k`; method="BH"; n=length(p))`; see GB2 package Graf and Nedyalkova [13]. Hence,

$$\text{if any } U_i > DL, \quad \text{for } i = 1, ..., k, \qquad H_0 \text{ is rejected.}$$

The U-plot can be plotted as

$$x_{axis} = 1 : k \quad \text{versus} \quad y_{axis} = U_i, \quad \text{with decision limit } DL.$$

H_0 is rejected if any point outside DL and this will identify where the differences occur.

4. SIMULATION STUDY

The proposed method using Bonferroni (Bonf.) approximation and Benjamini–Hochberg (BH) method is compared with Levene–Brown–Forsythe methods in terms of type I error $p(\text{reject } H_0 \mid H_0 \text{ is true})$ and power of the test $p(\text{reject } H_0 \mid H_0 \text{ is false}) = 1 - p(\text{accept } H_0 \mid H_0 \text{ is false}) = 1 - \text{type II error}$.

With respect to type I error, the following steps are used in simulation:

1. Construct the desired design $k = 3, 8$, $n_i = 10, 20, 50$ and nominal $\alpha = 0.05$.
2. Simulate data from a required distribution with equal variances. The normal distribution as original distribution, Laplace distribution as symmetric heavy-tailed distribution and $\chi^2\,(df = 2)$ as asymmetric heavy-tailed distributions are used.
3. Calculate U_i-Bonf., U_i-BH, Levene–Brown–Forsythe for each design.
4. Compute the decision limit for U_i-Bonf., U_i-BH and p-values for Levene–Brown–Forsythe.
5. Create a dummy variable by giving 1 for reject and 0 else.
6. Repeat R times and compute the mean for each design.

The results for these procedures are given in Table 2. It can be concluded about type I error that:

1. Levene test and U_i-Bonferroni using mean as location are giving a good empirical type I error in the case of normal distribution;
2. Brown–Forsythe and U_i-BH using median as location are giving a good empirical type I error in the case of chi square distribution;
3. Brown–Forsythe and U_i-Bonferroni using trimmed mean as location are giving a good empirical type I error in the case of Laplace distribution.

In general, Brown–Forsythe and U_i-BH using median as location tend to have adequate type I error control across all used distribution shapes and this is consistent with results of Wang *et al.* [27] and Nguyen *et al.* [19].

Table 2: Empirical type I error using U_i-Bonferroni (Bonf.), U_i-BH, Levene–Brown–Forsythe (LBF) methods, nominal $\alpha = 0.05$ from normal, χ^2 and Laplace distributions based on 10000 replications.

k	n_i	Bonf.	BH	LBF	Bonf.	BH	LBF	Bonf.	BH	LBF
		Mean, Normal (100,5)			Mean, Chisq ($df = 2$)			Mean, Laplace (0,4)		
3	10	0.056	0.06	0.064	0.176	0.185	0.195	0.065	0.067	0.074
	20	0.05	0.053	0.056	0.166	0.172	0.181	0.056	0.058	0.065
	50	0.048	0.051	0.053	0.16	0.168	0.178	0.047	0.05	0.054
8	10	0.071	0.073	0.074	0.314	0.322	0.37	0.103	0.094	0.101
	20	0.055	0.059	0.059	0.271	0.28	0.34	0.081	0.082	0.08
	50	0.053	0.058	0.058	0.255	0.263	0.31	0.061	0.063	0.06
		Median, Normal (100,5)			Median, Chisq ($df = 2$)			Median, Laplace (0,4)		
3	10	0.03	0.032	0.032	0.042	0.047	0.048	0.03	0.03	0.031
	20	0.034	0.037	0.036	0.041	0.044	0.044	0.037	0.04	0.043
	50	0.039	0.043	0.044	0.041	0.046	0.048	0.04	0.042	0.043
8	10	0.034	0.035	0.032	0.064	0.065	0.045	0.056	0.056	0.036
	20	0.036	0.037	0.034	0.056	0.056	0.044	0.051	0.051	0.042
	50	0.044	0.046	0.044	0.051	0.052	0.046	0.048	0.049	0.046
		Trimmed, Normal (100,5)			Trimmed, Chisq ($df = 2$)			Trimmed, Laplace (0,4)		
3	10	0.04	0.045	0.048	0.075	0.078	0.082	0.041	0.044	0.045
	20	0.044	0.046	0.049	0.059	0.062	0.066	0.038	0.042	0.043
	50	0.041	0.043	0.045	0.055	0.058	0.063	0.042	0.046	0.047
8	10	0.053	0.055	0.054	0.115	0.116	0.104	0.075	0.073	0.048
	20	0.046	0.048	0.047	0.08	0.082	0.072	0.064	0.061	0.046
	50	0.047	0.048	0.048	0.064	0.065	0.068	0.056	0.056	0.045

With respect to power of the test, the following steps are used in simulation:

1. Construct the desired design $k = 3, 8$, $n_i = 10, 20, 50$ and nominal $\alpha = 0.05$.
2. Simulate data from a required distribution with unequal variances. The used distributions are the normal distribution with variances 5, 5 and 10 ($k = 3$) and 5, 5, 5, 5, 10, 10, 25 and 25 ($k = 8$), Laplace distribution with $df = 2, 2, 10$ ($k = 3$) and $df = 2, 2, 2, 2, 1, 1, 5, 5$ ($k = 8$) and χ^2 with $df = 2, 2, 4$ ($k = 3$) and $df = 2, 2, 2, 2, 1, 1, 4, 4$ ($k = 8$).
3. Calculate U_i-Bonf., U_i-BH, Levene–Brown–Forsythe for each design.
4. Compute the decision limit for U_i-Bonf., U_i-BH and p-values for Levene–Brown–Forsythe.
5. Create a dummy variable by giving 1 for reject and 0 else.
6. Repeat R times and compute the mean for each design.

The results of these procedures are given in Table 3. It can be concluded that:

1. As k and n_i increase, the power becomes larger. If data is from normal and $k = 3$, n_i needs to be at least 20 to obtain good power while it will be much less if $k = 8$.
2. U_i are giving nearly power similar to Levene–Brown–Forsythe tests using the mean, trimmed mean and median.
3. U_i-BH gives slightly better results than U_i-Bonf. in terms of power.
4. With increasing the number of groups, U_i will be slightly better than Levene–Brown–Forsythe tests especially with using trimmed mean and median.

Table 3: Empirical power using U_i-Bonferroni (Bonf.), U_i-BH, Levene–Brown–Forsythe (LBF) methods, nominal $\alpha = 0.05$ from normal, χ^2 and Laplace distributions based on 10000 replications.

k	n_i	Bonf.	BH	Levene	Bonf.	BH	Levene	Bonf.	BH	Levene
		Mean, Normal var = 5, 5, 10			Mean, Chisq df = 2, 2, 4			Mean, Laplace scale = 5, 5, 10		
3	10	0.475	0.488	0.493	0.272	0.278	0.289	0.342	0.35	0.36
	20	0.835	0.847	0.832	0.378	0.387	0.399	0.604	0.612	0.622
	50	0.997	0.997	0.997	0.63	0.64	0.648	0.95	0.952	0.954
		var = 5, 5, 5, 5, 10, 10, 25, 25			df = 2, 2, 2, 2, 1, 1, 4, 4			scale = 5, 5, 5, 5, 10, 10, 20, 20		
8	10	0.997	0.997	0.999	0.575	0.598	0.695	0.907	0.924	0.968
	20	1	1	1	0.771	0.789	0.869	0.995	0.997	0.999
	50	1	1	1	0.978	0.981	0.993	1	1	1
		Median, Normal var = 5, 5, 10			Median, Chisq df = 2, 2, 4			Median, Laplace scale = 5, 5, 10		
3	10	0.348	0.355	0.361	0.114	0.116	0.121	0.225	0.228	0.236
	20	0.765	0.769	0.774	0.204	0.209	0.22	0.533	0.541	0.552
	50	0.998	0.998	0.998	0.518	0.526	0.538	0.943	0.945	0.947
		var = 5, 5, 5, 5, 10, 10, 25, 25			df = 2, 2, 2, 2, 1, 1, 4, 4			scale = 5, 5, 5, 5, 10, 10, 20, 20		
8	10	0.988	0.989	0.997	0.255	0.261	0.282	0.822	0.829	0.878
	20	1	1	1	0.492	0.513	0.639	0.991	0.993	0.999
	50	1	1	1	0.945	0.956	0.986	1	1	1
		Trimmed mean, Normal var = 5, 5, 10			Trimmed mean, Chisq df = 2, 2, 4			Trimmed mean, Laplace scale = 5, 5, 10		
3	10	0.42	0.43	0.435	0.162	0.168	0.177	0.272	0.277	0.285
	20	0.78	0.786	0.791	0.25	0.26	0.265	0.559	0.57	0.575
	50	0.997	0.997	0.997	0.561	0.57	0.579	0.945	0.948	0.95
		var = 5, 5, 5, 5, 10, 10, 25, 25			df = 2, 2, 2, 2, 1, 1, 4, 4			scale = 5, 5, 5, 5, 10, 10, 20, 20		
8	10	0.994	0.995	0.998	0.335	0.344	0.401	0.854	0.86	0.891
	20	1	1	1	0.566	0.588	0.704	0.992	0.995	0.999
	50	1	1	1	0.956	0.966	0.989	1	1	1

5. APPLICATION

Kvamme *et al.* [16] used Levene test and Brown–Forsythe robust version of Levene test to compare the dispersion of the apertures of the chalupa pots that vary in the method they arrange ceramic production from 3 locations, Dalupa (ApDg), Dangtalan (ApDg) and Paradijon (ApP). The data consists of 343 observations: ApDg that has 55 observations, ApDl that has 171 observations and ApP: that has 117 observations; see Gastwirth *et al.* [12].

Table 6 shows the mean, median and standard deviation (st. deviation) for pot data. The largest standard deviation is 12.73 (ApDg) followed by 8.13 (ApP) while the smallest standard deviation is 5.83 (ApP). Table 4 gives the results of Levene–Brown–Forsythe tests for pot data. The p-values of three tests are showing that the dispersion in every of 3 measured characteristics of the pots in different areas are statistically significant at 0.01 and 0.05.

Table 4: Levene–Brown–Forsythe tests for pot data.

	Mean	Trimmed mean	median
Test statistics	7.716	6.567	6.794
p-value	0.0005	0.0016	0.0013

On the other hand, Figure 1 illustrates the results of U-plot at both significance levels 0.01 and 0.05. Since the number of observations are not equal, the height of DL will be different.

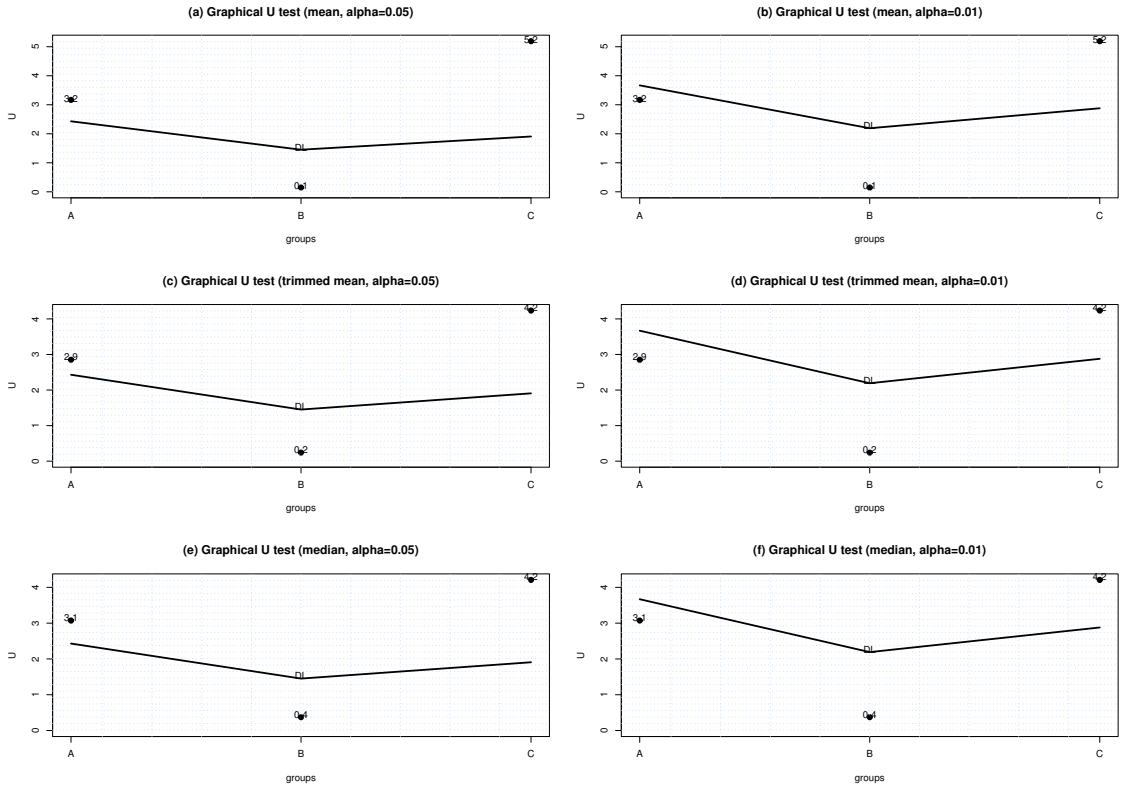

Figure 1: U plot for pot data using mean, trimmed mean and median as location measure.

For example, by using the quantile function of beta prime distribution of the second type, median as location measure and $\alpha = 0.05$, the decision limit is

$$DL = \text{qgb2}\left(1 - \frac{0.05}{3},\ p = 1,\ q = \frac{(343 - (55, 171, 117))(340)}{343(2)}, \alpha = \frac{1}{2},\ \beta = \frac{340}{2}\right).$$

This gives

$$DL = (2.43, 1.45, 1.91).$$

At 0.05, the values of U_1 and U_3 are outside the DL while the value of U_3 is outside DL for 0.01 based on mean, trimmed mean and median as location measures. Therefore, the dispersion in each of the three measured characteristics of the pots in different regions are statistically significant at 0.01 and 0.05 and the most different in dispersion comes from group 3.

The data for the second application is shown in Table 5 where these data are simulated from chi square distribution with $df = 1, 2, 2, 2, 2, 2, 2, 2$. The data consists of 8 groups and in every group, there are 20 observations.

Table 5: Simulated data from χ^2 $(df = 1, 2, 2, 2, 2, 2, 2, 2)$ distribution.

k1	k2	k3	k4	k5	k6	k7	k8
0.27	6.14	3.73	1.13	3.22	1.93	1.07	0.83
1.46	0.1	3.48	0.39	6.28	0.46	2.25	3.89
0.6	1.75	8.23	0.47	1.89	2.35	0.86	0.66
0.49	0.82	1.09	1.53	0.41	2.1	0.92	1.89
0.78	1.7	0.04	5.22	5.78	1.14	1.73	3.27
1.92	0.35	7.03	1.09	2.5	0.94	3.26	4.75
0.11	3.76	8.03	2	0.89	4.12	2.92	5.46
4.9	3.04	0.51	2.6	4.2	5.52	4.31	0.43
1.47	1.68	4.07	0.73	2.2	3.36	1.11	6.3
0.08	3.44	3.5	2.02	0.95	2.75	4.84	5.47
0.64	2.95	0.42	0.44	7.2	0.12	1.38	7.63
0.48	0.1	0.4	0.92	3.45	0.33	0.5	3.25
0.4	0.53	0.63	0.93	2.37	2.18	0.4	4.51
5.37	0.15	2.8	2.73	3.74	1.75	2.24	1.11
0.05	2.16	0.14	3.34	1.29	2.93	1.25	1.4
1.18	0.07	9.48	3.32	0.35	3.45	5.39	2.93
0.01	1.27	0.49	0.47	0.67	1.47	0.48	1.36
0.18	0.67	2.98	3.33	1.68	0.07	0.43	0.32
1.09	2.17	0.2	2.13	0.44	2.25	1.89	1.98
5.07	2.91	2.26	0.82	1.67	0.53	0.26	6.12

Table 6 shows the mean, median and standard deviation (st. deviation) for χ^2 simulated data. The largest standard deviation is 3.02 (k3) followed by 1.24 (k8) while the smallest standard deviation is 1.31 (k4) followed by second smallest 1.54 (k7).

Table 7 gives the results of Levene–Brown–Forsythe tests for simulated data from chi square distribution. The p-values of Levene–Brown–Forsythe tests are showing that the variances in each of the eight groups are statistically significant at 0.01 and 0.05.

With respect to U plot, Figure 2 displays the results of U-plot at both significance levels 0.01 and 0.05 and using mean, trimmed and median as location measures. Since the number of observations are equal, the height of DL will be the same. For example, by using

the quantile function of beta prime distribution of the second type and $\alpha = 0.01$, the decision limit can be computed as

$$DL = \text{qgb2}\left(1 - \frac{0.01}{8},\ p = 1,\ q = \frac{(160-20)(160-8)}{160(7)},\ \alpha = \frac{1}{2},\ \beta = \frac{160}{2}\right) = 1.35.$$

At 0.05 and 0.01, the value of U_1 is outside the DL using mean, trimmed mean and median as location measures. Therefore, the assumption of homogeneity of variances is rejected and the most different in dispersion comes from group 3.

Table 6: Summary statistics for Pot and simulation data.

	Pot data			Simulation data							
	ApDg	ApDl	ApP	k1	k2	k3	k4	k5	k6	k7	k8
#	55	171	117	20	20	20	20	20	20	20	20
Mean	170.5	163	128.6	1.33	1.79	2.98	1.78	2.56	1.99	1.87	3.18
median	170	165	130	0.62	1.69	2.53	1.33	2.04	2.02	1.31	3.09
st. devation	12.739	8.127	5.829	1.72	1.58	3.02	1.31	2.02	1.44	1.54	2.24

Table 7: Levene–Brown–Forsythe tests for simulated data.

	Mean	Trimmed mean	median
Test statistics	3.316	2.876	2.859
p-value	0.0026	0.0075	0.0078

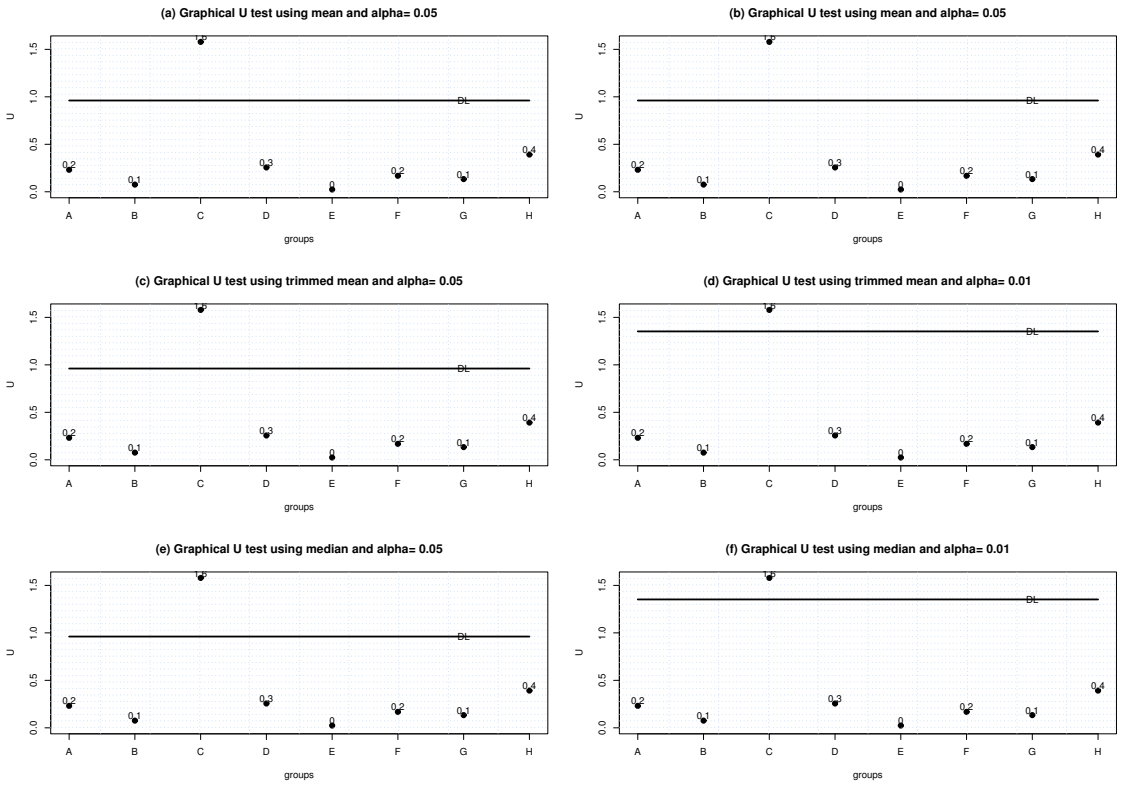

Figure 2: U plot for simulated data from chi square distribution using mean, trimmed mean and median as location measure.

6. DISCUSSION

The Levene–Brown–Forsythe test can be rewritten as

$$W = \sum_{i=1}^{k} \frac{n_i(Z_{i.} - Z_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i.})^2/(n-k)}.$$

This can be interpreted as an aggregate way to test whether the level factor mean absolute deviations differ from the overall mean absolute deviation. In terms of the null hypothesis, it tests for the equality of the mean absolute deviations for different factor levels. In terms of alternative hypothesis, it tests that at least two mean absolute deviations for factor levels are not equal. The U_i tests can be rewritten as

$$U_i = \frac{n_i(Z_{i.} - Z_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i.})^2/(n-k)}, \qquad i = 1, 2, ..., k.$$

These are simultaneous tests that show every level mean absolute deviation and the decision limit on the graph. If a value of any factor level mean absolute deviation is outside the decision limit, there is evidence that the level factor mean absolute deviation represented by that value is significantly different from the overall mean absolute deviation. In other words, these plots show whether there is statistically significant evidence of each group mean absolute deviation from centre differing from the overall mean absolute deviation from centre. In terms of alternative hypothesis, it tests at least one mean absolute deviation for factor levels is not equal the overall mean absolute deviation.

7. CONCLUSION

Assessing the homogeneity of variance is a prevalent question in many statistical analyses such as regression and analysis of variance. A graphical U_i test for homoscedasticity is proposed as the ratio for the contribution of each between squares treatment to mean square error of all treatments where the sum of the U_i is Levene–Brown–Forsythe tests. The sampling distribution of U_i is derived as beta prime distribution of the second type. By using Bonferroni approximation and Benjamini–Hochberg method, the decision line had been obtained to decide about homogeneity of variances when all values of U_i are less than decision limit or heterogeneity of variances when any value of U_i lies outside the decision line.

Overall, the simulation results showed that the performance of U_i plot is similar to Levene–Brown–Forsythe tests using different designs of number of groups and the number of observations in terms of type I error and test power. Therefore, it can be concluded that U_i plot using mean and trimmed means as a location is suited to symmetric distributions and U_i plot using median as a location was suited to asymmetric distribution. Moreover, if there are no ideas about the shape of the data, the U_i based on median should be used as a general test where it gives a good control for type I error and reasonable power in case of asymmetric distributions while hold a reasonable type I error control and test power in symmetric distributions.

There are many advantages of using U_i plot:

(**a**) provides a powerful visual tool for testing homogeneity of variances;

(**b**) keeps the size and power of the test like Levene–Brown–Forsythe tests;

(**c**) does not need to pairwise comparisons where it could be considered as a complement method to original test.

ACKNOWLEDGMENTS

The author acknowledge the valuable remarks and suggestions made by the referees and editor which improved the presentation of the paper.

REFERENCES

[1] Abdi, H. (2007). *The Bonferroni and Sidak corrections for multiple comparisons*. In "Encyclopedia of Measurement and Statistics" (Neil Salkind), Thousand Oaks (CA), Sage.

[2] Aslam, M. and Khan, N. (2021). Normality test of temperature in Jeddah city using Cochran's test under indeterminacy, *MAPAN: Journal of Metrology Society of India*, **36**(3), 589–598.

[3] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing, *Journal of Royal Statistical Society B*, **57**(1), 289–300.

[4] Berger, A.K.; Gottdiener, J.S.; Yohe, M.A. and Guerro, J.L. (1999). Epidemiologic approach to quality assessment in echocardiographic diagnosis, *Journal of the American College of Cardiology*, **34**(6), 1831–1836.

[5] Brown, M.B. and Forsythe, A.B. (1974). Robust tests for equality of variances, *Journal of the American Statistical Association*, **69**(346), 364–367.

[6] Coelho, C. and Mexia, J. (2007). On the distribution of the product and ratio of independent generalized gamma-ratio random variables, *Sankhya: The Indian Journal of Statistics*, **69**(2), 221–255.

[7] Elamir, E.A.H. (2021). Simultaneous test for means: An Unblind way to the F-test in One-way analysis of variance, *Statistics, Optimization and Information Computing*, **X**, Month 2021, 0–14.

[8] Fligner, M.A. and Killeen, T.J. (1976). Distribution-free two-sample tests for scale, *Journal of the American Statistical Association*, **71**(353), 210–213.

[9] Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*, Sage Publication Inc.

[10] Frane, A.V. (2015). Are per-family type I error rates relevant in social and behavioral science, *Journal of Modern Applied Statistical Methods*, **14**(1), 12–23.

[11] Gastwirth, J.; Yulia, R. and Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice, *Statistical Science*, **24**(2), 343–360.

[12] Gastwirth, J.; Yulia, R.; Wallace, H.; Lyubchich, V.; Miao, W. and Noguchi, K. (2009). *Lawstat: tools for biostatistics, public policy and law*, Version 3.4. https://cran.r-project.org/web/packages/lawstat/index.html

[13] Graf, M. and Nedyalkova, D. (2015). *Generalized beta distribution of the second kind: properties, likelihood, estimation.* https://cran.r-project.org/web/packages/GB2/index.html

[14] Johnson, A.R. and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*, 6th Ed., Pearson, Prentice Hall.

[15] Kutner, M.; Nachtsheim, C.; Neter, J. and William, L. (2004). *Applied Linear Statistical Models*, 5th Ed., McGraw-Hill/Irwin.

[16] Kvamme, K.L.; Stark, M.T. and Longacre, M.A. (1996). Alternative procedures for assessing standardization in ceramic assemblages, *American Antiquity*, **61**(1), 116–126.

[17] Levene, H. (1960). *Robust testes for equality of variances.* In "Contributions to Probability and Statistics" (I. Olkin), Stanford Univ. Press, Palo Alto, CA. MR0120709, 278–292.

[18] Miller, R.G. (1986). *Beyond ANOVA: Basics of Applied Statistics*, Wiley, New York.

[19] Nguyen, D.; Kim, E.; Wang, Y.; Pham, T.V.; Chen, Y.-H. and Kromrey, J.D. (2019). Empirical comparison of tests for one-factor ANOVA under heterogeneity and non-normality: a Monte Carlo study, *Journal of Modern Applied Statistical Methods*, **18**(2), 2–30.

[20] Nordstokke, D.W. and Zumbo, B.D. (2010). A new nonparametric Levene test for equal variances, *Psicologica*, **31**(2), 401–430.

[21] Nordstokke, D.W.; Zumbo, B.D.; Cairns, S.L. and Saklofske, D.H. (2011). The operating characteristics of the nonparametric Levene test for equal variances with assessment and evaluation data, *Practical Assessment, Research and Evaluation*, **16**(5), 1–8.

[22] Plourdes, A. and Watkins, G.C. (1998). Crude oil prices between 1985 and 1994: How volatile in relation to other commodities?, *Resource and Energy Economics*, **20**(2), 245–262.

[23] R Core Team (2021). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[24] Sant, R. and Cowan, A.R. (1994). Do dividends signal earnings?the case of omitted dividends, *Journal of Banking and Finance*, **18**(6), 1113–1133.

[25] Shear, B.R.; Nordstokke, D.W. and Zumbo, B.D. (2018). A note on using the nonparametric Levene test when population means are unequal, *Practical Assessment, Research, and Evaluation*, **23**(2), 1–11.

[26] Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments, *Statistical Applications in Genetics and Molecular Biology*, **3**(1), 1–11.

[27] Wang, Y.; Rodriguez de Gil, P.; Chen, Y.; Kromrey, J.; Kim, E.; Pham, T.; Nguyen, D. and Romano, J. (2017). Comparing the performance of approaches for testing the homogeneity of variance assumption in One-Factor ANOVA models, *Educational and Psychological Measurement*, **77**(3), 305–329.

[28] Yigit, S. and Mendes, M. (2018). Which effect size measure is appropriate for one-way and two-way ANOVA models? A Monte Carlo simulation study, *REVSTAT – Statistical Journal*, **16**(3), 295–313.

REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 405–424
https://doi.org/10.57805/revstat.v21i3.174

Estimating the Parameters of Burr Type XII Distribution with Fuzzy Observations

Authors: Abbas Abdul Hussein
– Department of Mathematics, University of Thi-Qar, Thi-Qar, Iraq
abbas1989.aa1989@gmail.com

Riyadh Al-Mosawi ✉
– Department of Mathematics, University of Thi-Qar, Thi-Qar, Iraq
riyadhrm@gmail.com ; riyadhrm@utq.edu.iq

Received: February 2021 Revised: January 2022 Accepted: January 2022

Abstract:

- In this article, the classical as well as the Bayesian estimation problems of two-parameter Burr type XII distribution based on fuzzy data are considered. The maximum likelihood estimators via two methods, namely, Newton-Raphson and Expectation-Maximization algorithms are computed. The standard errors of the estimates are computed using the observed information matrix. For computing the Bayes estimators, three methods viz Lindley's approximation, Tierney–Kadane approximation and highest posterior density method are obtained. Monte-Carlo simulation experiments are conducted to investigate the performance of the proposed methods. Finally, the proposed methods are illustrated by using three different real data sets.

Keywords:

- *Bayesian estimation; Burr type XII distribution; expectation-maximization algorithm; fuzzy observations; Lindley's approximation; maximum likelihood estimation; Tierney–Kadane approximation.*

AMS Subject Classification:

- 62N02, 62N86.

✉ Corresponding author.

1. INTRODUCTION

The Burr type XII distribution was first introduced in the literature by Burr ([5]). It has gained special attention in the last two decades and applied in different fields including the area of reliability, failure time modeling and acceptance sampling plan and so on. The two-parameter Burr type XII distribution has the following probability density function

$$f(x;\alpha,\beta) = \alpha\beta x^{\alpha-1}(1+x^{\alpha})^{-(\beta+1)},\ x>0, \alpha>0, \beta>0, \tag{1.1}$$

where α and β represent the shape parameters. It is easy to see that when $\alpha = 1$, the Burr type XII reduces to the log-logistic distribution. Maximum likelihood and Bayesian inferential issues for the unknown parameters of Burr type XII distribution with different types of data were considered by several authors. See, for example, Wang *et al.* ([35]), Moore and Papadopoulos ([21]), Ghitany and Al-Awadhi ([11]), Mousa and Jaheen ([22]), Wahed ([34]), Li *et al.* ([17]), Jaheen and Okasha ([13]), Panahi and Asadi ([29]), Al-Baldawi *et al.* ([1]), Rao *et al.* ([31]), Belaghi *et al.* ([3]) and Hakim *et al.* ([12]).

All the earlier works on the estimation of the parameters of the Burr type XII distribution have been done under the assumption of precise data. In the classical estimation theory, we consider only one source of uncertainty available, namely randomness. However, in many practical situations, in addition to the randomness, we may face other source of uncertainties, namely, vague uncertainty. Vagueness occurs as a result of imprecisely recording or measuring the observations due to, for example, machine errors, human errors, etc. For instance, the lifetime of a specific electric device may be recorded as vague statements like "about 3 years", "approximately less than 2 years", "approximately 3 years", "approximately between 3 and 4 years" and so on.

In recent years, many papers extended the statistical methods to analysis of fuzzy data for different distributions. Among others, Denœux ([8]), for a general parametric statistical model, showed that the EM algorithm may be used for analyzing statistical problems involving fuzzy data. Pak *et al.* ([27]) investigated different classical and Bayesian methods for estimating the parameters of Weibull distribution when the available data are in the form of fuzzy numbers. Pak *et al.* ([28]) discussed different procedures for estimating the parameter of Rayleigh distribution under doubly type II censoring when the available observations are described by means of fuzzy information. They computed the maximum likelihood, highest posterior density and method of moments estimators. Makhdoom *et al.* ([20]) estimated the parameter of exponential distribution on the basis of type II censoring scheme when the available data are in the form of fuzzy numbers. The Bayes estimate of the unknown parameter was also obtained under the assumption of gamma prior. Khoolenjani and Shahsanaie ([15]) derived the maximum likelihood estimator of the mean of exponential distribution under type II censoring scheme when the lifetime observations are in the form of fuzzy numbers. They also obtained the estimate, via Bayesian method, of the unknown parameter. Pak ([23]) obtained the maximum likelihood estimation and Bayesian estimation for Lindley distribution when the available observations are reported in the form of fuzzy data. The classical and Bayesian inferences for the Pareto distribution of life time fuzzy observations was studied by Shafiq ([32]). Chaturvedi *et al.* ([6]) presented procedures of parameter estimation of the Rayleigh distribution based on type II progressively hybrid censored fuzzy lifetime data.

Classical as well as the Bayesian procedures for the estimation of unknown parameters were investigated. Pak and Mahmoudi ([26]) estimated the parameters of Lomax distribution when the available observations are described by means of fuzzy information. They computed the maximum likelihood and the Bayesian estimators. Basharat *et al.* ([2]) derived the distribution of a linear combination of two independent exponential random variables. The parameter estimates of the proposed distribution were obtained by using the maximum likelihood estimation method and the method of moments from fuzzy data. Finally, Pak *et al.* ([25]) provided Bayesian inference for the parameters of the generalized exponential model under asymmetric and symmetric loss functions when the observations are described in terms of fuzzy numbers.

To the best of our knowledge, there are no studies focused on the analysis of fuzzy data on the parameter estimation of two-parameter Burr type XII distribution. The main purpose of this paper is to investigate the inferential procedures for the distribution of the two parameters of Burr type XII, where the available data is in the form of fuzzy data. In Section 2, we review the basic notations and definitions of fuzzy set theory. In Section 3, we address the estimation of the unknown parameters of the maximum likelihood estimates using the Newton–Raphson and expectation-maximization (EM) algorithm. In Section 4, the Bayes estimates of the unknown parameters are obtained via Lindley's approximation, Tierney–Kadane approximation and highest posterior distribution estimation method under the assumption of Gamma priors. A Monte Carlo simulation study is conducted in Section 5, to assess the performance of the proposed estimators. For illustration, analyses of three datasets are provided. Finally, some conclusions are provided in Section 6.

2. BASIC DEFINITION OF FUZZY SETS

In this section, we review some basic definitions and notations of fuzzy sets and fuzzy probability theory used in this paper. Suppose a random experiment with a probability space $(\mathbb{R}^m, \mathcal{B}^m, P_\theta)$, where $\mathbb{R}^m$ is an m-dimensional Euclidean space, $\mathcal{B}^m$ is the smallest Borel σ-field defined on $\mathbb{R}^m$ and P_θ, $\theta \in \Theta$, is a probability measure defined on $\mathcal{B}^m$. In many applications, we have a situation that the outcome of the experiment cannot be observed exactly and only partial information is available. For example, the lifetime of a specific electric device may be recorded as "about 3 years", "approximately less than 2 years", "approximately 3 years", "approximately between 3 and 4 years" and so on. These lifetimes can be modeled and described in the form of fuzzy subset. A fuzzy set $\tilde{A}$ in $\mathbb{R}^m$ is characterized by a membership function $\mu_{\tilde{A}} : \mathbb{R}^m \to [0, 1]$, where $\mu_{\tilde{A}}(x)$, $x \in \mathbb{R}^m$, represents the degree of membership of x in $\tilde{A}$. A fuzzy event is a fuzzy set whose membership function is Borel measurable function. According to Zadeh ([36]) the probability of a fuzzy event $\tilde{A}$ is computed by

$$P(\tilde{A}) = \int \mu_{\tilde{A}}(x) dP_\theta. \tag{2.1}$$

The most common fuzzy subsets that are frequently encountered in fuzzy statistical analysis are the fuzzy numbers and among them, the triangular fuzzy numbers are the most common type. A triangular fuzzy number, written as $\tilde{x} = (a, b, c)$, has the following membership

function

$$\mu_{\tilde{x}}(x) = \begin{cases} \dfrac{x-a}{b-a}, & \text{if } a \le x \le b, \\ \dfrac{c-x}{c-b}, & \text{if } b \le x \le c, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, assume X to be a random variable with a probability density function (p.d.f.) $g(x;\theta)$ that is absolutely continuous with probability measure P_θ. The conditional probability of a crisp (non-fuzzy) set A given a fuzzy set $\tilde{B}$ is given by (see Denœux ([8]))

$$P(A|\tilde{B}) = \frac{\int_A \mu_{\tilde{B}}(x) g(x;\theta)dx}{\int \mu_{\tilde{B}}(x) g(x;\theta)dx}.$$

Consequently, the conditional density of X given $\tilde{B}$ can thus be computed by

$$g(x|\tilde{B}) = \frac{\mu_{\tilde{B}}(x) g(x;\theta)}{\int \mu_{\tilde{B}}(x) g(x;\theta)dx}.$$

3. MAXIMUM LIKELIHOOD ESTIMATION

Let $X_1, X_2, ..., X_n$ denote a random sample of size n from Burr type XII distribution with p.d.f. given in (1.1). Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ denote the corresponding random vector. If a realization $\mathbf{x}$ of $\mathbf{X}$ was exactly observed, the likelihood function can be written as

$$L(\alpha, \beta|\mathbf{x}) = (\alpha\beta)^n \prod_{i=1}^{n} x_i^{\alpha-1}(1+x_i^\alpha)^{-\beta-1}. \tag{3.1}$$

Suppose now $\mathbf{x}$ is not observed precisely, and only partial information about $\mathbf{x}$ is available in form of fuzzy observation $\tilde{\mathbf{x}} = (\tilde{x_1}, ..., \tilde{x_n})$ with Borel measurable membership function $\mu_{\tilde{\mathbf{x}}}(\mathbf{x}) = (\mu_{\tilde{x_1}}(x), ..., \mu_{\tilde{x_n}}(x))$. Then, based on fuzzy observation $\tilde{\mathbf{x}}$, the log-likelihood function reduces to

$$\begin{aligned} l(\alpha, \beta|\tilde{\mathbf{x}}) &= n\log\alpha + n\log\beta + \sum_{i=1}^{n} \log \int x^{\alpha-1}(1+x^\alpha)^{-\beta-1} \mu_{\tilde{x_i}}(x)dx \\ &= n\log\alpha + n\log\beta + \sum_{i=1}^{n} \log \int A(x)\mu_{\tilde{x_i}}(x)dx, \end{aligned} \tag{3.2}$$

where

$$A(x) = x^{\alpha-1}(1+x^\alpha)^{-\beta-1}. \tag{3.3}$$

The maximum likelihood estimate of the parameters α and β can be obtained by maximizing the log-likelihood $l(\alpha, \beta|\tilde{\mathbf{x}})$ with respect to α and β. First we need to prove the following result.

Theorem 3.1. *The MLEs of α and β for $\alpha > 0$ and $\beta > 0$ exist and unique.*

Proof: The detailed proof of the theorem is deferred in the Appendix. □

By taking the partial derivatives of the log-likelihood $l(\alpha, \beta|\tilde{\mathbf{x}})$ with respect to α and β and equating the resulted equations to zero, we get the following two normal equations

$$\frac{\partial l(\alpha,\beta|\tilde{\mathbf{x}})}{\partial\alpha} \equiv l_\alpha = \frac{n}{\alpha} + \sum_{i=1}^{n} \frac{\int A_\alpha(x)\mu_{\tilde{x}_i}(x)dx}{\int A(x)\mu_{\tilde{x}_i}(x)dx} = 0 \tag{3.4}$$

and

$$\frac{\partial l(\alpha,\beta|\tilde{\mathbf{x}})}{\partial\beta} \equiv l_\beta = \frac{n}{\beta} + \sum_{i=1}^{n} \frac{\int A_\beta(x)\mu_{\tilde{x}_i}(x)dx}{\int A(x)\mu_{\tilde{x}_i}(x)dx} = 0, \tag{3.5}$$

where

$$A_\alpha(x) \equiv \frac{\partial A(x)}{\partial\alpha} = (1+x^\alpha)^{-\beta-2}x^{\alpha-1}\log(x)[1-\beta x^\alpha],$$
$$A_\beta(x) \equiv \frac{\partial A(x)}{\partial\beta} = -x^{\alpha-1}(1+x^\alpha)^{-\beta-1}\log(1+x^\alpha).$$

Since there are no closed forms to the normal equations (3.4) and (3.5), iterative numerical methods can be used to obtain the MLEs. In this section, we propose two methods to compute the MLEs of α and β, namely; Newton–Raphson method and EM method.

3.1. Newton–Raphson algorithm

The Newton–Raphson (NR) method is a numerical approach that is commonly used to compute MLEs of the unknown parameters. In this method, the solution of the likelihood function is obtained through an iterative procedure. First, we obtain the second-order derivatives of the log-likelihood with respect to α and β in order to implement the NR method:

$$l_{\alpha\alpha} = \frac{-n}{\alpha^2} + \sum_{i=1}^{n} \frac{\int A(x)\mu_{\tilde{x}_i}(x)dx \int A_{\alpha\alpha}(x)\mu_{\tilde{x}_i}(x)dx - (\int A_\alpha(x)\mu_{\tilde{x}_i}(x)dx)^2}{(\int A(x)\mu_{\tilde{x}_i}(x)dx)^2}, \tag{3.6}$$

$$l_{\beta\beta} = \frac{-n}{\beta^2} + \sum_{i=1}^{n} \frac{\int A(x)\mu_{\tilde{x}_i}(x)dx \int A_{\beta\beta}(x)\mu_{\tilde{x}_i}(x)dx - (\int A_\beta(x)\mu_{\tilde{x}_i}(x)dx)^2}{(\int A(x)\mu_{\tilde{x}_i}(x)dx)^2}, \tag{3.7}$$

$$l_{\alpha\beta} = \sum_{i=1}^{n} \frac{\int A(x)\mu_{\tilde{x}_i}(x)dx \int A_{\alpha\beta}(x)\mu_{\tilde{x}_i}(x)dx - \int A_\alpha(x)\mu_{\tilde{x}_i}(x)dx \int A_\beta(x)\mu_{\tilde{x}_i}(x)dx}{(\int A(x)\mu_{\tilde{x}_i}(x)dx)^2}, \tag{3.8}$$

where

$$A_{\alpha\alpha}(x) = x^{\alpha-1}(\log(x))^2(1+x^\alpha)^{-\beta-3}\Big[x^{2\alpha}(\beta+1)(\beta+2) - 3x^\alpha(\beta+1)(1+x^\alpha) + (1+x^\alpha)^2\Big],$$
$$A_{\beta\beta}(x) = x^{\alpha-1}(1+x^\alpha)^{-\beta-1}(\log(1+x^\alpha))^2$$
$$A_{\alpha\beta}(x) = x^{2\alpha-2}(1+x^\alpha)^{-\beta-2}\log(x)[(\beta+1)x\log(1+x^\alpha) - (1+x^\alpha)\log(1+x^\alpha) - x].$$

Assume $\alpha^{(k)}$ and $\beta^{(k)}$ are the values of α and β at the k-th iteration. Then at $(k+1)$-th iteration, the updated values of α and β are obtained as

$$\begin{pmatrix} \alpha^{(k+1)} \\ \beta^{(k+1)} \end{pmatrix} = \begin{pmatrix} \alpha^{(k)} \\ \beta^{(k)} \end{pmatrix} - \begin{pmatrix} l_{\alpha\alpha} & l_{\alpha\beta} \\ l_{\alpha\beta} & l_{\beta\beta} \end{pmatrix}^{-1}_{\alpha=\alpha^{(k)},\beta=\beta^{(k)}} \begin{pmatrix} l_\alpha \\ l_\beta \end{pmatrix}_{\alpha=\alpha^{(k)},\beta=\beta^{(k)}},$$

which is equivalent to

$$\alpha^{(k+1)} = \alpha^{(k)} - \frac{l_\alpha l_{\beta\beta} - l_\beta l_{\alpha\beta}}{l_{\alpha\alpha} l_{\beta\beta} - l_{\alpha\beta}^2}\Big|_{\alpha=\alpha^{(k)},\beta=\beta^{(k)}}, \tag{3.9}$$

$$\beta^{(k+1)} = \beta^{(k)} - \frac{l_\beta l_{\alpha\alpha} - l_\alpha l_{\alpha\beta}}{l_{\alpha\alpha} l_{\beta\beta} - l_{\alpha\beta}^2}\Big|_{\alpha=\alpha^{(k)},\beta=\beta^{(k)}}. \tag{3.10}$$

The iteration process then continues until convergence, i.e., $|\alpha^{(k+1)} - \alpha^{(k)}| + |\beta^{(k+1)} - \beta^{(k)}| < \varepsilon$, for some pre-specified $\varepsilon > 0$.

To estimate the standard error of maximum likelihood estimators, $\hat{\alpha}$ and $\hat{\beta}$, we use the observed information matrix method. The variance-covariance matrix of the MLEs of α and β is defined as

$$\Sigma = \begin{bmatrix} \mathrm{var}(\hat{\alpha}) & \mathrm{cov}(\hat{\alpha},\hat{\beta}) \\ \mathrm{cov}(\hat{\alpha},\hat{\beta}) & \mathrm{var}(\hat{\beta}) \end{bmatrix},$$

and can be estimated by using the inverse of the observed information matrix

$$I(\hat{\alpha},\hat{\beta}) = \begin{pmatrix} -l_{\alpha\alpha} & -l_{\alpha\beta} \\ -l_{\alpha\beta} & -l_{\beta\beta} \end{pmatrix}_{\alpha=\hat{\alpha},\beta=\hat{\beta}}, \tag{3.11}$$

where $l_{\alpha\alpha}, l_{\beta\beta}$ and $l_{\alpha\beta}$ are given in (3.6),(3.7) and (3.8), respectively. Then the $100(1-\gamma)\%$ Wald confidence intervals of α and β using the observed information matrix can be constructed, respectively, as

$$\hat{\alpha} \pm z_{\gamma/2}\sqrt{\mathrm{var}(\hat{\alpha})} \text{ and } \hat{\beta} \pm z_{\gamma/2}\sqrt{\mathrm{var}(\hat{\beta})},$$

where z_p is the upper p-th percentile of the standard normal distribution.

It is known that Newton–Raphson method is very sensitive to the initial values of parameters. In addition, the calculation of the second-order derivatives of the log-likelihood based on fuzzy data sometimes can be rather tedious. So we propose to use an alternative method to the Newton–Raphson method which is the EM algorithm.

3.2. EM Algorithm

In this subsection, we propose to use the EM algorithm to calculate the MLEs of the unknown parameters.

The EM algorithm, proposed by Dempster *et al.* ([7]), is a very powerful technique used in parameter estimation based on incomplete or missing information data. As stated by Pradhan and Kundu ([30]), the EM algorithm is an iterative method and each iteration consists of two main steps; Expectation(E)-step and Maximization(M)-step. In E-step, we form

the "pseudo-likelihood" function by replacing the incomplete or missing observations in the likelihood function with their corresponding expected values. In the M-step, we maximize the "pseudo-likelihood" function with respect to the parameters. Let us denote the observed data set by $\tilde{\mathbf{X}} = (\tilde{X}_1, ..., \tilde{X}_n)$ and let the complete data denoted by $\mathbf{X} = (X_1, ..., X_n)$. Define $\mathbf{Z} = (Z_1, ..., Z_n)$ where Z_i represents the conditional expectation of the complete observation X_i given the corresponding fuzzy observation $\tilde{X}_i$ with membership function $\mu_{\tilde{x}_i}(x)$. Observe that

$$Z_i = E(X_i|\tilde{X}_i) = \frac{\int x f(x;\alpha,\beta)\mu_{\tilde{x}_i}(x)dx}{\int f(x;\alpha,\beta)\mu_{\tilde{x}_i}(x)dx}, \qquad i = 1, ..., n. \tag{3.12}$$

Then the pseudo likelihood function takes the form

$$L^c(\alpha,\beta|\mathbf{z}) = (\alpha\beta)^n \prod_{i=1}^{n} z_i^{\alpha-1}(1+z_i^{\alpha})^{-\beta-1}, \tag{3.13}$$

with pseudo log-likelihood function

$$l^c(\alpha,\beta|\mathbf{z}) = n\log\alpha + n\log\beta + (1-\alpha)\sum_{i=1}^{n}\log(z_i) - (\beta+1)\sum_{i=1}^{n}\log(1+z_i^{\alpha}). \tag{3.14}$$

By taking the partial derivatives of l^c with respect to α and β, respectively, and equating the resulted equations to zero we obtain the following equations:

$$\frac{n}{\alpha} + \sum_{i=1}^{n}\log(z_i) - (\beta+1)\sum_{i=1}^{n}\frac{z_i^{\alpha}\log(z_i)}{(1+z_i^{\alpha})} = 0, \tag{3.15}$$

$$\frac{n}{\beta} - \sum_{i=1}^{n}\log(1+z_i^{\alpha}) = 0. \tag{3.16}$$

Therefore the EM algorithm is given by the following iterative process:

Step 1. Given starting values of α and β, say $\alpha^{(0)}$ and $\beta^{(0)}$, and take k=0.

Step 2. At the $(k+1)$-th iteration,

Step 2.1. E-step. Evaluate $\mathbf{Z} = (Z_1, ..., Z_n)$, where $Z_i \equiv Z_i(\alpha^{(k)}, \alpha^{(k)})$ is computed using the expression (3.12) with α and β are replaced by $\alpha^{(k)}$ and $\beta^{(k)}$, respectively.

Step 2.2. M-step. Solve the equations (3.15) and (3.16) and obtain the next values $\alpha^{(k+1)}$ and $\beta^{(k+1)}$ of α and β, respectively.

Step 3. If $|\alpha^{(k+1)} - \alpha^{(k)}| + |\beta^{(k+1)} - \beta^{(k)}| < \varepsilon$, for some pre-specified value $\varepsilon > 0$, then set $\alpha^{(k+1)}$ and $\beta^{(k+1)}$ as the maximum likelihood estimators of α and β, otherwise, set $k = k+1$ and go to **Step 2**.

Estimating the standard errors and constructing the confidence intervals in this section are similar to that given in Section 2 with NR estimates are replaced by EM estimates.

4. BAYESIAN ESTIMATION

In this section, we estimate the unknown parameters of Burr type XII distribution using Bayesian method under squared error loss function. The Bayes estimators are obtained using three different methods; Lindley's approximation, Tierney–Kadane approximation and highest posterior density methods. Assume that the parameters α and β have independent gamma priors such that $\alpha \sim \pi_1(\alpha) = Gamma(a,b)$ and $\beta \sim \pi_2(\beta) = Gamma(c,d)$. Based on the above priors, the joint posterior density function of α and β given the data can be written as follows

$$\pi(\alpha,\beta|\tilde{\mathbf{x}}) = \frac{\alpha^{n+a-1}\beta^{n+c-1}e^{-b\alpha-d\beta}\prod_{i=1}^{n}\int_0^\infty x^{\alpha-1}(1+x^\alpha)^{-\beta-1}\mu_{\tilde{x}_i}(x)dx}{\int_0^\infty\int_0^\infty \alpha^{n+a-1}\beta^{n+c-1}e^{-b\alpha-d\beta}\prod_{i=1}^{n}\int_0^\infty x^{\alpha-1}(1+x^\alpha)^{-\beta-1}\mu_{\tilde{x}_i}(x)dx d\alpha d\beta}. \tag{4.1}$$

Then, under a squared error loss function, the Bayes estimate of any function of α and β, say $g(\alpha,\beta)$, is given by

$$E(g(\alpha,\beta)|\tilde{\mathbf{x}}) = \int_0^\infty\int_0^\infty g(\alpha,\beta)\pi(\alpha,\beta|\tilde{\mathbf{x}})d\alpha d\beta. \tag{4.2}$$

Note that Equation (4.2) cannot be obtained analytically; therefore, in the following, we propose to use three methods, namely; Lindley's approximation and Tierney–Kadane approximation and highest posterior density methods to solve it and compute the Bayes estimators.

4.1. Lindley's Approximation

Lindley's approximation was proposed by Lindley ([18]) to approximate the integrals involved in Bayes estimator. Lindley proposed a ratio of integrals of the form

$$E(g(\alpha,\beta)|\tilde{\mathbf{x}}) = \frac{\int_0^\infty\int_0^\infty g(\alpha,\beta)e^{Q(\alpha,\beta)}d\alpha d\beta}{\int_0^\infty\int_0^\infty e^{Q(\alpha,\beta)}d\alpha d\beta} \tag{4.3}$$

that can be approximated by

$$\begin{aligned}\hat{g}(\alpha,\beta) &= g(\hat{\alpha},\hat{\beta}) + \frac{1}{2}\Big[(\hat{g}_{\alpha\alpha}+2\hat{g}_\alpha\hat{\rho}_\alpha)\hat{\sigma}_{\alpha\alpha} + (\hat{g}_{\alpha\beta}+2\hat{g}_\beta\hat{\rho}_\alpha)\hat{\sigma}_{\alpha\beta} + (\hat{g}_{\alpha\beta}+2\hat{g}_\alpha\hat{\rho}_\beta)\hat{\sigma}_{\alpha\beta}\\ &+ (\hat{g}_{\beta\beta}+2\hat{g}_\beta\hat{\rho}_\beta)\hat{\sigma}_{\beta\beta}\Big] + \frac{1}{2}\Big[(\hat{g}_\alpha\hat{\sigma}_{\alpha\alpha}+\hat{g}_\beta\hat{\sigma}_{\alpha\beta})(l_{\alpha\alpha\alpha}\hat{\sigma}_{\alpha\alpha}+2\hat{l}_{\alpha\alpha\beta}\hat{\sigma}_{\alpha\beta}+\hat{l}_{\alpha\beta\beta}\hat{\sigma}_{\beta\beta})\\ &+ (\hat{g}_\alpha\hat{\sigma}_{\alpha\beta}+\hat{g}_\beta\hat{\sigma}_{\beta\beta})(\hat{l}_{\alpha\alpha\beta}\hat{\sigma}_{\alpha\alpha}+2\hat{l}_{\alpha\beta\beta}\hat{\sigma}_{\alpha\beta}+\hat{l}_{\beta\beta\beta}\hat{\sigma}_{\beta\beta})\Big],\end{aligned} \tag{4.4}$$

where

$$Q(\alpha,\beta) = \log[\pi_1(\alpha)\pi_2(\beta)] + \log L(\alpha,\beta|\tilde{\mathbf{x}}) \equiv \rho(\alpha,\beta) + \ell(\alpha,\beta|\tilde{\mathbf{x}}).$$

The expressions $\hat{l},\hat{g},\hat{\rho}$ and $\hat{\sigma}$ denote, respectively, the functions l,g,ρ and σ evaluated at $\hat{\alpha}$ and $\hat{\beta}$, the MLEs of α and β. Here, the expressions $\hat{g}_\alpha$, $\hat{g}_\beta$, $\hat{g}_{\alpha\alpha}$, $\hat{g}_{\alpha\beta}$ and $\hat{g}_{\beta\beta}$ denote the first and the second order partial derivatives of g with respect α and β evaluated at the MLEs of α and β. First note that, the expressions of $l_\alpha, l_\beta, l_{\alpha\alpha}, l_{\beta\beta}$ and $l_{\alpha\beta}$ are given in (3.4), (3.5),

(3.6), (3.7) and (3.8), respectively. The third order of partial derivatives of the log-likelihood function with respect to α and β are given by

$$\begin{aligned}
l_{\alpha\alpha\alpha} &= \frac{2n}{\alpha^3} + \sum_{i=1}^{n} \frac{C_i^2 C_{i,\alpha\alpha\alpha} - 3C_i C_{i,\alpha} C_{i,\alpha\alpha} + 2C_{i,\alpha}^3}{C_i^3},\\
l_{\beta\beta\beta} &= \frac{2n}{\beta^3} + \sum_{i=1}^{n} \frac{C_i^2 C_{i,\beta\beta\beta} - 3C_i C_{i,\beta} C_{i,\beta\beta} + 2C_{i,\beta}^3}{C_i^3},\\
l_{\alpha\beta\beta} &= \sum_{i=1}^{n} \frac{C_i^2 C_{i,\alpha\beta\beta} - 2C_i C_{i,\beta} C_{i,\alpha\beta} - C_i C_{i,\alpha} C_{i,\beta\beta} + 2C_{i,\alpha} C_{i,\beta}^2}{C_i^3},\\
l_{\alpha\alpha\beta} &= \sum_{i=1}^{n} \frac{C_i^2 C_{i,\alpha\alpha\beta} - 2C_i C_{i,\alpha} C_{i,\alpha\beta} - C_i C_{i,\alpha\alpha} C_{i,\beta} + 2C_{i,\alpha}^2 C_{i,\beta}}{C_i^3},
\end{aligned}$$

where

$$\begin{aligned}
C_i &= \int A(x)\mu_{\tilde{x}_i}(x)dx,\\
C_{i,\alpha} &= \int A_\alpha(x)\mu_{\tilde{x}_i}(x)dx, C_{i,\alpha\alpha} = \int A_{\alpha\alpha}(x)\mu_{\tilde{x}_i}(x)dx, C_{i,\alpha\alpha\alpha} = \int A_{\alpha\alpha\alpha}(x)\mu_{\tilde{x}_i}(x)dx,\\
C_{i,\beta} &= \int A_\beta(x)\mu_{\tilde{x}_i}(x)dx, C_{i,\beta\beta} = \int A_{\beta\beta}(x)\mu_{\tilde{x}_i}(x)dx, C_{i,\beta\beta\beta} = \int A_{\beta\beta\beta}(x)\mu_{\tilde{x}_i}(x)dx,\\
C_{i,\alpha\beta} &= \int A_{\alpha\beta}(x)\mu_{\tilde{x}_i}(x)dx, C_{i,\alpha\alpha\beta} = \int A_{\alpha\alpha\beta}(x)\mu_{\tilde{x}_i}(x)dx, C_{\alpha\beta\beta} = \int A_{\alpha\beta\beta}(x)\mu_{\tilde{x}_i}(x)dx,
\end{aligned}$$

and

$$\begin{aligned}
A_{\alpha\alpha\alpha}(x) &= x^{2\alpha-1}(\beta+1)(\log(x))^3(1+x^\alpha)^{-\beta-4}\Big[-x^{2\alpha}(\beta+2)(\beta+3)\\
&\quad + 6x^\alpha(1+x^\alpha)(\beta+2) - 7(1+x^\alpha)^2\Big] + x^{\alpha-1}(\log(x))^3(1+x^\alpha)^{-\beta-1},\\
A_{\beta\beta\beta}(x) &= -x^{\alpha-1}(\log(1+x^\alpha))^3(1+x^\alpha)^{-\beta-1},\\
A_{\alpha\beta\beta}(x) &= x^{\alpha-1}\log(1+x^\alpha)\log(x)(1+x^\alpha)^{-\beta-2}\Big[-x^\alpha(\beta+1)\log(1+x^\alpha),\\
&\quad + 2x^\alpha + \log(1+x^\alpha)(1+x^\alpha)\Big],\\
A_{\alpha\alpha\beta}(x) &= (\beta+1)(\log(x))^2 x^{2\alpha-1}(1+x^\alpha)^{-\beta-3}\Big[-x^\alpha(\beta+2)\log(1+x^\alpha) + x^\alpha\\
&\quad + 3(1+x^\alpha)\log(1+x^\alpha)\Big] + (\log(x))^2 x^{2\alpha-1}(1+x^\alpha)^{-\beta-3}\Big[x^\alpha(\beta+2)\\
&\quad - 3(1+x^\alpha)\Big] - (\log(x))^2 x^{\alpha-1}(1+x^\alpha)^{-\beta-1}\log(1+x^\alpha).
\end{aligned}$$

The function ρ given by

$$\rho(\alpha,\beta) = (a-1)\log(\alpha) - b\alpha + (c-1)\log(\beta) - d\beta$$

has the following partial derivatives:

$$\begin{aligned}
\rho_\alpha &= \frac{\partial\rho(\alpha,\beta)}{\partial\alpha} = \frac{a-1}{\alpha} - b,\\
\rho_\beta &= \frac{\partial\rho(\alpha,\beta)}{\partial\beta} = \frac{c-1}{\beta} - d.
\end{aligned}$$

In addition

$$\begin{pmatrix} \sigma_{\alpha\alpha} & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_{\beta\beta} \end{pmatrix} = \begin{pmatrix} -l_{\alpha\alpha} & -l_{\alpha\beta} \\ -l_{\alpha\beta} & -l_{\beta\beta} \end{pmatrix}^{-1}.$$

If $g(\alpha,\beta)=\alpha$, we obtain $g_\alpha = 1$ and $g_{\alpha\alpha} = g_\beta = g_{\beta\beta} = g_{\alpha\beta} = 0$. Thus the Bayes estimator using Lindley's approximation is given by

$$\begin{aligned}\hat{\alpha} \;=\; & \hat{\alpha}_{MLE} + \hat{\rho}_\alpha\hat{\sigma}_{\alpha\alpha} + \hat{\rho}_\beta\hat{\sigma}_{\beta\alpha} + \frac{1}{2}\Big[\hat{\sigma}_{\alpha\alpha}(\hat{l}_{\alpha\alpha\alpha}\hat{\sigma}_{\alpha\alpha} + \hat{l}_{\alpha\alpha\beta}\hat{\sigma}_{\alpha\beta} + \hat{l}_{\alpha\alpha\beta}\hat{\sigma}_{\beta\alpha} + \hat{l}_{\alpha\beta\beta}\hat{\sigma}_{\beta\beta}) \\ & + \; (\hat{\sigma}_{\beta\alpha})(\hat{l}_{\alpha\alpha\beta}\hat{\sigma}_{\alpha\alpha} + \hat{l}_{\alpha\beta\beta}\hat{\sigma}_{\alpha\beta} + \hat{l}_{\alpha\beta\beta}\sigma\hat{}_{\beta\alpha} + \hat{l}_{\beta\beta\beta}\hat{\sigma}_{\beta\beta})\Big].\end{aligned}$$

If $g(\alpha,\beta)=\beta$, we obtain $g_\beta = 1$ and $g_{\alpha\alpha} = g_\alpha = g_{\beta\beta} = g_{\alpha\beta} = 0$. Then the Bayes estimates of β is given by

$$\begin{aligned}\hat{\beta} \;=\; & \hat{\beta}_{MLE} + \hat{\rho}_\alpha\hat{\sigma}_{\beta\alpha} + \hat{\rho}_\beta\hat{\sigma}_{\beta\beta} + \frac{1}{2}\Big[\hat{\sigma}_{\alpha\beta}(\hat{l}_{\alpha\alpha\alpha}\hat{\sigma}_{\alpha\alpha} + \hat{l}_{\alpha\alpha\beta}\hat{\sigma}_{\alpha\beta} + \hat{l}_{\alpha\alpha\beta}\hat{\sigma}_{\beta\alpha} + \hat{l}_{\alpha\beta\beta}\hat{\sigma}_{\beta\beta}) \\ & + \; (\hat{\sigma}_{\beta\beta})(\hat{l}_{\alpha\alpha\beta}\hat{\sigma}_{\alpha\alpha} + \hat{l}_{\alpha\beta\beta}\hat{\sigma}_{\alpha\beta} + \hat{l}_{\alpha\beta\beta}\hat{\sigma}_{\beta\alpha} + \hat{l}_{\beta\beta\beta}\hat{\sigma}_{\beta\beta})\Big].\end{aligned}$$

4.2. Tierney–Kadane approximation

In this subsection, we utilize another approximation of the integral (4.2) to compute the Bayes estimators. Using Laplace transformation, Tierney and Kadane [33] proposed an alternative method to approximate the ratio of integrals. The advantage of using Tierney–Kadane method is that it requires only the first and the second derivatives of the posterior density. The posterior expectation of a $g(\alpha,\beta)$ can be written as

$$E(g(\alpha,\beta|\tilde{\tilde{x}})) = \frac{\int_0^\infty\int_0^\infty e^{nH^*(\alpha,\beta)}d\alpha d\beta}{\int_0^\infty\int_0^\infty e^{nH(\alpha,\beta)}d\alpha d\beta}, \tag{4.5}$$

where

$$\begin{aligned}H(\alpha,\beta) \;&=\; \frac{1}{n}\Big[(a-1)\log(\alpha) - b\alpha + (c-1)\log(\beta) - d\beta + l(\alpha,\beta|\tilde{\mathbf{x}})\Big], \\ H^*(\alpha,\beta) \;&=\; H(\alpha,\beta) + \frac{1}{n}\log(g(\alpha,\beta)).\end{aligned}$$

Then the integral given in Equation (4.5) can be approximated by

$$\hat{g}(\alpha,\beta) = \Big(\frac{\det\sum^*}{\det\sum}\Big)^{\frac{1}{2}}\exp\{n[H^*(\bar{\alpha}^*,\bar{\beta}^*) - H(\bar{\alpha},\bar{\beta})]\}, \tag{4.6}$$

where $(\bar{\alpha}^*,\bar{\beta}^*)$ and $(\bar{\alpha},\bar{\beta})$ maximize H^* and H, respectively, $\sum^*$ and $\sum$ are the negatives of the inverse Hessian matrix of H^* and H evaluated at $(\bar{\alpha}^*,\bar{\beta}^*)$ and $(\bar{\alpha},\bar{\beta})$, respectively. Therefore $(\bar{\alpha},\bar{\beta})$ can be obtained by solving the following two equations

$$\begin{aligned}H_\alpha \;&=\; \frac{\partial H(\alpha,\beta)}{\partial\alpha} \;=\; \frac{a-1}{\alpha} - b + l_\alpha(\alpha,\beta|\tilde{\mathbf{x}}) \;=\; 0, \\ H_\beta \;&=\; \frac{\partial H(\alpha,\beta)}{\partial\beta} \;=\; \frac{c-1}{\beta} - d + l_\beta(\alpha,\beta|\tilde{\mathbf{x}}) \;=\; 0,\end{aligned}$$

and from the second derivatives of H(α, β), the determinant of the negative of the inverse Hessian of H (α, β) at $(\bar{\alpha}, \bar{\beta})$ is given by

$$\det \sum = \left(\bar{H}_{\alpha\alpha}\bar{H}_{\beta\beta} - \bar{H}^2_{\alpha\beta}\right)^{-1},$$

where

$$\begin{aligned}
\bar{H}_{\alpha\alpha} &\equiv \frac{\partial \bar{H}_{\alpha}}{\partial \alpha} = -\frac{a-1}{\bar{\alpha}^2} + l_{\alpha\alpha}(\bar{\alpha}, \bar{\beta}|\tilde{\mathbf{x}}),\\
\bar{H}_{\beta\beta} &\equiv \frac{\partial \bar{H}_{\beta}}{\partial \beta} = -\frac{a-1}{\bar{\beta}^2} + l_{\beta\beta}(\bar{\alpha}, \bar{\beta}|\tilde{\mathbf{x}}),\\
\bar{H}_{\alpha\beta} &\equiv \frac{\partial \bar{H}_{\alpha}}{\partial \beta} = l_{\alpha\beta}(\bar{\alpha}, \bar{\beta}|\tilde{\mathbf{x}}).
\end{aligned}$$

Similarly, for the function $H^*(\alpha, \beta)$, the determinant of the negative of the inverse Hessian of $H^*(\alpha, \beta)$ evaluated at $(\bar{\alpha^*}, \bar{\beta^*})$ is given by

$$\det \sum{}^{*} = (\bar{H}^*_{\alpha\alpha}\bar{H}^*_{\beta\beta} - \bar{H}^{*2}_{\alpha\beta})^{-1}.$$

For g$(\alpha, \beta) = \alpha$, we get

$$H^*_{\alpha}(\alpha, \beta) = H(\alpha, \beta) + \frac{1}{n}\log(\alpha)$$

and consequently, we have

$$\begin{aligned}
H^*_{\alpha,\alpha} &= \frac{\partial H^*(\alpha, \beta)}{\partial \alpha} = H_{\alpha} + \frac{1}{n\alpha},\\
H^*_{\alpha,\beta} &= \frac{\partial H^*(\alpha, \beta)}{\partial \beta} = H_{\beta},\\
H^*_{\alpha,\alpha\beta} &= \frac{\partial H^*(\alpha, \beta)}{\partial \alpha\beta} = H_{\alpha\beta},\\
H^*_{\alpha,\alpha\alpha} &= \frac{\partial H^*_1}{\partial \alpha} = H_{\alpha\alpha} - \frac{1}{n\alpha^2},\\
H^*_{\alpha,\beta\beta} &= \frac{\partial H^*_2}{\partial \beta} = H_{\beta\beta}.
\end{aligned}$$

For $g(\alpha, \beta) = \beta$, we have

$$H^*_{\beta}(\alpha, \beta) = \frac{1}{n}\log(\beta) + H(\alpha, \beta)$$

and

$$\begin{aligned}
H^*_{\beta,\alpha} &= \frac{\partial H^*(\alpha, \beta)}{\partial \alpha} = H_{\alpha},\\
H^*_{\beta,\beta} &= \frac{\partial H^*(\alpha, \beta)}{\partial \beta} = H_{\beta} + \frac{1}{n\beta},\\
H^*_{\beta,\alpha\beta} &= \frac{\partial H^*(\alpha, \beta)}{\partial \alpha\beta} = H_{\alpha\beta},\\
H^*_{\beta,\alpha\alpha} &= \frac{\partial D^*_1}{\partial \alpha} = H_{\alpha\alpha},\\
H^*_{\beta,\alpha\alpha} &= \frac{\partial D^*_2}{\partial \beta} = H_{\beta\beta} - \frac{1}{n\beta^2}.
\end{aligned}$$

Finally, substituting the above expressions in (4.6), we obtain the Bayes estimates of α and β.

4.3. Highest posterior density estimation

The highest posterior density estimation is another popular method used to compute the Bayes estimates. The highest posterior density (HPD) estimate represents the mode of the posterior density. The Bayes estimates using HPD method can be obtained by solving the equations

$$\frac{\partial \pi(\alpha,\beta|\tilde{\mathbf{x}})}{\partial \alpha} = \frac{n+a-1}{\alpha} - b + \frac{\int A_\alpha(x)\mu_{\tilde{x}_i}(x)dx}{\int A(x)\mu_{\tilde{x}_i}(x)dx} = 0, \tag{4.7}$$

$$\frac{\partial \pi(\alpha,\beta|\tilde{\mathbf{x}})}{\partial \beta} = \frac{n+c-1}{\beta} - d + \frac{\int A_\beta(x)\mu_{\tilde{x}_i}(x)dx}{\int A(x)\mu_{\tilde{x}_i}(x)dx} = 0. \tag{4.8}$$

It can be seen that, the solutions of the above two equation cannot be obtained explicitly and, similar to the maximum likelihood method, numerical methods like Newton–Raphson can be used to solve them.

5. SIMULATION EXPERIMENTS

In this section, we conduct Monte-Carlo simulation experiments to show how the various approaches work with different sample sizes. The performance of the proposed approaches was compared on the basis of their expected biases, root mean square error, average of standard errors and of 95% confidence intervals. The true values of the parameters (α,β) are assumed to be $(1.25, 1.5), (1.5, 0.5)$ and $(0.5, 0.75)$, respectively. The sample sizes are chosen as $n = 25, 50$ and 100 to represent small, moderate and large samples, respectively. Each observation from Burr type XII, x_i, was then fuzzified with the corresponding membership function $\mu_{\tilde{x}_i}(x)$, where

$$\mu_{\tilde{x}_i}(x) = \begin{cases} \dfrac{x-(x_i-a_i)}{a_i}, & \text{if } x_i - a_i \le x \le x_i, \\ \dfrac{(x_i+a_i)-x}{a_i}, & \text{if } x_i \le x \le x_i + a_i, \\ 0, & \text{otherwise,} \end{cases} \tag{5.1}$$

and $a_i = 0.05x_i$ (see, for example, Pak and Chatrabgoun ([24]), Pak *et al.* ([27]), Chaturvedi ([6])). That is the observer is unable to provide exact value of observation and an interval of plausible values $[x_i - a_i, x_i + a_i]$ is provided. For example the triangular fuzzy number $(0.1805, 0.1995)$ represents the observed value 0.19 i.e. the interval of plausible values of 0.19 is $[0.1805, 0.1995]$. Then, we compute the MLEs of α and β for the fuzzy sample via Newton–Raphson (NR) and Expectation-Maximization (EM) algorithm. The process is replicated 1000 times. In each replication, we compute the average of biases (Bias), sample standard error (SSE) and the root mean squared error (RMSE) using the expressions

$$\text{Bias}(\theta) = \frac{1}{k}\sum_{i=1}^{k}(\theta_i - \theta_0),$$

$$\mathrm{SSE}(\theta) = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\theta_i - \bar{\theta})^2}$$

and

$$\mathrm{RMSE}(\theta) = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\theta_i - \theta_0)^2},$$

where θ represents α or β, θ_0 is the true value of θ, $\bar{\theta}$ is the mean of the estimates of θ and k is the number of replications. Moreover, to compute the estimated standard error (ESE) for the MLEs, we use the observed information matrix given in (3.11). Approximated 95% confidence intervals for the MLE are constructed using the observed information matrix. Moreover, in each iteration, we compute the Bayes estimators using Lindley's approximation, Tierney–Kadane approximation and highest posterior density (HP) methods. At the end, we compute the averages of the absolute biases, sample standard deviation, estimated standard deviation, root mean squared error and 95% confidence intervals. For computing Bayes estimators, we consider gamma priors for α and β with hyperparameters (a, b) and (c, d) , respectively. To make the comparison meaningful, it is assumed that the priors are non-informative $a = b = c = d = 0$ but these priors are improper priors hence we have tried $a = b = c = d = 0.001$ to get proper priors. However, these results are same as those obtained for improper priors. The simulation results of the MLEs and Bayes estimators are reported in Tables 1–2. We have utilized R-4.0.3 software to compute the proposed estimators. The stopping criteria for the algorithms is based on the sum of the absolute differences between two consecutive values of parameters estimates less than 10^{-4}.

From Table 1, we observe that the biases for all estimators, in general, are reasonably small which indicate that the estimated values are close to the true parameter values. As expected, the biases of all estimators become better when the sample size increases. The values of sample standard error (SSE) of the MLEs are approximately close to estimated standard error (ESE) for all the cases and hence the estimated standard error can be used to estimate the standard error of the estimators. In addition, the Bias, SSE, ESE, RMSE and the length of 95% confidence inetrvals of all MLEs are decreasing when sample sizes increasing for all the cases. The estimated coverage probabilities of 95% confidence intervals (CP) are very close to the nominal level for all the cases. Hence, the performance of the MLEs are satisfactory in terms of the biases, standard errors and coverage probabilities of the estimates. Moreover, the Bias of the computed MLEs estimators using EM algorithm for most of the cases are slightly higher than that of the MLEs computed using EM-algorithm. In addition, the central processing time CPU required for NR per iteration is shorter than that of EM algorithm. Figure 1 demonstrates the histograms for the MLEs of α and β when $n = 100$ for the three sets of values. The histograms show approximately normal distribution of the MLEs of α and β.

From Table 2, the biases of the Bayesian estimates of all three methods are also reasonably small. It is clear that the Bias and RMSE are decreasing for increasing values of sample sizes. Moreover, the Bias and RMSE of the Bayes estimates obtained under highest posterior density (HP) are smaller than that of Lindley's method (LN) and Tierney–Kadane approximation (TK). Hence we recommend to use HP method for computing Bayes estimator. From the above results, we conclude that the estimation methods proposed in the article to compute the MLEs and Bayes estimators perform very well.

Table 1: Simulation results for MLEs of α and β.

n			Bias	RMSE	ESE	SSE	95% CI	Length	CP
25	$\alpha = 1.25$	NR	0.068	0.258	0.218	0.249	(0.95, 1.82)	0.87	93.6
		EM	0.069	0.255	0.218	0.245	(0.93, 1.81)	0.88	93.8
	$\beta = 1.50$	NR	0.072	0.353	0.321	0.345	(1.05, 2.35)	1.30	94.1
		EM	0.075	0.350	0.320	0.342	(1.01, 2.30)	1.29	94.8
50	$\alpha = 1.25$	NR	0.029	0.162	0.150	0.159	(1.02, 1.61)	0.59	94.5
		EM	0.031	0.160	0.152	0.160	(1.04, 1.60)	0.56	94.6
	$\beta = 1.50$	NR	0.027	0.226	0.220	0.224	(1.15, 2.03)	0.88	95.0
		EM	0.029	0.225	0.218	0.222	(1.11, 1.98)	0.87	95.0
100	$\alpha = 1.25$	NR	0.013	0.106	0.104	0.105	(1.07, 1.48)	0.41	94.9
		EM	0.015	0.108	0.103	0.105	(1.08, 1.46)	0.38	94.8
	$\beta = 1.50$	NR	0.012	0.153	0.154	0.153	(1.24, 1.85)	0.61	95.0
		EM	0.012	0.155	0.156	0.155	(1.24, 1.83)	0.59	95.1
25	$\alpha = 1.50$	NR	0.155	0.511	0.415	0.484	(1.01, 2.72)	1.71	93.8
		EM	0.157	0.515	0.413	0.486	(1.02, 2.72)	1.70	93.6
	$\beta = 0.50$	NR	0.007	0.143	0.135	0.146	(0.29, 0.87)	0.58	94.8
		EM	0.007	0.144	0.138	0.144	(0.30, 0.87)	0.57	94.3
50	$\alpha = 1.50$	NR	0.062	0.290	0.270	0.291	(1.12, 2.19)	1.07	94.7
		EM	0.067	0.297	0.269	0.290	(1.12, 2.19)	1.07	94.5
	$\beta = 0.50$	NR	0.002	0.098	0.098	0.097	(0.34, 0.73)	0.39	95.0
		EM	0.003	0.097	0.097	0.097	(0.34, 0.73)	0.39	95.0
100	$\alpha = 1.50$	NR	0.026	0.188	0.182	0.186	(1.21, 1.93)	0.72	95.2
		EM	0.027	0.187	0.180	0.185	(1.20, 1.93)	0.73	95.4
	$\beta = 0.50$	NR	0.002	0.069	0.069	0.070	(0.38, 0.66)	0.28	94.9
		EM	0.001	0.071	0.072	0.071	(0.38, 0.65)	0.27	94.9
25	$\alpha = 0.50$	NR	0.080	0.280	0.211	0.267	(0.51, 1.37)	0.86	93.4
		EM	0.082	0.270	0.210	0.265	(0.51, 1.38)	0.87	93.8
	$\beta = 0.75$	NR	0.006	0.140	0.140	0.143	(0.30, 0.86)	0.56	94.4
		EM	0.007	0.145	0.138	0.145	(0.30, 0.86)	0.56	94.2
50	$\alpha = 0.50$	NR	0.033	0.149	0.134	0.145	(0.56, 1.10)	0.54	94.5
		EM	0.034	0.150	0.132	0.143	(0.56, 1.10)	0.56	94.6
	$\beta = 0.75$	NR	0.002	0.096	0.095	0.095	(0.34, 0.73)	0.39	95.0
		EM	0.002	0.097	0.097	0.097	(0.34.0.73)	0.39	95.2
100	$\alpha = 0.50$	NR	0.013	0.094	0.096	0.093	(0.60, 0.97)	0.37	95.2
		EM	0.013	0.092	0.091	0.094	(0.60, 0.96)	0.36	95.5
	$\beta = 0.75$	NR	0.003	0.069	0.070	0.069	(0.38, 0.66)	0.28	94.9
		EM	0.002	0.070	0.069	0.070	(0.38, 0.65)	0.27	94.6

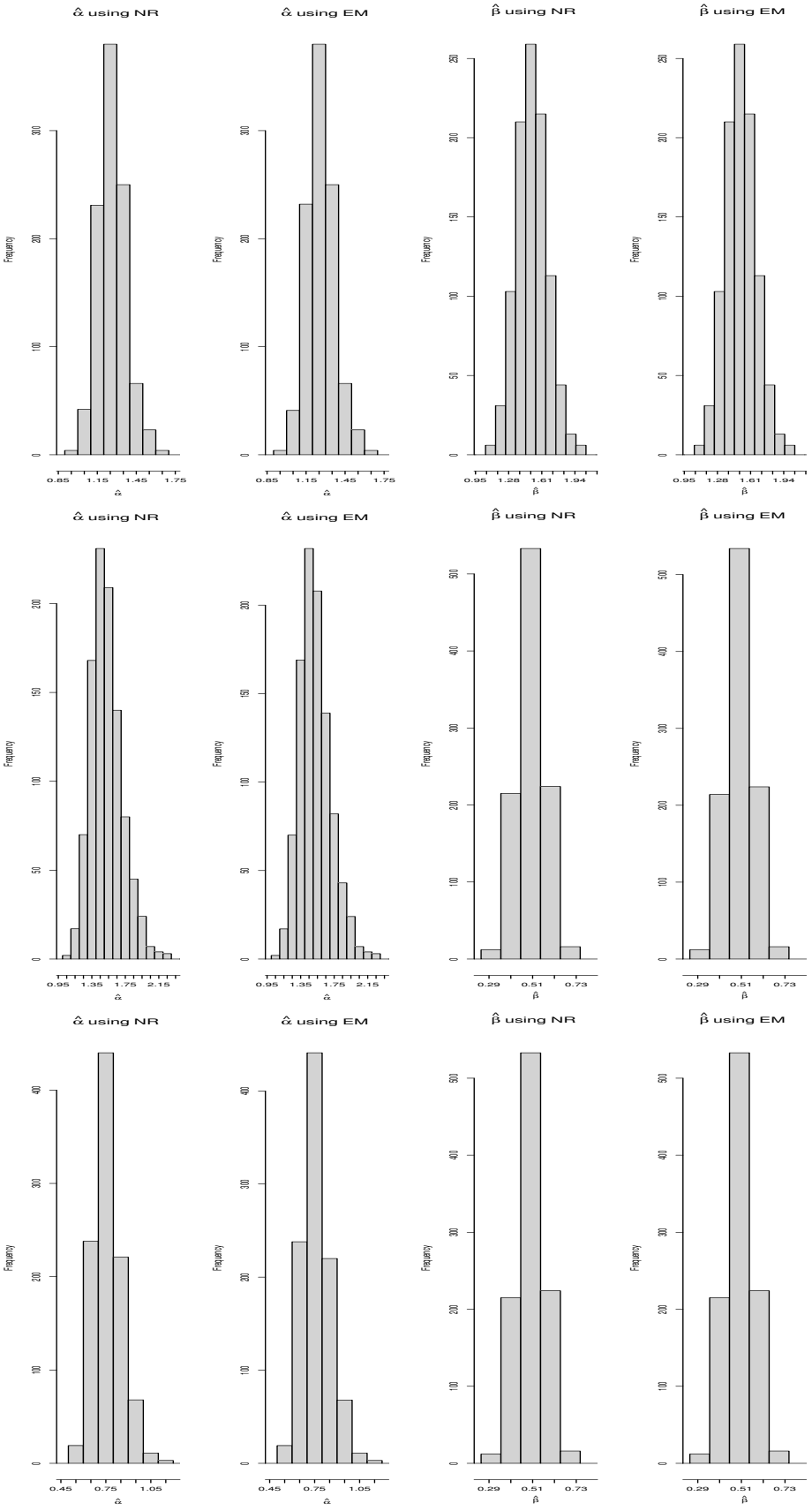

Figure 1: Histograms of the estimated values of the MLEs, $\hat{\alpha}$ and $\hat{\beta}$, for $n = 100$. The first line for ($\alpha = 1.25$, $\beta = 1.5$), the second line for ($\alpha = 1.5$, $\beta = 0.5$) and the third line for ($\alpha = 0.5$, $\beta = 0.75$).

Table 2: Simulation results for Bayesian estimates of α and β.

n		LN	TK	HPD	LN	TK	HPD
		$\alpha = 1.25$			$\beta = 1.5$		
25	Bias	0.068	0.065	0.040	0.068	0.067	0.014
	RMSE	0.259	0.264	0.247	0.348	0.345	0.328
50	Bias	0.029	0.029	0.016	0.025	0.022	0.003
	RMSE	0.162	0.160	0.158	0.224	0.220	0.218
100	Bias	0.013	0.012	0.006	0.011	0.009	−0.001
	RMSE	0.106	0.105	0.105	0.153	0.152	0.151
		$\alpha = 1.5$			$\beta = 0.5$		
25	Bias	0.191	0.199	0.140	0.015	0.015	−0.009
	RMSE	0.529	0.556	0.504	0.141	0.143	0.140
50	Bias	0.084	0.080	0.058	0.006	0.004	−0.006
	RMSE	0.309	0.309	0.295	0.096	0.096	0.095
100	Bias	0.030	0.034	0.022	0.007	0.004	−0.002
	RMSE	0.191	0.190	0.187	0.069	0.071	0.068
		$\alpha = 0.5$			$\beta = 0.75$		
25	Bias	0.103	0.105	0.075	0.014	0.015	−0.009
	RMSE	0.298	0.306	0.276	0.142	0.144	0.141
50	Bias	0.042	0.040	0.029	0.005	0.006	−0.006
	RMSE	0.152	0.154	0.147	0.096	0.092	0.095
100	Bias	0.018	0.017	0.011	0.004	0.002	−0.002
	RMSE	0.095	0.092	0.093	0.069	0.070	0.068

6. APPLICATION EXAMPLES

In this section, we analyze three real data sets to explain how the proposed approaches can be applied in real data analysis. We are assuming that each observation in any of these datasets, x_i, is reported as a fuzzy numbers with membership function given in (5.1). For computing Bayes estimators in this section, we assume gamma priors with hyperparameters $a = b = c = d = 0.001$. This choice of hyperparameters will make the priors proper. However, we have tried to consider different values of hyperparameters, for example, we have considered the cases $a = b = c = d = 1$, and $a = 2$, $b = 1$, $c = 2$, $d = 1$ and the results are not much different than that we have obtained from that case, and are not reported due to the space.

Example 1. The first data set was considered and analyzed by Zimmer *et al.* ([37]) and Lio *et al.* ([19]). The dataset contains the 19 times in minutes to oil breakdown of an insulating fluid under high test voltage (34 kV). The data set is listed as follows: 0.19, 0.78, 0.96, 0.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50, 7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71, 72.89. Lio *et al.* ([19]) showed that the two-parameters Burr type XII fits the data set very well. The MLEs of (α, β) using Newton–Raphson method are $(1.440, 0.354)$ with standard errors $(0.435, 0.126)$ and 95% confidence intervals $(0.588, 2.292)$ and $(0.106, 0.601)$, respectively, and MLEs using EM algorithm are $(1.436, 0.357)$ with estimated standard error $(0.431, 0.127)$ and 95% confidence intervals (0.590, 2.281) and $(0.108, 0.606)$. In addition, the Bayes estimates of (α, β) are (1.427, 0.338) using Lindley's approximation, $(1.507, 0.364)$ using Tierney–Kadane approximation and $(1.427, 0.338)$ using highest posterior density method.

Example 2. Lawless ([16]) reported the time between failure of air conditioning equipment in a particular type of aircraft. These observations are:

0.500, 0.875, 1.083, 1.125, 1.208, 1.208, 2.00, 2.375,
2.458, 2.917, 3.083, 6.375, 13.583, 16.083, 20.917.

Kayal *et al.* ([14]) concluded that Burr type XII model fits the data set quite good. The MLEs of (α, β) using Newton–Raphson method are $(3.571, 0.275)$ with standard errors $(1.488, 0.127)$ and 95% confidence intervals $(0.654, 6.487)$ and $(0.026, 0.524)$, respectively, and MLEs using EM algorithm are $(3.500, 0.284)$ with estimated standard error $(1.434, 0.129)$ and 95% confidence intervals $(0.690, 6.311)$ and $(0.031, 0.537)$, respectively. In addition, the Bayes estimates of (α, β) are $(3.519, 0.260)$ using Lindley's approximation, $(3.921, 0.289)$ using Tierney–Kadane approximation and $(3.519, 0.260)$ using highest posterior density method.

Example 3. In this example, we analyze a dataset that represents the survival time of animals observed due to different dosage of poison administered (see Box and Cox ([4])). The observations are listed as:

0.18, 0.21, 0.22, 0.22, 0.23, 0.23, 0.23, 0.24, 0.25, 0.29, 0.29, 0.30,
0.30, 0.31, 0.31, 0.31, 0.33, 0.35, 0.36, 0.36, 0.37, 0.38, 0.38, 0.40,
0.40, 0.43, 0.43, 0.44, 0.45, 0.45, 0.45, 0.46, 0.49, 0.56, 0.61, 0.62,
0.63, 0.66, 0.71, 0.71, 0.72, 0.76, 0.82, 0.88, 0.92, 1.02, 1.10, 1.24.

Kayal *et al.* ([14]) analyzed the above data and they concluded that the data might have come from a two-parameter Burr type XII distribution. The MLEs of (α, β) using Newton–Raphson method are $(2.346, 4.938)$ with standard errors $(0.231, 0.822)$ and 95% confidence intervals $(1.893, 2.798)$ and $(1.887, 2.785)$, respectively, and MLEs using EM algorithm are $(2.336, 5.075)$ with estimated standard error $(0.229, 0.850)$ and 95% confidence intervals $(3.326, 6.550)$ and $(3.408, 6.742)$, respectively. In addition, the Bayes estimates of (α, β) are (2.373, 4.928) using Lindley's approximation, $(2.338, 4.923)$ using Tierney–Kadane approximation and $(2.304, 4.761)$ using highest posterior density method.

7. CONCLUSION

In this article, we have considered both classical and Bayesian analysis of fuzzy survival time observations when the lifetime of the items follows two-parameter Burr type XII distribution. The MLEs do not have explicit forms. Thus, Newton–Raphson and Expectation-Maximization algorithms have been used to compute the MLEs and both of them work quite well. The Bayes estimates under the squared error loss function also do not exist in explicit form. In this case, we have proposed to use Lindley's approximation, Tierney–Kadane approximation and highest posterior density method to compute the Bayes estimates when the two unknown parameters have independent gamma priors. However, we have considered gamma priors, but a more general prior, namely a prior which has the log-concave p.d.f. may be used, and the method can be easily incorporated in that case. Moreover, in Bayesian estimation, we proposed to use a very well-known symmetric loss function which is the squared-error loss function. However, we may extend the results of the paper by adopting other loss function like LINEX. Another direction for extension is to consider censored fuzzy observations like type II progressively censored fuzzy observations.

A. Proof of Theorem 3.1

Recall that, the log-likelihood function of α and β is given by

$$l(\alpha,\beta|\tilde{\mathbf{x}}) = n\log\alpha + n\log\beta + \sum_{i=1}^{n}\log\int A(x)\mu_{\tilde{x_i}}(x)dx,$$

where

$$A(x) = x^{\alpha-1}(1+x^{\alpha})^{-\beta-1}. \tag{A.1}$$

Observe that, for fixed $\beta > 0$, we have

$$\lim_{\alpha\to 0} l(\alpha,\beta|\tilde{\mathbf{x}}) = \lim_{\alpha\to\infty} l(\alpha,\beta|\tilde{\mathbf{x}}) = -\infty$$

and, for fixed $\alpha > 0$, we have

$$\lim_{\beta\to 0} l(\alpha,\beta|\tilde{\mathbf{x}}) = \lim_{\beta\to\infty} l(\alpha,\beta|\tilde{\mathbf{x}}) = -\infty.$$

We can see that

$$\frac{\partial^2\log(A(x))}{\partial\alpha^2} = -\frac{(\beta+1)(\log(x))^2x^{\alpha}}{(1+x^{\alpha})^2} < 0$$

for fixed $\beta > 0$, i.e., $A(x)$ is strictly log-concave in α for fixed $\beta > 0$. Similarly, we can prove that $A(x)$ is log-concave in β for fixed $\alpha > 0$. By Prekopa–Leindler inequality (see Gardner [10]) we obtain that $\int A(x)\mu_{\tilde{x_i}}(x)dx$ is strictly log-concave in α (or β) for fixed $\beta > 0$ (or $\alpha > 0$). Therefore, for fixed α (or β), $l(\alpha,\beta|\tilde{\mathbf{x}})$ is strictly concave and unimodal function with respect to β (or α). Moreover,

$$\lim_{\substack{\alpha\to 0\\ \beta\to 0}} l(\alpha,\beta|\tilde{\mathbf{x}}) = \lim_{\substack{\alpha\to 0\\ \beta\to\infty}} l(\alpha,\beta|\tilde{\mathbf{x}}) = \lim_{\substack{\alpha\to\infty\\ \beta\to 0}} l(\alpha,\beta|\tilde{\mathbf{x}}) = \lim_{\substack{\alpha\to\infty\\ \beta\to\infty}} l(\alpha,\beta|\tilde{\mathbf{x}}) = -\infty.$$

The rest of the proof is the same as that of Dey *et al.* ([9]).

ACKNOWLEDGMENTS

The authors are grateful to the anonymous referees and the associate editor for making many helpful comments and suggestions on an earlier version of this manuscript which resulted in this improved version.

REFERENCES

[1] AL-Baldawi, T.H.K.; Rasheed, H.A. and Jaseim, S.H. (2015). Using generalized square loss function to estimate the shape parameter of the Burr type XII distribution, *International Journal of Advanced Research*, **3**(5), 393–398.

[2] Basharat, H.; Mustafa, S.; Mahmood, S. and Jun, Y.B. (2019). Inference for the linear combination of two independent exponential random variables based on fuzzy data, *Hacettepe Journal of Mathematics and Statistics*, **48**(6), 1859–1869.

[3] Belaghi, R.A.; Noori Asl, M.N. and Singh, S. (2017). On estimating the parameters of the Burr XII model under progressive type-I interval censoring, *Journal of Statistical Computation and Simulation*, **87**(16), 3132–3151.

[4] Box, G.E. and Cox, D.R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society: Series B (Methodological)*, **26**(2), 211–243.

[5] Burr, I.W. (1942). Cumulative frequency functions, *The Annals of Mathematical Statistics*, **13**(2), 215–232.

[6] Chaturvedi, A.; Singh, S.K. and Singh, U. (2018). Statistical inferences of type II progressively hybrid censored fuzzy data with Rayleigh distribution, *Austrian Journal of Statistics*, **47**(3), 40–62.

[7] 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 (Methodological)*, **39**(1), 1–22.

[8] Denœux, T. (2011). Maximum likelihood estimation from fuzzy data using the EM algorithm, *Fuzzy Sets and Systems*, **183**(1), 72–91.

[9] 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.

[10] Gardner, R. (2002). The Brunn-Minkowski inequality, *Bulletin of the American Mathematical Society*, **39**(3), 355–405.

[11] Ghitany, M. and Al-Awadhi, S. (2002). Maximum likelihood estimation of Burr XII distribution parameters under random censoring, *Journal of Applied Statistics*, **29**(7), 955–965.

[12] Hakim, A.R.; Novita, M. and Fithriani, I. (2019). Using Jeffrey prior information to estimate the shape parameter of Burr distribution, *Journal of Physics: Conference Series, IOP Publishing*, **1218**(1).

[13] Jaheen, Z.F. and Okasha, H.M. (2011). E-Bayesian estimation for the Burr type XII model based on type II censoring, *Applied Mathematical Modelling*, **35**(10), 4730–4737.

[14] Kayal, T.; Tripathi, Y.M.; Rastogi, M.K. and Asgharzadeh, A. (2017). Inference for Burr XII distribution under type-I progressive hybrid censoring, *Communications in Statistics – Simulation and Computation*, **46**(9), 7447–7465.

[15] Khoolenjani, N.B. and Shahsanaie, F. (2016). Estimating the parameter of exponential distribution under type II censoring from fuzzy data, *Journal of Statistical Theory and Applications*, **15**(2), 181–195.

[16] Lawless, J.F. (2011). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons.

[17] Li, X.; Shi, Y.; Wei, J. and Chai, J. (2007). Empirical Bayes estimators of reliability performances using LINEX loss under progressively type II censored samples, *Mathematics and Computers in Simulation*, **73**(5), 320–326.

[18] Lindley, D.V. (1980). Approximate Bayesian methods, *Trabajos de Estadística y de Investigación Operativa*, **31**(1), 223–245.

[19] Lio, Y.; Tsai, T.R. and Wu, S.J. (2010). Acceptance sampling plans from truncated life tests based on the Burr type XII percentiles, *Journal of the Chinese institute of Industrial Engineers*, **27**(4), 270–280.

[20] Makhdoom, I.; Nasiri, P. and Pak, A. (2016). Estimating the parameter of exponential distribution under type II censoring from fuzzy data, *Journal of Modern Applied Statistical Methods*, **15**(2) 495–509.

[21] Moore, D. and Papadopoulos, A.S. (2000). The Burr type XII distribution as a failure model under various loss functions, *Microelectronics Reliability*, **40**(12), 2117–2122.

[22] Mousa, M.A. and Jaheen, Z. (2002). Statistical inference for the Burr model based on progressively censored data, *Computers & Mathematics with Applications*, **43**(10), 1441–1449.

[23] Pak, A. (2017). Statistical inference for the parameter of Lindley distribution based on fuzzy data, *Brazilian Journal of Probability and Statistics*, **31**(3), 502–515.

[24] Pak, A. and Chatrabgoun, O. (2016). Inference for exponential parameter under progressive type II censoring from imprecise lifetime, *Electronic Journal of Applied Statistical Analysis*, **9**(1), 227–245.

[25] Pak, A.; Khoolenjani, N.B.; Alamatsaz, M.H. and Mahmoudi, M.R. (2020). Bayesian method for the generalized exponential model using fuzzy data, *International Journal of Fuzzy Systems*, **22**(4), 1243-1260.

[26] Pak, A. and Mahmoudi, M.R. (2018). Estimating the parameters of Lomax distribution from imprecise information, *Journal of Statistical Theory and Applications*, **17**(1), 122–135.

[27] Pak, A.; Parham, G.A. and Saraj, M. (2013a). Inference for the Weibull distribution based on fuzzy data, *Revista Colombiana de Estadistica*, **36**(2), 337–356.

[28] Pak, A.; Parham, G.A. and Saraj, M. (2013b). On estimation of Rayleigh scale parameter under doubly type II censoring from imprecise data, *Journal of Data Science*, **11**(2), 305–322.

[29] Panahi, H. and Asadi, S. (2011). Analysis of the type II hybrid censored Burr type XII distribution under LINEX loss function, *Applied Mathematical Sciences*, **5**, 3929–3942.

[30] Pradhan, B. and Kundu, D. (2014). Analysis of interval-censored data with Weibull lifetime distribution, *Sankhya B*, **76**(1), 120–139.

[31] Rao, G.S.; Aslam, M. and Kundu, D. (2015). Burr-XII distribution parametric estimation and estimation of reliability of multicomponent stress-strength, *Communications in Statistics – Theory and Methods*, **44**(23), 4953–4961.

[32] Shafiq, M. (2017). Classical and Baysian inference of Pareto distribution and fuzzy life times, *Pakistan Journal of Statistics*, **33**(1), 15–25.

[33] Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, **81**(393), 82–86.

[34] Wahed, A.S. (2006). Bayesian inference using Burr model under asymmetric loss function: an application to carcinoma survival data, *Journal of Statistical Research*, **40**(1), 45–57.

[35] Wang, F.; Keats, J.B. and Zimmer, W.J. (1996). Maximum likelihood estimation of the Burr XII parameters with censored and uncensored data, *Microelectronics Reliability*, **36**(3), 359–362.

[36] Zadeh, L.A. (1968). Probability measures of fuzzy events, *Journal of Mathematical Analysis and Applications*, **23**(2), 421–427.

[37] Zimmer, W.J.; Keats, J.B. and Wang, F. (1998). The Burr XII distribution in reliability analysis, *Journal of Quality Technology*, **30**(4), 386–394.

REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 425–445
https://doi.org/10.57805/revstat.v21i3.414

Likelihood-Based Prediction of Future Weibull Record Values

Authors: Grigoriy Volovskiy ✉
– Institute of Statistics, RWTH Aachen University,
D-52062 Aachen, Germany
volovskiy@isw.rwth-aachen.de

Udo Kamps
– Institute of Statistics, RWTH Aachen University,
D-52062 Aachen, Germany
kamps@isw.rwth-aachen.de

Received: December 2020 Revised: January 2022 Accepted: January 2022

Abstract:

- Point prediction of future record values from a sequence of independent and identically distributed two-parameter Weibull random variables using the maximum likelihood method is considered. Two possible likelihood functions for prediction, the predictive and the observed predictive likelihood functions, are considered and the associated predictors are derived. Mean squared error and Pitman closeness criterion are used for comparing the prediction procedures.

Keywords:

- *point prediction; maximum observed likelihood predictor; maximum likelihood predictor; upper record values; Pitman closeness; Weibull distribution*

AMS Subject Classification:

- 62F99, 62M20.

✉ Corresponding author.

1. INTRODUCTION

In many practical applications such as materials testing, meteorology and hydrology, only record data is available for statistical analysis. Then, for a sequence of successive, increasing record values, an appropriate model are upper record values, first studied in [5]. Suppose that $X_1, X_2, ...$ is an infinite sequence of independent and identically distributed (i.i.d.) continuous random variables with cumulative distribution function (cdf) F. An observation is called an *(upper) record value* provided it is greater than all previously observed values. More specifically, defining the record times as

$$L(1) = 1, \quad L(n+1) = \min\{j > L(n) \mid X_j > X_{L(n)}\}, \quad n \in \mathbb{N},$$

the sequence $(R_n)_{n\in\mathbb{N}} = (X_{L(n)})_{n\in\mathbb{N}}$ is referred to as the sequence of (upper) record values based on $(X_n)_{n\in\mathbb{N}}$; see [1], [15]. The structure of record values also appears in the context of minimal repair of a system, and, under mild conditions, the epoch times of a non-homogeneous Poisson process and upper record values are equal in distribution; see [11]. If not the record values themselves, but the successively k-th largest values $(R_n^{(k)})_{n\in\mathbb{N}}$, $k \in \mathbb{N}$, in an i.i.d. sequence of random variables are of interest, the appropriate description is provided by the model of k-th *record values* introduced in [9].

We consider the problem of providing a prediction value for the occurrence of a future Weibull record value R_s based on the first r, $r < s$, (observed) Weibull record values $R_\star = (R_1, ..., R_r)$. In addition to the modeling of repairable systems mentioned above, the Weibull record values model has been used in the literature to model reliability growth (see [7]) and software reliability (see [13]). The point prediction problem for Weibull record values has recently been studied in [16], where, in particular, the maximum likelihood predictor of R_s based on $R_\star$ was derived. In fact, predictive analysis of Weibull record values dates as back as [14] and [10], where exact prediction intervals for R_s were constructed. For Bayesian predictive analysis of Weibull record values, the reader is referred to, e.g., [3], [21], [22]. For statistical inference based on record values from Weibull distributions and application, we also refer to [23] and [20].

The *maximum likelihood prediction* procedure is frequently examined in the literature and commonly applied in the context of an ordered data model such as the model of upper record values; see [12]. The maximum likelihood prediction procedure derives a predictor of a r.v. Y based on a possibly p-dimensional random vector X with joint pdf $f_\theta^{X,Y}$, $\theta \in \Theta$, by maximizing the predictive likelihood function L_{rv} of Y and θ given that $X = x$, which takes the form

$$L_{\mathrm{rv}}(y, \theta|x) = f_\theta^{X,Y}(x, y),$$

with respect to θ and y. The functions π_{MLP} and $\hat{\theta}_{\mathrm{ML}}$ are called, respectively, maximum likelihood predictor (MLP) of Y and predictive maximum likelihood estimator (PMLE) of θ, if, for any $x \in \mathbb{R}^p$,

$$L_{\mathrm{rv}}(\pi_{\mathrm{MLP}}(x), \hat{\theta}_{\mathrm{ML}}(x)) = \max_{(y,\theta)\in\mathbb{R}\times\Theta} L_{\mathrm{rv}}(y, \theta|x).$$

Recently, a new likelihood-based general-purpose prediction procedure, the so-called *maximum observed likelihood prediction* method has been introduced and studied in [19]; see also [18].

By means of this procedure, a predictor of Y based on X is obtained by maximizing the observed predictive likelihood function L_{obs} defined by

$$L_{\text{obs}}(y, \theta|x) = f_\theta^{X|Y}(x|y)$$

with respect to θ and y. Then, any functions π_{MOLP} and $\hat{\theta}_{\text{MOL}}$ are referred to, respectively, as maximum observed likelihood predictor (MOLP) of Y and predictive maximum observed likelihood estimator (PMOLE) of θ, provided that, for any $x \in \mathbb{R}^p$,

$$L_{\text{obs}}(\pi_{\text{MOLP}}(x), \hat{\theta}_{\text{MOL}}(x)) = \max_{(y,\theta)\in\mathbb{R}\times\Theta} L_{\text{obs}}(y, \theta|x).$$

If, in a general parametric family $\{F_\theta \mid \theta \in \Theta\}$ of continuous cdfs, the s-th record R_s is predicted based on $R_\star = (R_1, ..., R_r)$, then the maximum observed likelihood predictor is given by

$$\pi_{\text{MOLP}}^{(s)} = F^{-1}_{\hat{\theta}(R_\star)}\Big(1 - (1 - F_{\hat{\theta}(R_\star)}(R_r))^{\frac{s-1}{r}}\Big), \tag{1.1}$$

where the function $\hat{\theta}$ is such that

$$\Psi(\hat{\theta}(r_\star), r_\star) = \max_{\substack{\theta\in\Theta:\\ (\theta,r_\star)\in\mathcal{Z}_r}} \Psi(\theta, r_\star); \tag{1.2}$$

see [19, Theorem 3.3], [18, Theorem 5.3]. In equation (1.2), the function Ψ is given by

$$\Psi(\theta, r_\star) = \prod_{i=1}^{r} \frac{f_\theta(r_i)/(1 - F_\theta(r_i))}{\ln(1 - F_\theta(r_r))}, \quad (\theta, r_\star) \in \mathcal{Z}_r, \tag{1.3}$$

with $\mathcal{Z}_r = \{(\theta, r_1, ..., r_r) \in \Theta \times \mathbb{R}^r_< \mid (r_1, ..., r_r) \in (\alpha(F_\theta), \omega(F_\theta))^r_<\}$, where, for an interval $I \subseteq \mathbb{R}$ and $n \in \mathbb{N}$, $I^n_< = \{(x_1, ..., x_n) \in I^n \mid x_1 < x_2 < \cdots < x_n\}$, and $\alpha(F)$ and $\omega(F)$ denote the left and right endpoints of the support of a cdf F.

In order to facilitate building some intuition for the difference between the predictive likelihood and the observed predictive likelihood function-based prediction procedures, let us slightly rewrite the associated likelihood functions. First, observe that the predictive likelihood function can be constructed by taking the product of the conditional density function of Y given X and the density function of X, that is

$$L_{\text{rv}}(y, \theta|x) = f_\theta^{Y|X}(y|x)\, f_\theta^X(x).$$

Thus, in maximizing the predictive likelihood function the information on the variability in Y as described by the conditional density function $f_\theta^{Y|X}$ is reduced to the mode of the conditional density of Y given X yielding a prediction value, which, given the observed data, is the most probable value of Y under a model that best fits the observed data as well as the prediction value. In principal, any functional of the conditional density of Y given X could be used to derive a prediction value of Y but the choice of the mode has the appealing advantage of allowing to formally extend the maximum likelihood method from the parametric to the predictive inference setup. Next, we have that

$$L_{\text{obs}}(y, \theta|x) = f_\theta^X(x)\, \frac{f_\theta^{Y|X}(y|x)}{f_\theta^Y(y)},$$

which shows that the maximum observed likelihood prediction procedure excludes the variability in Y from consideration and effectively turns the prediction problem into an estimation problem for the model $f_\theta^{X|Y}(x|y)$, $\theta \in \Theta$, $y > x_p$. In this model the variability of the observed data depends on the value of the quantity of interest, that is Y, which allows to draw inference on Y purely within the classical maximum likelihood framework that is to perform optimization with respect to quantities that are model parameters. Alternatively, the above representation can be interpreted to suggest that the prediction procedure yields a prediction value, which is associated with the highest relative increase in the conditional density of Y given X compared to the unconditional density of Y.

In contrast to the MOLP, a general expression such as in (1.1) does not seem to exist for the MLP. Moreover, from expression (1.1), we find when predicting the very next record value ($s = r + 1$) that the MOLP becomes trivial in the sense that the last observed record value serves as predictor for the next one.

We examine the MLP and the MOLP of future Weibull record values, derive representations and compare their performance via the mean squared error and the Pitman closeness criterion. A predictor π_1 of R_s is said to be Pitman closer to Y than a predictor π_2 if

$$P(|\pi_1 - R_s| < |\pi_2 - R_s|) > \frac{1}{2}, \tag{1.4}$$

and, if (1.4) holds, π_1 is said to be preferable to π_2 in Pitman closeness sense.

2. LIKELIHOOD-BASED PREDICTORS FOR WEIBULL RECORD VALUES

Let $(R_n)_{n\in\mathbb{N}}$ be the sequence of Weibull record values. The density, cumulative distribution and quantile functions of the two-parameter Weibull distribution $\mathcal{W}eibull(\sigma, p)$ with scale parameter $\sigma \in \mathbb{R}_+$ and shape parameter $p \in \mathbb{R}_+$ are given by

$$\begin{aligned} f_\theta(x) &= \frac{p}{\sigma}\Big(\frac{x}{\sigma}\Big)^{p-1} \exp\Big\{-\Big(\frac{x}{\sigma}\Big)^p\Big\}, \quad x \in \mathbb{R}_+, \\ F_\theta(x) &= 1 - \exp\Big\{-\Big(\frac{x}{\sigma}\Big)^p\Big\}, \quad x \in \mathbb{R}_+, \\ F_\theta^{-1}(x) &= \sigma(-\ln(1-x))^{\frac{1}{p}}, \quad x \in [0,1), \end{aligned} \tag{2.1}$$

where $\theta = (\sigma, p) \in \mathbb{R}_+^2$ is the vector of the distributional parameters. For $r, s \in \mathbb{N}$, $r < s - 1$, we derive the MOLP as well as the MLP of the future record R_s based on $R_\star = (R_1, ..., R_r)$. The density functions of the distribution of $R_\star$ as well as of the conditional distribution of R_s given $R_r = r_r$, $r_r \in (-\infty, \omega(F_\theta))$, can be stated in terms of f_θ and F_θ as follows (see [1]):

$$f_\theta^{R_\star}(r_1, ..., r_r) = \left(\prod_{i=1}^{r-1} \frac{f_\theta(r_i)}{1 - F_\theta(r_i)}\right) f_\theta(r_r)\, \mathbb{1}_{[\alpha(F_\theta),\omega(F_\theta))^r}(r_1, ..., r_r), \quad (r_1, ..., r_r) \in \mathbb{R}^r_<, \tag{2.2}$$

$$f_\theta^{R_s|R_r}(r_s|r_r) = \frac{1}{(s-r-1)!} \frac{f_\theta(r_s)}{1 - F_\theta(r_r)} \left(-\ln\left(\frac{1 - F_\theta(r_s)}{1 - F_\theta(r_r)}\right)\right)^{s-r-1} \mathbb{1}_{(r_r,\omega(F_\theta))}(r_s), \quad r_s \in \mathbb{R}. \tag{2.3}$$

The MOLP in the Weibull case can be explicitly stated.

Proposition 2.1. *For $s \geq 3$, let $R_1, ..., R_s$ be the first s Weibull record values. For $r \in \mathbb{N}$, $2 \leq r < s-1$, the unique MOLP of R_s and the PMOLE of p based on $R_\star$ are given by*

$$\pi^{(s)}_{\mathrm{MOLP}} = \left(\frac{s-1}{r}\right)^{1/\hat{p}_{\mathrm{MOL}}} R_r \quad \text{and} \quad \hat{p}_{\mathrm{MOL}} = -\frac{r}{\ln\left(\prod_{i=1}^{r} \frac{R_i}{R_r}\right)}.$$

Proof: With f_θ and F_θ as above, the function $\Psi(\cdot|r_\star)$, $r_\star = (r_1, ..., r_r) \in (0,\infty)^r_<$, in (1.2) reads

$$\Psi(\theta|r_\star) = p^r \left(\prod_{i=1}^{r} \frac{r_i}{r_r}\right)^{p-1} \frac{1}{r_r^r}, \quad \sigma \in \mathbb{R}_+, \ p \in \mathbb{R}_+. \tag{2.4}$$

The function Ψ does not depend on the scale parameter σ, thus, we only need to find a maximizing function with respect to p. Let

$$\hat{\theta}(r_\star) = \left(\hat{\sigma}(r_\star), -r/\ln\left(\prod_{i=1}^{r} r_i/r_r\right)\right),$$

where $\hat{\beta}$ is an arbitrary measurable function on $\mathbb{R}^r_<$ with values in $\mathbb{R}_+$. Then, $\hat{\theta}$ satisfies (1.2) with $\Psi(\cdot|r_\star)$ given by (2.4). Together with

$$F_\theta^{-1}\left(1 - (1 - F_\theta(R_r))^{\frac{s-1}{r}}\right) = \left(\frac{s-1}{r}\right)^{\frac{1}{p}} R_r,$$

we find the stated form of the MOLP. □

Remark 2.1.

(**i**) The PMOLE and the MLE of p coincide. For the MLEs of σ and p we refer to [14].

(**ii**) The MOLP can also be written as $\pi^{(s)}_{\mathrm{MOLP}} = (s-1)^{1/\hat{p}}\hat{\sigma}$, where $\hat{p}$ and $\hat{\sigma}$ are the MLEs of p and σ, respectively.

The maximum likelihood predictor of a future Weibull record value was derived in [16] (see also [18, Section 5.3.5]). The respective result is contained in the following theorem.

Proposition 2.2. *For $s \geq 3$, let $R_1, ..., R_s$ be the first s Weibull record values. For $r \in \mathbb{N}$, $2 \leq r < s-1$, the unique MLP of R_s based on $R_\star$ is given by*

$$\pi^{(s)}_{\mathrm{MLP}} = s^{-1/\hat{p}_{\mathrm{ML}}}\hat{\sigma}_{\mathrm{ML}}.$$

Here, $\hat{\sigma}_{\mathrm{ML}}$ and $\hat{p}_{\mathrm{ML}}$ are the PMLEs of σ and p.

The PMLE of σ takes the form

$$\hat{\sigma}_{\mathrm{ML}} = \left(\frac{s + 1/\hat{p}_{\mathrm{ML}} - 1}{s(r + 1/\hat{p}_{\mathrm{ML}})}\right)^{1/\hat{p}_{\mathrm{ML}}} R_r,$$

while the PMLE of p is obtained as the unique positive solution of

$$p^2 \ln\left(\prod_{i=1}^{r-1} \frac{R_i}{R_r}\right) + (r+1)p = \ln\left(\frac{r + 1/p}{s + 1/p - 1}\right)$$

with respect to $p \in \mathbb{R}_+$. For $s = r+1$, the MLP takes the form $\pi^{(s)}_{\mathrm{MLP}} = R_r$.

The following remark collects, in particular, some results concerning the existence of the MOLP and the MLP in the case of the three-parameter Weibull distribution, where, in (2.1), x is replaced by $x-\mu$ for some location parameter μ.

Remark 2.2.

(**i**) It is straightforward to see that the MLP can also be expressed in the form

$$\pi_{\mathrm{MLP}}^{(s)} = \left(\frac{s+1/\hat{p}_{\mathrm{ML}}-1}{r+1/\hat{p}_{\mathrm{ML}}}\right)^{1/\hat{p}_{\mathrm{ML}}} R_r.$$

(**ii**) In case the underlying distribution depends on an unknown location parameter $\mu \in \mathbb{R}$, neither the MLP nor the MOLP exists. Indeed, consider first the derivation of the MOLP. Then, for $r_\star = (r_1, ..., r_r) \in \mathbb{R}^r_<$, we want to determine the global maximum of the function

$$\Psi(\mu, p|r_\star) = p^r \left(\prod_{i=1}^{r-1} \frac{r_i-\mu}{r_r-\mu}\right)^{p-1} \frac{1}{(r_r-\mu)^r}, \quad (\mu,\sigma) \in (-\infty, r_1) \times \mathbb{R}_+.$$

We have

$$\Psi\left(\mu, -r/\ln\left(\prod_{i=1}^{r-1} \frac{r_i-\mu}{r_r-\mu}\right)\middle| r_\star\right) \sim h\left(-\ln\left(\prod_{i=1}^{r-1} \frac{r_i-\mu}{r_r-\mu}\right)\right) \frac{r^r e^{-r}}{(r_r-r_1)^r},$$

as $\mu \underset{\mu<r_1}{\to} r_1$, where $h(x) = e^x/x^r$, $x \in \mathbb{R}_+$. Since $\lim\limits_{x\to\infty} h(x) = \infty$,

$$\lim_{\substack{\mu\to r_1 \\ \mu<r_1}} \Psi\left(\mu, -r/\ln\left(\prod_{i=1}^{r-1} \frac{r_i-\mu}{r_r-\mu}\right)\middle| r_\star\right) = \infty.$$

Hence, function Ψ does not possess a finite global maximum.

Next, consider the derivation of the MLP. There, for $(r_1, ..., r_r) \in \mathbb{R}^r_<$, we want, in particular, to maximize the function

$$G(\mu, p) = p^{r+1} \left(\prod_{i=1}^{r-1} \frac{r_i-\mu}{r_r-\mu}\right)^{p-1} \frac{1}{(r_r-\mu)^{r+1}} \frac{(r+1/p)^{r+1/p}}{(s+1/p-1)^{s+1/p-1}},$$
$$(\mu,\sigma) \in (-\infty, r_1) \times \mathbb{R}_+.$$

Since

$$\frac{(r+1/p)^{r+1/p}}{(s+1/p-1)^{s+1/p-1}} = \exp\left\{-(s-r-1)\ln(s+1/p-1) - \frac{(s-r-1)}{rp+1} + o(1)\right\},$$

as $p \to 0$, we have that

$$G\left(\mu, -r/\ln\left(\prod_{i=1}^{r-1} \frac{r_i-\mu}{r_r-\mu}\right)\right) \sim g\left(-\ln\left(\prod_{i=1}^{r-1} \frac{r_i-\mu}{r_r-\mu}\right)\right) \frac{r^{r+1} e^{-(s-1)}}{(r_r-r_1)^{r+1}},$$

as $\mu \underset{\mu<r_1}{\longrightarrow} r_1$, where $g(x) = e^{x\left(1-\frac{(s-r-1)\ln(x(\frac{1}{r+1}+\frac{s-1}{x}))}{x}\right)} \Big/ x^{r+1}$, $x \in \mathbb{R}_+$.
Since $\lim\limits_{x\to\infty} g(x) = \infty$, we conclude that

$$\lim_{\substack{\mu\to r_1 \\ \mu<r_1}} G\left(\mu, -r/\ln\left(\prod_{i=1}^{r-1} \frac{r_i-\mu}{r_r-\mu}\right)\right) = \infty.$$

Hence, function G does not possess a finite global maximum.

3. EVALUATION IN TERMS OF THE BIAS AND THE MSE

In the following, $\mathcal{G}amma(a,b)$, $a,b\in\mathbb{R}_+$, denotes the gamma distribution with parameters a,b with density function $f(x)=b^a x^{a-1}\exp\{-bx\}/\Gamma(a)$, $x>0$, where $\Gamma(a)$ is the gamma function evaluated at a.

Lemma 3.1. *For $s\geq 3$, let $R_1,\ldots,R_s$ be the first s Weibull record values. For $r\in\mathbb{N}$, $2\leq r<s-1$, the bias of the MOLP of R_s based on $R_\star$ is finite if and only if $\frac{1}{r}\ln\left(\frac{s-1}{r}\right)<p$, in which case it is given by*

$$\mathrm{E}(R_s-\pi^{(s)}_{\mathrm{MOLP}})=\sigma\frac{\Gamma\left(s+\frac{1}{p}\right)}{\Gamma(s)}\left(1-\frac{\prod_{i=r}^{s-1}\left(1+\frac{1}{pi}\right)^{-1}}{\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1}}\right).$$

If $\frac{1}{r}\ln\left(\frac{s-1}{r}\right)\geq p$, then $\mathrm{E}(R_s-\pi^{(s)}_{\mathrm{MOLP}})=-\infty$.

Proof: To prove the statement, we derive the expression for the expectation of $\pi^{(s)}_{\mathrm{MOLP}}$ and use that the integral is linear if one of the integrand functions is integrable (cf. [17, p. 135]). By [14, p. 42], R_r and $\hat{p}_{\mathrm{MOL}}$ are independent and $pr/\hat{p}_{\mathrm{MOL}}\sim\mathcal{G}amma(r-1,1)$. Using results in [1, section 2.7.1], we conclude that

$$\begin{aligned}
\mathrm{E}(\pi^{(s)}_{\mathrm{MOLP}}) &= \mathrm{E}\left(\left(\frac{s-1}{r}\right)^{1/\hat{p}_{\mathrm{MOL}}}R_r\right)=\mathrm{E}(R_r)\mathrm{E}\left(\left(\frac{s-1}{r}\right)^{1/\hat{p}_{\mathrm{MOL}}}\right)\\
&= \mathrm{E}(R_r)\mathrm{E}\left(\exp\left\{\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\frac{pr}{\hat{p}_{\mathrm{MOL}}}\right\}\right)=\mathrm{E}(R_r)\frac{1}{\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1}}\\
&= \sigma\frac{\Gamma\left(r+\frac{1}{p}\right)}{\Gamma(r)}\frac{1}{\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1}}=\sigma\frac{\Gamma\left(s+\frac{1}{p}\right)}{\Gamma(s)}\frac{\prod_{i=r}^{s-1}(1+\frac{1}{pi})^{-1}}{\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1}}\\
&= \mathrm{E}(R_s)\frac{\prod_{i=r}^{s-1}(1+\frac{1}{pi})^{-1}}{\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1}},
\end{aligned}$$

where in the fourth equality we used the expression for the moment generating function of the $\mathcal{G}amma(r-1,1)$ distribution to evaluate $E\left(\exp\left\{\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\frac{pr}{\hat{p}}\right\}\right)$, which is finite if and only if $\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)<1$, as well as the fact that $\Gamma(x+1)=\Gamma(x)x$, $x\in\mathbb{R}_+$. Now, linearity of the integral yields the desired conclusion. □

Lemma 3.2. *For $s\geq 3$, let $R_1,\ldots,R_s$ be the first s Weibull record values. For $r\in\mathbb{N}$, $2\leq r<s-1$, the MSE of the MOLP of R_s based on $R_\star$ is finite if and only if $\frac{2}{r}\ln(\frac{s-1}{r})<p$, in which case it is given by*

$$\mathrm{MSE}(\pi^{(s)}_{\mathrm{MOLP}})=\sigma^2\frac{\Gamma(s+\frac{2}{p})}{\Gamma(s)}\left(1-2\frac{\prod_{i=r}^{s-1}(1+\frac{1}{pi})^{-1}}{\left(1-\frac{1}{pr}\ln(\frac{s-1}{r})\right)^{r-1}}+\frac{\prod_{i=r}^{s-1}(1+\frac{2}{pi})^{-1}}{\left(1-\frac{2}{pr}\ln(\frac{s-1}{r})\right)^{r-1}}\right).$$

Proof: To prove the statement, we use that

$$(R_s - \pi^{(s)}_{\mathrm{MOLP}})^2 = R_s^2 - 2R_sR_r\left(\frac{s-1}{r}\right)^{1/\hat{p}_{\mathrm{MOL}}} + (\pi^{(s)}_{\mathrm{MOLP}})^2$$

as well as the fact that the integral is linear if the integrand can be written as a sum of an integrable and a quasi-integrable function (cf. [17, p. 135]). By [1, Theorem 3.3.1], we have

$$\mathrm{E}(R_s^2) = \sigma^2\frac{\Gamma(s+\frac{2}{p})}{\Gamma(s)},$$

$$\mathrm{E}(R_rR_s) = \sigma^2\frac{\Gamma(s+\frac{2}{p})}{\Gamma(s)}\prod_{i=r}^{s-1}\left(1+\frac{1}{pi}\right)^{-1},$$

and a similar argument as in the proof of Lemma 3.1 yields

$$\mathrm{E}((\pi^{(s)}_{\mathrm{MOLP}})^2) = \begin{cases} \sigma^2\dfrac{\Gamma(r+\frac{2}{p})}{\Gamma(r)}\dfrac{1}{\left(1-\frac{2}{pr}\ln(\frac{s-1}{r})\right)^{r-1}}, & \dfrac{2}{r}\ln\left(\dfrac{s-1}{r}\right) < p, \\ \infty, & \dfrac{2}{r}\ln\left(\dfrac{s-1}{r}\right) \geq p, \end{cases}$$

$$= \begin{cases} \sigma^2\dfrac{\Gamma(s+\frac{2}{p})}{\Gamma(s)}\dfrac{\prod_{i=r}^{s-1}(1+\frac{2}{pi})^{-1}}{\left(1-\frac{2}{pr}\ln(\frac{s-1}{r})\right)^{r-1}}, & \dfrac{2}{r}\ln\left(\dfrac{s-1}{r}\right) < p \\ \infty, & \dfrac{2}{r}\ln\left(\dfrac{s-1}{r}\right) \geq p. \end{cases}$$

Combining these results, we conclude that

$$\mathrm{MSE}(\pi^{(s)}_{\mathrm{MOLP}}) = \mathrm{E}((R_s - \pi^{(s)}_{\mathrm{MOLP}})^2)$$

$$= \begin{cases} \sigma^2\dfrac{\Gamma(s+\frac{2}{p})}{\Gamma(s)}\left(1 - 2\dfrac{\prod_{i=r}^{s-1}(1+\frac{1}{pi})^{-1}}{\left(1-\frac{1}{pr}\ln(\frac{s-1}{r})\right)^{r-1}} + \dfrac{\prod_{i=r}^{s-1}(1+\frac{2}{pi})^{-1}}{\left(1-\frac{2}{pr}\ln(\frac{s-1}{r})\right)^{r-1}}\right), & \dfrac{2}{r}\ln\left(\dfrac{s-1}{r}\right) < p, \\ \infty, & 1 < \dfrac{pr}{\ln(\frac{s-1}{r})} \leq 2. \end{cases}$$

Finally, it remains to show that $\mathrm{MSE}(\pi^{(s)}_{\mathrm{MOLP}}) = \infty$ for $p \leq \frac{1}{r}\ln(\frac{s-1}{r})$. Lemma 3.1 implies that $\mathrm{E}(|R_s - \pi^{(s)}_{\mathrm{MOLP}}|) = \infty$ for $p \leq \frac{1}{r}\ln(\frac{s-1}{r})$. By the well-known embedding theorem for Lebesgue spaces (cf. [17, Example 8.4.9 (2)]), we find $\mathrm{E}((R_s - \pi^{(s)}_{\mathrm{MOLP}})^2) = \infty$ for $p \leq \frac{1}{r}\ln(\frac{s-1}{r})$. □

Table 1 contain the biases and MSEs of the MLP (estimated from 10^7 Monte Carlo replications) and the MOLP for various values of r, s and p, and with $\sigma = 1$. Results in boldface represent all best results in terms of the MSE among the prediction methods, provided the best result is achieved by the MLP. The simulation results indicate that the MOLP exhibits superior performance based on the MSE in most cases. There are a few exceptions though, which suggest that the MLP has a lower MSE in cases when p, r are small ($p = 0.5$, $r = 3$) and $s = r + 2$. It should be noted that the MSE of the MOLP can become large (or even infinite) for small values of p in $(0, 1)$, small values of r and a higher gap between r and s (see Table 1). This is due to the fact that, in theses cases, $(2/r)\ln((s-1)/r)$ is close to p from below (or exceeds p), which yields large (or infinite) MSEs by means of Lemma 3.2. However, the situation of a small r combined with a large gap between r and s is not meaningful in practice. Moreover, one can observe that the difference in performance becomes smaller as the sample size increases.

Table 1: Values of the biases and MSEs of the MLP (estimated from 10^7 Monte Carlo replications) and the MOLP of R_s based on $\mathcal{W}eibull(1,p)$ record values $R_1,...,R_r$ for selected r and $s \in \{r+2, r+3, r+4, r+5\}$ for $p \in \{0.5, 1.5, 2, 2.5\}$. Boldface: MSEs of MLP provided it is minimal among the MSEs of both predictors.

r \ s	Predictor	$r+2$								$r+3$							
		$p=0.5$		$p=1.5$		$p=2$		$p=2.5$		$p=0.5$		$p=1.5$		$p=2$		$p=2.5$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
3	MLP	15.299	579.159	0.679	0.808	0.408	0.292	0.287	0.145	24.711	**1,309.337**	0.914	1.363	0.534	0.471	0.369	0.228
	MOLP	11.629	569.880	0.571	0.730	0.348	0.264	0.246	0.132	14.406	2,590.372	0.689	1.257	0.414	0.427	0.290	0.205
5	MLP	19.869	984.181	0.524	0.548	0.292	0.171	0.196	0.077	29.635	**2,008.898**	0.670	0.881	0.365	0.266	0.241	0.118
	MOLP	15.387	885.968	0.446	0.496	0.251	0.155	0.170	0.071	18.515	2,042.598	0.508	0.787	0.284	0.237	0.190	0.105
10	MLP	30.423	2,409.340	0.372	0.323	0.186	0.082	0.117	0.033	40.279	4,281.257	0.439	0.486	0.217	0.120	0.135	0.047
	MOLP	25.197	2,169.620	0.332	0.300	0.168	0.076	0.106	0.031	28.332	3,842.880	0.357	0.441	0.180	0.110	0.113	0.043
15	MLP	40.638	4,442.341	0.309	0.239	0.145	0.053	0.088	0.020	50.312	7,340.106	0.350	0.348	0.163	0.076	0.098	0.028
	MOLP	35.132	4,065.112	0.283	0.226	0.134	0.051	0.082	0.019	38.237	6,601.259	0.298	0.322	0.141	0.071	0.086	0.026
20	MLP	50.727	7,064.397	0.273	0.194	0.123	0.039	0.072	0.014	60.330	11,192.520	0.302	0.277	0.134	0.056	0.079	0.019
	MOLP	45.099	6,562.983	0.254	0.185	0.115	0.038	0.068	0.013	48.183	10,183.050	0.265	0.260	0.119	0.052	0.071	0.018

r \ s	Predictor	$r+4$								$r+5$							
		$p=0.5$		$p=1.5$		$p=2$		$p=2.5$		$p=0.5$		$p=1.5$		$p=2$		$p=2.5$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
3	MLP	36.179	**2,526.103**	1.145	2.028	0.655	0.676	0.447	0.320	49.757	—	1.372	2.801	0.771	0.903	0.520	0.419
	MOLP	14.526	60,224.680	0.799	1.990	0.475	0.632	0.331	0.293	8.623	—	0.901	2.981	0.532	0.880	0.369	0.394
5	MLP	41.340	**3,612.705**	0.817	1.276	0.437	0.373	0.285	0.162	55.020	**5,967.848**	0.964	1.733	0.507	0.492	0.328	0.211
	MOLP	20.992	4,742.856	0.568	1.161	0.315	0.335	0.210	0.145	22.353	11,142.580	0.626	1.623	0.344	0.449	0.229	0.190
10	MLP	51.544	6,913.748	0.508	0.672	0.247	0.163	0.153	0.063	64.231	10,459.760	0.579	0.883	0.278	0.209	0.170	0.080
	MOLP	31.323	6,464.899	0.382	0.608	0.191	0.148	0.120	0.057	34.083	10,444.290	0.407	0.802	0.203	0.190	0.127	0.073
15	MLP	61.189	11,164.670	0.392	0.469	0.181	0.101	0.108	0.037	73.097	16,080.470	0.435	0.604	0.198	0.128	0.118	0.046
	MOLP	41.279	10,142.230	0.313	0.432	0.147	0.094	0.089	0.034	44.221	14,965.170	0.328	0.555	0.154	0.118	0.093	0.042
20	MLP	70.786	16,400.880	0.331	0.368	0.146	0.073	0.085	0.025	82.231	22,872.340	0.360	0.467	0.158	0.092	0.091	0.031
	MOLP	51.232	14,927.090	0.275	0.343	0.124	0.068	0.073	0.024	54.225	21,027.620	0.285	0.434	0.128	0.085	0.075	0.029

4. COMPARISON IN TERMS OF PITMAN'S MEASURE OF CLOSENESS

Since the MLP $\pi^{(s)}_{\mathrm{MLP}}$ of R_s based on $R_\star$ is not given in closed form, we are not able to derive an analytic expression for the Pitman efficiency of the MOLP $\pi^{(s)}_{\mathrm{MOLP}}$ relative to $\pi^{(s)}_{\mathrm{MLP}}$. We therefore aim at establishing a lower bound on the Pitman efficiency. The following lemmas are required to establish the desired result.

Lemma 4.1. *For $s \geq 3$, let $R_1, ..., R_s$ be the first s Weibull record values. For $r \in \mathbb{N}$, $2 \leq r < s-1$, the MOLP of R_s is always greater than the MLP of R_s.*

Proof: Indeed, we know that, for $(r_1, ..., r_r) \in (0, \infty)^r_<$, the PMLE of p satisfies the equation

$$p\left(p \ln\left(\prod_{i=1}^{r-1} \frac{r_i}{r_r}\right) + r + 1\right) = \ln\left(\frac{r + 1/p}{s + 1/p - 1}\right).$$

Note that $\ln\left(\frac{r+1/p}{s+1/p-1}\right) < 0$, $p \in \mathbb{R}_+$. Consequently, since $\ln(\prod_{i=1}^{r-1} r_i/r_r) < 0$, the solution of the above equation is always greater than $-(r+1)/\ln(\prod_{i=1}^{r-1} r_i/r_r)$ $(= \frac{r+1}{r}\hat{p}_{\mathrm{MOL}})$. Now, for $\alpha, \beta \in \mathbb{R}$, $0 < \alpha < \beta$, consider the functions

$$f_{\alpha,\beta}(t) = \frac{\beta + t}{\alpha + t}, \quad g_{\alpha,\beta}(t) = f_{\alpha,\beta}(t)^t, \quad t \in (-\alpha, \infty).$$

Differentiating $f_{\alpha,\beta}$ and $g_{\alpha,\beta}$ yields

$$f'_{\alpha,\beta}(t) = -\frac{\beta - \alpha}{(\alpha + t)^2}, \quad g'_{\alpha,\beta}(t) = f_{\alpha,\beta}(t)\left(\ln(f_{\alpha,\beta}(t)) + t\frac{f'_{\alpha,\beta}(t)}{f_{\alpha,\beta}(t)}\right), \quad t \in (-\alpha, \infty).$$

Obviously, $f'_{\alpha,\beta}(t) < 0$, $t \in (-\alpha, \infty)$. Hence, $f_{\alpha,\beta}$ is a strictly decreasing function. Furthermore, for $t \in (-\alpha, \infty)$,

$$\begin{aligned}
\ln(f_{\alpha,\beta}(t)) + t\frac{f'_{\alpha,\beta}(t)}{f_{\alpha,\beta}(t)} &= \ln\left(1 + \frac{\beta - \alpha}{\alpha + t}\right) + t\frac{\alpha - \beta}{(\alpha + t)(\beta + t)} \\
&> \frac{\beta - \alpha}{\beta + t} + t\frac{\alpha - \beta}{(\alpha + t)(\beta + t)} \\
&= \frac{\alpha(\beta - \alpha)}{(\alpha + t)(\beta + t)} > 0,
\end{aligned}$$

where we used the inequality $x/(x+1) < \ln(1+x)$, for $x > -1$, $x \neq 0$. Thus, $g_{\alpha,\beta}$ is a strictly increasing function. Using the preceding results, we obtain

$$\begin{aligned}
\pi^{(s)}_{\mathrm{MLP}} &= g_{r,s-1}(1/\hat{p}_{\mathrm{ML}})R_r \\
&< g_{r,s-1}(1/\hat{p}_{\mathrm{MOL}})R_r \\
&= f_{r,s-1}(1/\hat{p}_{\mathrm{MOL}})^{1/\hat{p}_{\mathrm{MOL}}}R_r \\
&< f_{r,s-1}(0)^{1/\hat{p}_{\mathrm{MOL}}}R_r = \pi^{(s)}_{\mathrm{MOLP}}.
\end{aligned}$$

□

Lemma 4.2. *For $s \geq 3$, let $R_1, ..., R_s$ be the first s Weibull record values. For $r \in \mathbb{N}$, $2 \leq r < s-1$, the probability of R_s exceeding its MOLP based on $R_\star$ is given by*

$$P(\pi^{(s)}_{\mathrm{MOLP}} < R_s) = \sum_{j=r}^{s-1} \sum_{i=0}^{s-1-j} (-1)^i \binom{s-1}{j} \binom{s-1-j}{i} \frac{1}{\left(1 + \frac{i+j}{r} \ln\left(\frac{s-1}{r}\right)\right)^{r-1}}.$$

In particular, this probability is independent of the distributional parameters σ and p.

Proof: Observe that

$$\pi^{(s)}_{\mathrm{MOLP}} < R_s \quad \Longleftrightarrow \quad \frac{s-1}{r} < \left(\frac{R_s}{R_r}\right)^{\hat{p}_{\mathrm{MOL}}}.$$

Let G denote the cumulative distribution function of $(R_s/R_r)^{1/\hat{p}_{\mathrm{MOL}}}$. By the results in [14, section 4], G admits the representation

$$G(t) = \int_0^\infty H\left(\frac{r}{s-r}\left(t^{\frac{z}{2r}} - 1\right)\middle| 2(s-r), 2r\right) f_{\chi^2}(z|2(r-1))dz, \quad t \in (1, \infty).$$

Here, for $n, m \in \mathbb{N}$, $H(\cdot|n,m)$ denotes the cumulative distribution function of the F distribution with parameters n and m, and $f_{\chi^2}(\cdot|n)$ denotes the density function of the χ^2 distribution with parameter n. First, note that

$$H\left(\frac{r}{s-r}\left(t^{\frac{z}{2r}} - 1\right)\middle| 2(s-r), 2r\right) = I_{1-t^{-\frac{z}{2r}}}(s-r, r) = 1 - I_{t^{-\frac{z}{2r}}}(r, s-r).$$

Consequently,

$$P(\pi^{(s)}_{\mathrm{MOLP}} < R_s) = 1 - G\left(\frac{s-1}{r}\right) = \int_0^\infty I_{\left(\frac{r}{s-1}\right)^{\frac{z}{2r}}}(r, s-r) f_{\chi^2}(z|2(r-1))dz.$$

Furthermore, since the parameters of the regularized incomplete beta function are integers, we have, by the relation of the regularized incomplete beta function to the binomial expansion (see [8, (6.6.4)]),

$$I_x(r, s-r) = \sum_{j=r}^{s-1} \binom{s-1}{j} x^j (1-x)^{s-1-j} = \sum_{j=r}^{s-1} \sum_{i=0}^{s-1-j} (-1)^i \binom{s-1}{j} \binom{s-1-j}{i} x^{i+j}, \quad x \in (0,1).$$

From the preceding results we infer that

$$\begin{aligned}
P(\pi^{(s)}_{\mathrm{MOLP}} < R_s) &= \int_0^\infty I_{\left(\frac{r}{s-1}\right)^{\frac{z}{2r}}}(r, s-r) f_{\chi^2}(z|2(r-1))dz \\
&= \sum_{j=r}^{s-1} \sum_{i=0}^{s-1-j} (-1)^i \binom{s-1}{j} \binom{s-1-j}{i} \int_0^\infty \left(\frac{r}{s-1}\right)^{\frac{i+j}{2r} z} f_{\chi^2}(z|2(r-1))dz \\
&= \sum_{j=r}^{s-1} \sum_{i=0}^{s-1-j} (-1)^i \binom{s-1}{j} \binom{s-1-j}{i} \int_0^\infty e^{\frac{i+j}{2r} \ln\left(\frac{r}{s-1}\right) z} f_{\chi^2}(z|2(r-1))dz \\
&= \sum_{j=r}^{s-1} \sum_{i=0}^{s-1-j} (-1)^i \binom{s-1}{j} \binom{s-1-j}{i} \frac{1}{\left(1 + \frac{i+j}{r} \ln\left(\frac{s-1}{r}\right)\right)^{r-1}},
\end{aligned}$$

where in the last equality we used the expression for the moment generating function of the $\chi^2(2(r-1))$ distribution to evaluate the integrals $\int_0^\infty e^{\frac{i+j}{2r} \ln\left(\frac{r}{s-1}\right) z} f_{\chi^2}(z|2(r-1))dz$. This concludes the proof. □

Remark 4.1. The proof of Lemma 4.2 yields a finite sum representation of the cumulative distribution function G of $(R_s/R_r)^{1/\hat{p}_{\mathrm{MOL}}}$:

$$G(t) = 1 - \sum_{j=r}^{s-1} \sum_{i=0}^{s-1-j} (-1)^i \binom{s-1}{j} \binom{s-1-j}{i} \frac{1}{\left(1 + \frac{i+j}{r} \ln(t)\right)^{r-1}}, \quad t \in (1, \infty).$$

By exploiting the presence of alternating binomial sums in the above representation, a more compact representation of G can be obtained. More precisely, we have that

$$G(t) = 1 - \sum_{j=r}^{s-1} (-1)^{s-1-j} \binom{s-1}{j} \Delta^{s-1-j} f_{r,j,t}, \quad t \in (1, \infty),$$

where

$$f_{r,j,t}(i) = \frac{1}{\left(1 + \frac{i+j}{r} \ln(t)\right)^{r-1}}, \quad 0 \le i \le s-1-j,$$

and, for $j = r, ..., s-1$, the $(s-1-j)$-th forward difference is computed for $i = 0$. Using this finite sum representation allows to avoid applying numeric integration for evaluation of G (cf. [14, section 6]). Since alternating sums can be numerically problematic, for an efficient and accurate implementation of G, it is advisable to use high precision arithmetic. See `sumBinomMpfr()` in `R` package `Rmpfr` and its documentation.

Proposition 4.1. *For $s \ge 3$, let $R_1, ..., R_s$ be the first s Weibull record values. For $r \in \mathbb{N}$, $2 \le r < s-1$, let $\pi^{(s)}_{\mathrm{MOLP}}$ and $\pi^{(s)}_{\mathrm{MLP}}$ be the MOLP and the MLP of R_s based on $R_\star$, respectively. Then*

$$P(|\pi^{(s)}_{\mathrm{MOLP}} - R_s| < |\pi^{(s)}_{\mathrm{MLP}} - R_s|) > \sum_{j=r}^{s-1} \sum_{i=0}^{s-1-j} (-1)^i \binom{s-1}{j} \binom{s-1-j}{i} \frac{1}{\left(1 + \frac{i+j}{r} \ln\left(\frac{s-1}{r}\right)\right)^{r-1}}.$$

Proof: Due to Lemma 4.1,

$$P(|\pi^{(s)}_{\mathrm{MOLP}} - R_s| < |\pi^{(s)}_{\mathrm{MLP}} - R_s|) > P(R_s > \pi^{(s)}_{\mathrm{MOLP}}).$$

Hence, Lemma 4.2 yields the desired result. □

Figure 1 contains the contour plots of the lower bound on the Pitman efficiency $\mathrm{PE}(\mathrm{MOLP}, \mathrm{MLP}) = P\left(|R_s - \pi^{(s)}_{\mathrm{MOLP}}| < |R_s - \pi^{(s)}_{\mathrm{MLP}}|\right)$ of the MOLP of R_s relative to the MLP of R_s based on $R_\star$ for r, s such that $2 \le r \le 20$ and $r+1 < s \le r+10$. Table 2 contains values of the lower bound on as well as estimated Pitman efficiencies for selected r and s, and, in the case of estimated Pitman efficiencies, for shape parameter values $p = 0.5, 1.5, 2, 2.5$. Observe that while the lower bound on the Pitman efficiencies does not depend on the distributional parameters, the Pitman efficiencies do depend on the shape parameter p.

Each estimated Pitman efficiency was computed based on 10^6 simulated samples of Weibull record values. From the contour plot of the lower bound on the Pitman efficiency, the MOLP seems to be superior to the MLP in terms of Pitman closeness for r, s such that $2 \leq r \leq 20$ and $r+1 < s \leq r+10$. The estimated Pitman efficiencies presented in Table 2 as well as additional simulation results suggest that for fixed r and s the Pitman efficiency is a decreasing function of p. Furthermore, the simulation results indicate that the lower bound from Proposition 4.1 is the tighter, the bigger r and the smaller $s-r$ are.

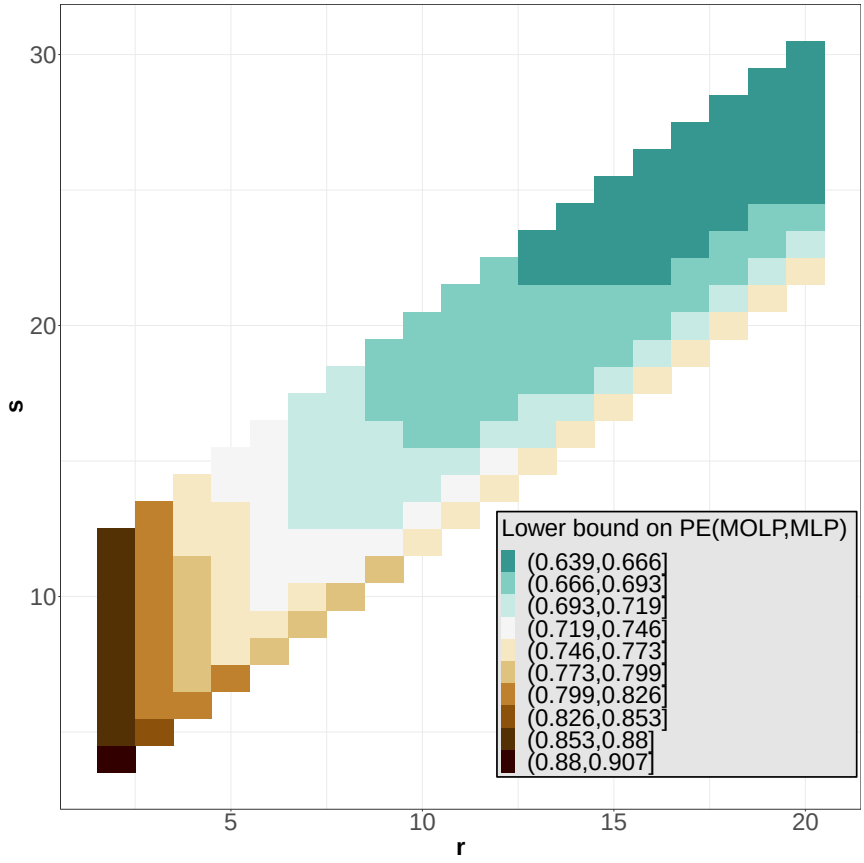

Figure 1: Contour plot of the lower bound on the Pitman efficiency $\mathrm{PE}(\mathrm{MOLP}, \mathrm{MLP}) = P(|R_s - \pi^{(s)}_{\mathrm{MOLP}}| < |R_s - \pi^{(s)}_{\mathrm{MLP}}|)$ of the MOLP of R_s relative to the MLP of R_s based on Weibull record values $R_1, ..., R_r$ for r, s such that $2 \leq r \leq 20$ and $r+1 < s \leq r+10$.

The superior performance of the MOLP in terms of the Pitman efficiency compared to the MLP even for small values of r and $p < 1$, for which the MOLP performs poorly if evaluated in terms of the MSE (see Table 1) is in line with the intuition underlying the Pitman criterion: namely being only affected by the bias and not accounting for the variability in the predictors. Note also that according to Table 1, despite MOLP's inferior performance in terms of the MSE, it still has a lower bias than the MLP, which supports its superior performance in terms of the Pitman criterion as evidenced by the values in Table 2.

Table 2: Values of the lower bound (first row in each section) as well as estimated Pitman efficiencies $\mathrm{PE}(\mathrm{MOLP}, \mathrm{MLP}) = P(|R_s - \pi^{(s)}_{\mathrm{MOLP}}| < |R_s - \pi^{(s)}_{\mathrm{MLP}}|)$ of the MOLP of R_s relative to the MLP of R_s based on Weibull record values $R_1, ..., R_r$ for selected r and s, and, in the case of estimated Pitman efficiencies, for $p \in \{0.5, 1.5, 2, 2.5\}$.

r \ s	p	$r+2$	$r+3$	$r+4$	$r+5$	$r+10$
2		0.906	0.875	0.877	0.874	0.873
	0.5	0.927	0.920	0.917	0.915	0.909
	1.5	0.923	0.918	0.918	0.918	0.920
	2	0.921	0.916	0.917	0.917	0.920
	2.5	0.920	0.915	0.916	0.916	0.920
5		0.805	0.771	0.759	0.754	0.745
	0.5	0.852	0.836	0.832	0.831	0.831
	1.5	0.838	0.818	0.812	0.812	0.818
	2	0.836	0.814	0.809	0.808	0.812
	2.5	0.834	0.813	0.807	0.805	0.809
10		0.771	0.727	0.708	0.698	0.681
	0.5	0.807	0.778	0.770	0.766	0.769
	1.5	0.794	0.760	0.746	0.742	0.739
	2	0.792	0.757	0.743	0.738	0.734
	2.5	0.790	0.755	0.742	0.736	0.731
15		0.760	0.711	0.689	0.677	0.654
	0.5	0.788	0.751	0.738	0.732	0.729
	1.5	0.776	0.735	0.718	0.710	0.701
	2	0.775	0.733	0.716	0.707	0.696
	2.5	0.773	0.731	0.713	0.705	0.693
20		0.754	0.703	0.679	0.665	0.639
	0.5	0.776	0.735	0.719	0.711	0.704
	1.5	0.767	0.722	0.702	0.692	0.677
	2	0.766	0.720	0.700	0.690	0.673
	2.5	0.765	0.719	0.699	0.687	0.671

5. ASYMPTOTIC RESULTS

In the present section we establish two asymptotic results concerning the behavior of the bias as well as the asymptotic distribution of the prediction error of the MOLP. Hereby, we consider sequences $(r_n)_{n=1}^{\infty}, (s_n)_{n=1}^{\infty} \in \mathbb{N}^{\mathbb{N}}$, satisfying

$$r_n < s_n \text{ for all } n \in \mathbb{N} \text{ and } \lim_{n\to\infty} r_n = \infty.$$

However, by an abuse of notation, we write $r, s \to \infty$ when taking limits with respect to n. Also, we will suppress n in the notation.

Proposition 5.1. *For $s \geq 3$, let $R_1, ..., R_s$ be the first s Weibull record values. The MOLP $\pi^{(s)}_{\mathrm{MOLP}}$ of R_s based on $R_\star$ is asymptotically unbiased in the sense that if $\lim_{s,r\to\infty} s/r = \lambda$, for some $\lambda > 1$, then*

$$\frac{\mathrm{E}(\pi^{(s)}_{\mathrm{MOLP}})}{\mathrm{E}(R_s)} \longrightarrow 1, \quad r, s \to \infty.$$

Proof: Observe that, under the stated assumptions, the condition $\frac{1}{pr}\ln\left(\frac{s-1}{r}\right) < 1$ is satisfied for r large enough, in which case, by the proof of Lemma 3.1,

$$\frac{\mathrm{E}(\pi^{(s)}_{\mathrm{MOLP}})}{\mathrm{E}(R_s)} = \frac{\prod_{i=r}^{s-1}(1+\frac{1}{pi})^{-1}}{\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1}}.$$

For $x \in (-1,1)$, set $\rho(x) = \sum_{k=2}^{\infty}(-1)^{k+1}\frac{x^k}{k}$. Then $\log(1+x) = x + \rho(x)$, $x \in (-1,1)$, and

$$\begin{aligned}\prod_{i=r}^{s-1}\left(1+\frac{1}{pi}\right)^{-1} &= \exp\left\{-\sum_{i=r}^{s-1}\log\left(1+\frac{1}{pi}\right)\right\}\\ &= \exp\left\{-\frac{1}{p}\sum_{i=r}^{s-1}\frac{1}{i}\right\}\exp\left\{-\sum_{i=r}^{s-1}\rho(1/pi)\right\}.\end{aligned}$$

Since $\rho(x) = O(x^2)$ as $x \to 0$, we have $\rho(1/pi) = O(1/i^2)$ as $i \to \infty$. Consequently, $\lim_{r,s\to\infty}\sum_{i=r}^{s-1}\rho(1/pi) = 0$. Moreover, $\lim_{r,s\to\infty}\sum_{i=r}^{s-1}1/i = \ln(\lambda)$. Hence, we obtain that

$$\lim_{r,s\to\infty}\prod_{i=r}^{s-1}\left(1+\frac{1}{pi}\right)^{-1} = \left(\frac{1}{\lambda}\right)^{\frac{1}{p}}.$$

The claim now follows from the fact that

$$\lim_{r,s\to\infty}\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1} = \exp\left\{-\frac{1}{p}\ln(\lambda)\right\} = \left(\frac{1}{\lambda}\right)^{\frac{1}{p}}. \qquad \square$$

Remark 5.1. The shifted Stirling's approximation for the (real) Gamma function reads (see [2, formula (5.11.7)])

$$\Gamma(x+a) = \sqrt{2\pi}e^{-x}x^{x+a-\frac{1}{2}}e^{o(1)}, \quad \text{as } x \to \infty.$$

Hence,

$$\frac{\Gamma(s+1/p)}{\Gamma(s)} = s^{1/p}e^{o(1)}, \quad \text{as } s \to \infty.$$

To prove that $\pi^{(s)}_{\mathrm{MOLP}}$ is unbiased in the limit, i.e., $\lim_{r,s\to\infty}\mathrm{E}(R_s - \pi^{(s)}_{\mathrm{MOLP}}) = 0$, where r and s are supposed to satisfy $\lim_{r,s\to\infty}\frac{s}{r} = \lambda > 1$, one has to prove that

$$s^{1/p}\left(1-\frac{\prod_{i=r+1}^{s}(1+\frac{1}{pi})^{-1}}{\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1}}\right) \to 0, \quad r,s \to \infty,$$

which is equivalent to showing that

$$s^{1/p}\left(\left(1-\frac{1}{pr}\ln\left(\frac{s-1}{r}\right)\right)^{r-1} - \prod_{i=r+1}^{s}\left(1+\frac{1}{pi}\right)^{-1}\right) \to 0, \qquad r,s \to \infty.$$

Numerical computation indicates that this is true.

We continue with a result concerning the asymptotic distribution of the prediction error of the MOLP.

Proposition 5.2. *For $s \geq 3$, let $R_1, ..., R_s$ be the first s Weibull record values. The prediction error of the MOLP $\pi^{(s)}_{\mathrm{MOLP}}$ of R_s based on $R_\star$ has an asymptotic normal distribution. More specifically, we have that*

$$\alpha_s(\sigma, p)\left(R_s - \left(\frac{s-1}{r}\right)^{\frac{1}{\hat{p}_{\mathrm{MOL}}}} R_r\right) \longrightarrow \mathcal{N}\big(0, \lambda + \lambda \ln^2(\lambda) - 1\big), \qquad r, s \to \infty,$$

where it is assumed that there exists a $\lambda \in (1, \infty)$ such that $\lim_{r,s\to\infty}(\lambda - s/r)\sqrt{r} = 0$, and the sequence of normalizing constants is given by $\alpha_s(\sigma, p) = \frac{p}{\sigma} s^{\frac{1}{2}-\frac{1}{p}}$.

Proof: First, recall that, by result (7) in [14], $\hat{p}_{\mathrm{MOL}}$ is independent of R_r and R_s and $pr/\hat{p}_{\mathrm{MOL}} \sim \mathit{Gamma}(r-1, 1)$. Let $(Y_n)_{n=1}^{\infty}$ and $(Z_n)_{n=1}^{\infty}$ be two independent sequences of i.i.d. random variables, Y_1, $Z_1 \sim \mathcal{E}xp(1)$. By [1, equation (2.3.3)], for any $r, s \in \mathbb{N}$, $r < s$, the identity $(R_r, R_s) \stackrel{d}{=} \sigma((\sum_{i=1}^{r} Y_i)^{1/p}, (\sum_{i=1}^{s} Y_i)^{1/p})$ holds true. Combining these results, we conclude that

$$\begin{aligned}
\frac{1}{\sigma}\left(R_s - \left(\frac{s-1}{p}\right)^{\frac{1}{\hat{p}_{\mathrm{MOL}}}} R_r\right) &\stackrel{d}{=} \left(\sum_{i=1}^{r} Y_i + \sum_{i=r+1}^{s} Y_i\right)^{1/p} - \left(\left[\frac{s-1}{r}\right]^{1/p}\right)^{\frac{1}{r}\sum_{i=1}^{r-1} Z_i} \left(\sum_{i=1}^{r} Y_i\right)^{1/p} \\
&= \left(s + \left\{\frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}}\sqrt{\frac{r}{s}} + \frac{\sum_{i=r+1}^{s} Y_i - (s-r)}{\sqrt{s-r}}\sqrt{\frac{s-r}{s}}\right\}\sqrt{s}\right)^{1/p} - s^{1/p} \\
&\quad + s^{1/p} - \lambda^{1/p}\left(r + \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}}\sqrt{r}\right)^{1/p} \\
&\quad + \left\{\lambda^{1/p} - \left(\left[\frac{s-1}{r}\right]^{1/p}\right)^{\frac{1}{r}\left((r-1) + \frac{\sum_{i=1}^{r-1} Z_i - (r-1)}{\sqrt{r-1}}\sqrt{r-1}\right)}\right\} \\
&\quad \times \left(r + \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}}\sqrt{r}\right)^{1/p} \\
&=: A + B + C. \qquad (5.1)
\end{aligned}$$

We first treat the terms denoted by A. By [6, Theorem 5.6.1],

$$\frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}}, \frac{\sum_{i=r+1}^{s} Y_i - (s-r)}{\sqrt{s-r}} \xrightarrow{d} \mathcal{N}(0,1), \qquad r, s \to \infty. \qquad (5.2)$$

Hence, applying Theorem 2.8 in [4] as well as Slutsky's lemma, we conclude that

$$\frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}}\sqrt{\frac{r}{s}} + \frac{\sum_{i=r+1}^{s} Y_i - (s-r)}{\sqrt{s-r}}\sqrt{\frac{s-r}{s}} \xrightarrow{d} \mathcal{N}(0,1), \qquad r, s \to \infty. \qquad (5.3)$$

By expanding the function $f(x) = (1+x)^{1/p}$, $x > -1$, around $x = 0$, $f(x) = 1 + \frac{x}{p} + o(x)$, as $x \to 0$. Consequently,

$$\begin{aligned} A &= \left(s + \left\{ \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{\frac{r}{s}} + \frac{\sum_{i=r+1}^{s} Y_i - (s-r)}{\sqrt{s-r}} \sqrt{\frac{s-r}{s}} \right\} \sqrt{s} \right)^{1/p} - s^{1/p} \\ &= s^{1/p} \left[\left(1 + \left\{ \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{\frac{r}{s}} + \frac{\sum_{i=r+1}^{s} Y_i - (s-r)}{\sqrt{s-r}} \sqrt{\frac{s-r}{s}} \right\} \frac{1}{\sqrt{s}} \right)^{1/p} - 1 \right] \\ &= \frac{s^{\frac{1}{p}-\frac{1}{2}}}{p} \left\{ \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{\frac{r}{s}} + \frac{\sum_{i=r+1}^{s} Y_i - (s-r)}{\sqrt{s-r}} \sqrt{\frac{s-r}{s}} \right\} (1 + o_P(1)), \quad r, s \to \infty. \end{aligned} \tag{5.4}$$

Similarly, we have

$$\begin{aligned} B &= s^{1/p} - \lambda^{1/p} \left(r + \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{r} \right)^{1/p} \\ &= -s^{1/p} \left[\left(\lambda \frac{r}{s} + \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{\lambda^2 \frac{r}{s}} \frac{1}{\sqrt{s}} \right)^{1/p} - 1 \right] \\ &= -s^{1/p} \left[\left(1 + o(1/\sqrt{s}) + \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{\lambda^2 \frac{r}{s}} \frac{1}{\sqrt{s}} \right)^{1/p} - 1 \right], \quad r, s \to \infty \\ &= -\frac{s^{\frac{1}{p}-\frac{1}{2}}}{p} \left(\frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{\lambda^2 \frac{r}{s}} + o(1) \right) (1 + o_P(1)), \quad r, s \to \infty, \end{aligned} \tag{5.5}$$

where the second-last equality is due to $\lim_{r,s\to\infty} (\lambda - s/r)\sqrt{r} = 0$. Finally, we treat the terms denoted by C. First, we rewrite C as

$$\begin{aligned} C &= \left\{ \lambda^{1/p} - \left(\left[\frac{s-1}{r} \right]^{1/p} \right)^{\frac{1}{r}\left((r-1) + \frac{\sum_{i=1}^{r-1} Z_i - (r-1)}{\sqrt{r-1}} \sqrt{r-1} \right)} \right\} \left(r + \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{r} \right)^{1/p} \\ &= \frac{s^{\frac{1}{p}-\frac{1}{2}}}{p} \left\{ \lambda^{1/p} - \left(\left[\frac{s-1}{r} \right]^{1/p} \right)^{\frac{1}{r}\left((r-1) + \frac{\sum_{i=1}^{r-1} Z_i - (r-1)}{\sqrt{r-1}} \sqrt{r-1} \right)} \right\} \\ &\quad \times \frac{\left(r + \frac{\sum_{i=1}^{r} Y_i - r}{\sqrt{r}} \sqrt{r} \right)^{1/p} - r^{1/p}}{s^{\frac{1}{p}-\frac{1}{2}}/p} \\ &\quad + \frac{s^{\frac{1}{p}-\frac{1}{2}}}{p} \left\{ \lambda^{1/p} - \left(\left[\frac{s-1}{r} \right]^{1/p} \right)^{\frac{1}{r}\left((r-1) + \frac{\sum_{i=1}^{r-1} Z_i - (r-1)}{\sqrt{r-1}} \sqrt{r-1} \right)} \right\} \left(\frac{r}{s} \right)^{1/p} p\sqrt{s} \end{aligned}$$

$$=: D + E. \tag{5.6}$$

By [6, Theorem 5.6.1] and Slutsky's lemma,

$$\lambda^{1/p} - \left(\left[\frac{s-1}{r} \right]^{1/p} \right)^{\frac{1}{r}\left((r-1) + \frac{\sum_{i=1}^{r-1} Z_i - (r-1)}{\sqrt{r-1}} \sqrt{r-1} \right)} \xrightarrow{P} 0, \qquad r, s \to \infty. \tag{5.7}$$

Next, by (5.2), we have that

$$\begin{aligned}\frac{\left(r+\frac{\sum_{i=1}^{r}Y_i-r}{\sqrt{r}}\sqrt{r}\right)^{1/p}-r^{1/p}}{s^{\frac{1}{p}-\frac{1}{2}}/p} &= \frac{\left(r+\frac{\sum_{i=1}^{r}Y_i-r}{\sqrt{r}}\sqrt{r}\right)^{1/p}-r^{1/p}}{r^{\frac{1}{p}-\frac{1}{2}}/p}\frac{r^{\frac{1}{p}-\frac{1}{2}}}{s^{\frac{1}{p}-\frac{1}{2}}}\\ &= \frac{\left(1+\frac{\sum_{i=1}^{r}Y_i-r}{\sqrt{r}}\frac{1}{\sqrt{r}}\right)^{1/p}-1}{r^{-\frac{1}{2}}/p}\frac{r^{\frac{1}{p}-\frac{1}{2}}}{s^{\frac{1}{p}-\frac{1}{2}}}\\ &= \left(\frac{\sum_{i=1}^{r}Y_i-r}{\sqrt{r}}+o_P(1)\right)\lambda^{\frac{1}{2}-\frac{1}{p}}, \quad r,s\to\infty\\ &= O_P(1), \qquad r,s\to\infty. \end{aligned} \tag{5.8}$$

From (5.7) and (5.8) it readily follows that

$$ps^{\frac{1}{2}-\frac{1}{p}}D = o_P(1), \qquad r,s\to\infty. \tag{5.9}$$

Next, consider the terms denoted by E. We have $(r/s)^{1/p}p\sqrt{s}\sim p\lambda^{\frac{1}{2}-\frac{1}{p}}\sqrt{r}$, as $r,s\to\infty$. Furthermore, setting

$$W_r = \frac{1}{p}\left(\left(\frac{\sum_{i=1}^{r-1}Z_i-(r-1)}{\sqrt{r-1}}\sqrt{\frac{r-1}{r}}\right)\frac{1}{\sqrt{r}}-\frac{1}{r}\right)\ln\left(\frac{s-1}{r}\right)$$

and observing that $W_r = o_P(1)$, as $r,s\to\infty$, we conclude that

$$\begin{aligned}\left(\left[\frac{s-1}{r}\right]^{1/p}\right)^{\frac{1}{r}\left((r-1)+\frac{\sum_{i=1}^{r-1}Z_i-(r-1)}{\sqrt{r-1}}\sqrt{r-1}\right)} &= \left(\frac{s-1}{r}\right)^{\frac{1}{p}}\exp\{W_r\}\\ &= \left(\frac{s-1}{r}\right)^{\frac{1}{p}}(1+W_r+W_r^2+o_P(W_r^2)),\end{aligned}$$

as $r,s\to\infty$. Hence, using these asymptotic relations as well as the fact that $W_r^2 = o_P(1/\sqrt{r})$, as $r\to\infty$, we obtain

$$\begin{aligned}ps^{\frac{1}{2}-\frac{1}{p}}E &= \left\{\lambda^{1/p}-\left(\left[\frac{s-1}{r}\right]^{1/p}\right)^{\frac{1}{r}\left((r-1)+\frac{\sum_{i=1}^{r-1}Z_i-(r-1)}{\sqrt{r-1}}\sqrt{r-1}\right)}\right\}\left(\frac{r}{s}\right)^{1/p}p\sqrt{s}\\ &= \left\{\lambda^{1/p}-\left(\left[\frac{s-1}{r}\right]^{1/p}\right)^{\frac{1}{r}\left((r-1)+\frac{\sum_{i=1}^{r-1}Z_i-(r-1)}{\sqrt{r-1}}\sqrt{r-1}\right)}\right\}p\lambda^{\frac{1}{2}-\frac{1}{p}}\sqrt{r}, \qquad r,s\to\infty\\ &= p\lambda^{\frac{1}{2}-\frac{1}{p}}\left\{\left(\lambda^{\frac{1}{p}}-\left(\frac{s-1}{r}\right)^{\frac{1}{p}}\right)\sqrt{r}\right.\\ &\quad\left.-\left(\frac{s-1}{r}\right)^{\frac{1}{p}}\sqrt{r}W_r-\left(\frac{s-1}{r}\right)^{\frac{1}{p}}\sqrt{r}W_r^2+o_P(W_r^2)\sqrt{r}\right\}, \qquad r,s\to\infty\\ &= p\,\lambda^{\frac{1}{2}-\frac{1}{p}}\left\{-\left(\frac{s-1}{r}\right)^{\frac{1}{p}}\sqrt{r}W_r+o_P(1)\right\}, \qquad r,s\to\infty. \end{aligned} \tag{5.10}$$

Finally, combining (5.1), (5.4), (5.5), (5.6), (5.9), (5.10), we arrive at

$$\begin{aligned}\frac{1}{\sigma}\left(R_s-\left(\frac{s-1}{p}\right)^{\frac{1}{\hat{p}_{\mathrm{MOL}}}}R_r\right)&\\ \stackrel{d}{=}\ \frac{s^{\frac{1}{p}-\frac{1}{2}}}{p}&\left\{\frac{\sum_{i=1}^{r}Y_i-r}{\sqrt{r}}\left(\sqrt{\frac{r}{s}}-\sqrt{\lambda^2\frac{r}{s}}\right)+\frac{\sum_{i=r+1}^{s}Y_i-(s-r)}{\sqrt{s-r}}\sqrt{\frac{s-r}{s}}\right\}\\ -\ \frac{s^{\frac{1}{p}-\frac{1}{2}}}{p}&\left\{p\lambda^{\frac{1}{2}-\frac{1}{p}}\left(\frac{s-1}{r}\right)^{\frac{1}{p}}\sqrt{r}W_r+o_P(1)\right\},\qquad r,s\to\infty.\end{aligned}$$

Since $\sqrt{r}W_r \stackrel{d}{\to} \frac{1}{p}\ln(\lambda)\mathcal{N}(0,1)$, as $r,s\to\infty$, Theorem 2.8 in [4] as well as Slutsky's lemma yield that

$$\frac{p}{\sigma}s^{\frac{1}{2}-\frac{1}{p}}\left(R_s-\left(\frac{s-1}{p}\right)^{\frac{1}{\hat{p}_{\mathrm{MOL}}}}R_r\right)\stackrel{d}{\longrightarrow}\left(\sqrt{\frac{1}{\lambda}}-\sqrt{\lambda}\right)X_1+\sqrt{1-\frac{1}{\lambda}}X_2-\sqrt{\lambda}\ln(\lambda)X_3,$$

as $r,s\to\infty$, where X_1, X_2, $X_3 \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Since

$$\left(\frac{1}{\lambda}-\sqrt{\lambda}\right)X_1+\sqrt{1-\frac{1}{\lambda}}X_2-\sqrt{\lambda}\ln(\lambda)X_3\sim\mathcal{N}(0,\lambda+\lambda\ln^2(\lambda)-1),$$

the statement is proved. □

Remark 5.2. From the preceding result we obtain an approximate prediction interval for R_s with nominal coverage probability $1-\alpha$, which is given by

$$\pi^{(s)}_{\mathrm{MOLP}}\pm z_{1-\alpha/2}\frac{\sqrt{(s/r)(1+\ln^2(s/r))-1)}}{\alpha_s(\hat{\sigma}_{\mathrm{MLE}},\hat{p}_{\mathrm{MLE}})},$$

where $\hat{\sigma}_{\mathrm{MLE}}$ is the MLE of σ, $\hat{\sigma}_{\mathrm{MLE}}=R_r/r^{1/\hat{p}_{\mathrm{MLE}}}$ and $z_{1-\alpha/2}$ denotes the respective quantile of $\mathcal{N}(0,1)$. For the MLEs of σ and p we refer to [14].

6. CONCLUSION

For predicting future record values based on a sequence of observed upper record values with an underlying Weibull distribution, we derive two likelihood-based predictors, namely the maximum likelihood predictor and the maximum observed likelihood predictor. Expressions for the predictors are derived along with properties in terms of bias and mean squared error. The predictors are compared via Pitman's measure of closeness and their performance is examined in a simulation study.

ACKNOWLEDGMENTS

The authors would like to thank an Associate Editor for careful reading and valuable comments, which led to an improved presentation.

REFERENCES

[1] Arnold, B.C.; Balakrishnan, N. and Nagaraja, H.N. (1998). *Records*, John Wiley & Sons, New York.

[2] Askey, R.A. and Roy, R. (2010). *Gamma function*. In "NIST Handbook of Mathematical Functions" (F.W.J. Olver, D.W. Lozier, R.F. Boisvert, C.W. Clark, Eds.), New York, 135–144.

[3] Bar-Lev, S.K.; Lavi, I. and Reiser, B. (1992). Bayesian inference for the power law process, *Ann. Inst. Statist. Math.*, **44**(4), 623–639.

[4] Billingsley, P. (1999). *Convergence of Probability Measures*, John Wiley & Sons, New York.

[5] Chandler, K.N. (1952). The distribution and frequency of record values, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **14**(2), 220–228.

[6] Cramer, E. (2003). *Contributions to generalized order statistics*, Habilitation thesis, University of Oldenburg.

[7] Crow, L.H. (1982). Confidence interval procedures for the Weibull process with applications to reliability growth, *Technometrics*, **24**(1), 67–72.

[8] Davis, P.J. (1964). *Gamma function and related functions*. In "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables" (M. Abramowitz, and I.A. Stegun, Eds.), Washington, D.C., 253–266.

[9] Dziubdziela, W. and Kopociński, B. (1976). Limiting properties of the k-th record values, *Appl. Math. (Warsaw)*, **2**(15), 187–190.

[10] Engelhardt, M. and Bain, L.J. (1978). Prediction intervals for the Weibull process, *Technometrics*, **20**(2), 167–169.

[11] Gupta, R.C. and Kirmani, S.N.U.A. (1988). Closure and monotonicity properties of nonhomogeneous Poisson processes and record values, *Probab. Engrg. Inform. Sci.*, **2**(4), 475–484.

[12] Kaminsky, K.S. and Rhodin, L.S. (1985). Maximum likelihood prediction, *Ann. Inst. Stat. Math.*, **37**(3), 507–517.

[13] Kyparisi, J. and Singpurwalla, N.D. (1985). *Bayesian inference for the Weibull process with applications to assessing software reliability growth and predicting software failures*. In "Computer Science and Statistics: Proceedings of the Sixteenth Symposium on the Interface" (L. Billard, Ed.), Amsterdam, North Holland, 57–64.

[14] Lee, L. and Lee, S.K. (1978). Some results on inference for the Weibull process, *Technometrics*, **20**(1), 41–45.

[15] Nevzorov, V.B. (2001). *Records: Mathematical Theory*, American Mathematical Society, Providence, RI.

[16] Raqab, M.Z.; Alkhalfan, L.A.; Bdair, O.M. and Balakrishnan, N. (2019). Maximum likelihood prediction of records from 3-parameter Weibull distribution and some approximations, *J. Comput. Appl. Math.*, **356**, 118–132.

[17] Schmidt, K.D. (2011). *Maß und Wahrscheinlichkeit*, Springer-Verlag, Berlin.

[18] Volovskiy, G. (2018). *Likelihood-based Prediction in Models of Ordered Data*, PhD thesis, RWTH Aachen University.

[19] Volovskiy, G. and Kamps, U. (2020). Maximum observed likelihood prediction of future record values, *TEST*, **29**(4), 1072–1097.

[20] Wang, B.X. and Ye, Z.-S. (2015). Inference on the Weibull distribution based on record values, *Comput. Statist. Data An.*, **83**, 26–36.

[21] Wang, Y.-P. and Lu, Z.-Z. (2011). Bayesian inference and prediction analysis of the power law process based on a gamma prior distribution, *Comm. Statist. Simulation Comput.*, **40**(9), 1383–1401.

[22] Wang, Y.-P. and Lu, Z.-Z. (2015). Bayesian inference and prediction analysis of the power law process based on a natural conjugate prior, *J. Statist. Comput. Simul.*, **85**(5), 881–898.

[23] Zakerzadeh, H. and Jafari, A.A. (2015). Inference on the parameters of two Weibull distributions based on record values, *Stat. Methods Appl.*, **24**(1), 25–40.

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.