ISSN 1645-6726

REVSTAT

Statistical Journal

vol. 23 - n. 2 - April 2025

Instituto Nacional de Estatística
Statistics Portugal

1935 2025

REVSTAT — Statistical Journal, vol.23, n. 2 (April 2025)

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

Continues: Revista de Estatística = ISSN 0873-4275.

ISSN 1645-6726 ; e-ISSN 2183-0371

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 23, Number 2, April 2025, 161–187
https://doi.org/10.57805/revstat.v23i2.515

Stationary Underdispersed INAR(1) Models Based on the Backward Approach

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: August 2022 Revised: March 2023 Accepted: March 2023

Abstract:

- Most of the stationary first-order autoregressive integer-valued (INAR(1)) models in the literature have been developed using the idea of Binomial thinning. Two approaches have been adopted to establish the distributional properties of a stationary INAR(1) process: the forward approach and the backward approach. In the forward approach, the marginal distribution of the process is specified and an appropriate distribution for the innovation sequence is sought. Whereas in the backward setting, the roles are reversed. The common distribution of the innovation sequence is specified and the marginal distribution of the process is studied. In this article we focus on the backward approach. Our motivation is mainly theoretical, in the context of statistical distribution theory. We establish a number of basic properties of a specific infinite convolution of distributions on $\mathbb{Z}_+$. We then proceed to interpret our results in the context of stationary INAR(1) models whose innovation has a finite mean. As an application, we present new distributional properties for some stationary INAR(1) processes that show underdispersion, including two new models with q-series innovation distributions.

Keywords:

- *integer-valued time series; the Binomial thinning operator; q-series; Poissonian Binomial distribution; Heine distribution.*

AMS Subject Classification:

- 60E99, 62M10.

✉ Corresponding author

1. INTRODUCTION

The area of integer-valued time series has attracted a lot of interest in research and practice during the last 35 years. It started with the pioneering work of McKenzie (1985, 1986, 1988) and Al-Osh and Alzaid (1987), and Alzaid and Al-Osh (1990). Most of the existing models are based on the Binomial thinning operator of Steutel and van Harn (1979). Models under a variety of generalized thinning operators have been proposed by several authors. We refer to the review articles of McKenzie (2003) and Scotto *et al.* (2015) for details and additional references.

The Binomial thinning (cf. Steutel and van Harn, 1979) of a $\mathbb{Z}_+$-valued random variable X denoted by $\alpha \odot X$, is defined as

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

where $\alpha \in (0,1)$ and $\{Y_i\}$ is a sequence of independent identically distributed (iid) Bernoulli(α) rv's independent of X. The operation $\odot$ incorporates the discrete nature of the variates and acts as the analogue of the standard multiplication used in the continuous time series models.

Assume that $0 < \alpha < 1$, and $(\varepsilon_t, t \geq 1)$ is an iid sequence of $\mathbb{Z}_+$-valued rv's. A sequence $(X_t, t \geq 0)$ of $\mathbb{Z}_+$-valued rv's is said to be an INAR(1) process if

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

such that 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 Bernoulli(α) rv's, independent of $\{\varepsilon_t\}$, such that

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

In (1.2), $\{\varepsilon_t\}$ is referred to as the innovation sequence and α as the coefficient of the process $\{X_t\}$.

The main focus of this paper is on stationary INAR(1) models. The basic question of interest in this case is the choice of the marginal distribution of the process and that of its innovation.

Two approaches to this question prevail. One, which we will refer to as the forward approach, consists in selecting a specific marginal distribution for the process and then searching for the proper innovation distribution. The other approach, referred to as the backward approach, consists of the exact opposite: start out with the marginal distribution of the innovation sequence and then search for the proper marginal distribution of the process.

Both approaches have been widely used in the literature. For the forward approach, we refer to the review articles cited above and references therein. For models based on the backward approach, we cite a number of fairly recent articles: Jung *et al.* (2005), Pedeli and Karlis (2011), Weiß (2013), Schweer and Weiß (2014), Schweer and Wichelhaus (2015), Bourguignon and Vasconcellos (2015), and Kim and Lee (2017).

In the current work, we adopt the backward approach to develop INAR(1) models driven by (1.2) and whose innovation has finite mean. Our motivation is mainly theoretical, in the context of statistical distribution theory. We establish a number of basic properties of a specific infinite convolution of distributions on $\mathbb{Z}_+$. These results are then used to obtain most of the needed properties of the marginal and the conditional distributions of a stationary INAR(1) model. That is the object of Section 2. As an application, we present new distributional properties for some stationary INAR(1) models that show underdispersion, including two new INAR(1) models with q-series innovation distributions. More specifically, in Sections 3-7 we study in details the models whose innovation follow the Bernoulli distribution, the Binomial distribution, a q-series called the Poissonian Binomial distribution, the logarithmic distribution, and the Heine distribution, another q-series, respectively. We note that the INAR(1) models with Bernoulli, binomial and logarithmic innovations have been discussed in Bourguignon and Vasconcellos (2015). Our results provide additional properties for these processes. We also note the backward approach has been used by the authors in a related article (see Aly and Bouzar, 2021) that develops INAR(1) models with compound Poisson innovations.

We will use throughout the rest of this paper the notation $\overline{a} = 1 - a$, $a \in (0, 1)$.

We designate by $\mu_r^{(u)}(\kappa_r^{(u)})$ and $\mu_{[r]}^{(u)}(\kappa_{[r]}^{(u)})$ the r-th moment (cumulant) and the r-th factorial moment (factorial cumulant) of the pmf $\{u_r\}$, respectively. We will make use of the formulas (see Johnson *et al.*, 2005, Sections 1.2.7 and 1.2.8):

$$\mu_r^{(u)} = \sum_{j=1}^{r} S(r,j)\mu_{[j]}^{(u)} \quad \text{and} \quad \kappa_r^{(u)} = \sum_{j=1}^{r} S(r,j)\kappa_{[j]}^{(u)}, \tag{1.3}$$

where $S(r,j)$ are the Stirling numbers of the second kind defined as $S(0,0) = 1, S(0,k) = S(r,0) = 0$ and

$$S(r,j) = \frac{1}{j!}\sum_{k=0}^{j}(-1)^{j-k}\binom{j}{k}k^r. \tag{1.4}$$

2. BASIC RESULTS ON THE BACKWARD APPROACH

Our goal in this section is to establish several properties of a specific infinite convolution of distributions on $\mathbb{Z}_+$. We then proceed to interpret our results in the context of stationary INAR(1) models whose innovation has finite mean.

Theorem 2.1. *Let $\Psi(z)$ be the pgf of a pmf $\{f_r\}$. Assume $\Psi'(1) < \infty$, i.e., $\{f_r\}$ has finite mean. Then, the function*

$$\varphi(z) = \prod_{i=0}^{\infty}\Psi(1 - \alpha^i + \alpha^i z) \tag{2.1}$$

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

$$\varphi(z) = \varphi(1 - \alpha + \alpha z)\Psi(z), \qquad z \in [0,1]. \tag{2.2}$$

Proof: First, we recall some basic results on pgf's (we refer to Feller, 1968). For $k \geq 0$, let $q_k = \sum_{i=k+1}^{\infty} f_i$ be the sequence of the tail probabilities corresponding to $\{f_k\}$ and let

$$Q(z) = \sum_{k=0}^{\infty} q_k z^k,$$

be the generating function of $\{q_k\}$. We have $1 - \Psi(z) = (1-z)Q(z)$, $z \in [0,1]$, and $Q(1) = \sum_{k=0}^{\infty} q_k = \sum_{k=0}^{\infty} k f_k = \Psi'(1) < \infty$. Define $h_i(z) = 1 - \Psi(1 - \alpha^i + \alpha^i z)$. It follows that $h_i(z) = \alpha^i (1-z) Q(1 - \alpha^i + \alpha^i z)$. Noting that Q is increasing over $[0,1], 0 \leq 1 - z \leq 1$, and $Q(1)$ is finite, we conclude that $0 \leq h_i(z) \leq Q(1)\alpha^i$ and $\sum_{i=n+1}^{\infty} h_i(z) \leq Q(1) \sum_{i=n+1}^{\infty} \alpha^i$. This implies $\sum_{i=n+1}^{\infty} h_i(z)$ converges uniformly to 0 over the interval $[0,1]$. For every $n \geq 0$, define

$$\varphi_{n+1}(z) = \prod_{i=0}^{n} \Psi(1 - \alpha^i + \alpha^i z), \tag{2.3}$$

which can be rewritten as $\varphi_{n+1}(z) = \prod_{i=0}^{n} (1 - h_i(z))$. It follows by Theorem 1, p. 381, in Knopp (1990), that the sequence $\{\varphi_{n+1}(z)\}$ converges uniformly over the interval $[0,1]$ to

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

Next, we show that $\lim_{z \uparrow 1} \varphi(z) = 1$. Define $r_n(z) = \prod_{i=n+1}^{\infty} \Psi(1 - \alpha^i + \alpha^i z)$ and let $\delta > 0$ be arbitrary. By the uniform convergence of $\{\varphi_{n+1}(z)\}$ to $\varphi(z)$, there exists a positive integer $N(\delta)$ such that for any $n > N(\delta)$, $\sup_{z \in [0,1]} |r_n(z) - 1| < \delta$. Note that $\varphi_{n+1}(\cdot)$ of (2.3) satisfies $\varphi_{n+1}(1) = 1$ and $\varphi_{n+1}(z) \leq 1$. Since

$$|\varphi(z) - 1| = |\varphi_{n+1}(z)(r_n(z) - 1) + \varphi_{n+1}(z) - 1|,$$

it follows that for any $n > N(\delta)$, $|\varphi(z) - 1| \leq \delta + |\varphi_{n+1}(z) - 1|$, which in turn implies

$$\limsup_{z \uparrow 1} |\varphi(z) - 1| = \limsup_{z \uparrow 1} (1 - \varphi(z)) \leq \delta + \liminf_{z \uparrow 1} (1 - \varphi_{n+1}(z)) \leq \delta.$$

Since $\varphi(z)$ is the limit of the sequence of pgf's $\{\varphi_{n+1}(z)\}$, we conclude that $\varphi(z)$ is a pgf by the Continuity Theorem. Equation (2.2) is easily shown to hold. □

Theorem 2.2. *Let $\{f_r\}$ be a pmf with finite mean and with pgf $\Psi(z)$. Let $\{p_r\}$ be the pmf with pgf $\varphi(z)$ of (2.1) and let $\{f_r^{(i)}\}$ be the pmf with pgf $\Psi(1 - \alpha^i + \alpha^i z)$, $i \geq 0$. The following assertions are true:*

1. *$f_r^{(0)} = f_r$ and*

$$f_r^{(i)} = \begin{cases} f_0 + \sum_{n=1}^{\infty} (1 - \alpha^i)^n f_n, & \text{if } r = 0 \\ \alpha^{ir} \sum_{n=r}^{\infty} \binom{n}{r} f_n (1 - \alpha^i)^{n-r}, & \text{if } r \geq 1. \end{cases} \tag{2.4}$$

2.

$$p_r = \lim_{k \to \infty} \left(f^{(0)} * f^{(1)} * \cdots * f^{(k-1)} \right)_r, \tag{2.5}$$

where $f^{(0)} * f^{(1)} * \cdots * f^{(k-1)}$ designates the k-factor convolution of the pmf's $\{f_r^{(0)}\}$, $\{f_r^{(1)}\}, \ldots, \{f_r^{(k-1)}\}$.

3. *Assume the factorial cumulant generating function (fcgf)* $\ln\Psi(1+t)=\sum_{r=1}^{\infty}\kappa_{[r]}^{(f)}\frac{t^r}{r!}$ *of the pmf* $\{f_r\}$ *exists for* $|t|<\rho_0$ *for some* $\rho_0>0$*. Then, for every* $r\ge 1$, $\kappa_{[r]}^{(p)}$ *and* $\kappa_r^{(p)}$ *are finite and are given by (cf. (1.3)–(1.4))*

$$\kappa_{[r]}^{(p)}=\frac{\kappa_{[r]}^{(f)}}{1-\alpha^r}\quad\text{and}\quad\kappa_r^{(p)}=\sum_{j=1}^{r}S(r,j)\frac{\kappa_{[j]}^{(f)}}{1-\alpha^j}. \tag{2.6}$$

4. *If* $\{f_r\}$ *has a finite second cumulant, then the mean* $\mu^{(p)}$*, the variance* $(\sigma^{(p)})^2$ *and the dispersion index of* $I^{(p)}$ *of* $\{p_r\}$ *are obtained in terms of their* $\{f_r\}$ *counterparts,* $\mu^{(f)}$, $(\sigma^{(f)})^2$ *and* $I^{(f)}$ *as follows:*

$$\mu^{(p)}=\frac{\mu^{(f)}}{1-\alpha},\quad(\sigma^{(p)})^2=\frac{(\sigma^{(f)})^2+\alpha\mu^{(f)}}{1-\alpha^2}\quad\text{and}\quad I^{(p)}=1+\frac{I^{(f)}-1}{1+\alpha}. \tag{2.7}$$

Proof: The proof of (2.4) is straightforward. Since $\varphi(z)=\lim_{k\to\infty}\prod_{i=0}^{k-1}\varphi_k(z)$, with $\varphi_k(z)$ of (2.3), we obtain (2.5) by the Continuity Theorem and (2.4). Since the fcgf $\ln\Psi(1+t)$ of the pmf $\{f_r\}$ exists for $|t|<\rho_0$, we have $\ln\Psi(1+\alpha^i t)=\sum_{r=1}^{\infty}\alpha^{ir}\kappa_{[r]}^{(f)}\frac{t^r}{r!}$. It follows by (2.1) that $\ln\varphi(1+t)=\sum_{i=0}^{\infty}\ln\Psi(1+\alpha^i t)$. One can show by a standard argument that the series $\ln\varphi(1+t)$ converges uniformly in the interval $|t|\le\rho$ for every $0<\rho<\rho_0$. Therefore, by Weierstrass Theorem, p. 430 in Knopp (1990), we have

$$\ln\varphi(1+t)=\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\alpha^{ir}\kappa_{[r]}^{(f)}\frac{t^r}{r!}=\sum_{r=1}^{\infty}\frac{\kappa_{[r]}^{(f)}}{1-\alpha^r}\frac{t^r}{r!}\qquad(|t|<\rho_0),$$

proving the first part of (2.6). The second part of (2.6) is deduced from (1.3)–(1.4). The formulas in (2.7) follow from (2.2). □

Remark 2.1. We make a number of useful remarks.

1. Equations (2.6) and (2.7) are known (see Weiß, 2013). We note that as a function of α the dispersion index $I^{(p)}$ is increasing and concave down if the innovation distribution is underdispersed.
2. As noted in Weiß (2013), it is easily seen from (2.7), that $\{p_r\}$ of (2.5) is underdispersed (i.e., $(\sigma^{(p)})^2<\mu^{(p)}$) if and only if $\{f_r\}$ is underdispersed.
3. There are no simple formulas linking the r-th moment $\mu_r^{(p)}$ and the r-th factorial moment $\mu_{[r]}^{(p)}$ of $\{p_r\}$ to their $\{f_r\}$ counterparts. However, if either $\kappa_{[r]}^{(p)}$ or $\kappa_r^{(p)}$ can be calculated for every $r\ge 1$, then one can compute $\mu_r^{(p)}$ and $\mu_{[r]}^{(p)}$ recursively using standard formulas that link moments and cumulants (see Johnson *et al.*, 2005, Sections 1.2.7 and 1.2.8, and Smith, 1995).

Next, we interpret Theorem 2.1 and Theorem 2.2 in the context of INAR(1) modeling.

If the INAR(1) process $\{X_t\}$ of (1.2) is stationary, then its marginal pgf $\varphi_X(z)$ and the common pgf $\Psi(z)$ of the innovation sequence $\{\varepsilon_t\}$ must satisfy the functional equation (2.2) with $\varphi(z)=\varphi_X(z)$.

The backward approach we have adopted in this paper translates as follows: one chooses a pgf $\Psi(\cdot)$ and solve for $\varphi_X(\cdot)$ that satisfies (2.2). It can be shown that in this case $\varphi_X(z) = \lim_{n\to\infty} \varphi_n(z)$, with $\varphi_n(z)$ of (2.3), provided that the limit exists and is a pgf.

The backward approach leads to the following existence theorem for a stationary INAR(1) process.

Theorem 2.3. *Let $\alpha \in (0,1)$. Any pgf $\Psi(z)$ such that $\Psi'(1) < \infty$ gives rise to a stationary INAR(1) process $\{X_t\}$ defined on some probability space $(\Omega, \mathcal{F}, P)$ and driven by equation (1.2). Its marginal pgf is*

$$\varphi_X(z) = \prod_{i=0}^{\infty} \Psi(1 - \alpha^i + \alpha^i z). \tag{2.8}$$

Proof: Since $\Psi'(z) < 1$, by Theorem 2.1 $\varphi_X(z)$ is a pgf that satisfies equation (2.2). By Proposition 2.1 in Bouzar and Jayakumar (2008), there exists a stationary INAR(1) process $\{X_t\}$ on some probability space $(\Omega, \mathcal{F}, P)$ such that its marginal distribution and that of its innovation sequence $\{\varepsilon_t\}$ have respective pgf's $\varphi_X(z)$ and $\Psi(z)$. □

The following additional results (we refer to Al-Osh and Alzaid (1987) and McKenzie (1988)) are needed in the sequel. An INAR(1) model driven by (1.2) is necessarily a homogeneous Markov chain with the 1-step transition probabilities,

$$P(X_t = k | X_{t-1} = l) = \sum_{j=0}^{\min(l,k)} \binom{l}{j} \alpha^j (1-\alpha)^{l-j} P(\varepsilon = k - j). \tag{2.9}$$

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

$$X_{t+k} \overset{d}{=} \alpha^k \circ X_t + \sum_{j=1}^{k} \alpha^{j-1} \circ \varepsilon_{t+k-j+1} \tag{2.10}$$

and the k-step autocorrelation of $\{X_t\}$ is

$$\mathrm{Corr}(X_t, X_t + k) = \alpha^k. \tag{2.11}$$

It follows from (2.10) that the conditional pgf of X_{t+k} given X_t satisfies

$$\varphi_{X_{t+k}|X_t}(z) = \left(1 - \alpha^k + \alpha^k z\right)^{X_t} \times \prod_{i=0}^{k-1} \Psi(1 - \alpha^i + \alpha^i z). \tag{2.12}$$

Therefore, given $X_t = n$, the distribution of X_{t+k} is the convolution of a Binomial(n, α^k) distribution and the pmf $\left\{\left(f^{(0)} * f^{(1)} * \cdots * f^{(k-1)}\right)_r\right\}$ of Theorem 2.2.

Remark 2.2. It is a well known fact that the INAR(1) process (1.2) is a branching process with a Bernoulli(α) offspring distribution and an immigration sequence of iid random variables with common pgf $\Psi(z)$. It follows by Theorem in Heathcote (1965) that Theorem 2.3 holds under the weaker condition $\sum_{k=0}^{\infty} q_k (k+1)^{-1} < \infty$, where $\{q_k\}$ is the sequence of tail probabilities of $\Psi(z)$. For a more general result, we refer to Theorem 2, p. 264, in Athreya and Ney (2004).

3. STATIONARY INAR(1) MODELS WITH BERNOULLI INNOVATIONS

In this section and subsequent ones, we describe the properties of the marginal and conditional distributions of stationary INAR(1) processes with specific innovation sequences. We obtain useful representations of the marginal pgf's of these models as well as formulas for moments and cumulants of their marginal distributions.

We start out with the case of Bernoulli innovations. First, we recall a definition.

Let $q, c \in (0,1)$ and $m \geq 1$. Kemp (1987) (see also Johnson *et al.*, 2005, p. 467) introduced and studied the Poissonian Binomial(m,q,c) distribution as the distribution of a finite convolution of Bernoulli(cq^i) distributions, $i = 0,1,2,...,m-1$ with pgf

$$\Psi(z) = \prod_{i=0}^{m-1} \left(1 - cq^i(1-z)\right) \tag{3.1}$$

and pmf

$$q_r(m,q,c) = \sum_{k=r}^{m} (-1)^{k-r} \binom{k}{r} c^k q^{\binom{k}{2}} \prod_{l=0}^{k-1} \frac{1-q^{m-l}}{1-q^{l+1}} \quad , r = 0,1,...,m. \tag{3.2}$$

We will expand more on this distribution in Section 5.

The main result of this section follows next. Its proof is long and is deferred to the Appendix.

Theorem 3.1. *Let $\{X_t\}$ be the stationary INAR(1) process driven by (1.2) and with a Bernoulli(p) innovation sequence for some $p \in (0,1)$. Then:*

1. *The marginal pmf $\{p_r\}$ of $\{X_t\}$ is the weak limit of Poissonian Binomial(n,α,p) (see (3.1) and (3.2)) as $n \to \infty$ and is given by*

$$p_r = \lim_{n\to\infty} q_r(n,\alpha,p) = \sum_{k=r}^{\infty} (-1)^{k-r} \binom{k}{r} \frac{p^k \alpha^{\binom{k}{2}}}{\prod_{l=1}^{k}(1-\alpha^l)}, r \geq 0; \tag{3.3}$$

2. *The tail probabilities $P(X_t \geq r) = \sum_{j=r}^{\infty} p_j$ of X_t are obtained by the formula*

$$P(X_t \geq r) = \sum_{k=r}^{\infty} (-1)^{k-r} \binom{k-1}{r-1} \frac{p^k \alpha^{\binom{k}{2}}}{\prod_{l=1}^{k}(1-\alpha^l)}, \quad r \geq 1; \tag{3.4}$$

3. *The marginal pgf $\varphi_X(z)$ of $\{X_t\}$ admits two useful representations:*

$$\varphi_X(z) = 1 + \sum_{n=1}^{\infty} \frac{p^n (z-1)^n \alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)} \tag{3.5}$$

and

$$\varphi_X(z) = \exp\left\{ -\sum_{n=1}^{\infty} \frac{p^n}{n(1-\alpha^n)} (1-z)^n \right\}. \tag{3.6}$$

Additional properties of $\{X_t\}$ are given next.

By (2.9), the 1-step transition probability is given by

$$P(X_t = k|X_{t-1} = l) = \begin{cases} 0, & k > l+1 \\ p\alpha^{k-1}, & k = l+1 \\ \alpha^{k-1}\overline{\alpha}^{l-k}\left\{p\binom{l}{k-1}\overline{\alpha} + \overline{p}\binom{l}{k}\alpha\right\}, & k \le l \end{cases}. \tag{3.7}$$

By (2.12), the conditional pgf of X_{t+k} given X_t satisfies

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

Therefore, given $X_t = n$, the distribution of X_{t+k} is the convolution of a Binomial(n, α^k) distribution and the Poissonian Binomial(k, α, p) distribution of (3.2).

Next, we derive the factorial moments $(\mu_{[r]}^{(p)}, r \ge 1)$ of X_t. Using the version (3.5) of $\varphi_X(z)$, we deduce that

$$\varphi_X(1+t) = 1 + \sum_{r=1}^{\infty} \frac{r!p^r\alpha^{\binom{r}{2}}}{\prod_{i=1}^{r}(1-\alpha^i)} \cdot \frac{t^r}{r!}.$$

Since the series converges everywhere, the factorial moments (the coefficients of $t^r/r!$) and therefore the moments of X_t (by (1.3)–(1.4)) of all orders are finite and are given by

$$\mu_{[r]}^{(p)} = \frac{r!p^r\alpha^{\binom{r}{2}}}{\prod_{i=1}^{r}(1-\alpha^i)} \quad \text{and} \quad \mu_r^{(p)} = \sum_{j=1}^{r} S(r,j)\frac{j!p^j\alpha^{\binom{j}{2}}}{\prod_{i=1}^{j}(1-\alpha^i)} \qquad (r \ge 1). \tag{3.8}$$

By (2.7), the mean, the variance and the index of dispersion of X_t are

$$\mu_X = \frac{p}{1-\alpha}, \quad \sigma_X^2 = \frac{p(1-p)+\alpha p}{1-\alpha^2} \quad \text{and} \quad I_X = 1 - \frac{p}{1+\alpha}.$$

As expected, the marginal distribution of $\{X_t\}$ is underdispersed. We note that I_X is decreasing and linear affine in p and increasing and concave down in α (see Figure 1).

By (3.6), the fcgf of X_t is given by

$$\ln \varphi_X(1+t) = \sum_{r=1}^{\infty} \frac{(-1)^{r+1}(r-1)!p^r}{(1-\alpha^r)} \cdot \frac{t^r}{r!}.$$

Since the series above converges everywhere, the factorial cumulants and the cumulants of X_t of all orders are finite and given by (applying (1.3)–(1.4)):

$$\kappa_{[r]}^{(p)} = (-1)^{r+1}\frac{(r-1)!p^r}{(1-\alpha^r)} \quad \text{and} \quad \kappa_r^{(p)} = \sum_{j=0}^{r} S(r,j)(-1)^{j+1}\frac{(j-1)!p^j}{(1-\alpha^j)} \quad (r \ge 1). \tag{3.9}$$

Remark 3.1. We note that if the innovation sequence $\{\varepsilon_t\}$ has the Power-Law distribution of the first kind $(PL_1(\lambda, p))$, i.e., $\varepsilon_t \sim$ Poisson$(\lambda) *$ Bernoulli(p), $0 < p < 1$, then its marginal distribution will result from the convolution of a Poisson$(\frac{\lambda}{1-\alpha})$ and the pmf $\{p_r\}$ of (3.3) in Theorem 3.1. The $PL_1(\lambda, p)$ law was discussed in Section 2.3 of Weiß (2013). Additional distributional properties of this law such as moments and cumulants, can be obtained from Theorem 3.1 and subsequent results in this section.

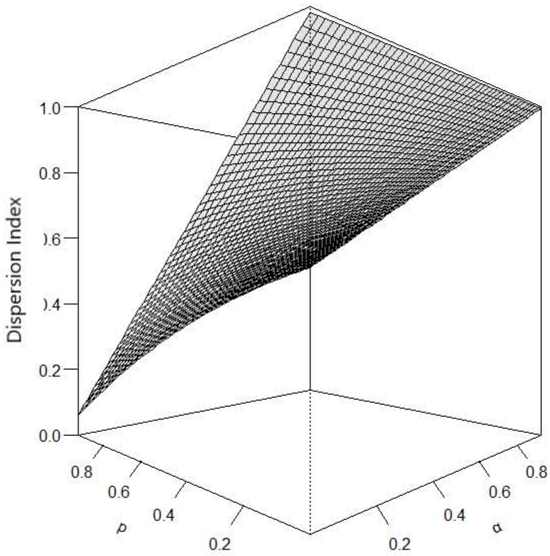

Figure 1: Variance-mean ratio of the marginal distribution of an INAR(1) process with a Bernoulli innovation.

4. STATIONARY INAR(1) MODELS WITH BINOMIAL INNOVATIONS

The treatment is essentially similar to the Bernoulli case ($m = 1$). We summarize the main results with minimal justifications for the most part.

Theorem 4.1. *Let $\{X_t\}$ be the stationary INAR(1) process driven by (1.2) and with a Binomial(m, p) innovation sequence for some positive integer m and some $p \in (0, 1)$. Then:*

1. *The marginal pmf $\{p_r\}$ of $\{X_t\}$ is the m-fold convolution of the marginal distribution (3.3) of the INAR(1) process with a Bernoulli(p) innovation.*
2. *The marginal pgf $\varphi_X(z)$ of $\{X_t\}$ admits two representations:*

$$\varphi_X(z) = \left[1 + \sum_{n=1}^{\infty} \frac{p^n (z-1)^n \alpha^{\binom{n}{2}}}{\prod_{l=1}^{n} (1-\alpha^l)}\right]^m \tag{4.1}$$

and

$$\varphi_X(z) = \exp\left\{ -m \sum_{n=1}^{\infty} \frac{p^n}{n(1-\alpha^n)} (1-z)^n \right\}. \tag{4.2}$$

Proof: Straightforward. We omit the details. □

We proceed to give additional properties of $\{X_t\}$.

By (2.9), the 1-step transition probability of $\{X_t\}$ is

$$P(X_t = k | X_{t-1} = l) = p^k \overline{p}^{m-k} \overline{\alpha}^l \sum_{j=\max(k-m,0)}^{\min(l,k)} \binom{l}{j} \binom{m}{k-j} \left(\frac{\alpha \overline{p}}{p \overline{\alpha}}\right)^j, \quad k \leq l+m. \tag{4.3}$$

By (2.12), the conditional pgf of X_{t+k} given X_t satisfies

$$\varphi_{X_{t+k}|X_t}(z) = \left(1-\alpha^k+\alpha^k z\right)^{X_t} \times \left[\prod_{i=0}^{k-1}(1-p\alpha^i(1-z))\right]^m.$$

Therefore, the conditional distribution of X_{t+k} given $X_t = n$ is the convolution of a Binomial(n,α^k) distribution and the m-fold convolution of the Poissonian Binomial(k,α,p) distribution of (3.2).

Since the power series expansion of $\varphi_X(1+t)$ for $\varphi_X(z)$ of (4.1) is not easily computable, we proceed to derive simpler recurrence formulas for the factorial moments $(\mu^{(p)}_{[r]}, r \geq 1)$ of X_t by using instead the representation (4.2).

Let

$$\phi(z) = \sum_{n=1}^{\infty} b_n(1-z)^n, \quad b_n = \frac{p^n}{n(1-\alpha^n)}. \tag{4.4}$$

The series (4.4) converges uniformly over the interval $(0,1)$ due to the fact that for every $n \geq 1$ and $z \in (0,1)$, $b_n(1-z)^n \leq b_n$, and that $\sum_{n=1}^{\infty} b_n$ converges. It follows that $\phi'(z)$ and subsequent higher order derivatives exist and converge uniformly over $(0,1)$ (see Knopp, 1990). The r-th derivative of $\phi(z)$ admits the representation

$$\phi^{(r)}(z) = (-1)^r \sum_{n=r}^{\infty} \frac{p^n}{1-\alpha^n}\frac{(n-1)!}{(n-r)!}(1-z)^{n-r} \qquad (r \geq 1). \tag{4.5}$$

Uniform convergence allows for the interchange of limit (as $z \uparrow 1$) and summation in (4.5). Hence,

$$\phi^{(r)}(1) = (-1)^r \frac{(r-1)!p^r}{1-\alpha^r} \qquad (r \geq 1). \tag{4.6}$$

Since $\ln \varphi_X(z) = -m\phi(z)$, it follows that $\varphi'_X(z) = -m\varphi_X(z)\phi'(z)$. An induction argument shows that the r^{th} derivative, $\varphi^{(r)}_X(z)$, of $\varphi_X(z)$ can be obtained by the following forward recursion (with $\varphi^{(0)}_X(z) = \varphi_X(z)$ and $\binom{0}{0} = 1$):

$$\varphi^{(r)}_X(z) = -m \sum_{j=0}^{r-1} \binom{r-1}{j} \varphi^{(j)}_X(z)\phi^{(r-j)}(z). \tag{4.7}$$

Therefore, the factorial moments $\mu^{(p)}_{[r]} = \varphi^{(r)}_X(1)$, $r \geq 1$, are finite and satisfy the recurrence relation (with $\mu^{(p)}_{[0]} = 1$),

$$\mu^{(p)}_{[r]} = -m \sum_{j=0}^{r-1} \binom{r-1}{j} \mu^{(p)}_{[j]}\phi^{(r-j)}(1) \qquad (r \geq 1). \tag{4.8}$$

By (2.7), the mean, the variance and the index of dispersion of X_t are

$$\mu_X = \frac{mp}{1-\alpha}, \quad \sigma^2_X = \frac{mp(1+\alpha-p)}{1-\alpha^2} \quad \text{and } I_X = 1 - \frac{p}{1+\alpha},$$

implying the marginal of $\{X_t\}$ is underdispersed.

We note that the dispersion indexes for INAR(1) processes with Bernoulli and binomial innovations are identical. However, as pointed out by the referee, the additional parameter m of the model with Binomial innovations gives further flexibility for the parameterization of the INAR(1) model. For example, we may estimate α using the sample autocorrelation function of order one, $ACF(1)$ (cf. (2.11)), and $\lambda = mp$ using the sample mean $\widehat{\lambda}$. Thus, the remaining degree of freedom, p in $\left\{\widehat{\lambda}, \frac{\widehat{\lambda}}{2}, \frac{\widehat{\lambda}}{3}, ...\right\}$ can be used to adjust the dispersion index.

The moments $(\mu_r^{(p)}, r \geq 1)$ of X_t are finite and can be obtained from their factorial counterparts via (4.8) and equations (1.3)–(1.4).

Finally, and similarly to the Bernoulli case, the factorial cumulants and the cumulants of X_t are obtained via the pgf representation (4.2) and equations (1.3)–(1.4):

$$(4.9) \qquad \kappa_{[r]}^{(p)} = \frac{m(-1)^{r+1}(r-1)!p^r}{1-\alpha^r} \quad \text{and} \quad \kappa_r^{(p)} = m\sum_{j=0}^{r} S(r,j)(-1)^{j+1}\frac{(j-1)!p^j}{1-\alpha^j} \quad (r \geq 1).$$

5. STATIONARY INAR(1) MODELS WITH POISSONIAN BINOMIAL INNOVATIONS

In this section, we develop a stationary INAR(1) process with a Poissonian Binomial innovation sequence with pgf and pmf given respectively in (3.1) and (3.2), for some positive integer m and some real numbers $q, c \in (0,1)$. This distribution belongs to the family of discrete q-series distributions with finite range. It results from the convolution of m independent Bernoulli(p_j) distributions , $j = 1, 2, ..., m$, where the p_j's vary according to the geometric progression $p_j = cq^{j-1}$. For more on q-distributions, we refer to the monograph Charalambides (2016).

We recall for further reference that a Poissonian Binomial(m, q, c) is underdispersed with mean, variance, and dispersion index given by (see Kemp, 1987)

$$(5.1) \qquad \mu_\varepsilon = \frac{(1-q^m)c}{1-q}, \quad \sigma_\varepsilon^2 = \frac{(1-q^m)c}{1-q} - \frac{(1-q^{2m})c^2}{1-q^2}, \quad \text{and} \quad I_\varepsilon = 1 - \frac{(1+q^m)c}{1+q}.$$

Theorem 5.1. *Let $\{X_t\}$ be the stationary INAR(1) process driven by (1.2) and with a Poissonian Binomial(m, q, c) innovation sequence for some positive integer m and some real numbers $q, c \in (0,1)$.*

1. *The marginal pgf $\varphi_X(z)$ of $\{X_t\}$ admits the following representations:*

$$(5.2) \qquad \varphi_X(z) = \prod_{j=0}^{m-1}\left[1 + \sum_{n=1}^{\infty}\frac{(cq^j)^n(z-1)^n\alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)}\right]$$

and

$$(5.3) \qquad \varphi_X(z) = \exp\left\{-\sum_{n=1}^{\infty}\frac{1-q^{mn}}{1-q^n}\frac{c^n}{n(1-\alpha^n)}(1-z)^n\right\}.$$

2. *The marginal pmf* $\{p_r\}$ *of* $\{X_t\}$ *is the convolution of the pmf's* $(\{p_r^{(j)}\}, 0 \le j \le m-1)$,

$$p_r = (p^{(0)} * p^{(1)} * \cdots * p^{(m-1)})_r \quad (r \ge 0), \tag{5.4}$$

where

$$p_r^{(j)} = \sum_{k=r}^{\infty} (-1)^{k-r} \binom{k}{r} \frac{(cq^j)^k \alpha^{\binom{k}{2}}}{\prod_{l=1}^{k}(1-\alpha^l)}, \quad r \ge 0. \tag{5.5}$$

Proof: Let $\Psi(z)$ be the pgf of the Poissonian Binomial(m, q, c) distribution as given in (3.1). Then,

$$\Psi(1 - \alpha^i + \alpha^i z) = \prod_{j=0}^{m-1} (1 + c\alpha^i q^j (z-1)),$$

which is the pgf of a Poissonian Binomial$(m, q, c\alpha^i)$. By Theorem 2.1, the marginal pgf $\varphi_X(z)$ is

$$\varphi_X(z) = \prod_{j=0}^{m-1} \prod_{i=0}^{\infty} (1 + c\alpha^i q^j (z-1)).$$

Noting that $\prod_{i=0}^{\infty} (1 + c\alpha^i q^j (z-1))$ is the marginal pgf of a stationary INAR(1) process with Bernoulli(cq^j) innovations, representations (5.2) and (5.3) follow from (3.5) and (3.6), respectively. By Theorem 3.1 and (3.3), for each $j \ge 0$, the pmf with pgf

$$\varphi_j(z) = 1 + \sum_{n=1}^{\infty} \frac{(cq^j)^n (z-1)^n \alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)}$$

is $\{p_r^{(j)}\}$ of (5.5). Therefore, part 3 and (5.5) follow from (5.2). □

Some additional properties of $\{X_t\}$ are presented next.

By (2.9), the 1-step transition probability of $\{X_t\}$ is given by

$$P(X_t = k | X_{t-1} = l) = \sum_{j=\max(k-m,0)}^{\min(l,k)} \binom{l}{j} \alpha^j (1-\alpha)^{l-j} q_{k-j}(m, q, c), \quad k \le l + m. \tag{5.6}$$

By (2.12), the conditional pgf of X_{t+k} given X_t satisfies

$$\varphi_{X_{t+k}|X_t}(z) = \left(1 - \alpha^k + \alpha^k z\right)^{X_t} \times \prod_{j=0}^{m-1} \left[\prod_{i=0}^{k-1} (1 - (cq^j)\alpha^i (1-z))\right].$$

Therefore, the conditional distribution of X_{t+k} given $X_t = n$ is the convolution of a Binomial(n, α^k) distribution and the Poissonian Binomial(k, α, cq^j) distributions, $j = 0, 1, ..., m-1$.

By (5.2), the power series expansion of $\varphi_X(1+t)$ exists but is not easily computable. We proceed as in the Binomial case (Section 4) to derive the factorial moments $(\mu_{[r]}^{(p)}, r \geq 1)$ of X_t via the representation (5.3) of $\varphi_X(z)$ and a recurrence relation.

Let

$$\phi_1(z) = -\ln \varphi_X(z) = \sum_{n=1}^{\infty} \frac{1-q^{mn}}{1-q^n} \frac{c^n}{n(1-\alpha^n)}(1-z)^n. \tag{5.7}$$

The argument we used to derive (4.5)–(4.8) in Section 4 carries over almost verbatim. We state the main steps without further explanations. The r-th derivative of $\phi_1(z)$ admits the representation

$$\phi_1^{(r)}(z) = (-1)^r \sum_{n=r}^{\infty} \frac{(1-q^{mn})c^n}{(1-q^n)(1-\alpha^n)} \frac{(n-1)!}{(n-r)!}(1-z)^{n-r} \qquad r \geq 1. \tag{5.8}$$

Hence,

$$\phi_1^{(r)}(1) = (-1)^r \frac{(1-q^{mr})c^r}{(1-q^r)(1-\alpha^r)}(r-1)! \qquad (r \geq 1). \tag{5.9}$$

Since $\ln \varphi_X(z) = -\phi_(z)$, the r^{th} derivative, $\varphi_X^{(r)}(z)$, of $\varphi_X(z)$ can be obtained by the following forward recursion (with $\varphi_X^{(0)}(z) = \varphi_X(z)$ and $\binom{0}{0} = 1$):

$$\varphi_X^{(r)}(z) = -\sum_{j=0}^{r-1} \binom{r-1}{j} \varphi_X^{(j)}(z)\phi_1^{(r-j)}(z). \tag{5.10}$$

The factorial moments $\mu_{[r]}^{(p)} = \varphi_X^{(r)}(1)$, $r \geq 1$, are finite and satisfy the recurrence relation (with $\mu_{[0]}^{(p)} = 1$),

$$\mu_{[r]}^{(p)} = -\sum_{j=0}^{r-1} \binom{r-1}{j} \mu_{[j]}^{(p)} \phi_1^{(r-j)}(1) \qquad (r \geq 1). \tag{5.11}$$

The moments of X_t, $\mu_r^{(p)} = E(X_t^r)$, $r \geq 1$, are finite and can be obtained from their factorial counterparts via equations (1.3)–(1.4).

By (2.7) and (5.1), the marginal distribution of $\{X_t\}$ is underdispersed with mean, variance and dispersion index given by

$$\mu_X = \frac{(1-q^m)c}{(1-\alpha)(1-q)}, \quad \sigma_X^2 = \frac{(1-q^m)c}{(1-\alpha)(1-q)} - \frac{(1-q^{2m})c^2}{(1-\alpha^2)(1-q^2)}$$

and

$$I_X = 1 - \frac{(1+q^m)c}{(1+\alpha)(1+q)}.$$

We note that I_X is increasing in α and m $(m \geq 2)$ and decreasing in c. Moreover, it is concave down in q with concavity becoming more pronounced as c increases (see Figure 2).

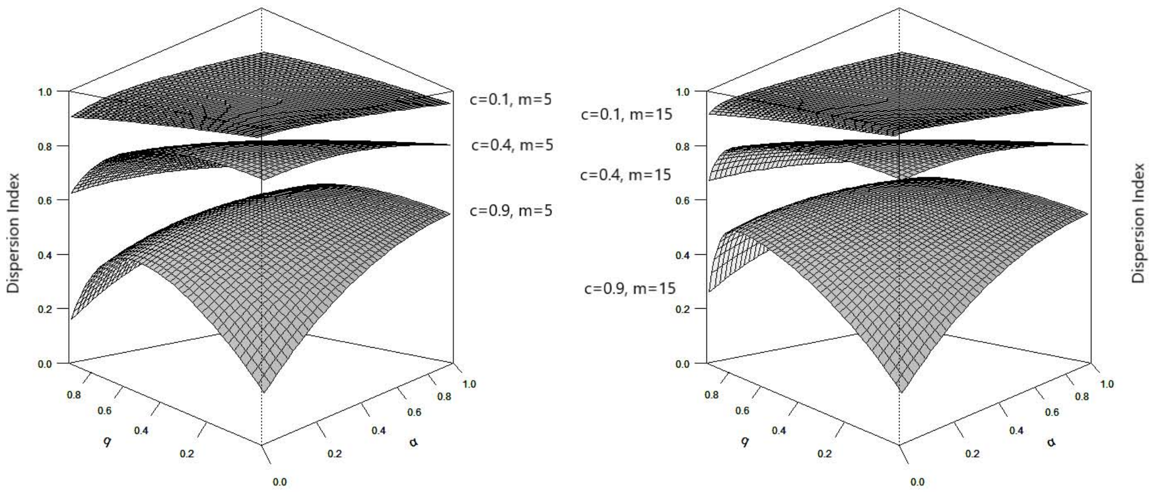

Figure 2: Variance-mean ratio of the marginal distribution of an INAR(1) process with a Poissonian Binomial innovation.

Similarly to the Bernoulli and the Binomial cases (Sections 3 and 4, respectively), the factorial cumulants and the cumulants of X_t are obtained in straightforward fashion from the pgf representation (5.3) and equations (1.3)–(1.4):

$$\kappa_{[r]}^{(p)} = \frac{(-1)^{r+1}(r-1)!(1-q^{mr})c^r}{(1-q^r)(1-\alpha^r)} \qquad (r \geq 1) \tag{5.12}$$

and

$$\kappa_r^{(p)} = \sum_{j=0}^{r} S(r,j) \frac{(-1)^{j+1}(j-1)!(1-q^{mj})c^j}{(1-q^j)(1-\alpha^j)}, \qquad r \geq 1. \tag{5.13}$$

6. STATIONARY INAR(1) MODELS WITH LOGARITHMIC INNOVATIONS

We revisit in this section the INAR(1) process with a logarithmic(p) distribution introduced by Bourguignon and Vasconcellos (2015). Most of the discussion will focus on the underdispersion version of the process.

We start out by recalling a few facts about the logarithmic distribution (see Johnson *et al.*, 2005). The pmf of the logarithmic(p) distribution is given by $f_r = \frac{p^r}{-r \ln \overline{p}}$, $r \geq 1$, where $p \in (0,1)$. Its pgf, mean, variance and dispersion index are given respectively by

$$\Psi(z) = \frac{\ln(1-pz)}{\ln \overline{p}}, \tag{6.1}$$

$$\mu^{(f)} = -\frac{p}{\overline{p}\ln\overline{p}}, \quad (\sigma^{(f)})^2 = -\frac{p(p+\ln\overline{p})}{(\overline{p}\ln\overline{p})^2} \quad \text{and} \quad I^{(f)} = \frac{p+\ln\overline{p}}{\overline{p}\ln\overline{p}}. \tag{6.2}$$

Note that the logarithmic distribution is underdispersed if $p < 1 - 1/e$, equidispersed if $p = 1 - 1/e$ and overdispersed if $p > 1 - 1/e$.

The factorial moments and moments of $\{f_r\}$ are respectively (cf. (1.3)–(1.4))

$$(6.3) \qquad \mu_{[r]}^{(f)} = -\frac{p^r(r-1)!}{(1-p)^r \ln \overline{p}} \quad \text{and} \quad \mu_r^{(f)} = -\frac{1}{\ln \overline{p}} \sum_{j=1}^{r} S(r,j) \frac{p^j(j-1)!}{(1-p)^j} \qquad (r \geq 1).$$

We will also refer to the logarithmic-with-zeros(c,p) distribution, with $c, p \in (0,1)$, that arises as a two-mixture of the Dirac measure δ_0 sitting at 0 and the logarithmic(p) distribution with respective mixing probabilities c and $1-c$ (see Johnson *et al.*, 2005, Sections 7.1 and 8.2). Its pgf is $P(z) = c + (1-c)\ln[(1-pz)/\ln \overline{p}]$.

Lemma 6.1. *Let $p \in (0,1)$ and let $\{f_r\}$ be the pmf of a logarithmic(p) distribution with pgf $\Psi(z)$ of (6.1). Then for every $i \geq 0$, the pmf, $\{f_r^{(i)}\}$ of (2.4) with pgf $\Psi(1-\alpha^i+\alpha^i z)$ is a logarithmic-with-zeros(b_i, q_i) distribution with $b_i = 1 - \frac{\ln \overline{q}_i}{\ln \overline{p}}$ and $q_i = \frac{p\alpha^i}{1-p(1-\alpha^i)}$ $(q_0 = p, b_0 = 0)$, i.e.,*

$$(6.4) \qquad f_r^{(i)} = b_i \delta_0(\{r\}) + (1-b_i) \frac{q_i^r}{-r \ln \overline{q}_i}.$$

Noting $f_r^{(0)} = f_r$, the k-factor convolution of the pmf's $\{f_r^{(0)}\}, \{f_r^{(1)}\}, \ldots, \{f_r^{(k-1)}\}, k \geq 2$, is a finite mixture of convolutions of logarithmic distributions, namely,

$$(6.5) \qquad \Big(f^{(0)} * f^{(1)} * \cdots * f^{(k-1)}\Big)_r = C_{I,0}\, g_r^{(0)} + \sum_{l=1}^{k-1} \sum_{\mathbf{j} \in \mathbf{J}_l} C_{\mathbf{j},l} \Big(g^{(0)} * g^{(j_1)} * g^{(j_2)} * \cdots * g^{(j_l)}\Big)_r,$$

where $\{g_r^{(j)}\}$ is the pmf of the logarithmic(q_j), $I = \{1, 2, \ldots, k-1\}$, $\mathbf{J}_l$ is the collection of ordered l-tuples $\mathbf{j} = (j_1, j_2, \ldots, j_l)$, $1 \leq j_1 < j_2 < \cdots < j_l \leq k-1$ and $\mathbf{j}_u = \{j_1, j_2, \ldots, j_l\}$ is the corresponding unordered l-tuple. The mixing probabilities are

$$(6.6) \qquad C_{I,0} = \prod_{j=1}^{k-1} b_j \quad \text{and} \quad C_{\mathbf{j},l} = \Big(\prod_{j \in I \setminus \mathbf{j}_u} b_j\Big)\Big(\prod_{h=1}^{l} (1 - b_{j_h})\Big).$$

Proof: If $i = 0$, (6.4) is true since $\{f_r^{(0)}\} = \{f_r\}$. Assume $i \geq 1$. By (2.4),

$$f_0^{(i)} = \frac{-1}{\ln \overline{p}} \sum_{n=1}^{\infty} \frac{\big(p(1-\alpha^i)\big)}{n} = \frac{\ln(1-p(1-\alpha^i))}{\ln \overline{p}},$$

and for $r \geq 1$,

$$f_r^{(i)} = -\frac{(p\alpha^i)^r}{r \ln \overline{p}} \sum_{n=r}^{\infty} \binom{n-1}{r-1} \big(p(1-\alpha^i)\big)^{n-r}.$$

Using the power series expansion $(1-t)^{-r-1} = \sum_{n=r}^{\infty} \binom{n}{r} t^{n-r}$, with $t = p(1-\alpha^i)$, it follows that

$$f_r^{(i)} = -\frac{1}{r \ln \overline{p}} \left[\frac{p\alpha^i}{1-p(1-\alpha^i)}\right]^r \qquad (r \geq 1).$$

Setting $q_i = \frac{p\alpha^i}{1-p(1-\alpha^i)}$, it is easily verified that $f_r^{(i)}$ satisfies (6.4). The second part of the Lemma and equations (6.5) and (6.6) are proved by a tedious but straightforward induction argument. The details are omitted. □

Theorem 6.1. *Let $\{X_t\}$ be the stationary INAR(1) process driven by (1.2) and with a logarithmic(p) innovation sequence for some $p \in (0,1)$. Then:*

- **(i)** *The marginal distribution $\{p_r\}$ of $\{X_t\}$ is the weak limit, as $k \to \infty$, of the sequence of pmf's $\big(f^{(0)} * f^{(1)} * \cdots * f^{(k-1)}, k \geq 1\big)$, where $f^{(0)} * f^{(1)} * \cdots * f^{(k-1)}$ is described by equations (6.5) and (6.6).*
- **(ii)** *The marginal pgf $\varphi_X(z)$ of $\{X_t\}$ admits the representation*

$$\varphi_X(z) = \prod_{i=0}^{\infty}\Big[1 - \frac{1}{\ln \overline{p}} \cdot \ln \frac{\overline{q}_i}{1 - q_i z}\Big] \qquad (0 < z \leq 1), \tag{6.7}$$

with $p_0 = \varphi_X(0) = 0$.

Proof: Part (i) is a direct consequence of (2.5) and Lemma 6.1. For part (ii), (6.7) follows from (2.8) and the fact that when $z = 0$ the first factor in (6.7) is equal to 0 . □

Next, we provide additional properties of $\{X_t\}$, some of which appeared in Bourguignon and Vasconcellos (2015).

By (2.9), the 1-step transition probability of $\{X_t\}$ is given by

$$P(X_t = k | X_{t-1} = l) = -\frac{p^k}{\ln \overline{p}} \sum_{j=0}^{\min(l,k-1)} \binom{l}{j} \frac{(\alpha/p)^j \, \overline{\alpha}^{l-j}}{k-j}, \quad k, l \geq 1.$$

By (2.12), the conditional pgf of X_{t+k} given X_t satisfies

$$\varphi_{X_{t+k}|X_t}(z) = \Big(1 - \alpha^k + \alpha^k z\Big)^{X_t} \times \prod_{i=0}^{k-1}\Big[1 - \frac{1}{\ln \overline{p}} \cdot \ln \frac{\overline{q}_i}{1 - q_i z}\Big].$$

Therefore, given $X_t = n$, the distribution of X_{t+k} is the convolution of a Binomial(n, α^k) distribution and the finite mixture of convolutions of logarithmic distributions described by (6.5) and (6.6).

By (2.7) and (6.3), the mean μ_X, the variance σ_X^2 and the dispersion index I_X of the marginal distribution of $\{X_t\}$ are given by

$$\mu_X = -\frac{p}{\overline{p}(1-\alpha)\ln \overline{p}}, \quad \sigma_X^2 = -\frac{p(p + \ln \overline{p} + \alpha(\overline{p}\ln \overline{p})^2)}{(\overline{p}\ln \overline{p})^2(1-\alpha^2)} \quad \text{and} \quad I_X = 1 + \frac{p(1 + \ln \overline{p})}{\overline{p}(1+\alpha)\ln \overline{p}}. \tag{6.8}$$

Note that the distribution of X_t is underdispersed if and only if $p < 1 - 1/e$. The graph of I_X restricted to that range is given in Figure 3 below. I_X is increasing and concave down in α and decreasing and concave up in p.

Unlike the previously encountered models, the representations of the functions $\varphi_X(z)$ and $\ln \varphi_X(z)$ of the distribution of X_t are too complex to lead to manageable formulas for moments and cumulants of X_t. Instead, we proceed as in Weiß (2013), Section 4.2, and use a number of recurrence formulas that will compute these quantities in the following order:

1. Compute the r-th cumulant $\kappa_r^{(f)}$ of ε_t using the formula, derived in Smith (1995),

$$\kappa_r^{(f)} = \mu_r^{(f)} - \sum_{i=1}^{r-1} \binom{r-1}{i} \kappa_{r-i}^{(f)} \mu_i^{(f)}, \tag{6.9}$$

along with (6.3) (recall $\kappa_1^{(f)} = \mu^{(f)}$ and $\kappa_2^{(f)} = (\sigma^{(f)})^2$, cf. (6.2)).

2. Compute the r-th factorial cumulant $\kappa_{[r]}^{(f)}$ of ε_t using the formula (see Johnson *et al.*, 2005, Sections 1.2.7 and 1.2.8)

$$\kappa_{[r]}^{(f)} = \sum_{j=0}^{r} s(r,j) \kappa_j^{(f)}, \tag{6.10}$$

where $s(r,j)$ is the Stirling number of the first kind satisfying the recurrence relation

$$s(r+1,j) = s(r,j-1) - ns(r,j), \tag{6.11}$$

with $s(n,0) = 0$ and $s(1,1) = 1$.

3. Compute the r-th factorial cumulant, $\kappa_{[r]}^{(p)}$, and cumulant, $\kappa_r^{(p)}$, of X_t using (2.6).
4. Use the formulas in Johnson *et al.* (2005) to compute the moments $\{\mu_r\}$ of X_t, equation (1.252), p. 54, and its factorial moments $\{\mu_{[r]}\}$, equation (1.244), p. 53.

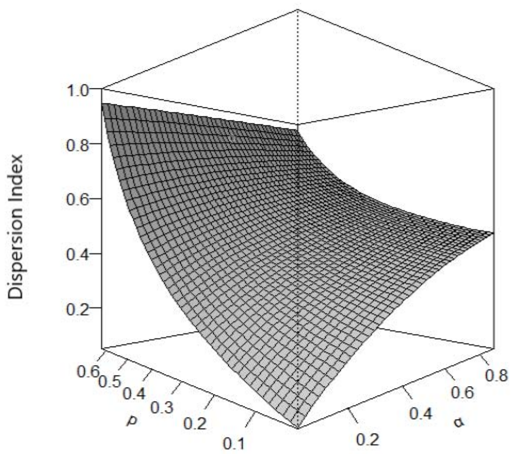

Figure 3: Variance-mean ratio of the marginal distribution of an INAR(1) process with an underdispersed logarithmic innovation.

Remark 6.1. Since both the innovation and the marginal distribution of the INAR(1) process in this section have support in $\mathbb{N}^* = \{1, 2, 3, ...\}$, it may be more suitable to adopt a measure of dispersion different from the one relative to the Poisson distribution. We briefly discuss two approaches and submit the relevant graphs as an illustration. For a random variable Y with support in $\mathbb{N}^*$ such that $0 < P(Y = 1) < 1$, we suggest using dispersion indexes of the zero-shifted distribution of Y, that is the distribution of $Y - 1$. We consider two such indexes: the standard one, we denote by I_{Y-1} (as in previous sections) and the one introduced in Abid *et al.* (2021), we denote by $I_{Y-1}^{(geo)}$, relative to the zero-shifted geometric distribution (with the usual interpretation of under/equi/over-dispersion). The formulas are as follows:

$$I_{Y-1} = \frac{Var(Y)}{E(Y)-1} \quad \text{and} \quad I_{Y-1}^{(geo)} = \frac{I_Y}{E(Y)-1}. \tag{6.12}$$

It is easily seen that for any $p \in (0,1)$, the zero-shifted version of the logarithmic(p) distribution is overdispersed relative to both the Poisson and the zero-shifted geometric distributions. Applying the formulas (6.12), along with (6.8), to the zero-shifted version of the marginal distribution of the stationary INAR(1) with a logarithmic innovation, we obtain the following graphs for the two indexes:

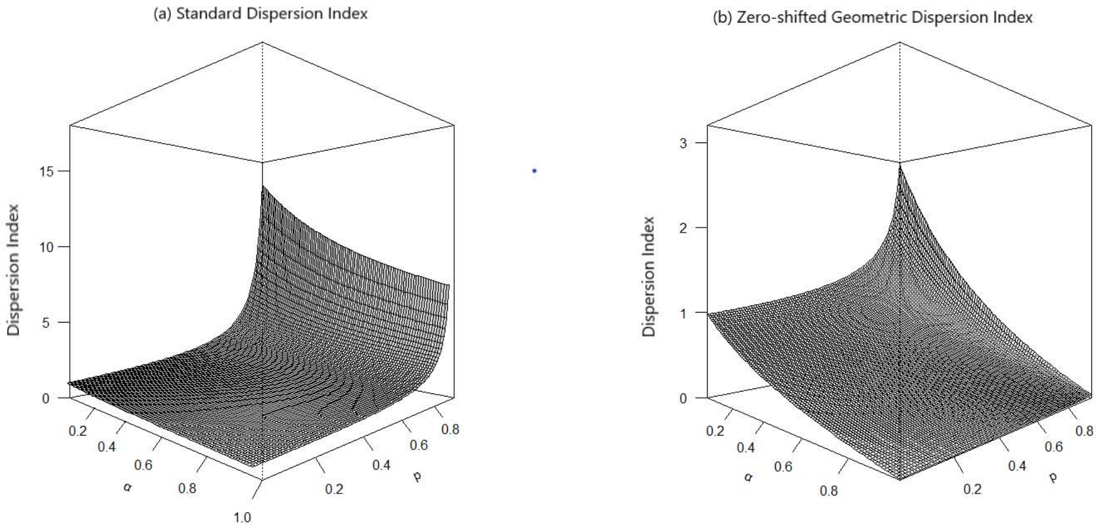

Figure 4: Two variance-mean ratios of the zero-shifted marginal distribution of an INAR(1) process with a Logarithmic innovation.

Both graphs show increase in p and decrease in α. Note also that for every $\alpha \in (0,1)$ and $p \in (0,1)$, I_{X-1} exhibits all three dispersion states, whereas $I_{X-1}^{(geo)}$ shows underdispersion for every $\alpha \geq 0.58$ and $p \in (0,1)$.

7. STATIONARY INAR(1) PROCESSES WITH HEINE INNOVATIONS

A distribution on $\mathbb{Z}_+$ is said to have the Heine distribution (Heine(λ, q)) with parameters $\lambda > 0$ and $q \in (0,1)$ if its pgf and pmf are respectively given by

$$\Psi(z) = \prod_{j=0}^{\infty} \big(1 - \beta_j + \beta_j z\big) \tag{7.1}$$

and

$$f_0 = \prod_{j=0}^{\infty} (1-\lambda q^j)^{-1} \quad \text{and} \quad f_r = \frac{\lambda^r q^{r(r-1)/2}}{\prod_{l=1}^{r}(1-q^l)} f_0 \quad (r \geq 1), \tag{7.2}$$

where $\beta_j = \frac{\lambda q^j}{1+\lambda q^j}$ for $j \geq 0$.

The Heine distribution was introduced by Benkherouf and Bather (1988). It is a q-series distribution with infinite support. Many of its properties were studied in Kemp (1992). More details can be found in these references and in Johnson *et al.* (2005), Section 10.8.2. The Heine distribution is underdispersed and its mean and variance are

$$\mu = \sum_{r=0}^{\infty} \frac{\lambda q^r}{1+\lambda q^r} \quad \text{and} \quad \sigma^2 = \sum_{r=0}^{\infty} \frac{\lambda q^r}{(1+\lambda q^r)^2}. \tag{7.3}$$

We recall a few facts about double infinite products. Let $\{a_{mn}\}$ be a double sequence. The double infinite product $\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}(1+a_{ij})$ is defined as the limit of the double sequence $P_{mn} = \prod_{i=0}^{m}\prod_{j=0}^{n}(1+a_{ij})$ as $m, n \to \infty$. If $\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}|a_{ij}| < \infty$, then the double infinite product $\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}(1+a_{ij})$ converges. Moreover, if $\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}(1+a_{ij})$ and the iterated infinite products

$$\prod_{i=0}^{\infty}\Big[\prod_{j=0}^{\infty}(1+a_{ij})\Big] \quad \text{and} \quad \prod_{j=0}^{\infty}\Big[\prod_{i=0}^{\infty}(1+a_{ij})\Big]$$

are all convergent, then they necessarily have the same value.

The main result of this section is given next. Its proof is deferred to the Appendix.

Theorem 7.1. *Let $\{X_t\}$ be the stationary INAR(1) process driven by (1.2) and with a Heine(λ, q) innovation sequence for some $\lambda > 0$ and $0 < q < 1$. Then the marginal pgf $\varphi_X(z)$ of $\{X_t\}$ admits the following representations:*

1.

$$\varphi_X(z) = \prod_{j=0}^{\infty}\Big[1 + \sum_{n=1}^{\infty} \frac{\beta_j^n (z-1)^n \alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)}\Big], \tag{7.4}$$

where β_j is as in (7.1).

2.

$$\varphi_X(z) = \exp\Big\{-\sum_{n=1}^{\infty} \frac{B_n}{n(1-\alpha^n)}(1-z)^n\Big\} \tag{7.5}$$

with $B_n = \sum_{j=0}^{\infty} \beta_j^n$, $n \geq 1$.

3. *The marginal pmf $\{p_r\}$ of $\{X_t\}$ is*

$$p_r = \lim_{k\to\infty} (q^{(0)} * q^{(1)} * \cdots * q^{(k-1)})_r \quad (r \geq 0), \tag{7.6}$$

where

$$q_r^{(j)} = \sum_{k=r}^{\infty} (-1)^{k-r} \binom{k}{r} \frac{\beta_j^k \alpha^{\binom{k}{2}}}{\prod_{l=1}^{k}(1-\alpha^l)}, \quad r \geq 0. \tag{7.7}$$

Next, we discuss additional properties of $\{X_t\}$.

The 1-step transition probability can be obtained from (2.9). Given there are no notable simplifications of the formulas, we omit the details and refer to the following discussion by setting $k = 1$.

By (2.12), the conditional distribution of X_{t+k} given $X_t = n$ results from the convolution of $k+1$ distributions, namely a Binomial(n, α^k) distribution and the distributions $(\{g_r^{(i)}\}, 0 \le i \le k-1)$ defined as follows:

$$g_r^{(i)} = \alpha^{ir} \sum_{l=0}^{\infty} \binom{r+l}{r} (1-\alpha)^l f_{r+l},$$

where $\{f_r\}$ is the pmf of the Heine(λ, q) distribution (7.2).

The factorial cumulants and the cumulants of X_t are obtained in straightforward fashion from the pgf representation (7.5) and equations (1.3)–(1.4):

$$\kappa_{[r]}^{(p)} = (-1)^{r+1} \frac{(r-1)!B_r}{1-\alpha^r} \quad \text{and} \quad \kappa_r^{(p)} = \sum_{j=0}^{r} S(r,j)(-1)^{j+1} \frac{(j-1)!B_j}{(1-\alpha^j)} \qquad (r \ge 1). \tag{7.8}$$

By (2.7) and (7.3), the mean, the variance and the index of dispersion of X_t are given by

$$\mu_X = \frac{1}{1-\alpha} \sum_{r=0}^{\infty} \frac{\lambda q^r}{1+\lambda q^r}, \quad \sigma_X^2 = \frac{1}{1-\alpha^2} \sum_{r=0}^{\infty} \frac{\lambda q^r}{1+\lambda q^r} \left([1+\lambda q^r]^{-1} + \alpha \right)$$

and

$$I_X = \sum_{r=0}^{\infty} \frac{\lambda q^r}{1+\lambda q^r} \left([1+\lambda q^r]^{-1} + \alpha \right) \Big/ \left[(1+\alpha) \sum_{r=0}^{\infty} \frac{\lambda q^r}{1+\lambda q^r} \right].$$

Since the Heine distribution is underdispersed, the INAR(1) process with a Heine innovation is underdispersed. We note that I_X is increasing in α and q and decreasing in λ (see Figure 5).

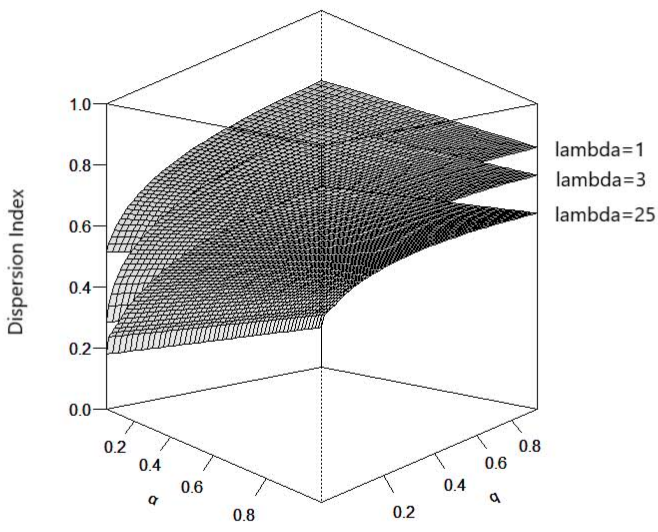

Figure 5: Variance-mean ratio of the marginal distribution of an INAR(1) process with a Heine innovation.

Similarly to the way (6.9)–(6.10) were derived, the moments and factorial moments of X_t can be computed using the formulas

$$\mu_r^{(p)} = \sum_{j=0}^{r-1} \binom{r-1}{j} \kappa_{r-j}^{(p)} \mu_j \quad \text{and} \quad \mu_{[r]}^{(p)} = \sum_{j=0}^{r} s(r,j) \mu_j^{(p)}, \tag{7.9}$$

with initial conditions $\mu_0^{(p)} = 1$ and $\mu_1^{(p)} = \kappa_1^{(p)}$, and where $\{s(r,j)\}$ are the Stirling numbers of the first kind of (6.11).

In turn, the factorial moments $\mu_{[r]}^{(p)}, r \geq 1$, of X_t can be obtained via the formula (see Johnson *et al.*, 2005, Section 1.2.7):

$$\mu_{[r]}^{(p)} = \sum_{j=0}^{r} s(r,j) \mu_j^{(p)}, \tag{7.10}$$

where $\{s(r,j)\}$ are the Stirling numbers of the first kind of (6.11).

8. CONCLUSION

In this article we formalized a theoretical approach to study the distributional properties of a stationary INAR(1) process based on Binomial thinning when the innovation distribution is known. We established a number of basic properties of a specific infinite convolution of distributions on $\mathbb{Z}_+$ and interpreted our results in the context of stationary INAR(1) models whose innovation has a finite mean. As an application, we presented new distributional properties for some stationary INAR(1) models that show underdispersion, including two new INAR(1) models with q-series innovation distributions. Simulations and statistical analysis for some of these models will be the object of the authors future work. Another direction of research would be to extend the results in this paper by using more general thinning operators.

A. APPENDIX

This section is devoted to the proof of Theorem 3.1 and Theorem 7.1. We first establish a Lemma.

Lemma A.1. *Assume $n \geq 2$ and $a_i \in (0,1)$ for $i = 0,1,2,...,n-1$. Then,*

1.

$$\prod_{i=0}^{n-1}(1-a_i) = 1 + \sum_{k=1}^{n}(-1)^k\Big[\sum_{0\leq j_1<j_2<\cdots<j_k\leq n-1}\prod_{l=1}^{k} a_{j_l}\Big]. \tag{A.1}$$

2.

$$\sum_{0\leq j_1<j_2<\cdots<j_k\leq n-1}\alpha^{j_1}\alpha^{j_2}\cdots\alpha^{j_k} = \alpha^{\binom{k}{2}}\prod_{l=0}^{k-1}\frac{1-\alpha^{n-l}}{1-\alpha^{l+1}} \tag{A.2}$$

for every $k \in \{1,...,n\}$.

Proof: (1) follows by a straightforward induction.

(2) We also proceed by induction. The result is trivially true for $n = 2$ (forces $k = 1$). Assume the assertion is true up to n. Equation (A.2) holds for $n+1$ and $k = n+1$, as in this case

$$\sum_{0\leq j_1<j_2<\cdots<j_{n+1}\leq n}\alpha^{j_1}\alpha^{j_2}\cdots\alpha^{j_k} = \alpha^{\sum_{k=0}^{n}k} = \alpha^{\binom{n+1}{2}} = \alpha^{\binom{n+1}{2}}\prod_{l=0}^{n}\frac{1-\alpha^{n+1-l}}{1-\alpha^{l+1}}.$$

Assume now $k \in \{1,2,...,n\}$. Setting $J = (j_1, j_2, ..., j_k) \in \mathbb{N}^k$, it is clear that

$$\{J \in \mathbb{N}^k : 0 \leq j_1 < j_2 < \cdots < j_k \leq n\} = A \cup B,$$

where $A = \{J \in \mathbb{N}^k : 0 \leq j_1 < j_2 < \cdots < j_k \leq n-1\}$ and $B = \{J \in \mathbb{N}^k : 0 \leq j_1 < j_2 < \cdots < j_{k-1} \leq n-1, j_k = n\}$. Therefore,

$$\sum_{0\leq j_1<j_2<\cdots<j_k\leq n}\prod_{l=1}^{k}\alpha^{j_l} = \sum_{J\in A}\prod_{l=1}^{k}\alpha^{j_l} + \sum_{J\in B}\alpha^n\prod_{l=1}^{k-1}\alpha^{j_l}.$$

Using the induction hypothesis, it follows that

$$\sum_{J\in A}\prod_{l=1}^{k}\alpha^{j_l} = \alpha^{\binom{k}{2}}\prod_{l=0}^{k-1}\frac{1-\alpha^{n-l}}{1-\alpha^{l+1}}$$

and

$$\sum_{J\in B}\alpha^n\prod_{l=1}^{k-1}\alpha^{j_l} = \alpha^n\alpha^{\binom{k-1}{2}}\prod_{l=0}^{k-2}\frac{1-\alpha^{n-l}}{1-\alpha^{l+1}},$$

which implies

$$\sum_{0\leq j_1<j_2<\cdots<j_k\leq n}\prod_{l=1}^{k}\alpha^{j_l} = \frac{\prod_{l=0}^{k-2}(1-\alpha^{n-l})\Big[(1-\alpha^{n-k+1})\alpha^{\binom{k}{2}} + (1-\alpha^k)\alpha^n\alpha^{\binom{k-1}{2}}\Big]}{\prod_{l=0}^{k-1}(1-\alpha^{l+1})}.$$

Now, noting that $\binom{k}{2} = \binom{k-1}{2} + k - 1$, it is easily seen that

$$(1-\alpha^{n-k+1})\alpha^{\binom{k}{2}} + (1-\alpha^k)\alpha^n\alpha^{\binom{k-1}{2}} = \alpha^{\binom{k}{2}}(1-\alpha^{n+1}).$$

Therefore, (A.2) holds for $n+1$. □

Proof of Theorem 3.1: Let $\{X_t\}$ be the stationary INAR(1) process with a Bernoulli(p) innovation sequence. By Theorem 2.3 and (2.8), its marginal pgf is

$$\varphi_X(z) = \prod_{i=0}^{\infty}(1 - p\alpha^i(1-z)). \tag{A.3}$$

Since $\varphi_X(z) = \lim_{n\to\infty} \prod_{i=0}^{n-1}(1-p\alpha^i(1-z))$, we conclude by the continuity theorem that the marginal pmf $\{p_r\}$ of $\{X_t\}$ is the weak limit of a sequence of Poissonian Binomial distributions of (3.1) and (3.2), with $m = n$, $q = \alpha$ and $c = p$. Let $r \geq 0$. We define a purely atomic measure, we denote meas_r, on $\mathbb{N}_r = \{r, r+1, r+2, ...\}$ and its power set $\mathcal{P}(\mathbb{N}_r)$ as follows:

$$\mathrm{meas}_r(\{k\}) = \frac{p^k\alpha^{\binom{k}{2}}}{\prod_{l=1}^{k}(1-\alpha^l)}, \quad (k \geq r), \tag{A.4}$$

with $\mathrm{meas}_0(\{0\}) = 1$. It is clear that $\sum_{k=r}^{\infty}\mathrm{meas}_r(\{k\}) < \infty$. Therefore, meas_r is a finite measure. Define now the sequence of functions $\{f_n(\cdot)\}$ on $\mathbb{N}_r$ by

$$f_n(k) = \begin{cases} (-1)^{k-r}\binom{k}{r}\prod_{l=0}^{k-1}(1-\alpha^{n-l}) & \text{if } k = r, r+1, ..., n \\ 0 & \text{if } k > n. \end{cases}$$

Define $h(k) = \binom{k}{r}$ on $\mathbb{N}_r$. It is clear that $|f_n(k)| \leq h(k)$ (recall $\alpha \in (0,1)$) and that $\sum_{k=r}^{\infty} h(k)\,\mathrm{meas}_r(\{k\}) < \infty$ by the ratio test). Moreover, for every $k \in \mathbb{N}_r$,

$$f(k) = \lim_{n\to\infty} f_n(k) = (-1)^{k-r}\binom{k}{r}.$$

Rewriting $p_r^{(n)}$ in terms of the discrete integral of $f_n(k)$ on the measure space $(\mathbb{N}_r, \mathcal{P}(\mathbb{N}_r), \mathrm{meas}_r)$ and calling on the dominated convergence theorem, we have

$$p_r = \lim_{n\to\infty}\int_{\mathbb{N}_r} f_n(k)\,\mathrm{meas}_r(dk) = \int_{\mathbb{N}_r} f(k)\,\mathrm{meas}_r(dk),$$

which is precisely (3.3) and thus part 1 of the Theorem is established. To show part 2, note that

$$P(X_t \geq r) = \sum_{j=r}^{\infty}\sum_{k=j}^{\infty}(-1)^{k-j}\binom{k}{j}\frac{p^k\alpha^{\binom{k}{2}}}{\prod_{l=1}^{k}(1-\alpha^l)}.$$

Since the double series above converges absolutely, interchanging summations is allowed, leading to

$$P(X_t \geq r) = \sum_{k=r}^{\infty}\left(\sum_{j=r}^{k}(-1)^{k-j}\binom{k}{j}\right)\frac{p^k\alpha^{\binom{k}{2}}}{\prod_{l=1}^{k}(1-\alpha^l)}.$$

We have by induction on k that $\sum_{j=r}^{k}(-1)^{k-j}\binom{k}{j} = (-1)^{k-r}\binom{k-1}{r-1}$, thus establishing (3.4).

For part 3, we note first that $\varphi_X(z)$ of (A.3) can be rewritten as

$$\varphi_X(z) = \exp\left\{\sum_{i=0}^{\infty} \ln(1 - p\alpha^i(1 - z))\right\}. \tag{A.5}$$

The representation (3.6) of $\varphi_X(z)$ follows by way of the power series expansion of

$$-\ln(1 - x) = \sum_{n=1}^{\infty} x^n/n, \quad 0 \le x < 1$$

applied to $x = p\alpha^i(1 - z)$ in (A.5).

To prove (3.5), we first note that by letting $a_i = p\alpha^i(1 - z)$ in (A.1) and using (A.2), we obtain the following expression for $\varphi_n(z)$ of (2.3):

$$\varphi_{n-1}(z) = 1 + \sum_{k=1}^{n} p^k (z - 1)^k \alpha^{\binom{k}{2}} \prod_{l=0}^{k-1} \frac{1 - \alpha^{n-l}}{1 - \alpha^{l+1}} \tag{A.6}$$

and therefore,

$$\varphi_X(z) = \lim_{n\to\infty}\left[1 + \sum_{k=1}^{n} (z - 1)^k \prod_{l=0}^{k-1}(1 - \alpha^{n-l}) \frac{p^k \alpha^{\binom{k}{2}}}{\prod_{l=1}^{k}(1 - \alpha^l)}\right]. \tag{A.7}$$

We proceed as in the proof of (3.3). We define a sequence of functions $g_n(k)$ on the finite measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \text{meas}_0)$, where meas_0 is defined in (A.4):

$$g_n(k) = \begin{cases} 1 & \text{if } k = 0 \\ (z - 1)^k \prod_{l=0}^{k-1}(1 - \alpha^{n-l}) & \text{if } 1 \le k \le n \\ 0 & \text{if } k > n. \end{cases}$$

It is easily seen that $|g_n(k)| \le 1$ (recall $\alpha \in (0, 1)$ and $z \in [0, 1]$) and that

$$g(k) = \lim_{n\to\infty} g_n(k) = \begin{cases} 1 & \text{if } k = 0 \\ (z - 1)^k & \text{if } k \ge 1. \end{cases}$$

Rewriting (A.7) in terms of the discrete integral on the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \text{meas}_0)$ and calling on the Dominated Convergence Theorem, we have

$$\varphi_X(z) = \lim_{n\to\infty} \int_{\mathbb{N}} g_n(k)\, \text{meas}_0(dk) = \int_{\mathbb{N}} g(k)\, \text{meas}_0(dk),$$

which is precisely (3.5). □

Proof of Theorem 7.1: First, we note that $0 < \beta_j < 1$ for any $j \ge 0$. Moreover, for any $n \ge 1$,

$$B_n = \sum_{j=0}^{\infty} \frac{\lambda^n (q^n)^j}{(1 + \lambda q^j)^n} \le \frac{\lambda^n}{1 - q^n} < \infty. \tag{A.8}$$

The pgf $\Psi(z)$ of the innovation sequence of $\{X_t\}$ (cf. (7.1)) yields

$$\Psi(1-\alpha^i+\alpha^i z)=\prod_{j=0}^{\infty}\big(1-\beta_j\alpha^i(1-z)\big)\qquad (i\ge 0).$$

It follows by Theorem 2.3 and (2.8) that

$$\varphi_X(z)=\prod_{i=0}^{\infty}\Big[\prod_{j=0}^{\infty}\big(1-\beta_j\alpha^i(1-z)\big)\Big]. \tag{A.9}$$

Clearly, the right hand side of (A.9) converges. A straightforward argument shows that the double infinite product $\prod\limits_{i=0}^{\infty}\prod\limits_{j=0}^{\infty}\big(1-\beta_j\alpha^i(1-z)\big)$ converges. In order to be able to interchange the order of the infinite products in (A.9), it remains to show that the iterated infinite product

$$\prod_{j=0}^{\infty}\Big[\prod_{i=0}^{\infty}\big(1-\beta_j\alpha^i(1-z)\big)\Big]=\prod_{j=0}^{\infty}P_j(z)$$

converges, where for each $j\ge 0$,

$$P_j(z)=\prod_{i=0}^{\infty}\big(1-\beta_j\alpha^i(1-z)\big).$$

Note that for each $j\ge 0$, $P_j(\cdot)$ has the form of the pgf of the marginal of an INAR(1) process with a Bernoulli(β_j) innovation (see (A.3)). Therefore, by Theorem 3.1 and (3.5),

$$P_j(z)=1+\sum_{n=1}^{\infty}\frac{\beta_j^n(z-1)^n\alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)}. \tag{A.10}$$

For $j\ge 0$, denote

$$\zeta_j(z)=\sum_{n=1}^{\infty}\frac{\beta_j^n(z-1)^n\alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)}.$$

By (A.8) and $0\le z\le 1$, we have

$$\sum_{j=0}^{\infty}|\zeta_j(z)|\le\sum_{n=1}^{\infty}\sum_{j=0}^{\infty}\frac{\beta_j^n\alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)}=\sum_{n=1}^{\infty}\frac{B_n\alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)}\le\sum_{n=1}^{\infty}a_n,$$

where

$$a_n=\frac{\lambda^n}{1-q^n}\frac{\alpha^{\binom{n}{2}}}{\prod_{l=1}^{n}(1-\alpha^l)}.$$

Since $\alpha,q\in(0,1)$, we have

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=\lim_{n\to\infty}\frac{\lambda(1-q^n)\alpha^n}{(1-q^{n+1})(1-\alpha^{n+1})}=0.$$

By the ratio test, $\sum\limits_{j=0}^{\infty}|\zeta_j(z)|$ converges uniformly over $z\in[0,1]$. This in turn implies (see Knopp, 1990) that $\prod\limits_{j=0}^{\infty}P_j(z)$ converges uniformly over $z\in[0,1]$. The representation (7.4) then follows by interchanging the order of the infinite products in (A.9) and by using (A.10).

We now prove (7.5). By the first part of the proof, we have

$$\varphi_X(z) = \prod_{j=0}^{\infty}\Big[\prod_{i=0}^{\infty}\big(1-\beta_j\alpha^i(1-z)\big)\Big].$$

Applying the representation (3.6) to $\prod_{i=0}^{\infty}\big(1-\beta_j\alpha^i(1-z)\big)$ with $p=\beta_j$, we have

$$\varphi_X(z) = \exp\left\{-\sum_{j=0}^{\infty}\Big[\sum_{n=1}^{\infty}\frac{\beta_j^n}{n(1-\alpha^n)}(1-z)^n\Big]\right\}. \tag{A.11}$$

This implies that the double series in (A.11) is convergent. Since its terms are nonnegative (as $0 \le z \le 1$), the order of summation can be interchanged (by Cauchy's criterion for double series). This establishes the representation (7.5). By Theorem 3.1 and (3.3), $P_j(z)$ of (A.10) is the pgf of the pmf $\{q_r^{(j)}\}$ of (7.7). Therefore, part 3 and (7.6) follow from the representation (7.4), Theorem 2.2 and (2.5). □

ACKNOWLEDGMENTS

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

REFERENCES

Abid, R., Kokonendji, C. C., and Masmoudi, A. (2021). On Poisson-exponential-Tweedie models for ultra-overdispersed count data. *AStA Advances in Statistical Analysis*, 105:1–23.

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

Aly, E.-E. A. A. and Bouzar, N. (2023). On some stationary INAR(1) processes with compound Poisson distributions. *REVSTAT – Statistical Journal*, 21:321–345.

Alzaid, A. A. and Al-Osh, M. A. (1990). An integer-valued p-th-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 27:314–324.

Athreya, K. B. and Ney, P. E. (2004). *Branching Processes*. Dover Publications (originally published in 1972 by Springer-Verlag Berlin-Heidelberg).

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

Bourguignon, M. and Vasconcellos, K. L. (2015). First-order non-negative integer-valued autoregressive processes with power series innovations. *Brazilian Journal of Probability and Statistics*, 29: 71–93.

Bouzar, N. and Jayakumar, K. (2008). Time series with discrete semi-stable marginals. *Statistical Papers*, 49:619–635.

Charalambides, C. A. (2016). *Discrete q-Distributions*. John Wiley & Sons Inc.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. Vol. 1, John Wiley & Sons Inc.

Heathcote, C. R. (1965). Corrections and comments on the paper "A branching process involving immigration". *Journal of the Royal Statistical Society: Series B*, 27:213–217.

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

Jung, R. C., Ronning, G., and Tremayne, A. R. (2005). Estimation in conditional first order autoregression with discrete support. *Statistical Papers*, 46:195–224.

Kemp, A. W. (1987). A Poissonian binomial model with constrained parameters. *Naval Research Logistics*, 34:853–858.

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

Kim, H. and Lee, S. (2017). On first-order integer-valued autoregressive process with Katz family innovations. *Journal of Statistical Computation and Simulation*, 87:546–562.

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

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

McKenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric marginal distributions. *Advances in Applied Probability*, 18:679–705.

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

McKenzie, E. (2003). Discrete variate time series. In: *Handbook of Statistics*; *Stochastic Processes: Modelling and Simulation*, C. R. Rao and D. Shanbhag, Eds., Elsevier Science, Amsterdam.

Pedeli, X. and Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical Modelling*, 11:325–349.

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

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

Scotto, M. G., Weiß, C. H., and Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, 15:590–618.

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

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

Weiß, C. H. (2013). Integer-valued autoregressive models for counts showing underdispersion. *Journal of Applied Statistics*, 40:1931–1940.

REVSTAT – Statistical Journal
Volume 23, Number 2, April 2025, 189–205
https://doi.org/10.57805/revstat.v23i2.523

A Multinomial Asymptotic Representation of Zenga's Discrete Index, Its Influence Function and Data-Driven Applications

Authors: Tchilabalo Abozou Kpanzou ✉
– Department of Mathematics, University of Kara, Togo
t.kpanzou@univkara.net

Diam Ba
– Department of Applied Mathematics, Gaston Berger University, Sénégal
diamba79@gmail.com

Gandasor Bonyiri Onesiphore Da
– Department of Applied Mathematics, Gaston Berger University, Sénégal
gandasor@gmail.com

Gane Samb Lo
– Department of Applied Mathematics, Gaston Berger University, Sénégal
gane-samb.lo@ugb.edu.sn

Received: September 2022 Revised: March 2023 Accepted: March 2023

Abstract:

- In this paper, we consider the Zenga index, one of the most recent inequality indices. We keep the finite-valued original form and address the asymptotic theory. The asymptotic normality is established through a multinomial representation. The influence function is also given. The results are simulated and applied to Senegalese income data.

Keywords:

- *inequality measures; asymptotic behaviour; asymptotic representations; functional empirical process.*

AMS Subject Classification:

- 62G05, 62G20, 62G07, 91B82, 62P20.

✉ Corresponding author

1. INTRODUCTION

Over the years, a number of measures of inequality have been developed. Examples include the generalized entropy, the Atkinson, the Gini, the quintile share ratio and the Zenga measures (see, e.g., Zenga, 1984; Zenga, 1990; Cowell and Flachaire, 2007; Cowell *et al.*, 2009; Kpanzou *et al.*, 2013; Kpanzou, 2014; Hulliger and Schoch, 2009). Recently, Mergane and Lo (2013) gathered a significant number of inequality measures under the name of Theil-like family. Such inequality measures are very important in capturing inequality in income distributions. They also have applications in many other branches of Science, e.g., in ecology (see, e.g., Magurran, 1991), sociology (see, e.g., Allison, 1978), demography (see, e.g., White, 1986) and information science (see, e.g., Rousseau, 1993).

The inequality measure of Zenga (2006) is one of the most recent ones. It is receiving a considerable attention from researchers for its novelty indeed, but for its interesting properties. Papers dealing with that measure cover theoretical aspects including asymptotic theory and statistical inference (Greselin *et al.*, 2010b; Emad-Eldin and Marilou, 1999) and applied works to income data (Greselin *et al.*, 2010a, 2014), etc.

In this paper, we focus on the discrete form of the inequality measure as introduced by Zenga (2006). We justify the asymptotic study of the discrete and finite form by a number of reasons. In some situations, only aggregated data exist. Although this is hardly conceivable today, it is still possible and it is highly probable that the researcher does not have access to the original data and has in hand only data in form of frequency tables. Some other times, frequency tables may be available while the full data is destroyed or lost. Right now, in Gambia, health data collected from the health centers are stored in daily books and the national health direction extracts frequency tables from those books and this type of data is the only one available in their computerized system. So one of the main reasons to work on the finite discrete data is the lack of accessibility to the full data for one reason or another. The second main reason is that an asymptotic theory on such kind of data will give the structure of the limit results with also no severe conditions. By replacing the discrete finite probability law of the aggregated data by a general probability law, we get the precise general asymptotic case. From that simplified study, we see what might be expected in general theory before we proceed with it.

Here, we suppose that the full data have been summarized into a frequencies table as given in Table 1, where each class (c_{i-1}, c_i) is represented by a single point x_i^*, usually taken as the middle of the class, $x_i^* = (c_{i-1} + c_i)/2$ (other possible choices are the mean or the median of the observations falling in the class).

We may thus adopt to approximately reconstitute the $n \geq 1$ data as

$$\underbrace{x_1^*}_{n_1\ times} \cdots \underbrace{x_j^*}_{n_j\ times} \cdots \underbrace{x_m^*}_{n_m\ times}$$

In the sequel, we suppose that the data themselves are discrete and take a pre-determined number of m values. First, we will give an asymptotic theory which will be in the form of representation in multinomial laws, instead of a representation in Brownian Bridges as in general. Next, the influence function (IF) will be derived by direct computations and this usually allows to again find the asymptotic variance and sometimes, as in our case, to find a different but equivalent expression of that variance.

Table 1: Frequencies Table.

classes (c_{i-1}, c_i)	representatives x_i^*	frequencies n_i
(c_0, c_1)	x_1^*	n_1
(c_1, c_2)	x_2^*	n_2
$\vdots$	$\vdots$	$\vdots$
(c_{m-1}, c_m)	x_m^*	n_m
Total	x_i^*	n

The work presented here will be applied to income data available in an aggregated form. At the same time, it serves as a paving way to a more general approach.

Let us suppose that the income variable X is discrete and takes the m $(m > 1)$ ordered values $-\infty = x_0 < x_1 < ... < x_m < x_{m+1} = +\infty$ with the probabilities $p_j > 0$, $j \in \{1, ..., m\}$ with $p_1 + p_2 + ... + p_m = 1$. If the income is continuously observed, we have a sequence of random replications $X_1, X_2, ...$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For each $n \geq 1$, the empirical distribution of X on the sample is characterized by the empirical frequencies

$$n_0 = 0, \quad n_j = \#\{h \in \{1, ..., n\}, \quad X_h = x_j\}, \quad j \in \{1, ..., m\},$$

and their normalized and cumulative forms given respectively by

$$f_0 = 0, \ f_j = \frac{n_j}{n}, \ j \in \{1, ..., m\},$$

and

$$n_0^* = f_0^* = 0, n_j^* = \sum_{h=1}^{j} n_h, \ f_j^* = \sum_{h=1}^{j} f_h, \quad j \in \{1, ..., m\},$$

with

$$\sum_{j=1}^{m} n_j = n, \quad \sum_{j=1}^{m} f_j = 1, \quad n_m^* = n, \quad f_m^* = 1.$$

We also define

$$p_0^* = 0, \quad p_j^* = \sum_{h=1}^{j} p_h, \ p_m^* = 1.$$

The empirical and discrete Zenga (2006)'s index is given by

$$Z_{d,n} = 1 - \sum_{j=1}^{m-1} f_j \frac{(n_j^*/n)^{-1} \sum_{1 \leq h \leq j} n_h x_h}{(1 - (n_j^*/n))^{-1} \sum_{j+1 \leq h \leq m} n_h x_h},$$

which is obtained by summing Formula (3.1) in Zenga (2006) over $j \in \{1, ..., m\}$ and presented as a synthetic measure of inequality. The empirical cumulative distribution function (cdf) based on the sample of size $n \geq 1$ is

$$F_n(x) = \frac{1}{n} \sum_{h=1}^{m} n_h 1_{[x_h, x_{h+1}[}(x), \quad x \in \mathbb{R},$$

and is the non-parametric estimator of the true cdf

$$F(x) = \sum_{h=1}^{m} p_j 1_{[x_h, x_{h+1}[}(x), \quad x \in \mathbb{R}.$$

We also have the empirical probability generated by the sample, given by

$$P_{X,n}(A) = \frac{1}{n} \sum_{j=1}^{m} 1_A(x_j).$$

We may express $Z_{d,n}$ in terms of the empirical probability measure by

$$Z_{d,n} = 1 - \sum_{j=1}^{m-1} \mathbb{P}_{X,n}(x_j) \frac{\left(\int 1_{]0,x_j]}(t) d\mathbb{P}_{X,n}(t)\right)^{-1} \left(\int t 1_{]0,x_j]}(t) d\mathbb{P}_{X,n}(t)\right)}{\left(\int 1_{]x_j,+\infty[}(t) d\mathbb{P}_{X,n}(t)\right)^{-1} \left(\int t 1_{]x_j,+\infty[}(t) d\mathbb{P}_{X,n}(t)\right)}.$$

Finally by considering the discrete measure $\nu = \sum_{1 \leq j \leq n} \delta_{x_j}$, where δ_{x_j} is the Dirac measure concentrated at x_j with mass one, we may also write

$$Z_{d,n} = 1 - \int \frac{\left(\int 1_{]0,s]}(t) d\mathbb{P}_{X,n}(t)\right)^{-1} \left(\int t 1_{]0,s]}(t) d\mathbb{P}_{X,n}(t)\right)}{\left(\int 1_{]s,+\infty[}(t) d\mathbb{P}_{X,n}(t)\right)^{-1} \left(\int t_{]s,+\infty[}(t) d\mathbb{P}_{X,n}(t)\right)} \mathbb{P}_{X,n}(s) d\nu(s).$$

It is clear, by the convergence in law of a sequence of probability measures $\mathbb{P}_{X,n}$ to $\mathbb{P}_X = \mathbb{P}X^{-1}$ (the probability law of X), that $Z_{d,n}$ converges to

$$Z_d = 1 - \int \frac{\left(\int 1_{]0,s]}(t) d\mathbb{P}_X(t)\right)^{-1} \left(\int t 1_{]0,s]}(t) d\mathbb{P}_X(t)\right)}{\left(\int 1_{]s,+\infty[}(t) d\mathbb{P}_X(t)\right)^{-1} \left(\int t 1_{]s,+\infty[}(t) d\mathbb{P}_X(t)\right)} \mathbb{P}_X(s) d\nu(s).$$

In this simple setting, the convergence is easily justified because of the finiteness of the summations and of the functions. In terms of cdf and mathematical expectation, we have

$$Z_d = 1 - \int_{x_1}^{x_m} \frac{\frac{1}{F(s)} \int_0^s t d\mathbb{P}_X(t)}{\frac{1}{1-F(s)} \int_s^\infty t d\mathbb{P}_X(t)} \mathbb{P}_X(s) d\nu(s).$$

The integral in the last expression should be read as

$$\int_{x_1}^{x_m -} \frac{\frac{1}{F(s)} \int_0^s t d\mathbb{P}_X(t)}{\frac{1}{1-F(s)} \int_s^\infty t d\mathbb{P}_X(t)} \mathbb{P}_X(s) d\nu(s) = \int 1_{[x_1, x_m[}(s) \frac{\frac{1}{F(s)} \int_0^s t d\mathbb{P}_X(t)}{\frac{1}{1-F(s)} \int_s^\infty t d\mathbb{P}_X(t)} \mathbb{P}_X(s) d\nu(s),$$

so that neither $1 - F(s)$ nor $F(s)$ never vanishes on the integration domain.

On one hand, we are going to draw an asymptotic normality theory of $Z_{d,n}$ using the m-multivariate binomial laws. On the other hand, the sensitivity of a statistic $T(F)$ and the impact of extreme observations on it are also two recurrent questions in the research in the field (see Cowell and Flachaire, 2007).

In that context, the asymptotic variance of the plug-in estimator $T(F_n)$ of the statistic $T(F)$ is of the form $\sigma^2 = \int L(x, T(F))^2 dF(x)$. From this, we may say that the influence function behaves in nonparametric estimation as the score function does in the parametric setting (see Wasserman, 2006, p. 19). To define the notion of *IF*, let us consider the contaminated probability law $\mathbb{P}_X^{-(\varepsilon)}$ of $\mathbb{P}_X$ at x with mass $\varepsilon > 0$, defined by

$$\mathbb{P}_X^{(\varepsilon)} = (1-\varepsilon)\mathbb{P}_X + \varepsilon\delta_x \tag{1.1}$$

and a functional of $\mathbb{P}_X$, namely $T(\mathbb{P}_X)$. The influence function of the functional T at x, if it exists, is given by

$$IF(T, x) = \lim_{\varepsilon \to 0} \frac{T(\mathbb{P}_X^{(\varepsilon)}) - T(\mathbb{P}_X)}{\varepsilon}. \tag{1.2}$$

The previous remarks motivate us to derive the *IF* of $Z_d(\mathbb{P}_X)$ and to compare it with the asymptotic variance the Zenga's plug-in estimator.

Before we proceed to our task, we point out that asymptotic normality results for Zenga's index are available in the literature, among them those of Greselin *et al.* (2010b) and Emad-Eldin and Marilou (1999).

Here is how the paper is organized, we give our asymptotic results as described above in Section 2 in Theorems 2.1 and 2.2, and the proofs of these theorems are given in the Appendices A and B. Section 3 is devoted to simulation studies and data-driven application to Senegalese Data. We conclude in Section 4.

2. ASYMPTOTIC THEORY FOR THE DISCRETE ZENGA MEASURE

2.1. Asymptotic normality

Let us begin with the following reminder. For each $m \geq 1$, the random vector $(n_1, ..., n_m)$ follows a m-dimensional multimonial law of parameters $n \geq 1$ and $p = (p_1, ..., p_m)^t$. In such a case, a classical result of weak convergence (see, e.g., Lo *et al.*, 2016), as $n \to +\infty$, is the following

$$\begin{aligned} \left(\frac{n_1 - np_1}{\sqrt{np_1}}, ..., \frac{n_m - np_m}{\sqrt{np_m}}\right)^t &\equiv (N_{1,n}, ..., N_{m,n})^t \\ &\rightsquigarrow Z = (Z_1, ..., Z_m)^t \sim \mathcal{N}_m(0, \Sigma), \end{aligned}$$

the variance-covariance matrix $\Sigma = (\sigma_{h,k})_{1 \leq h,k \leq m}$ of Z is defined, for $(h, k) \in \{1, ..., m\}^2$, $h \neq k$, by

$$\sigma_{hh} = \mathbb{E}(Z_h^2) = 1 - p_h \textit{ and } \sigma_{hk} = \mathbb{E}(Z_h Z_k) = -\sqrt{p_h p_k}.$$

We invoke the Skorohod-Wichura Theorem (see Wichura, 1970) to suppose that Z is defined on the same probability space and that

$$(N_{1,n},...,N_{m,n})^t \to_{\mathbb{P}} Z, \ as \ n \to +\infty.$$

Let us give some notation. Define vectors $C = (c_1, ..., c_m)^t$ such that

$$c_j = \sqrt{p_j}\frac{(1/p_j^*)\mu_{(j)}}{(1/(1-p_j^*))\mu^{(j)}}1_{(j\neq m)}, j \in \{1,...,m\},$$

for $j \in \{1, ..., m-1\}$, $i \in \{1,2\}$, $D_{j,i} = (d_{j,i,1}, ..., d_{j,i,m})^t$ such that

$$d_{j,1,h} = (x_h\sqrt{p_h})1_{(h\leq j)}, \quad d_{j,2,h} = -(x_h\sqrt{p_h})1_{(h\geq j+1)},$$

$$\gamma_{j,1} = p_j\frac{(1/p_j^*)}{(1/(1-p_j^*))\mu^{(j)}}, \quad \gamma_{j,2} = p_j\frac{(1/p_j^*)}{(1/(1-p_j^*))}\frac{\mu_{(j)}}{\left(\mu^{(j)}\right)^2},$$

and let $E_j = (e_{j,1}, ..., e_{j,m})^t$ be the vector defined by its components as follows:

$$e_{j,h} = -(\sqrt{p_h})1_{(h\leq j)}.$$

Finally, let us define

$$-H = C + \sum_{j=1}^{m-1}\Big(\gamma_{j,1}D_{j,1} + \gamma_{j,2}D_{j,2} + \left(p_j^*\right)^{-2}E_j\Big).$$

Theorem 2.1. *Under the notation given above, we have, as $n \to +\infty$,*

$$\sqrt{n}(Z_{d,n} - Z_d) \rightsquigarrow \mathcal{N}_m\big(0, H^t\Sigma H\big).$$

Proof: The proof is given in Appendix A.

2.2. Influence function of Z_d

Theorem 2.2. *Under the notations given above, the influence function of Z_d is given, for $x_1 \leq x \leq x_m$, by*

$$\begin{aligned} IF(Z_d, x) &= \int \mathbb{P}_X(s)\Big(\frac{R_1(s)}{R_2(s)^2(1-F(s)})1_{]s,+\infty]}(x) - \frac{1}{R_2(s)F(s)}1_{]0,s]}(x)\Big)xd\nu \\ &+ \int \mathbb{P}_X(s)\Big(\frac{R_1(s)}{R_2(s)F(s)}1_{]0,s]}(x) - \frac{R_1(s)}{R_2(s)(1-F(s))}1_{]s,+\infty]}(x)\Big)d\nu \\ &- \int \delta_x(s)\frac{R_1(s)}{R_2(s)}d\nu + \int \mathbb{P}_X(s)\frac{R_1(s)}{R_2(s)}d\nu. \end{aligned}$$

Proof: The proof is given in Appendix B.

3. SIMULATION AND DATA-DRIVEN APPLICATIONS

3.1. Simulation study

Quality of the convergence: We choose a Probability distribution of yearly income supported by $m = 10$ points with lower endpoint $x_1 = 4,515,000$ XOF $(9,030$ nearly) and upper endpoint $x_m = 9,000,000$ XOF $(170,490$ nearly), characterized as in Table 2.

Table 2: Underlying Probability Law.

Values	x_1	x_2	x_3	x_4	x_5	...
	4.515×10^6	13.485×10^6	22.455×10^6	31.425×10^6	40.395×10^6	...
$\mathbb{P}(X = x_i)$	0.05	0.05	0.05	0.05	0.1	...

Values	...	x_6	x_7	x_8	x_9	x_{10}
	...	49.365×10^6	58.335×10^6	67.305×10^6	76.275×10^6	85.245×10^6
$\mathbb{P}(X = x_i)$	...	0.1	0.2	0.2	0.1	0.1

Table 3 shows the good performance of the nonparametric estimation of the Zenga index for sample size from $n = 100$ to $n = 1500$. Such sizes are comparable with those of sample survey from populations counted in dozen of millions.

Table 3: Mean Errors (ERM), Mean Square Errors (MSE).

Size	100	200	500	750	1000	1500
ERM	3.6×10^{-3}	-5.36×10^{-3}	10^{-3}	-8.41×10^{-4}	4.56×10^{-5}	-1.44×10^{-3}
MSE	6.4×10^{-2}	3.35×10^{-2}	2.49×10^{-2}	2.16×10^{-2}	1.9×10^{-2}	1.64×10^{-2}

Figure 1 shows the pretty good asymptotic normality approximation of the centered and normalized empirical Zenga's estimator.

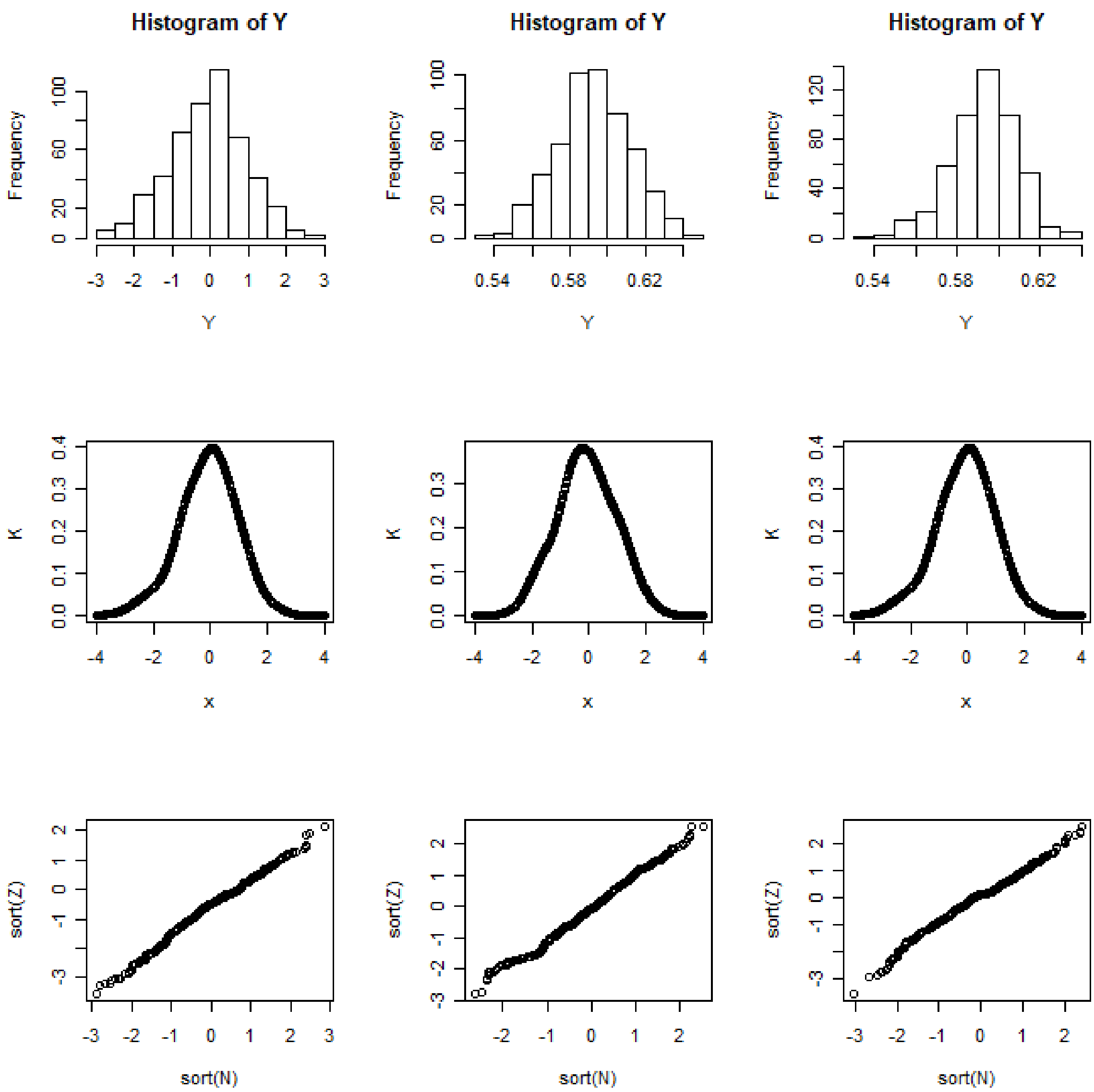

Figure 1: Histograms, Parzen Estimators and QQ-plots for sample sizes 500, 1000 and 1500 from left to right.

3.2. Data-driven applications

We use the income Data in Senegal (2001-2002) from the database related to ANSD (2019). The incomes are given by households. We should use an adult-equivalence scale to consider to be able to compare households. The notion of adult-equivalence has already been described in Lo (2016) and implemented on different sets of data, among them the data just described above. The data are available for the whole country (Senegal) and for the 10 areas given in the following order:

(OA): Dakar, Diourbel, Fatik, Kaolack, Louga, Saint-Louis, Tamba, Thies, Ziguinchor, Kolda.

Dakar is the most urbanized area of Senegal and includes the capital of the country, also named Dakar. It concentrates almost 23.1% of the population.

The Zenga and the Gini indices have been computed for the 11 areas from the aggregate data, and are display in Table 4. Note that these values are given in percentage (%).

Table 4: Zenga and Gini index measures for Senegal's administrative areas (2000).

Index	Senegal	Dakar	Diourbel	Fatick	Kaolack	Louga	...
Zenga	80.65	93.33	81.34	92.54	81.11	84.00	...
Gini	75.00	80.90	75.26	80.39	75.16	16.25	...

Index	...	Saint-Louis	Tamba	Thies	Ziguinchor	Kolda
Zenga	...	87.69	86.64	82.61	82.11	80.24
Gini	...	78.83	77.26	75.72	75.52	47.86

Through the values in this table, the 11 areas are ordered from the least inequality index to the greatest as follows:

Ordering by Zenga's index: Kolda (1), Senegal (2), Kaolack (3), Diourbel (4), Ziguinchor (5), Thies (6), Louga (7), Tamba (8), Saint-Louis (9), Fatick (10), Dakar (11).

Ordering by Gini's index: Louga (1), Kolda (2), Senegal (3), Kaolack (4), Diourbel (5), Ziguinchor (6), Thies (7), Tamba (8), Saint-Louis (9), Fatick (10), Dakar (11).

These orderings are illustrated in Figure 2.

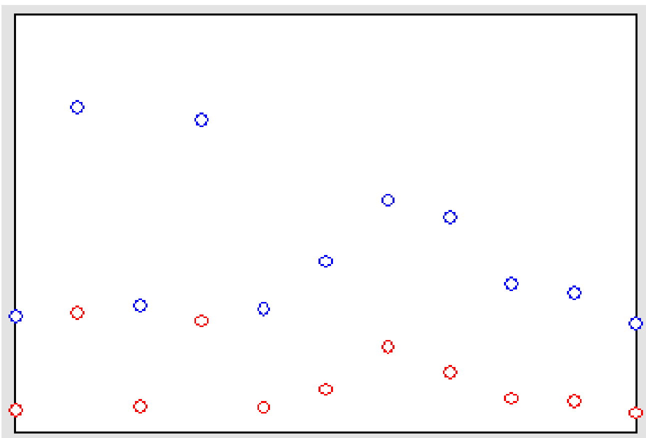

Figure 2: The areas are given in the horizontal line and are ordered according to the ranking **(AO)** above. Blue: Zenga's index. Red: Gini's index.

The most striking fact is that both indices do not order the areas in an exact similar way. The most unfair areas (with the greatest values of the inequality index) are the same with the same ordering, from areas 8 to 11. From areas 1 to 7, the ordering has slightly changed but the case of Louga is remarkable. It is ranked first by Gini and seventh by Zenga.

One may think that the inequality should be greater in urban areas than in rural zones. Indeed we see that with the areas of Thies, Saint-Louis, Dakar. But Factik and Tamba are so urbanized areas. Investigating why the inequality indices (both Zenga and Gini) are high should be investigated in accordance with local realities.

4. CONCLUSION

In this paper, we have considered the discrete Zenga index for which we derived the influence function and studied the asymptotic theory. The asymptotic normality is established through a multinomial representation. Through simulation, we confirmed the asymptotic normality result obtained theoretically. The results are also applied to Senegalese income data.

A. APPENDIX — Proof of Theorem 2.1

Let us fix $n \geq 1$. We have

$$Z_{n,d} = 1 - \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{n}{n_j^*} - 1 \right) \frac{\sum_{1 \leq h \leq j} n_h x_h}{\sum_{j+1 \leq h \leq m} n_h x_h}.$$

We define

$$Z_{d,n}^* = \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{n}{n_j^*} - 1 \right) \frac{\sum_{1 \leq h \leq j} n_h x_h}{\sum_{j+1 \leq h \leq m} n_h x_h}.$$

and for $1 \leq j \leq m-1$,

$$\mu_{(j)} = \sum_{h=1}^{j} p_h x_h \quad \textit{and} \quad \mu^{(j)} = \sum_{h=j+1}^{m} p_h x_h.$$

We have

$$\begin{aligned}
\frac{\sum_{1 \leq h \leq j} n_h x_h}{\sum_{j+1 \leq h \leq m} n_h x_h} - \frac{\mu_{(j)}}{\mu^{(j)}} &= \frac{\sum_{1 \leq h \leq j} n_h x_h}{\sum_{j+1 \leq h \leq m} n_h x_h} - \frac{n\mu_{(j)}}{\sum_{j+1 \leq h \leq m} n_h x_h} + \frac{n\mu_{(j)}}{\sum_{j+1 \leq h \leq m} n_h x_h} - \frac{\mu_{(j)}}{\mu^{(j)}} \\
&= \frac{\sum_{h=1}^{j} x_h N_{h,n} \sqrt{p_h}}{\sqrt{n} \sum_{j+1 \leq h \leq m} n_h x_h / n} - \frac{\mu_{(j)} \sum_{h=j+1}^{m} x_h N_{h,n} \sqrt{p_h}}{\sqrt{n} \mu^{(j)} \left(\sum_{j+1 \leq h \leq m} n_h x_h / n \right)}.
\end{aligned}$$

Then

$$\begin{aligned}
Z_{d,n}^* &= \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{n}{n_j^*} - 1 \right) \frac{\mu_{(j)}}{\mu^{(j)}} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{n}{n_j^*} - 1 \right) \left(\frac{\sum_{h=1}^{j} x_h N_{h,n} \sqrt{p_h}}{\sum_{j+1 \leq h \leq m} n_h x_h / n} - \frac{\mu_{(j)} \sum_{h=j+1}^{m} x_h N_{h,n} \sqrt{p_h}}{\mu^{(j)} \left(\sum_{j+1 \leq h \leq m} n_h x_h / n \right)} \right) \\
&=: Z_{d,n}^*(1) + R_n(1,1).
\end{aligned}$$

We also have

$$\begin{aligned}
\left(\frac{n}{n_j^*} - 1 \right) - \left(\frac{1}{p_j^*} - 1 \right) &= \left(\frac{n}{n_j^*} - 1 \right) - \left(\frac{n}{\sum_{h=1}^{j} n p_h} - 1 \right) \\
&= - \frac{\sum_{h=1}^{j} n_h - \sum_{h=1}^{j} p_h}{\left(\sum_{h=1}^{j} p_h \right) \left(\sum_{h=1}^{j} n_h \right)} \\
&= - \frac{1}{\sqrt{n}} \frac{\sum_{h=1}^{j} \sqrt{p_h} N_{h,n}}{\left(\sum_{h=1}^{j} p_h \right) \left(\sum_{h=1}^{j} n_h / n \right)}.
\end{aligned}$$

This leads to

$$\begin{aligned}
Z_{d,n}^*(1) &= \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{1}{p_j^*} - 1 \right) \frac{\mu_{(j)}}{\mu^{(j)}} - \sum_{j=1}^{m-1} \frac{n_j}{n} \frac{n \sqrt{n} \sum_{h=1}^{j} \sqrt{p_h} N_{h,n}}{\left(\sum_{h=1}^{j} n_h \right) \left(\sum_{h=1}^{j} n p_h \right)} \frac{\mu_{(j)}}{\mu^{(j)}} \\
&=: Z_{d,n}^*(2) + R_n(1,2).
\end{aligned}$$

Finally, we have

$$\begin{aligned} Z^*_{d,n}(2) &= \sum_{j=1}^{m-1} p_j \left(\frac{1}{p_j^*}-1\right)\frac{\mu_{(j)}}{\mu^{(j)}} + \frac{1}{n}\sum_{j=1}^{m-1} \frac{\sqrt{np_j}N_{j,n}}{}\left(\frac{1}{p_j^*}-1\right)\frac{\mu_{(j)}}{\mu^{(j)}} \\ &= \sum_{j=1}^{m-1} p_j \left(\frac{1}{p_j^*}-1\right)\frac{\mu_{(j)}}{\mu^{(j)}} + \frac{1}{\sqrt{n}}\sum_{j=1}^{m-1} \sqrt{p_j}N_{j,n}\left(\frac{1}{p_j^*}-1\right)\frac{\mu_{(j)}}{\mu^{(j)}} \qquad \text{(L2)} \\ &= \sum_{j=1}^{m-1} \frac{(1/p_j^*)\mu_{(j)}}{(1/(1-p_j^*))\mu^{(j)}} + \frac{1}{\sqrt{n}}\sum_{j=1}^{m-1} \sqrt{p_j}N_{j,n}\left(\frac{1}{p_j^*}-1\right)\frac{\mu_{(j)}}{\mu^{(j)}} \\ &=: Z^*_d + R_n(3). \end{aligned}$$

It is clear that

$$Z_d = 1 - Z^*_d.$$

We finally get

$$\sqrt{n}(Z^*_{d,n} - Z^*_d) = \sqrt{n}R_n(1) + \sqrt{n}R_n(2) + \sqrt{n}R_n(3).$$

By using the convergence (strong and weak) on binomial probabilities, we get

$$\begin{aligned} \sqrt{n}R_n(1,1) &= \sum_{j=1}^{m-1} \frac{n_j}{n}\left(\frac{n}{n_j^*}-1\right)\left(\frac{\sum_{h=1}^{j}\big(x_h\sqrt{p_h}\big)N_{h,n}}{\sum_{j+1\leq h\leq m} n_h x_h/n} - \frac{\mu_{(j)}\sum_{h=j+1}^{m}\big(x_h\sqrt{p_h}\big)N_{h,n}}{\mu^{(j)}\Big(\sum_{j+1\leq h\leq m} n_h x_h/n\Big)}\right) \\ &\to_{\mathbb{P}} \sum_{j=1}^{m-1} p_j \frac{(1/p_j^*)}{(1/(1-p_j^*))}\left(\frac{\sum_{h=1}^{j}\big(x_h\sqrt{p_h}\big)Z_h}{\mu^{(j)}} - \frac{\mu_{(j)}\sum_{h=j+1}^{m}\big(x_h\sqrt{p_h}\big)Z_h}{\big(\mu^{(j)}\big)^2}\right). \quad (A1) \end{aligned}$$

Next

$$\begin{aligned} \sqrt{n}R_n(1,2) &= -\frac{\sum_{h=1}^{j}\sqrt{p_h}N_{h,n}}{\Big(\sum_{h=1}^{j} p_h\Big)\Big(\sum_{h=1}^{j} n_h/n\Big)} \\ &\to_{\mathbb{P}} -\frac{\sum_{h=1}^{j}\sqrt{p_h}Z_h}{(p_j^*)^2}. \quad (A2) \end{aligned}$$

Finally,

$$\begin{aligned} \sqrt{n}R_n(3) &= \sum_{j=1}^{m-1} \sqrt{p_j}\left(\frac{1}{p_j^*}-1\right)\frac{\mu_{(j)}}{\mu^{(j)}}N_{j,n} \\ &\to_{\mathbb{P}} \sum_{j=1}^{m-1} \sqrt{p_j}\frac{(1/p_j^*)\mu_{(j)}}{(1/(1-p_j^*))\mu^{(j)}}Z_j. \quad (A3) \end{aligned}$$

By combining Developments (A1), (A2) and (A3), we get

$$
\begin{aligned}
\sqrt{n}(Z^*_{d,n} - Z^*_d) \;\to\; & \sum_{j=1}^{m-1} p_j \frac{(1/p^*_j)}{(1/(1-p^*_j))} \left(\frac{\sum_{h=1}^{j} \big(x_h \sqrt{p_h}\big) Z_h}{\mu^{(j)}} - \frac{\mu_{(j)} \sum_{h=j+1}^{m} \big(x_h \sqrt{p_h}\big) Z_h}{\big(\mu^{(j)}\big)^2} \right) \\
& - \frac{\sum_{h=1}^{j} \sqrt{p_h} Z_h}{(p^*_j)^2} + \sum_{j=1}^{m-1} \sqrt{p_j} \frac{(1/p^*_j)\mu_{(j)}}{(1/(1-p^*_j))\mu^{(j)}} Z_j \\
= \; & \left(\sum_{j=1}^{m-1} \langle \gamma_{j,1} D_{j,1}, Z \rangle + \langle \gamma_{j,2} D_{j,2}, Z \rangle + \langle (p^*_j)^{-2} E_j, Z \rangle \right) + \langle C, Z \rangle.
\end{aligned}
$$

We conclude that

$$
\sqrt{n}(Z^*_{d,n} - Z^*_d) \to_{\mathbb{P}} H^t Z. \quad \square
$$

B. APPENDIX — Proof of Theorem 2.2

Let us write, for $s \in \mathcal{R}$,

$$R_1(s) = R_1(s, \mathbb{P}_X) = \frac{\int t 1_{]0,s]}(t) d\mathbb{P}_X(t)}{\int 1_{]0,s]}(t) d\mathbb{P}_X(t)}$$

and

$$R_2(s) = R_2(s, \mathbb{P}_X) = \frac{\int t 1_{]s,+\infty[}(t) d\mathbb{P}_X(t)}{\int]s, +\infty[d\mathbb{P}_X(t)}.$$

We have

$$Z_d(\mathbb{P}_X) = Z_d = 1 - \int \frac{R_1(s)}{R_2(s)} \mathbb{P}_X(s) d\nu(s).$$

By using Formula (1.1), we have

$$\frac{d(\mathbb{P}_X^{(\varepsilon)} - \mathbb{P}_X)}{\varepsilon} = -d\mathbb{P}_X + d\delta_x.$$

For short, we write

$$R_i(s, \mathbb{P}_X) = R_i(s) \text{ and } R_i(s, \mathbb{P}_X^{(\varepsilon)}) = R_i(s, \varepsilon), i \in \{1, 2\}.$$

We have

$$\begin{aligned} Z_d(\mathbb{P}_X^{(\varepsilon)}) - Z_d(\mathbb{P}_X) &= -(1-\varepsilon) \int \mathbb{P}_X(s) \frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} d\nu - \varepsilon \int \delta_x(s) \frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} d\nu \\ &\quad + \int \mathbb{P}_X(s) \frac{R_1(s)}{R_2(s)} d\nu \\ &= - \int \mathbb{P}_X(s) \left(\frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} - \frac{R_1(s)}{R_2(s)} \right) d\nu \\ &\quad + \varepsilon \int \mathbb{P}_X(s) \frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} d\nu - \varepsilon \int \delta_x(s) \frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} d\nu. \end{aligned}$$

Let us apply the definition of the *IF* as in Formula (1.2). Since $\mathbb{P}_X^{(\varepsilon)} \to \mathbb{P}_X$ as $\varepsilon \to 0$ (the convergence being meant as a convergence in law), we have no problem to see that

$$\begin{aligned} \lim_{\varepsilon \to 0} \frac{Z_d(\mathbb{P}_X^{(\varepsilon)}) - Z_d(\mathbb{P}_X)}{\varepsilon} &= \int \mathbb{P}_X(s) \frac{R_1(s)}{R_2(s)} d\nu - \int \delta_x(s) \frac{R_1(s)}{R_2(s)} d\nu \\ &\quad - \int \mathbb{P}_X(s) \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} - \frac{R_1(s)}{R_2(s)} \right) d\nu. \end{aligned} \tag{B.1}$$

So we have to find the influence function of $R_1(s)/R_2(s)$. By formally representing the differentiation of a functional $T(\mathbb{P}_X)$ by

$$\frac{\partial T(\mathbb{P}_X)}{\partial \lambda},$$

we have that the influence function of $R_1(s)/R_2(s)$ is given by

$$IF(R_1(s)/R_2(s), x) = \frac{R_2(s)\frac{\partial R_1(s)}{\partial \lambda} - R_1(s)\frac{\partial R_2(s)}{\partial \lambda}}{R_2(s)^2}.$$

But

$$\begin{aligned} R_1(s,\varepsilon) - R_1(s) &= \frac{\int t 1_{]0,s]}(t) d\mathbb{P}_X(t)}{\int 1_{]0,s]}(t) d\mathbb{P}_X^{(\varepsilon)}(t)} - \frac{\varepsilon \int t 1_{]0,s]}(t) d\mathbb{P}_X(t)}{\int 1_{]0,s]}(t) d\mathbb{P}_X^{(\varepsilon)}(t)} \\ &+ \frac{\varepsilon \int t 1_{]0,s]}(t) d\delta_x(t)}{\int 1_{]0,s]}(t) d\mathbb{P}_X^{(\varepsilon)}(t)} - \frac{\int t 1_{]0,s]}(t) d\mathbb{P}_X(t)}{\int 1_{]0,s]}(t) d\mathbb{P}_X(t)} \\ &= \frac{\int t 1_{]0,x_j]}(t) d(\mathbb{P}_X^{(\varepsilon)}(t) - \mathbb{P}_X(t))}{\int 1_{]0,s]}(t) d\mathbb{P}_X^{(\varepsilon)}(t)} \\ &- \frac{\int 1_{]0,s]}(t) d(\mathbb{P}_X^{(\varepsilon)}(t) - \mathbb{P}_X(t))}{\left(\int 1_{]0,s]}(t) d\mathbb{P}_X^{(\varepsilon)}(t)\right)\left(\int 1_{]0,s]}(t) d\mathbb{P}_X(t)\right)} \int t 1_{]0,s]}(t) d\mathbb{P}_X(t). \end{aligned}$$

We get

$$\begin{aligned} \lim_{\varepsilon \to 0} \frac{R_1(s,\varepsilon) - R_1(s)}{\varepsilon} &= \frac{\int t 1_{]0,s]}(t) d(-\mathbb{P}_X(t) + \delta_x)}{\int 1_{]0,s]}(t) d\mathbb{P}_X(t)} \\ &- \frac{\int 1_{]0,s]}(t) d(-\mathbb{P}_X(t) + \delta_x)}{\left(\int 1_{]0,s]}(t) d\mathbb{P}_X(t)\right)^2} \int t 1_{]0,s]}(t) d\mathbb{P}_X(t) \\ &= \frac{-\left(\int t 1_{]0,s]}(t) d\mathbb{P}_X(t)\right) + x 1_{]0,s]}(x)}{\int 1_{]0,s]}(t) d\mathbb{P}_X(t)} \\ &- \frac{-\left(\int 1_{]0,s]}(t) d\mathbb{P}_X(t)\right) + 1_{]0,s]}(x)}{\left(\int 1_{]0,s]}(t) d\mathbb{P}_X(t)\right)^2} \int t 1_{]0,s]}(t) d\mathbb{P}_X(t) \end{aligned}$$

and so

$$\frac{\partial R_1(s)}{\partial \lambda} = -R_1(s) + \frac{x 1_{]0,s]}(x)}{F(s)} + R_1(s) - \frac{R_1(s)}{F(s)} 1_{]0,s]}(x).$$

By treating $R_2(s)$ in the same manner, we have (we should not forget that we differentiate in the probability)

$$\begin{aligned} \frac{\partial R_1(s)}{\partial \lambda} &= \frac{x 1_{]0,s]}(x)}{F(s)} - \frac{R_1(s)}{F(s)} 1_{]0,s]}(x), \\ \frac{\partial R_2(s)}{\partial \lambda} &= \frac{x 1_{]s,+\infty]}(x)}{1 - F(s)} - \frac{R_2(s)}{1 - F(s)} 1_{]s,+\infty]}(x). \end{aligned}$$

Thus

$$\begin{aligned} \lim_{\varepsilon \to 0} \frac{R_1(s,\varepsilon) - R_1(s)}{\varepsilon} &= \left(\frac{1_{]0,s]}(x)}{R_2(s)F(s)} - \frac{R_1(s) 1_{]s,+\infty]}(x)}{R_2^2(s)(1 - F(s))}\right) x \\ &+ \left(\frac{R_1(s)}{R_2(s)(1 - F(s))} 1_{]s,+\infty]}(x) - \frac{R_1(s)}{R_2(s)F(s)} 1_{]0,s]}(x)\right). \end{aligned}$$

By replacing this limit with its expression in Equation (B.1) we get

$$\begin{aligned}\lim_{\varepsilon\to 0}\frac{Z_d(\mathbb{P}_X^{(\varepsilon)})-Z_d(\mathbb{P}_X)}{\varepsilon} &= \int \mathbb{P}_X(s)\frac{R_1(s)}{R_2(s)}d\nu - \int \delta_x(s)\frac{R_1(s)}{R_2(s)}d\nu \\ &+ \int \mathbb{P}_X(s)\left(\frac{R_1(s)}{R_2(s)^2(1-F(s)}) 1_{]s,+\infty]}(x) - \frac{1}{R_2(s)F(s)}1_{]0,s]}(x)\right)x d\nu \\ &+ \int \mathbb{P}_X(s)\left(\frac{R_1(s)}{R_2(s)F(s)}1_{]0,s]}(x) - \frac{R_1(s)}{R_2(s)(1-F(s))}1_{]s,+\infty]}(x)\right)d\nu.\end{aligned}$$

From this, the proof is concluded. □

ACKNOWLEDGMENTS

The authors thank the Editor-in-Chief and the anonymous reviewers for their very valuable comments and suggestions that helped to improve the quality of the paper.

REFERENCES

Agence Nationale de la Statistique et de la Démographie (ANSD) (2019). Enquête Sénégalaise Auprès des Ménages 2 (ESAM II, 2001-2002).
Report Website: `http://anads.ansd.sn/index.php/catalog/47/study-description`

Allison, P. D. (1978). Measures of inequality. *American Sociological Review*, 43:478–484.

Cowell, F. A. and Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141:1044–1072.

Cowell, F. A., Flachaire, E., and Bandyopadhyay, S. (2009). Goodness-of-Fit: An Economic Approach. *Distributional Analysis Research Programme (DARP 101)*. Discussion Paper. Department of Economics, University of Oxford.

Emad-Eldin, A. A. A. and Marilou, O. H. (1999). Nonparametric inference for Zenga's measure of income inequality. *Metron – International Journal of Statistics*. LVII(1-2):69–84.

Greselin, F., Pasquazzi, L., and Zitikis, R. (2010a). Zenga's new index of economic inequality, its estimation, and analysis of incomes in Italy. *Journal of Probability and Statistics*, 2010:1–26.

Greselin, F., Pasquazzi, L., and Zitikis, R. (2010b). Asymptotic theory for Zenga's new index of economic inequality. In "*The Proceedings of the 45th Scientific Meeting of the Italian Statistical Society*" (CLEUP Eds.), Padova.

Greselin, F., Pasquazzi, L., and Zitikis, R. (2014). Heavy tailed capital incomes: Zenga index, statistical inference, and ECHP data analysis. *Extremes*, 17:127–155.

Hulliger, B. and Schoch, T. (2009). Robust estimation of the quintile share ratio with bias reduction. Presented at the Swiss Statistics Meeting, October 30, 2009, Geneva.

Kpanzou, T. A. (2014). Asymptotic distribution of the quintile share ratio estimator. *Afrika Statistika*, 9(1):659–670.

Kpanzou, T. A., de Wet, T., and Neethling, A. (2013). Semi-parametric estimation of inequality measures. *South African Statistical Journal*, 47:33–48.

Lo, G. S. (2016). VB and R codes using Households databases available in the NSI's: A prelude to statistical applied studies. *African Journal to Applied Statistics*, 3:121–156. http://dx.doi.org/10.16929/ajas/206

Lo, G. S., Ngom, M., and Kpanzou, T. A. (2016). *Weak Convergence (IA). Sequences of Random Vectors*. SPAS Books Series, Saint-Louis, Senegal – Calgary, Canada.

Magurran, A. E. (1991). *Ecological Diversity and its Measurement*. Chapman and Hall.

Mergane, P. D. and Lo, G. S. (2013). On the functional empirical process and its application to the mutual influence of the Theil-like inequality measure and the growth. *Applied Mathematics*, 4: 986–1000. http://dx.doi.org/10.4236/am.2013.47136

Rousseau, R. (1993). Measuring concentration: sampling design issues as illustrated by the case of perfectly stratified samples. *Scientometrics*, 28:3–14.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer.

Wichura, M. J. (1970). On the construction of almost sure uniformly convergent random variable with given weakly convergent images laws. *Annals of Mathematical Statistics*, 41(1):284–291.

White, M. J. (1986). Segregation and diversity measures in population distribution. *Population Index*, 52:198–221.

Zenga, M. (1984). Proposta per un indice di concentrazione basato sui rapporti tra quantili di popolazione e quantili di reddito. *Giornale degli Economisti e Annali di Economia*, 43:301–326.

Zenga, M. (1990). Concentration curves and concentration indexes derived from them. In "Income and Wealth Distribution, Inequality and Poverty. Studies in Contemporary Economics" (C. Dagum, M. Zenga, Eds.), Springer, Berlin, Heidelberg, 94–110. https://doi.org/10.1007/978-3-642-84250-4_7

Zenga, M. (2006). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means, (provisional work). Rapporto di Ricerca 115, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Universita degli Studi du Milano-Bicocca.

REVSTAT – Statistical Journal
Volume 23, Number 2, April 2025, 207–221
https://doi.org/10.57805/revstat.v23i2.517

Generalizing the Heat Equation

Author: Christos P. Kitsos
– Department of Informatics, University of West Attica, Greece
xkitsos@uniwa.gr

Received: August 2022 Revised: March 2023 Accepted: March 2023

Abstract:

- The target of this paper is to generalize the Heat Equation, highly related with the Normal distribution. Therefore a generalization of the Normal distribution, the γ-order Normal distribution is introduced, which influences the entropy type information measures and offers the generalization of the Heat Equation.

Keywords:

- *Heat Equation; Wiener Process; Logarithm Sobolev Inequality; γ-order Normal.*

AMS Subject Classification:

- 60E05, 62B10, 62E99, 62P35.

1. INTRODUCTION

The most well known continuous Markov stochastic process $\{X(t); t \in [0, \infty)\}$ is the Brownian motion due to the name of the Scottish botanist Robert Brown (1773–1858), who in 1827 observed the phenomenon: minute particles, who were executing a continuous jittery and erratic motion. As Brownian motion was studied by Norbert Wiener (1894–1964) it is also known as Wiener Process. The basic framework is that for the stochastic process, as above, $X(t)$ is considered the x component of a particle, always as a function of time. Let at the time $t_0, X(t_0) = x_0$ and let the conditional probability density of $X(t + t_0)$ given $X(t_0) = x_0$ to be $p(x, t|x_0)$. In principle we postulate the probability law governing the tradition, is stationary in time and therefore $p(x, t|x_0)$ does not depend on t_0. Therefore the density function $p(x, t|x_0)$ we stipulate that for "small t" $X(t + t_0) \approx X(t_0)$ i.e. we assume $\lim_{t\to 0} p(x, t|x_0) = 0$, see for details (Levy, 1948; Feller, 1950; Papoulis, 1981). The Brownian motion can be applied to continuous time optimization in Economics, see Ross (1970) among others.

Since Albert Einstein (1879–1955) explained the behavior of the stochastic process physically, Einstein (1905), and he proved that eventually holds

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}, \tag{1.1}$$

with $D = 2RT/Nf$ the so called diffusion coefficient with R being the gas constant, T the temperature, N the Avogadro number and f the coefficient of friction, the diffusion equation (1.1) attracted a special interest. We mention that Schrödinger (1915) also investigated the Brownian Motion and worked with the Normal Inverse Gaussian (NIG), see Lahcene (2019). From an Analysis point of view is considered as a partial differential equation, Wazwaz (2002), among others, from physical point of view, known as Heat Equation (HE)-modeling the proportion of the amount of heat divided by the "amount" (precisely the mass) of the material, with a proportionality factor, which under a proper scale can be $D = 1/2$ i.e. (1.1) is reduced to

$$\frac{\partial^2 p}{\partial x^2} = 2\frac{\partial p}{\partial t}. \tag{1.2}$$

Eventually Probability theory is also involved as we can easily verify that the unique solution of (1.2), under the boundary conditions

- (**a**) $\lim_{t\to 0} p(x, t|x_0) = 0, \quad x \neq x_0$;
- (**b**) $p(x, t|x_0)$ is a density function in x thus
$p(x, t|x_0) \geq 0$ and $\int_{-\infty}^{\infty} p(x, t|x_0)dx = 1$;

is:

$$p(x, t|x_0) = \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{1}{2t}(x - x_0)^2\right\}, \tag{1.3}$$

i.e. if, without loss of generality, we assume that $x_0 = 0$, $p(x, t|x_0)$ coincides with the (distribution function of the) standard Normal distribution

$$\phi(x) = \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{1}{2}\left(\frac{x}{\sqrt{t}}\right)^2\right\} := \phi(x; 0, t). \tag{1.4}$$

The target of this paper is to work with a general form of (1.4) introducing an extra parameter, γ say, and form $\phi_\gamma(x; 0, t)$ which is discussed in Section 3, and evaluate a constant term $K = K(x, t; \gamma)$ so that (1.2) to be generalized as:

$$\frac{\partial^2 \phi_\gamma}{\partial x^2} = K \frac{\partial \phi_\gamma}{\partial t}. \tag{1.5}$$

Relation (1.5) is the introduced Generalized Heat Equation (GHE). This is discussed in Section 4. We shall prove that for $\gamma = 2$, relation (1.2) is obtained. Section 2 is devoted to explain the line of thought of the introduction of the extra parameter γ as a background to Section 3 where the γ-order Normal distribution is introduced.

2. BACKGROUND

The fact that we are facing the Heat Equation from a statistical point of view, provides evidence that no Physics is involved in this paper, but certainly an appropriate background from the Analysis point of view has been adopted to discuss the problem. Moreover it is emphasized that both in Physics and Analysis there is a strong theoretical insight with different approaches facing the Thermal or Heat equations. The pioneering work of Feller (1950) and Lévy (1948) covers the theoretical statistical background. Somebody might think that these are "old references", but they are the pioneering work in the field, the "back-bone" of the probability line of thought for the subject. We are referring to such pioneer work in this paper, as actually there are not that many from a Probability point of view. Moreover, we worked with the statistical oriented papers adopting an Analysis approach.

We shall use throughout this paper the statistical notation for example we let p to be the number of the involved parameters, and not n as n is the sample size in Statistics. We avoid to adopt a as a is the significance level. We are adopting γ instead, as the extra parameter of the γ-order Normal, see Section 3. Recall that in all entropy type research problems, as well, the three lines of thought: Statistics, Physics, Analysis are also met. Both Poincaré (Beckner, 1989) and Logarithm Sobolev inequalities (Carlen, 1991) are involved in such problems, where the energy $Ener_\mu(f)$ of a local integrable function f with $f \in L^2(\mathbb{R}^n, \mu)$ can be defined as:

$$Ener_\mu(f) := \mathbf{E}[|\nabla f|^2], \tag{2.1}$$

with ∇f the gradient of f and $\mathbf{E}$ the expectation of measure μ, i.e. $\mathbf{E}(f) = \int_{\mathbb{R}} f d\mu$, see for details Kitsos and Toulias (2010).

Let X be a random variable with probability density function (pdf) f on $\mathbb{R}^p$. Recall that, in principle, a function $f : \mathbb{R}^p \mapsto \mathbb{R}$ is said to be in a Sobolev space $W^{1,\gamma}(\mathbb{R}^p), \gamma \geq 1$, (Brezis, 1983; Takasi and Takashi, 2011), if:

- **(i)** $f \in L^\gamma(\mathbb{R}^p)$;
- **(ii)** its gradient $\nabla f \in L^\gamma(\mathbb{R}^p)$.

For $f^{1/2} \in W^{1,2}(\mathbb{R}^p)$, Fisher's entropy type information of f, $J(X)$ say, is defined:

$$J(X) = 4 \int_{\mathbb{R}^p} |\nabla \sqrt{f}|^2 dx,$$

see Carlen (1991). It is easy to see that:

$$\begin{aligned} J(X) &= \int_{\mathbb{R}^p} |\nabla \log f|^2 dx = \int_{\mathbb{R}^p} |\nabla f|^2 f^{-1} dx \\ &= \int_{\mathbb{R}^p} (\nabla f)(\nabla \log f) dx. \end{aligned} \tag{2.2}$$

Considering (2.2), recall that the Shannon entropy is defined as:

$$H(X) = -\int_{\mathbb{R}^p} f \log f dx, \tag{2.3}$$

while considering the family of densities of X parameterized by $\theta \in \Theta$, with Θ be a (compact, when limiting results are requested) subset of $\mathbb{R}^p$, the Fisher's (parametric) information matrix is:

$$I(\theta) := I_2(\theta) := \mathbf{E}_\theta |\nabla_\theta \log f_\theta(x)|^2. \tag{2.4}$$

The Vajda (parametric) information measure, Vajda (1973), is defined as an extension of (2.4)

$$I_\gamma(\theta) := \mathbf{E}_\theta |\nabla_\theta \log f_\theta(X)|^\gamma, \gamma \geq 1. \tag{2.5}$$

Comparing $I_2(\theta)$ and the general $I_\gamma(\theta)$, see (2.4) and (2.5), it is easy for somebody to think to obtain, based on $J = J_2(X)$ in (2.2), a general form, $J_\gamma(X)$ say. Such a procedure was also considered in the sense that Rényi's entropy, Rényi (1961), generalized Kullback-Leibler information, Kullback and Leibler (1951). For a recent study on Rényi's divergence measure see Jose and Abdul Sathar (2022).

Under this line of thought Kitsos and Tavoularis (2009) worked and generalized $J(X)$ to $J_\gamma(X)$, as well as Shannon exponential entropy $N_\gamma(X)$ through Shannon entropy $H(X)$, see (2.3). Indeed:

It is easy to consider that the entropy-type Fisher's information measure, see (2.2) can be extended to:

$$\begin{aligned} J_\gamma(X) &= \int_{\mathbb{R}^p} |\nabla \log f|^\gamma dx \\ &= \int_{\mathbb{R}^p} |\nabla f|^\gamma f^{1-\gamma} dx, \end{aligned} \tag{2.6}$$

with $\gamma \geq 1$ and f the pdf of a random variable X with $f^{1/\gamma} \in W^{1,\gamma}(\mathbb{R}^p)$. For $\gamma = 2$ we can verify that:

$$J_2(X) = J(X). \tag{2.7}$$

Moreover Shannon's exponentially entropy, $N_\gamma(X)$, can be extended and defined as:

$$N_\gamma(X) = const(\gamma, p) \exp\left\{ \frac{\gamma}{p} H(X) \right\}. \tag{2.8}$$

It is crucial that due to the generalized form of the entropy power, an extension of the Cramer-Rao inequality can be obtained, relating $J_\gamma(X)$ and $N_\gamma(X)$. This has been also obtained from the Heisenberg uncertainty inequality, see Beckner (1995). Indeed the following theorem holds, Kitsos and Tavoularis (2009).

Theorem 2.1. *It holds for the introduced entropy-type measures in (2.6) and (2.8) that* $J_\gamma(X)N_\gamma(X) \geq p$.

Due to Theorem 2.1, the well-known Cramer-Rao inequality is obtained, when $\gamma = 2$, while an example for the exponential family is discussed in Kitsos and Tavoularis (2009).

It has been pointed out, Beckner (1989), that the Logarithm Sobolev Inequality (LSI) can be also interpreted as sharpening the uncertainty principle, related through Theorem 1 to $J_\gamma(X)$ and $N_\gamma(X)$. That is why LSI was adopted and a new distribution (3.2) emerged, as it is discussed in Section 3.

3. THE γ-ORDER NORMAL DISTRIBUTION

Due to the discussion in Section 2 we consider the del Pino *et al.* (2004) extension of the del Pino and Dolbeault (2003) work for $g \in W^{1,\gamma}(\mathbb{R}^p)$ and $1 < \gamma < p$ with $\|g\|_\gamma = 1$ of the form:

$$\int_{\mathbb{R}^p} |g|^\gamma \log|g| \leq \frac{p}{\gamma^2}\log\left(c_\gamma \int_{\mathbb{R}^p} |\nabla g|^\gamma dx\right), \tag{3.1}$$

where the optimal constant $c_\gamma = c(\gamma, p)$ equals:

$$c(p,\gamma) = \frac{\gamma}{p}\left(\frac{\gamma-1}{e}\right)^{\gamma-1}\pi^{-\gamma/2}\left(\frac{\Gamma(\frac{p}{2}+1)}{\Gamma(p\frac{\gamma-1}{\gamma}+1)}\right)^{\gamma/p}, \gamma \in \mathbb{R}\setminus[0,1].$$

For the del Pino and Dolbeault (2003) LSI, externals are precisely Gaussians, Carlen (1991), while for del Pino *et al.* (2004) as it was pointed out by Kitsos and Tavoularis (2009) and presented by Kitsos and Toulias (2010), Toulias and Kitsos (2014) it is:

$$\phi_\gamma(x) = \phi_\gamma(x;\mu,\Sigma) = c(p,\gamma)[det\Sigma]^{-1/2}\exp\left\{-\frac{\gamma-1}{\gamma}Q^{\frac{1}{2}\frac{\gamma}{\gamma-1}}(x)\right\}, \tag{3.2}$$

with $x \in \mathbb{R}^p, Q(x) = \langle(x-\mu),\Sigma^{-1}(x-\mu)\rangle,\ c(p,\gamma) = \pi^{-\frac{p}{2}}\left(\frac{\gamma-1}{\gamma}\right)^{p\frac{\gamma-1}{\gamma}}\frac{\Gamma(\frac{p}{2}+1)}{\Gamma(p\frac{\gamma-1}{\gamma}+1)},\ \gamma \in \mathbb{R}\setminus[0,1]$ and $\langle a,b\rangle = ab^\top$ the inner product in $\mathbb{R}^n$, for $a, b \in \mathbb{R}^n$.

Notice that with $\gamma = 2$ we obtain the classical multivariate Normal, while with $\Sigma = I\sigma^2$, $I = \text{diag}\{1\} \in \mathbb{R}^{p\times p}$ and $p = 1$ we obtain the standard Normal distribution $\phi(x)$ with $\mu = 0$:

$$\begin{aligned}\phi_2(x;0,1) &= \frac{\Gamma(\frac{1}{2}+1)}{\Gamma(\frac{1}{2}+1)}\frac{(1/2)^{1/2}}{\sqrt{\pi}}\exp\left\{-\frac{1}{2}x^2\right\}\\ &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\} = \phi(x).\end{aligned} \tag{3.3}$$

The introduced γ-order Normal distribution $N(\mu,\Sigma;\gamma)$ is a Kotz-type distribution, as it was pointed out by Kitsos and Tavoularis (2009). In principle, heavy-tailed distributions are those probability distributions whose tails are not exponentially bounded, they have "heavier" tails than we usually assume, practically more that 0.05 as the standard Normal. The Normal

Inverse Gaussian $NIG(\cdot)$ is a very nice attempt to take into consideration the tails of the distribution and tries to cover the "fat tails" problem, which appears mainly in Finance studies, see Atkinson (1982), Barndorff-Nielsen (1997), Eberlein and Keller (1995), Rydberg (1997). As far as Environmental Economics concerned, for the uncertainty see Halkos and Kitsos (2018) where (3.2) was used. Notice that the Brownian motion will be normally distributed at all points in time, while a Lévy process, which is Generalized Hyperbolic (GH) distribution, can be GH at one point and might fail to be GH at another point in time, Podgórski and Wallin (2015). Both the GH and the γ-order GN are closed to affine transformations, while the Generalized Laplace (GL) and the NIG obey to the "closed under convolution" principle.

Although the existent background of the γ-order GN, $GN \sim N(\mu, \sigma^2 I; \gamma)$ depends on LSI, from which it "emerged" then the extra shape parameter γ provides an easy generalization of the multivariate Normal. The involved "international constant" $\left(\frac{\gamma-1}{\gamma}\right)^{\frac{\gamma-1}{\gamma}}$, Takasi and Takashi (2011), plays an important role in our development and is very essential that it comes through LSI. When $\gamma \rightarrowtail 1$ the Uniform distribution is obtained and when $\gamma \rightarrowtail \infty$ the Laplace distribution is obtained. Therefore we believe that γ-GN covers more prons than cons and this distribution outperforms comparing the other two. Moreover, it is a generalization of the distribution directly connected to the Heat Equation. It is rather helpful to obtain results, for Fisher's entropy type information. So there is a well working set of applications based on γ-GN and everything is reduced to the classical Normal when $\gamma = 2$. It is true that from NIG we can also reach Normal distribution but the NIG depends on four parameters, $NIG(\tau, \alpha, \delta, \mu)$, where the parameters are τ : for tail heaviness, α : asymmetry, δ : scale parameter, μ : location parameter. So there are 4 parameters involved and a rather complicated probability density function (pdf) based on modified Bessel function, of second kind. So it is rather difficult to be completely clear to those who are not mathematicians. Notice that the NIG belongs to the GH type distribution, while the γ-GN is a Kotz type distribution. The log-Normal (LN) it seems easier to be handled, but there is no shape parameter — actually the shape changes due to the fact that the logarithm is applied.

The introduced γ-ordered generalized Normal offers the possibility to approximate heavy-tailed distributions. The cumulative distribution function $\Phi_\gamma(z)$, for $z = \frac{x-\mu}{\sigma}$, for the $\phi_\gamma(x; \mu, \sigma^2)$ as in (3.2) has been obtained due to the following Theorem, see Toulias and Kitsos (2014).

Theorem 3.1. *Let X be a random variable from the univariate $N_\gamma(\mu, \sigma^2)$ with pdf $\phi_\gamma(x; \mu, \sigma^2)$ and F_γ the cdf. Let Φ_γ be the cdf of the standardized $z = \frac{1}{\sigma}(x-\mu) \sim N_\gamma(0,1)$. Then:*

$$\Phi_\gamma(z) = \frac{1}{2} + \frac{\sqrt{\pi}}{2\Gamma(\gamma_0+1)\Gamma(\gamma_1+1)} Erf_{\gamma_1}[\gamma_2 z], \tag{3.4}$$

with Erf being the usual error function and

$$\gamma_0 = \frac{\gamma-1}{\gamma}, \quad \gamma_1 = \frac{\gamma}{\gamma-1}, \quad \gamma_2 = \gamma_0^{\gamma_0}, \quad z = \frac{x-\mu}{\sigma}, \quad x \in \mathbb{R}.$$

Based on (3.4) a number of calculations have been proceeded and a part is presented here for different γ and p values, see Table 1. For a graph of a Bivariate 10-order generalized

Normal $N(0,1;10)$ see Kitsos and Toulias (2010) or Kitsos and Toulias (2010), while for the Kullback-LeibLer (K-L) information of two p-variate density functions from $N_\gamma(\mu_1, I\sigma^2)$ and $N_\gamma(\mu_2, I\sigma^2)$ see Kitsos and Toulias (2012).

Table 1: Probability mass $\mathbf{P}(\|X\| \leq 1)$, where $X \sim N(0, \sigma^2 I_p; \gamma)$, for various p and γ. * Uniform, ** Normal, *** Laplace.

γ	p		
	1	2	3
-2	0.6084	0.8100	0.8995
-1	0.5940	0.7737	0.8603
1^*	1.0000	1.0000	1.000
2^{**}	0.6827	0.9545	0.9973
5	0.6470	0.8953	0.9724
10	0.6390	0.8792	0.9614
$\pm\infty^{***}$	0.6320	0.8666	0.9510

Considering (3.2), the (elliptical contoured) γ-order Normal is reduced to a spherical contoured when $\Sigma = \sigma^2 I_p$, while when only one variable is involved and $\mu = 0$ then:

$$\phi_\gamma(x; 0, \sigma^2) = \frac{\Gamma(\frac{1}{2}+1)}{\Gamma(\gamma+1)} \frac{\gamma_2}{\sqrt{\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\}. \tag{3.5}$$

The distribution described in (3.5), $N_\gamma(0, \sigma^2)$ say, is fundamental for generalizing the Heat Equation as in (1.2). We introduce and prove the problem in the next section.

4. GENERALIZING THE HEAT EQUATION

Consider a Brownian motion $\{X(t); t \geq 0\}$ and assume that every increment is γ-order Normal distribution with mean 0 and variance $\sigma^2 t, \sigma$ is fixed, i.e. the definition of the Brownian motion is valid under the γ-order Normal distribution.

As usually $X(0) = 0$ and $X(t)$ is continuous at $t = 0$. We can assume without loss of generality that $\sigma = 1$, or that the Brownian motion is standard, i.e. $N_\gamma(0, t)$ is considered. That is, as we are working under $N_\gamma(0, t)$, considering (3.5) with $\sigma = \sqrt{t}$ the corresponding γ-order generalizing distribution function is:

$$\phi_\gamma(x; 0, t) = \frac{\lambda}{\sqrt{\pi t}} \exp\left\{-\gamma_0\left(\frac{x}{\sqrt{t}}\right)^{\gamma_1}\right\}, \tag{4.1}$$

with

$$\lambda = \frac{\Gamma(\frac{1}{2}+1)}{\Gamma(\gamma_0+1)}\gamma_2,$$

$$\gamma_0 = \frac{\gamma-1}{\gamma} > 0, \quad \gamma_1 = \frac{\gamma}{\gamma-1}, \quad \gamma_2 = \left(\frac{\gamma-1}{\gamma}\right)^{\frac{\gamma-1}{\gamma}} = \gamma_0^{\gamma_0}.$$

Lemma 4.1. *Let $\phi_\gamma = \phi_\gamma(x;0,t)$ as in (4.1). Then it holds:*

$$\frac{\partial \phi_\gamma}{\partial t} = \phi_\gamma(x;0,t)A(x;t,\gamma), \tag{4.2}$$

where $A(x;t,\gamma)$ is a well defined function.

Proof of Lemma 4.1: From (4.1) differentiate with respect to t to get:

$$\frac{\partial \phi_\gamma}{\partial t} = \frac{\lambda}{\sqrt{\pi}}(t^{-1/2})' \exp\left\{-\gamma_0\left(\frac{x}{\sqrt{t}}\right)^\gamma\right\} + \frac{\lambda}{\sqrt{\pi}}t^{-1/2}\left(\exp\left\{-\gamma_0\left(\frac{x}{\sqrt{t}}\right)^\gamma\right\}\right)'.$$

We let:

$$Q(x,t) = -\gamma_0\left(\frac{x}{\sqrt{t}}\right)^{\gamma_1}. \tag{4.3}$$

Then:

$$\begin{aligned}\frac{\partial \phi_\gamma}{\partial t} &= -\frac{\lambda}{\sqrt{\pi}}t^{-3/2}\exp\{Q(x,t)\} \\ &\quad + \frac{\lambda}{\sqrt{\pi}}t^{-1/2}\exp\{Q(x,t)\}\cdot Q'(x,t) \\ &= \frac{\lambda}{\sqrt{\pi t}}\exp\{Q(x,t)\}\left(-\frac{1}{2}t^{-1} + \frac{1}{2}x^{\gamma_1}t^{-\frac{\gamma_1+2}{2}}\right) \\ &= \phi_\gamma(x;0,t)A(x;t,\gamma). \qquad (4.4)\end{aligned}$$

□

Lemma 4.2. *Let $\phi_\gamma = \phi_\gamma(x;0,t)$ as in (4.1). Then the following are true:*

$$\frac{\partial \phi_\gamma}{\partial x} = \phi_\gamma(x;0,t)B_1(x;t,\gamma),$$

$$\frac{\partial^2 \phi_\gamma}{\partial x^2} = \phi_\gamma(x;0,t)B_2(x;t,\gamma),$$

with $B_1(x;t,\gamma), B_2(x;t,\gamma)$ well defined functions.

Proof of Lemma 4.2: Recall (4.3) and that $\gamma_0\gamma_1 = 1$. Differentiating with respect to x it is

$$\begin{aligned}\frac{\partial \phi_\gamma}{\partial x} &= \frac{\lambda}{\sqrt{\pi t}}\exp\{Q(x,t)\}'_x \\ &= \frac{\lambda}{\sqrt{\pi t}}\exp\{Q(x,t)\}(Q(x,t))' \\ &= \phi_\gamma(x;0,t)\left(-\gamma_0\gamma_1 t^{-\frac{1}{2}\gamma_1}x^{\gamma_1-1}\right) \\ &= \phi_\gamma(x;0,t)\left(-t^{-\frac{1}{2}\gamma_1}x^{\gamma_1-1}\right) \\ &= \phi_\gamma(x;0,t)B_1(x;t,\gamma). \qquad (4.5)\end{aligned}$$

Thus, from (4.5) we get:

$$\begin{aligned}\frac{\partial^2 \phi_\gamma}{\partial x^2} &= \phi'_\gamma(x;0,t)B_1(x;t,\gamma) + \phi_\gamma(x;0,t)B'_1(x;t,\gamma) \\ &= \phi_\gamma(x;0,t)B_1^2(x;t,\gamma) + \phi_\gamma(x;0,t)\left(-t^{-\frac{1}{2}\gamma_1}\frac{1}{\gamma_1-1}x^{\gamma_1-2}\right) \\ &= \phi_\gamma(x;0,t)B_2(x;t,\gamma). \qquad (4.6)\end{aligned}$$

□

Therefore we can state and prove the following Theorem, generalizing the Heat Equation (1.2) as in (4.4).

Theorem 4.1. *There exists a well defined function $K = K(x;t,\gamma)$ such that*

$$\frac{\partial^2 \phi_\gamma}{\partial x^2} = K \frac{\partial \phi_\gamma}{\partial t}. \tag{4.7}$$

Proof of Theorem 4.1: From (4.4) it is, as $\gamma_1 = \frac{\gamma}{\gamma-1}$,

$$A(x;t,\gamma) = -\frac{1}{2}t^{-1} + \frac{1}{2}x^{\frac{\gamma}{\gamma-1}}/t^{\frac{3\gamma-1}{2(\gamma-1)}}. \tag{4.8}$$

From (4.5) and (4.6) respectively we get

$$B_1(x;t,\gamma) = -t^{-\frac{\gamma}{2(\gamma-1)}}x^{\frac{1}{\gamma-1}}, \tag{4.9}$$

$$\begin{aligned} B_2(x;t,\gamma) &= B_1^2(x;t,\gamma) + \left(-t^{-\frac{1}{2}\frac{\gamma}{\gamma-1}}\frac{1}{\gamma-1}x^{\frac{2-\gamma}{\gamma-1}}\right) && (4.10)\\ &= \left(-t^{-\frac{\gamma}{2(\gamma-1)}}x^{\frac{1}{\gamma-1}}\right)^2 + \left(-t^{-\frac{1}{2}\frac{\gamma}{\gamma-1}}\frac{1}{\gamma-1}x^{\frac{2-\gamma}{\gamma-1}}\right)\\ &= t^{-\frac{\gamma}{\gamma-1}}x^{\frac{2}{\gamma-1}} - \frac{1}{\gamma-1}t^{-\frac{1}{2}\frac{\gamma}{\gamma-1}}x^{\frac{2-\gamma}{\gamma-1}}. && (4.11)\end{aligned}$$

Therefore:

$$\frac{\dfrac{\partial^2\phi_\gamma}{\partial x^2}}{\dfrac{\partial\phi_\gamma}{\partial t}} = \frac{t^{-\frac{\gamma}{\gamma-1}}x^{\frac{2}{\gamma-1}} - \frac{1}{\gamma-1}t^{-\frac{1}{2}\frac{\gamma}{\gamma-1}}x^{\frac{2-\gamma}{\gamma-1}}}{\dfrac{1}{2}\left(-\dfrac{1}{t} + \dfrac{x^{\frac{\gamma}{\gamma-1}}}{t^{\frac{3\gamma-2}{2(\gamma-1)}}}\right)} := K(x;t,\gamma) := K, \tag{4.12}$$

i.e. $\dfrac{\partial^2\phi_\gamma}{\partial x^2} = K\dfrac{\partial\phi_\gamma}{\partial t}$ and K is well defined as in (4.12). □

Corollary 4.1. *With $\gamma = 2$ the classical homogeneous Heat Equation is true, i.e. (1.2) holds.*

Proof of Corollary 4.1: From (4.12) with $\gamma = 2$ we obtain

$$K(x;t,2) = \frac{\dfrac{x^2}{t^2} - \dfrac{1}{t}}{\dfrac{1}{2}\left(-\dfrac{1}{t} + \dfrac{x^2}{t^2}\right)} = 2. \tag{4.13}$$ □

We can arrive at the same result through $\phi_2(x;0,t)$, verifying relation (1.2). Indeed:

$$\frac{\partial\phi_2}{\partial t} = \frac{1}{2}\phi_2(x;0,t)\left(-\frac{1}{t} + \frac{x^2}{t^2}\right),$$

$$\frac{\partial\phi_2}{\partial x} = \phi_2(x;0,t)\left(-\frac{x}{t}\right),$$

$$\frac{\partial^2\phi_2}{\partial x^2} = \phi_2(x;0,t)\left(\frac{x^2}{t^2} - \frac{1}{t}\right) = 2\frac{\partial\phi_2}{\partial t}.$$

Corollary 4.2. *With $t = 1$ it holds*

$$\frac{\partial \phi_\gamma}{\partial x} = \phi_\gamma(x)\Big(-x^{\frac{1}{\gamma-1}}\Big). \tag{4.14}$$

Proof of Corollary 4.2: From (4.5) and (4.9)

$$\begin{aligned}\frac{\partial \phi_\gamma}{\partial x} &= \phi_\gamma(x;0,1)B_1(x;1,\gamma)\\ &= \phi_\gamma(x;0,1)\Big(-x^{\frac{1}{\gamma-1}}\Big).\end{aligned}$$

□

Corollary 4.3. *From (4.14) with $\gamma = 2$ the well known relation*

$$\phi_2'(x) = -x\phi_2(x) \tag{4.15}$$

holds.

Consider the Generalized Heat Equation (GHE) as in (4.7). In Table 2 and Figures 1 and 2 we present, for given values of γ, the corresponding γ-order GN, $\phi_\gamma(x;0,t)$ as in (4.1), with the value of λ as in (4.1). The calculations were performed through MATLAB and the corresponding $K = K_\gamma = K(x;t,\gamma)$ is also evaluated and presented.

Table 2: Evaluating $\phi_\gamma(x;0,t)$ as (4.1) with $\lambda = \frac{\Gamma(\frac{1}{2}+1)}{\Gamma(\gamma_0+1)}$ and K_γ for given values of γ, see (4.1) and (4.7) the GHE.

γ	ϕ_γ	K_γ
1.5	$\frac{\lambda}{\sqrt{\pi t}}\exp\Big\{-0.3\Big(\frac{x}{\sqrt{t}}\Big)^3\Big\}$	$-\dfrac{\frac{1}{2}t^{3/2}-0.5\Big(\frac{x}{\sqrt{t}}\Big)^2 xt}{t^{3/2}\Big(-2\Big(\frac{x}{\sqrt{t}}\Big)+\Big(\frac{x}{\sqrt{t}}\Big)^4\Big)}$
1.9	$\frac{\lambda}{\sqrt{\pi t}}\exp\Big\{-0.4736842105\Big(\frac{x}{\sqrt{t}}\Big)^{2.1}\Big\}$	$-\dfrac{\frac{1}{2}t^{3/2}-0.5\Big(\frac{x}{\sqrt{t}}\Big)^{1.1} xt}{t^{3/2}\Big(-1.1\Big(\frac{x}{\sqrt{t}}\Big)^{0.1}+\Big(\frac{x}{\sqrt{t}}\Big)^{2.2}\Big)}$
2	$\frac{1}{\sqrt{2\pi t}}\exp\Big\{-0.5\Big(\frac{x}{\sqrt{t}}\Big)^2\Big\}$	2
2.1	$\frac{\lambda}{\sqrt{\pi t}}\exp\Big\{-0.5\Big(\frac{x}{\sqrt{t}}\Big)^{1.9}\Big\}$	$\dfrac{\frac{1}{2}t^{3/2}-0.5\Big(\frac{x}{\sqrt{t}}\Big)^{0.9} xt}{t^{3/2}\Big(0.9\Big(\frac{x}{\sqrt{t}}\Big)^{-0.09}-\Big(\frac{x}{\sqrt{t}}\Big)^{1.81}\Big)}$
2.5	$\frac{\lambda}{\sqrt{\pi t}}\exp\Big\{-0.6\Big(\frac{x}{\sqrt{t}}\Big)^{1.6}\Big\}$	$\dfrac{\frac{1}{2}t^{3/2}-0.5\Big(\frac{x}{\sqrt{t}}\Big)^{0.6} xt}{t^{3/2}\Big(0.66\Big(\frac{x}{\sqrt{t}}\Big)^{-0.33}-\Big(\frac{x}{\sqrt{t}}\Big)^{1.33}\Big)}$
3	$\frac{\lambda}{\sqrt{\pi t}}\exp\Big\{-2/3\Big(\frac{x}{\sqrt{t}}\Big)^{3/2}\Big\}$	$\dfrac{t^{3/2}-xt\sqrt{\frac{x}{\sqrt{t}}}}{t^{3/2}-2xt\sqrt{\frac{x}{\sqrt{t}}}}$

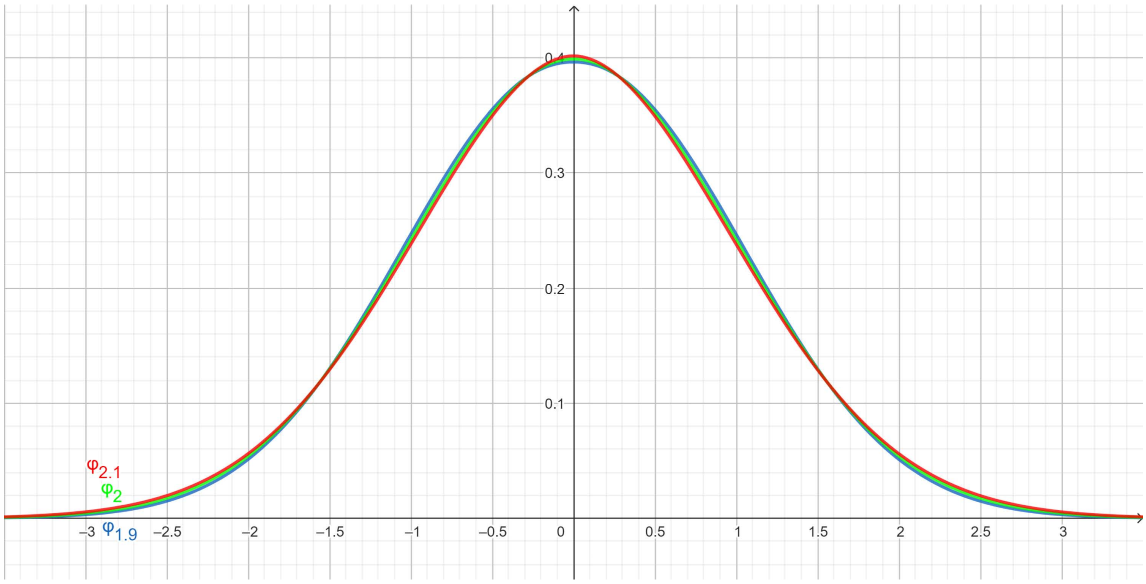

Figure 1: Plots of $\phi_\gamma(x;0,1)$ for values of γ close to 2. Namely $\gamma = 1.9, 2$ and 2.1.

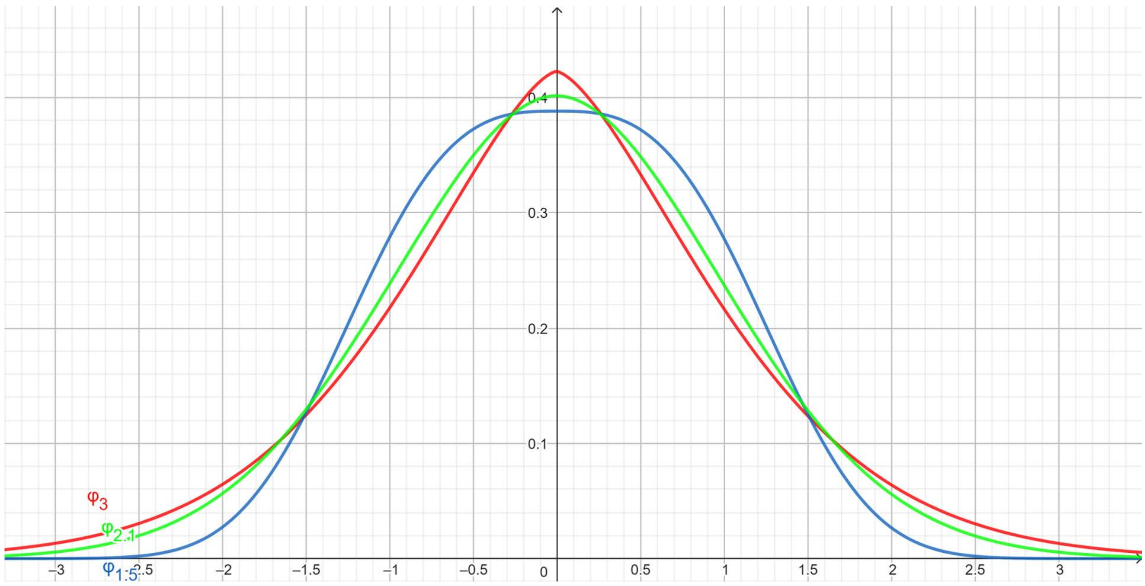

Figure 2: Plots of $\phi_\gamma(x;0,1)$ for values of γ. Namely $\gamma = 1.5, 2.1$ and 3.

Therefore not only a general form of the Heat Equation was provided due to Theorem 4.1, but also minor results can be generalized due to the γ-order Normal. Notice that when $t = 1$ (equivalently $\sigma^2 = 1$) the values of ϕ_γ and K_γ are simplified.

It is clear that for values of $\gamma = 1.9, 2, 2.1$ i.e close to 2, the corresponding graphs are close to the usual Normal, see Figure 1, but for values of $\gamma = 1.5, 2.1, 3$ i.e the corresponding graphs provide evidence for their "fat tails", see Figure 2.

5. DISCUSSION

From a statistical point of view the Heat Equation has faced little attention. Most of the work is referring to the Brownian motion process as one of the two Lévy processes — the other one is the Poisson process. In Medical Physics, in their recent paper, Mashekova *et al.* (2022), working on breast cancer and the involved mathematical equations, came across the Heat Equation, due to the fact that the thermal diffusivity, (α) in their notation, the time t and the penetration depth (δ) linked, in their notation, as $\delta = 3.65\sqrt{\alpha t}$ satisfy the transient conduction equation for the temperature $T = T(x,t)$ in one dimension

$$\frac{\partial^2 T(x,t)}{\partial x^2} = \frac{1}{\alpha}\frac{\partial T(x,t)}{\partial t}.$$

For a given value of the thermal diffusivity and the depth of tumor the cold stress time is evaluated. They also provided a review of the mathematical models concerning the subject. That is, the Heat Equation participated in problems facing the underground application rather through Mathematics, than Statistics, as the Gaussian is rather more familiar that any other distribution. But the Statistical background is solid and we only try to offer a more general Gaussian to face applications with "fat tails", and still obey to the GHE, as in (4.6).

From a physical point of view the appropriate name of the Heat Equation is diffusion equation with a source term. From an Analysis point of view the Heat Equation is known as a parabolic differential equation: briefly it describes the distribution of the heat flow (into/out) of a material in a given space over time. The proportional factor is the specific heat capacity of the material. The Heat Equation is a typical example of a continuity equation and it was related to the Gaussian. Schrödinger (1915) was investigating the Brownian motion and he came across, as we already mentioned, to the Normal Inverse Gaussian (NIG), see Seshadri (1997), Lahcene (2019). The Brownian movement provides food for thought to continuous optimization models in Economics, Ross (1970) among others. Let $F_\gamma(x)$ to be the cdf of the rv $X \sim N(0,1;\gamma)$. For given different shape values for γ, the corresponding probability values have been evaluated, see also Halkos and Kitsos (2018), in Table 3.

Table 3: Values of cdf of γ-order generalized Normal with $\mu = 0$, $\sigma = 1$.

γ	$GN(0,1;\gamma)$	
	$F_\gamma(-3)$	$F_\gamma(2)$
2	0.0013	0.9772
3	0.0071	0.9598
10	0.0193	0.9396
$-1/10$	0.1656	0.8111

Let $F_{NIG}(x)$ to be the cdf of a rv $X \sim NIG(\tau,\alpha,\delta,\mu) = NIG(\tau,0,1,0)$. For given values of tail heaviness taken from Table 3 and keeping asymmetry parameter as well as location parameter equal to zero, while the scale parameter is one, we present in Table 4 the values of the cdf $F_{NIG}(x)$, with $x = -3, 2$ as for the $N(0,1;\gamma)$.

Table 4: Values of cdf of $NIG(\tau, 0, 1, 0)$ for different values of heavy tailness parameter τ.

τ	$NIG(\tau, 0, 1, 0)$	
	$F_{NIG}(-3)$	$F_{NIG}(2)$
0.0014	0.1018	0.9718
0.0193	0.0953	0.9701

Although the $NIG(\cdot)$ did not appear to our procedure, still it is clear that it provides "heavy tails" and this is evidence that the shape parameter γ in $N(0, 1; \gamma)$ "adjusts" the value of tail as well as shape.

Now, with the γ-order Normal, we described a general family of distributions with a particular extra shape parameter γ. The shape parameter can describe "fat tails distributions", "close" to what is known as Gaussian or Normal, see Table 2. Consider that with $\sigma = \sqrt{t} = 1$ the values of $\phi(0, 1; \gamma) = \phi_\gamma$ and $K_\gamma = K(x, 1; \gamma)$ are simplified. In limited cases, as $\gamma \to 1$ the Uniform distribution can be obtained or the Laplace, when $\gamma \to \infty$, Toulias and Kitsos (2014). Notice that the "international constant" $\gamma_0^{\gamma_0}$ plays an important role to the described formulation. That certainly needs more investigation as the statistical generalization might offer chance for food of thought under Mathematical or Physical considerations. Moreover the generalization of the Normal provide generalized entropy type information measures, Kitsos and Toulias (2012) and possible engineering implementations, Papoulis (1981). Therefore we open a subject which can offer a number of different lines of thought to work in future. We shall try to continue the statistical line of generalization we are creating.

ACKNOWLEDGMENTS

I would like to thank Prof. N. Halidias, Univ. of Aegean, for useful discussions during the preparation of this paper. The comments of the referees are very much appreciated as eventually improved this paper.

REFERENCES

Atkinson, A. C. (1982). The simulation of generalized inverse Gaussian and hyperbolic random variables. *SIAM Journal on Scientific and Statistical Computing*, 3(4):502–515.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13.

Beckner, W. (1989). A generalized Poincaré inequality for Gaussian measure. *Proc. of the Amer. Math. Soc.*, 105:397–400.

Beckner, W. (1995). Pitt's inequality and the uncertainty principle. *Proc. of the Amer. Math. Soc.*, 123(6):1897–1905.

Brezis, H. (1983). *Analyse Fonctionnelle: Théorie et Applications*. Masson, Paris.

Carlen, E. A. (1991). Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 101(1):194–211.

del Pino, M. and Dolbeault, J. (2003). The optimal Euclidean L^p-Sobolev logarithmic inequality. *Journal of Functional Analysis*, 197(1):151–161.

del Pino, M., Dolbeault, J., and Gentil, I. (2004). Nonlinear diffusions, hypercontractivity and the optimal L^p-Euclidean logarithmic Sobolev inequality. *Journal of Mathematical Analysis and Applications*, 293(2):375–388.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17:549.

Feller, W. (1950). *An Introduction to Probability Theory and Its Applications*. Wiley.

Halkos, G. and Kitsos, C. P. (2018). Uncertainty in environmental economics: The problem of entropy and model choice. *Elsevier*, 60(C):127–140.

Jose, J. and Abdul Sathar, E. I. (2022). Rényi entropy of k-records: properties and applications. *REVSTAT – Stat. J.*, 20(4):481–500.

Kitsos, C. P. and Tavoularis, N. K. (2009). Logarithmic Sobolev inequalities for information measures. *IEEE Transactions on Information Theory*, 55(6):2554–2561.

Kitsos, C. P. and Toulias, T. L. (2010). Entropy inequalities for the generalized Gaussian. *Proceedings of the 32nd International Conference on Information Technology Interfaces*, June 21–24, 2010, Cavtat/Dubrovnik, Croatia.

Kitsos, C. P. and Toulias, T. L. (2010). New information measures for the generalized normal distribution. *Inf.*, 1:13–27.

Kitsos, C. P. and Toulias, T. L. (2012). Bounds for the generalized entropy − type information. *J. of Communications and Computer*, 9(1):56–64.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.

Lahcene, B. (2019). On extended normal inverse Gaussian distribution: theory, methodology, properties and applications. *American Journal of Applied Mathematics and Statistics*, 7(6):224–230.

Levy, P. S. (1948). *Processus Stochastiques et Mouvement Brownien*. Gauthier − Villars, Paris.

Mashekova, A., Zhao, Y., Ng, E. Y. K., Zarikas, V., Fok, S. C., and Mukmetov, O. (2022). Early detection of the breast cancer using infrared technology – A comprehensive review. *Thermal Science and Engineering Progress*, 27(1):101142.

Papoulis, A. (1981). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill.

Podgórski, K. and Wallin, J. (2015). Convolution-invariant subclasses of generalized hyperbolic distributions. *Communications in Statistics – Theory and Methods*, 45(1):98–103.

Ross, S. M. (1970). *Applied Probability Models with Optimization Applications*. Dover Pub, New York.

Rényi, A. (1961). On measures of entropy and information. In *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, Berkeley, 1:547–561.

Rydberg, T. H. (1997). The normal inverse Gaussian Lévy process: simulation and approximation. *Communications in Statistics. Stochastic Models*, 13(4):887–910.

Seshadri, V. (1997). Halphen's Laws. In *Encyclopedia of Statistical Sciences* (Kotz, S., Read, C.B., Banks, D.L.), New York: Wiley, 1:302–306.

Schrödinger, E. (1915). Zur Theorie Der Fall-und Steigversuche an Teilchen Mit Brownscher Bewegung. *Physikalische Zeitschrift*, 16:289–295.

Takasi, S. and Takashi, S. (2011). *Applied Analysis: Mathematical Methods in Natural Sciences*. Imperial College Press.

Toulias, T. L. and Kitsos, C. P. (2014). On the properties of the generalized normal distribution. *Discussiones Mathematicae Probability and Statistics*, 34(1-2):35–49.

Vajda, I. (1973). χ^2-divergence and generalized Fischer's information. In *Proc. 6th Prague Conf. Inf. Theory, Stat. Decision Function Random Processes*, 873–886.

Wazwaz, A.-M. (2002). *Partial Differential Equations*. Swets & Zertilinger B.V. Lisse.

REVSTAT – Statistical Journal
Volume 23, Number 2, April 2025, 223–233
https://doi.org/10.57805/revstat.v23i2.537

Randomly Weighted Averages on Multivariate Dirichlet Distributions with Generalized Parameters

Authors: Hazhir Homei ✉
– Department of Mathematics, Statistics, and Computer Science, University of Tabriz, Tabriz, Iran
homei@tabrizu.ac.ir

Saralees Nadarajah
– Department of Mathematics, Howard University, Washington DC 20059, USA
saraleesan.nadarajah@howard.edu

Ameneh Taherkhani
– Department of Mathematics, Statistics, and Computer Science, University of Tabriz, Tabriz, Iran
a.taherkhani@tabrizu.ac.ir

Received: February 2022 Revised: April 2023 Accepted: April 2023

Abstract:

- We study the distributional properties of the product of random stochastic matrices by using the Dirichlet distribution. Our observations widely generalize some known results on the randomly weighted averages about multivariate Dirichlet distributions. We present a new method for approximating the distribution that can be used for the multivariate random variables with a slight change.

Keywords:

- *Cauchy composition test; random stochastic matrices; real lifetime.*

AMS Subject Classification:

- 65Cxx, 62E15.

✉ Corresponding author

1. INTRODUCTION

The product of random stochastic matrices is one of the concepts that has attracted the attention of many mathematicians (Touri, 2012). The behavior of this concept in science and engineering has been investigated under some specific assumptions. The studies and applications of this concept are closely related to the studies of averaging dynamics. Let us note that some applications are presented by Touri (2012).

Here, we consider the independent random matrices each of which has independent rows and are identically distributed with the Dirichlet distribution, and investigate some distributional and statistical properties of the product of those random matrices. For this purpose, we have addressed the mixture random variables and followed their traces in applied fields. Here we list some important applications:

Cauchy Composition Test. Let p_i be the individual p-value for $i = 1, 2, ..., d$. We define the Cauchy combination test statistic as

$$T = \sum_{i=1}^{d} \omega_i \tan\{(0.5 - p_i)\pi\}, \tag{1.1}$$

where the weights ω_i's are nonnegative and $\sum_{i=1}^{d} \omega_i = 1$. Given that p_i is uniformly distributed between 0 and 1 under the null, the component $\tan\{(0.5 - p_i)\pi\}$ follows a standard Cauchy distribution. To overcome these challenges, Liu and Xie (2020) propose a new test that takes advantage of the Cauchy distribution. Their test statistic has a simple form and is defined as a weighted sum of Cauchy transformation of individual p-values. We emphasize that some of the results have been examined in a state where W_i are random variables. Here, we define this statistic as a vector and obtain the properties of the statistic under some specific assumptions. Of course, some of the results obtained in our study indicate that these properties also apply to other distributions in addition to the Cauchy distribution. For the notation and discussions we refer the reader to Liu and Xie (2020).

Real Lifetime. Suppose that $X_1, ..., X_r$ are random variables, and the coefficient of the impact of environment on lifetime, $Y_1, ..., Y_r$, are independent random variables with Gamma distribution that indicate the lifetime in the laboratory conditions. The following random variable is called real lifetime: $T = \sum_{i=1}^{r} Y_i X_i$. In this paper, the random variable T presented in the research by Homei and Nadarajah (2018) has been generalized to vector random variables. In this paper, some results and characterizations have been studied in vector random variables for T; we also answer some questions asked there (in particular, we have generalized Homei and Nadarajah (2018, Theorem 2.2) to the multivariate case in Theorem 3.4 below).

Solving Some Differential Equations. Using the properties of the Beta distribution, Homei (2021) and Hadad *et al.* (2021) solved some differential equations. Indeed, the given equations may be solved by very long calculations. Let us recall that Theorem 1 of Homei (2015) identifies the distribution of (the 1-dimensional) Z from the distributions of X_i's by means of the following differential equation:

$$\frac{(-1)^{n^*-1} d^{n^*-1}}{(n^*-1)! dz^{n^*-1}} S(F_Z, z) = \prod_{i=1}^{n} \frac{(-1)^{m_i-1}}{(m_i-1)!} \frac{d^{m_i-1}}{dz^{m_i-1}} S(F_{X_i}, z), \tag{1.2}$$

where F_Z denote the cumulative distribution function of a random variable Y and $S(F_Z, z)$ is Stieltjes transform defined by

$$S(H, z) = \int \frac{1}{z - x} H(dx), \quad z \in \mathbf{C} \cap (\mathsf{supp} H)^c. \tag{1.3}$$

Here, $\mathsf{supp} H$ is the support of H; see Homei (2015).

Random Convex Combination. A stochastic linear combination

$$\hat{C}_1 \cdot Z_1 + \hat{C}_2 \cdot Z_2 + \cdots + \hat{C}_m \cdot Z_m \tag{1.4}$$

of random variables $Z_1, ..., Z_m$ where $\hat{C}_i, 1 \leq i \leq m$, are random variables such that

- **(i)** $\hat{C}_i \geq 0,\ 1 \leq i \leq m,$ and
- **(ii)** $\sum_{i=1}^m \hat{C}_i = 1, a.s.,$

is called a random convex combination of the random variables $Z_1, ..., Z_m$ (for more details see Homei, 2015).

Of course, another form of real lifetime is provided by Homei (2015), which is not far from the statistic defined by Liu and Xie (2020). Let $Z_i, i = 1, ..., n$, be the lifetime measured in a laboratory and $0 \leq C_i \leq 1$ be the random effect of the environment on it, so $C_i Z_i \leq Z_i$ and thus $\sum_{i=1}^n C_i Z_i$ is the average lifetime in the environment, see Homei (2021), Homei and Nadarajah (2018). If Y_i is the real lifetime in the i-th area, $C_i = \dfrac{Y_i}{\sum_{i=1}^n Y_i}$ is the random effect ratio in the i-th area. Therefore, it is clear that a good choice for the distribution of $\mathbf{C} = \langle C_1, ..., C_n \rangle$ can be Dirichlet distribution. It is important that the product of random stochastic matrices connect us directly to stochastic linear combination.

The structure of the paper is as follows: the next subsection gives the motivation of the research in this paper, and lists some innovations. In Section 2, the mean, variance and moments of the randomly weighted averages on random vectors with Dirichlet distributions are obtained. In Section 3, a new method for calculating the distribution of randomly weighted averages are presented, and the distribution of this random vector is calculated under some specific assumptions. In Section 4 using simulation we suggest an approximation for the distribution of randomly weighted averages.

1.1. Motivation and Innovation

For obtaining the distribution of the randomly weighted averages, one suggested method is using Stieltjes transforms, which is a very complex strategy. In this paper, a novel method for obtaining those distributions in the multivariate case is presented. For evaluating this new method we have reproved Theorem 3.4, for which we have also proved some new auxiliary theorems. In spite of the fact that some special cases of those auxiliary results are well known theorems, but these results in the general form that are presented here seem to be new, with much simpler and more elementary proofs. For keeping the copyright of the results, some preliminary versions had been put on the arXiv (Homei, 2016), and re-emphasized in Homei and Nadarajah (2018) that we would generalize the earlier results in the future. At the end we also presented some of them in a national conference (Homei, 2017b).

1.1.1. Some History and Earlier Results

Two ideas of Soltani and Roozegar (2020), i.e., (i) defining the random coefficients, and (ii) using Stieltjes transformations for obtaining the distribution of the randomly weighted averages, have been taken from Soltani and Homei (2009). In fact, the class of randomly weighted averages defined in those two papers are much more restricted in comparison to the ones defined in this paper, since their defined coefficients have Dirichlet distributions with limited parameters (which are positive integer numbers). But in this paper, the random vectors are taken to have Dirichlet distributions without any limitations on their parameters; notice that the assumption $\sum_i \alpha_i = 1$ in Soltani and Roozegar (2020, Theorem 3.1) shows that the results of Soltani and Roozegar (2020) (both Theorems 2 and 5) are weaker than ours here.

To make a long story short, the main result of Soltani and Roozegar (2020), which is their Theorem 2.1, is not useful for a large part of the class of randomly weighted averages defined in the present paper; and their Theorem 3.1 is a very special case of our Theorem 3.4 below, which was published before in Homei (2016). Therefore, our method is much stronger, and even more elementary at the same time; indeed, Theorem 2.1 in Soltani and Roozegar (2020) should be improved for being usable in our class of randomly weighted averages. A complete and general proof for that theorem is in the possession of the first author, and is planned for a publication in future. For further references, we invite the readers to consult Homei (2017a) and Pitman (2018, p. 57).

2. PRODUCT MOMENTS OF RANDOM CONVEX COMBINATION

The concept of product of random stochastic matrices motivated us to discuss the distributional properties of random convex combinations. These properties include the product moments, and the mean and the variance of components. Throughout the paper, $\langle W_1, ..., W_r \rangle$ is called random coefficient vector of environmental effect in r-position. As mentioned in the introduction, the rows are independent and have Dirichlet distribution in random stochastic matrices.

Theorem 2.1. *Suppose that the independent random vectors $\mathbf{X}_1, ..., \mathbf{X}_r$ have identical distributions with mean μ and variance $\boldsymbol{S}$ and that the random vector $W = \langle W_1, ..., W_r \rangle$ is independent from $\mathbf{X}_1, ..., \mathbf{X}_r$ such that $\sum_{j=1}^r W_j = 1$, a.s. Then the mean and the variance of $\boldsymbol{Z} = \sum_{j=1}^r W_j \mathbf{X}_j$ are*

$$E(\boldsymbol{Z}) = \mu \quad \text{and} \quad \mathrm{Var}(\boldsymbol{Z}) = \sum_{j=1}^{r} EW_j^2 \boldsymbol{S},$$

where $\boldsymbol{S}$ is variance-covariance matrix.

Proof: By using the double conditional expectation and the conditional variance, the result is proved. □

The following theorem results in product moments of random convex combination when we consider the random vectors by Dirichlet distribution.

Theorem 2.2. *Suppose that the independent random vectors $\mathbf{X}_1, ..., \mathbf{X}_r$ have, respectively,*

$$\text{Dirichlet}(n_{11}, ..., n_{1k}),\ ...,\ \text{Dirichlet}(n_{r1}, ..., n_{rk})$$

distributions and that the random vector $W = \langle W_1, ..., W_r\rangle$ is independent from $\mathbf{X}_1, ..., \mathbf{X}_r$ and has $\text{Dirichlet}(\alpha_1, ..., \alpha_r)$ distribution.

Then the product moments in $(s_1, ..., s_k)$ of $\mathbf{Z} = \sum_{j=1}^{r} W_j \mathbf{X}_j$ are

$$\begin{aligned} E(L_1^{s_1} L_2^{s_2} \cdots L_k^{s_k}) &= \frac{\Gamma(\alpha)}{\Gamma(\alpha+h)} \sum_{h_1} \cdots \sum_{h_k} \Bigg(\prod_{j=1}^{k} \binom{s_j}{h_{1j}, ..., h_{rj}} \\ &\quad \times \prod_{i=1}^{r} \frac{\Gamma(\alpha_i + h_i)}{\Gamma(\alpha_i)} \frac{\Gamma(n_i)}{\Gamma(n_i + h_i)} \prod_{i=1}^{r} \prod_{j=1}^{k} \frac{\Gamma(n_{ij} + h_{ij})}{\Gamma(n_{ij})} \Bigg), \end{aligned}$$

where L_j's are components of vector Z, $\sum_{i=1}^{r} h_i = h$ and $\sum_{i=1}^{r} \alpha_i = \alpha$.

Proof: We find the general moments $(s_1, s_2, ..., s_k)$ of Z as follows:

$$\begin{aligned} E(L_1^{s_1} L_2^{s_2} \cdots L_k^{s_k}) &= E\Bigg(\prod_{j=1}^{k} \Big(\sum_{i=1}^{r} W_i \mathbf{X}_{ij} \Big)^{s_j} \Bigg) \\ &= E\Bigg(\prod_{j=1}^{k} \Big(\sum_{h_j} \binom{s_j}{h_{1j}, h_{2j}, ..., h_{nj}} \times \prod_{i=1}^{r} (W_i X_{ij})^{h_{ij}} \Big) \Bigg), \end{aligned}$$

where the expression $\sum h_j$ denotes the summation over all the nonnegative integers $h_j = (h_{1j}, h_{2j}, ..., h_{rj})$ subject to

$$\sum_{i=1}^{r} h_{ij} = s_j, \quad (j = 1, 2, ..., k).$$

This can be rearranged as

$$\begin{aligned} &E\Bigg(\sum_{h_1} \sum_{h_2} \cdots \sum_{h_k} \Bigg(\prod_{j=1}^{k} \binom{s_j}{h_{1j}, h_{2j}, ..., h_{rj}} \prod_{j=1}^{k} \prod_{i=1}^{r} (W_i \mathbf{X}_{ij})^{h_{ij}} \Bigg) \Bigg) = \\ &\qquad E\Bigg(\sum_{h_1} \sum_{h_2} \cdots \sum_{h_k} \Bigg(\prod_{j=1}^{k} \binom{s_j}{h_{1j}, h_{2j}, ..., h_{rj}} \prod_{i=1}^{r} W_i^{h_{i.}} \prod_{j=1}^{k} \prod_{i=1}^{r} X_{ij}^{h_{ij}} \Bigg) \Bigg), \end{aligned}$$

where $h_{i.} = \sum_{j=1}^{k} h_{ij}$ and we have this equal to

$$\sum_{h_1} \cdots \sum_{h_k} \Bigg(\prod_{j=1}^{k} \binom{s_j}{h_{1j}, h_{2j}, ..., h_{rj}} E\Big(\prod_{i=1}^{r} W_i^{h_{i.}} \Big) E\Big(\prod_{j=1}^{k} \prod_{i=1}^{r} X_{ij}^{h_{ij}} \Big) \Bigg), \tag{2.1}$$

now we find two expectations in equation (2.1):

$$E\Big(\prod_{i=1}^{r} W_i^{h_{i.}} \Big) = \frac{\Gamma(\sum_{i=1}^{r} \alpha_i)}{\Gamma(\sum_{i=1}^{r} (\alpha_i + h_{i.}))} \times \prod_{i=1}^{r} \frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)}.$$

By using the Dirichlet distribution, we have

$$E\left(\prod_{i=1}^{r} W_i^{h_{i.}}\right) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+h)} \times \prod_{i=1}^{r} \frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)}. \tag{2.2}$$

Also, we have

$$E\left(\prod_{j=1}^{k}\prod_{i=1}^{r} X_{ij}^{h_{ij}}\right) = \prod_{i=1}^{r} E\left(\prod_{j=1}^{k} X_{ij}^{h_{ij}}\right) = \prod_{i=1}^{r}\left(\frac{\Gamma(\sum_{j=1}^{k} n_{ij})}{\Gamma(\sum_{j=1}^{k}(n_{ij}+h_{ij}))} \times \prod_{j=1}^{k} \frac{\Gamma(n_{ij}+h_{ij})}{\Gamma(n_{ij})}\right).$$

Now we have $\sum_{j=1}^{k} n_{ij} = n_{i.}$ and $\sum_{j=1}^{k} h_{ij} = h_{i.}$, so the above is equal to

$$\prod_{i=1}^{r}\left(\frac{\Gamma(n_{i.})}{\Gamma(n_{i.}+h_{i.})} \times \prod_{j=1}^{k} \frac{\Gamma(n_{ij}+h_{ij})}{\Gamma(n_{ij})}\right) \tag{2.3}$$

and by using the Dirichlet distribution we have

$$E\left(\prod_{j=1}^{k} X_{ij}^{h_{ij}}\right) = \frac{\Gamma(\sum_{j=1}^{k} \alpha_j^{(i)})}{\Gamma(\sum_{j=1}^{k} \alpha_j^{(i)} + h_{i.})} \prod_{j=1}^{k} \frac{\Gamma(\alpha_j^{(i)} + h_{ij})}{\Gamma(\alpha_j^{(i)})}.$$

So, by using (2.2) and (2.3) in (2.1) the above is equal to

$$\sum_{h_1}\cdots\sum_{h_k}\left(\prod_{j=1}^{k}\binom{s_j}{h_{1j}, h_{2j}, ..., h_{rj}} \frac{\Gamma(\alpha)}{\Gamma(\alpha+h)} \prod_{i=1}^{r} \frac{\Gamma(\alpha_i+h_{i.})}{\Gamma(\alpha_i)} \prod_{i=1}^{r} \frac{\Gamma(n_{i.})}{\Gamma(n_{i.}+h_{i.})} \prod_{j=1}^{k} \frac{\Gamma(n_{ij}+h_{ij})}{\Gamma(n_{i})}\right)$$
$$= \frac{\Gamma(\alpha)}{\Gamma(\alpha+h)} \sum_{h_1}\cdots\sum_{h_k} \prod_{j=1}^{k}\binom{s_j}{h_{1j}, ..., h_{rj}} \prod_{i=1}^{r} \frac{\Gamma(\alpha_i+h_{i.})}{\Gamma(\alpha_i)} \frac{\Gamma(n_{i.})}{\Gamma(n_{i.}+h_i)} \prod_{i=1}^{r}\prod_{j=1}^{k} \frac{\Gamma(n_{ij}+h_{ij})}{\Gamma(n_{ij})}.$$

Therefore the proof of the product moments on $(s_1, ..., s_k)$ is complete. □

The moments of Z on (s_1, s_2, s_3) are given in Table 1.

Table 1: The moment of Z.

(n_{11}, n_{12}, n_{13})	(n_{21}, n_{22}, n_{23})	(n_{31}, n_{32}, n_{33})	$(\alpha_1, \alpha_2, \alpha_3)$	(s_1, s_2, s_3)	$E(Z)$
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	0.001851852
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(1,1,2)	0.002469136
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(1,3,1)	0.003086420
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(1,2,3)	0.004938272
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(2,4,5)	0.025925926
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(2,2,2)	0.006172840
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(2,2,1)	0.003703704
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(2,2,3)	0.008641975
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(3,3,3)	0.017901235
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(3,3,1)	0.006790123
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(3,3,5)	0.029012346
(1,1,1)	(1,1,1)	(1,1,1)	(1,1,1)	(4,3,5)	0.038271605

3. SOME CHARACTERIZATIONS

In this section, some characterizations of random stochastic linear combinations (real lifetime, random convex combination) in Dirichlet random vectors are introduced.

Theorem 3.1. *Let the k-variate random vectors $\mathbf{X}_1, ..., \mathbf{X}_r$ be independent with common distributions, and let $Y_1, ..., Y_r$ be independent with* $\text{Gamma}(k\alpha, \frac{1}{\mu})$ *distributions, and independent from $\mathbf{X}_1, ..., \mathbf{X}_r$. Then the components of $\boldsymbol{T} = \sum_{i=1}^{r} Y_i \mathbf{X}_i$ have independent* $\text{Gamma}(r\alpha, \frac{1}{\mu})$ *distributions if and only if $\mathbf{X}_i$ have* $\text{Dirichlet}(\alpha, ..., \alpha)$ *distributions $(i = 1, ..., n)$.*

Proof: First we find the moment generating function of T

$$\begin{aligned}
E\big(e^{t'T}\big) &= E\big(e^{t' \sum_{i=1}^{r} Y_i \mathbf{X}_i}\big) \\
&= \prod_{i=1}^{r} E\big(e^{t' Y_i \mathbf{X}_i}\big) \\
&= \prod_{i=1}^{r} E\Big(E\big(e^{t' Y_i \mathbf{X}_i} \,|\, \mathbf{X}_i\big)\Big) \\
&= \prod_{i=1}^{r} E\left(\left(\frac{1}{1 - t'\mathbf{X}_i}\right)^{\alpha}\right) \\
&= E^r\left(\frac{1}{1 - t'\mathbf{X}_i}\right)^{\alpha}.
\end{aligned}$$

The second side of equation is well-known Stieltjes transformation; for more application and properties of this transformation see Homei (2012), Homei (2015), Homei (2016), Homei (2017b), and Cifarelli and Regazzini (1979).

The last statement is the Stieltjes transformation which is unique, by which both of the if part and the only if part can be easily proved. □

The following theorem is a generalization Theorem 1 of Yeo and Milne (1989) that leads us to the next theorems.

Theorem 3.2. *Suppose that $\mathbf{U}$ and V are independent (absolutely continuous) non-negative random variables, respectively, such that $\mathbf{U}$ has bounded support and $\mathbf{Z} = \mathbf{U}V$. Then for arbitrary positive α_i, $i = 1, ..., k$, and any two of the following three conditions imply the third:*

- **(i)** $\mathbf{Z} \sim \big\langle \text{Gamma}_1(\alpha_1, \frac{1}{\mu}), ..., \text{Gamma}_k(\alpha_k, \frac{1}{\mu})\big\rangle$ *where* $\text{Gamma}(\alpha_i, \frac{1}{\mu})$ *are independent;*
- **(ii)** $\mathbf{U} \sim \text{Dirichlet}(\alpha_1, ..., \alpha_k)$*;*
- **(iii)** $V \sim \text{Gamma}(\alpha^+, \frac{1}{\mu}), \quad \alpha^+ = \sum \alpha_i$.

Proof: For proving (i),(ii) ⇒ (iii), it suffices to note that the sum of the components of $\mathbf{Z}$ and $\mathbf{U}V$ are identically distributed; i.e., $V \overset{\text{d}}{=} \text{Gamma}(\alpha^+, \frac{1}{\mu})$, since random vectors $\mathbf{Z}$ and $\mathbf{U}V$ have identical distribution.

Now we prove the implication (i),(iii) ⇒ (ii). By using the general moments of $(s_{11}, ..., s_{kk})$ we have

$$E(Z_{11}^{s_{11}} Z_{12}^{s_{12}} \cdots Z_{kk}^{s_{kk}}) = E(V^{\sum_{i=1}^{k} s_i}) \cdot E\big((U_{11})^{s_{11}} (U_{12})^{s_{12}} \cdots (U_{kk})^{s_{kk}}\big).$$

By substituting the gamma moments it can be shown that $\mathbf{U}$ has Dirichlet distribution.

Finally, assume that (ii) and (iii) are satisfied, then we can obtain the distribution of $\mathbf{U}V$ by using the transformation method of random variables (or change of variables),

$$f(z_1, ..., z_r) = \frac{1}{\prod_{i=1}^{r} \Gamma(\alpha_i)\, \mu^{(\sum_{i=1}^{r} \alpha_i)}}\, e^{-\frac{\sum_{i=1}^{r} z_i}{\mu}} \prod_{i=1}^{r} z_i^{\alpha_i - 1}.$$

So, the proof is complete. □

Remark 3.1. Throughout this paper we set $\alpha^+ = \sum \alpha_i$.

Theorem 3.3. *Let* $\mathbf{X}$ *be any random vector with bounded support and* Y *be independent random variable of* $\mathbf{X}$ *with* Gamma$(\sum_{j=1}^{r} \alpha_j, \frac{1}{\mu})$ *distribution. If*

$$\sum_{i=1}^{r} Y_i \mathbf{X}_i \overset{\mathrm{d}}{=} Y\mathbf{X}, \tag{3.1}$$

where $Y_i (i = 1, ..., r)$ *are independent random variables with* Gamma$(\alpha_i, \frac{1}{\mu})$ *distribution, then* $\mathbf{X}$ *and the randomly linear combination* $\mathbf{Z} = \sum_{i=1}^{r} W_i \mathbf{X}_i$ *have identical distribution, where the random vector* $\mathbf{W} = \langle W_1, ..., W_r \rangle$ *is independent from* $\mathbf{X}_1, ..., \mathbf{X}_r$ *and has* Dirichlet$(\alpha_1, ..., \alpha_k)$ *distribution.*

Proof: First we define $Y^+ = \sum_{i=1}^{r} Y_i$, which has Gamma$(\alpha^+, \frac{1}{\mu})$ distribution, then by using (3.1) we have

$$Y^+ \cdot \frac{\sum_{i=1}^{r} Y_i \mathbf{X}_i}{Y^+} \overset{\mathrm{d}}{=} Y\mathbf{X},$$

the fraction $\frac{\sum_{i=1}^{r} Y_i \mathbf{X}_i}{Y^+}$ has the same distribution as $\mathbf{Z}$, so we can rewrite the above expression in the form of

$$Y^+ \mathbf{Z} \overset{\mathrm{d}}{=} Y\mathbf{X}. \tag{3.2}$$

The random vectors $Y^+\mathbf{Z}$ and $Y\mathbf{X}$ in (3.2) have the same moments, so

$$E\big((Y^+) {Z_1}^{k_1} \cdot (Y^+)^{k_2} Z_2^{k_2} \cdots (Y^+)^{k_r} Z_r^{k_r}\big) = E(Y^{k_1} X_1^{k_1} \cdot Y^{k_2} X_2^{k_2} \cdots Y^{k_r} X_r^{k_r}),$$

and we have

$$E\big((Y^+)^{k^+}\big) E(Z_1^{k_1} Z_2^{k_2} \cdots Z_r^{k_r}) = E\Big(Y^{k^+} \big(E(X_1^{k_1} \cdot X_2^{k_2} \cdots X_r^{k_r})\big)\Big),$$

where $k^+ = \sum_{j=1}^{r} \sum_{i=1}^{r} k_{ij}$.

Considering the same distribution of Y^+ and Y, we can omit the first expectations from both sides of the equation

$$E(Z_1^{k_1} Z_2^{k_2} \cdots Z_r^{k_r}) = E(X_1^{k_1} \cdot X_2^{k_2} \cdots X_r^{k_r})$$

as a result of having bounded support variables, the equation of the same moments of two variables conduces to the same distribution, so the proof is completed and $\mathbf{X}$ and $\mathbf{Z}$ have identical distributions. □

The following theorem is a generalization of Theorem 2.2 (Homei and Nadarajah, 2018) and we want to provide another perspective to prove Theorem 2.1 of Homei (2021).

Theorem 3.4. *If* $\mathbf{X}_1, ..., \mathbf{X}_r$ *are independent* k*-variate random vectors with respectively* $\text{Dirichlet}(n_1^{(1)}), ..., \text{Dirichlet}(n_k^{(r)})$ *distributions, for some* k*-dimensional vectors* $n_i^{(j)} = \langle n_1^{(j)}, ..., n_k^{(j)} \rangle$ $(j = 1, ..., r)$*, and the random vector* $\boldsymbol{W} = \langle W_1, ..., W_r \rangle$ *is independent from* $\mathbf{X}_1, ..., \mathbf{X}_r$ *and has the distribution*

$$\text{Dirichlet}\left(\sum_{i=1}^{k} n_i^{(1)}, ..., \sum_{i=1}^{k} n_i^{(r)}\right),$$

then the randomly linear combination $\boldsymbol{Z} = \sum_{i=1}^{r} W_i \mathbf{X}_i$ *has the distribution*

$$\text{Dirichlet}\left(\sum_{j=1}^{r} n_1^{(j)}, ..., \sum_{j=1}^{n} n_k^{(j)}\right).$$

Proof: Let Y_j $(j = 1, ..., r)$ be independent random variables and independent from $\langle \mathbf{X}_1, ..., \mathbf{X}_r \rangle$ that have the distribution $\text{Gamma}(\sum_{i=1}^{k} n_i^{(j)}, \frac{1}{\mu})$, respectively. It can be seen, by some classic ways (e.g. $E(e^{t'T}) = [\Psi(t)]^{(\sum_j n_j)}$ — see table 2 in Kerov and Tsilevich, 2004), that the distribution of $\mathbf{T} = \sum_j \mathbf{T}_j = \sum_j Y_j \mathbf{X}_j$ is the same as the distribution of $\mathbf{T}_j$ with the parameter $(\sum_j n_j, ..., \sum_j n_j)$. We can also write $\mathbf{T} \overset{\mathrm{d}}{=} Y\mathbf{X}$ in which Y has the Gamma distribution with the parameter $(\sum_j \sum_{i=1}^{k} n_i^{(j)}, \frac{1}{\mu})$, and Y and $\mathbf{X}$ are independent from each other. By using Theorem 3.3 the proof is complete. □

4. SIMULATION AND ITS APPLICATION

We present a new method for approximating the distribution that can be used for the multivariate random variables with a slight change in the approximation of Homei and Nadarajah (2018). The previous method of approximation was for a univariate random variable, but we want to introduce a method that is able to approximate the multivariate random variables. Calculating the distribution of $\mathbf{Z}$ may not be easy; in this section we suggest a distribution which may be close to the real distribution of $\mathbf{Z}$. See Homei and Nadarajah (2018) for more details.

Let $\mathbf{X}_1$ and $\mathbf{X_2}$ be independent random vectors with $\text{Dirichlet}(1, 1, 1)$ distribution, and let w be a random variable independent from them with $\text{Beta}(1, 1)$ distribution. Calculating the distribution of

$$\mathbf{Z} = w\mathbf{X}_1 + (1 - w)\mathbf{X_2}$$

could be cumbersome, and it could be that there is no closed form. For estimating the distribution of $\mathbf{Z}$ we suggest the following. Firstly, we generate $\mathbf{X}_1$ and $\mathbf{X_2}$ data by the means of Python software package. Then we compose them in accordance with the definition of $\mathbf{Z}$. We simulate random numbers with size 8000 of $\mathbf{Z}$, and assume that it has the Dirichlet distribution; we estimate their parameters by the method of maximum likelihood estimation.

As a result, the values of $\mathbf{Z}$ will be observable. Our suggested approximation has the Dirichlet$(2.7, 2.8, 3)$ distribution and we expect it to be useful as mentioned in Homei and Nadarajah (2018). Figure 1 shows the approximated distribution of $\mathbf{Z}$, i.e., Dirichlet$(2.7, 2.8, 3)$.

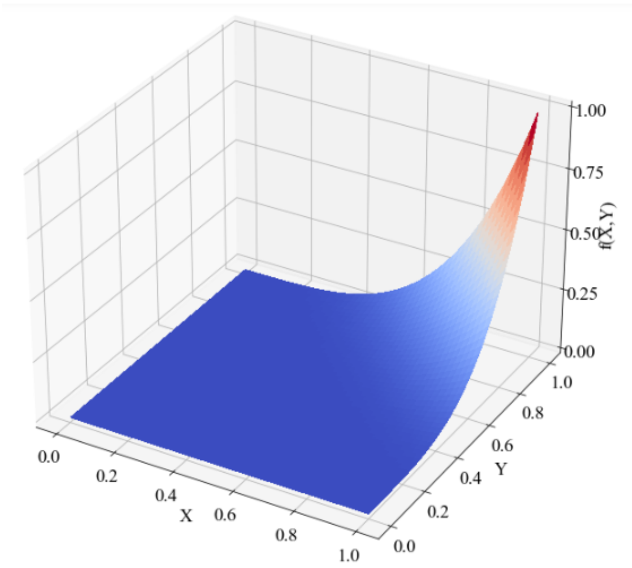

Figure 1: The approximate distribution is Dirichlet$(2.7, 2.8, 3)$.

5. CONCLUSIONS

In this paper, a novel method for obtaining the distribution of the randomly weighted averages on random vectors is presented, which is simpler and more elementary than the others. Beside that one can obtain the distribution of $T = \sum \mathbf{X}_i Y_i$ by that method, which is left to be done in the future. In case this distribution appears to be complicated, we will approximate it by simulation, and we will study some distributional properties of T in general. The four examples illustrated in the Introduction (Cauchy Composition Test, Real Lifetime, Solving Some Differential Equations, and Random Convex Combination) are some visible applications of the research in this paper.

ACKNOWLEDGMENTS

We would like to thank the Referee#2 of the first round for the most useful comments and suggestions, and we also thank the editors and Referee#A of the third round for carefully reading the manuscript along with our responses to the referee-reports in the previous rounds. We are wholeheartedly grateful to Professor J. Pitman for justly evaluating our works, which highly encouraged us to continue this research.

REFERENCES

Cifarelli, D. M. and Regazzini, E. (1979). A general approach to Bayesian analysis of nonparametric problems. The associative mean values within the framework of the Dirichlet process: II (in Italian). *Rivista di Matematica per le Scienze Economiche e Sociali*, 2:95–111.

Hadad, H., Homei, H., Behzadi, M., and Farnoosh, R. (2021). Solving some stochastic differential equation using Dirichlet distributions. *Computational Methods for Differential Equations*, 9(2): 393–398.

Homei, H. (2012). Randomly weighted averages with beta random proportions. *Statistics and Probability Letters*, 82(8):1515–1520.

Homei, H. (2015). A novel extension of randomly weighted averages. *Statistical Papers*, 56(4):933–946.

Homei, H. (2016). Randomly weighted averages: a multi-variate case. arXiv:1603.00596 [math.ST]. `https://arxiv.org/abs/1603.00596`

Homei, H. (2017a). Characterizations of arcsin and related distributions based on a new generalized unimodality. *Communications in Statistics — Theory and Methods*, 46(2):1024–1030.

Homei, H. (2017b). Randomly weighted averages: a multi-variate case. In: *Abstracts of the 11th Seminar on Probability and Stochastic Processes*, Qazvin, Iran (30–31 August 2017), Imam Khomeini International University, page 49.

Homei, H. (2021). The stochastic linear combination of Dirichlet distributions. *Communications in Statistics – Theory and Methods*, 50(10):2354–2359.

Homei, H. and Nadarajah, S. (2018). On products and mixed sums of gamma and beta random variables motivated by availability. *Methodology and Computing in Applied Probability*, 20(2): 799–810.

Kerov, S. V. E. and Tsilevich, N. V. (2004). The Markov-Krein correspondence in several dimensions. *Journal of Mathematical Sciences*, 121(3):2345–2359.

Liu, Y. and Xie, J. (2020). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structure. *Journal of the American Statistical Association*, 115(529):393–402.

Pitman, J. (2018). Random weighted averages, partition structures and generalized arcsine laws. arXiv 1804.07896, to appear in *Probability Surveys*.

Soltani, A. R. and Homei, H. (2009). Weighted averages with random proportions that are jointly uniformly distributed over the unit simplex. *Statistics and Probability Letters*, 79(9):1215–1218.

Soltani, A. R. and Roozegar, R. (2020). Averages for multivariate random vectors with random weights. *REVSTAT – Statistical Journal*, 18(4):453–460.

Touri, B. (2012). *Product of Random Stochastic Matrices and Distributed Averaging*. Springer-Verlag, Berlin.

Yeo, G. F. and Milne, R. K. (1989). On characterizations of exponential distributions. *Statistics and Probability Letters*, 7(4):303–305.

REVSTAT – Statistical Journal
Volume 23, Number 2, April 2025, 235–251
https://doi.org/10.57805/revstat.v23i2.513

Rotatable Response Surface Designs for $s_1^{n_1} \times s_2^{n_2}$ Incorporating Neighbour Effects

Authors: Ashutosh Dalal
– ICAR-Indian Agricultural Statistics Research Institute,
New Delhi, India
ashutosh.dalal97@gmail.com

Seema Jaggi
– Indian Council of Agricultural Research,
New Delhi, India
seema.jaggi@icar.gov.in

Eldho Varghese ✉
– ICAR-Central Marine Fisheries Research Institute,
Kochi, Kerala, India
eldhoiasri@gmail.com ; eldho.varghese@icar.gov.in

Arpan Bhowmik
– ICAR-Indian Agricultural Research Institute,
Gogamukh, Assam
arpan.stat@gmail.com ; arpan.bhowmik@icar.gov.in

Cini Varghese
– ICAR-Indian Agricultural Statistics Research Institute,
New Delhi, India
cini.varghese@icar.gov.in

Anindita Datta
– ICAR-Indian Agricultural Statistics Research Institute,
New Delhi, India
anindita.datta@icar.gov.in

Received: August 2022 Revised: June 2023 Accepted: June 2023

Abstract:

- This paper describes the response surface methodology for mixed-level factors of the form $s_1^{n_1} \times s_2^{n_2}$ when experimental units experiences overlap effects from the adjacent neighbouring units. Conditions have been derived for the near orthogonal estimation of the parameters of the model and ensuring the constancy in the prediction variance. A method of construction of designs satisfying derived conditions has been developed. Some particular cases of $s_1^{n_1} \times s_2^{n_2}$ has also described. An R package named rsdNE has also been developed for the generation of these designs.

Keywords:

- *mixed-level response surface design; neighbour effect; rotatable design.*

AMS Subject Classification:

- 49A05, 78B26.

✉ Corresponding author

1. INTRODUCTION

The response surface methodology (RSM) is a widely used statistical method for modelling and analyzing a process in which the response of interest is affected by various input variables, and the objective of this method is to optimize the response (Box and Wilson, 1951). The main idea of RSM is to use a sequence of designed experiments to obtain an optimal response. If $x_1, x_2, ..., x_v$ are v independent variables, y is the response variable, and N is the total number of observations, then the response function can be approximated in some region of the polynomial model given by

$$y_u = f(x_{1u}, x_{2u}, ..., x_{vu}) + e_u,$$

where $u = 1, 2, ..., N$, y_u is the response from u-th treatment combination and x_{iu} is the level of the i-th ($i = 1, 2, ..., v$) factor in the u-th combination. The function $f(\cdot)$ describes the form in which the response and the input variables are related. e_u is the random error associated with the u-th observation that is independently and normally distributed as $e_u \sim N(0, \sigma^2)$. For details on RSM, one can refer to Box and Draper (1987), Khuri and Cornell (1996), Myers (1995), Myers *et al.* (2009).

In RSM, it is mainly assumed that the observations are independent and that neighbouring units have no effect. But in agricultural field trials, the neighbouring effect or overlap effect is very prominent (Bartlett, 1938, 1978; Dalal *et al.*, 2022; Draper and Guttman, 1980; Papadakis, 1937). For example, if a chemical treatment is sprayed on one plot, wind drift may allow the spray to spread to adjacent/neighbouring plots, or preceding soil preparation may allow sterilized soil from one plot to mingle with non-sterile soil from the next plot. As a result, it is vital to assume that the response received from a given plot is influenced not only by the treatment combinations used on that plot but also by the treatment combinations used on the plots next to it. So, it gives a great scope to take into account the neighbouring effects in RSM. Because of these neighbouring effects, the variation of treatment differences arises. If the neighbour effect is present and is included in the model, there is a considerable reduction in the residual sum of squares, and the response is predicted with more precision (Jaggi *et al.*, 2010). Over the years, work on different aspects of RSM with neighbour effects from immediate adjacent neighbours has been done for factors with same levels (Sarika *et al.*, 2009, 2013; Varghese *et al.*, 2013, 2016, 2019, 2020). The response surface model with neighbour effects up to distance 2 is also available in the literature (Kumar *et al.*, 2020). Some attempts were also made to develop asymmetrical response surface designs of the form $2^n \times 3$ and $2^n \times 3^n$ in the presence of neighbour effects (Verma *et al.*, 2021).

Here, a general methodology for constructing response surface designs of the form $s_1^{n_1} \times s_2^{n_2}$ has been developed, incorporating neighbour effects and an `R` package named `rsdNE` has also been developed for the generation of these designs.

2. RESPONSE SURFACE MODEL WITH NEIGHBOUR EFFECTS

Consider a response surface with n_1 factors at s_1 levels each and n_2 factors at s_2 levels each resulting in $s_1^{n_1} \times s_2^{n_2}$ combinations. The form $f(x_u)$ of considered here is as follows:

$$
\begin{aligned}
f(x_u) \;=\; & \beta_0 + \sum_{i=1}^{n_1} \beta_i x_i + \sum_{i=n_1+1}^{n_1+n_2} \beta_i x_i + \sum_{i=1}^{n_1} \beta_{ii} {x_i}^2 + \sum_{i=n_1+1}^{n_1+n_2} \beta_{ii} {x_i}^2 + \cdots + \\
& + \sum_{i=1}^{n_1} \beta_{ii\ldots i} {x_i}^{s_1-1} + \sum_{i=n_1+1}^{n_1+n_2} \beta_{ii\ldots i} {x_i}^{s_2-1}, \qquad (2.1)
\end{aligned}
$$

where β_0, β_i's [associated with linear terms of n_1+n_2 factors], β_{ii} [associated with quadratic term of n_1+n_2 factors] and $\beta_{i\ldots i}$ [associated with (s_i-1)-th order term of n_i-th factor] are parameters to be estimated. Thus, the total number of parameters to be estimated in above model is $p = n_1(s_1-1) + n_2(s_2-1) + 1$. The response model incorporating the effects from immediate left and right neighbouring units is defined as follows:

$$
y_{u'} = \sum_{u=1}^{N} g_{uu'} f(x_u) + e_{u'}, \qquad u' = 1, 2, \ldots, N, \qquad (2.2)
$$

where

$$
\begin{aligned}
g_{uu'} &= 1, && \text{if } u = u', \\
&= \alpha, \quad |\alpha| < 1, && \text{if } |u - u'| = 1, \text{ i.e., units are physically adjacent}, \\
&= 0, && \text{otherwise}.
\end{aligned}
$$

Here, α represents the neighbour effect from left and right neighbouring units and ranges from 0 to 1 (Draper and Guttman, 1980; Sarika *et al.*, 2009). $f(x_u)$ is as given in (2.1).

It is to be mentioned that the design layout of the experiment for estimating this model will consist of two extra units as border units at each end. Observations are not taken from border units and thus are not modelled.

Model (2.2) can be rewritten as

$$
\mathbf{Y} = \mathbf{GXY} + \mathbf{e}, \qquad (2.3)
$$

where $\mathbf{G} = g_{uu'}$ is the $N \times (N+2)$ symmetric neighbour matrix, $\mathbf{X}$ is a $(N+2) \times p$ matrix of N points (runs) with two extra border units (treatment combinations applied on these units are the treatment combinations from the N points), p is the number of parameters, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters to be estimated and $\mathbf{e}$ is $N \times 1$ vector of errors which follows $N(0, \sigma^2\mathbf{I})$.

The ordinary least squares estimate of $\boldsymbol{\beta}$, in the presence of neighbour effects with known $\mathbf{G}$ is

$$
\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y},
$$

where $\mathbf{Z} = \mathbf{GX}$ and $D(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}$.

2.1. Response Surface Methodology for $3^2 \times 4^2$ with Neighbour Effects

Consider $n_1 = n_2 = 2$, $s_1 = 3$, $s_2 = 4$, i.e., 2 factors at levels three and 2 factors at levels four. The form of $f(x_u)$ is as follows:

$$\begin{aligned} f(x_u) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ &\quad + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{44} x_4^2 + \beta_{333} x_3^3 + \beta_{444} x_4^3 \end{aligned} \tag{2.4}$$

The matrix $\mathbf{X}$ in (2.3) is here as follows of order $(N+2) \times 11$ with two border units:

$$\mathbf{X}_{(N+2)\times 11} = \left[\begin{array}{ccccccc} 1 & x_{1N} & x_{2N} & \cdots & x_{4N}^2 & x_{3N}^3 & x_{4N}^3 \\ \hline \mathbf{1} & \mathbf{X_1} & \mathbf{X_2} & \cdots & \mathbf{X}_4^2 & \mathbf{X}_3^3 & \mathbf{X}_4^3 \\ \hline 1 & x_{11} & x_{21} & \cdots & x_{41}^2 & x_{31}^3 & x_{41}^3 \end{array}\right],$$

where $\mathbf{X}_i^b = \begin{bmatrix} x_{i1}^b & x_{i2}^b & \cdots & x_{iN}^b \end{bmatrix}'$, $\mathbf{X}_i^1 = \mathbf{X}_i$, $i = 1, 2, 3, 4$ and $b = 1, 2, 3$; and $\mathbf{1}_N = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}'$. The structure of $\mathbf{G}$ matrix is

$$\mathbf{G}_{N\times(N+2)} = \begin{bmatrix} \alpha & 1 & \alpha & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \alpha & 1 & \alpha & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \alpha & 1 & \alpha & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \alpha & 1 & \alpha & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \alpha & 1 & \alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & \alpha & 1 & \alpha \end{bmatrix}.$$

Further, $\mathbf{Z} = \mathbf{GX} =$

$$\begin{bmatrix} \theta' & x_{11} + \alpha(x_{1N} + x_{12}) & \cdots & x_{31}^2 + \alpha(x_{3N}^2 + x_{32}^2) & \cdots & x_{41}^3 + \alpha(x_{4N}^3 + x_{42}^3) \\ \theta' & x_{12} + \alpha(x_{11} + x_{13}) & \cdots & x_{32}^2 + \alpha(x_{31}^2 + x_{33}^2) & \cdots & x_{42}^3 + \alpha(x_{41}^3 + x_{43}^3) \\ \theta' & x_{13} + \alpha(x_{12} + x_{14}) & \cdots & x_{33}^2 + \alpha(x_{32}^2 + x_{34}^2) & \cdots & x_{43}^3 + \alpha(x_{42}^3 + x_{44}^3) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta' & x_{1(N-1)} + \alpha(x_{1N} + x_{1(N-2)}) & \cdots & x_{3(N-1)}^2 + \alpha(x_{3N}^2 + x_{3(N-2)}^2) & \cdots & x_{4(N-1)}^3 + \alpha(x_{4N}^3 + x_{4(N-2)}^3) \\ \theta' & x_{1N} + \alpha(x_{11} + x_{1(N-1)}) & \cdots & x_{3N}^2 + \alpha(x_{31}^2 + x_{3(N-1)}^2) & \cdots & x_{4N}^3 + \alpha(x_{41}^3 + x_{4(N-1)}^3) \end{bmatrix},$$

where $\theta' = 1 + 2\alpha$.

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} N\theta_1 & \theta' \sum_{u=1}^{N} a_{1u} & \theta' \sum_{u=1}^{N} a_{2u} & \cdots & \theta' \sum_{u=1}^{N} c_{4u} \\ & \sum_{u=1}^{N} a^2{}_{1u} & \sum_{u=1}^{N} a_{1u}a_{2u} & \cdots & \sum_{u=1}^{N} a_{1u}c_{4u} \\ & & \sum_{u=1}^{N} a^2{}_{2u} & \cdots & \sum_{u=1}^{N} a_{2u}c_{4u} \\ & & & \cdots & \cdots \\ & & & & \sum_{u=1}^{N} c^2{}_{4u} \end{bmatrix},$$

where $\theta_1 = (1+2\alpha)^2$, $a_{iu} = x_{iu} + \alpha[x_{i(u-1)\bmod(N)} + x_{i(u+1)\bmod(N)}]$, $b_{iu} = x^2{}_{iu} + \alpha[x^2{}_{i(u-1)\bmod(N)} + x^2{}_{i(u+1)\bmod(N)}]$, $c_{iu} = x^3{}_{iu} + \alpha[x^3{}_{i(u-1)\bmod(N)} + x^3{}_{i(u+1)\bmod(N)}]$.

The following conditions are required for near orthogonal estimation of parameters:

(**a**) $\displaystyle\sum_{u=1}^{N} \prod_{i=1}^{4} x_{iu}^{w_i} = 0$ for $w_i = 0, 1, 2, 3$ or 4 and $\sum w_i < 6$;

(**b**) $\displaystyle\sum_{u=1}^{N} x_{iu}^2 = \delta_i$;

(**c**) $\displaystyle\sum_{u=1}^{N} x_{iu}^4 = \gamma_i$, where $i = 1, 2, 3$ and 4;

(**d**) $\displaystyle\sum_{u=1}^{N} x_{iu}^2 x_{ju}^2 = \lambda_1$ and $\displaystyle\sum_{u=1}^{N} x_{iu}^3 x_{ju}^3 = \lambda_2$, where $i \neq j$;

(**e**) $\displaystyle\sum_{u=1}^{N} x_{lu}^6 = \beta_1$, where $l = 3, 4$.

Therefore, $\mathbf{Z'Z}$ matrix can now be written as

$$\begin{bmatrix}
N\theta_1 & 0 & 0 & 0 & 0 & \theta_1 \sum_{u=1}^{N} b_{1u} & \theta_1 \sum_{u=1}^{N} b_{2u} & \theta_1 \sum_{u=1}^{N} b_{3u} & \theta_1 \sum_{u=1}^{N} b_{4u} & 0 & 0 \\
 & \sum_{u=1}^{N} a^2{}_{1u} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & \sum_{u=1}^{N} a^2{}_{2u} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & \sum_{u=1}^{N} a^2{}_{3u} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & \sum_{u=1}^{N} a^2{}_{4u} & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & \sum_{u=1}^{N} b^2{}_{1u} & \sum_{u=1}^{N} b_{1u}b_{2u} & \sum_{u=1}^{N} b_{1u}b_{3u} & \sum_{u=1}^{N} b_{1u}b_{4u} & 0 & 0 \\
 & & & & & & \sum_{u=1}^{N} b^2{}_{2u} & \sum_{u=1}^{N} b_{2u}b_{3u} & \sum_{u=1}^{N} b_{2u}b_{4u} & 0 & 0 \\
 & & & & & & & \sum_{u=1}^{N} b^2{}_{3u} & \sum_{u=1}^{N} b_{3u}b_{4u} & 0 & 0 \\
 & & & & & & & & \sum_{u=1}^{N} b^2{}_{4u} & 0 & 0 \\
 & & & & & & & & & \sum_{u=1}^{N} c^2{}_{3u} & \sum_{u=1}^{N} c_{3u}c_{4u} \\
 & & & & & & & & & & \sum_{u=1}^{N} c^2{}_{4u}
\end{bmatrix}.$$

Let's consider a matrix $\mathbf{E}$ given by

$$\begin{bmatrix}
E_{11} & \sum_{u=1}^{N} a^2{}_{1u}a^2{}_{2u} & \sum_{u=1}^{N} a^2{}_{1u}a^2{}_{3u} & \sum_{u=1}^{N} a^2{}_{1u}a^2{}_{4u} & 0 & 0 \\
 & \sum_{u=1}^{N} a^4{}_{2u} & \sum_{u=1}^{N} a^2{}_{2u}a^2{}_{3u} & \sum_{u=1}^{N} a^2{}_{2u}a^2{}_{4u} & 0 & 0 \\
 & & \sum_{u=1}^{N} a^4{}_{3u} & \sum_{u=1}^{N} a^2{}_{3u}a^2{}_{4u} & 0 & 0 \\
 & & & \sum_{u=1}^{N} a^4{}_{4u} & 0 & 0 \\
 & & & & \sum_{u=1}^{N} a^6{}_{3u} & \sum_{u=1}^{N} a^3{}_{3u}a^3{}_{4u} \\
 & & & & & \sum_{u=1}^{N} a^6{}_{4u}
\end{bmatrix},$$

E_{ij}'s are the element position of i^{th} row and j^{th} column of E matrix.

Here, $E_{11} = \sum_{u=1}^{N} a^4{}_{1u} - \frac{\theta_1}{N}\left[\sum_{u=1}^{N} (a^2{}_{1u})^2 + \sum_{u=1}^{N} (a^2{}_{2u})^2 + \sum_{u=1}^{N} (a^2{}_{3u})^2 + \sum_{u=1}^{N} (a^2{}_{4u})^2\right]$.

Now, $(\mathbf{Z'Z})^{-1}$matrix can be written as follows:

$$\begin{bmatrix} \eta & 0 & 0 & 0 & 0 & \frac{\sum_{j=1}^{4} C_{j1}L_{1j}}{-\Delta K_{11}} & \frac{\sum_{j=1}^{4} C_{j2}L_{1j}}{-\Delta K_{11}} & \frac{\sum_{j=1}^{4} C_{j3}L_{1j}}{-\Delta K_{11}} & \frac{\sum_{j=1}^{4} C_{j4}L_{1j}}{-\Delta K_{11}} & 0 & 0 \\ & \frac{1}{\sum_{u=1}^{N} a_{1u}^2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & \frac{1}{\sum_{u=1}^{N} a_{2u}^2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & \frac{1}{\sum_{u=1}^{N} a_{3u}^2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & \frac{1}{\sum_{u=1}^{N} a_{4u}^2} & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & E^{11} & E^{12} & E^{13} & E^{14} & E^{15} & E^{16} \\ & & & & & & E^{22} & E^{23} & E^{24} & E^{25} & E^{26} \\ & & & & & & & E^{33} & E^{34} & E^{35} & E^{36} \\ & & & & & & & & E^{44} & E^{45} & E^{46} \\ & & & & & & & & & E^{55} & E^{56} \\ & & & & & & & & & & E^{66} \end{bmatrix},$$

where $\eta = \frac{L_{11}\sum_{j=1}^{4} C_{j1}L_{1j}}{-\Delta N\theta_1^2} + \frac{L_{12}\sum_{j=1}^{4} C_{j2}L_{1j}}{-\Delta N\theta_1^2} + \frac{L_{13}\sum_{j=1}^{4} C_{j3}L_{1j}}{-\Delta N\theta_1^2} + \frac{L_{14}\sum_{j=1}^{4} C_{j4}L_{1j}}{-\Delta N\theta_1^2} + \frac{1}{K_{11}}$, Δ is determinant of $\mathbf{E}$ matrix, E^{ij}'s are the elements of $\mathbf{E}^{-}$ matrix, C_{ij}'s are co-factors of $\mathbf{E}$ and $L_{1j} = \theta_1 \sum_{u=1}^{N} a^2{}_{ju}$ ($j = 1, 2, 3$ and 4).

The expressions of the variance and covariance terms of the estimates are given below:

$$\text{Var}(\hat{\beta}_0) = \eta\sigma^2, \qquad \text{Var}(\hat{\beta}_{11}) = E^{11}\sigma^2,$$

$$\text{Var}(\hat{\beta}_1) = \frac{1}{\sum_{u=1}^{N} a_{1u}^2}\sigma^2, \qquad \text{Var}(\hat{\beta}_{22}) = E^{22}\sigma^2,$$

$$\text{Var}(\hat{\beta}_2) = \frac{1}{\sum_{u=1}^{N} a_{2u}^2}\sigma^2, \qquad \text{Var}(\hat{\beta}_{33}) = E^{33}\sigma^2,$$

$$\text{Var}(\hat{\beta}_3) = \frac{1}{\sum_{u=1}^{N} a_{3u}^2}\sigma^2, \qquad \text{Var}(\hat{\beta}_{44}) = E^{44}\sigma^2,$$

$$\text{Var}(\hat{\beta}_4) = \frac{1}{\sum_{u=1}^{N} a_{4u}^2}\sigma^2, \qquad \text{Var}(\hat{\beta}_{333}) = E^{55}\sigma^2,$$

$$\text{Var}(\hat{\beta}_{444}) = E^{66}\sigma^2,$$

$$\text{Cov}(\hat\beta_0, \hat\beta_{11}) = \frac{\sum_{j=1}^{4} C_{j1}L_{1j}}{-\Delta K_{11}}\sigma^2, \qquad \text{Cov}(\hat\beta_{22}, \hat\beta_{33}) = E^{23}\sigma^2,$$

$$\text{Cov}(\hat\beta_0, \hat\beta_{22}) = \frac{\sum_{j=1}^{4} C_{j2}L_{1j}}{-\Delta K_{11}}\sigma^2, \qquad \text{Cov}(\hat\beta_{22}, \hat\beta_{333}) = E^{25}\sigma^2,$$

$$\text{Cov}(\hat\beta_0, \hat\beta_{33}) = \frac{\sum_{j=1}^{4} C_{j3}L_{1j}}{-\Delta K_{11}}\sigma^2, \qquad \text{Cov}(\hat\beta_{33}, \hat\beta_{44}) = E^{34}\sigma^2,$$

$$\text{Cov}(\hat\beta_0, \hat\beta_{44}) = \frac{\sum_{j=1}^{4} C_{j4}L_{1j}}{-\Delta K_{11}}\sigma^2, \qquad \text{Cov}(\hat\beta_{33}, \hat\beta_{333}) = E^{35}\sigma^2,$$

$$\text{Cov}(\hat\beta_{44}, \hat\beta_{333}) = E^{45}\sigma^2,$$

$$\text{Cov}(\hat\beta_{44}, \hat\beta_{444}) = E^{46}\sigma^2,$$

$$\text{Cov}(\hat\beta_{22}, \hat\beta_{44}) = E^{24}\sigma^2.$$

The estimated response at any point $\mathbf{x}_0$ is given below:

$$\hat y_0 = \hat\beta_0 + \sum_{i=1}^{4} \hat\beta_i x_{i0} + \sum_{i=1}^{4} \hat\beta_{ii} x_{i0}^2 + \sum_{i=3}^{4} \hat\beta_{iii} x_{i0}^3,$$

with the variance of estimated response as

$$V(\hat y_0) = \sigma^2 \begin{bmatrix} V_0 + x_{10}^2 V_1 + x_{20}^2 V_2 + x_{30}^2 V_3 + x_{40}^2 V_4 + V_{11}x_{10}^4 + V_{22}x_{20}^4 + V_{33}x_{30}^4 + V_{44}x_{40}^4 + V_{333}x_{30}^6 \\ +V_{444}x_{40}^6 + 2C_{0,11}x_{10}^2 + 2C_{0,22}x_{20}^2 + 2C_{0,33}x_{30}^2 + 2C_{0,44}x_{40}^2 + 2C_{22,33}x_{20}^2x_{30}^2 + 2C_{22,44}x_{20}^2x_{40}^2 \\ +2C_{22,333}x_{20}^2x_{30}^6 + 2C_{33,44}x_{30}^4x_{40}^4 + 2C_{33,333}x_{30}^{10} + 2C_{44,333}x_{40}^4x_{30}^6 + 2C_{44,444}x_{40}^{10} \end{bmatrix},$$

where $V_i = V(\hat\beta_i)$, $V_{ij} = V(\hat\beta_{ij})$ and so on. Similarly, $C_{i,j} = \text{Cov}(\hat\beta_i, \hat\beta_i)$, $C_{i,jk} = \text{Cov}(\hat\beta_i, \hat\beta_{jk})$ and so on.

It can be seen that the variance of the estimated response at any point $\mathbf{x}_0$ depends on the distance of that point from the design centre. The design obtained for fitting this model with mixed levels of factors will be called here as Mixed Level Response Surface Design with Neighbour Effects. From this one can check, at any point $\mathbf{x}_0$, the variance of the estimated response from the centre. If this is equal, then the design obtained is called as Mixed Level Rotatable Design with Neighbour Effects (MLRDNE) and if different, then partially rotatable. Similarly, this can be extended for $s_1^{n_1} \times s_2^{n_2}$.

The above discussions lead to the following theorem:

Theorem 2.1. *A $s_1^{n_1} \times s_2^{n_2}$ mixed-level factorial arranged in reverse lexicographic order along with the circular borders at each end will lead to MLRDNE under the model defined in (2.2) with $f(x_u)$ as defined in (2.1), provided either s_1 or s_2 should be an odd number.*

The proof of Theorem 2.1 has been given in the subsequent section with the help of examples.

3. METHOD OF CONSTRUCTING MLRDNE FOR $s_1^{n_1} \times s_2^{n_2}$

Consider n_1 factors having s_1 levels and n_2 factors having s_2 levels. The $n_1 + n_2$ columns of $\mathbf{X}$ corresponding to $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_{n_1}, \mathbf{X}_{n_1+1}, ..., \mathbf{X}_{n_1+n_2}$ of MLRDNE in $N = \max(n_1, n_2) \times s_1^{n_1} \times s_2^{n_2}$ points are developed as follows:

$$\mathbf{X} = \left[\mathbf{O}_{N\times n_1} \;\; \mathbf{Q}_{N\times n_2}\right],$$

$$\mathbf{O} = \begin{bmatrix} \mathbf{o}_1 & \mathbf{o}_2 & \cdots & \mathbf{o}_{n_1} \\ \mathbf{o}_{n_1} & \mathbf{o}_1 & \cdots & \mathbf{o}_{n_1-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{o}_2 & \mathbf{o}_3 & \cdots & \mathbf{o}_1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_{n_2} \\ \mathbf{q}_{n_2} & \mathbf{q}_1 & \cdots & \mathbf{q}_{n_2-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_1 \end{bmatrix}.$$

$\mathbf{o}_i$ and $\mathbf{q}_j$ are the vectors of order $N' \times 1$ (where, $N' = s_1^{n_1} \times s_2^{n_2}$):

$$\begin{aligned} \mathbf{o}_i &= \mathbf{1}_{s_1^p \times s_2^{n_2}} \otimes (\text{column vector of } s_1 \text{ levels}) \otimes \mathbf{1}_{s_1^{n_1-i}}, \quad p = i-1, \quad i = 1, 2, ..., n_1(>1), \\ \mathbf{q}_j &= \mathbf{1}_{s_1^{n_1} \times s_2^q} \otimes (\text{column vector of } s_2 \text{ levels}) \otimes \mathbf{1}_{s_2^{n_2-j}}, \quad q = j-1, \quad j = 1, 2, ..., n_2(>1). \end{aligned}$$

If $n_1 = 1$, then $\mathbf{O} = [\mathbf{o}_1 \quad \mathbf{o}_1]'$ and if $n_2 = 1$, then $\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_1]'$.

The other columns of $\mathbf{X}$ are generated as per the model and values of s_1 and s_2.

D-efficiency of the design can be calculated using the formula

$$\frac{|\mathbf{Z}'\mathbf{Z}|^{\frac{1}{p}}}{N},$$

where p is number of parameters considered in a model and N is total number of runs (Verma *et al.*, 2021).

Example 3.1. Let $n_1 = n_2 = 2$, $s_1 = 2$ (levels: $1, -1$) and $s_2 = 5$ (levels: $2, 1, 0, -1, -2$), i.e., $2^2 \times 5^2$. The first four columns of $\mathbf{X}$ corresponding to $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{X}_4$ and $2(2^2 \times 5^2) = 200$ points are as follows:

$$\mathbf{X} = \left[\mathbf{O}_{200\times 2} \;\; \mathbf{Q}_{200\times 2}\right].$$

The other columns of $\mathbf{X}$ are generated as per model with 11 columns ($\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_3^2, \mathbf{X}_4^2, \mathbf{X}_3^3, \mathbf{X}_4^3, \mathbf{X}_3^4, \mathbf{X}_4^4$). For $\alpha = 0.5$,

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 800 & \mathbf{0}_2' & \mathbf{0}_2' & 1600\mathbf{1}_2' & \mathbf{0}_2' & 5440\mathbf{1}_2' \\ & 100\mathbf{I}_2 & \mathbf{O}_2 & \mathbf{O}_2 & \mathbf{O}_2 & \mathbf{O}_2 \\ & & 930\mathbf{I}_2 & \mathbf{O}_2 & 2910\mathbf{I}_2 & \mathbf{O}_2 \\ & & & \mathbf{A} & \mathbf{O}_2 & \mathbf{B} \\ & & & & 10074\mathbf{I}_2 & \mathbf{O}_2 \\ & & & & & \mathbf{C} \end{bmatrix},$$

where $\mathbf{A} = 1470\mathbf{I}_2 + 3200\mathbf{J}_2$, $\mathbf{B} = 6438\mathbf{I}_2 + 10880\mathbf{J}_2$, $\mathbf{C} = 28686\mathbf{I}_2 + 36992\mathbf{J}_2$; $\mathbf{I}_2$ is identity matrix order 2×2, $\mathbf{J}_2$ is matrix of 1's of order 2×2, is column vector of 1's of order 2×1, is a column vector of 0's of order 2×1 and is matrix of 0's of order 2×2.

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 0.02 & \mathbf{0}_2' & \mathbf{0}_2' & -0.02\mathbf{1}_2' & \mathbf{0}_2' & 0.004\mathbf{1}_2' \\ & 0.01\mathbf{I}_2 & \mathbf{O}_2 & \mathbf{O}_2 & \mathbf{O}_2 & \mathbf{O}_2 \\ & & 0.011\mathbf{I}_2 & \mathbf{O}_2 & -0.003\mathbf{I}_2 & \mathbf{O}_2 \\ & & & 0.004\mathbf{I}_2 & \mathbf{O}_2 & -0.009\mathbf{I}_2 \\ & & & & 0.001\mathbf{I}_2 & \mathbf{O}_2 \\ & & & & & 0.002\mathbf{I}_2 \end{bmatrix}.$$

The variance of estimated response is $V(\hat{y}_0) = 0.042\sigma^2$ for all points in $\mathbf{X}$ and so design is rotatable.

The method discussed in Section 3 leads to the following proposition:

Proposition 3.1. *As the $V(\hat{y}_0)$ is the same for all points in $\mathbf{X}$ which are equidistant from the design centre for $s_1^{n_1} \times s_2^{n_2}$ mixed-level factorial constructed as per the method given in Section 3, the design satisfies Theorem 2.1 and hence rotatable.*

For $\alpha = 0.5$, the eigen values of $(\mathbf{Z}'\mathbf{Z})^{-1}$ are 0.06, 0.042, 0.012, 0.012, 0.01, 0.01, 0.003, 9.1×10^{-5}, 9.1×10^{-5}, 3.3×10^{-5}, 9×10^{-6} and D-efficiency of this design is 2.979. At $\alpha = 0.1$, D-efficiency of this design is 1.958.

Remark 3.1. If s_1 and s_2 both are even, the design is partially rotatable.

3.1. Particular Cases of $s_1^{n_1} \times s_2^{n_2}$

3.1.1. For $n_2 = 1$, i.e., $s_1^{n_1} \times s_2$

Consider n_1 factors having s_1 levels and one factor having s_2 levels. The $n_1 + 1$ columns of $\mathbf{X}$ corresponding to $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_{n_1}, \mathbf{X}_{n_1+1}$ of MLRDNE in $N = n_1(s_1^{n_1} \times s_2)$ points are developed as follows:

$$\mathbf{X} = \left[\mathbf{O}_{N\times n_1} \ \ \mathbf{Q}_{N\times 1}\right].$$

Example 3.2. Let $n_1 = 2$ with $s_1 = 3$ (levels: $1, 0, -1$) and $s_2 = 4$ (levels: $3, 1, -1, -3$), i.e., $3^2 \times 4$. The first three columns of $\mathbf{X}$ corresponding to $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ with $2(3^2 \times 4) = 72$ points are as follows:

$$\mathbf{X} = \left[\mathbf{O}_{72\times 2} \ \ \mathbf{Q}_{72\times 1}\right],$$

$$\mathbf{o_i} = \mathbf{1}_{3^p \times 4} \otimes [1 \quad 0 \quad -1]' \otimes \mathbf{1}_{3^{2-i}}, \quad p = i-1, \quad i = 1, 2,$$
$$\mathbf{q}_1 = \mathbf{1}_{3^2} \otimes [3 \quad 1 \quad -1 \quad -3]'$$
$$\mathbf{Q}_{72\times 1} = [\mathbf{q}_1 \quad \mathbf{q}_1]'.$$

The other columns of $\mathbf{X}$ are generated as per model with 8 columns $(\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_1^2, \mathbf{X}_2^2, \mathbf{X}_3^2, \mathbf{X}_3^3)$.

For $\alpha = 0.5$,

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 288 & \mathbf{0}_2' & 0 & 192\mathbf{1}_2' & 1440 & 0 \\ & 66\mathbf{I}_2 & \mathbf{0}_2 & \mathbf{O}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ & & 1296 & \mathbf{0}_2' & 0 & 10512 \\ & & & \mathbf{A} & 960\mathbf{1}_2 & \mathbf{0}_2 \\ & & & & 11424 & 0 \\ & & & & & 92304 \end{bmatrix},$$

where $\mathbf{A} = 22\mathbf{I}_2 + 128\mathbf{J}_2$.

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 0.05 & \mathbf{0}_2' & 0 & -0.03\mathbf{1}_2' & -0.001 & 0 \\ & 0.002\mathbf{I}_2 & \mathbf{0}_2 & \mathbf{O}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ & & 0.01 & \mathbf{0}_2' & 0 & -0.001 \\ & & & 0.04\mathbf{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ & & & & 0.002 & 0 \\ & & & & & 0.0001 \end{bmatrix}.$$

Thus,

$$\begin{aligned} V(\hat{\beta}_0) &= 0.05\sigma^2, & V(\hat{\beta}_1) &= V(\hat{\beta}_2) = 0.02\sigma^2, \\ V(\hat{\beta}_3) &= 0.01\sigma^2, & V(\hat{\beta}_{11}) &= V(\hat{\beta}_{22}) = 0.04\sigma^2, \\ V(\hat{\beta}_{33}) &= 0.0002\sigma^2, & V(\hat{\beta}_{333}) &= 0.0001\sigma^2, \end{aligned}$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_{11}) = \text{Cov}(\hat{\beta}_0, \hat{\beta}_{22}) = -0.03\sigma^2, \quad \text{Cov}(\hat{\beta}_0, \hat{\beta}_{33}) = -0.001\sigma^2.$$

The variance of estimated response,$V(\hat{y}_0) = 0.056\sigma^2$ for all points in $\mathbf{X}$. Hence, the design is rotatable.

The method discussed in Section 3.1.1 leads to the following proposition:

Proposition 3.2. *As the $V(\hat{y}_0)$ is the same for all points in $\mathbf{X}$ which are equidistant from the design centre for $s_1^{n_1} \times s_2$ mixed-level factorial constructed as per the method given in Section 3.1.1, the design satisfies Theorem 2.1 and hence rotatable.*

For $\alpha = 0.5$, the eigen values of $(\mathbf{Z}'\mathbf{Z})^{-1}$ are $0.09, 0.04, 0.02, 0.02, 0.01, 0.005, 8.4 \times 10^{-5}$, 1×10^{-5} and D-efficiency of this design is 3.663. For $\alpha = 0.1$, D-efficiency of this design is 1.94.

3.1.2. For $n_2 = 0$, i.e., $s_1^{n_1}$

Consider n_1 factors having s_1 levels. The n_1 columns of $\mathbf{X}$ corresponding to $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_{n_1}$ of MLRDNE in $N = n_1 s_1^{n_1}$ points are developed as follows:

$$\mathbf{X} = \left[\mathbf{O}_{N \times n_1}\right].$$

The other columns of $\mathbf{X}$ are generated as per the model and values of s_1.

Example 3.3. Let $n_1 = 4$ with $s_1 = 3$ (levels: 1, 0, −1), i.e., 3^4. The first four columns of $\mathbf{X}$ corresponding to $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{X}_4$ and $4 \times 3^4 = 324$ points are as follows:

$$\mathbf{X} = \begin{bmatrix}\mathbf{O}_{324\times 4}\end{bmatrix}.$$

The other columns of $\mathbf{X}$ are generated as per the model with nine columns $(\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_1^2, \mathbf{X}_2^2, \mathbf{X}_3^2, \mathbf{X}_4^2)$. For α= 0.3,

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 8.294 & \mathbf{0}_4' & 552.961\mathbf{1}_4' \\ & 380.341\mathbf{1}_4' & \mathbf{O}_4 \\ & & \mathbf{B} \end{bmatrix},$$

where $\mathbf{B} = 368.64\mathbf{J}_4 + 126.78\mathbf{I}_4$, $\mathbf{I}_4$ is identity matrix order 4×4, $\mathbf{J}_4$ is matrix of 1's of order 4×4, $\mathbf{1}_4$ is column vector of 1's of order 4×1, $\mathbf{0}_4$ is a column vector of 0's of order 4×1 and $\mathbf{O}_4$ is matrix of 0's of order 4×4.

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 0.015 & \mathbf{0}_4' & -0.005\mathbf{1}_4' \\ & 0.003\mathbf{1}_4' & \mathbf{O}_4 \\ & & 0.008\mathbf{I}_4 \end{bmatrix}.$$

The variance of estimated response, $V(\hat{y}_0) = 0.0152\sigma^2$ for all points in $\mathbf{X}$. Hence, the design is rotatable.

The method discussed in Section 3.1.2 leads to the following proposition:

Proposition 3.3. *As the $V(\hat{y}_0)$ is the same for all points in $\mathbf{X}$ which are equidistant from the design centre for $s_1^{n_1}$ mixed-level factorial constructed as per the method given in Section 3.1.2, the design satisfies Theorem 2.1 and hence rotatable.*

When $\alpha = 0.3$, the eigen values of $(\mathbf{Z}'\mathbf{Z})^{-1}$ are 0.02, 0.008, 0.008, 0.008, 0.003, 0.003, 0.003, 0.003, 0.0004 and D-efficiency for this design is 0.785. If $\alpha = 0.8$, then D-efficiency of this design is 1.831.

4. $s_1^{n_1} \times s_2$ MLRDNE IN SMALLER RUNS

The MLRDNE for $s_1^{n_1} \times s_2$ in a smaller number of runs can be obtained by taking $n_1 + 1$ columns of $\mathbf{X}$ corresponding to $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_{n_1+1}$ in $N = s_1^{n_1} \times s_2$ points are as follows:

$$\mathbf{X} = \begin{bmatrix}\mathbf{C}_{N\times n_1} & \mathbf{d}\end{bmatrix},$$

where $\mathbf{d}$ = (column vector of s_2 levels)$\otimes \mathbf{1}_{s_1^{n_1}}$

$$\mathbf{C}_{N\times n_1} = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_{n_1}\end{bmatrix},$$

$\mathbf{c}_i = \mathbf{1}_{s_1^j \times s_2} \otimes$ (column vector of s_1 levels)$\otimes \mathbf{1}_{s_1^{n_1-i}}$, where $i = 1, ..., n_1$ and $j = i - 1$. Each vector $\mathbf{c}_i$ is of order $N \times 1$. The other columns of $\mathbf{X}$ are generated as per the model and values of s_1 and s_2.

Example 4.1. For $n_1 = 2$ with $s_1 = 3$ (levels: $1, 0, -1$) and $s_2 = 4$ (levels: $3, 1, -1, -3$), i.e., $3^2 \times 4$, the first three columns of X corresponding to $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ and $3^2 \times 4 = 36$ points are as follows:

$$\mathbf{X} = \left[\mathbf{C}_{36\times 2} \ \ \mathbf{d}\right],$$

where

$$\mathbf{d} = [3 \quad 1 \quad -1 \quad -3]' \otimes \mathbf{1}_9,$$
$$\mathbf{c}_i = \mathbf{1}_{3^j\times 4} \otimes [1 \quad 0 \quad -1]' \otimes \mathbf{1}_{3^{2-i}}.$$

The other columns of $\mathbf{X}$ are generated as per the model with eight columns $(\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_1^2, \mathbf{X}_2^2, \mathbf{X}_3^2, \mathbf{X}_3^3)$ and for $\alpha = 0.7$,

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 207.36 & 0 & 0 & 0 & 138.24 & 138.24 & 103.68 & 0 \\ & 81.12 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 2.16 & 0 & 0 & 0 & 0 & 0 \\ & & & 922.56 & 0 & 0 & 0 & 7473.6 \\ & & & & 119.2 & 92.16 & 691.2 & 0 \\ & & & & & 92.88 & 691.2 & 0 \\ & & & & & & 8197.12 & 0 \\ & & & & & & & 65519.04 \end{bmatrix},$$

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 0.64 & 0 & 0 & 0 & -0.025 & -0.93 & -0.002 & 0 \\ & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 0.46 & 0 & 0 & 0 & 0 & 0 \\ & & & 0.014 & 0 & 0 & 0 & -0.002 \\ & & & & 0.04 & 0 & 0 & 0 \\ & & & & & 1.39 & 0 & 0 \\ & & & & & & 0.0003 & 0 \\ & & & & & & & 0.0002 \end{bmatrix}.$$

Variance of the estimated response, $V(\hat{y}_0) = 0.655\sigma^2$ for all points in $\mathbf{X}$ and thus the design is rotatable. However, it can be seen that the variances of a particular order of estimates are not same unlike Example 3.2.

For $\alpha = 0.7$, the eigen values of $(\mathbf{Z}'\mathbf{Z})^{-1}$ are 2.015, 0.463, 0.051, 0.014, 0.012, 0.007, 0.0001, 0.00001 and this design having the D-efficiency 2.77. When $\alpha = 0.2$, D-efficiency of this design is 2.192.

A list of MLRDNE is given in Appendix 2 containing n_1, n_2, s_1, s_2 and the variance of estimated response at $\alpha = 0$, 0.3, 0.5, 0.7 and 0.9. It is seen that in the presence of the neighbour effect and when the value of α increases, the variance of estimated response is in decreasing order.

5. R PACKAGE FOR THE GENERATION OF MLRDNE

An `R` package named `rsdNE` has been developed for the generation of MLRDNE. It also computes the variance of parameter estimates and variance of the predicted response. The package is made available at `https://cran.r-project.org/package=rsdNE`. This package includes sym(), asym1(), asym2() functions that generates response surface designs which are rotatable under a polynomial model of a given order without interaction term incorporating neighbour effects. A few Snapshots of the package has been given in the Appendix 1.

6. DISCUSSION

This article attempts to provide a series of rotatable response surface designs when immediate (left and right at distance 1) neighbour effects are suspected in an experiment. A general procedure for the construction of Mixed Level Rotatable Design with Neighbour Effects (MLRDNE) of the form $s_1^{n_1} \times s_2^{n_2}$ is given. For these designs it is seen that as the value of α increases the variance of the estimated response decreases and on the other hand, D-efficiency increases. The `R` package `rsdNE` developed would help the experimenter to generate MLRDNE design along with the variance of the parameter estimates as well as the variance of predicted response.

APPENDIX 1

```
> library(rsdNE)
> asym1(1,1,0.5)
[1] "X matrix"
     [,1] [,2] [,3] [,4]
[1,]    1   -1   -1    1
[2,]    1    1    1    1
[3,]    1    1    0    0
[4,]    1    1   -1    1
[5,]    1   -1    1    1
[6,]    1   -1    0    0
[7,]    1   -1   -1    1
[8,]    1    1    1    1
[1] "Z_prime_Z matrix"
     [,1] [,2] [,3] [,4]
[1,]   24    0    0   16
[2,]    0   12    0    0
[3,]    0    0    1    0
[4,]   16    0    0   11
[1] "inv(Z_prime_Z) matrix"
       [,1]       [,2] [,3] [,4]
[1,]  1.375 0.00000000    0   -2
[2,]  0.000 0.08333333    0    0
[3,]  0.000 0.00000000    1    0
[4,] -2.000 0.00000000    0    3
[1] "total number of runs" "6"
[1] "variance of estimated response" "1.4583"
```

Figure 1: 2 factors having 2 and 3 levels, i.e., 2×3 when $\alpha = 0.5$.

```
R 4.2.2 · ~/
> asym2(3,2,4,2,0.5)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
 [1,]    1   -1   -1   -3   -3    1    1    9    9   -27   -27
 [2,]    1    1    1    3    3    1    1    9    9    27    27
 [3,]    1    1    0    3    1    1    0    9    1    27     1
 [4,]    1    1   -1    3   -1    1    1    9    1    27    -1
 [5,]    1    0    1    3   -3    0    1    9    9    27   -27
 [6,]    1    0    0    1    3    0    0    1    9     1    27
 [7,]    1    0   -1    1    1    0    1    1    1     1     1
 [8,]    1   -1    1    1   -1    1    1    1    1     1    -1
 [9,]    1   -1    0    1   -3    1    0    1    9     1   -27
[10,]    1   -1   -1   -1    3    1    1    1    9    -1    27
[11,]    1    1    1   -1    1    1    1    1    1    -1     1
[12,]    1    1    0   -1   -1    1    0    1    1    -1    -1
[13,]    1    1   -1   -1   -3    1    1    1    9    -1   -27
[14,]    1    0    1   -3    3    0    1    9    9   -27    27
[15,]    1    0    0   -3    1    0    0    9    1   -27     1
[16,]    1    0   -1   -3   -1    0    1    9    1   -27    -1
[17,]    1   -1    1   -3   -3    1    1    9    9   -27   -27
[18,]    1   -1    0    3    3    1    0    9    9    27    27
[19,]    1   -1   -1    3    1    1    1    9    1    27     1
[20,]    1    1    1    3   -1    1    1    9    1    27    -1
[21,]    1    1    0    3   -3    1    0    9    9    27   -27
[22,]    1    1   -1    1    3    1    1    1    9     1    27
```

Figure 2: **X** matrix for $3^2 \times 4^2$ MLRDNE.

```
[1] "Z_prime_matrix"
      [,1] [,2] [,3]  [,4]  [,5] [,6] [,7]  [,8]  [,9]  [,10]  [,11]
 [1,] 1152    0    0     0     0  768  768  5760  5760      0      0
 [2,]    0  264    0     0     0    0    0     0     0      0      0
 [3,]    0    0  264     0     0    0    0     0     0      0      0
 [4,]    0    0    0  2808     0    0    0     0     0  21816      0
 [5,]    0    0    0     0  2808    0    0     0     0      0  21816
 [6,]  768    0    0     0     0  600  512  3840  3840      0      0
 [7,]  768    0    0     0     0  512  600  3840  3840      0      0
 [8,] 5760    0    0     0     0 3840 3840 38592 28800      0      0
 [9,] 5760    0    0     0     0 3840 3840 28800 38592      0      0
[10,]    0    0    0 21816     0    0    0     0     0 180792      0
[11,]    0    0    0     0 21816    0    0     0     0      0 180792
```

Figure 3: **Z′Z** matrix for $3^2 \times 4^2$ MLRDNE.

```
[1] "inv(Z_prime_Z) matrix"
               [,1]        [,2]        [,3]          [,4]          [,5]         [,6]
 [1,]  0.0160752748 0.000000000 0.000000000  0.0000000000  0.0000000000 -0.007575758
 [2,]  0.0000000000 0.003787879 0.000000000  0.0000000000  0.0000000000  0.000000000
 [3,]  0.0000000000 0.000000000 0.003787879  0.0000000000  0.0000000000  0.000000000
 [4,]  0.0000000000 0.000000000 0.000000000  0.0056985294  0.0000000000  0.000000000
 [5,]  0.0000000000 0.000000000 0.000000000  0.0000000000  0.0056985294  0.000000000
 [6,] -0.0075757576 0.000000000 0.000000000  0.0000000000  0.0000000000  0.011363636
 [7,] -0.0075757576 0.000000000 0.000000000  0.0000000000  0.0000000000  0.000000000
 [8,] -0.0005106209 0.000000000 0.000000000  0.0000000000  0.0000000000  0.000000000
 [9,] -0.0005106209 0.000000000 0.000000000  0.0000000000  0.0000000000  0.000000000
[10,]  0.0000000000 0.000000000 0.000000000 -0.0006876362  0.0000000000  0.000000000
[11,]  0.0000000000 0.000000000 0.000000000  0.0000000000 -0.0006876362  0.000000000
              [,7]          [,8]          [,9]         [,10]         [,11]
 [1,] -0.007575758 -5.106209e-04 -0.0005106209  0.000000e+00  0.000000e+00
 [2,]  0.000000000  0.000000e+00  0.0000000000  0.000000e+00  0.000000e+00
 [3,]  0.000000000  0.000000e+00  0.0000000000  0.000000e+00  0.000000e+00
 [4,]  0.000000000  0.000000e+00  0.0000000000 -6.876362e-04  0.000000e+00
 [5,]  0.000000000  0.000000e+00  0.0000000000  0.000000e+00 -6.876362e-04
 [6,]  0.000000000  0.000000e+00  0.0000000000  0.000000e+00  0.000000e+00
 [7,]  0.011363636  0.000000e+00  0.0000000000  0.000000e+00  0.000000e+00
 [8,]  0.000000000  1.021242e-04  0.0000000000  0.000000e+00  0.000000e+00
 [9,]  0.000000000 -2.834515e-20  0.0001021242  0.000000e+00  0.000000e+00
[10,]  0.000000000  0.000000e+00  0.0000000000  8.850763e-05  0.000000e+00
[11,]  0.000000000  0.000000e+00  0.0000000000  0.000000e+00  8.850763e-05
[1] "total number of runs" "288"
[1] "variance of estimated response" "0.0231"
>
```

Figure 4: $\mathbf{Z'Z})^{-1}$ matrix, N and $V(\hat{y}_0)$ for $3^2 \times 4^2$ MLRDNE.

APPENDIX 2

Table 1: MLRDNE.

				$V(\hat{y}_0)$				
n_1	n_2	s_1	s_2	$\alpha = 0$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
2	0	4	—	0.2188	0.1555	0.1166	0.0856	0.0638
2	0	5	—	0.1800	0.1212	0.0894	0.0653	0.0484
2	0	6	—	0.1528	0.0994	0.0723	0.0525	0.0390
2	1	2	4	0.2083	0.1663	0.1645	0.1361	0.0975
2	1	2	5	0.1750	0.1448	0.1395	0.1145	0.0822
2	1	2	6	0.1667	0.1304	0.1229	0.1002	0.0720
2	1	3	4	0.1111	0.0723	0.0556	0.0427	0.0330
2	1	3	5	0.1000	0.0625	0.0475	0.0363	0.0280
2	1	3	6	0.0926	0.0560	0.0422	0.0320	0.0246
2	1	4	2	0.1250	0.0844	0.0626	0.0458	0.0341
2	1	4	3	0.0937	0.0605	0.0445	0.0325	0.0241
2	1	4	5	0.0688	0.0414	0.0300	0.0218	0.0162
2	1	4	6	0.0625	0.0366	0.0263	0.0191	0.0142
2	1	5	2	0.1000	0.0647	0.0474	0.0345	0.0256
2	1	5	3	0.0733	0.0458	0.0333	0.0242	0.0179
2	1	5	4	0.0600	0.0364	0.0263	0.0190	0.0141
2	1	5	6	0.0467	0.0269	0.0192	0.0139	0.0103
2	1	6	2	0.0833	0.0525	0.0379	0.0275	0.0205
2	1	6	3	0.0602	0.0369	0.0265	0.0192	0.0142
2	1	6	4	0.0486	0.0290	0.0207	0.0150	0.0111
2	1	6	5	0.0417	0.0243	0.0173	0.0125	00093
2	2	2	5	0.0550	0.0476	0.0424	0.0336	0.0243
2	2	3	4	0.0382	0.0295	0.0231	0.0175	0.0134
2	2	3	5	0.0289	0.0213	0.0164	0.0124	0.0094
2	2	4	5	0.0188	0.0133	0.0099	0.0073	0.0054
2	2	5	6	0.0106	0.0071	0.0052	0.0038	0.0028

ACKNOWLEDGMENTS

We are grateful to the editor and anonymous reviewers for the constructive comments that have led to considerable improvement in the article. We also acknowledge the facilities provided by ICAR-IASRI, New Delhi, P.G. School, IARI, New Delhi, ICAR-CMFRI, Kochi for conducting the research. Financial assistance received in the form of Junior Research Fellowship (AD) and Lal Bahadur Shashtri Young Scientist Award Project (EV) from Indian Council of Agricultural Research, Govt. of India is also duly acknowledged.

REFERENCES

Bartlett, M. S. (1938). The approximate recovery of information from replicated field experiments with large blocks. *Journal of Agricultural Science*, 28:418–427.

Bartlett, M. S. (1978). Nearest neighbour models in the analysis of field experiments. *Journal of the Royal Statistical Society: Series B*, 40(2):147–158.

Box, G. E. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(1):1–38.

Box, G. E. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces*. New York, Wiley.

Dalal, A., Jaggi, S., Varghese, E., Bhowmik, A., Varghese, C., and Datta, A. (2022). `NBBDesigns` and `rsdNE`: R packages for the generation of designs and analysis of data incorporating neighbour effects. Paper presented in the useR virtual conference, VUMC, Nashville, 23^{rd} June, 2022. `https://user2022.r-project.org/program/talks/#session-37-unique-applications-and-methods`

Draper, N. R. and Guttman, I. (1980). Incorporating overlap effects from neighbouring units into response surface models. *Applied Statistics*, 29(2):128–134.

Jaggi, S., Sarika and Sharma, V. K. (2010). Response surface analysis incorporating neighbour effects from adjacent units. *Indian Journal of Agricultural Sciences*, 80(8):719–723.

Khuri, A. I. and Cornell, J. A. (1996). *Response Surface Designs and Analyses*. Marcel Dekker, New York.

Kumar, J., Jaggi, S., Varghese, E., Bhowmik, A., and Varghese, C. (2020). First order rotatable designs incorporating differential neighbour effects from experimental units up to distance 2. *Metrika*, 83:923–935.

Myers, R.H. (1995). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley and Sons, New York.

Myers, R. H., Montgomery, D. C., and Andersson-Cook, C. M. (2009). *Response Surface Methodology – Process and Product Optimization Using Designed Experiments*. New York, John Wiley Publication.

Papadakis, J. S. (1937). Mithode statistique pour des expiriences sur champ. *Bulletin Scientifique, Institut d'Amelioration des Plantes a Thessaloniki*, (Greece) 23:12–28.

Sarika, Jaggi, S., and Sharma, V. K. (2009). Second-order response surface model with neighbour effects. *Communications in Statistics – Theory and Methods*, 38(9):1393–1403.

Sarika, Jaggi, S., and Sharma, V. K. (2013). First order rotatable designs incorporating neighbour effects. *ArsCombinatoria*, 112:145–159.

Varghese, E., Jaggi, S., and Sarika (2013). Response surface model with neighbour effects and correlated observations. *Model Assisted Statistics and Applications*, 8(1):41–49.

Varghese, E., Jaggi, S., and Sharma, V. K. (2016). Rotatable response surface designs in the presence of differential neighbour effects from adjoining experimental units. *Calcutta Statist. Association Bulletin*, 67(3-4):163–186.

Varghese, E., Bhowmik, A., Jaggi, S., Varghese, C., and Lall, S. (2019). On the construction of response surface designs with minimum level changes. *Utilitas Mathematica*, 110:293–303.

Varghese, E., Kumar, J., Jaggi, S., Varghese, C., and Bhowmik, A. (2020). A note on constructing small rotatable designs under first order response surface interference model. *Utilitas Mathematica*, 115:171–180.

Verma, A., Jaggi, S., Varghese, E., Varghese, C., Bhowmik, A., Datta, A., and Hemavathi, M. (2021). On the construction of mixed level rotatable response surface designs when experimental unit experiences overlap effects. *Communication in Statistics – Simulation and Computation*, DOI: 10.1080/03610918.2021.1890123.

REVSTAT – Statistical Journal
Volume 23, Number 2, April 2025, 253–271
https://doi.org/10.57805/revstat.v23i2.563

Bimodal and Multimodal Extensions of the Normal and Skew Normal Distributions

Authors: Emilio Gómez-Déniz
– Department of Quantitative Methods in Economics and TiDES Institute,
University of Las Palmas de Gran Canaria,
Spain
emilio.gomez-deniz@ulpgc.es

Enrique Calderín-Ojeda ✉
– Department of Economics, University of Melbourne,
Australia
enrique.calderin@unimelb.edu.au

José M. Sarabia
– Department of Quantitative Methods, CUNEF Universidad,
Spain
josemaria.sarabia@cunef.edu

Received: February 2022 Revised: June 2023 Accepted: June 2023

Abstract:

- A transformation of a density function is introduced to derive two families of continuous densities, the first symmetric and the second not-necessarily symmetric, exhibiting both unimodality and bimodality. Their respective density functions are provided in closed form, allowing us to simply obtain moments and related quantities. We focus on the case where the normal distribution is considered, although it can be applied to other models, such as the logistic and Cauchy distributions. This transformation is also extended to derive a family of asymmetric unimodal and bimodal distributions via Azzalini's scheme. An example related to environmental science illustrate these models' practical performance.

Keywords:

- *multimodality; old faithful geyser data; skewness; unimodality; univariate distribution.*

AMS Subject Classification:

- 62P05, 97M30, 91B30.

✉ Corresponding author

1. INTRODUCTION

We use an old theorem proven over ninety years ago to obtain bimodal and multimodal extensions of the normal distribution and the skew-normal distribution. One can almost certainly say that the normal distribution constitutes the queen of the comprehensive family of the continuous probability distributions. Since the end of the 19th century, numerous researchers, such as the distinguished F. Y. Edgeworth, and also Chas. H. Kummel, Arthur L. Bowley, Morgan W. Crofton, among many others derived modifications of the normal law to discuss situations where the empirical data presented some asymmetry that the normal distribution could not explain. A review of the normal distribution and some of its modifications can be found in Patel and Campbell (1984).

Bimodal distributions arise in nature in many different scenarios. Perhaps, one of the most relevant phenomena that can be explained with distributions is the disease patterns. For example, the incidence of some types of cancers by age displays a major mode for young adults and minor mode for older adults (see Anderson *et al.*, 2006). In addition, the occurrence of bimodality has also implications in geoscience (see Hirota *et al.*, 2011). Finding appropriate probabilistic models that can explain bivariate datasets is an issue of vital importance. In this work, we propose an extension of the normal and skew-normal densities that may be unimodal or bimodal. This new family of distributions that arises from an old Theorem provided by Slobin (1927) comprises flexible parametric families of continuous distributions that are useful in statistical practice.

In the last years, different techniques to extend the normal family have been deemed in the statistical literature: the skew-normal distribution in Azzalini (1985) (see also Azzalini, 1986), the Balakrishnan skew-normal density in Sharafi and Behboodian (2008) (more details in Teimouri and Nadarajah, 2016), the generalization proposed by Arnold and Beaver (2002), the Sinh-arcsinh family introduced by Jones and Pewsey (2009), the generalized normal one in García *et al.* (2010), Gómez-Déniz *et al.* (2021) and Gómez-Déniz *et al.* (2021), and the recently proposed models provided by Venegas *et al.* (2018) and Sulewski (2022), among others. Some other works related to the normal and skew normal densities are Arellano-Valle *et al.* (2004), Arellano-Valle *et al.* (2005) and Gómez *et al.* (2007). For a comprehensive review of the skew normal families the reader is referred to Azzalini (2013).

The density function introduced here resembles some important properties satisfied by the normal distribution. The first family is symmetric with positive real support. The second family is asymmetric and defined on the positive real numbers. In general, both families show bimodality. An overview of this work that will undoubtedly help the reader to understand better the elements that are not so essential is illustrated in the flowchart displayed in Figure 4.

The rest of this paper is structured as follows. In Section 2 we derive the methodology based on the use of a result provided in Slobin (1927) to derive the new family of distribution. Here, expressions for the mean, variance, and other features for the general model are also provided. Next, we also examine the special case of considering the classical normal distribution as the parent distribution. Then, to break the symmetry of the latter case, we introduce the skew-normal distribution as the baseline model. In Section 3, the parameter estimation problem is discussed. Some illustrative examples related to environmental issues, in particular in geoscience, are analyzed in Section 4. Finally, closing comments and modifications of the models proposed are shown in the last Section.

2. THE PROPOSED MODEL

This section gives the main results of this paper, from which we derive the two families of probability density functions that will be described later. The first family is introduced in the second theorem of this section. Although any distribution with support on the real line can be used as a candidate of this new distribution, the normal case is the one we are examining in this section. It can be simply shown after a change of variable that this model is connected to the generalized inverse Gaussian distribution. This probabilistic family is symmetric and has two modal values that are equidistant with respect to the axis of symmetry. The second family presents the advantage of having an asymmetric density function. We begin with the following Theorem found in Slobin (1927) that is required for the main result of this work.

Theorem 2.1 (Slobin, 1927). *Let the function $\omega(x) = x - 1/x$, $x \neq 0$. Then, if the function $m(x)$ is a function integrable on $\mathbb{R} = (-\infty, \infty)$ and if the function $m(\omega(x))$ is also integrable in $\mathbb{R} = (-\infty, \infty)$, we have that*

$$\int_{-\infty}^{\infty} m(\omega(x))\, dx = \int_{-\infty}^{\infty} m(x)\, dx. \tag{2.1}$$

Following the same arguments that the ones provided in the proof of the above Theorem given in Slobin (1927), it is simple to observe that (2.1) is also valid for $\omega_\alpha(x) = x - \alpha/x$, being $\alpha \geq 0$. The following result provides an alternative and more simple proof than the one given in Slobin (1927) for this case. Previously we need the following Lemma, which is provided in Behboodian (1978).

Lemma 2.1 (Behboodian, 1978). *Let X be a symmetric random variable, and let $y = h(x)$ be an odd real-valued function. Then, the random variable $Y = h(X)$ is also symmetric.*

As a result of this Lemma, if X is a symmetric random variable then the random variable $Y = \omega_\alpha(X)$ is also symmetric. In the next result we derive an expression for the density function of $Y = \omega_\alpha(X)$.

Theorem 2.2. *Let $f(x)$ be a probability density function (pdf hereafter) symmetric about 0 and consider the function $f(\omega_\alpha(x))$, with $\omega_\alpha(x) = x - \alpha/x$, being $\alpha \geq 0$. Then, if $df(\omega_\alpha(x))/(d\alpha)$ is also a symmetric function we have that $\int_{-\infty}^{\infty} f(\omega_\alpha(x))\, dx = 1$.*

Proof: Since $f(x)$ is symmetrical and $\omega_\alpha(x)$ is an odd function, using Lemma 2.1 we have that $f(\omega_\alpha(x))$ is also symmetrical. Now, consider the function $\nu(\alpha) = \int_{-\infty}^{\infty} f(\omega_\alpha(x))\, \mathrm{d}x$ for which we have that

$$\nu'(\alpha) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\nu(\alpha) = -\int_{-\infty}^{\infty} \frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}\alpha} f(\omega_\alpha(x))\, \mathrm{d}x = 0,$$

because $\mathrm{d}f(\omega_\alpha(x))/(\mathrm{d}\alpha)$ is symmetrical (by assumption). Therefore, $\nu(\alpha)$ is constant and since $\nu(0) = 1$ we have the result. □

Based on the use of Theorem 2.2 we can build a family of pdf's by taking

$$g_\alpha(x) = \begin{cases} f(\omega_\alpha(x)), & x \neq 0, \\ f(0), & x = 0, \end{cases} \tag{2.2}$$

where $\alpha \geq 0$. Note that this is a two piece-wise pdf.

The following proposition displays some essential properties related to this distribution.

Proposition 2.1. *The pdf given in (2.2) satisfies the following properties:*

- **(i)** *$g_\alpha(x)$ is symmetric about zero. That is, $g_\alpha(x) = g_\alpha(-x)$ for all $x \in \mathbb{R}$. In fact, the random variable $Z = -X$ follows the same distribution that X.*
- **(ii)** *$g_0(x) = f(x)$.*
- **(iii)** *$g_\alpha(0) = f(0)$ for all $\alpha \geq 0$.*
- **(iv)** *$\mathbb{E}(X^{2\kappa+1}) = 0$, $\kappa \in \{0, 1, ...\}$. That is, all odd raw moments are zero.*
- **(v)** *The random variables $Y = \omega_\alpha(X)$ and $Z = g_\alpha(X)$ are uncorrelated and therefore $\mathrm{cov}(Y, Z) = 0$, provided that all the first and second moments of Y and Z exist.*

Proof: Properties (i)–(iv) are direct. To show (v), observe that $\omega_\alpha(x)$ is an odd function, $g_\alpha(x)$ is an even real-valued (measurable) function and the random variable $T = YZ$ satisfies that $T(-x) = \omega_\alpha(-x)g_\alpha(-x) = -\omega_\alpha(x)g_\alpha(x) = -T(x)$, therefore is an odd function. Thus, $\mathrm{cov}(Y, Z) = \mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z) = 0$, because $\mathbb{E}(Y) = 0$ (due to Lemma 2.1, Y is symmetrical) and $\mathbb{E}(YZ) = 0$ ($T = YZ$ is an odd function). For more details see Behboodian (1978). □

2.1. THE NORMAL CASE

Natural choices for $f(x)$ to be plugged into (2.2) are the Cauchy distribution, the Student's t distribution, and the normal distribution that will be the one considered in the rest of this work, i.e. $f(x) = \phi(x)$, being $\phi(x)$ the pdf of the standard normal distribution. Then, it is simple to see that

$$g_\alpha(x) = \begin{cases} \phi(\omega_\alpha(x)), & x \neq 0, \\ \phi(0), & x = 0, \end{cases} \tag{2.3}$$

is a genuine pdf for $\alpha \geq 0$. Note that the special case $\alpha = 0$ represents the standard normal distribution. Simple algebra provides that the distribution is symmetric about zero and has mean and variance given by 0 and $1 + \alpha$, respectively. The distribution is always bimodal, with two modes in $x = -\sqrt{\alpha}$ and $x = \sqrt{\alpha}$. To see this, observe that

$$g'_\alpha(x) = -g_\alpha(x)\Big(x - \frac{\alpha}{x}\Big)\Big(1 + \frac{\alpha}{x^2}\Big) = 0$$

for $x = \pm\sqrt{\alpha}$. Now, it is simple to see that $g''_\alpha(\pm\sqrt{\alpha}) < 0$. The antimode is obviously $x = 0$. Henceforward, we will write $X \sim \mathrm{BN}(\alpha)$ when the random variable X follows the pdf given in (2.3), denoting that is a bimodal generalization of the normal distribution.

The entropy does not depend on α and is equivalent to the one of the standard normal distribution. Observe that $\lim_{x\to 0^+} g_\alpha(x) = \lim_{x\to 0^-} g_\alpha(x) = \phi(0)$ and thus the pdf defined in (2.3) is a continuous function.

Figure 1 displays the graphs of the pdf given in (2.3) for selected values of parameter $\alpha \geq 0$. The α parameter, the only parameter of the distribution, clearly indicates two fundamental things: first, if it takes the value zero, we are in the case of the standard normal distribution; second, a value other than zero provides a distribution with two modes that are equidistant with respect to the axis of symmetry. The distance between the modes increases with the value of α.

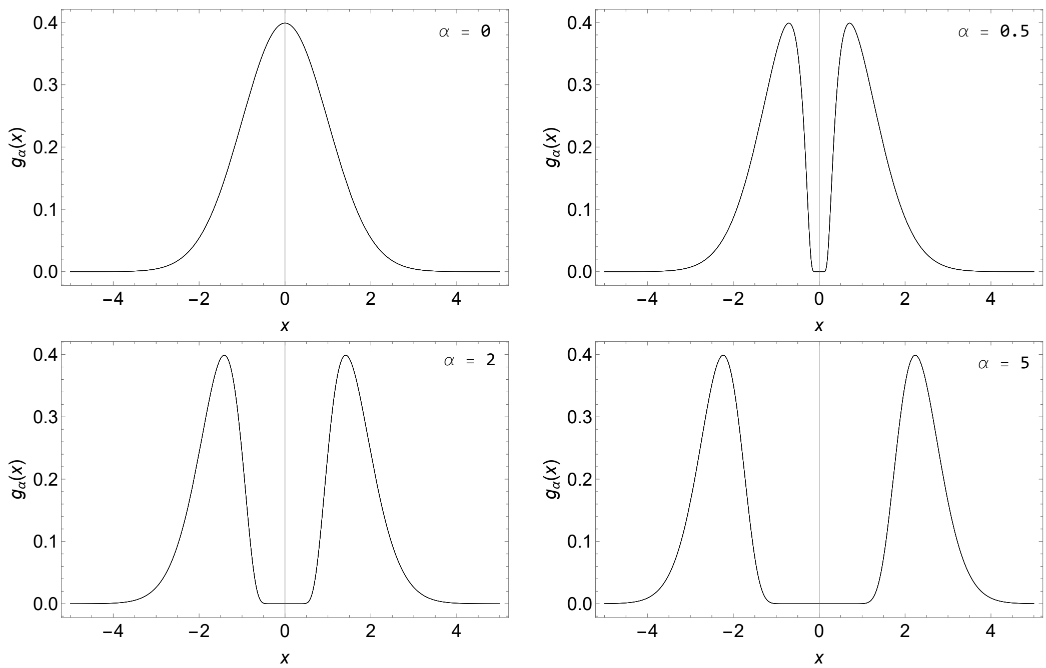

Figure 1: Plots of the pdf $g_\alpha(x)$ for selected values of the parameter α.

2.2. Connection with others distributions

The following result connects the proposed distribution with the generalized inverse Gaussian distribution. Recall that a continuous variable $Z > 0$ follows a generalized inverse Gaussian distribution (see Jørgensen, 1982 and Johnson *et al.*, 1995, Chapter 15) with parameters $a > 0$, $b > 0$ and $r \in \mathbb{R}$ if its pdf is given by

$$(2.4) \qquad f(z) = \frac{(a/b)^{r/2}}{2K_r(\sqrt{ab})} z^{r-1} \exp\left[-\frac{1}{2}\left(az + \frac{b}{z}\right)\right], \quad z > 0,$$

where $K_\nu(s)$ gives the modified Bessel function of the second kind. Furthermore, if Z follows a generalized inverse Gaussian distribution, then $1/Z$ follows a reciprocal generalized inverse Gaussian distribution. Additionally, simple computation provides that the random variable $1/X^2$ follows a reciprocal generalized inverse Gaussian distribution.

Proposition 2.2. *Let $X \sim BN(\alpha)$ with the pdf given in (2.3). Then, the random variable $V = X^2$ follows a generalized inverse Gaussian distribution with parameters $a = 1$, $b = \alpha^2$ and $r = 1/2$.*

Proof: Since $\mathrm{d}x = 1/(2\sqrt{v})\mathrm{d}v$ we have that

$$\begin{aligned} g_\alpha(v) &= \frac{1}{2\sqrt{2v\pi}} \exp\left[-\frac{1}{2}\left(\sqrt{v} - \frac{\alpha}{\sqrt{v}}\right)^2\right] \\ &= \frac{v^{-1/2}\exp(\alpha)}{2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(v + \frac{\alpha^2}{v}\right)\right]. \end{aligned} \tag{2.5}$$

Now, having into account that $K_{1/2}(\alpha) = \exp(-\alpha)\sqrt{\pi/(2\alpha)}$, the result follows by comparing (2.5) with (2.4). □

Proposition 2.3. *Let $X \sim BN(\alpha)$ with the pdf given in (2.3). Then, it is verified that $\mathbb{E}(X^\kappa) = 0$ if κ (positive or negative) is odd while the even moments (positive or negative) are given by*

$$\mathbb{E}(X^{2\kappa}) = \sqrt{\frac{2\alpha^{1+2\kappa}}{\pi}} \exp(\alpha) K_{\kappa+\frac{1}{2}}(\alpha), \quad \kappa \in \{0, 1, ...\}. \tag{2.6}$$

Proof: Since the distribution given in (2.3) is symmetrical, then all odd-order moments are equal to zero. To see that (2.6) is true, then it is simple to see that the distribution is symmetrical since we have that

$$\mathbb{E}(X^\kappa) = 2\int_0^\infty \phi(\omega_\alpha(x)\,\mathrm{d}x$$

and by making the change of variable $u = x^2$ we get

$$\mathbb{E}(X^\kappa) = \frac{2\exp(\alpha)}{\sqrt{2\pi}} \int_0^\infty u^{(\kappa-1)/2} \exp\left[-\frac{1}{2}\left(u + \frac{\alpha^2}{u}\right)\right] \mathrm{d}u \tag{2.7}$$

from which the result follows immediately by arranging parameters in (2.7) and identifying it with the pdf of the generalized inverse Gaussian distribution given in (2.4). □

In particular, if $\kappa = 1$ we get the second row moment of the distribution, which coincides with the variance, given by $\mathrm{var}(X) = 1 + \alpha$. Furthermore, if $\kappa = -1$ by using (2.6) we have that

$$\mathbb{E}\left(\frac{1}{X^2}\right) = \frac{1}{\alpha}, \quad \alpha \neq 0, \tag{2.8}$$

and

$$\mathbb{E}\left[\left(X - \frac{\alpha}{X}\right)^{2\kappa}\right] = (2\kappa - 1)!!, \tag{2.9}$$

where $n!! = n(n-2)(n-4)\cdots 2 \cdot 1$ represents the double factorial.

Note that property given in (2.9) is shared with the standard normal distribution. Using the series representation of the exponential function, we derive the moment generating function of the distribution, which is given by

$$\mathrm{M}_X(t) = \mathbb{E}[\exp(tX)] = \sum_{j=0}^{\infty} \frac{t^{2j}}{(2j)!} \sqrt{\frac{2\alpha^{1+2j}}{\pi}} \exp(\alpha) K_{j+\frac{1}{2}}(\alpha).$$

Proposition 2.4. *The cumulative distribution function (cdf henceforward), $G_\alpha(x) = \Pr(X \le x)$, for a continuous random variable following the pdf given in (2.3) is*

$$G_\alpha(x) = \frac{1}{2}[\Phi(\omega_\alpha(x)) + \Phi(\tau_\alpha(x))\exp(2\alpha)], \quad x < 0, \tag{2.10}$$

$$G_\alpha(x) = 1 - \frac{1}{2}\left[\bar{\Phi}(\omega_\alpha(x)) + \bar{\Phi}(\tau_\alpha(x))\exp(2\alpha)\right], \quad x > 0, \tag{2.11}$$

and $G_\alpha(0) = 1/2$, where $\tau_\alpha(x) = x + \alpha/x$ and $\bar{\Phi}(z) = 1 - \Phi(z)$ is the survival function of the standard normal distribution.

Proof: The proof is obtained in the following way. Let $G_\alpha(-x) = \Pr(X \le -x)$. Thus,

$$G_\alpha(-x) = \int_{-\infty}^{-x} \phi(\omega_\alpha(t))\,\mathrm{d}t = \int_{x}^{\infty} \phi(\omega_\alpha(t))\,\mathrm{d}t,$$

which can be written, after the change of variable $Y = X^2$, as

$$G_\alpha(-x) = \int_x^\infty \frac{\exp(\alpha)}{\sqrt{2y\pi}} \exp\left[-\frac{1}{2}\left(y + \frac{\alpha^2}{y}\right)\right] \mathrm{d}y.$$

Now, by using the cdf of the generalized inverse Gaussian distribution provided in Malinovskii (2017) we get, after simple algebra (2.10). Expression (2.11) is obtained in a similar way. □

A random variate X from the random variable with pdf given by (2.3) is derived as follows:

- Generate a random number u from the standard uniform distribution, $\mathrm{U}(0,1)$.
- Generate random variate v from the generalized inverse Gaussian distribution with parameters $a = 1$, $b = \alpha^2$ and $r = 1/2$.
- If $u < 0.5$ then $x = -\sqrt{v}$; otherwise $x = \sqrt{v}$.

2.3. Extensions

The major disadvantage of the family of distributions given in (2.3) lies in its symmetry and also in the fact that the two modes are equidistant with respect to the axis of symmetry. Since $f(\omega_\alpha(x)$ is a symmetric pdf, by using the representation provided by Azzalini (1985), we can consider the more flexible family of pdf's given by

$$g_{\alpha,\lambda}(x) = \begin{cases} 2\Phi(\lambda x)\phi(\omega_\alpha(x)), & x \ne 0, \\ \phi(0), & x = 0, \end{cases} \tag{2.12}$$

where $\alpha \ge 0$ and $\lambda \in \mathbb{R}$.

In practice $\Phi(\lambda x)$ can be replaced by $\Phi(\lambda\, m(x))$ for any odd function $m(\cdot)$ in order to ensure that (2.13) represents a proper density function. In particular, we can take $m(x) = \omega_\beta(x)$, $\beta \in \mathbb{R}$, to build the family of pdf's given by

$$g_{\alpha,\beta,\lambda}(x) = \begin{cases} 2\Phi(\lambda\, \omega_\beta(x))\phi(\omega_\alpha(x)), & x \neq 0, \\ \phi(0), & x = 0, \end{cases} \tag{2.13}$$

where $\alpha \geq 0$, $\beta \in \mathbb{R}$ and $\lambda \in \mathbb{R}$. See for instance Azzalini (2013). Observe that when $\alpha = \beta = 0$ the pdf given in (2.13) reduces to the skew normal density provided in Azzalini (1985). See also Azzalini (1986) and Azzalini and Bowman (1990), among others. Azzalini (1985), Azzalini (1986), Chiogna (1998), Henze (1986) and Gupta *et al.* (2004), among other papers, provide many properties of the skew normal density. The standard normal distribution is obtained for $\alpha = \lambda = 0$. A probabilistic representation of this family of distribution can be obtained in a similar fashion as the one provided in Azzalini (1986) and Henze (1986) (see also Azzalini, 2013).

To see that (2.13) represents a genuine pdf, we proceed in a similar way as we did in Theorem 2.2. In this case, we have to add that $\Phi(\cdot)$ is a bounded function with a derivative being a symmetric density function about zero. The family (2.13) contains the normal, the skew normal density and others for $\lambda \neq 0$. Furthermore, density (2.3) also appears by mixture (see the discussion of M. Cuadras about the work of Arnold and Beaver, 2002). To see this, note that if λ follows a symmetric distribution $\pi(\lambda)$, with $-\infty < \lambda < \infty$, then

$$\int_{-\infty}^{\infty} 2\Phi(\lambda\, \omega_\beta(x))\phi(\omega_\alpha(x))\pi(\lambda)\, \mathrm{d}\lambda = \phi(\omega_\alpha(x)).$$

Hereafter, we will write $X \sim \mathrm{GSN}(\alpha, \beta, \lambda)$ to denote that the pdf of the random variable X follows the pdf given in (2.13).

Generation of random variates from (2.13) is now easy via the following representation of the distribution. Let $X \sim \mathrm{BN}(\alpha)$ and $Z = X\, S_X$ where, conditionally on $X = x \neq 0$, we have

$$S_X = \begin{cases} +1 & \text{with probability } \Phi(\lambda\, \omega_\beta(x)), \\ -1 & \text{with probability } 1 - \Phi(\lambda\, \omega_\beta(x)). \end{cases}$$

Therefore, a random variate z from the random variable with density function given by (2.13) is derived as follows:

- Generate a random number u from the standard uniform distribution, $\mathrm{U}(0, 1)$.
- Generate random variate x from the distribution with pdf (2.3).
- Compute $\Phi(\lambda\, \omega_\beta(x))$.
- If $u < \Phi(\lambda\, \omega_\beta(x))$ then $z = x$; otherwise $z = -x$.

Then, the random variable Z has the density function given in (2.13). Figure 2 displays some plots of the pdf (2.13) for special values of the parameters.

It is straightforward to verify that properties (2.8) and (2.9) are satisfied also for the distribution (2.13). Some additional results of (2.13) are given below.

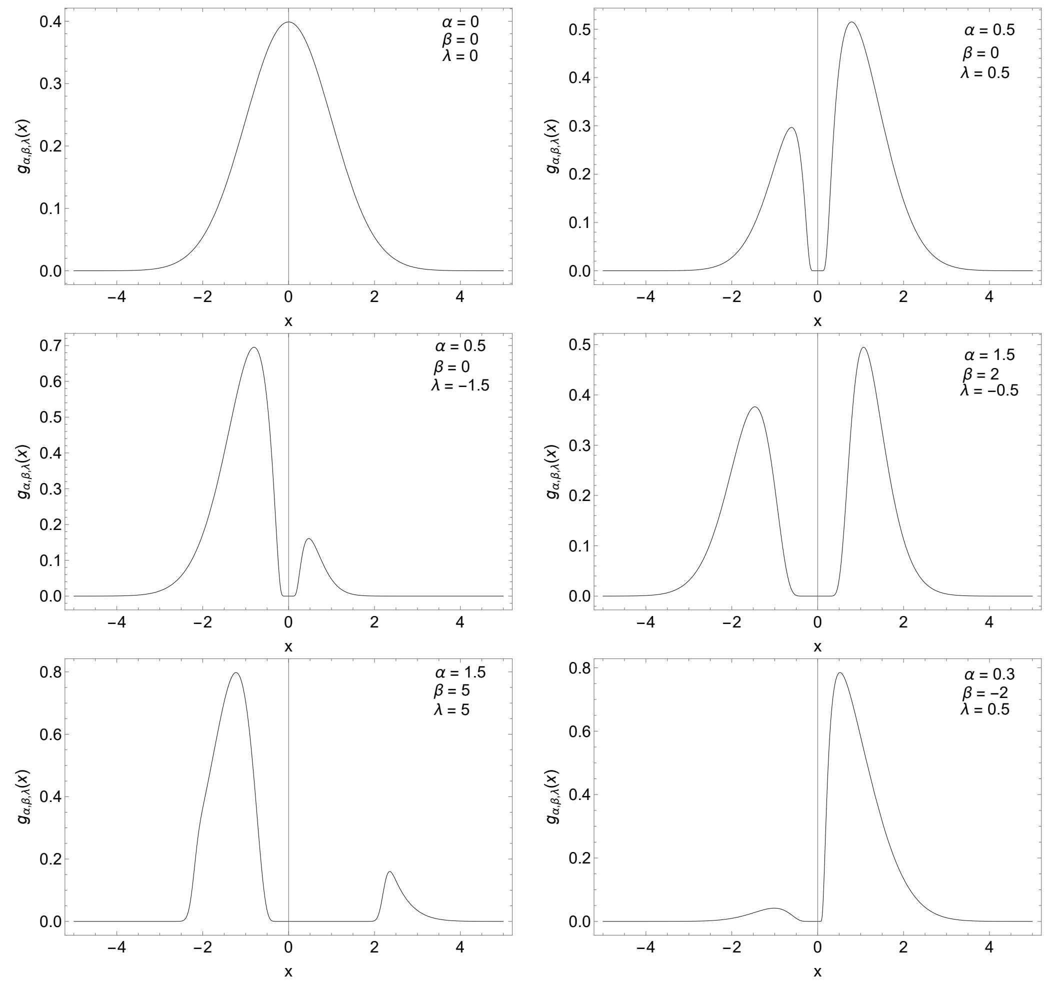

Figure 2: Plots of the pdf (2.13) for selected values of the parameters α, β and λ.

Proposition 2.5. *The following results are verified:*

(i) *If* $X \sim g_{\alpha,\beta,\lambda}(x)$ *then the random variable* $Z = -X \sim g_{\alpha,\beta,-\lambda}(z)$. *That is,* $g_{\alpha,\beta,\lambda}(-x) = g_{\alpha,\beta,-\lambda}(x)$ *for all* x.

(ii) *For all* $x \in \mathbb{R}$, *the cdf* $G_{\alpha,\beta,\lambda}(x) = \Pr(X \le x)$, *verifies:* $G_{\alpha,\beta,\lambda}(x) = G_{\alpha,\beta,-\lambda}(-x)$.

Proof: To see (i), observe that given $Z = -X$ we have that $|dz| = |dx|$. Now the result follows having into account that $\lambda\,\omega_\beta(-z) = \lambda(-z + \beta/z) = -\lambda(z - \beta/z) = -\lambda\,\omega_\beta(z)$ and $\phi(\omega_\alpha(-x)) = \phi(\omega_\alpha(x))$. Finally, (ii) follows from (i). □

Proposition 2.6. *As* $\lambda \to \infty$ *and* $\beta \to 0$ *the pdf given in* (2.13) *tends to* $g_\alpha(x) = 2\phi(\omega_\alpha(x))$, *i.e. a generalized half-normal density.*

Proof: It is derived as a result of writing (2.13) as

$$g_{\alpha,\beta,\lambda}(x) = 2\left(\int_{-\infty}^{\lambda\omega_\beta(x)} \phi(t)\,\mathrm{d}t\right)\phi(\omega_\alpha(x)),$$

and taking $\lambda \to \infty$. □

For $\lambda \to \infty$ and $\alpha \to 0^+$ the classical half-normal density is obtained.

If $X \sim \mathrm{GSN}(\alpha, \beta, \lambda)$ then its distribution function

$$G_{\alpha,\beta,\lambda}(x) = 2\int_{-\infty}^{x}\int_{-\infty}^{\lambda\omega_\beta(s)} \phi(t)\phi(\omega_\alpha(s))\,\mathrm{d}t\,\mathrm{d}s \tag{2.14}$$

can be represented as the cdf of a bivariate normal distribution. To see this take $\delta = \lambda/\sqrt{1+\lambda^2}$ and consider the change of variable

$$t = \frac{\eta + \delta\,\omega_\beta(s)}{\sqrt{1-\delta^2}}.$$

Then, some algebra provides that (2.14) can be rewritten as

$$G_{\alpha,\beta,\lambda}(x) = \frac{2}{\sqrt{1-\delta^2}}\int_{-\infty}^{x}\left(\int_{-\infty}^{0}\phi\left(\frac{\eta+\delta\,\omega_\beta(s)}{\sqrt{1-\delta^2}}\right)\mathrm{d}\eta\right)\phi(\omega_\alpha(s))\,\mathrm{d}s.$$

Unfortunately, we have not been able to find either the generating moment function or the ordinary moments of the distribution given in (2.13). Finally, by taking logarithm in (2.13), it is simple to verify that this pdf can have two modes which are the solutions of the equation

$$\lambda\left(1+\frac{\beta}{x^2}\right)\phi(\lambda\omega_\beta(x)) - \left(1+\frac{\alpha}{x^2}\right)\Phi(\lambda\omega_\beta(x))\,\omega_\alpha(x) = 0.$$

As most of the multimodal datasets considered in practice are defined on the positive real values, it is convenient to reparametrized the distribution given by (2.3) via a linear transformation, i.e. $Y = \mu + \sigma X$, where $X \sim g_\alpha(x)$, where $\alpha \geq 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$ given in (2.3) to obtain a more general family of densities. Its pdf is given by

$$g_{\alpha,\mu,\sigma}(x) = \begin{cases} \phi\left(\omega_\alpha\left(\dfrac{x-\mu}{\sigma}\right)\right), & x \neq \mu, \\ \phi_{\mu,\sigma}(\mu), & x = \mu. \end{cases} \tag{2.15}$$

For the sake of simplicity, we will consider the value $\mu = 0$ when estimating the parameters of the distribution, in that case the distribution coincides with (2.3). A value $x = 0$ is better identifiable in an empirical data source than another value that is unlikely to be an integer. For the case that $\mu = 0$, the parameter can be estimated by using a similar procedure as the one used in the composite models (see Calderín-Ojeda, 2015).

2.4. Extensions

A variant of the approach used to derived (2.13) can be simply implemented as follows:

$$g_{\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda}(x) = 2\Phi(\lambda\,\omega_{\beta_1,\beta_2}(x))\phi(\omega_{\alpha_1,\alpha_2}(x)) \tag{2.16}$$

for $x \neq 0$, $x \neq \sqrt{\beta_i}$, $x \neq \sqrt{\alpha_i}$, while $g_{\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda}(0) = \phi(0)$, $g_{\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda}(\sqrt{\alpha_i}) = \phi(\sqrt{\alpha_i})$, $g_{\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda}(\sqrt{\beta_i}) = \phi(\sqrt{\beta_i})$, where $\beta_i \in \mathbb{R}$, $\alpha_i \geq 0$ $(i = 1, 2)$ and

$$\omega_{\alpha_1,\alpha_2}(x) = x - \alpha_2 - \frac{\alpha_1}{x - \frac{\alpha_2}{x}},$$

$$\omega_{\beta_1,\beta_2}(x) = x - \beta_2 - \frac{\beta_1}{x - \frac{\beta_2}{x}}.$$

This modified family of distributions would allow us to obtain densities with more than two modal values. The extension of this distribution to generate multimodality is immediate. For the particular case (2.16), two graphs of the pdf have been plotted in Figure 3.

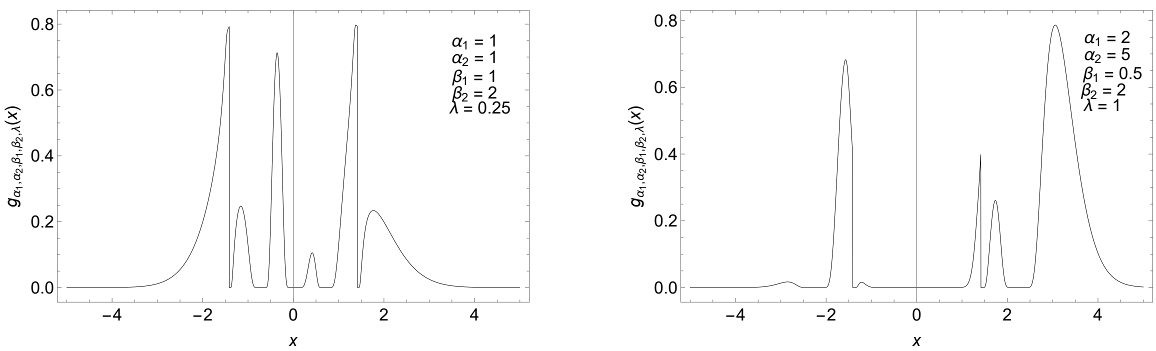

Figure 3: Plot of the probability density function (2.16) for selected values of the parameters α_i, β_i $(i = 1, 2)$ and λ.

This new multimodal family of probability distributions can be utilized to explain the size of the claims in cyber risk. In this regard, some multimodal and asymmetric distribution can be effortlessly applied to capture the multimodality and extremely skewed feature of the severity of the cyber breaches.

2.5. Summary of the proposed methodology

Before continuing with the usual elements of distribution theory, such as statistical inference and applications, it is essential to summarize the methodology we have carried out in this work in a diagram. Figure 4 shows a flowchart outlining the methods developed in this article. This diagram can help the reader observe the work's general perspective and allow, if desired, to ignore those elements that could be of lesser interest.

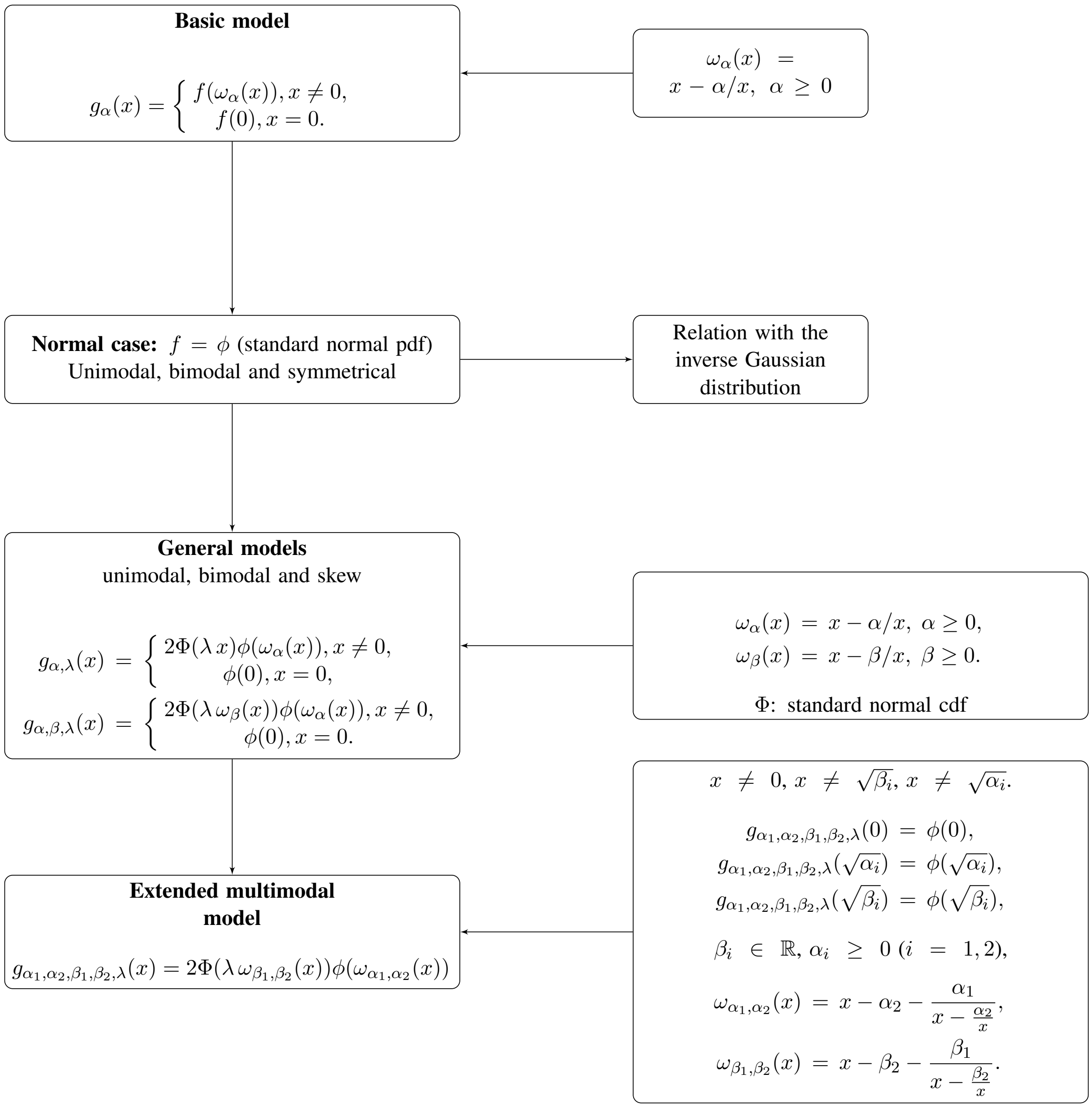

Figure 4: Flowchart showing the methodology proposed in this paper.

3. STATISTICAL INFERENCE

Let us consider a random sample of n observations $\boldsymbol{x} = (x_1, ..., x_n)$, in which there are n_0 observations that are zeros and n_1 non-zero observations; $n_0 + n_1 = n$. Now by using the pdf (2.3), the log-likelihood function is proportional to $\ell(\alpha; \boldsymbol{x}) \propto -1/2 \sum_{i \in \{1,...,n_1\}} (\omega_\alpha(x_i))^2$. By equating the first derivative with respect to α to zero, we get the maximum likelihood estimator of the parameter α is given by $\widehat{\alpha} = n_1 \left\{ \sum_{i \in \{1....,n_1\}} x_i^{-2} \right\}^{-1}$, $x_i \neq 0$. Now, by computing the second derivative of the log-likelihood function and its expectation, the corresponding standard error, that can be obtained from the Fisher's information entry, is $(n/\widehat{\alpha})^{-1/2}$. To obtain this result, it is necessary the expectation of $1/X^2$ with respect to the random variable with pdf (2.3) which is given by $1/\alpha$.

Let us now examine the pdf (2.15) with $\mu = 0$. In this case, the log-likelihood function is proportional to

$$\ell(\alpha, \sigma; \boldsymbol{x}) \propto -n \log \sigma - \frac{1}{2} \sum_{i \in \{1,\dots,n_1\}} (\omega_\alpha(x_i/\sigma))^2, \tag{3.1}$$

where n_1 is the number of non-zero observations in the sample. From (3.1) we derive the normal equations given by

$$\frac{n_1}{\sigma} - \alpha\sigma \sum_{i \in \{1,\dots,n_1\}} \left(\frac{1}{x_i}\right)^2 = 0, \tag{3.2}$$

$$\frac{n}{\sigma} - \sigma \sum_{i \in \{1,\dots,n_1\}} \left[\left(\frac{x_i}{\sigma^2}\right)^2 - \left(\frac{\alpha}{x_i}\right)^2\right] = 0. \tag{3.3}$$

After simple algebra, equations (3.2)–(3.3) provides the maximum likelihood estimators of the parameters which are given by

$$\begin{aligned}
\widehat{\alpha} &= \frac{n n_1}{\left(\sum_{i \in \{1,\dots,n_1\}} x_i^{-2}\right)\left(\sum_{i \in \{1,\dots,n_1\}} x_i^2\right) - n_1^2}, \\
\widehat{\sigma} &= \left\{\frac{1}{n}\left[\sum_{i \in \{1,\dots,n_1\}} x_i^2 - n_1^2 \left(\sum_{i \in \{1,\dots,n_1\}} x_i^{-2}\right)^{-1}\right]\right\}^{1/2}.
\end{aligned}$$

The second partial derivatives are provided by

$$\begin{aligned}
\frac{\partial \ell(\alpha, \sigma; \boldsymbol{x})}{\partial \alpha^2} &= -\sigma^2 \sum_{i \in \{1,\dots,n_1\}} \left(\frac{1}{x_i}\right)^2, \\
\frac{\partial \ell(\alpha, \sigma; \boldsymbol{x})}{\partial \alpha \partial \sigma} &= -2\alpha\sigma \sum_{i \in \{1,\dots,n_1\}} \left(\frac{1}{x_i}\right)^2, \\
\frac{\partial \ell(\alpha, \sigma; \boldsymbol{x})}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \sum_{i \in \{1,\dots,n_1\}} \left[\frac{3x_i^2}{\sigma^4} + \left(\frac{\alpha}{x_i}\right)^2\right].
\end{aligned}$$

Now, taking into account that $\mathbb{E}(X^2) = \sigma^2(1+\alpha)$ and $\mathbb{E}(1/X_i^2) = 1/(\alpha\,\sigma^2)$, it is a simple exercise to note that the Fisher's information matrix is

$$\mathcal{I}(\widehat{\alpha}, \widehat{\sigma}) = \begin{bmatrix} n_1/\widehat{\alpha} & 2n_1/\widehat{\sigma} \\ 2n_1/\widehat{\sigma} & (2n(2\widehat{\alpha}+1) - n_1)/\widehat{\sigma}^2 \end{bmatrix}.$$

Finally, when the pdf (2.13) is considered, the log-likelihood function is proportional to

$$\ell(\boldsymbol{\theta}; \boldsymbol{x}) \propto -n \log \sigma + \sum_{i \in \{1,\dots,n_1\}} \log \Phi(\lambda\,\omega_\beta(x_i/\sigma)) - \frac{1}{2} \sum_{i \in \{1,\dots,n_1\}} (\omega_\alpha(x_i/\sigma))^2, \tag{3.4}$$

where $\boldsymbol{\theta} = (\alpha, \beta, \lambda, \sigma)$ is the vector of parameters to be estimated.

In practice, although both normal equations and Fisher's information matrix can be obtained after tedious algebra, the estimates and the entries of this matrix can be achieved by directly maximizing the log-likelihood function given in (3.4). Moreover, this procedure can be extended, as it is seen in the numerical illustrations, for the case where a location parameter μ is included. Recall that the Fisher's information matrix of the skew-normal distribution proposed by Azzalini (1985) is singular for the skew parameter and, consequently, the maximum likelihood estimate of this parameter can be infinite with a positive probability. With respect to the singularity of the Fisher information matrix of the generalized skew normal (GSN) distribution with pdf (2.13), we could use the Theorem 3 in Rotnitzky *et al.* (2000) to derive a reparametrization of (2.13) and provide a solution to the singularity problem for (α, β, λ) as in Venegas *et al.* (2018). In order to show the asymptotic behaviour of the maximum likelihood estimator, we carry out the following simulation experiment where the algorithm illustrated in the previous section is used, a complete simulation analysis for the GSN distribution with density function (2.13) is carried out by generating $N := 1000$ samples of sizes $n := 50$, 100, 200 for different values of the parameters α, β and λ. The value of these parameters have been chosen for the sake of simplicity in estimation. For each parameter, the analysis computes the following measures:

- Average bias (AB) of the simulated estimates:

$$\mathrm{AB}(\Lambda^*) = \frac{1}{N} \sum_{j \in \{1,\dots,N\}} (\Lambda_j^* - \Lambda);$$

- Mean square error (MSE) of the simulated estimates:

$$\mathrm{MSE}(\Lambda^*) = \frac{1}{N} \sum_{j \in \{1,\dots,N\}} (\Lambda_j^* - \Lambda)^2;$$

where Λ_j^* represents the maximum likelihood estimate of each parameter in the j-th sample and Λ is the true value of the parameter. Table 1 shows the average bias and mean square errors of the parameter estimates for different values of α, β and λ for different values of n. In the first row of this table, the case of the skew parameter $\lambda = 0$ is considered, i.e. symmetric case. As expected, the mean square error decreases when n increases. Also, the average bias is positive and decreases with n. It is also noted that the MSE increases with the value of the parameter α. However, the mean square errors for the parameters β and λ seem to be influenced by the value considered for the parameter α. In general, the MSE's decrease with the sample size satisfying that $\lim_{n\to\infty} \mathrm{MSE}(\Lambda^*) = 0$, and therefore, the estimates are consistent in mean square error. It implies that the estimate gets closer and closer to the parameter's true value as data accumulates. Also, for large values of n, the maximum likelihood estimators are normally distributed with the mean equals to the true value of the parameter and variance equal to the reciprocal of the information function evaluated at the mean.

Table 1: Average bias (AB) and mean square error (MSE) of the maximum likelihood estimates for different values of the parameters of the GSN distribution for different samples sizes n with simulation size $N := 1000$.

n		$\alpha = 0.25$	$\beta = 0.5$	$\lambda = 0$	$\alpha = 1$	$\beta = 1$	$\lambda = 0$
50	AB	0.0015	—	—	0.0160	—	—
	MSE	0.0003	—	—	0.0224	—	—
100	AB	0.0013	—	—	0.0138	—	—
	MSE	0.0002	—	—	0.0108	—	—
200	AB	0.0000	—	—	0.0021	—	—
	MSE	0.0001	—	—	0.0049	—	—

n		$\alpha = 0.25$	$\beta = 0.5$	$\lambda = 0.5$	$\alpha = 1$	$\beta = 1$	$\lambda = 1$
50	AB	0.0008	0.0922	0.0552	0.0230	0.0057	0.0606
	MSE	0.0003	0.1302	0.1640	0.0211	0.0445	0.0756
100	AB	0.0001	0.1028	0.0472	0.0169	0.0072	0.0386
	MSE	0.0002	0.0795	0.1068	0.0105	0.0209	0.0419
200	AB	0.0001	0.0848	0.0336	0.0032	0.0075	0.0212
	MSE	0.0001	0.0606	0.0767	0.0054	0.0105	0.0209

n		$\alpha = 0.5$	$\beta = 0.25$	$\lambda = 0.25$	$\alpha = 0.75$	$\beta = 1.5$	$\lambda = 1.2$
50	AB	0.0040	−0.0224	−0.0024	0.0135	0.0322	0.0544
	MSE	0.0025	0.0470	0.0725	0.0092	0.0524	0.0890
100	AB	0.0030	−0.0181	0.0122	0.0054	0.0158	0.0320
	MSE	0.0012	0.0437	0.0407	0.0046	0.0284	0.0581
200	AB	0.0022	−0.0176	0.0234	−0.0003	0.0078	0.0276
	MSE	0.0007	0.0363	0.0219	0.0021	0.0143	0.0391

4. NUMERICAL ILLUSTRATIONS

In this section, some numerical applications of the GSN distribution given in (2.13) are carried out. The results are compared with those ones of the skew-normal distribution with parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda \in \mathbb{R}$, i.e. $\text{SN}(\mu, \sigma, \lambda)$.

The example considered uses the well-known old faithful geyser (Yellowstone Park, Wyoming, USA) data set. This data set consists of 299 measurements of the numerical eruption time in minutes and the waiting time to the next eruption (also in minutes). This popular dataset has been examined extensively in the literature. See, for example, Silverman (1986), Azzalini and Bowman (1990) and Dekking *et al.* (2005), among others. It is already known that these two datasets show bimodality. There are different versions of these datasets in the statistical literature. The one examined here is taken from the `R` package `MASS` available in the website

`https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/faithful.html`

Descriptive statistics of these two datasets are shown in Table 2.

Table 2: Descriptive statistics of the two variables considered in the Old Faithful dataset.

	Time eruption	Time waiting
Mean	3.461	72.314
Variance	1.313	192.296
min	0.833	43.000
max	5.450	108.000

The estimated values of the parameters for the two models are shown in Table 3 together with the standard errors (in brackets). This Table also includes the value of the maximum log-likelihood function ($\ell_{\max}$), the Akaike's information criterion (AIC) (see Akaike, 1974) and the consistent Akaike's information criteria (CAIC), proposed by Bozdogan (1987). The last measure of model selection was chosen to overcome the tendency of the AIC to overestimate the complexity of the underlying model since it lacks certain properties of asymptotic consistency as it does not directly depend on the sample size. Then, to calculate the CAIC, a correction factor based on the sample size is used to compensate for the overestimating nature of AIC. The CAIC is defined as twice $\ell_{\max}$ plus $k\,(1+\log(n))$, where k is the number of free parameters and n refers to the sample size. Note that a model with a lower AIC and CAIC values is preferred to one with a higher value. It is observable that the GSN distribution has a better performance than the skew normal (SN).

Table 3: Parameters estimates, standard errors (in brackets), maximum of the log-likelihood function ($\ell_{\max}$), AIC and CAIC values for the two variables considered in the old faithful geyser dataset.

	Time eruption		Time waiting	
	SN	GSN	SN	GSN
$\widehat{\lambda}$	10.310 (3.851)	0.676 (0.116)	−7.975 (1.512)	0.247 (0.078)
$\widehat{\alpha}$	— —	0.468 (0.058)	— —	0.551 (0.062)
$\widehat{\beta}$	— —	0.227 (0.096)	— —	−0.216 (0.334)
$\widehat{\mu}$	48.454 (0.944)	65.185 (0.258)	4.897 (0.049)	3.135 (0.009)
$\widehat{\sigma}$	27.597 (1.393)	13.088 (0.557)	1.837 (0.084)	0.956 (0.038)
$\ell_{\max}$ AIC CAIC	−1231.57 2469.13 2483.24	−1116.427 2242.85 2266.36	−425.737 857.474 871.575	−399.229 808.458 831.960

Graphs of the empirical smooth kernel density and theoretical distribution model (GSN) are shown in Figure 5. This former density function was derived by using the in-built function `SmoothKernelDistribution` in `Mathematica`® `v.12.0`. We used a smoothing Gaussian kernel and automatically computed bandwidth parameter. As it can be seen, the GSN is able to capture the bimodal nature of the empirical data although there is an underestimation produced by the adjustment of the proposed distributions. Maximization techniques were completed using `Mathematica`® `v.12.0` and corroborated with `WinRATS v.7.0` (the codes are available upon request) and the computer used was a Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz with 16,0 GB RAM and a processor based on x64 getting acceptable time of processing. Details about these two software can be found in Ruskeepaa (2009) and Brooks (2009), among others. The routines employed were standard, including among others the `FindMaximum` to compute the maximum likelihood estimates and the `Experimental‘CreateNumericalFunction` to obtain the Hessian matrix.

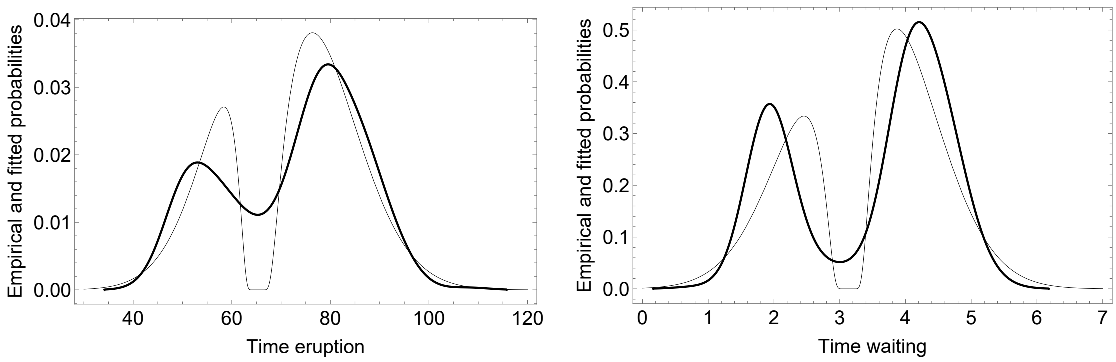

Figure 5: Smooth kernel density estimate of the empirical data (thick line) and the GSN (thin line) for the old faithful data set.

5. CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH

In this work, we have studied two families of distributions with support on the real line, the first symmetric and the second not necessarily symmetric. Both families can present more than one mode and include the normal distribution as a special case. In addition, the second one includes, as a particular case, the skew normal distribution. The model has been applied to environmental data, and it can also be used in other scenarios where bimodality is present.

One of the limitations of the distribution proposed in this work is based on the fact that the value that the first distribution takes at zero (at μ for the second model) is fixed, what make these models inflexible. This is an issue that that undoubtedly deserves to be deeply studied to guarantee a more versatile and flexible proposal than the ones presented in this work.

It should also be noted that the extension shown in the Subsection 2.4 requires a separate analysis outside this work's scope. This indeed constitutes a promising probabilistic family that allows to model multimodal data.

ACKNOWLEDGMENTS

EGD and ECO work were partially funded by grant PID2021-127989OB-I00 (Ministerio de Economía y Competitividad, Spain. Also ECO thanks Ministerio de Universidades and Next Generation EU for funding the Program "Ayudas María Zambrano". The authors thank the Associate Editor and three anonymous referees for their constructive comments and suggestions, which have greatly helped them to improve the paper.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Anderson, W., Pfeiffer, R., Dores, G., and Sherman, M. (2006). Bimodal breast cancer incidence patterns provide support for a dualistic model of mammary carcinogenesis. *Journal of Clinical Oncology*, 24(18):595–595.

Arellano-Valle, R., Gómez, H., and Quintana, F. (2004). A new class of skew-normal distributions. *Communications in Statistics: Theory and Methods*, 33(7):1465–1480.

Arellano-Valle, R., Gómez, H., and Quintana, F. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128(2):427–443.

Arnold, B. and Beaver, R. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, 11(1):1–54.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46:199–208.

Azzalini, A. (2013). *The Skew-Normal and Related Families*. Cambridge University Press, Cambridge.

Azzalini, A. and Bowman, A. (1990). A look at some data on the old faithful geyser. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(3):357–365.

Behboodian, J. (1978). Uncorrelated dependent random variables. *Mathematics Magazine*, 51(5): 303–304.

Bozdogan, H. (1987). The general theory and its analytical extension. *Psychometrika*, 52(3):345–370.

Brooks, C. (2009). *RATS Handbook to Accompany Introductory Econometrics for Finance*. Cambridge: Cambridge University Press.

Calderín-Ojeda, E. (2015). On the Composite Weibull-Burr Model to describe claim data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1(1):59–69.

Chiogna, M. (1998). Some results on the scalar skew-normal distribution. *J. Ital. Statist. Soc.*, 1:1–13.

Dekking, F., Kraaikamp, C., Lopuhaa, H., and Meester, L. (2005). *A Modern Introduction to Probability and Statistics Understanding Why and How*. Springer-Verlag, London.

García, V., Gómez-Déniz, E., and Vázquez-Polo, F. (2010). A new skew generalization of the normal distribution: properties and applications. *Computational Statistics and Data Analysis*, 54: 2021–2034.

Gómez, H., Venegas, O., and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, 18:395–407.

Gómez-Déniz, E., Arnold, B. C., Sarabia, J. M., and Gómez, W. H. (2021). Properties and applications of a new family of skew distributions. *Mathematics*, 9(87):1–18.

Gómez-Déniz, E., Sarabia, J. M., and Calderín-Ojeda, E. (2021). Bimodal normal distribution: extensions and applications. *Journal of Computational and Applied Mathematics* (in press).

Gupta, A., Nguyen, T., and Sanqui, J. (2004). Characterization of the skew-normal distribution. *Ann. Inst. Statist. Math.*, 56(2):351–360.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, 13(4):271–275.

Hirota, M., Holmgren, M., Nes, E. V., and Scheffer, M. (2011). Globa lresilience of tropical forest and savanna to critical transitions. *Science*, 334:232–235.

Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Wiley, New York.

Jones, M. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, 96(4):761–780.

Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution. Lecture Notes in Statistics*. Springer–Verlag, New York.

Malinovskii, V. K. (2017). Generalized inverse Gaussian distributions and the time of first level crossing. *arXiv:1708.08671v1.*

Patel, J. and Campbell, B. (1984). *Handbook of the Normal Distribution*. Marcel Dekker, Inc., New York and Basel.

Rotnitzky, A., Cox, D. R., Bottai, M., et al. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6:243–284.

Ribeiro-Reis, L. D., Cordeiro, G. M., and de Santana e Silvas, J. J. (2022). The Mc-Donald Chen distribution: A new bimodal distribution with properties and applications. *Chilean Journal of Statistics*, 13(1):91–111.

Ruskeepaa, H. (2009). *Mathematica Navigator: Mathematics, Statistics, and Graphics*. New York: Academic Press.

Sharafi, M. and Behboodian, J. (2008). The Balakrishnan skew-normal density. *Statistical Papers*, 49:769–778.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. London, UK: Chapman and Hall.

Slobin, H. (1927). A theorem on improper definite integrals. *The American Mathematical Monthly*, 34(5):265–266.

Sulewski, P. (2022). New members of the Johnson family of probability distributions: properties and application. *REVSTAT* (accepted).

Teimouri, M. and Nadarajah, S. (2016). On simulating Balakrishnan skew-normal variates. *Computational Statistics and Data Analysis*, 45(11):3310–3322.

Venegas, O., Salinas, H. S., Gallardo, D. I., Bolfarine, H., and Gómez, H. W. (2018). Bimodality based on the generalized skew-normal distribution. *Journal of Statistical Computation and Simulation*, 88(1):156–181.

REVSTAT – Statistical Journal
Volume 23, Number 2, April 2025, 273–290
https://doi.org/10.57805/revstat.v23i2.444

Performance Assessment of Sandwich and Block Bootstrap Estimators for Temporally Dependent Bivariate Extremes

Author: Alexander Engberg
– Department of Statistics, Uppsala University, Sweden
alexander.engberg@statistics.uu.se

Received: March 2023 Revised: June 2023 Accepted: June 2023

Abstract:

- Ignoring temporal dependence when modelling sequences of extreme observations yields underestimated standard errors which can lead to inaccurate risk assessment of extreme phenomena such as floods and economic crises. One remedy is to inflate standard errors with a sandwich or block bootstrap estimator. In this study, the performance of four such standard error estimators is investigated, through simulation, when modelling extremes from bivariate sequences. The results show that under strong temporal dependence, all considered estimators seriously underestimate standard errors, while under moderate to weak dependence both the sandwich and the bootstrap estimators can mitigate this underestimation.

Keywords:

- *block bootstrap; sandwich estimator; logistic model; Hüsler–Reiss model; bivariate extremes; temporal dependence.*

AMS Subject Classification:

- 62G32, 62M10.

1. INTRODUCTION

Extreme phenomena such as floods, droughts and economic crises cause devastating damage to our societies and tend to be driven by concatenated rare events in complex systems such as rivers or financial networks. Flooding in the lower part of a river system is often the result of simultaneous high flows further up in the system (Zheng *et al.*, 2013; Keef *et al.*, 2009). Severe droughts are the result of joint extremes of several meteorological variables (Zscheischler and Seneviratne, 2017). The risk of economic crises is highly dependent on the connections between different financial institutions (Zhou, 2010). In all these examples, the multivariate dependence between individual rare events determines the risk of an extreme phenomenon, and multivariate extreme-value theory provides a suitable framework for modelling such dependence. Standard extreme-value modelling approaches, however, rely on the assumption that data consists of independent and identically distributed (i.i.d.) random vectors, while in applications, temporal dependence in data is common. Failure to account for temporal dependence leads to underestimating the uncertainty of estimated quantities, such as the strength of multivariate dependence, and, in consequence, inaccurate risk assessment. Against this background, the present study is concerned with methods to account for the effect that temporal dependence in data has on the estimated parameter uncertainty when modelling bivariate dependence between rare events.

An often more realistic assumption than that of i.i.d. observations is that data constitutes a stationary sequence of random vectors. Under stationarity, temporal dependence can be explicitly modelled in conjunction with the multivariate model, or accounted for after model estimation by adjusting standard errors with block bootstrap estimators (Künsch, 1989; Liu and Singh, 1992; Politis and Romano, 1994) or sandwich covariance matrix estimators (Godambe, 1960; White, 1982; Smith, 1990). Explicit modelling might yield the most accurate results, but models can be application specific or intractable, and if the model fit is poor, results may be misleading. To adjust the standard errors with a block bootstrap or sandwich estimator is more robust and often simpler. However, block bootstraps are computationally expensive and sensitive to the choice of block length, and while sandwich estimators are computationally cheaper, they might be subject to numerical difficulties.

The goal of this study is to, through simulation, investigate and compare the performance of four standard error estimators when modelling the bivariate dependence between extreme values from bivariate stationary sequences. Focus is on the logistic extreme-value model due to its relative simplicity, which makes it a good model of reference. The results are, however, generalized both to higher dimensions and to additional models. The considered estimators are, first, the sandwich estimators of Godambe (1960) and Smith (1990), where the former is combined with the Newey–West estimator (Newey and West, 1987) to account for temporal dependence. Second, the moving block- and stationary bootstrap estimators of Künsch (1989) and Liu and Singh (1992), and Politis and Romano (1994), respectively. Focus is on how the uncertainty of the estimated dependence strength is affected by temporal dependence in data, and to what extent the considered methods can mitigate underestimation.

To the author's knowledge, no comparative performance assessment of sandwich and block bootstrap estimators has been conducted in the context of modelling extremes from bivariate (or multivariate) stationary sequences. The performance of some estimators have, however, been examined in specific settings. Fawcett and Walshaw (2007) show that the sandwich estimator of Smith (1990) yields notably less biased standard errors of parameter estimates when modelling univariate threshold exceedances from first order Markov chains; Northrop (2015) uses the sandwich estimator of White (1982) and the stationary bootstrap (Politis and Romano, 1994) to adjust standard errors of different extremal index estimators for univariate block-maxima showing that the bootstrap generally provides a good bias reduction; and Huser and Wadsworth (2019) show that the stationary bootstrap adequately captures temporal dependence in rare events when modelling extremes from spatial processes with a censored likelihood.

In Section 2, a brief overview of bivariate extreme-value theory is given. In Section 3, the block bootstrap and sandwich estimators are presented in more detail, and in Section 4, the performance of the estimators is assessed through simulation. The outcomes of the study are discussed in Section 5.

2. BIVARIATE EXTREMES

In this section, an overview of the bivariate extreme-value theory used in this study is given. For a more comprehensive account of multivariate extreme-value theory, see Davison and Huser (2015), Coles (2001, Ch. 8), and Beirlant *et al.* (2004, Ch. 8–9).

2.1. Block-maxima

Let $\{(X_i, Y_i)\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. bivariate random vectors with joint cumulative distribution function $F(x,y)$, and denote the vector of component-wise maxima $\mathbf{M}_n = (M_{x,n}, M_{y,n}) = \big(\max_{i=1,\dots,n}(X_i), \max_{i=1,\dots,n}(Y_i)\big)$. It should be noted that the maximum might not occur for the same index i in both X_i and Y_i which means that the joint maxima $\mathbf{M}_n$ might not correspond to an actual observation. Furthermore, assume that there exist sequences $(a_{x,n}, a_{y,n}) > 0$ and $(b_{x,n}, b_{y,n})$, and a non-degenerate distribution $G(x,y)$, such that the distribution of normalized maxima converges as

$$\lim_{n\to\infty} \Pr\left(\frac{M_{x,n}-b_{x,n}}{a_{x,n}} \le x,\ \frac{M_{y,n}-b_{y,n}}{a_{y,n}} \le y\right) = G(x,y), \tag{2.1}$$

where $G(x,y)$ has non-degenerate marginals $G_X(x)$ and $G_Y(y)$. Then, $G(x,y)$ is a bivariate extreme-value distribution, and the marginals belong to the generalized extreme-value (GEV) family of distributions, which has the distribution function

$$G_Z(z) = \begin{cases} \exp\left[-\left(1+\xi\left(\dfrac{z-\mu}{\tau}\right)\right)_+^{-1/\xi}\right], & \xi \neq 0, \\[2ex] \exp\left[-\exp\left(-\dfrac{z-\mu}{\tau}\right)\right], & \xi = 0, \end{cases} \tag{2.2}$$

defined on $\{z\colon 1+\xi(z-\mu)/\tau > 0\}$, and where $c_+ = \max(c,0)$, and $\mu\in\mathbb{R}$, $\tau\in\mathbb{R}_+$ and $\xi\in\mathbb{R}$ are location, scale and shape parameters which can be different for the respective marginals.

A customary assumption is that the univariate marginals of $\{(X_i, Y_i)\}_{i\in\mathbb{N}}$ follow the unit Fréchet distribution (corresponding to GEV$(\mu = \tau = \xi = 1)$). This assumption is not restrictive because the marginals can be transformed with the probability integral transform to approximately follow the unit Fréchet distribution once the parameters of the GEV distribution have been estimated. With unit Fréchet marginals the joint distribution function $G(x, y)$ can be expressed as

$$G(x, y) = \exp\bigl[-V(x, y)\bigr], \qquad x > 0, \quad y > 0, \tag{2.3}$$

(de Haan and Resnick, 1977). The function V can be written as

$$V(x, y) \;=\; 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right)\, \mathrm{d}H(w), \tag{2.4}$$

where H is a distribution function that determines the bivariate dependence structure and satisfies the mean constraint

$$\int_0^1 w\; \mathrm{d}H(w) \;=\; \int_0^1 (1-w)\; \mathrm{d}H(w) \;=\; 1/2, \tag{2.5}$$

(Coles, 2001; Ledford and Tawn, 1997).

In practice, vectors of component-wise maxima $\mathbf{M}_n$ are obtained by splitting the data sequence into large disjoint blocks and extracting the maxima from each block, such as annual maximum daily rainfall. For sufficiently large blocks the block-maxima can be viewed as approximate realizations from a GEV distribution. Maximum likelihood estimation is often used to estimate the model parameters (Prescott and Walden, 1980), and the marginal and dependence parameters can be estimated either separately or simultaneously. Simultaneous estimation has the benefit that information can be shared across marginals, although at a higher computational cost. The procedure of modelling maxima from large disjoint blocks of data is referred to as the *block-maxima* modelling approach.

2.2. Threshold exceedances

In many cases, rare events occur in clusters, such as consecutive days of heavy rainfall or high temperatures, which the analysis of block-maxima overlooks. A more efficient use of data is achieved with the *peaks over thresholds* (POT) approach, in which exceedances over some high thresholds, u_x and u_y, are studied. More specifically, for a random variable X with distribution function $F_X(x)$, define the right endpoint of the distribution as $x_F = \sup\{x\colon F_X(x) < 1\}$. Then, if the distribution of normalized block-maxima from $F_X(x)$ converge to a GEV distribution, the limiting distribution of suitably normalized conditional excesses $\bigl((X-u)/a(u) \mid X > u\bigr)$, $a(u) > 0$, as $u \nearrow x_F$, is a generalized Pareto (GP) distribution with distribution function

$$H(z) \;=\; \begin{cases} 1 - \left(1 + \xi\, \dfrac{z}{\tilde{\tau}}\right)_+^{-1/\xi}, & \xi \neq 0, \\[2ex] 1 - \exp\left(-\dfrac{z}{\tilde{\tau}}\right), & \xi = 0, \end{cases} \tag{2.6}$$

defined on $\{z\colon z > 0 \text{ and } (1 + \xi z/\tilde{\tau}) > 0\}$, with parameters $\tilde{\tau} \in \mathbb{R}_+$ and $\xi \in \mathbb{R}$, and where $\tilde{\tau} = \tau + \xi u$ (Balkema and de Haan, 1974; Pickands, 1975). Thus, for high thresholds, a GP likelihood provides an appropriate model for threshold excesses.

In practice, for bivariate sequences $\{(X_i, Y_i)\}_{i\in\mathbb{N}}$, GP likelihoods are fitted to each marginal, and the marginal distributions are transformed to unit Fréchet scale. To model the bivariate dependence structure it can be shown that with unit Fréchet marginals, if the thresholds are sufficiently high and $(X_i \geq u_x,\ Y_i \geq u_y)$, the limiting joint distribution of conditional excesses is approximately the multivariate extreme-value distribution in (2.3) (Beirlant *et al.*, 2004, p. 276). This fact was used by Ledford Tawn (1996) as the basis for a censored likelihood for threshold exceedances in which contributions of observations are censored from below at the thresholds. To specify the censored likelihood in the bivariate case, let λ_x and λ_y be some small probabilities and set $r_x = -1/\ln(1-\lambda_x)$ and $r_y = -1/\ln(1-\lambda_y)$ which corresponds to marginal thresholds transformed to unit Fréchet scale. Then the bivariate censored likelihood can be expressed as

$$L(\theta; x, y) = \begin{cases} \dfrac{\partial^2}{\partial x\,\partial y}\exp\bigl[-V(x,y)\bigr], & \text{if } x > r_x,\ y > r_y, \\ \dfrac{\partial}{\partial x}\exp\bigl[-V(x,r_y)\bigr], & \text{if } x > r_x,\ y \leq r_y, \\ \dfrac{\partial}{\partial y}\exp\bigl[-V(r_x,y)\bigr], & \text{if } x \leq r_x,\ y > r_y, \\ \exp\bigl[-V(r_x,r_y)\bigr], & \text{if } x \leq r_x,\ y \leq r_y. \end{cases} \tag{2.7}$$

2.3. Models

To form tractable models from the distribution in (2.3), a common approach is to approximate the family of distributions with some parametric sub-family. A frequently used alternative is the logistic extreme-value model (Gumbel, 1961) which has the exponent function

$$V(x,y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}, \qquad x > 0,\ y > 0. \tag{2.8}$$

The parameter $0 < \alpha \leq 1$ governs the dependence with $\alpha = 1$ corresponding to independence and the limiting case $\alpha \to 0$ to complete dependence. The logistic model is relatively easy to work with and is therefore chosen as a model of reference in this study. It is also related to more flexible models, such as the asymmetric logistic model presented below.

To assess the generality of the results, the asymmetric logistic (Tawn, 1988) and the Hüsler–Reiss (Hüsler and Reiss, 1989) models are also considered. Assuming unit Fréchet marginals, the asymmetric logistic model has the exponent function

$$V(x,y) = \frac{1-\psi_1}{x} + \frac{1-\psi_2}{y} + \left\{\left(\frac{\psi_1}{x}\right)^{1/\beta} + \left(\frac{\psi_2}{y}\right)^{1/\beta}\right\}^{\beta}, \qquad x > 0,\ y > 0. \tag{2.9}$$

Here, dependence is governed by the parameter $0 < \beta \leq 1$, and the asymmetry parameters $0 \leq \psi_1, \psi_2 \leq 1$. As such, this model has two additional parameters compared to the (symmetric) logistic model and, including the marginals, it has a total of 9 parameters to be estimated for block-maxima, and 7 for POT. Independence can be achieved when either $\beta = 1$ or $\psi_1 = 0$ or $\psi_2 = 0$, or when $\psi_1 = \psi_2 = \beta = 1$, while complete dependence is obtained in the limit when $\psi_1 = \psi_2 = 1$ and $\beta \to 0$. The exponent function of the Hüsler–Reiss model has the form

$$V(x,y) \;=\; \frac{1}{x}\,\Phi\!\left[\frac{1}{r} + \frac{r}{2}\,\ln\!\left(\frac{y}{x}\right)\right] + \frac{1}{y}\,\Phi\!\left[\frac{1}{r} + \frac{r}{2}\,\ln\!\left(\frac{x}{y}\right)\right], \qquad x>0,\;\; y>0, \tag{2.10}$$

where $\Phi(\cdot)$ is the standard normal distribution function. The dependence is governed by $r > 0$, where independence is obtained in the limit as $r \to 0$ and complete dependence corresponds to $r \to \infty$.

3. EXTREMES OF DEPENDENT SEQUENCES

In this section, the modelling of extreme values from temporally dependent sequences is discussed, and the considered sandwich and block bootstrap estimators are described.

3.1. Asymptotic dependence

When analysing extreme values from bivariate (or multivariate) random vectors, the primary interest is often to describe the dependence in the tail, referred to as asymptotic dependence. As standard association metrics often perform poorly in the tail, asymptotic dependence can better be characterized by the tail dependence coefficient χ (Coles *et al.*, 1999). For two random variables X and Y with continuous distributions F_X and F_Y, the tail dependence coefficient is defined as

$$\chi \;=\; \lim_{u\to 1} P\big(F_X(X) > u \mid F_Y(Y) > u\big). \tag{3.1}$$

The variables X and Y are considered asymptotically dependent if $\chi > 0$ and asymptotically independent if $\chi = 0$.

3.2. Modelling of extremes from dependent sequences

Since the block-maxima and POT approaches in their simplest form rely on the assumption of i.i.d. observations, applying them to temporally dependent stationary sequences results in model misspecification. A common practice is therefore to remove, or at least diminish, the dependence before model estimation. For block-maxima, provided that long-range dependence in the sequence is weak at extreme levels, maxima are often close to independent and the limiting distribution still belongs to the GEV family (Leadbetter, 1974). Thus, applying the block-maxima approach to maxima from stationary sequences yields valid parameter point estimates, although different than if data had been independent (Coles, 2001, p. 96). However, if some temporal dependence remains, standard errors associated with the parameter estimates are underestimated.

In the POT approach, exceedances from stationary sequences tend to form clusters of dependent extremes above the thresholds. In consequence, the resulting sequence of exceedances constitutes a series of concatenated clusters, with temporal dependence retained within the clusters. Perhaps the most common remedy is declustering, which entails identifying clusters of extremes and only using the maximum from each cluster in the analysis (Davison and Smith, 1990). However, Fawcett and Walshaw (2007) suggest that declustering might induce serious bias in estimates, and showed that fitting GP likelihoods to all exceedances from first order Markov chains yields close to unbiased parameter estimates but underestimated standard errors. A better approach than declustering might therefore be to fit the model to all exceedances and to account for temporal dependence afterwards.

The underestimated standard errors can be inflated with sandwich or block bootstrap estimators to yield more correct uncertainty measurements. In Sections 3.3 and 3.4, the four standard error estimators considered in this study are presented in more detail.

3.3. Sandwich estimators

Under model misspecification, and given regularity conditions, the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, of the vector of model parameters, $\boldsymbol{\theta}$, converges as

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{d} N\big(\mathbf{0}, \tilde{\mathrm{I}}(\boldsymbol{\theta})^{-1}\big) \qquad \text{as } n\to\infty, \tag{3.2}$$

with $\tilde{\mathrm{I}}(\boldsymbol{\theta})^{-1} = \mathrm{H}(\boldsymbol{\theta})^{-1}\mathrm{J}(\boldsymbol{\theta})\,\mathrm{H}(\boldsymbol{\theta})^{-1}$. Here, $\tilde{\mathrm{I}}(\boldsymbol{\theta})^{-1}$ is the sandwich covariance matrix, $\mathrm{H}(\boldsymbol{\theta}) = -\mathbb{E}\big[\nabla^2\ell(\boldsymbol{\theta};x,y)\big]$ is the Fisher information matrix and $\mathrm{J}(\boldsymbol{\theta}) = \mathbb{V}\big[\nabla\ell(\boldsymbol{\theta};x,y)\big]$ is the variance of the score vector, with $\ell(\boldsymbol{\theta};x,y)$ being the model log-likelihood, and ∇ and ∇^2 the first and second order gradients (Davison, 2003, p. 147). The negative Hessian matrix

$$-\mathsf{H}_{\hat{\boldsymbol{\theta}}} = -\frac{\partial^2\,\ell(\boldsymbol{\theta};x,y)}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{3.3}$$

provides an estimate of $\mathrm{H}(\boldsymbol{\theta})$, and to estimate $\mathrm{J}(\boldsymbol{\theta})$, while also taking temporal dependence into account, the covariance matrix of Newey and West (1987) is used which has the form

$$\mathsf{J}_{\hat{\boldsymbol{\theta}}} = \mathsf{J}_{\hat{\boldsymbol{\theta}},0} + \sum_{j=1}^{m} w_{j,m}\big(\mathsf{J}_{\hat{\boldsymbol{\theta}},j} + \mathsf{J}_{\hat{\boldsymbol{\theta}},j}^{\mathsf{T}}\big), \tag{3.4}$$

with

$$\mathsf{J}_{\hat{\boldsymbol{\theta}},j} = \sum_{i=1}^{n-j} \nabla\ell(\hat{\boldsymbol{\theta}};x_i,y_i)\,\nabla\ell(\hat{\boldsymbol{\theta}};x_{i+j},y_{i+j})^{\mathsf{T}} \quad \text{and} \quad w_{j,m} = 1-\left(\frac{j}{m+1}\right). \tag{3.5}$$

Here $\mathsf{J}_{\hat{\boldsymbol{\theta}},j}$ is the heavy tailed sample autocovariance (Feigin and Resnick, 1999) at lag j, and $w_{j,m}$ are weights that decline as j increases. The first term on the right hand side of (3.4) is the non-adjusted variance of the score vector, while the second term accounts for temporal dependence up to lag m. An optimal choice of m is determined by the dependence strength and the sequence length. In this study, however, m is chosen based on the actual asymptotic dependence in the data, by studying plots of $\hat{\chi}(u)$ for a range of lags. The "standard" sandwich covariance matrix estimator is then

$$\hat{\tilde{\mathrm{I}}}(\boldsymbol{\theta})^{-1} = (-\mathsf{H}_{\hat{\boldsymbol{\theta}}})^{-1}\,\mathsf{J}_{\hat{\boldsymbol{\theta}}}(-\mathsf{H}_{\hat{\boldsymbol{\theta}}})^{-1}. \tag{3.6}$$

If data consists of blocks of observations, such as annual summer temperatures, with dependence within, but near independence between the blocks, Smith (1990) proposed a blocked sandwich estimator in which the sequence of n observations is split into K approximately independent blocks. The log-likelihood is divided accordingly as

$$\ell(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}) = \sum_{k=1}^{K} h_k(\boldsymbol{\theta}; \mathbf{x}_k, \mathbf{y}_k), \tag{3.7}$$

where $h_k(\boldsymbol{\theta}; \mathbf{x}_k, \mathbf{y}_k)$, $k = 1, ..., K$, is the contribution to the log-likelihood from the k-th block. The score vector can be expressed as

$$\nabla\ell(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}) = \sum_{k=1}^{K} \nabla h_k(\boldsymbol{\theta}; \mathbf{x}_k, \mathbf{y}_k), \tag{3.8}$$

and score covariance estimator

$$\mathsf{J}_{\hat{\boldsymbol{\theta}},\text{Block}} = \sum_{k=1}^{K} \nabla h_k(\hat{\boldsymbol{\theta}}; \mathbf{x}_k, \mathbf{y}_k)\, \nabla h_k(\hat{\boldsymbol{\theta}}; \mathbf{x}_k, \mathbf{y}_k)^{\mathsf{T}}. \tag{3.9}$$

The blocked sandwich estimator is formed by replacing $\mathsf{J}_{\hat{\boldsymbol{\theta}}}$ with $\mathsf{J}_{\hat{\boldsymbol{\theta}},\text{Block}}$ in (3.6). This estimator was developed for spatial data with dependence in space but near independence in time, but Fawcett and Walshaw (2007) showed that when modelling first order Markov chains with GP likelihoods, the blocked sandwich estimator provided a notable improvement compared to non-adjusted standard errors, even when the blocks were dependent.

3.4. Block bootstrap estimators

The bootstrap of Efron (1979) provides a method to approximate the sampling distribution of a statistic through repeated random sampling of individual observations with replacement. In its original form, however, the method relies on the assumption that data are i.i.d. and fails to reproduce dependence in temporally dependent sequences (Singh, 1981). Block bootstrap methods, on the other hand, sample blocks of length $l \in \{1, 2, ..., n-1\}$ of consecutive observations which retain some temporal dependence in data and, provided that

$$l \to \infty \quad \text{and} \quad n^{-1} l \to 0 \quad \text{as} \quad n \to \infty, \tag{3.10}$$

they produce asymptotically valid approximations of the underlying sequence for weakly dependent sequences. In this study, the moving block bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992), and the stationary bootstrap (SB) of Politis and Romano (1994) are considered, and descriptions of the methods are provided below. For a more comprehensive presentation of block bootstrap see e.g. Lahiri (2003) and Kreiss and Lahiri (2012).

To describe the MBB method, first, let $X_1, ..., X_n$ denote observations from a stationary sequence. To generate an MBB sample, $X_1^*, ..., X_n^*$, start by determining a block length l and define the set of overlapping blocks $\{\mathrm{B}_1, ..., \mathrm{B}_{n-l+1}\}$, where $\mathrm{B}_i = (X_i, ..., X_{i+l-1})$. Then draw $b = \min(k \geq 1\colon\ kl \geq n)$ blocks with replacement from $\{\mathrm{B}_1, ..., \mathrm{B}_{n-l+1}\}$, and concatenate the blocks. The first n observations form the MBB sample.

The SB method is similar to the MBB, but the block lengths are random, drawn from a Geometric(p) distribution with $p = 1/l \in (0, 1]$ such that $l = 1/p$ is the mean, rather than the actual, block length. To form an SB sample, first draw a block length L_1 from the Geometric(p) distribution. Second, draw a block, B_{L_1}, of L_1 consecutive observations randomly from the original sequence. Repeat the procedure until $L_1 + \cdots + L_b \geq n$ and concatenate the $B_{L_1}, ..., B_{L_b}$ blocks. The first n observations in the sequence form the SB sample. As the name suggests, samples generated by the SB are stationary, in contrast to those generated by the MBB. However, in terms of bias and variance, Lahiri (1999) showed that MBB and SB variance estimators share asymptotic bias, but that the asymptotic variance of the SB estimator is considerably larger than that of the MBB.

To compute a block bootstrap standard error associated with an estimate of the dependence parameter θ of one of the considered extreme-value models, first generate a bootstrap sample $\{(X_i^*, Y_i^*)\}_{i=\mathbb{N}}$ with the MBB or SB. Second, fit GEV or GP likelihoods to the marginals, and a bivariate extreme-value likelihood to the joint sample to obtain a bootstrap estimate $\hat{\theta}^*$. Repeat the procedure N times and estimate the standard error as

$$\hat{\sigma}_{\hat{\theta}} = \left(\frac{1}{N-1} \sum_{j=1}^{N} (\hat{\theta}_j^* - \bar{\theta}^*)^2 \right)^{1/2}, \tag{3.11}$$

where $\bar{\theta}^*$ is the average of the N bootstrap estimates $\hat{\theta}_j^*$.

The bias and variance of block bootstrap estimators are largely affected by the choice of block length; longer blocks yield a better approximation of the temporal dependence and thus lower bias, but also fewer blocks to sample from and thereby higher variance (Davison and Hinkley, 1997, p. 397). An investigation of block length estimators is outside the scope of this study. However, it is investigated through simulation how the choice of block length affects the standard error estimates.

4. SIMULATION STUDY

In this section, the implementation of, and the results from, the simulation study are presented. All numerical experiments are performed in `R` (R Core Team, 2023).

4.1. Data generating process

The considered data generating process is a bivariate process with first order temporal dependence imposed by a logistic extreme-value copula with parameter $0 < \varphi \leq 1$, and bivariate dependence imposed by one of the extreme-value models in Section 2.3. To achieve this, first, a bivariate sequence $\{(X_i, Y_i)\}_{i\in\mathbb{N}}$ of independent vectors is generated with algorithms from Stephenson (2003) which are implemented in the `R` package `evd` (Stephenson, 2002). Second, temporal dependence is imposed on each marginal separately by inverse transform sampling of x_{t+1} from the conditional distribution of x_{t+1} given x_t and φ, i.e.

$$G_{X_{t+1}\,|\,X_t=x_t}\big(x_{t+1} \mid \varphi\big) = \Pr\big(X_{t+1} \leq x_{t+1} \mid X_t = x_t,\, \varphi\big), \tag{4.1}$$

where G is the logistic extreme-value distribution formed by (2.8). As such, the joint distribution function of two values (x_t, x_{t+1}) from the marginal $\{X_i\}$ is

$$F_{X_t,X_{t+1}}(x_t, x_{t+1}; \varphi) = \exp\left[-\left(x_t^{-1/\varphi} + x_{t+1}^{-1/\varphi}\right)^{\varphi}\right], \tag{4.2}$$

and correspondingly for $\{Y_i\}$. For simplicity, the same temporal dependence strength is used for both marginals. The marginal distributions of $\{(X_i, Y_i)\}$ are then transformed with the probability integral transform to GEV$(\mu=0,\ \tau=1,\ \xi=0.1)$ for block-maxima, and GP$(\tau=1,\ \xi=0.1)$ for POT. To investigate if the heaviness of the marginal tails have an effect on the standard error estimator's performance, parts of the simulation were replicated with marginal distributions GEV$(\mu=0,\ \tau=1,\ \xi=-0.1)$ for block-maxima, and GP$(\tau=1,\ \xi=-0.1)$ for POT. Changing the value of ξ from 0.1 to -0.1, however, had a negligible effect on the results and is therefore not further discussed.

To assess the strength of temporal dependence in the generated data, the tail dependence coefficient χ in (3.1) is estimated for $\{X_i\}$ and $\{Y_i\}$, respectively. For the POT approach, χ is estimated only from the exceedances as those are the only actually informative observations. The empirical estimator

$$\hat{\chi}(u) = \frac{\sum_{i=k+1}^{n} I\big(Z_i > \hat{F}_Z^{-1}(u)\big)\, I\big(Z_{i-k} > \hat{F}_Z^{-1}(u)\big)}{\sum_{i=k+1}^{n} I\big(Z_i > \hat{F}_Z^{-1}(u)\big)}, \tag{4.3}$$

is used, where I is the indicator function and $\hat{F}^{-1}$ is the inverse empirical distribution function (Coles, 2001, p. 165). Estimates are computed with $u = 0.95$ at lags $k = 1, ..., 5$, and in Figure 1 the averages over 1000 estimates from the logistic model for different temporal dependence strengths are presented. Corresponding estimates from the asymmetric logistic and Hüsler–Reiss models look similar. When temporal dependence is weak ($\varphi = 0.9$), the estimated asymptotic dependence is close to 0 already at lag 2, while under strong temporal dependence ($\varphi = 0.3$) there is still notable dependence in the tail at lag 5.

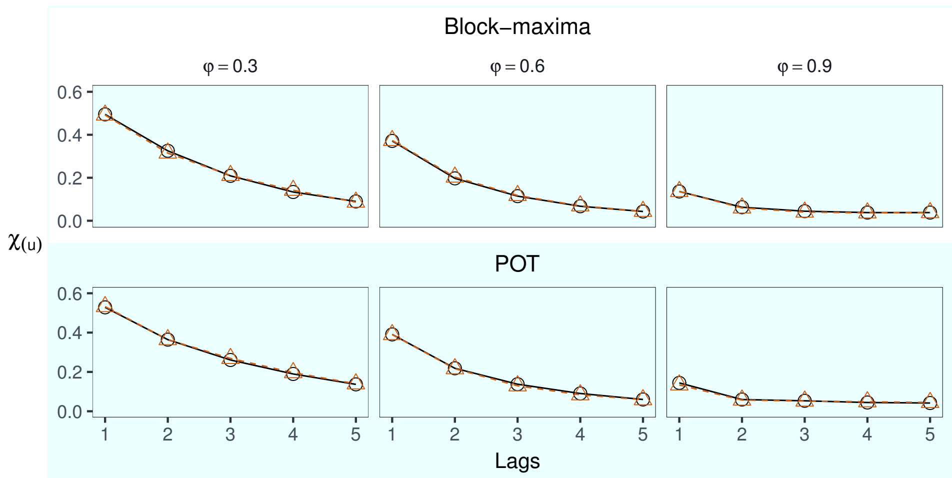

Figure 1: Average estimates of χ under different levels of temporal dependence for the marginals $\{X_i\}$ (solid, circles) and $\{Y_i\}$ (dashed, triangles) at lags 1 to 5, computed from 1000 simulated data sets from the logistic model.

4.2. Estimation

To compute the parameter estimates and associated standard errors, a bivariate extreme-value model with GEV (block-maxima) or GP (POT) marginals is fitted with maximum likelihood estimation to each generated data sequence. The model fitting is done in one step, estimating marginal and dependence parameters simultaneously. Once the model is fitted, standard errors are computed with each method described in Sections 3.3 and 3.4. Furthermore, 95% confidence intervals are computed by using either the normal distribution approximation of the maximum likelihood estimator for the sandwich estimators or basic bootstrap confidence intervals for the block bootstrap estimators (Davison and Hinkley, 1997, p. 194). For comparison, confidence intervals are also computed with the non-adjusted "naive" standard errors

$$\hat{\sigma}_{\text{Naive}} = -\frac{\partial^2 \ell(\theta; x, y)}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}}, \tag{4.4}$$

where $\ell(\theta; x, y)$ is the log-likelihood of the considered model with bivariate dependence parameter θ. As a benchmark, the sample standard deviation of the maximum likelihood estimator (MLE) is used

$$\sigma_R = \left(\frac{1}{R-1} \sum_{r=1}^{R} \left(\hat{\theta}_r - \bar{\theta}\right)^2 \right)^{1/2}. \tag{4.5}$$

Here $\hat{\theta}_r$ is the parameter estimate from the r-th data sequence and $\bar{\theta}$ is the average over R estimates.

4.3. Main simulation: logistic model

This section presents the main simulation study performed with the logistic extreme-value model with bivariate dependence parameter $0 < \alpha \leq 1$. Complementary results from the asymmetric logistic and the Hüsler–Reiss models are presented in Section 4.4.

To investigate how the standard error estimators perform under different dependence strengths, data sets are generated from the bivariate logistic model with dependence parameter $\alpha \in \{0.3, 0.7\}$ (strong, moderate) and temporal dependence parameter $\varphi \in \{0.1, 0.2, ..., 0.9\}$ (strong to weak). In each scenario $R = 1000$ independent sequences of length $n = 100$ are generated for block-maxima, and $n = 2000$ for POT. In the POT approach, the thresholds are set to the 0.95 quantiles such that there are 100 exceedances in each marginal. The block length for the blocked sandwich estimator is specified such that the data sequences are divided into 10 equally sized blocks, i.e. of size 10 for block-maxima and 200 for POT. Different block lengths were considered but with small effects on the results and, as such, the chosen lengths are deemed to be adequate.

To assess the effect that the choice of bootstrap block length has on estimation, root mean squared error (RMSE) is computed for the bootstrap estimators for each block length $l \in \{2, 3, ..., 12\}$ for block-maxima, and $l \in \{2, 4, ..., 24\}$ for POT, with 300 bootstrap replicates. This is done for strong and weak temporal dependence, and the difference in block lengths between block-maxima and POT is due to the different lengths of the underlying sequences (100 and 2000). The results are presented in Figure 2 together with associated 95% bootstrap percentile confidence intervals. Under weak dependence, the shortest possible block length ($l_{\mathrm{MBB}} = l_{\mathrm{SB}} = 2$) yields the lowest RMSE for both estimators. This is unsurprising as dependence is weak and close to 0 at lag 2 (see Figure 1). Under strong dependence RMSE decreases with block length and has a minimum at $l_{\mathrm{MBB}} = 10$ and $l_{\mathrm{SB}} = 9$ for block-maxima, and $l_{\mathrm{MBB}} = 24$ and $l_{\mathrm{SB}} = 18$ for POT. Since the lowest RMSE for the moving block bootstrap was achieved with $l_{\mathrm{MBB}} = 24$, longer block lengths were investigated but with no notable improvements of the results.

In Section 4.3.1, block bootstrap results are computed with the block lengths that yield the lowest RMSE, henceforth referred to as the "minimum RMSE block lengths".

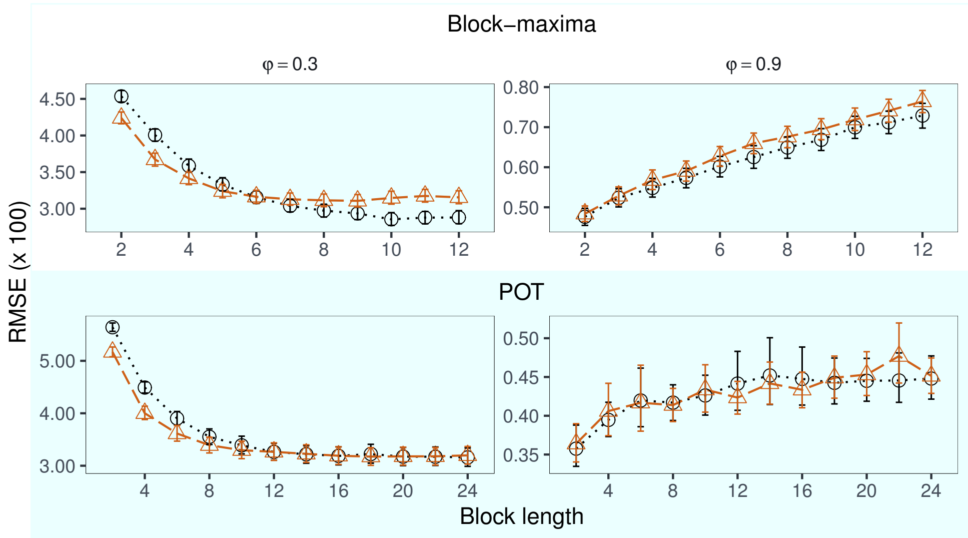

Figure 2: RMSE with 95% bootstrap percentile confidence intervals under strong (left panel) and weak (right panel) dependence for the moving block bootstrap (dotted, circles) and the stationary bootstrap (long-dash, triangles) estimators, computed from 1000 simulated data sets with 300 bootstrap replicates.

4.3.1. Results for the bivariate logistic model

In this section, results from the main simulation with the bivariate logistic model are presented. Only the results from the scenarios with strong bivariate dependence ($\alpha = 0.3$) are shown, as these summarise the general outcomes of the study. In the following, "dependence" refers to temporal dependence unless stated otherwise.

Figure 3 presents coverage probabilities of 95% confidence intervals for the bivariate logistic dependence parameter α, computed from data sets with dependence $\varphi \in \{0.1, 0.2, \dots, 0.9\}$. Here, the effect of ignoring temporal dependence is clearly seen as the naive estimator has poor coverage probabilities under strong dependence. Noteworthy is that the sandwich and block bootstrap estimators also have coverage probabilities far from 0.95 under strong dependence; although they still provide a clear improvement compared to the naive estimator.

As expected, when dependence weakens all estimators approach the nominal coverage probability. The standard sandwich estimator slightly outperforms the other estimators for the POT approach, but overall the differences are quite small.

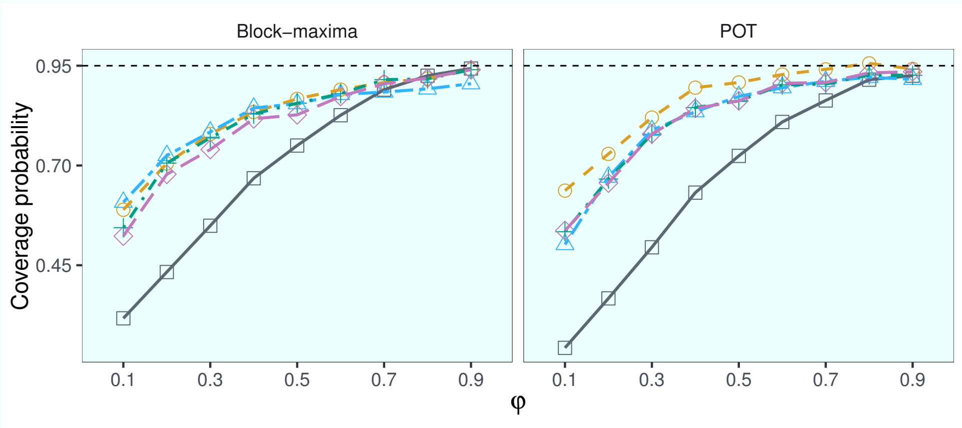

Figure 3: Coverage probabilities of 95% confidence intervals from the bivariate logistic model under strong ($\alpha = 0.3$) bivariate dependence for the naive (solid, squares), standard sandwich (dashed, circles), blocked sandwich (two-dash, triangles), MBB (dash-dotted, plus) and SB (long-dash, diamonds) estimators, computed from 1000 simulated data sets.

In Figure 4, box-plots of $\hat{\sigma}_r/\sigma_R$ are illustrated, where $\hat{\sigma}_r$ is the estimate from one replicate and σ_R is the benchmark in (4.5). This gives a view of the sampling distributions of the estimators in relation to the benchmark σ_R. Under strong dependence, the estimates are overall negatively biased. The block bootstraps perform slightly better than the standard sandwich estimator, while the blocked sandwich estimates are almost as biased as the naive ones for POT. Under weak dependence, the distributions of all estimates are more or less centred around the benchmark with similar variability.

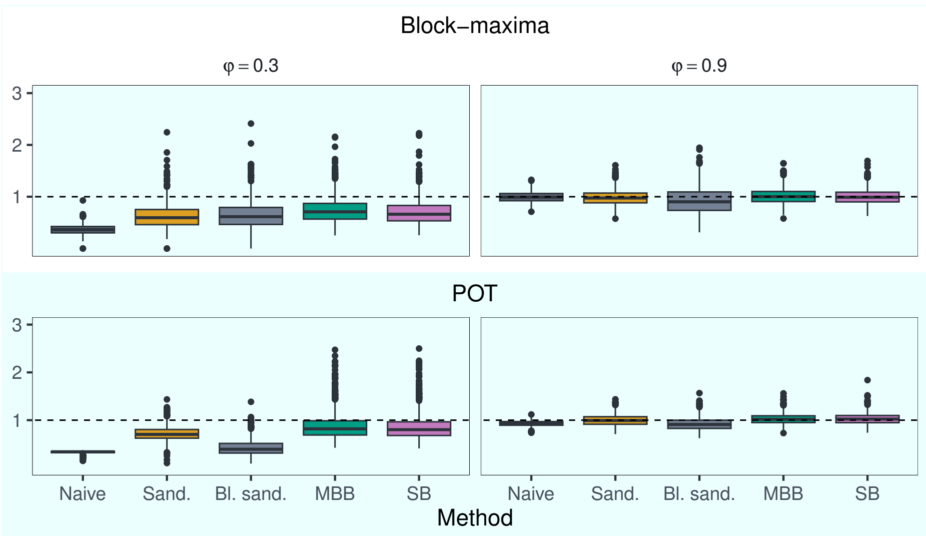

Figure 4: Distributions of $\hat{\sigma}_r/\sigma_R$ for strong (left panel) and weak (right panel) temporal dependence, under strong ($\alpha = 0.3$) bivariate dependence computed from 1000 simulated data sets. Two large estimates of $\sigma_{\text{Sand.}}$ were excluded for visibility purposes.

To summarise, when dependence in data is moderate to weak all considered standard error estimators perform similarly and quite well. Under strong dependence, however, all estimators are negatively biased, and confidence intervals computed from the estimates have coverage probabilities far from the nominal coverage probability. It is known that both sandwich and block bootstrap estimators might perform poorly when dependence in data is strong, which is confirmed by the results of this study. Furthermore, the standard sandwich estimator might provide inaccurate results if the data sequence is too short. This is supported by results from complementary numerical experiments (not shown) where substantially increasing the size of data sets clearly improved the coverage probabilities. Hence, altogether the results suggest that if the number of observations is large, the standard sandwich estimator can provide acceptable results under strong dependence, while the considered block bootstrap estimators should only be used when temporal dependence in data is moderate to weak.

4.3.2. Results for the logistic model in higher dimensions

In order to assess how the results from Section 4.3.1 generalize to higher dimensions, numerical experiments are conducted with the logistic model in dimensions $d = 5$ and $d = 10$, in the scenarios with temporal dependence $\varphi \in \{0.3, 0.6, 0.9\}$ and multivariate dependence $\alpha = 0.3$. The data generating process is the same as described in Section 4.1 although extended to higher dimensions. Furthermore, the marginal distributions are directly transformed to unit Fréchet to avoid the estimation of a large number of marginal parameters. Thus, only the multivariate logistic dependence parameter α is estimated. The multivariate logistic log-likelihood is fitted efficiently using a representation from Shi (1995), and only the block-maxima approach is considered as the results from the bivariate logistic model were similar for the block-maxima and the POT approaches.

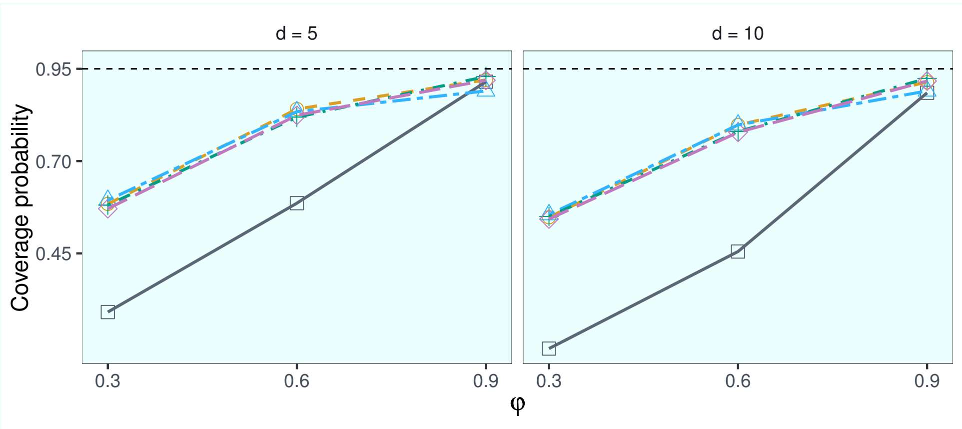

Figure 5: Coverage probabilities of 95% confidence intervals from the multivariate logistic model in dimension $d = 5$ (left) and $d = 10$ (right) under strong ($\alpha = 0.3$) multivariate dependence for the naive (solid, squares), standard sandwich (dashed, circles), blocked sandwich (two-dash, triangles), MBB (dash-dotted, plus) and SB (long-dash, diamonds) estimators, computed from 1000 simulated data sets.

Figure 5 shows coverage probabilities of 95% confidence intervals for the multivariate logistic dependence parameter α computed from 1000 independent data sets. The relation between the estimators is similar to what was observed in the bivariate case. The coverage probabilities are, however, lower and decrease with increasing dimension, due to increasing bias. Thus, the results suggest that even greater care must be taken when modelling extremes from temporally dependent sequences in higher dimensions.

4.4. Results for the bivariate asymmetric logistic and Hüsler–Reiss models

To evaluate if the results from the bivariate logistic model in Section 4.3.1 translates to other models, numerical experiments are conducted with the asymmetric logistic and Hüsler–Reiss models. Data is generated as previously, described in Section 4.1, but the bivariate dependence is imposed with either the asymmetric logistic model in (2.9), or the Hüsler–Reiss model in (2.10). Again, only the block-maxima approach is considered in the scenarios with temporal dependence $\varphi \in \{0.3, 0.6, 0.9\}$.

With the asymmetric logistic model, dependence is governed by the parameter $0 < \beta \leq 1$ and the asymmetry parameters $0 \leq \psi_1, \psi_2 \leq 1$. Numerical experiments are performed with $\beta = 0.3$, $\psi_1 = 0.5$ and $\psi_1 = 0.8$ which corresponds to quite strong bivariate dependence. The Hüsler–Reiss model has a single dependence parameter $r > 0$ which, in the experiments, is set to $r = 1.5$ which corresponds to moderate bivariate dependence. The results are presented in Figure 6, which shows coverage probabilities of 95% confidence intervals for the dependence parameter β of the asymmetric logistic model, and r of the Hüsler–Reiss model, computed from 1000 independent data sets. Corresponding results from the bivariate logistic model with $\alpha = 0.3$ are also shown for comparison.

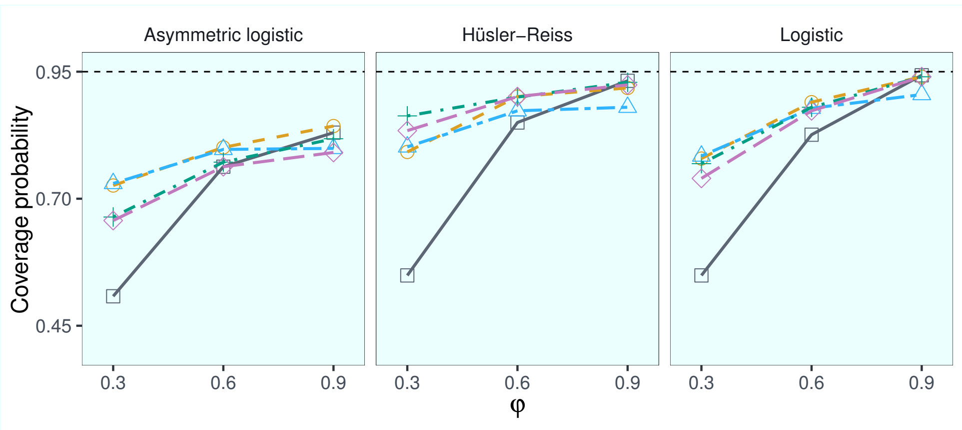

Figure 6: Coverage probabilities of 95% confidence intervals under different levels of temporal dependence for the naive (solid, squares), standard sandwich (dashed, circles), blocked sandwich (two-dash, triangles), MBB (dash-dotted, plus) and SB (long-dash, diamonds) estimators. Values are computed from 1000 simulated data sets from the asymmetric logistic model (left) with parameter $\beta = 0.3$, the Hüsler–Reiss model (middle) with parameter $r = 1.5$, and logistic model (right) with parameter $\alpha = 0.3$.

The relation between the estimator's performance is similar for all models, although the block bootstraps perform somewhat worse than the sandwich estimators for the asymmetric logistic model under strong dependence. It is noteworthy that the coverage probabilities of all estimators for the asymmetric logistic model are far from the nominal coverage probability, even when temporal dependence is weak. A possible explanation is the added complexity from the two additional parameters in the model, which has a total of 9 parameters to be estimated compared to 7 for the bivariate logistic and Hüsler–Reiss models. The estimator's performance for the Hüsler–Reiss model is similar to that of the logistic model and, hence, the conclusions from Section 4.3.1 seem to also hold for the Hüsler–Reiss model.

5. DISCUSSION

The goal of this study was to investigate and compare the performance of four sandwich and block bootstrap standard error estimators when modelling the bivariate dependence between extremes from bivariate stationary sequences. The performance of the estimators was assessed through coverage probabilities of 95% confidence intervals, and by comparing the estimators sampling distributions to a benchmark. Focus was on the bivariate logistic extreme-value model, but the results were generalized both to higher dimensions and to additional models.

In the cases with the logistic and Hüsler–Reiss models, when temporal dependence in data is moderate to weak, all considered estimators perform quite well and provide viable alternatives to account for underestimation. With the asymmetric logistic model, however, all estimators performed poorly under both strong and weak dependence, and extra care should be taken if modelling extremes from temporally dependent sequences with this model. Under strong temporal dependence, all estimators have coverage probabilities far from the nominal coverage probability and are notably biased due to an inability to fully capture the strong dependence in data. Furthermore, for the standard sandwich estimator, the considered data sequences seem too short, at least in the block-maxima case, for the estimator to perform well.

Altogether, the results suggest that when encountered with smaller data sets with strong temporal dependence, none of the considered estimators are preferable. Instead, it may be more successful either to model the dependence explicitly or, if feasible, to decrease the dependence by e.g. declustering. There also exist bootstrapping techniques that might provide more accurate standard errors, such as residual bootstraps where a parametric model is first fitted to data and the residuals resampled, or bootstrap methods tailored for Markov processes. An investigation of additional bootstrapping methods is, however, left for future research.

ACKNOWLEDGMENTS

I would like to thank Jennifer Wadsworth for valuable ideas and discussions during this project. I would also like to thank the anonymous referees for valuable suggestions that improved the paper.

REFERENCES

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2. https://doi.org/10.1214/aop/1176996548.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, UK.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2:339–365. https://doi.org/10.1023/A:1009963131610.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press, New York.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, New York.

Davison, A.C. and Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, 2(1):203–235. https://doi.org/10.1146/annurev-statistics-010814-020133.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–442. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x.

de Haan, L. and Resnick, S.I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, 40:317–337. https://doi.org/10.1007/BF00533086.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1): 1–26. https://doi.org/10.1214/aos/1176344552.

Fawcett, L. and Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics*, 18:173–188. https://doi.org/10.1002/env.810.

Feigin, P. D. and Resnick, S. I. (1999). Pitfalls of fitting autoregressive models for heavy-tailed time series. *Extremes*, 1:391–422. https://doi.org/10.1023/A:1009946321478.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211.

Gumbel, E. J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association*, 56(294):335–349. https://doi.org/10.2307/2282259.

Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525):434–444. https://doi.org/10.1080/01621459.2017.1411813.

Hüsler, J. and Reiss, R. D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286. https://doi.org/10.1016/0167-7152(89)90106-5.

Keef, C., Svensson, C., and Tawn, J. A. (2009). Spatial dependence in extreme river flows and precipitation for Great Britain. *Journal of Hydrology*, 378(3-4):240–252. https://doi.org/10.1016/j.jhydrol.2009.09.026.

Kreiss, J. P. and Lahiri, S. N. (2012). *Bootstrap methods for time series*. In “Handbook of Statistics” (T.S. Rao, S.S. Rao and C.R. Rao, Eds.), Volume 30, Elsevier, 3–26.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241. https://doi.org/10.1214/aos/1176347265.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404. https://doi.org/10.1214/aos/1018031117.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York.

Leadbetter, M. R. (1974). On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 28:289–303.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187. `https://doi.org/10.1093/biomet/83.1.169`.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 59(2):475–499. `https://doi.org/10.1111/1467-9868.00080`.

Liu, R. Y. and Singh, K. (1992). *Moving blocks jackknife and bootstrap capture weak dependence*. In "Exploring the Limits of the Bootstrap" (R. LePage and L. Billard, Eds.), Wiley, 225–248.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Northrop, P. J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes*, 18:585–603. `https://doi.org/10.1007/s10687-015-0221-5`.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131. `https://doi.org/10.1214/aos/1176343003`.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313. `https://doi.org/10.1080/01621459.1994.10476870`.

Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724. `https://doi.org/10.2307/2335152`.

R Core Team (2023). R: A language and environment for statistical computing. *R Foundation For Statistical Computing*, Vienna, Austria. `https://www.R-project.org/`.

Shi, D. (1995). Multivariate extreme value distribution and its Fisher information matrix. *Acta Mathematicae Applicatae Sinica*, 11(4):421–428. `https://doi.org/10.1007/BF02007180`.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, 9(6): 1187–1195. `https://doi.org/10.1214/aos/1176345636`.

Smith, R. L. (1990). Regional estimation from spatially dependent data. *Preprint*.

Stephenson, A. (2002). Evd: Extreme value distributions. *The R Journal*, 2(2):31–32.

Stephenson, A. (2003). Simulating multivariate extreme value distributions of logistic type. *Extremes*, 6:49–59. `https://doi.org/10.1023/A:1026277229992`.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika*, 75(3): 397–415. `https://doi.org/10.1093/biomet/75.3.397`.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25. `https://doi.org/10.2307/1912526`.

Zheng, F., Westra, S., and Sisson, S. A. (2013). Quantifying the dependence between extreme rainfall and storm surge in the coastal zone. *Journal of Hydrology*, 505(15):172–187. `https://doi.org/10.1016/j.jhydrol.2013.09.054`.

Zhou, C. (2010). Dependence structure of risk factors and diversification effects. *Insurance: Mathematics and Economics*, 46(3):531–540. `https://doi.org/10.1016/j.insmatheco.2010.01.010`.

Zscheischler, J. and Seneviratne, S. I. (2017). Dependence of drivers affects risks associated with compound events. *Science Advances*, 3(6):e1700263. `https://doi.org/10.1126/sciadv.1700263`.

REVSTAT – Statistical Journal
Volume 23, Number 2, April 2025, 291–300
https://doi.org/10.57805/revstat.v23i2.567

Some Information Properties of Order Statistics of Skew-normal Distribution

Authors: Parisa Hasanalipour
– Department of Statistics, Ferdowsi University of Mashhad, Iran
p.hasanalipour116919@gmail.com

Mostafa Razmkhah ✉
– Department of Statistics, Ferdowsi University of Mashhad, Iran
razmkhah_m@um.ac.ir

Gholamreza Mohtashami Borzadaran
– Department of Statistics, Ferdowsi University of Mashhad, Iran
grmohtashami@um.ac.ir

Received: January 2022 Revised: June 2023 Accepted: June 2023

Abstract:

- The skew-normal distribution and some of its extensions have been considered in the last two decades in view of distribution theory and the associated properties. However, less attention has been paid to other aspects of this family of distributions. In this paper, we focus on the information properties of this distribution and the distributions of order statistics of a simple random sample from the skew-normal distribution. The Shannon's entropy as well as Kullback–Leibler divergence between the order statistics of two independent skew-normal distributions are studied. Some interesting properties of the information measures of different order statistics are presented.

Keywords:

- *Kullback–Leibler information; order statistic; Shannon's entropy.*

AMS Subject Classification:

- 62B10, 62F10, 62G30.

✉ Corresponding author

1. INTRODUCTION AND PRELIMINARIES

The celebrated normal distribution has been known in all fields of data analysis for centuries. Its popularity has been derived from its analytical simplicity and the associated Central Limit Theorem. There are numerous situations in which the assumption of normality is not validated by the data. So, some families of near-normal distributions have played a crucial role in data analysis. Azzalini (1985) introduced the skew-normal (SN) distribution and studied some of its properties. This class of distributions includes the normal distribution and possesses several properties which coincide or are close to the properties of the normal family. The random variable X is said to have skew-normal distribution, denoted by $X \sim \mathrm{SN}(\lambda)$, if it has the following probability density function (pdf):

$$\phi(x;\lambda) = 2\,\phi(x)\,\Phi(\lambda x), \qquad x \in (-\infty, \infty), \tag{1.1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cumulative distribution function (cdf) of standard normal distribution, respectively. The skewness parameter λ varies on the real line and controls the skewness of the distribution. From (1.1), the cdf of the $\mathrm{SN}(\lambda)$ distribution can be expressed as

$$\Phi(x;\lambda) = \int_{-\infty}^{x} 2\,\phi(t)\,\Phi(\lambda t)\,dt = \Phi(x) - 2\int_{x}^{\infty}\int_{0}^{\lambda t} \phi(u)\,\phi(t)\,du\,dt. \tag{1.2}$$

The standard normal distribution is a special case of the SN distribution, such that SN(0) coincide with the normal distribution. Moreover, as λ tends to infinity, $\phi(x;\lambda)$ tends to the half-normal density. Also, the pdf of SN distribution is a log-concave function. Many extensions of skew-normal distribution have been proposed by different authors, and some inferential aspects of them have been investigated; see, for example, Arellano-Valle *et al.* (2010), Gómez and Salinas (2010), Hasanalipour and Sharafi (2012), Azzalini and Capitanio (2014), Hasanalipour *et al.* (2017), Hasanalipour and Razmkhah (2020, 2022) and Arnold *et al.* (2021).

Let $X_1, ..., X_n$ be a random sample of $\mathrm{SN}(\lambda)$ distribution; moreover, let $X_{1:n} \le \cdots \le X_{n:n}$ denote the corresponding order statistics. Then, for $1 < i < n$, the pdf of $X_{i:n}$ is given by

$$\phi_{i:n}(x;\lambda) = c_{i,n}\,\phi(x;\lambda)\,\Phi^{i-1}(x;\lambda)\,\big[1 - \Phi(x;\lambda)\big]^{n-i}, \tag{1.3}$$

where $c_{i,n} = i\binom{n}{i}$, also, $\phi(x;\lambda)$ and $\Phi(x;\lambda)$ are as defined in (1.1) and (1.2), respectively. For more details about order statistics and their applications, one may refer to David and Nagaraja (2003) and Arnold *et al.* (2008).

The mathematical theory of communication introduced by Shannon (1948) describes logarithmic measures of information and has stimulated a tremendous amount of study in engineering fields. It is a branch of applied probability and statistics relevant to statistical inference and therefore, it should be of essential interest to statisticians. The Shannon's Entropy (SE) of a random variable X with pdf $f(\cdot)$ is given by:

$$H(X) = -\int_{-\infty}^{\infty} f(x)\,\log f(x)\,dx. \tag{1.4}$$

Entropy is a measure of average uncertainty in a random variable, and also it is considered as a measure of the randomness of a probabilistic system.

The KL divergence measuring the degree of divergence between two probability distributions, is another information index considered in this paper. By assuming X and Y have pdfs $f(\cdot)$ and $g(\cdot)$, respectively, the KL divergence of $f(\cdot)$ with respect to $g(\cdot)$ is defined as

$$K(X\,|\,Y) = \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx. \tag{1.5}$$

Note that $K(X\,|\,Y)$ becomes zero when $f(x) = g(x)$, almost everywhere. Several authors have studied the properties of the information measures of ordered data in the fields of estimation, reliability analysis, quality control, goodness of fit tests, characterization of probability distributions, and many other problems. See, for example, Ebrahimi *et al.* (2004, 1999, 2010), Zarezadeh and Asadi (2010), Arellano-Valle *et al.* (2013, 2017), Kayal and Kumar (2017), Ardakani *et al.* (2018), and Jose and Abdul Sathar (2021).

In this paper, we study some information properties of the order statistics of simple random samples from the SN distribution. First, Shannon's entropy of the SN distribution is studied. It is proved that this measure is symmetric to the skewness parameter, such that the maximum entropy occurs when the skewness parameter is zero or equivalently when the distribution is normal. The results are extended to the distributions of order statistics of a simple random sample from the SN distribution. The relation between the entropies of lower and upper order statistics from different SN distributions with opposite signs skewness parameters is stated. The average entropy of distributions of order statistics and data distribution is also compared. Then, the Kullback–Leibler (KL) divergence between the distribution of different order statistics is investigated, and some interesting results are obtained, theoretically or numerically.

The rest of this paper is organized as follows. The Shannon's entropy of SN distribution and the distribution of order statistics are investigated in Section 2. The KL divergence and some results are studied in Section 3. Eventually, some conclusions are stated in Section 4.

2. THE ENTROPY OF THE SKEW-NORMAL DISTRIBUTION

Let X have SN(λ) distribution. Then, using (1.4), the Shannon's entropy of X is given by

$$H(X;\lambda) = -\int_{-\infty}^{\infty} \phi(x;\lambda) \log\big(\phi(x;\lambda)\big)\, dx. \tag{2.1}$$

Theorem 2.1. *The maximum entropy model in the skewed normal family is the normal distribution.*

Proof: Assuming $X \sim \mathrm{SN}(\lambda)$, the proof includes three parts:

- **(i)** $H(X;\lambda) = H(X;-\lambda)$;
- **(ii)** $\lim_{\lambda\to\pm\infty} H(X;\lambda) = \frac{1}{2} + \log\sqrt{\frac{\pi}{2}}$;
- **(iii)** $H(X;\lambda)$ is increasing for $\lambda < 0$ and it is decreasing for $\lambda > 0$.

A well-known property of SN distribution is that if $X \sim \phi(x;\lambda)$, then $-X = Y \sim \phi(y;-\lambda)$, hence, $H(X;\lambda) = H(X;-\lambda)$ and the proof of part (i) is complete. That is, the entropy of SN distribution does not depend on the sign of the skewness parameter.

To prove part (ii), note that equation (6) of Arellano-Valle *et al.* (2017) with $\Phi(\lambda x)$ gives the following relationship between the entropies of $\phi(x;\lambda)$ and $\phi(x)$:

$$H(X;\lambda) - H(X;\lambda=0) = -\log 2 - E\big[\log \Phi(\lambda X)\big],$$

where $H(X;\lambda=0) = \frac{1}{2}\log(2\pi e)$ is the Shannon entropy of standard normal distribution. Hence,

$$H(X;\lambda) = \frac{1}{2} + \log\sqrt{\frac{\pi}{2}} - E\big[\log \Phi(\lambda X)\big]. \tag{2.2}$$

On the other hand,

$$E\big[\log \Phi(\lambda X)\big] = \int_{-\infty}^{0} \phi(x;\lambda)\log \Phi(\lambda x)\,dx + \int_{0}^{\infty} \phi(x;\lambda)\log \Phi(\lambda x)\,dx. \tag{2.3}$$

Note that for $x > 0$ as $\lambda \to \infty$, we get $\Phi(\lambda x) \to 1$, hence, the second term in (2.3) tends to zero when $\lambda \to \infty$. Moreover, for $x < 0$ as $\lambda \to \infty$, we have $\Phi(\lambda x) \to 0$ and by using L'Hopital's rule $\Phi(\lambda x)\log\Phi(\lambda x) \to 0$, hence, the first term in (2.3) also tends to zero. Therefore, $E\big[\log \Phi(\lambda X)\big] \to 0$ as $\lambda \to \infty$. Further, according to part (i), it is deduced that $\lim_{\lambda\to-\infty} H(X;\lambda) = \lim_{\lambda\to\infty} H(X;\lambda)$; hence, part (ii) is also proved.

Finally, to prove part (iii), using (2.2), we get

$$\begin{aligned}\frac{\partial}{\partial\lambda} H(X;\lambda) &= -\int_{-\infty}^{\infty} 2\,x\,\phi(x)\,\phi(\lambda x)\log\Phi(\lambda x)\,dx - \int_{-\infty}^{\infty} 2\,x\,\phi(x)\,\phi(\lambda x)\,dx \\ &= -\sqrt{\frac{2}{\pi}}\,\frac{1}{\sqrt{1+\lambda^2}}\,E\big(Y\log\Phi(\lambda Y)\big),\end{aligned}$$

where $Y \sim N\Big(0, \frac{1}{1+\lambda^2}\Big)$. By Stein's lemma, if $Y \sim N(0,\sigma^2)$, then $E\big[Yg(Y)\big] = \sigma^2 E\big[g'(Y)\big]$, provided that $g(\cdot)$ is a function for which both expectations $E\big[Yg(Y)\big]$ and $E\big[g'(Y)\big]$ exist. Therefore,

$$\frac{\partial}{\partial\lambda} H(X;\lambda) = -\sqrt{\frac{2}{\pi}}\,\frac{\lambda}{(1+\lambda^2)^{\frac{3}{2}}}\,E\left[\frac{\phi(\lambda Y)}{\Phi(\lambda Y)}\right].$$

On the other hand, $E\left[\frac{\phi(\lambda Y)}{\Phi(\lambda Y)}\right] > 0$ for all λ. Hence, $\frac{\partial}{\partial\lambda} H(X;\lambda)$ is negative (positive) for $\lambda > (<)\ 0$. Therefore, the entropy of SN(λ) distribution is a unimodal symmetric function of λ that maximizes at $\lambda = 0$. This completes the proof. □

Corollary 2.1. *The skewness parameter orders the entropy and variance in the skewed normal family similarly, in that they both increase for $\lambda < 0$ and decrease for $\lambda > 0$. Such ordering behavior which holds for parts of supports of parameter was studied by Ebrahimi et al. (1999) for the beta family.*

Remark 2.1. It is clear that when $\lambda \to \infty$ (or $\lambda \to -\infty$), the $\phi(x;\lambda)$ tends to the positive (or negative) half-normal distribution with pdf $2\,\phi(x)$, for $x > 0$ (or $x < 0$) (Azzalini, 1985). It is not difficult to show that $H(X^+) = H(X^-) = \frac{1}{2} + \log\sqrt{\frac{\pi}{2}}$, where X^+ and X^- stand for positive and negative half-normal distributions, respectively. Comparing to part (ii) of the proof of Theorem 2.1, it can be concluded that limit of entropy is equal to the entropy of a limiting distribution; precisely, $\lim_{\lambda\to\infty} H(X;\lambda) = H(X^+)$ and $\lim_{\lambda\to-\infty} H(X;\lambda) = H(X^-)$. On the other hand, from part (iii) of the proof of Theorem 2.1, it is concluded that the entropy of SN(λ) distribution decreases to entropy of the limiting positive (or negative) half-normal case for $\lambda > 0$ (or $\lambda < 0$). Such information properties were investigated in details by Ardakani *et al.* (2018) for symmetric families that include the normal distribution as special cases.

Using (2.1) and employing the numerical computations, the behavior of $H(X;\lambda)$ with respect to λ is shown in Figure 1. This figure confirms that the maximum entropy occurs for the case of $\lambda = 0$, which coincides with the case of standard normal distribution.

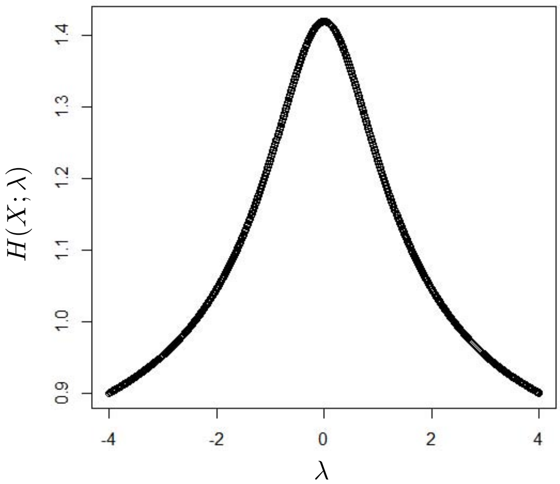

Figure 1: Plot of $H(X;\lambda)$ with respect to λ.

Now, we focus on entropy of order statistics of SN distribution. Using (1.1) and (1.3) and doing some algebraic calculations, one can show that the entropy of the i-th order statistic of the SN distribution is

$$\begin{aligned} H(X_{i:n};\lambda) &= -\log c_{i,n} - \log\frac{2}{\sqrt{2\pi}} \\ &\quad - E\Big(\log\Phi\big(\lambda\Phi^{-1}(W;\lambda)\big)\Big) + \frac{1}{2}\,E\Big(\big(\Phi^{-1}(W;\lambda)\big)^2\Big) \\ &\quad - i\,(i-1)\Big(\psi(i)-\psi(n+1)\Big) - i\,(n-i)\Big(\psi(n-i+1)-\psi(n+1)\Big), \end{aligned} \tag{2.4}$$

where the random variable W has the beta distribution with parameters i and $(n-i+1)$, denoted by $W \sim \mathrm{Beta}(i, n-i+1)$.

Remark 2.2. Similar to part (i) of the proof of Theorem 2.1, it is easy to deduce that entropy of the i-th order statistic from a SN distribution equals the entropy of the $(n-i+1)$-th order statistic from a different SN distribution with opposite sign skewness parameter. That is, for given i and λ, we have

$$H(X_{i:n};\lambda) = H(X_{n-i+1:n};-\lambda). \tag{2.5}$$

From (2.4) and by using numerical computations, the behavior of $H(X_{i:n};\lambda)$ with respect to λ is shown in Figure 2 for $n=5$ and $i=1,\ldots,5$; for other values of n and i, similar figures are obtained which are omitted due to similarity. From this figure and similar ones, the following results are deduced:

- For given n and a fixed i, the entropy $H(X_{i:n};\lambda)$ is a symmetric increasing-decreasing function of λ, such that the maximizer decreases when i goes from 1 up to n. For example, the order statistics $X_{1:5}, X_{2:5}, X_{3:5}, X_{4:5}, X_{5:5}$ get their maximum entropy at $\lambda = 1.4, 0.6, 0, -0.6, -1.4$, respectively.
- For given n, the entropies $H(X_{i:n};\lambda)$ and $H(X_{n-i+1:n};\lambda)$ have the same maximum entropy that confirms the relation (2.5).
- For given n, the maximum entropy of $X_{i:n}$ (or equivalently $X_{n-i+1:n}$) is decreasing in i for $i \le \frac{n}{2}$.

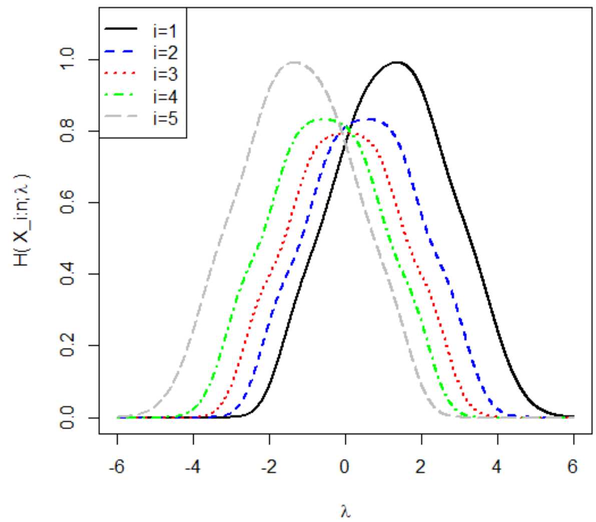

Figure 2: Plot of $H(X_{i:5};\lambda)$ with respect to λ for $i=1,\ldots,5$.

For more investigation about the entropy of order statistics of a simple random sample of size n from the SN(λ) distribution, let us define the average uncertainty of these statistics as follows:

$$\bar{H}_n(X;\lambda) = \frac{1}{n}\sum_{i=1}^{n} H(X_{i:n};\lambda). \tag{2.6}$$

This measure can be used to compare the average entropy of distributions of order statistics of a simple random sample of size n with the entropy of a single observation or data distribution.

Using (2.6), the values of $\bar{H}_n(X;\lambda)$ are calculated for some positive values of λ and some choices of n. To compare them with entropy of the SN distribution, numerical values of $H(X;\lambda)$ are also computed. The results are presented in Table 1. From this table, it is deduced that:

- Since, according to Remark 2.2, we get $\bar{H}_n(X;\lambda) = \bar{H}_n(X;-\lambda)$, it is concluded that both $\bar{H}_n(X;\lambda)$ and $H(X;\lambda)$ are symmetric increasing-decreasing functions in λ for $-\infty < \lambda < \infty$, such that $\max_\lambda \bar{H}_n(X;\lambda) = \bar{H}_n(X;0)$ and $\max_\lambda H(X;\lambda) = H(X;0)$.
- $H(X;\lambda) > \bar{H}_n(X;\lambda)$, for all values of n and λ.

Table 1: Values of $\bar{H}_n(X;\lambda)$ for some choices of λ and n.

n	λ								
	0.5	1	2	3	4	5	6	7	7.5
2	1.1576	1.0326	0.9391	0.8462	0.7570	0.6128	0.2274	0.0718	0.0265
3	1.0211	0.9961	0.8036	0.7208	0.6705	0.5446	0.2189	0.0636	0.0241
5	0.8384	0.8034	0.7123	0.6395	0.5774	0.4483	0.1941	0.0537	0.0211
10	0.5486	0.5136	0.4243	0.4114	0.4087	0.3033	0.1397	0.0411	0.0171
$H(X;\lambda)$	1.3507	1.2257	1.0456	0.9528	0.9001	0.7094	0.3461	0.0862	0.0307

3. KULLBACK–LEIBLER DIVERGENCE

This section discusses some information between distributions of the i-th and the j-th order statistics from SN distributions with different skewness parameters. Ebrahimi *et al.* (2004) showed that the discrimination information between a given order statistic and data distribution of the same population is distribution free. They also proved that the discrimination information among different order statistics of the same distribution is distribution free. The question arises here is that what relationship exists between the order statistics of two different distributions. To study this important subject, suppose that $X_1, ..., X_n$ and $Y_1, ..., Y_m$ are simple random samples from different SN distributions with parameters λ_1 and λ_2, respectively. Also, let $X_{i:n}$ be the i-th order statistic from X, and $Y_{j:m}$ be the j-th order statistic from Y. Using (1.5), the discrimination information between the distribution of $X_{i:n}$ and $Y_{j:m}$ is obtained as

$$\begin{aligned} K_{\lambda_1,\lambda_2}(X_{i:n} \,|\, Y_{j:m}) &= \int_{-\infty}^{\infty} \phi_{i:n}(x;\lambda_1) \log \frac{\phi_{i:n}(x;\lambda_1)}{\phi_{j:m}(x;\lambda_2)}\, dx \\ &= -H(X_{i:n};\lambda_1) - \log c_{j,m} - E\Big(\log \phi\big(\Phi^{-1}(W,\lambda_1);\lambda_2\big)\Big) \\ &\quad - (j-1)\; E\Big(\log \Phi\big(\Phi^{-1}(W,\lambda_1);\lambda_2\big)\Big) \\ &\quad - (m-j)\; E\bigg(\log\Big(1-\Phi\big(\Phi^{-1}(W,\lambda_1);\lambda_2\big)\Big)\bigg), \end{aligned} \tag{3.1}$$

where $W \sim \text{Beta}(i, n-i+1)$ and $\Phi^{-1}(\cdot,\lambda)$ stands for the inverse function of $\Phi(\cdot;\lambda)$.

Remark 3.1. When $\lambda_1 = \lambda_2$, that is, in the situation in which both samples come from the same distribution, it is trivial that in the special case of $n = m$, the discrimination information between the distribution of $X_{i:n}$ and $Y_{i:n}$ is zero, i.e., $K_{\lambda_1,\lambda_1}(X_{i:n}, Y_{i:n}) = 0$, for $1 \le i \le n$. But, when $\lambda_1 \ne \lambda_2$, the KL information between the distribution of the order statistics of different distributions is positive.

Using (3.1), the KL information between the distribution of sample maxima may be obtained as

$$\begin{aligned} K_{\lambda_1,\lambda_2}(X_{n:n}\,|\,Y_{n:n}) \;=\; & (n-1)\,E(\log W) + E\Big(\log\phi\big(\Phi^{-1}(W,\lambda_1);\lambda_1\big)\Big) \\ & - E\Big(\log\phi\big(\Phi^{-1}(W,\lambda_1);\lambda_2\big)\Big) - (n-1)\,E\Big(\log\Phi\big(\Phi^{-1}(W,\lambda_1);\lambda_2\big)\Big). \end{aligned}$$

Similarly, the KL information between the distribution of sample minima is given by

$$\begin{aligned} K_{\lambda_1,\lambda_2}(X_{1:n}\,|\,Y_{1:n}) \;=\; & (n-1)\,E\big(\log(1-W)\big) + E\Big(\log\phi\big(\Phi^{-1}(W,\lambda_1);\lambda_1\big)\Big) \\ & - E\Big(\log\phi\big(\Phi^{-1}(W,\lambda_1);\lambda_2\big)\Big) - (n-1)\,E\Big(\log\Big(1-\Phi\big(\Phi^{-1}(W,\lambda_1);\lambda_2\big)\Big)\Big). \end{aligned}$$

Remark 3.2. Let $X_1, ..., X_n$ and $Y_1, ..., Y_m$ be two independent random samples from SN(λ_1) and SN(λ_2) distributions, respectively. Then, using the fact that if $X \sim \phi(x;\lambda)$, then $-X = Y \sim \phi(y;-\lambda)$, it can be simply shown that the KL divergence between two lower sample quantiles from the SN distribution with given skewness parameter equals that of upper sample quantiles from the SN distribution with opposite sign skewness parameter. More precisely, for each $1 \le i \le n$ and $1 \le j \le m$, we get

$$\begin{aligned} K_{\lambda_1,\lambda_2}(X_{i:n}\,|\,Y_{j:m}) &= K_{-\lambda_1,-\lambda_2}(X_{n-i+1:n}\,|\,Y_{m-j+1:m}), \\ K_{\lambda_1,-\lambda_2}(X_{i:n}\,|\,Y_{j:m}) &= K_{-\lambda_1,\lambda_2}(X_{n-i+1:n}\,|\,Y_{m-j+1:m}). \end{aligned}$$

The values of $K_{\lambda_1,\lambda_2}(X_{i:n}\,|\,Y_{j:m})$ may be numerically obtained using (3.1). Figure 3 shows the behavior of $K_{\lambda_1,\lambda_2}(X_{i:n}\,|\,Y_{j:m})$ to λ_2 when $\lambda_1 = 1$, $n = 6$ and $m = 3$. In fact, the KL between the pdfs of $X_{1:6}$ (or $X_{6:6}$) and $Y_{j:3}$ for $j = 1, 2, 3$ are plotted in the left (right) hand side of this figure. Analogous results may be obtained for other values of n and m.

From Figure 3, the following results are deduced:

- For given λ_1, the KL divergence is a decreasing-increasing function of λ_2.
- From the left plot, it is observed that for given λ_1, the KL divergence between the minimum of X sample and the minimum of Y sample tends to zero when λ_2 tends to λ_1. Though, the KL of $X_{1:n}$ and $Y_{j:m}$, for $j > 1$, becomes zero for smaller λ_2.
- From the right plot, it is observed that for given λ_1, the KL between $X_{n:n}$ and $Y_{m:m}$ becomes zero when λ_2 tends to λ_1, however, the KL of $X_{n:n}$ and $Y_{j:m}$, for $j < m$, is zero for larger λ_2.
- The above results mean that for given λ_1, there exists a value such λ_2^j that the pdf of a sample quantile $X_{i:n}$ of SN(λ_1) distribution closes to the pdf of the sample quantile $Y_{j:m}$ of SN(λ_2^j) distribution, such that λ_2^j decreases with respect to j.

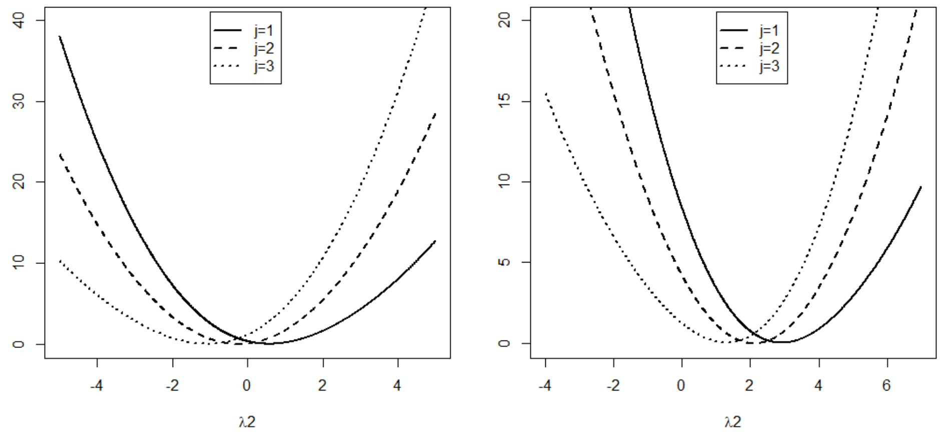

Figure 3: Plots of $K_{1,\lambda_2}(X_{i:6} \,|\, Y_{j:3})$ with respect to λ_2 for $i = 1$ (the left plot) and $i = 6$ (the right plot).

4. CONCLUSIONS

In this paper, some information properties of SN distribution and its order statistics were studied. Shannon entropy and KL criteria were investigated, and some theoretical and numerical results were obtained. The behavior of entropy of the SN(λ) distribution to λ was studied, and it was shown that the maximum entropy model in the skewed normal family is the normal distribution. Moreover, it was deduced that limit entropy equals the entropy of the limiting distribution when λ tends to infinity. It was also shown that for fixed sample size, the entropy of a given order statistic is symmetric an increasing-decreasing function of λ in which the maximizer of entropy of $X_{i:n}$ decreases when i goes from 1 up to n; further, the maximum entropy of sample quantiles are decreasing to the maximum entropy of sample median. The maximum entropy plays an important role in choosing the best order statistics in specifying the outliers or determining the control limits in statistical quality control. Also, it is possible to compare the uncertainty of the distribution of k-out-of-n systems for different values of k or n. Some relations were also obtained for the KL divergence between distributions of the order statistics of two independent SN(λ_1) and SN(λ_2) distributions with respect to the variations of skewness parameters and the ranks of order statistics for given sample sizes. It was shown that for given λ_1, the KL divergence is a decreasing-increasing function of λ_2; moreover, for any given λ_1, there exists a value λ_2^j such that the pdf of the sample quantile $X_{i:n}$ of SN(λ_1) distribution closes to the pdf of the sample quantile $Y_{j:m}$ of SN(λ_2^j) distribution, such that λ_2^j decreases for j.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referees and the associate editor for their useful comments and constructive criticisms on the original version of this manuscript which led to this considerably improved version. The research of the corresponding author is supported by a grant from Ferdowsi University of Mashhad (N. 2/55138).

REFERENCES

Ardakani, O. M., Ebrahimi, N., and Soofi, E. S. (2018). Ranking forecasts by stochastic error distance, information, and reliability measures. *International Statistical Review*, 83:442–468.

Arellano-Valle, R. B., Contreras-Reyes, J. E., and Genton, M. G. (2013). Shannon entropy and mutual information for multivariate skew-elliptical distributions. *Scandinavian Journal of Statistics*, 40: 42–62.

Arellano-Valle, R. B., Cortes, M. A., and Gomez, H. W. (2010). An extension of the epsilon-skew-normal distribution. *Communications in Statistics – Theory and Methods*, 39:912–922.

Arellano-Valle, R. B., Contreras-Reyes, J. E., and Stehlk, M. (2017). Generalized skew-normal negentropy and its application to fish condition factor time series. *Entropy*, 19:528–546.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (2008). *A First Course in Order Statistics*. SIAM, Philadelphia.

Arnold, B. C., Gallardo, D. I., and Gómez, H. W. (2021). A new bivariate Birnbaum-Saunders type distribution based on the skew generalized normal model. *REVSTAT – Statistical Journal*, 21:1–20.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178.

Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. Cambridge University Press.

David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*. John Wiley, New York.

Ebrahimi, N., Soofi, E. S., and Zahedi, H. (2004). Information properties of order statistics and spacing. *IEEE Transactions on Information Theory*, 50:177–183.

Ebrahimi, N., Maasoumi, E., and Soofi, E.S. (1999). Ordering univariate distributions by entropy and variance. *Journal of Econometrics*, 90:317–336.

Ebrahimi, N., Soofi, E.S., and Soyer, R. (2010). Information measures in perspective. *International Statistical Review*, 78:383–412.

Gómez, H. W. and Salinas, H. S. (2010). Information matrix for generalized skew-normal distributions. *Proyecciones (Antofagasta)*, 29:83–92.

Hasanalipour, P. and Razmkhah, M. (2020). Testing skew-symmetry based on extreme ranked set sampling. *Statistical Papers*, 62:2311–2332.

Hasanalipour, P. and Razmkhah, M. (2022). Inference on skew-normal distribution based on Fisher information in order statistics. *Communications in Statistics – Simulation and Computation*, 51: 1525–1541.

Hasanalipour, P. and Sharafi, M. (2012). A new generalized Balakrishnan skew-normal distribution. *Statistical Papers*, 53:219–228.

Hasanalipour, P., Hassani Oskouei, L., Jabbari Khamnei, H., and Khorshidian, K. (2017). The Beta skew-generalized normal distribution. *Communications in Statistics – Theory and Methods*, 46:2270–2276.

Jose, J. and Abdul Sathar, E. I. (2021). Renyi entropy of k-records: properties and applications. *REVSTAT – Statistical Journal*, 20:481–500.

Kayal, S. and Kumar, S. (2017). Estimating Renyi entropy of several exponential distributions under an asymmetric loss function. *REVSTAT – Statistical Journal*, 15:501–522.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Zarezadeh, S. and Asadi, M. (2010). Results on residual Renyi entropy of order statistics and record values. *Information Sciences*, 180:4195–4206.

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.

 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 — *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 email-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, in .pdf format;
- 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, 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.